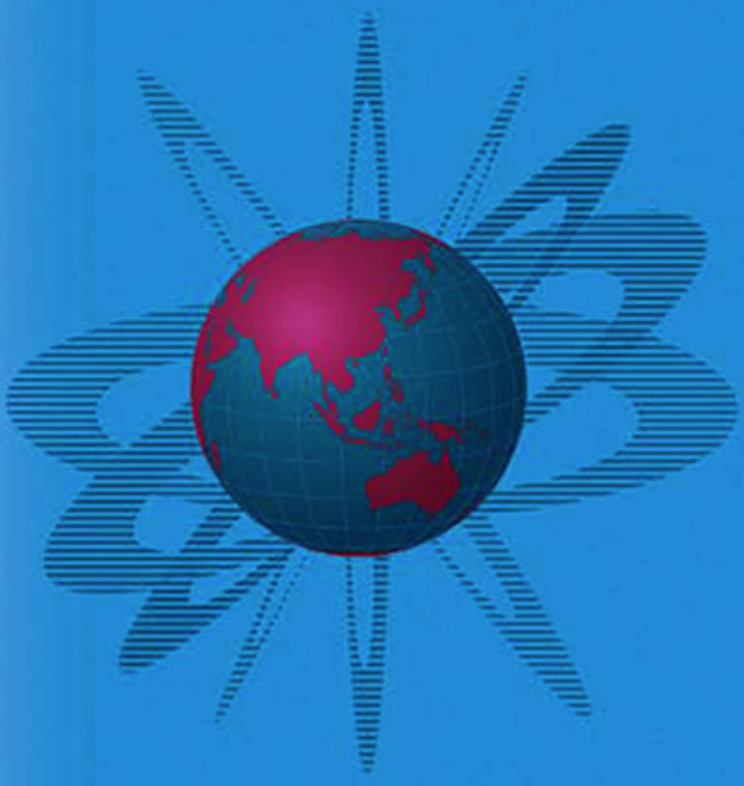
Handbooks in Information Systems

Volume 2

National Security



Edited by:
H. Chen, T.S. Raghu,
R. Ramesh, A. Vinze and D. Zeng

Series Editor: Andrew B. Whinston

HANDBOOKS IN INFORMATION SYSTEMS
VOLUME 2

Handbooks in Information Systems

Editor

Andrew B. Whinston

Volume 2

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford – Paris – San Diego
San Francisco – Singapore – Sydney – Tokyo

National Security

Edited by

H. Chen
University of Arizona

T.S. Raghu
Arizona State University

R. Ramesh
SUNY at Buffalo

A. Vinze
Arizona State University

D. Zeng
University of Arizona

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford – Paris – San Diego
San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

First edition 2007

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-51996-2 (this volume)
ISSN: 1574-0145 (series)

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Contents

This page intentionally left blank

Preface

In the post 9/11 world, security issues are viewed with a new sense of urgency and criticality. National and international security as a domain of study has assumed significant importance in recent years. Given the size and scope of this topic, it is important that a multi-faceted approach be adopted. As such, in this book we have assembled insights from a representative sample of academicians and practitioners, and addressed this topic from a variety of perspectives ranging from technologies, economics and social studies, organizational and group behavior, and policy making. Frameworks, methods, and systems approaches from the fields of information and communication technology and information systems play an increasingly important role in this emerging field of security studies. Technological advances address challenges in information collection, sharing, surveillance and analysis, infrastructure protection, and related information presentation and decision aiding. While technology provides critical drivers, organizational advances are also much needed to foster collaborative arrangements between federal, state, and local agencies as well as the private sector. Policy considerations and their implications in technology development and adoption need to be studied carefully to balance the needs of security and protection of citizens' rights.

While the literature on security studies is expanding quickly, much of it is fragmented and often narrowly focused within certain specific domains and approaches. This edited volume is intended to address in a comprehensive and integrated manner three major areas of national and international security research from an information systems centric perspective: legal and policy frameworks; intelligence and security informatics; and emergency preparedness and infrastructure protection. The discussions are replete with real-world case studies and examples that present the concepts using an integrated, action-oriented and theory-based approach to validate the frameworks presented and provide specific insights on the technical approaches and organizational issues under investigation.

Intended audience and use

The intended audience for this book includes:

- Graduate and advanced undergraduate level students in Information Sciences, Information Systems, Computer Science, Systems Engineering, Social Studies, and Public Policy.

- Researchers engaged in national security related research from a wide range of perspectives including but not limited to informatics, decision sciences, organizational behavior and social studies, and public administration.
- Public officials in all security-related areas, such as law enforcement, public health, anti-terrorism, and policy making.
- Private sector practitioners engaged in ongoing relationships with federal, state, and local agencies on projects related to national security.

This book is intended to be used as both textbook and comprehensive research handbook. The contributors to this edited volume are renowned experts in their respective fields. Most of the chapters contained in this book provide an updated comprehensive survey of the related field and also specific findings from cutting-edge innovative research. To facilitate its adoption as a textbook, all the authors have included specific discussion questions that could be utilized to stimulate discussions and potentially further research in the area.

An overview of the book

The chapters in the book are divided into three main sections—Legal and Policy Frameworks, Intelligence and Security Informatics, and Emergency Preparedness and Infrastructure Protection. The chapters addressing legal and policy frameworks are intended to address broadened information sharing practices and policies among government agencies and between government agencies and the private sector. Such information sharing naturally raises concerns of privacy, confidentiality, and trust. The chapters in this section discuss the significant issues in addressing security-related concerns through well-articulated legal and policy frameworks. The chapters in the section on intelligence and security informatics address the development of use of advanced information technologies, computer science, and algorithms for national/international and homeland security related applications. The discussion here provides a good sample of various types of intelligence and security informatics techniques and application contexts. The section on emergency preparedness and infrastructure protection explores advanced research solutions, practices, and solutions to emergency preparedness and response problems. Governmental agencies have critical responsibilities for preparing citizens and communities for emergencies through increased surveillance, monitoring, and protective measures. Protection of critical infrastructure including the information and computer network infrastructure is also a pressing matter. The chapters in this section present technical frameworks and carefully chosen case studies to illustrate challenges and propose integrated solutions.

Chapter summaries

Legal and policy frameworks

The first chapter in this section by Brenner and Clarke (''Should Commercial Misuse of Private Data be a Crime?'') tackles the all-important issue of privacy from a legal and policy perspective. The myriad set of activities over the Internet and other communications media often leaves consumers vulnerable to the breach of privacy. Yet, such information can be valuable to law enforcement purposes as well as commercial purposes. With this orientation, the chapter proposes a relation-based privacy sharing mechanism that allows for criminal and civil sanctions to be imposed for breaches of privacy. The authors define relationship types and associated expectations of privacy in this chapter. In the context of recent disclosures and controversies over sharing of telephone conversations with NSA and the controversy surrounding sharing of data stored by Internet search engine companies with the FBI, this chapter makes a pertinent contribution to our understanding of data privacy in this century.

The second chapter in this section by Venkatraman, Rao, and Dewitt (''National Information Technology (IT) Security Policies: An Overview of Issues'') analyzes the security policies across multiple countries, and identifies commonalities and uniqueness across these policies. In the wake of increased awareness of national security implications of IT infrastructure, several countries have revamped their IT policies considerably. The unique contribution of this chapter lies in analyzing the possible policy obstacles and facilitators to information sharing across national boundaries. The authors make specific recommendations, taking into consideration the economic, cultural, and structural differences across countries, in formulating international security policies.

The third chapter by Huang, Behara, and Hu (''Economics of Information Security Investment'') summarizes economic approaches for analyzing investment decisions in securing inter-connected IT systems. The authors differentiate between different attack scenarios, and discuss applicable economic theories for investment analysis. The chapter reviews applicability of expected utility theory, game theory, and insurance-based approaches in specific security breach scenarios. Free-riding is a common problem associated with inter-connected multi-organizational IT systems. In addition to technical challenges, lack of resources make achievement of ''perfect security'' an impossible proposition. Given this, optimal investment and policy decisions will have to be made on sound economic analyses of specific problem contexts. The chapter makes valuable contribution to our understanding of the economic issues pertaining to security investments.

Intelligence and security informatics

The first chapter in this section by Razdan, Farin, Soo-Bae, and Chaudhari (''State of 3D Face Biometrics for Homeland Security Applications'') presents a detailed description of the 3D biometrics research program of PRISM group at Arizona State University (ASU). Biometrics research is extremely important in bolstering authentication and recognition functions performed by homeland security personnel. The PRISM group at ASU has focused on facial biometrics. The chapter provides an extensive overview of the state of the art in this area before providing details of the 3D face recognition techniques being developed by their group. There are still challenges to be overcome in 3D face biometrics—from both computational and operational perspectives. However, the technology holds promise for future applications in homeland security contexts.

The second chapter in this section by Went (''The Necessity of Fuzzy Logic for Identity Matching'') reviews the practical challenges of dealing with identity data matching on a large scale. At many immigration checkpoints, identity information presented by individuals has to be matched against large databases (with potentially millions of records). This presents operational challenges to personnel at the physical checkpoint; this is important since technologies that use biometric and other identification methods need to be able to do accurate identity matching within a short amount of time. The chapter presents a fuzzy algorithmic approach to do efficient search in large data stores. An interesting perspective of the chapter is that traditional database queries (i.e., exact match for the criteria selected) are typically ineffective in identity matching problems—what one needs is an approximate match that has low false acceptance and rejection rates. The author prescribes a multi-pronged approach to identity matching that can enhance the efficiency of database queries in retrieving relevant records.

The third chapter in this section by Berndt, Hevner, and Griffiths (''Managing Real-Time Bioterrorism Surveillance Data'') discusses important real-time data management and analysis challenges prevalent in many security-related applications, which involve processing information from a distributed array of sensors and information sources for surveillance purposes. The authors identify the technical challenges to information systems development in these applications with the specific emphasis on bioterrorism surveillance. A summary of related technological approaches and a new framework based on real-time data warehousing, which has been validated through a retrospective case study, is also presented. Another interesting contribution of this chapter is a comparative analysis between fully automated and human-in-the-loop pattern recognition for surveillance via a case study using Florida wildfire data.

The fourth chapter by Zeng, Chen, and Chang (''Spatio-Temporal Data Analysis in Security Informatics'') investigates spatio-temporal data

analysis models and related computational methods. As spatial and temporal coordinates of events play a central role in support of various surveillance and decision-aiding functions in a wide range of security applications, spatio-temporal data analysis has emerged as an important research area in intelligent and security informatics. This chapter provides an overview of this active research area. It focuses on a specific type of spatio-temporal data analysis method concerning identification of unusual cluster of events, and presents a new approach based on support vector machines, which have some very useful computational properties. The authors also discuss two security informatics case studies: crime analysis and biosurveillance.

The fifth chapter by Burgoon, Jensen, Kruse, Meservy, and Nunamaker (''Deception and Intention Detection'') studies an important class of application unique to security applications: how to develop an automated or semi-automated framework to detect deceptive intention or behavior? This paper is based on some of the most recent research results from a large ongoing research project conducted at the University of Arizona. After a short but comprehensive review of the existing work in deception and intention detection, the authors focus on a particular type of intention detection in screenings, where the detection method must be unobtrusive without the cooperation or even knowledge of the individuals who are being scrutinized. They report a preliminary study of a newly developed approach that makes use of data spanning video, audio, and textual modalities.

The last two chapters in this section are about groups. The sixth chapter by Baumes, Goldberg, Magdon-Ismail, and Wallace (''Identification of Hidden Groups in Communications'') reports an emerging line of research concerned with discovering groups that communicate in a large communication network, yet try to hide the nature of their communication patterns including their group membership. This study has obvious applications in intelligence analysis and other security applications. In this chapter, the authors present specific motivations behind hidden group identification, and report analytical models and related algorithms.

The seventh chapter by Reid, Chen, and Xu (''Social Network Analysis for Terrorism Research'') studies terrorist/extremist groups using advanced social network analysis and visualization techniques. Such techniques have gained wide use in recent years in a spectrum of science, engineering, management, and social studies including intelligence and security informatics. This chapter presents two case studies in real-world security informatics applications. In the first case study, various kinds of relations between and among various types of actors, both individuals and groups, of an international terrorist network, are examined. In the second case study, the authors analyze how terrorist groups' websites relate to each other and form patterns (e.g., clusters), generating useful insights about the relations among terrorist groups.

Emergency preparedness and infrastructure protection

The first three chapters in this section address the public health preparedness issues. The first chapter by Weisbuch (''Disaster Response and the Local Public Health Department'') emphasizes that local actions are paramount to emergency preparedness in public health contexts. The chapter describes three separate public health incidents and outlines how local public health department in a large city became an integral part of emergency preparedness team, and developed collaborative partnerships with other non-health emergency management entities. As the chapter demonstrates, while local public health departments will play a central role in case of bioterrorism incidents, it will still play a key role in all other emergency preparedness and response situations.

The second chapter by Burton and Ipe (''Challenges of Bioterrorism Preparedness for Organizational Processes and Resources'') takes a process-oriented view to emergency preparedness in a bioterrorism context. The authors argue that information structures generated by intentional and non-intentional disease outbreaks can be very different. These differences can impact missing information and associated decision-making ability of public health personnel. The findings of the chapter point to the need for sustained and comprehensive strategies for public health preparedness as opposed to funding strategies that are intended to support specific programmatic goals as a short-term reaction to bioterrorism threats.

The third chapter by Swaminathan, Brown, Long, and Gerner-Smidt (''PulseNet Provides Early Warning for Foodborne Disease Outbreaks'') presents a summary of a collaborative approach to building a cross-organizational system for investigating communicable diseases at national and international levels. The chapter describes the PulseNet system built by the Centers for Disease Control and Prevention (CDC) that enables sharing of DNA profiles of disease agents by state, county, and city public health laboratories across the nation. More recently, an international effort has been launched to enable information sharing between public health departments across countries. The benefits of such a network are quantifiable both in terms of the effectiveness of information gathering as well as efficiency in responding to communicable diseases.

The fourth chapter by Adam, Paliwal, Atluri, Chun, Cooper, Paczkowski, Bornhövd, Nassi, Schaper, and Ellenberger (''Government Agency Interoperation in Security Applications'') focuses on technical challenges with developing an agile and effective situational awareness platform for information integration and dissemination. The particular application context studied is incident management at state and local levels. Drawing from lessons learned from a large collaborative project, the authors, consisting of both academic researchers and government partners, make a strong argument for a semantics-driven incident management framework. They present components of this framework and discuss a prototypical implementation.

The process- and information-centric view of public health preparedness and incident management resonates with the objectives of the fifth chapter by Behara and Bhattacharya (''Process-Centric Risk Management Framework for Information Security''). A hallmark of the approach outlined in this chapter is to consider the people, technology, information, and work-processes in assessing and quantifying risk measures. Given that information and work-processes are highly integrated in today's information-based economy, approach to securing infrastructure resources should take a work-process driven perspective. A key differentiator of this risk management framework is the emphasis on process in addition to the traditional asset-based assessment of risks.

The last two chapters in this section are centered around the protection of the information and computer network infrastructure. The sixth chapter by Li and Ye (''Intrusion Detection and Information Infrastructure Protection'') provides a comprehensive introduction to recent intrusion detection work. Intrusion detection has become a critical technology to protect computer systems from various attack schemes. Such attacks could be used by groups with malicious intents or enemy military forces to disrupt the critical information network infrastructure on which modern life and work depends. In this chapter, the authors argue that the next-generation intrusion detection technology should be framed as part of the larger information- and network-centric warfare. They also advocate a holistic perspective on intrusion detection, which considers the entire lifecycle of the relevant data and detection algorithms in a complex and distributed environment.

The last, seventh chapter by Qu and Hariri (''Anomaly-Based Security Framework for Network-Centric Systems'') is closely related to the previous chapter but with a particular emphasis on attacks on networks. The motivation behind their work is that most existing methods for analyzing and preventing network attacks rely on an offline analysis and that there is a critical need for a proactive self-protection mechanism. In this chapter, the authors present such an online mechanism by employing anomaly analysis to protect against known and unknown network attacks. In their work, an information–theoretic anomaly detection method has been developed and compared against existing approaches.

Concluding remarks

National security is clearly the topic of our times. IT is intricately weaved into all aspects of security and increasingly needs to be monitored and understood for optimal benefits. One outcome of this book is bringing attention to the existence of an information supply chain, which holds the various entities engaged in a collaborative fashion. While the concepts of physical supply chain have greatly influenced the nature of logistics collaborations in public and private sectors, the criticality of maintaining an

information supply chain is becoming increasingly apparent. Our conceptualization of information supply chain envisions an integrated collection of technologies to provide secure and integrated decisional environments that enable public, private, and academic partners to collectively sense and respond to opportunities and challenges in this increasingly networked ecosystem. The informational and technology related issues raised by the authors of this book stress the importance of the information supply chain. The contributing authors in this book have taken a first step to define the underlying issues related to national security. We hope that this book initiates ongoing discussions on this topic.

We thank all the authors of the chapters for their commitment to this endeavor, and their timely response to our incessant requests for revisions. The editors would like to recognize the contributions of staff at the Center for Advancing Business through Information Technology (CABIT), in particular, the patient editing and collating services of Kimberly Linton, Shweta Bhandari, and Swetha Mikkilineni. Co-editors Chen and Zeng would like to acknowledge and appreciate the following grant funding that has contributed to their work (NSF #EIA9983304, #ITR 0326348, #ITR 0428241, #IIS 0429364). In addition, the editors wish to thank the editorial staff at Elsevier for their professional assistance and patience, especially Gerard Wanrooy, Philip Tite, and Julie Walker.

Part I:
Legal and Policy Frameworks

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 1

Should Commercial Misuse of Private Data be a Crime?

Susan W. Brenner

University of Dayton School of Law, 300 College Park, Dayton, OH 45469-2772, USA

Leo L. Clarke

Thomas M. Cooley Law School, 111 Commerce Avenue SW, Grand Rapids, MI 49503, USA

Abstract

Computer technology lets people acquire and retain knowledge, communicate instantly and globally, purchase goods and services, engage in hobbies, and participate in politics and cultural affairs in less time and with less expense than was once dreamed possible. This revolution has had certain consequences for an individual's legal expectations of privacy. It has become increasingly common for vendors, service providers, and government agencies ("Collectors") to collect and retain data about transactions ("Data") involving individuals ("Consumers"), who often disclose intimate details of their lives and lifestyles. The retention of this Data by private parties creates a serious risk that unauthorized third parties, such as hackers, will obtain that Data without the Consumer's consent. The chapter analyzes the implications of this phenomenon. After outlining how technology impacts on our traditional notions of privacy, it argues that Consumers should not lose their privacy interest in the Data they share with Collectors in the context of a trust-based relationship. The chapter urges the adoption of a concept of relation-based shared privacy that will protect Consumers' privacy interest in personal Data without interfering with the benefits that result from advances in technology. The chapter concludes with a consideration of the sanctions that should be imposed for breaches of this doctrine of relation-based privacy.

1 Introduction

We live in a world of pervasive, ubiquitous data collection and retention. In this chapter, the phrases "ubiquitous technology" and "ubiquitous computing" are used interchangeably to refer to technologies that are woven into the fabric of everyday life (Winters, 2004). Ubiquitous computing involves having computing devices essentially everywhere in the home, office, or public area, as well as easy and natural ways for people to interact with them. Wireless technologies, sensors, radio frequency identification (RFID) tags and machine-to-machine communications are all examples (Blau, 2004). This article focuses on "communicative" technologies instead of, say, industrial or agricultural technologies. Its concern is with technologies that can be used to generate information, collect information, and/or share information.

Modern computer technology permits us to acquire and retain knowledge, communicate instantly and globally, purchase goods and services, engage in hobbies, and participate in politics and cultural affairs, all in less time and with less expense than once dreamed possible. One major effect of this revolution has been a serious reduction in an individual's rights and expectations of privacy. It has become increasingly common for vendors, service providers, and government agencies ("Collectors") to collect and retain data about transactions ("Data") involving individuals ("Consumers"), who in the course of those transactions often disclose more intimate details of their lives and lifestyles than would have ever been imaginable or acceptable just a decade ago. In turn, this retention creates an unprecedented risk that third parties such as criminal hackers or multinational corporations will obtain that Data without the Consumer's consent. This risk arises because a Collector in possession of Data could either fail to adequately protect that Data or decide to disclose it for the Collector's own gain. In this chapter, we call it "misuse" when a Collector permits or through its negligence allows third parties to use Data without the Consumer's consent.

In this chapter we first consider, in Part II, how advances in technology have transformed the context in which we consider the notion of privacy. In the past, Consumers disclosed Data to Collectors such as educational, religious, and medical institutions either orally or in scattered paper documents. Now, transferring the Data in a digital format allows the Collector to sort and report it in ways never before possible. With the increasing computerization of services to our homes such as security surveillance, cable television, and even "smart houses," Data that relates to events within our homes that was previously available only to family members or servants is now communicated to Collectors' databases. Similarly, the increasing sophistication of remote sensing and database technology means that the amount of Data available to providers of utility and telecommunications services has dramatically increased.

Part II sets the stage for the inquiry addressed in Part III: Do we lose our privacy interest in that Data because it is now more efficient to collect it in a database where it can be searched and sorted in a myriad of ways? We argue that pervasive technology should not deprive Consumers of their privacy interests in Data that they have shared with Collectors in the context of a trust-based relationship. Rather, a notion of relation-based shared privacy will offer adequate protection to Consumers' Data without interfering with the benefits to be obtained through advances in collection and database technology.

Finally, in Part IV, we consider the nature of the sanctions that should be imposed for misuse of private Data. Courts have long recognized that one appropriating the private information of another should be liable for damages under the civil law (Warren and Brandeis, 1890). We analyze, however, whether the potential for misuse of Data in a world of pervasive technology would justify criminal sanctions. This suggestion may on first consideration seem unsupportable since the law is generally parsimonious with criminal liability. Logically, however, there could be at least two reasons to use criminal liability: (1) Civil liability alone is not sufficient to ensure Data privacy and (2) Data privacy is an interest significant enough to justify defining violations as an affront to the state.

Before we proceed to that analysis, however, we lay the groundwork by considering how pervasive technology affects our understanding and need for Data privacy.

2 The need for privacy in an era of ubiquitous information technology

There has always been a close relationship between technology and the legal system's recognition of a right to privacy. Privacy evolved as a "bricks and mortar" concept, and therefore was constrained by the needs of the real world. Now that computer technology has given us the ability to transcend the strictures of the real world, we have the ability to substitute virtual realities for the physical world; we can communicate instantaneously with almost anyone from almost anywhere; we use technologies to make our lives easier, to earn our living, and even for our own amusement. It follows that our concepts of privacy must adapt if we are to maintain the protections afforded to real-world interests as they are affected by cyber technologies. The question becomes, how should our legal constructs adapt to those changes?

Samuel Warren and Louis Brandeis's famous 1890 article *The Right to Privacy* is a good starting point for this analysis (Warren and Brandeis, 1890). They argued for a common law cause of action for invasion of an individual's privacy, which differed from the then extant concepts of privacy because it (i) was directed at private parties and (ii) did not involve a zero-sum approach to privacy. They demonstrated that an individual's

ability to exercise some control over how the private sector gathers, disseminates, and uses personal information is fundamental to an ordered society (Warren and Brandeis, 1890).

Warren and Brandeis faced several conceptual difficulties in articulating their new right to informational privacy. For our purposes, the most fundamental difficulty went to the essence of the principle: What is "private?" Historically, privacy had been interpreted as incorporating a zero-sum conception of privacy in which only two states exist: private or not private. However, this simplistic notion did not work for Warren and Brandeis because they were concerned with how new technologies affected traditional understandings of privacy. Since it was Warren and Brandeis's goal to control the collection, dissemination, and use of information about individuals, they sought to redefine "privacy" in order to make it more consistent with and analogous to a property right. The eventual adoption by virtually every state of the Warren–Brandeis analysis demonstrates that such a redefinition was an essential consequence of the evolution in these particular technologies.

The need for a similar redefinition is even more pressing today, as demonstrated by the national debate over privacy interests in Data maintained by health care providers, financial institutions, and retail merchants (American Civil Liberties Union, 2004). A commentator recently captured the public's concern over Data privacy:

> Law enforcement and intelligence services don't need to design their own surveillance systems …. They only have to reach out to the companies that already track us so well while promising better service, security, efficiency, and, perhaps most of all, convenience. It takes less and less effort each year to know what each of us is about. When we were at the coffee shop and where we went in our cars. What we wrote online, who we spoke to on the phone, the names of our friends and their friends and all the people they know. When we rode the subway, the candidates we supported, the books we read, the drugs we took, what we had for dinner, how we like our sex.
>
> More than ever before, the details about our lives are no longer our own. They belong to the companies that collect them, and the government agencies that buy or demand them in the name of keeping us safe …. (O'Harrow, 2005, p. 300)

A recent survey of likely U.S. voters found that over 70% favored more legislation to protect the privacy of their Internet-related communications and Data (Cyber Security Industry Alliance, 2005). What created this heightened public concern? No doubt it is the creeping realization that Data retention by the businesses from which we purchase the vast majority of our goods and services is not only pervasive, it is unavoidable. Such pervasive technology affects us not only when we venture into the public marketplaces, but it is also intruding into our homes at an increasing rate. As computer technology becomes a more and more embedded feature in every aspect of our lives, our homes are becoming equipped with technology that can be used to eavesdrop on our conversations and track our activities, even though such Data collection and retention is not a primary purpose motivating its use (Kannellos, 2005). Indeed, such technology

continues to be successful in the marketplace only because its information collection aspects are overshadowed by the benefits it provides to Consumers (Raisinghani et al., 2004).

For example, efforts are underway to develop "aware homes" that incorporate intelligent, embedded systems, which interact with the occupants and with outside technology (Ward, 2004). Similar systems will become features of offices, hotel rooms, and other environments (European Commission, 2001, pp. 4–7). While the potential for abuses of the information-gathering capabilities of such products is particularly dramatic, the *nature and sensitivity* of the information gathered is often not inherently different from that acquired by mundane Collectors such as grocery and clothing retailers.

Pervasive technology raises difficult issues about privacy, especially for those who are not users of advanced technology. "Old-century" people may think that their communications and activities are private only insofar as they shield them from observation by others. Such a view tends to associate "privacy" with enclaves such as our homes, cars, and offices. Those who are accustomed to using new technology, however, are rapidly experiencing a decline in the privacy traditionally associated with these enclaves. Cell phones have basically eliminated phone booths, vehicles are equipped with surveillance technology, wireless networks and cellular communications, and information concerning much of what goes on in our homes can be obtained by third parties. Offices may be somewhat more secure, but much of our work takes place outside our offices; "road warriors" equipped with the latest wireless communication conduct business from—and on their way to and from—other offices, and other places. The notion of "private enclaves" as places separate and apart from the world, areas in which our activities and communications are not subject to observation, is disappearing. In this world of ubiquitous and ambient technology, "an invisible and comprehensive surveillance network" has been created, the constituent parts of which are operated by private Collectors. This network has effectively eradicated the distinction between "public" and "private" spaces (O'Harrow, 2005, p. 291). Information that was historically secluded behind physical barriers now has the potential to leak into the "public" domain.

In a sense that, then, the issue becomes whether the notion of privacy is limited to the right to *prevent* others from gaining access to Data, or whether it can be construed as the mere right to *control* access (Brenner, 2003, pp. 398–402). For example, when a Consumer gives her credit card to a waiter to pay for dinner, she knows she has given him access to the information that (a) she has a Visa card; (b) the card numbers are 3333 4444 5555 6666; and (c) the expiration date is 07/2006. But it is reasonable for her to assume that she is making a controlled disclosure of that information, that the waiter will use it only to process the credit that will pay her bill.

The same is true if she gives the sponsor of a conference her Social Security number so the sponsor can reimburse her for her expenses as a

speaker. Here, too, it is reasonable for her to assume that she is making a controlled disclosure of the information in which it will be used for the intended purpose and no other. This expectation of controlled disclosure is an unarticulated, perhaps unrealized, but reasonable and essential component of the commercial transactions and other activities we pursue in our everyday lives. Its source is a world in which the clerks at the general store could note a customer's purchases and predilections but had no way, aside from provincial gossip, to distribute that information to others.

As we all know, things have changed. Data has become an implement we employ for various purposes, including identification, authorization, and payment. We can use cash to pay debts, but it will be extraordinarily inconvenient to buy a new car with cash. And cash simply cannot be used for certain types of transactions; the Consumer must use Data in the form of a credit card. Other types of Data, such as Social Security numbers, may be necessary for the Consumer to identify herself. Routine aspects of a Consumer's life, such as purchases, become Data; and his activities, such as jogging, can be transformed into Data. None of this Data is private in the sense of being accessible only by the Consumer. The Consumer understands others have access to it, but she expects that their access will be limited and that her disclosures of information about herself are controlled.

3 Relation-based shared privacy

We thus reach the question: how to determine when a Collector's misuse of such a right of access violates the Consumer's privacy interest. The continued use of real-world, zero-sum notions of privacy in a world of pervasive transactional Data collection would substantially reduce our ability to maintain a cloak of privacy around much of what might be called our private lives. We suggest that whether Data in the hands of a Collector should still be considered as subject to a Consumer's privacy claim should depend on the nature of the relationship between the Collector and the Consumer. If the parties have entered into a relationship that demonstrates an intention to share the Data, each party has an independent interest in keeping that Data private. We therefore contend that Data should be subject to privacy protection against misuse if the general purposes that lead (1) Collectors to store and mine it and (2) Consumers to permit that storage and manipulation, reflect the parties' legitimate expectation that the Collector will not exercise sole dominion over the Data. Whether this shared privacy interest exists in a specific case should be determined from the nature of the transactions involved and the expressions of the parties regarding their relationships.

We call our approach "relation-based shared privacy." In this Part, we define the types of relationships and the nature of the privacy expectations that should produce an expectation of privacy for stored transactional Data

maintained by a Collector. Next, we identify three parameters for determining whether Data should be subjected to privacy protection. Each parameter derives from the underlying competing interests: the Consumer's privacy interest in the information and society's interest in a free flow of information.

We start with the premise that one can share information without contemplating that the information will be disclosed to the *public* or even to other third persons. We do not suggest, however, that privacy protection should depend on a case-by-case evaluation of the subjective or objective intent of parties who disclose information. Rather, protection can be based solely on the existence of defined relationships from which we can conclude that society does or should recognize a privacy interest. For example, if we look into old-century analogs, we see that society has long recognized that many of these disclosures take place in the course of defined relationships, such as wife/husband, patient/doctor, client/attorney, and penitent/priest, where society's interest in maintaining the free flow of information justifies even an evidentiary privilege. We can also identify other relationships that have enough societal significance, if only from the viewpoint of personal autonomy and economic efficiency, to justify protecting the disclosing party's interest in confidentiality. Trade secret protection and enforcement of confidentiality (non-disclosure) agreements are just two examples of doctrines that recognize "shared privacy" interests.

The notion of shared privacy does not depend on the existence of express confidentiality agreements. For example, when servants in the home were more common, it would be unreasonable to conclude that the presence of a servant destroyed the privacy of a conversation between family members. The servants understood that a condition of their employment was that the conversations stayed in the room. It would make no sense from a societal viewpoint to hold that a conversation was not private just because family members failed to dismiss the servant from the room before conversing (Brenner, 2005, p. 25).

In other words, the existence of a relationship of a given nature can demonstrate that the disclosing party expected that the information disclosed would be kept confidential *and* that her expectation was reasonable. In the absence of the relationship, the information would not be private. For example, a conversation between a husband and wife in front of a butler serving dinner in the family dining room would remain private, while the same conversation in the presence of a waiter in a restaurant would not, in the absence of other circumstances, be private.

We see two key differences in these situations. One is that the first conversation takes place in the home, where there is a greater expectation of privacy; this spatial consideration does not apply to the present context. The more important difference for present purposes is that the spouses have an existing relationship with the butler based on at least in part the trust that the butler will respect the confidentiality of family conversations. In other words, the nature of the trust is that neither the spouses nor the butler

feels that the butler is free to disclose the conversation outside the home. This conclusion is based on the historic understanding that the privacy of the home encompassed family members, servants, and guests (Oysted v. Shed, 1816). No such trust-based relationship exists with the waiter at least in the absence of other circumstances. The situation would be different if, for example, the spouses are regular customers of the restaurant and the waiter is their usual waiter who is familiar with their habits.

We conclude that the Data maintained by a Collector with respect to a Consumer should be protected from misuse under the following conditions:

(a) the Consumer has provided, and the Collector has collected and retained, the Data in the course of a relationship that permits a reasonable inference that the Data would not be practicably available but for the Data collection and mining capabilities of "pervasive technology;"
(b) the Collector maintains the Data (i) at least in part for the direct benefit of the Consumer and (ii) the Consumer has direct access to at least a material part of the Data; and
(c) the Collector has agreed not to disclose the Data to third parties without the Consumer's consent.

3.1 Relation-based

Ubiquitous technology requires a re-evaluation of the appropriate balancing of private and public interests for privacy purposes. Collectors should be able to allow third parties to have access to Data that the Consumer has set adrift in the stream of commerce in the sense that the same information would have been disclosed to casual observers or employees of the Collector in comparable real-world transactions. On the other hand, the mere fact that a Collector possesses Data should not permit the Collector to use it at its whim. The problem lies in attempting to identify the factors that should be taken into account in determining the appropriate balance in cases between these two extremes.

One way to determine the application of privacy protection in the world of pervasive technology is to compare such technological transactions with analogous real-world transactions. For example, one factor to consider is the "visibility" or "publicity" of the transaction that created the Data at issue. The Consumer who buys an automobile tire at a retail store has no expectation that the *fact* of her purchase is private because the seller's employees and other customers can see the purchase; also, anyone seeing her car can infer that she had purchased that brand of tire. The purchase is not private in any sense. The tire purchaser therefore cannot complain if a third party obtains Data from the retail store, or from manufacturer, confirming the fact that she bought that tire or from obtaining related transactional Data such as the time, date, and price of the purchase.

The same rule should hold true for Data identifying a single transaction occurring through the use of pervasive technology if sufficient indicia identifying that transaction are inherently public. For example, Data regarding a tire purchase does not become "private" just because the purchaser completes the transaction in the privacy of her home through cheaptire.com and puts the tire on her car in her own garage with the garage door closed. Even though she may hide from public view many of the aspects of the transaction, the telltale sign (the tire on the car) is still visible to the public, so the Collector should have free use of the Data for the same reason as stated above for a retail store transaction.

Different concerns are presented, however, when a Consumer and a Collector each manifest an intention to maintain the privacy of transactions that otherwise might be public. For example, a Consumer who desires to purchase prescription medicine and wishes to maintain her privacy might be entitled to privacy protection if she purchases the medicine through a secure website that promises confidentiality and that delivers the medicine in a plain wrapper.

The intention to maintain privacy is readily inferable when a Consumer, in the course of creating or continuing a relationship that anticipates at least the strong likelihood of multiple transactions, provides "personal profile" Data that the Collector combines in a database with transactional Data. Such a relation can be found in the delivery of "personal profile" Data to the Collector with an expectation on the part of the Consumer and the Collector that the Collector will combine the profile Data with transactional Data. (Examples of "profile" Data would be the Consumer's Social Security number, weight, birth date, or mother's maiden name. Each item of profile Data may be public in some sense, but it can be private when aggregated.)

This combination of personal information with the details of multiple transactions creates a corpus of information that bystanders could not observe, and thus supports a conclusion that the Consumer and Collector have entered into a "private" relationship. Moreover, the Consumer's willingness to allow the Collector to combine personal and transactional Data into a database strongly supports the inference that the Consumer reposes enough trust in the Collector's goods or services that she anticipates repeated dealings with the Collector. The trust we refer to is not trust that the Collector will not disclose information. Rather, it is the Consumer's trust in the value of the Collector's products such that the Consumer anticipates continued dealing with the Collector.

The combination of a corpus of complex information and the prospect of repeated dealings is sufficient to create and sustain an expectation of privacy beyond that created by mere contract; it also creates the possibility that an aggregation of Data can compromise Consumer privacy.

We should note that the Consumer may have an expectation of privacy even if such Data pertained to a transaction occurring outside the context

of pervasive technology. For example, mail order and phone order transactions are not observable by third parties any more than Internet transactions, and the records maintained by the Collector may not differ between the two types of transactions. To the extent that database technology is employed in such old-world transactions, our argument, as set forth below, may apply to those transactions as well because Consumers should be encouraged to participate fully in modern society without requiring a forfeiture of important societal interests. We also note that drawing distinctions in any of these types of transactions based on comparisons of database contents to the *potential* recollections of Collector employees is not persuasive, especially when transactions are completed solely on the basis of digital transmissions and computer generated documents and records. For example, given computer technology, no employee even completes an address label in an Internet purchase transaction.

Why should the existence of such a relationship with no historic legal substance or grounding have significance in the criminal law? The answer to this question requires us to revisit the notions of privacy discussed in Part II. Pervasive technology changes the focus of privacy from a Consumer's right of physical control over space or tangible property into a right to impose sanctions for disclosure of information in databases over which the Consumer has no physical control or access. While the notion of physical control is a reasonable approach to implementing privacy when we deal with spaces and things, it is meaningless with respect to a modern information-based economy. Requiring physical control, therefore, would effectively place Data beyond privacy protection without any balancing of societal costs. In short, neither control nor rights of physical access can provide a limiting principle that will distinguish privacy-protected interests in Data. Instead, we need a surrogate that will enable us to avoid both the total abrogation of privacy protection to Data in a world of pervasive technology and an unprincipled *ad hoc* application that turns on mere formalistic notions of privacy.

Focusing on the existence of a "trust" relation between the Consumer and the Collector, even though the trust may be merely inferential and minimal, enables us to evaluate the reasonableness of a Consumer's claim that Data remains private and is entitled to societal protection. Prior to the implementation of pervasive Data collection, retention, and aggregation technology, there was not a realistic possibility that a Collector could disclose Data reflecting a Consumer's personal profile information *and* the details of numerous specific transactions between the Consumer and the Collector. We can therefore confidently say that Consumers in most circumstances had a reasonable, empirically based expectation that the *aggregate* Data reflecting those transactions were not available to third parties or even to other departments of a Collector. When the Consumer "trusts" the Collector and its products enough to anticipate the potential for such aggregation of information, it is unreasonable to conclude that the

Consumer in providing Data is indifferent to its use. In a very real sense, the trust in the Collector's products reflects trust in the integrity of the Collector to maintain the privacy of the Data provided.

Unless we are ready to adopt the view that privacy protection should continually narrow as technology increasingly permits information to be stored and correlated, there is little reason to conclude that a Consumer *should* expect that Data becomes public just because it is mined and aggregated. That is, Data inaccessible in the real world should not lose privacy protection just because it *can be* accessed in a world of pervasive technology. As a matter of societal values, a Consumer *reasonably* expects that it will not be so disclosed simply because she has chosen to conduct her affairs by using more efficient pervasive technology to conduct transactions with "trusted" parties.

3.2 Shared interest based on direct consumer benefit and access

Not all Data possessed by a Collector in the course of a "trust" relation will be entitled to privacy protection. There is still a role for assumption of risk. A Consumer should be held to have assumed the risk that Data collected at the sole instigation and for the sole benefit of the Collector is beyond protection because the Consumer effectively set the Data adrift in the stream of commerce. For example, before the advent of Data mining, businesses collected Data for internal marketing, inventory control, product quality, regulatory, and warranty liability purposes. The Collector's use of that Data indirectly benefited Consumers in general, whether by lower prices or higher quality. Usually, however, the Data itself was not manipulated and re-disclosed to assist the Consumer in making additional purchase decisions or obtaining service. (To the extent the information was so used by salespeople, for example, our notion of shared privacy might apply.) Stated differently, individual Consumers received no direct benefit from the collection of the Data. Therefore, one could not reasonably conclude that the Consumer had provided the underlying information with the expectation that the Collector would use the Data for the Consumer's own purposes and benefit. In short, the Consumer had given up any privacy "interest" in the information.

In contrast, Consumers who provide "profile" Data to Collectors generally do so because that profile information, when combined with transactional Data, saves the Consumer time and/or money. A significant amount of those savings can derive from the ability of database technology to aggregate or isolate Data to provide the Consumer with new information or insights regarding her dealings with the Collector. In this context, it seems reasonable to conclude that the Consumer has a *shared interest* with the Collector in the Data because Consumers are induced to provide the relevant information at least in part on the ground that they will also benefit. Because the Consumer retains an interest in the Data, the Collector

should not have a unilateral right to disclose the Data to third parties or to use the Data as it pleases. Moreover, society has an interest in protecting that Data from misuse because disclosure would discourage Consumers from sharing Data that allows them to make more intelligent and more efficient transactional decisions.

This shared interest is particularly evident when the Collector enters Consumer-provided Data into a database that allows the Consumer direct access to information about the Consumer's dealings with the Collector. The right and value of direct access to information regarding past transactions and related financial information is one of the great benefits of Internet-accessible Data mining. For example, by going to "My Account" on a electricity utility's website, a Consumer can review her past electricity usage, compare it to average usage statistics, estimate potential energy saving from replacing her water heater, and evaluate the effect of various pricing options in light of her particular energy usage patterns. Such access permits a Consumer to use the Data for her own purposes unrelated to any benefit to the Collector. For example, a Consumer might consult information on orbitz.com regarding past flights and hotel stays in connection with purchasing travel services on expedia.com or directly from an airline. Therefore, the independent usage strengthens the notion of a shared interest by both the Collector and the Consumer.

Direct access also reinforces the significance of the "relation" element because it demonstrates the existence of a more permanent relationship between the Collector and the Consumer. The Collector incurs the expense of creating and maintaining the database to increase the likelihood that the Consumer will enter into additional transactions with the Collector. It is this repetition that creates the aggregation of Data, which in turn increases the risk of an invasion of privacy and invalidates an analogy to observation of real-world transactions. In short, direct access is a significant limiting characteristic of relation-based shared privacy because it is a strong, investment-backed evidence that the Collector and the Consumer are parties not just to a transaction, but also to a private relationship.

3.3 Confidentiality representation or agreement

Parties in the world of pervasive technology rely on contractual promises to control access to Data. A societal/criminal privacy protection for Data should be found only if the Collector breaches a promise to the Consumer to maintain the confidentiality of her Data. Society has no interest in protecting the interests of those who do not value privacy enough to satisfy this simple element. We do not mean to suggest that Consumers must draft their own confidentiality agreements or even have read, much less fully appreciate, a Collector's website "terms of use" regarding privacy and Data usage. Instead, it is likely that market forces will be sufficient to attract privacy-conscious/valuing Consumers to Collectors who unilaterally

represent that they will not disclose Consumer-related Data to third parties without the Consumer's consent. Thus, this element of relation-based shared privacy is satisfied if the Collector includes such a confidentiality undertaking in its customer agreement or website terms of use. It should also be sufficient to show that a third party credentialing service has certified that the Collector's privacy procedures include a commitment not to disclose Data.

4 Criminal liability

There are two justifications for using criminal liability to punish a misuse of a relation-based shared privacy interest in transactional Data: one is that civil liability alone is insufficient; this "true crime" rationale is discussed Section IV(A) (Sayre, 1933, p. 84). It is based on the premise that a violation of Data privacy inflicts personal "harm" of the type society cannot tolerate. The second rationale is based on the premise that Data privacy is a systemic interest, which is significant enough to justify defining violations as an affront to the state; this "public welfare offense" rationale is discussed in Section 4.2 (Sayre, 1933, p. 84).

4.1 "True crime"

Criminal law is concerned with "mak[ing] people do what society regards as desirable and …prevent[ing] them from doing what society considers to be undesirable" (LaFave, 2003, Section 1.5). It maintains order, which is essential if a society is to carry out the processes necessary for its survival (Brenner, 2004, pp. 6–12). The traditional model of criminal law targets "true crime," which consists of one individual's inflicting a proscribed type of "harm" upon another (Brenner, 2004, pp. 15–25). A society's "crimes" each target a specific "harm." The "harms" encompass conduct a society cannot tolerate; every society therefore criminalizes "harms" against persons (murder, rape, etc.) and against property (theft, arson, etc.) because a society cannot survive if its members are free to prey upon each other and each other's property (Brenner, 2004, pp. 15–25). Societies also outlaw "harms" against morality (e.g., adultery, blasphemy, etc.) and against the state (e.g., riot, treason, etc.); but crimes against persons and property have been the core of criminal law (Brenner, 2004, pp. 15–25).

4.1.1 "Harm"

The first step in applying this model to Data privacy is identifying the "harm" that warrants the use of criminal sanctions. The use of such sanctions does not have to be the only avenue of recourse against a perpetrator; victims of such traditional crimes such as murder and fraud can sue their victimizers (Bernoskie v. Zarinsky, 2001). But if we are to authorize the use

of criminal sanctions to secure Data privacy we must identify a "harm" of sufficient severity to justify such a measure, one that cannot be adequately addressed by the use of civil liability. For example, the drafters of the Model Penal Code, which is the source of criminal statutes in most of the states of the United States, did not include a criminal libel provision in the Code because they believed "penal sanctions cannot be justified merely by the fact that defamation is … damaging to a person in ways that entitle him to maintain a civil suit … [W]e reserve the criminal law for harmful behavior, which exceptionally disturbs the community's sense of security" (American Law Institute, 1961, Section 250.7 cmt. at 44). We submit that our discussion in Part II above demonstrates that the societal benefits of pervasive technology should not come at the cost of a loss of privacy. Therefore, such a loss would be a "harm" which, when inflicted upon individuals, "disturbs the community's sense of security" in an exceptional way.

Another option would be to use a property analogy and develop a "misuse of personal Data" offense as a crime against property. The rationale would be that personal Data is intangible property that "belongs" to the person whom it concerns in the same way intellectual property "belongs" to its author; the theory is that while personal Data is not a commodity over which I can exercise exclusive possession and control, it "belongs" to me in the sense that I can dictate with whom it is to be shared and how they can utilize it.

4.1.2 Elements

The second step in using a "true crime" approach to apply the traditional model of criminal law to Data privacy is designing a "misuse of personal Data" offense. Traditional offenses have three basic elements: (a) conduct; (b) mental state; and (c) a prohibited result (American Law Institute, 1961, Section 1.13).

The first issue we need to address, then, is the conduct that constitutes a "misuse" of Data. This seems relatively simple, since concerns about Data's being misused focus on two activities: (i) collecting personal Data and (ii) using it. The "conduct" element of the offense will therefore encompass collecting and/or using personal Data; and we will assume "personal Data" is Data that satisfies the notion of relation-based shared privacy.

But we still need to define other terms: What does it mean to "collect" personal Data? It does not encompass "observation"; when a Consumer's neighbor sees her come home, the neighbor has not "collected" Data establishing that the Consumer lives at 555 Lamont Terrace, Dayton, Ohio. Collection requires that the Data become a *commodity*, something that can be preserved, manipulated, and distributed; under this definition, photographing a police officer jogging on a high school track or noting the officer's jogging there and posting that information on a website would seem to constitute "collecting" personal Data (Brenner, 2003, pp. 398–402).

The waiter skimming credit card numbers would clearly be ''collecting'' Data, and the same is true for websites or real-world establishments that record and compile customer information.

What about ''using'' Data—how should we define that? In criminal statutes, ''using'' has been defined as ''[t]aking or exercising control over property; or ….[m]aking any … disposition, or transfer of property.'' (Florida Annotated Statutes, Section 825.101(10)). The waiter who skims credit card numbers would be ''exercising control over property,'' as defined above; and by the same token, commercial establishments' logging and compiling Data about my buying habits would be ''taking control over property,'' and their selling the information to other commercial entities would clearly qualify as transferring that property.

The next offense element we need to define is *mens rea*. Since the traditional model of criminal law tends to limit liability to advertent conduct, we will do the same. Under the influence of the Model Penal Code, American criminal law relies on two advertent mental states: purposefully and knowingly (American Law Institute, 1961, Section 2.02(2)). One commits an offense purposefully when it is her conscious objective to cause the ''harm'' constituting the crime; one does so knowingly when she is aware it is practically certain her conduct will cause such ''harm'' (American Law Institute, 1961, Section 2.02(2)). When a statute makes it a crime ''knowingly'' to inflict a proscribed harm, an offender can be convicted even though she acted purposefully (American Law Institute, 1961, Section 2.02(5)). Using purposefulness as the required mental state for the misuse of personal Data offense would circumscribe its reach because it could only be used to prosecute those whose *goal* it was to engage in the unauthorized collection and use of personal Data. When the ''harm'' an offense addresses is conduct, it is reasonable to use the purposefulness as a standard because the object is to sanction only those who deliberately engage in that conduct. But when an offense addresses a ''harm'' consisting of a specific result, there is a strong argument for expanding the requisite culpability level to include ''knowing'' as well as ''purposeful'' conduct (LaFave, 2003, Section 1.4(b)). Doing so allows the prosecution of those who are aware that their conduct will almost certainly produce the proscribed result, as well as of those whose goal it is to cause that result. Therefore, since the misuse of personal Data is concerned with results, not with conduct, the appropriate level of culpability should be ''knowing'' misuse.

The final offense element is the result, the ''misuse'' of personal Data as defined above. The only issue we need to consider is whether to limit the scope of the offense to instances in which misuse actually occurred (substantive offense) or extend it by including an ''attempt to misuse personal Data'' offense, i.e., an incomplete or inchoate offense (LaFave, 2003, Section 11.1). Pragmatically, it seems an attempt offense would seldom be used, since it is unlikely that law enforcement officers would successfully frustrate the efforts of those bent on misusing another's personal Data with

any degree of frequency. On the other hand, such an option would allow prosecution when such efforts were interrupted.

4.2 "Public welfare offenses"

"Public welfare" offenses are the product of a very different approach to the imposition of criminal liability. To understand this approach, it is helpful to consider a specific offense, i.e., antitrust. Antitrust prosecutions differ from traditional prosecutions in that they are predicated on the infliction of a systemic "harm" while traditional criminal proceedings are predicated on the infliction of "harm" to individual victims. In a traditional criminal proceeding, the state acts to vindicate its obligation to protect the individual members of the social system it represents (American Bar Association, 1993). In a criminal antitrust proceeding, the state acts to vindicate its obligation to ensure the viability of an essential component of a social system (U.S. Department of Justice, 1997). The "harm" at issue is an erosion of the principle of competition; criminal antitrust proceedings target "systemic" crimes, i.e., crimes that impact on a nation's infrastructure, instead of "individual" crimes.

4.2.1 United States v. Park

"Public welfare" offenses emerged at the beginning of the twentieth century (Sayre, 1933, pp. 67–68). They were the product of a "shift in emphasis from the protection of individual interests, which marked nineteenth century criminal administration to the protection of public and social interests" (Barber, 1992, p. 110). The first "public welfare" offenses "regulated liquor sales, adulterated food and drug sales, narcotic sales, misbranded articles, … and criminal nuisances including injuries to public health and safety. Such crimes were not … true crimes because the act that constituted the offense was not intrinsically wrong" (Barber, 1992, p. 111). They differed from "true crimes" in another respect; "public welfare" offenses did not require *mens rea* (U.S. v. Balint, 1922).

As "public welfare" offenses evolved, they also eliminated the need for culpable conduct, as is illustrated by another Supreme Court case (U.S. v. Park, 1975). John Park was CEO and President of Acme Markets, Inc., a "retail food chain" that had "36,000 employees, 874 retail outlets, 12 general warehouses, and 4 special warehouses" (U.S. v. Park, 1975). Its headquarters, and Park's office, were in Philadelphia. Both Park and the company were prosecuted for violating 21 U.S. Code Section 331(k), which makes it a federal crime to let food that has been shipped in interstate commerce and is being held for sale become adulterated. The charges were based on food that was held for sale at Acme's Baltimore warehouse. Federal inspectors determined that the warehouse was "accessible to rodents" that were contaminating the food stored there and that the contamination constituted adulteration under 21 U.S. Code Section 331(k) (U.S. v. Park, 1975). Acme

pled guilty, but Park went to trial, claiming he was not ''personally responsible'' for the contamination.

Park argued that while all of Acme's employees ''were in a sense under his general direction,'' the responsibility for ensuring sanitary conditions at the warehouse belonged to the Baltimore division vice president; Park said he had checked and was told the vice president was taking ''corrective action.'' (U.S. v. Park, 1975). Park was convicted and appealed to the Supreme Court; it upheld his conviction, concluding that his ''responsible'' position in the company's corporate structure justified holding him liable for not preventing the contamination (U.S. v. Park, 1975). As the Court explained, the statute imposed not only a ''duty to seek out and remedy violations when they occur'' but also a duty to ''implement measures that will ensure that violations will not occur.'' The requirements of foresight and vigilance … are … demanding, and perhaps onerous, but they are no more stringent than the public has a right to expect of those who voluntarily assume positions of authority in business enterprises whose services and products affect the health and well being of the public that supports them (U.S. v. Park, 1975).

In *Park*, the Supreme Court upheld the imposition of criminal liability in the absence both of culpability (strict liability) and personal participation in unlawful conduct (vicarious liability) (LaFave, 2003, Section 13.4(c)). The rationale for eliminating these otherwise fundamental requisites of Anglo-American criminal liability is that ''public welfare'' offenses target systemic ''harms'' that are important enough to justify placing the risk of preventing a particular ''harm'' on those who are in a position to do so (LaFave, 2003, Section 13.4(c)). The need to this results from the fact that ''public welfare'' offenses target conduct that occurs in a business setting where it can be difficult, if not impossible, to prove personal moral fault on the part of specific employees (LaFave, 2003, Section 13.4(c)). To eliminate the potential unfairness inherent in this approach, liability cannot be imposed if someone was ''powerless'' to prevent the harm (U.S. v. Park, 1975).

4.2.2 Data crime

If we decide the interest in controlling the collection and dissemination of Consumers' personal Data is a systemic interest sufficient to warrant the creation of a ''public welfare'' offense, we can use the approach outlined above to protect that interest. This approach offers certain advantages: One is that prosecutions can proceed without the government having to prove that an individual ''harm'' occurred; it is enough to prove that conditions were in place which *could have resulted* in such a ''harm.'' In *Park*, after all, the government did not have to prove that the contamination of the food in the Baltimore warehouse actually ''harmed'' anyone; it only needed to prove that the conditions created the proscribed systemic ''harm,'' i.e., a default in the company's duty to ensure the integrity of the food it held for sale. In this regard, ''public welfare'' offenses are similar to inchoate

offenses; that is, they allow the government to intercede, and impose liability, without having to wait until an individual "harm" occurs.

Under the "true crime" approach, prosecutors can do nothing unless and until they are approached by a victim who has suffered an actual, individual "harm" of the type proscribed by the misuse of Data offense described in Section IV(A). Once they are approached by such a victim, the prosecutors must investigate the case, file charges, assemble evidence, and assume the risk of persuading a jury (typically) beyond a reasonable doubt that the defendant "knowingly" engaged in the unauthorized collection and/or use of the victim's personal Data. The defendant, in turn, can present evidence showing that (a) no such misuse occurred and (b) if it did, it was not due to "knowing" conduct on his or her part. In so doing, the defense, of course, does not have to actually prove anything; all the defense has to do is raise a reasonable doubt in the jury's mind. Consequently, even if the prosecutors believe in their case, they may lose, or they may find it advisable to engage in a plea bargain, for a lesser crime, instead of going to trial.

This brings us to another advantage of the "public welfare" approach: "Public welfare" offenses put the risk of failing to discharge a statutorily-prescribed obligation on the entities and individuals who have chosen to engage in the activity targeted by the statute(s) at issue. In "true crime" prosecutions, defendants can challenge the prosecution's claims of unlawful conduct and motives and may gain an acquittal; in "public welfare" prosecutions, they can challenge neither, which means, as the *Park* case demonstrates, they are very likely to be convicted if prosecuted. The only viable defense in a "public welfare" prosecution is for the defendant to prove that he or she simply did not have the power to prevent the conditions that give rise to the charges. Since this will no doubt be impossible to do, it creates incentives to ensure that an organization does not default on the statutorily-prescribed obligations to which it is subject.

What might a "public welfare" misuse of Consumer Data offense look like? It would not include *mens rea*, but it would have to identify the individuals and entities to which the offense applied and the types of Data it protected. The latter would drive the former, because the offense's restrictions would be imposed only on those who deal with the types of Data it encompassed. We believe that an offense limited to misuse of Data in which Consumers have a relation-based shared privacy interest would be sufficient because these are the only types of Data that can reasonably be said to constitute a systemic interest sufficient to support the imposition of *Park*-style liability.

A "public welfare" Data offense would therefore target the activities of Collectors within the constraints set out in the previous paragraph. The gravamen of the offense would be doing any of these things except in accordance with authorization provided by the person having the particular Data. A violation could take either of two non-exclusive forms: it could consist of allowing Data actually to be collected, used and/or disseminated

without authorization; or, adhering more closely to *Park*, a violation could (also) occur when systems and procedures of an entity were found to be inadequate to prevent such an unauthorized collection, use and disclosure. The first option resuscitates the "true crime" element of actual, individual "harm", but this is not uncommon in "public welfare" offenses; the advantage here is that someone (the victim) is likely to bring the violation to law enforcement's attention. The second option has the advantage of letting law enforcement intervene before an actual, individual "harm" has occurred; but it would require providing law enforcement with investigators who can identify these inchoate violations, just as the *Park* investigators identified the contamination in the Baltimore warehouse. The advantage of combining these options is that dedicated investigators can identify actual "harms" that have gone unnoticed by the victims or that they were unable or unwilling to pursue.

5 Conclusion

This Chapter has analyzed the extent to which stored transactional Data should be considered "private" and how two varieties of criminal liability—the "true crime" and "public welfare" offense approaches—*could* be used to secure that privacy. We have not argued that either *should* be used for this purpose. Criminal sanctions have always been and should continue to be extraordinary measures that are invoked only when recurring conduct inflicts a type of "harm" that is egregious enough to warrant the creation of a new offense and cannot be adequately addressed by the use of civil liability.

Which of the two possible types of criminal liability should be adopted? *Park*-style criminal liability seems to be the more promising strategy. A "true crime" approach suffers from some of the same problems as a civil liability approach: Both are reactive strategies; neither is available unless and until an individual "harm" to Data privacy has occurred. Both rely on the individual victim to discover that the "harm" has occurred and to seek redress for it. In civil context, the victim seeks redress by bringing private litigation, whereas in the criminal context he or she reports the matter to the authorities and cooperates in their investigation and in the prosecution of the offender. And both require that the party pursuing redress, whether the plaintiff in a civil suit or the prosecution in a "true crime" proceeding, establish culpable conduct on the part of the violators. The violators, in turn, are free to raise evidentiary and other issues to convince a trier of fact that the "harm," if one occurred, is not their responsibility.

Park-style criminal liability, on the other hand, implements a preventative, not a reactive, strategy. If it were applied to Data privacy, law enforcement investigators could monitor Collectors' procedures for securing Data and preventing its unlawful collection or use. Having identified actual or potential violations, the investigators could employ criminal liability to

sanction those responsible. An essential component of such sanctions would be remediation, i.e., requiring the violator(s) to implement procedures that will prevent further violations. The imposition of such sanctions should encourage other, similarly situated entities to monitor their procedures to ensure they were adequate to prevent such violations.

"Public welfare" offenses emerged a century ago to protect tangible items and activities. Given the increasingly critical role intangible items, including Data, play in defining the "public welfare" in twenty-first century America, it seems both reasonable and prudent to use this approach to protect them as well.

6 Questions on the material

1. Is privacy a "zero-sum" concept, or can different persons share a privacy interest in the same information?
2. Does society have an interest in preventing disclosure of data provided by Consumers and collected in commercial databases?
3. What role does the "pervasiveness" of information technology (IT) play in determining whether data should be treated as private?
4. Should society's concept of privacy be affected by the ability of Collectors to aggregate and "mine" data by means of database technology?
5. Do Consumers give up their right to privacy in data simply by providing it to commercial Collectors?
6. Of what significance should it be that a Consumer provides data to a Collector with an expectation that the confidentiality of the information will be maintained?
7. Under what circumstances, if any, should a Collector be permitted to disclose to government agents data obtained as a result of a confidentiality agreement or trust-based relationship?

7 Questions for further research

1. Should "mere economic" data be provided less protection than data relating to personal attributes or proprietary commercial data?
2. How can disclosures of private data by transferees of that data be controlled by the Consumer?
3. As technology becomes more pervasive and ubiquitous, should privacy of non-aggregated data be presumed?
4. What are the transaction costs of maintaining privacy of commercially collected data?
5. Should privacy be dealt with on an international basis rather than national one?

6. If regulation of privacy is not uniform, should the application of regulation depend on the situs of the Consumer, the Collector, or the Government desiring the information?
7. In cases of conflicts in regulation, should the most or the least strict regulation apply?

References

American Bar Association. (1993). Standards for criminal justice. Standard 3–2.1, Commentary.

American Civil Liberties Union. (2004). The Surveillance-Industrial Complex, http://www.aclu.org/Files/OpenFile.cfm?id=16225.

American Law Institute (1961). *Model Penal Code*. American Law Institute, Philadelphia.

Barber, M. (1992). Fair warning: the deterioration of scienter under environmental criminal statute. *Loyola of Los Angeles Law Review* 26, 105.

Bernoskie v. Zarinsky (2001). 344 N.J. Super. 160, 781 A.2d 52, 53–45 (N.J. Super. 2001).

Blau, J. (2004). German group studies ubiquitous computing, data privacy. Network World, http://www.nwfusion.com/news/2004/1222germagroup.html.

Brenner, S. (2003). Complicit publication: when should the dissemination of ideas and data be criminalized? *Albany Law Journal of Science and Technology* 13, 273.

Brenner, S. (2004). Toward a criminal law for cyberspace: distributed security. *Boston University Journal of Science and Technology Law* 10, 2.

Brenner, S. (2005). The fourth amendment in an era of ubiquitous technology. *Mississippi Law Journal* 74, 1.

Cyber Security Industry Alliance. (2005). Internet Security Voter Survey, https://www.csialliance.org/resources/pdfs/CSIA_Internet_Security_Survey_June_2005.pdf.

European Commission, IST Advisory Group (2001). Scenarios for Ambient Intelligence in 2010, 4–7, http://www.newscenter.philips.com/assets/Downloadablefile//ISTAG_scenarios-3146-1215.pdf.

Florida Annotated Statutes. Section 825.101(10).

Kannellos, M. (2005). These walls (and teddy bears) have eyes. CNET News, http://news.com.com/These+walls+and+teddy+bears+have+eyes/2100-1040_3-5738029.html.

LaFave, W. (2003). *Substantive Criminal Law*. Thomson West, St. Paul, Minnesota.

O'Harrow, R. (2005). *No Place to Hide*. Free Press, New York.

Oysted v. Shed (1816). 13 Massachusetts 520, 522–523 (Massachusetts Supreme Court, 1816).

Raisinghani, M. S., A. Benoit, J. Ding, M. Gomez, K. Gupta, V. Gusila, D. Power, O. Schmedding (2004). Ambient intelligence: changing forms of human-computer interaction and their social implications. *Journal of Digital Information* 5, http://jodi.ecs.soton.ac.uk/Articles/v05/i04/Raisinghani/.

Sayre, F. (1933). Public welfare offenses. *Columbia Law Review* 33, 55.

U.S. Department of Justice. (1997). United States Attorneys' Manual, Section 7–1.100, http://www.usdoj.gov/usao/eousa/foia_reading_room/usam/title7/1mant.htm.

U.S. v. Balint (1922). 258 U.S. 250.

U.S. v. Park (1975). 421 U.S. 658.

Ward, M. (2004). Smart homes offer a helping hand. BBC News, http://news.bbc.co.uk/1/hi/technology/3715927.stm.

Warren, S., L. Brandeis (1890). The right to privacy. Harvard Law Review 4, 193, http://www.louisville.edu/library/law/brandeis/privacy.html.

Winters, N. (2004). Personal Privacy and Popular Ubiquitous Technology, http://www.uclic.ucl.ac.uk/ubiconf/materials/Papers/Niall%20Winters.pdf.

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 2

National Information Technology (IT) Security Policies: An Overview of Issues

Karthik Venkataraman and H. Raghav Rao
State University of New York at Buffalo, Buffalo, NY, USA

David Dewitt
York University, Toronto, Ont., Canada

Abstract

The growth of the Internet and the dependence on information technology (IT) has significantly affected progress in the last decade. As the use of electronic data networks spread globally, many traditional communication barriers have been broken and have sped up efforts to achieve growth in developing countries. In this context, new forms of international cooperation by world's nations are required to construct an efficient and equitable global information infrastructure. The chapter discusses the development of comprehensive IT policies that can be modified to meet the needs of specific nations, and adhere to a standard set of protocols and regulations. Three countries are chosen—USA, Canada, and India, as examples, and the authors use the standards of IT security policies and the efforts of the respective governments to compare and synthesize the need, potential problems, and benefits of creating a comprehensive and international IT policy. Multilateral cooperation over the IT policies across borders via international organizations would be advantageous to achieve common goals of peace and development. In this context, an attempt has been made to identify commonalities in the experiences of various countries and recognize and reconcile their diversity to enable responsive policy protocols to be designed.

1 Introduction

From 2000 to 2005, the Internet expanded at an average rate of 146.2% eclipsing all previous technology growth patterns. An estimated 888 million people use the Internet today (Internet Usage Statistics—The Big Picture, Internet World Stats). Many countries are becoming dependent on their information technology (IT) infrastructures. With the world's communication and information services becoming truly global, data is transferred almost instantaneously. Several core systems such as defense, industry, and financial sectors depend on the national and private IT infrastructure that is in place. Most of this infrastructure is subject to security issues, which with the emergence of the Internet, and ready access to public domain infrastructure, has exploded into a national and international concern. Research shows that the number of security incidents has increased dramatically from 20,000 incidents per year in 1995 to approximately 150,000 in 2003 (Incidents Reported—CERT Coordination Center).

We live in a world that is full of varied opinions and viewpoints, some of which may lead to conflict while others engender cooperation for mutual benefit. The approach to the management of shared but scarce resources in a multipolar world by one country often defines the response by other countries. Economic and security cooperation are the most common, however cooperation in the IT security arena is equally important. One way to achieve this is by implementing a national IT security policy which, while governing the guidelines for ensuring security to a nation's resources and interests, also allows for mutual cooperation and development. However, encompassing all issues within a national IT security policy is challenging due to many factors. Issues such as widely divergent national security policies, interoperability between networks and infrastructure compatibility are factors that lead to problems. Alliances between nations which may include intelligence protocols do not necessarily reflect full disclosure or sharing of information. They give rise to issues such as withholding of information and selective information exchange. The Internet carries its own share of security concerns including viruses, worms, and hackers. This in turn could lead to economic problems resulting from espionage or attacks on the economic infrastructure of a nation.

This chapter identifies a template of possible issues or obstacles that need to be addressed by an IT security policy, to enable governments to implement a framework for their specific needs. As a part of the chapter, we observe the national IT Security policies for Canada, USA, and India. (On observation, the USA would be considered as the powerhouse of IT infrastructure security, India as the developing infrastructure, and Canada as the middle range in between.) Issues such as information sharing, interoperability, governmental interaction, and economic cooperation are examples of defining factors. It is important to examine how these factors will allow sharing of data and information between countries. Though it is a lofty goal

to find a common ground to draft a pseudo-international policy that governs the cooperative IT infrastructure of these countries, this chapter lays out a basic framework that can be helpful in drafting IT security policies.

2 National information technology security policies

With the arrival of the Internet, data communication has become second nature. Financial institutions and private corporations are but a few in the millions of members transferring information across the world. All of these data are transmitted from location to location across network infrastructures put in place by Internet service providers, telecommunication companies, private organizations, and governing bodies. The entire infrastructure including the hardware and software is normally regulated and governed by national governments. Internationally, countries have economies in different stages of development. Some of them are growing whereas others are stagnant or even declining. For some countries, the new economy is harnessed by strong IT concentration and development. Any compromise to a nation's data infrastructure can be devastating and has to be understood in a larger framework of the core sectors of modern societies. Cyber security is not a standalone issue concerned with high-technology networks; it is significant enough to encompass all walks of life and powerful enough to disrupt any service of the modern society. These services include banking, finance, and insurance; government services; communication and technology; infrastructure; emergency services; water and energy; health services; transportation/logistics/distribution. Therefore cyber security is important to be analyzed for the disruptions it can bring into sectors of any nation beyond the traditional areas of policing and defense.

Cyber-terrorism is a reality in the post-9/11 age. The US government agency, Federal Bureau of Investiation (F.B.I.) defines cyber-terrorism as "the unlawful use of force or violence against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives… through the exploitation of systems deployed by the target" (Conway, 2002; Denning, 2001). Cyber-terrorism is a powerful political and economic weapon. Shutting down the power grid or the banking system can seriously undermine political stability and damage a nation's economy. A cyber attack that has the capacity to shutdown the 9/11 emergency response systems used by fire, medical, and law enforcement could compound damage from a physical attack whether it is from domestic or foreign sources. There is also a potential overlap with cyber crime; groups could steal credit card numbers or important data and cause damage to others while profiting from the act themselves (Lilley, 2003). The lack of security on the Internet that makes users vulnerable also has implications for espionage and warfare. A sophisticated opponent might hack into a system and unobtrusively

either collect intelligence or plant code that could be activated in a conflict (Ward, 2003).

Considering the amount of disarray that can be injected by cyber-terrorism, most countries have implemented precautionary measures and guidelines that govern their cyberspace. It is because of the visible nature of these network infrastructures that we have seen the need for strict and extensive policies to govern these systems. The terrorist attacks of September 11, 2001 have alerted nations across the world to strengthen their security in all relevant areas, including their borders and networks (Center for Strategic and International Studies, 2002; Overholt and Brenner, 2001; Vatis, 2001).

This being the case, several countries have made concerted efforts at establishing full-scale organizations to handle cyber security issues (Dunn, 2005):

- Australia has built up several organizations and made them responsible for critical infrastructure protection (CIP). Cyber security is considered to be an integral part of the country's overall counter-terrorism effort. The Critical Infrastructure Protection Group includes the Defense Signals Directorate, the Australian Security Intelligence Organization, and the Australian Federal Police, each of which focuses on operational military, security, and policing intelligence services.
- Austria has its own specific security measures to defend against outside attacks and to prevent the unauthorized usage of data. Cyber security is mainly perceived as an issue of data protection, there is no single authority responsible for CIP/CIIP, as the Official Austrian Data Security Website indicates.
- Canada has an Office of Critical Infrastructure Protection and Emergency Preparedness now known as Public Safety and Emergency Preparedness Canada (PSEPC). This follows a centralized model and is the key organization responsible for CIP as well as Civil Emergency Planning.
- Finland views cyber security as a data security issue and as a matter of economic importance that is closely related to the development of the Finnish information society. Several organizations such as the Communications Regulatory Authority, the Emergency Supply Agency, the Board of Economic Defense, and the Committee for Data Security deal with cyber security.
- France sees cyber security as both a high-tech crime issue and as an issue affecting the development of the information society. Cyber security issues are dealt with by the Secretary-General of National Defense.
- Italy has no single authority dealing with cyber security. A working group on cyber security has been set up at the Ministry for Innovation and Technologies. It includes representatives of all ministries involved in the management of critical infrastructures, and many Italian infrastructure operators as well as some research institutes.

- The CIP project in the Netherlands aims to develop a set of interrelated measures to protect the infrastructure of government and trade industry in 11 identified sectors. The Ministry for Interior and Kingdom Relations coordinates CIP policy across all sectors and responsible ministries.
- New Zealand's Centre for Critical Infrastructure Protection (CCIP), located at the Government Communications Security Bureau, is the central institution dealing with cyber security. New Zealand's security policy, including cyber security, is formulated by the Domestic and External Secretariat (DESS, which is the support secretariat for the Officials Committee for Domestic and External Security Coordination (ODESC)).
- Norway's national key player in civil emergency planning is the Directorate for Civil Defense and Emergency Planning. This is also a key player for CIP-related issues. It is subordinated to the Ministry of Justice and Police.
- Switzerland has a number of different organizational units that deal with CIP. The Reporting and Analysis Centre for Information Assurance (Melde und Analysestelle Informationssicherung (MELANI)) in conjunction with public–private partnerships are among the central pillars of Switzerland's CIP policy.
- United Kingdom's National Infrastructure Security Coordination Centre (NISCC) is the key interdepartmental organization dealing with CIP. The private sector and the intelligence community have strong ties with NISCC.
- United States of America has the Department of Homeland Security (DHS) playing the leading role in CIP and cyber security. Public–private partnerships, e.g., Information Sharing and Analysis Centers (ISACs), are also given important roles in CIP.

In the next section, we focus on the national IT security policies for Canada, USA, and India. These are examples chosen for analysis since these countries represent well-established and growing economies. Drawing an analogy to the product diffusion models (Unknown), we can categorize the USA as an innovator, Canada as an adopter, and India as an imitator as far as conception and implementation of the security policies are concerned.

USA: USA is one of the oldest democracies in the world, currently the only post-Cold War superpower and one of the five permanent members of Security Council. It continues to have the single largest national economy and is the source of much of the technological innovation, which has driven economic and social change since the early 20th century. The expansive and open economy, its immigrant-based growth and development, and its role as a global political and military actor has brought not only wealth and prosperity for many but also has led the United States to become involved

in numerous military interventions the world over while also, more recently, making it the target of attack at home. With its security compromised, especially during and since the Cold War and most dramatically with 9/11, many of the policy initiatives of the American government have served as guidelines for other countries and organizations. As a result, the United States has been one of the principal leaders in devising effective measures to combat vulnerabilities in the IT field.

Canada: Canada is part of the G-7 countries, a commonwealth member, and currently occupies fifth position in the United Nations' Human Development Index (Human Development Index Report). Canada is a country that has largely leveraged the effective use of IT in governance rather than being the primary innovator in the IT field. A stated government IT policy (Management of Information Technology Policy) is to use IT in renewing the way business is done by government.

India: By population India is the largest democracy and one of the fastest developing economies in the world. While its wealth and use of IT remains unevenly distributed with enormous problems of poverty and underdevelopment, it also has significantly increasing technological presence especially with its burgeoning software industry. It is a regional leader, and wishes to play a pivotal role in international affairs. The Indian economy, built on a strong science and technology education, research and development base, and a stable democratic political system, has been able to adjust quickly to new global political and economic factors, employing IT initiatives as a major instrument for growth and development.

Issues such as information sharing, interoperability, governmental interaction, and economic cooperation are examples of defining factors that need to be looked at in the development of policies. Though it might be difficult to draft a transnational policy that governs the cooperative IT infrastructure of each country, a basic framework can be drafted and incorporated by other developing nations. By comparing the progress made by these three countries, we are able to synthesize the core aspects of the problems faced by nations in different phases of their technological, social, institutional development. Our intent in this chapter is to contribute to thinking about common framework, which, while being sensitive to the differences among countries, would be applicable irrespective of the stages of technological progress.

2.1 United States of America: The innovator

The US is among the most IT-dependent countries, and has the most concerns about information warfare and national security dimensions of cyber security. It has both the private and public sector working hand in hand in cyber space. Information transfer is almost instantaneous and has increased in volume. This amount of growth fosters more need for control and monitoring. Post 9/11, existing cyber security policies were seen to be

inadequate. Since then, the United States has been cast into the role of policing its cyber infrastructure both in the public and private sector.

After 9/11, the United States government merged all its domestic security needs into one organization, the DHS. This branch of the federal administration governs and maintains the policies and guidelines regarding security measures undertaken (including cyberspace). The DHS has implemented the President's National Strategy to Secure Cyberspace. The Homeland Security Act of 2002 created the National Cyber Security Division (NCSD) under the Department's Information Analysis and Infrastructure Protection Directorate. The NCSD provides for 24×7 functions, including conducting cyberspace analysis, issuing alerts and warnings, improving information sharing, responding to major incidents, and aiding in national-level recovery efforts. This division represents a significant step toward advancing the US Federal government's interaction and partnership with industry and other organizations in this critical area. The NCSD tries to identify, analyze, and reduce cyber threats and vulnerabilities, disseminate threat warning information, coordinate incident response, and provide technical assistance in continuity of operations and recovery planning (Office of the Press Secretary, 2003).

The NCSD builds on the existing capabilities transferred to DHS from the former Critical Infrastructure Assurance Office, the National Infrastructure Protection Center, the Federal Computer Incident Response Center, and the National Communications System. The creation of the NCSD both strengthens government-wide processes for response and improves protection of critical cyber assets through maximizing and leveraging the resources of these previously separate offices. The division is organized around three units designed to:

- Identify risks and help reduce the vulnerabilities to the government's cyber assets and coordinate with the private sector to identify and help protect America's critical cyber assets.
- Oversee a consolidated Cyber Security Tracking, Analysis, and Response Center (CSTARC), which will detect and respond to Internet events; track potential threats and vulnerabilities to cyberspace; and coordinate cyber security and incident response with federal, state, local, private sector, and international partners.
- Create, in coordination with other appropriate agencies, cyber security awareness and education programs and partnerships with consumers, businesses, governments, academia, and international communities.

Consistent with law and policy, DHS's NCSD works closely with the Office of Management and Budget and National Institute of Standards and Technology regarding the security of federal systems. It also coordinates with federal law enforcement authorities, as seen appropriate. NCSD will leverage other DHS components including the Science and Technology Directorate, the US Secret Service and the Department's Privacy Officer.

The government of the United States of America aims to be the global leader with the vision to create a culture of security (The National Strategy to Secure Cyberspace, 2003).

The National Strategy to Secure Cyberspace was incorporated in order to present a formidable position for preventing cyber attacks and maintaining security for US interests. The strategy articulates five national priorities including:

- National Cyberspace Security Response System,
- National Cyberspace Security Threat and Vulnerability Reduction Program,
- National Cyberspace Security Awareness and Training Program,
- Securing the government's cyberspace, and
- National Security and International Cyberspace Security Cooperation.

The first priority focuses on improving response to cyber incidents and reducing the potential damage from such events. The second, third, and fourth priorities aim to reduce threats from, and vulnerabilities to, cyber attacks. The fifth priority is to prevent cyber attacks that could impact national security assets and to improve the international management of and response to such attacks (United States Embassy in Japan, 2003).

2.2 Canada: The adopter

Canada's national security policy is a strategic framework and action plan designed to ensure that Canada is prepared for and can respond to current and future threats (Operational Security Standard: Management of Information Technology Security (MITS)). The focus is on events and circumstances that generally require a national response as they are beyond the capacity of individuals, communities, or provinces to address alone. The former Prime Minister, Paul Martin initiated the security policy review and development, highlighting the importance of doing so while acknowledging the importance of protecting Canada's open, liberal-democratic ethos (Cross-Cultural Roundtable on Security). The national security policy articulated in *Securing an Open Society* and then contextualized in the then Liberal Government's foreign policy review, *Canada's International Policy Statement*, focuses on addressing three core national security interests:

1. Protecting Canada and Canadians at home and abroad.
2. Ensuring that Canada is not a base for threats to its allies.
3. Contributing to international security.

The national security document contains several measures to help build a more integrated security system that is consistent with the goals of the policy. The policy incorporates an Integrated Threat Assessment Center that ensures that all threat-related information is brought together, assessed and reaches

all those who need it in a timely and effective manner. The government also established a citizen-based advisory council called Cross-Cultural Roundtable on Security (Cross-Cultural Roundtable on Security), which is composed of citizens and security experts who are external to government ("Securing an Open Society: Canada National Security Policy, 2004).

The electrical blackout that affected Ontario and eight states in America in August 2003 demonstrated how dependent Canada is on critical infrastructure and how vulnerable it is to accidents or deliberate attack on cyber and physical security. Cyber attacks are growing concerns that have the potential to have an impact on a wide range of critical infrastructure that is connected through computer networks. The increasing complexity of the threats facing Canada required an integrated national security framework. The lack of integration in Canada's system was a key gap that was recognized by the Auditor General of Canada. Former Prime Minister Paul Martin announced a series of organizational changes that helped facilitate more effective integration, including the appointment of the Minister of Public Safety and Emergency Preparedness, with a new department supporting the core functions of security and intelligence, policing and enforcement, corrections and crime prevention, border services, immigration enforcement, and emergency management. A Cabinet Committee on Security, Public Health and Emergencies whose function was to coordinate government-wide responses to emergencies and to manage national security and intelligence issues, was also created. A National Security Advisor to the Prime Minister was also appointed to improve coordination and integration of security efforts among government departments.

Canada is currently building a fully integrated security system that is meant to ensure that the government can effectively respond to existing threats and quickly adapt to new ones, drawing from across government departments and agencies as well as levels of jurisdiction. Therefore the system is meant to be connected to key partners across levels of government, including provinces, territories, urban centers, and communities, as well as first line responders and the private sector. Canada's current approach to emergencies dates back to the Cold War era. The system is based on a highly decentralized and distributed division of responsibilities among first line responders, provinces, and territories, and lead departments at the federal level. There is currently a move for a more modern and integrated national support system. Interoperability of policies, systems, and personnel continues to be a major national challenge that must be tackled.

The Canadian government has identified two important elements for modernizing Canada's approach to emergency management. The first is an enhanced ability of law enforcement agencies to investigate cyber-incidents and other threats to national security. To this end, the overall statutory framework for the Government's emergency management activities—in particular the Emergency Preparedness Act—is being reviewed and modernized.

These requirements cover the areas of CIP, cyber security, disaster mitigation, information-sharing between federal departments, agreements with international and private sector partners, and protection of sensitive private sector information. Second, most of Canada's critical infrastructure is owned by the private sector or levels of government other than federal, and much of it is connected to international networks. To establish a basis for the federal, provincial and territorial governments, and the private sector to meet CIP challenges, the federal government has released a position paper setting out the key elements of the proposed Critical Infrastructure Protection Strategy for Canada (National Strategy for Critical Infrastructure Protection). The government plans to consult senior-level provincial, territorial, and private sector leaders to inform this strategy. Key international partners such as the United States are part of this consultation process. The government is working with provinces, territories, and the private sector to drive forward a national process that prioritizes substantial improvement of its national capabilities in CIP. To achieve a more proactive cyber security posture and to keep pace with the efforts of key allies, Canada plans to strengthen its capacity to predict and prevent cyber-attacks. It will substantially improve threat and vulnerability analyses for its systems, and strengthen its ability to defend its systems and respond to cyber-incidents. The government will also convene a high-level national task force, with public and private representation, to develop the National Cyber Security Strategy to reduce Canada's vulnerability to cyber-attacks and cyber-accidents. The federal government, ironically, is often not a lead player in emergency management (Securing an Open Society: Canada National Security Policy, 2004).

2.3 India: The imitator

India is a developing nation with a unique history of a relatively stable and functioning parliamentary democracy, and multilevel federal system, and an expanding economy driven by a diverse set of capacities, including advanced production and information technologies. It is the dominant economic force in South Asia. IT has been readily embraced by the private sector and is proving to be a booming venture for India's economy. Positioning itself as one of the leaders in outsourcing, many companies are looking to India for their information management needs. This is reflected in their rapidly growing infrastructure. Historically, India's infrastructure was hardly modern, stemming from archaic technologies passed down from developed nations. The 1990s technology boom and a technology conscious government enabled India to get on the cyberspace map. India has seen rapid improvements in telecommunications infrastructure, favorable government and tax incentives, and increased investments in technical education, along with an abundant supply of computer-literate people to make it attractive for offshore initiatives.

Government officials are serving as advocates and sponsors for offshore investment by undertaking a range of policy changes to encourage development of infrastructure. Since 1991, Indian government has embarked on liberalization of trade restrictions and has opened up several key sectors for private parties. Electronics and IT are the fastest growing segments of Indian industry both in terms of production and exports. Today, the electronics industry is completely de-licensed with the exception of aerospace and defense electronics. The liberalization in foreign investment and export–import policies of the entire economy, this sector is attracting considerable interest not only as a vast market but also as a potential production base by international companies. The Internet had been in India for many years in the form of ERNET. However, it was not easily accessible as it was meant for only the educational and research communities. In 1986, the Videsh Sanchar Nigam Limited (VSNL), a wholly government owned corporation was born, and in 1995 VSNL introduced Internet services in India. Seven years later in 2002, the Government of India, inline with their disinvestments plan, sold 26% of VSNL's equity to a strategic partner from private sector, which in turn stimulated competition and economic growth (Robinson and Kalakota, 2004).

Although Internet access was initially controlled by the government, there was no security policy that governed the cyber infrastructure until recently. The Government of India's focus on IT led to the creation of Department of Information Technology under the Ministry of Communication and Information Technology. With this new department, India will soon have benchmark norms for best practices in cyber security to enhance the country's security standards and improve its credibility in the global IT industry. Indian government introduced it first legislation and the Information Technology Act, 2000 came into force effective from 17th October 2000 (India's Information Technology Act, 2000). Indian companies already had robust cyber security practices, which they had adopted because of the nature of their business. A NASSCOM study has shown that almost all Indian companies offering IT services outsourcing had security policies that prevented malicious access, and anti-virus protection at multiple levels (Rules and Acts, 2003).

Furthermore, an external agency known as the Indian Computer Emergency Response Team (CERT-In) was initiated. This is a branch of an international forum of security policies governing cyberspace. The CERT-In operates under the auspices of, and with authority delegated by the Department of Information Technology, Ministry of Communications and Information Technology, and Government of India. The CERT-In works cooperatively, with information officers and system administrators of various sectors and organizational networks of its constituency. The purpose of the CERT-In is, "… to become the nation's most trusted referral agency of the Indian Community for responding to computer security incidents as and when they occur; the CERT-In will also assist members

of the Indian Community in implementing proactive measures to reduce the risks of computer security incidents." Its mission is to enhance the security of India's communications and information infrastructure through proactive action and effective collaboration. CERT-In will provide a single point of contact for reporting local problems. It is also geared to assist the organizational constituency and general computing community in preventing and handling computer security incidents. The CERT-In is also in the position to issue security guidelines, advisories, and timely advice. It plans to be a national repository of, and a referral agency for, cyber-intrusions. The objectives of CERT-In is to create awareness in the cyber community regarding information and computer security by issuing security guidelines and informing them of latest security threats, prevention measures, and solutions by issuing advisories and vulnerability notes (About CERT-In, 2004).

2.4 A comparison

Countries cannot exist without an infrastructure that allows them to be an entity in the global market, whether a member of the advanced industrial economies of the Organization for Economic Cooperation and Development (OECD) or part of the burgeoning developing countries in our contemporary world. Most countries have invested in some sort of network infrastructure and its citizens and businesses have made their ventures into cyberspace. Almost 15% of the world's population has Internet connectivity. Although in America, only an average of 70% of homes have Internet (http://www.internetworldstats.com/stats2.htm), the involvement is almost total (World Internet Usage and Population Statistics, 2005). The majority of business done worldwide is conducted in some manner over network infrastructures, and electronic business systems. Countries spanning every part of the globe have invested heavily in their infrastructure and embraced new technology readily. This has been an opportunity for entrepreneurs and small businesses spurring multinational investment and outsourcing. Developing nations such as India and many other Asian countries are seeing the economic benefit from this investment and are moving forward with examples learnt from their predecessors on the information superhighway. Whether a nation is a developed one or is developing, IT investment is a necessity for economic growth and for integration in the global context.

Although the threat of cyber terrorism may be the same for every nation in the world, the security policy they choose to follow can vary. Since 9/11 the United States government has been in a state of war and therefore is adamantly enforcing policies that are current and updated in order to maintain security. This is an increasingly contested position, with the Bush administration being challenged both by citizens' groups and from within Congress. Some argue that these policies, which include the Patriot Act,

infringe on the rights of its citizens. However the Bush administration has countered that these policies are in the interest of national security. In comparison both Canada and India are perceived to be more passive in their policies. Both these countries are concerned with their own security vulnerabilities as well as the effects that possible attacks on their allies would have on them. As the neighbor of the US, as each other's most important trading partner, having the most integrated military alliance, sharing the longest bilateral border and having much of their political, security, and economic factors integrated, shared, or coordinated, Canada is most directly affected by any aggressive actions taken against the United States on American soil. India, although thousands of kilometers away without any of the structural or systemic integration found between the US and Canada, is nevertheless entwined with American well-being in terms of future economic development, regional and global political stability, and the emerging importance of India being a major outsourcing ally to US business and IT.

The events of September 11th, 2001 required most western governments to examine and revise whatever security policies—both domestic and international—already existed. It soon became apparent, in light of the perceived global nature of the threat, that the countries of the Global South were not exempt. The *Al Qaeda* threat mobilized a range of responses, and even those governments sympathetic to the politics of fundamentalism had to adjust in order to manage their relations with other countries in their region and globally. While IT security and cyber-terrorism had previously been a recognized part of the critical infrastructure domain of the contemporary state and its economy, 9/11 soon heightened this awareness, not least because of the use of IT by the perpetrators of transnational terrorism but also because of the possible vulnerabilities of the cyber systems to attack.

Thus, like most others, Canada's cyber security and IT policies were developing, but with the events of 9/11 there has been a concerted effort to address these security vulnerabilities. India's policies, while also at least partially reactive to the 9/11 and post-9/11 events are further stimulated and enriched by the economic boom of offshore outsourcing from countries like the USA. The United States has had government guidelines and regulations in place for the last 30–40 years with implementations of standards for both academia and defense. India recently has created a Ministry for this purpose with its guidelines and policies developed by an external agency. This agency is working hand in hand with an international consortium of agencies to provide international advice and policy to its cyber infrastructure. Canada does have a strong IT infrastructure that focuses on day-to-day business of the government and the private sector. The three countries have initiated a centralized governance of the cyber-infrastructure and are providing guidelines to protect their individual interests. The table below develops a comparison of key measures (Table 1).

Table 1
Comparison of key measures amongst three countries—USA, Canada, and India

USA	Canada	India
Governing body		
Department of Homeland Defense	Minister of Public Safety and Emergency Preparedness	Ministry of Communication and Information Technology
NCSD	National Security Advisory Committee	CERT-In
History		
NIST	Canadian government	Private sector
CIA; FBI	Private sector	VSNL
Science and Technology Directorate		
Scope		
All Internet and Infrastructure	Primarily Government systems	Private sector initiatives
Driving force		
Post-9/11 preparedness	International move for security from cyber terrorism	Economic boom of offshoring
War on terrorism		
Compliance		
Self-governing	Self-governing	International CERT compliance
Working with private sector	Working with private sector	ISO 17799
External advisors		
Private sector	United States government	Private sector
Security advisors	Local governments	International security organizations
	Private sector	

3 Internationalization

Cyber security is not exclusively a national concern. With the growth of the Internet and its far-reaching boundaries, there is the need for international integration. With its expanse of data networks over internationally sanctioned infrastructure, the Internet has made global communication a reality. Over the last 20 years new technologies have evolved especially in developed nations. With this growth there is the obvious cyber security issue that must be overseen. It is especially evident in countries that possess higher dependency on their infrastructure, which may include cyber networks and data systems. Due to the nature of the systems in place and the inherent

issues with security it is hard to replicate system frameworks among nations as sometimes, national security is dependent on the cyber security infrastructure in place. However international treaties can provide a more secure infrastructure sharing and integration of security systems across borders. Latch-on mechanisms for security frameworks that provide a measure of continuity between cooperating countries need to be identified and defined.

There is a scope for cyber security cooperation agreement between friendly countries that share common concerns, interests, and boundaries. The methodology and practices developed and adopted by one country will need to be codified before they qualify for grafting to make them usable in another country.

An international organization such as the United Nations can develop detailed guidelines for other countries utilizing the lessons learned from advance first movers of the technology. A good example of current infrastructure sharing and governance is evident in the European Union's setup over the last 10 years. Systems such as immigration and customs are very well integrated among the participating nations in the union. National security is more of a joint concern than that of an individual nation. Recently India and the United States launched into a new phase of cyber security cooperation. The US–India Cyber security Forum convened both government and industry representatives from each country to identify areas for collaboration in combating cyber-crime, cyber security research and development, information assurance and defense cooperation, standards and software assurance, and cyber incident management and response. Cooperation between the United States and India is of growing importance as the American government and corporations utilize IT companies in India very strongly. This is a good way to usher in a new era of international cyber cooperation (India and the United States Launch New Phase of Cyber Security Cooperation, 2004).

3.1 Developing the framework

A policy is typically a document that outlines specific requirements or rules that must be met. In the information network security realm, policies are usually point-specific and cover a single area. For example, an "Acceptable Use" policy would cover the rules and regulations for appropriate use of the computing facilities. A standard is typically a collection of system-specific or procedural-specific requirements that must be met by everyone. For example, one might have a standard that describes how to harden a Windows 2000 workstation for placement on an external (DMZ) network. People must follow this standard exactly if they wish to install a Windows 2000 workstation on an external network segment. A guideline is typically a collection of system-specific or procedural specific suggestions for best practice. They are not requirements to be met, but are strongly

recommended. Effective security policies make frequent references to standards and guidelines that exist within an organization.

Hence the questions arise, is a policy or guideline going to ensure complete security? Does a simple policy maintain the well-being of any nation? It is unlikely that total security can ever be implemented. This does not mean that we are doomed or that we do not have the means. What is being addressed is our ability to make policies that ensure the highest level of security possible. Security policies are useless unless they are properly followed and implemented. Security policies must be constantly updated and maintained with information on new threats such as sabotage or cyber-terrorism, harmful crippling viruses or worms, denial of service attacks, or hacked hardware failures. The goal for any governing body is to be a step ahead of the forces that intend to cause harm. As computers and the Internet continue to pervade human life in everything from automobiles to kitchen appliances, the risk of terrorism grows. Individuals or groups can now use cyberspace to threaten international governments, or to terrorize the citizens of a country. Therefore many nations are now including their cyberspace and network infrastructure in their scope of national security policies.

Any development of a cyber security framework has to include a comprehensive policy and guidelines involving the dynamics of various factors and strive to be flexible to invigorate the challenges from each of these different aspects. The five different focal points of significance in the field of critical information infrastructure protection (CIIP) (Dunn, 2005) are:

1. *Critical infrastructure:* The critical infrastructure sectors are identified by each country taking into consideration the challenges and significance of each of the sectors to its society. There can be a common ground in this, but the priority accorded to each sector may be different factoring many unique aspects of each nation.
2. *Organizational structures:* The organizational framework characterizes the specific responsibilities of the different functionaries at each level of governance for a nation. Organizational hierarchy needs to be devised and has to be clearly communicated to ensure the chain of command and also make sure that each entity is accounted for its role in ensuring the effectiveness of the framework. The state (federal) level such as ministries, national offices, agencies, coordination groups, etc. has to be accounted for along with the clear guidelines for the lower state level and private sector companies, industry, etc.
3. *Early warning and prevention initiatives:* The national organizations responsible for cyber-early warning, namely cyber security-related information-sharing organizations such as CERTs, ISACs, etc. have to be constituted and their effectiveness reviewed from time to time. It is also important to enhance and achieve a high degree of trust and cooperation between these agencies of each nation to warn, prevent, and mitigate any threats effectively.

4. *Issues in legislation and prosecution:* An essential approach to deter virtual abuse and other offences against the information infrastructure is the development of effective regulations, laws, and criminal justice mechanisms. Though the privacy and human rights aspects have to be taken into consideration, the rule of the law has to be upheld to counter the increasing sophistication of cyber-terrorism. Moreover, a strict regulation may create trust in the new ICT and encourage the private sector and individuals to make better use of e-Commerce or e-Government services.
5. *Research and development:* The major players in the field of cyber security R&D are US and the European Union. The crucial issue will be to promote cross-national R&D and information exchange in the field of cyber security and enabling all nations to reap the benefits of research enhancements. A more effective forum for the identification and promotion of relevant research topics in cyber security will be a major boost toward this goal of cross-border cooperation. The inherently transnational nature of the information infrastructure and the growing international dependency on these systems make the topic of cyber security R&D a significant issue for international cooperation.

One of the major hurdles is the issue of sharing information. In spite of being allies, it does not require friendly countries to handover data and proprietary ideas on the basis of cyber policy sharing. Similarly there are many issues that can be addressed but may not have a complete and binding solution. An "International Security policy" is a framework that can be tweaked or modified to suit each nation's needs (Table of Estimated Budgets of Countries, 2004). In this section we shall discuss some of the issues that are relevant to drafting a *National/International* security policy.

- *A National/International IT security policy is a cyber security issue:* With the advent of the Internet and its free domain for data sharing, there has been an exponential growth in compromised security across computer networks. It is now a reality that new virus, worms, and hacker attacks compromise data integrity and cripple network infrastructure affecting the governance and business. According to the Network Associates website, makers of the famous anti-virus program McAfee, there are currently over 100,000 viruses alone that are making their rounds across the millions of networks and workstations in the computer universe (Definitions of Viruses, 2005). Such evolving issues for any network, corporate or public, primarily calls for systems in place to prevent vulnerability. Standard software and firewalls with built-in redundancy along with active monitoring and continuous patch updates can act as immediate defenses. A well-defined policy and cooperation in the international forum could yield a comprehensive

and effective solution involving cooperation on research and development, implementation and sharing of information, and prosecution of anti-social and terrorist members involved. An international task force could help participating nations implement a security policy to govern the active prevention and monitoring of security vulnerabilities in their networks with active sharing of information on current viruses and hackers. Therefore we outline three critical issues below:

1. *One of the foremost problems for nations deciding on security policies is to address the problem of vulnerabilities.*
2. *Hackers break into computer systems or networks, destroy data, steal copyrighted software, and perform other destructive or illegal acts. Access to classified information related to strategic national resources like military installations, power systems, and security systems can be dangerous if terrorist organizations take advantage of this. Policies regarding classified information need to be in place at all levels.*
3. *Distributed denial of service attacks are often attacks across national borders. As such they are examples of vulnerabilities showing the impact and the dependence on the network infrastructure across nations. The ability to act proactively to prevent, and the timely recovery from, such attacks are crucial. Perhaps it would be a prudent idea for developed countries to provide the financial backing for under-developed and developing nations to create national security policies that can counter such attacks.*

- *A National/International IT security policy is a military security and defense issue:* It is in a nation's best interest to hide and keep secret its military information ranging from maneuvers, top secret weapons to facilities and personnel involved. All nations therefore maintain military systems separate from public domain networks. However, the need to overlap these systems to provide tactical support and information sharing among friendly allied nations is increasing. An example of this need is the role of multinational troops in the Iraq war and Afghanistan where non-NATO members became allies to fight common enemies. Information sharing between the participating nations without jeopardizing data security is the key to successful execution of combat strategies. During these times it is important to create and draft temporary policies that govern such information transfer and also encompass the networks and infrastructures that will be used in such operations. Further issues include:
  1. *Establishing frameworks for information sharing in defense issues*
  2. *Creating secure network connectivity to allow information exchange in peace and combat periods*
  3. *Necessary goodwill to avoid charges of eavesdropping and spying between cooperating nations to ensure success of such arrangements and sustaining it over time.*

- *A National/International IT security policy is an interoperability issue:* The disparity between IT infrastructures and the willingness to share the technological know-how often create barriers in building cooperation. The priority accorded by a nation toward building the network infrastructure and the financial capability of that nation to do so are the two major factors in this regard. Developed countries like USA and Canada are capable of funneling huge finances to new technologies, while developing countries often cannot afford to do so. India is a special case in this discussion where infrastructure is gaining growth by contributions from the private sector. Another important factor toward the development of a well-planned infrastructure that provides interoperability is the use of commonly used technologies that move away from proprietary equipment. Many countries have national pride in providing jobs to their citizens. Therefore, it is important to deal with the issue of interoperability between participating nations and their infrastructures as well as standardization of the technologies while addressing implementation issues of secure protocols with encryption to provide secure transfer of data as pointed out here:
  1. *Ensure the interoperability of IT infrastructure to promote information exchange*
  2. *Developed countries should take lead in standardization of the basic information infrastructure allowing nations to cooperate better.*
- *A National/International IT security policy is a cultural and social issue*: Terminology and methodology can vary from land-locked neighbors let alone those who are oceans apart. In spite of such differences it is important for national policies to allow for friendly information sharing across infrastructure. As an example, USA and Canada are neighbors. Though at the outset their cultural and technological orientation seem to be similar, they differ in governmental structures, priorities, loyalties, foreign policies, role in UN, etc. Canada is still nominally headed by the British monarchy and has stronger ties with the United Kingdom. As governments, once again Canada has a socialist style of government. In order for these two countries to use a common infrastructure would require policies that allow the free flow of information that take into account the cultural differences. India is an extreme example, primarily because of the language barrier with a diversity that extends to 18 different languages spoken across its states. Though English is primarily spoken to do business, the language varies across geographical locations. Therefore:
  1. *IT policies should be based on a broad understanding of cultural and social differences between countries.*
  2. *IT policies should be flexible to accommodate different governance structures in implementation to reflect cultural priorities of the participating countries.*

- *A National/International IT security policy is an economic issue:* In the last 50 years the world has become a global village that has for the most part defined boundaries and is multipolar and economically interdependent. As with the events of September 11, 2001 or the December 2004 Tsunami disaster, it is obvious that any event across the globe has local and international consequences. However, economic strengths of countries are still a differentiating factor. Problems could occur when the financial capability of a participating nation cannot allow it to be an equal participant in the international scene. Developed nations should *enable* growth opportunities to encourage infrastructure development and economic growth. This discussion suggests that:
  1. *In formulating international security policies issues such as financial stature and ability should be addressed to ensure implementation feasibility. This can be achieved by requiring adequate budgetary commitments to participate in an international forum for data networks (What is Intellectual Property, 2004).*
  2. *Organizations such as the UN can mandate requirements and compliancy for nations to participate and benefit from IT on an international scale. Issues such as local tax structures and tariffs should be harmonized across boundaries.*
  3. *A study into the economic implications should be performed before implementing such international security policies.*
- *A National/International IT security policy is a structural issue*: Most governmental organizations have vertical organizational structures. This structure dictates that organizational units are in hierarchies that are top down when it comes to directive and authority. There is an issue with this organizational structure, when it comes to working across other vertically structured organizations. There are also problems for vertical structured organizations to work and be compatible with horizontally structured organizations. Interaction between different organizational structures leads to ''red tape'' and unnecessary obstacles.

From a national security policy level it is important to address the issue of organization structure and take into account the level of authority the concerned parties may have. In the United States in pre-9/11 times, independent security groups such as the FBI, CIA, and Secret Service worked independently of each other. With the development of the DHS this has some what converted the vertical nature of the organizations to have more horizontal governance, which enables them to easily transfer information and data. In countries such as Canada and India there are very few security organizations so this is less of a problem.

However the issue of governance and collaboration at the international level, particularly in the realm of cybersecurity is a much larger issue. It

is difficult to have everything under the purview of a single organization particularly when several different ideologies and interface issues are involved. A possible solution to enforce international security policies would be to have a blanket organization that governed the international implementation of a transnational IT security policy. For this purpose, within this organization we suggest a matrix organizational structure because it is difficult to see how organizations in different nations would report to an entity in a nation, or single entity even if it is a branch of the United Nations. Such a matrix would be able to better define decisional rights across nations and various organizations involved in security governance. Therefore we recommend that:

1. *The organizational structure in each nation has to incorporate the proven matrix to encourage cooperation between different departments in each nation. The vertical and horizontal structure of organization has to be reviewed and the most appropriate organizational structure needs to be determined for implementation on a wide-scale.*
2. *Once again an international organization such as the UN could be made to govern this and ensure compatibility between participating nations. Currently there are organizations that do ensure certain level of compliancy between nations. For example India enforces the CERT-In policy. CERT-In's guidelines for Secure Information transfer is followed by many other nations as well.*

- *A National/International IT security policy is an information sharing issue*: One of the prime issues in the implementation of international IT policies is the concept of what is shared. This problem is very important in the arena of military defense or financial information. Improper sharing can put businesses or the military in jeopardy and hurt the economies of the countries in which they are involved. However, there are peaceful countries willing to share more information if needed but are objected to by the paranoid. How does one determine what information can be shared? Does everyone have access? Does there need to be authority and clearance to get this information? These are questions that must be answered by the Security Policy that is implemented. Such policies regarding network access should be implemented on a national level for classified information, obviously on a bigger scale:
  1. *Cross border agreements on IT policies require governance structures and/or third party authority to implement regulations and authority for information sharing.*
  2. *All users under a IT security policy must comply and be aware of the use of accepted and authorized means to information sharing.*
- *A National/International IT security policy is an enforcement issue*: Different nations enforce laws using different methods. In the United States and Canada, law breakers are brought to justice after a long and arduous legal process. However the methods of deterrence may be

Table 2
Key factors in the framework for a cyber security policy

Issue	Description	Solution	Example
Content Security	Problems that are caused by viruses, worms, and hackers that danger information and allow access to classified information or systems.	Policy must incorporate a ''step-ahead'' methodology that can ensure preparedness and enforce constant monitoring to prevent vulnerability.	Compliance to Sarbanes–Oxley explicitly states the implementation of programs and control to prevent and detect fraud by intelligent content monitoring of all Internet activities including email, web traffic, peer-to-peer file sharing, instant messaging, Web-based email, email attachments, and bulletin board postings (Compliance to Sarbanes-Oxley).
Military Security	Certain military secrets and information can be shared by allied nations. Some of it can get into wrong hands.	Policy should ensure that authorized information is detailed and concerned parties are held responsible.	The Global information grid for information technology management (Issued by CIO of DoD on September 22, 1999) by US government outlines the security policy and forms the basis of further governance procedures and lays the foundation for integrating the different components together in this context (Global Information Grid).
Interoperability	Infrastructure can differ. Network systems may not work together. Methodology can differ.	Policy should enable a common format and compatible systems. International governance must be implemented to ensure bi-directional access to information.	The DoD Instruction 4630.8, Procedures for Interoperability and Supportability of Information Technology (IT) and National Security Systems (NSS), June 30, 2004 states the policies related to interoperability (''Procedures for Interoperability and Supportability of Information Technology (IT) and National Security Systems (NSS)'').

Culture and Social Situation	Cultural difference and language barriers can make information sharing and implementation difficult.	Policy should be clear and incorporate cultural sensitiveness to participating nations	The policy has to take into account the cultural perspectives as to how the affected users, owners of data, web page administrators, etc., will be affected as personal data storage and its repercussions often raise privacy concerns. The example of Fed government asking search companies to part with the search data was not well received by the sections of society which believes that government is intruding into the private lives of its citizens (Requirements vs. Firewall Designs).
Economic Structure	Financial power between nations can provide inequality in infrastructure strength and security.	Policy should enable equality of infrastructure. Treaties to ensure participating nations help one another financially and technologically.	The US–India Information and Communication Technology Working Group shows how the two countries can benefit each other in promoting economic growth by supporting IT standards that are open, interoperable, non-discriminatory, and demand-driven (Enhancing the US–India Economic Dialogue).
Organizational Structure	Difference in horizontal and vertical structured orgs and their working together.	Policy moves all participants to a Horizontal Structure of Organization.	The policies in geo-spatial intelligence are evolving by horizontal re-structuring of organizations. Horizontal integration, across all disciplines provides an opportunity to effectively increase the collection resources for intelligence community and is an example of horizontal integration in predominantly vertical structure embracing communities (The Next Evolution of Geospatial Intelligence).

Table 2. (*Continued*)

Issue	Description	Solution	Example
Information Sharing	Question of what type of information can be transferred.	Policy must give detailed description of what type of information is to be freely shared between parties.	The Justice Department is improving information sharing by the planned introduction of Justice Unified Telecommunications Network (JUTNet), a wide-area network that will handle data, voice, and images. Similarly Computer Emergency Response Team (CERT) in different countries is an effective mechanism of information sharing and is an efficient way to gather, assess, and disseminate information swiftly for government and the private sector alike (Justice to Improve Info Sharing, 2006).
Enforcement	Local laws can prevent equal punishment. Tax shelters can be exploited to prevent equality for infrastructure.	Policy should be enforced by committee approved by all participating nations.	FIRST is a premier organization and recognized global leader in incident response. Best practices and tools provided by such organizations for member countries can be adopted into the policy framework or can be a pre-cursor for adoption of such common international laws under UN framework to enforce critical laws to be abided by all member nations (FIRST Vision and Mission Statement).

different for policy breakers and law breakers in different countries. A hacker in Malaysia can decide to hack into a Canadian military database or a 15-year-old programmer can accidentally create a virus in his school lab in Birmingham, AL. Whatever the method or location of the security breach, it will have a global effect. Local law and national law may govern the method of punishment for such misuse of systems but should be overseen by national security policy. All enforcement should be implemented on an agreed method based on the nature of the crime, with participating countries having authority to determine any punishment and necessary measure to prevent further damage. An analysis of the recent examples of punishment history shows the lack of rigorous laws to prosecute those guilty of cyber security incidents. We, therefore, suggest:

1. *A rigorous IT legal policy has to incorporate the local as well as international concerns.*
2. *A broad legal framework can be developed which can serve as a guideline for many countries lacking exposure and experience to IT practices and issues as it is an evolving area.*

The above issues are just the tip of the iceberg that must be addressed in order to draft an easily adaptable National/International security policy that can be enabled and adapted by any interested nation. Below is a table describing a brief run down of the above-mentioned issues (Table 2). *The examples of each of these issues were prepared from the context of United States and may not be reflective of each country.*

4 Conclusion

It is an arduous task to make an IT policy that is all encompassing. However with adequate fieldwork and cooperation of nations, countries can achieve technology parity within a short time-frame. Organizations such as the United Nations can create an international Agency such as the UNESCO, WIPO, UNCTAD, or UNIDO, which can promote all IT-related developments from a global perspective. Possibly the most appropriate agency for such an initiative could be the UN Centre for Trade Facilitation and Electronic Business (UN/CEFACT).

Private organizations may also take the lead in this domain and provide a developed method that encompasses the overall requirements for a whole nation. A governing body could recommend guidelines and regulations for implementing such technology. Standardization can be achieved by conforming to the guidelines recommended by an international group. A general consensus may emerge if there is some initial development with worldwide seminars, symposiums, and tests. Every aspect of industry and commerce can benefit from a globally standardized infrastructure. Conflict

resolution can be achieved efficiently if the communication gap between nations is bridged. An international IT policy is possible if all involved see the need and have the will to make it happen.

The discussions in this chapter apply to post-conflict nations as well. While a detailed discussion is beyond the scope of this chapter, examples of Afghanistan and Iraq show our need to understand the post-conflict implementation of policies and infrastructure to allow these countries to benefit and turn a new leaf in the international arena. IT allows an efficient method to provide quicker growth to a stabilized state in this day and age. According to a World Bank survey (Carvalho & Melhem, 2005), in the 2 years after the fall of the Taliban in Afghanistan, private investors poured $130 million into telecommunications. Similarly the survey states that in October 2003 when three, 2-year phone licenses came up for bid in Iraq, more than 200 company consortia submitted bids. The willingness of the private sector to invest in IT when conflict ends not only helps business but creates a key government ally for restoring stability to conflict-affected countries. Less widely appreciated is the key role IT can play in improving government performance by linking local, municipal, and federal officials and systems in countries undergoing decentralization, such as Rwanda. Carefully planned implementation of technologies, starting with simple projects and then scaling up to more complex integrated systems allow for improved financial management and increased transparency and accountability—essential for countries trying to shed legacies of corruption and provide a public view of their systems (Bray, 2005). A policy should ensure internationally compatible infrastructures to enable proper flow of information. All security measures must be internationally monitored and all systems must be implemented to work not only overseas but with immediate neighbors which may include an upgrade of technologies in those neighboring countries. All government employees must be trained in use and regulations of the policy. Finally, policies for regular audit of systems must be implemented to ensure compliance and usability.

5 Questions

(1) Discuss the effectiveness of developing an IT framework for security policies for international implementation?
(2) How can the different factors which thwart the development of a common IT policy framework be addressed?
(3) Should UN handle the responsibility of forming a IT framework in the wake of increasing instances of cyber terrorism and dependence on IT in governance?
(4) How can countries cooperate in IT field, in spite of the conflicting foreign-policies in a world increasingly dependent on IT?

(5) Would the trade-interdependence between nations be helpful in future toward the goal of establishing a common security policy framework?
(6) Discuss the specific instances of cooperation between countries in IT field, which has benefited both countries?
(7) How can the privacy concerns of citizens be addressed when governments intrude into Internet search data stored by private corporations?
(8) How the adherences by each country to an international framework can effectively monitored? Discuss this by comparing the nuclear proliferation issues and its monitoring under UN.

Acknowledgments

The authors would like to thank the editors for their encouragement regarding this chapter. The research of the second author was carried out at York University, Canada, where he was a Fulbright scholar in Fall 2004. The authors would like to thank Pramod Kakkanath for research assistance.

References

About CERT-In (2004). From http://www.cert-in.org.in/
Bray, J. (2005). International Companies and Post-Conflict Reconstruction: Worldbank.
Carvalho, A., S. Melhem (2005). Attracting investment in post-conflict countries. From http://rru.worldbank.org/Discussions/topics/topic63.aspx
Center for Strategic and International Studies (2002). Cyber-terrorism and cyber-security [online]. Available on: http://www.csis.org/tech/pubs/cyber.htm
Compliance to Sarbanes–Oxley (2005). From www.s-ox.com/Feature/detail.cfm?ArticleID = 867
Conway, M. (2002). Reality bytes: cyberterrorism and terrorist 'use' of the Internet. From http://firstmonday.org/issues/issue7_11/conway/index.html
Cross-Cultural Roundtable on Security (2004). From http://ww2.psepc-sppcc.gc.ca/roundtable/index_e.asp
Definitions of Viruses (2005). From http://www.mcafee.com
Denning, D. (2001). Is cyber terror next? New York: U.S. Social Science Research Council. From http://www.ssrc.org/sept11/essays/denning.htm
Dunn, M. (10 June, 2005). A comparative analysis of cyber Security initiatives. Center for Security Studies, Swiss Federal Institute of Technology (ETH Zurich) WORLDWIDE for the WSIS Thematic Meeting on Cyber Security—International Telecommunication Union Retrieved 07/23/2005.
Enhancing the U.S.–India Economic Dialogue (2005). From http://www.usindiafriendship.net/viewpoints1/strategicpartnership.htm
FIRST Vision and Mission Statement (2003). From http://www.first.org/about/mission/mission.html
Global Information Grid (2002). From http://www.defenselink.mil/nii/org/cio/doc/gig7-8170-082400.pdf
Human Development Index Report (2005). From http://hdr.undp.org/statistics/data/country_fact_sheets/cty_fs_CAN.html
Incidents Reported–CERT Coordination Center. Retrieved 07/23, 2005. From http://www.cert.org/stats/cert_stats.html#incidents
India and the United States Launch New Phase of Cyber Security Cooperation (2004). From http://www.state.gov/r/pa/prs/ps/2004/38080.htm

India's Information Technology Act (2000). From http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan002090.pdf
Internet Usage Statistics—The Big Picture, Internet World Stats (2005). Retrieved 07/23/2005. From http://www.internetworldstats.com/stats.htm
Justice to Improve Info Sharing. 2006. From http://www.fcw.com/article91858-01-04-06-Web
Lilley, P. (2003). Credit card database hacked [online]. Available on: http://news.bbc.co.uk/1/hi/business/2774477.stm
Management of Information Technology Policy (1995). From http://www.tbs-sct.gc.ca/pubs_pol/ciopubs/tb_it/mit-gti_e.asp
National Strategy for Critical Infrastructure Protection (2005). From http://www.psepc-sppcc.gc.ca/prg/em/nciap/national_strategy-en.asp
Office of the Press Secretary (2003). Ridge creates new division to combat cyber threats [online]. Available on: www.dhs.gov/dhspublic/display?content = 916
Operational Security Standard: Management of Information Technology Security (MITS) (2004). From http://www.tbs-sct.gc.ca/pubs_pol/gospubs/TBM_12A/23RECON_e.asp
Overholt, M., and S. Brenner (2001). Introduction to cyber-terrorism [online]. Available on: http://cybercrimes.net/Terrorism/overview/page1.html
Procedures for Interoperability and Supportability of Information Technology (IT) and National Security Systems (NSS) (2004). From http://www.dtic.mil/whs/directives/corres/html/46308.htm
Requirements vs. Firewall Designs (2002). From http://fwtf.berkeley.edu/fwtf_report/Appendix_A.htm
Robinson, M., R. Kalakota (2004). *Offshore Outsourcing* 1st ed. Mivar Press, Alpharetta, GA.
Rules and Acts (2003). From http://www.mit.gov.in/
Securing an Open Society: Canada National Security Policy (2004). From Original—http://www.pco-bcp.gc.ca/docs/Publications/NatSecurnat/natsecurnat_e.pdf—Revision—http://www.pco-bcp.gc.ca/docs/ministers/deputypm/secure_e.pdf
Table of Estimated Budgets of Countries (2004). From http://www.cia.gov/cia/publications/factbook/fields/2056.html
The National Strategy to Secure Cyberspace (2003). From http://wbln0018.worldbank.org/html/FinancialSectorWeb.nsf/%28attachmentweb%29/USCyberStrategy/$FILE/US + Cyber + Strategy.pdf
The Next Evolution of Geospatial Intelligence (2005). From www.military-geospatial-technology.com/article.cfm?DocID = 1226
United States Embassy in Japan (2003). U.S.–Japan Joint Statement on Cyber Security [online]. Available on: http://japan.usembassy.gov/e/p/tp-20030909d2.html
Vatis, M.A. (2001). Cyber Terrorism: The State of U.S. Preparedness, September 26, 2001.
Ward, M. (2003). Cyber terrorism 'overhyped' [online]. Available on http://news.bbc.co.uk/1/hi/technology/2850541.stm
What is Intellectual Property (2004). From http://www.intellectual-property.gov.uk
World Internet Usage and Population Statistics (2005). From http://www.internetworldstats.com/stats.htm

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 3

Economics of Information Security Investment

C. Derrick Huang, Ravi S. Behara and Qing Hu

Department of Information Technology and Operations Management, College of Business, Florida Atlantic University, Boca Raton, FL 33431, USA

Abstract

Today, firms, large and small, are engaged in the defense against high-profile information security crimes and breaches. The economics of such investments, although often taking the backseat to the technical and procedural issues related to information security, are crucial to decision makers: given the fact that one may never achieve complete security, it is important that decision makers identify the optimal investment for their own circumstances. This chapter presents and analyzes the economics of information security investments in a systematic fashion. There are unique characteristics associated with information security investments, such as the imbalance between the offense and defense parties in a security event, and the free rider problems created by the interconnected nature of IT systems. We categorize the different types of security events based on the breach probability and the potential loss from a security breach. Appropriate economic theories are then employed to analyze applicable security cases—expected utility theory for common attacks, game theory for targeted theft of proprietary information, and cyberinsurance for information disasters—and studies in each category are summarized and discussed. Finally, we examine how these economic analyses can be applied to the three key steps of the management of information security investments: risk assessment, investment optimization and selection, and evaluation and monitoring of the investments.

1 Introduction

With the proliferation of Internet and other information technologies, information security has been a major concern of corporations large and small. This is partly due to the frequent occurrence of security incidents: in

a 2005 survey of 819 U.S. organizations, 68% report experiencing at least one electronic crime or intrusion (CSO, 2005). And the damages of such incidents can be high and far-reaching (Cambell et al., 2003; Farahmond et al., 2004). Based on a CSI/FBI survey in 2005, 639 respondents report a total of $130 million losses from information security incidents (Gordon et al., 2005). And it was estimated that firms experiencing breaches lost 2.1% of their stock value—an average loss in market capitalization of $1.65 billion per incident—within 2 days of the announcement (Cavusoglu et al., 2004b). As a result, firms are investing heavily in information and network security technologies to prevent IT security problems. It is estimated that US companies spent on average $196 per employee per year on security (Geer et al., 2003), and the total worldwide security spending is forecasted to reach $21.6 billion in 2006, a 17.6% growth per annum (AT&T, 2004).

Given the elevated level of information security risks brought about by the high probability of attacks and the potentially large losses associated with breaches, the focus has primarily been on devising the defenses from the technical and behavioral aspects of information security, and considerable advances have been made in those areas. The economics of such investments, however, poses considerable challenges to academics and practitioners alike. Because information security investments usually do not generate economic benefits in the sense of revenue generation or cost reduction, traditional economic analyses of the value of IT investment (Barua et al., 1991; Brynjolfsson and Hitt, 1996; Pavlou et al., 2005) as well as conventional accounting measures (Gordon and Loeb, 2002b) often do not apply to information security. Despite the difficulty, economic analysis is crucial in helping firms determine the optimal form and level of information security investment. Given the risky nature and the high level of uncertainty associated with information security, and the fact that a "completely secure organization" is an insurmountable, if not impossible, goal in today's networked economy, the critical question for a decision maker is, "What is the right amount of investment?" Investing less than that "right amount" will result in unacceptable security risk that could have been reduced. However, investments exceeding the optimal level do not bring justifiable returns in reducing security risks.

In this chapter, the economics of information security investment is presented and analyzed in a systematic fashion. We first discuss the unique characteristics associated with such investments, and establish a systematic way to categorize different types of security events based on their economic properties. This categorization then allows us to employ appropriate economic theories to applicable security cases. We summarize the findings in each category to date and discuss their practical implications. Finally, we examine how these economic analyses can be applied to the management of information security investments.

2 Characteristics of information security investment

Information security investments, be they technical or procedural, exhibit some unique economic characteristics. First, unlike most IT applications, whose benefits come from "making something happen," the value of information security investments lies in "preventing something from happening." Thus, while the payoff of the former can be calculated by, say, revenues generated or costs saved resulted from the use of the application, the return on information security investment can only be measured by the "avoided potential loss," which is indirect, unintuitive, and often problematic. The difficulty in measuring the return may partially explain the fact that, in the annual CSI/FBI survey, the percentage of respondents that use traditional metrics, such as return on investment (ROI), net present value (NPV), and internal rate of return (IRR), to quantify the benefits and costs of information security expenditures actually drop from 2004 to 2005, despite the increase in overall spending (Gordon et al., 2005). Even managerial performance evaluation systems such as the balanced scorecard do not apply to IT security very well.

Second, as long as the information systems of various parties and stakeholders are interconnected and information shared among them, the IT security problem cannot be isolated. IT security investment by one party is thus likely to influence and be influenced by the activities taken by others. One implication of the interdependent nature of information security is that the "free rider" problem, in which most "agents" (departments in a firm, for example) would usually not contribute to such efforts while expecting to enjoy others' contributions, is prominent in information security investment decisions (Varian, 2000). It has been found, for instance, that a higher degree of interdependency makes a firm invest less in security, despite the fact that doing so increases the risk it faces (Ogut et al., 2005). Also, gaining knowledge about such information as security breaches experienced by others can help a company plan for and defend against potential attacks. Indeed, it is found that there are strong economic incentives for firms to engage in security information sharing, and increased level of sharing also leads to higher social welfare (Gal-Or and Ghose, 2005). However, although information sharing reduces the optimal level of security investments by participating firms, it may also lead to underinvestment due to the aforementioned free-rider problem (Gordon et al., 2003a).

Finally, the state of "perfect security" does not exist. That is, no matter how much and how well a firm spends on information security, there is no guarantee that all the potentially damaging attacks and breaches will be blocked. This is illustrated vividly by Anderson (2001), where he shows that even very modest resources of the hackers can create huge threats to a corporation trying to fend off potential attacks, despite the latter's extensive monetary and technical resources. Given the fact that all firms have limited

resources, this lack of total security creates a constant struggle between "spending more for improved protection" and "spending enough to have adequate protection," and makes the quest for a level of *optimal* investment all the more important.

3 Economics of information security investment

In this section, we present the results of various studies on the economics of information security investment and discuss their theoretic and practical implications. Most of the ongoing research in this field is based on one of the three well-established theories in economics to evaluate information security investments. These are the expected utility theory, game theory, and insurance theory. Expected utility theory is an established approach to optimizing endogenous economic activities (the information security investment in this case), given exogenous factors (attackers' propensity in this case), to control risks caused by uncertainty (the probability of security breaches in this case) (Borch, 1968; Gerber and Pafumi, 1998). Game theory focuses on decision making under interdependent uncertainty when internal economic activities also influence the external entities (von Neumann and Morgenstern, 1953). Finally, when the uncertainty of a breach is not controllable by internal actions, insurance theory is concerned with the transfer of economic risks to a third party at a fair price (Borch, 1990). Together, the three theories address most aspects of economic decision making under uncertainty.

To address the issues discussed in the last section, the task of the economic studies based on those theories discussed above centers on the analysis of the optimal investment that can minimize the security risks in a cost-effective way. Following the definition by Schechter (2005) that security risks = (likelihood of security breach) × (cost of security breach), we classify the security events as well as the economic treatments by way of an analytic framework that highlights the two key parameters that characterize the security risks: the breach probability and the potential loss.

The breach probability describes the likelihood of a potentially damaging security event happening. It is not a given number; rather, it has to be derived (or estimated) from at least three factors (Huang et al., 2005a). First, it is a function of the probability of attack, an exogenous factor determined by the number, type, incentive, cost, etc., of the adversaries (hackers). Higher attack probability leads to higher breach probability. This attack probability is most likely beyond a firm's control, although sometimes the protective measures a firm implemented may lower the inclination for adversaries to attack, thus reducing the attack probability. The second factor is the intrinsic vulnerability associated with the IT systems deployed by the firm in question. Note that the IT system configuration is determined by a firm's business and operating requirements, and more connectivity and access result in higher level of vulnerability. Thus, when a

company's IT systems are accessible by its suppliers (such as Wal-mart's Retail Link), the vulnerability is higher than an isolated computer system. The breach probability increases with systems vulnerability. Lastly, the breach probability is also a function of the security measures that the firm deploys in order to reduce the vulnerability of the systems configuration and block the potential attacks. Such measure can take many forms, from technologies such as antivirus software, firewalls, and intrusion detection systems (IDS), to procedures such as user's security training, to policies such as authentication procedures and official policy for information security. Ideally, the more the firm spends on information security, the lower the breach probability.

The potential loss specifies the amount of damage that is likely to incur in a security breach. It can be regarded as the incentive for a firm to employ security measure: the higher the potential loss, the more likely a firm is to invest to protect against it. The potential loss can be short term (e.g., lost revenues due to a denial of service attack) or long term (e.g., loss of customers to competitors with better IT security), and it can be tangible (e.g., lost sales) or intangible (e.g., damaged reputation). Figure 1 lists examples of the potential losses. As in all economic analysis, different types of losses are converted into monetary terms for proper comparison and manipulation.

Based on these two key parameters, our framework for analysis is summarized in Fig. 2. It categorizes potentially damaging security events into four groups: "low-level security concerns," where both the potential loss and the breach probability are low; "common attack" with high breach probability but low potential loss; "targeted theft of proprietary information" with high breach probability and relatively high potential loss; and "information security disaster," where the breach probability is low but the

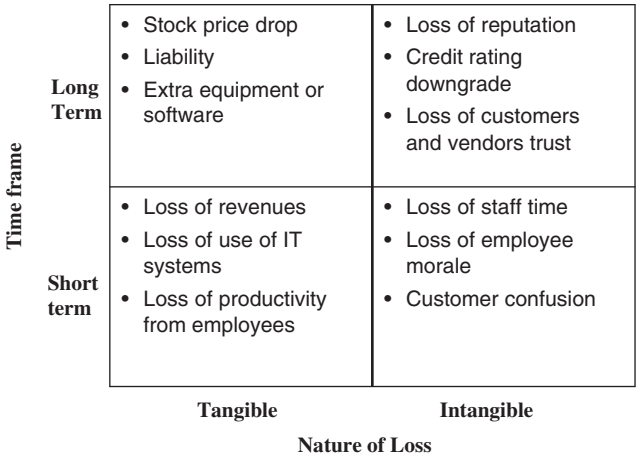


Fig. 1. Examples of potential losses from security breaches.

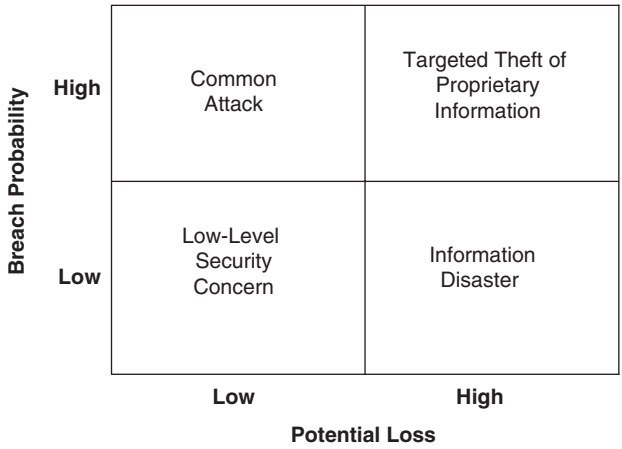


Fig. 2. Framework for economics of information security investment.

potential loss is very high. Below, we discuss the economics of information security investment associated with each type of the events, with the exception of the "low-level security problems," which, justifiably, often garner little management attention. The categories of the information security cases and the economics of the information security investment appropriate for each of the categories of security are summarized in Table 1.

3.1 Common attack

The security events that happen most often today are with relatively low (but not negligible) potential losses. These are the common attacks that corporations face on a daily or even hourly basis, such as virus, spyware, phishing, and spam email. The probability of these common attacks overwhelms other types of security incidents (CSO, 2005), but their consequences are generally limited: damages caused by such common attacks range from mere inconvenience, to temporary system downtime, to moderate monetary loss. Traditional decision analysis, where the optimal level of investment is determined via the risk-return analysis that is common in management and economics literature, can be used to analyze firms' investments to prevent this type of security problems. The expected utility theory (Fishburn, 1989; Friedman and Savage, 1952), a classical economic treatment, is well suited for such tasks.

Research in this stream is still in the initial stages, but the few studies have produced important results with real applications. Based on the assumption that a risk-neutral firm would maximize the expected benefits of the security investment by comparing the cost of the investment and the potential loss caused by possible security breaches, Gordon and Loeb (2002a) find that the optimal level of investment in information security does not always

Table 1
Summary of economics of information security investment findings

	Economic and risk characteristics	Applicable economic theory	Key results
Common attack	High breach probability, (relatively) low potential loss	Expected utility theory	• Optimal investment does not always increase with vulnerability. • Optimal investment is always smaller than the potential loss. • There exists a minimum potential loss below which the optimal investment is zero. • Higher level of risk aversion does not lead to higher level of security investment.
Targeted theft of proprietary information	High breach probability, (relatively) high potential loss	Game theory	• Investing in IDS provides a positive return to a firm only when the detection rate is higher than a critical value determined by the utility parameters of the attacker. • The optimal IDS configuration is to set the detection rate to the ratio of the benefit and cost parameter of the attacker.
Information disaster	Low breach probability, very high potential loss	Insurance theory	• Cyberinsurance and self-protection are complementary. • Cyberinsurance increases social welfare. • Insurers would face excessive high risk of simultaneous claims because of the dominance of a few IT platforms worldwide. • Sub-optimal premium pricing (too low *or* too high) leads to excessive risks taken up by the insurers.

increase with vulnerability, implying that a firm should not necessarily focus its security investment entirely on the most vulnerable information systems. Their analysis also shows that the optimal security investment would be far less than the potential loss if a security breach does happen, with a theoretical maximum of 36.8% of the potential loss.

Huang et al. (2005a,b) extend the above model by considering the risk profile of the decision maker of the firm in question and adopt the expected utility theory explicitly in finding the optimal level of information security investment. Assuming that the decision maker is risk averse—a commonly accepted assumption for firms with good performance (Fiegenbaum and Thomas, 1988; Jegers, 1991)—they find that there always exists a minimum potential loss below which the optimal information security investment is zero; above that minimum, optimal investment does increase with potential loss (Fig. 3). And, contrary to the risk-neutral case demonstrated by Gordon and Loeb, a risk-averse decision maker may continue to invest in information security until the spending is close to (but never exceeds) the potential loss in the case of common distributed attacks. Perhaps more interesting, their finding shows that a decision maker more averse to risk (i.e., with less appetite in accepting risk) does not necessarily spend more on information security. This finding runs counter to intuition, because the main benefit of information security investment is to reduce the corporate security risk. To explain this counterintuitive result, they argue that a decision maker faces many types of risks and has to balance them constantly. When one invests more, the risk of investment not performing ultimately outweighs the risk of information security. Thus, the optimal level of information security investment does not necessarily increase with decision maker's level of risk aversion. For managers, investing more in

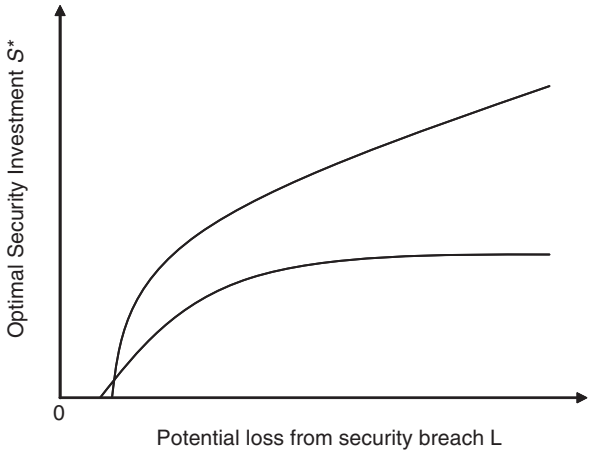


Fig. 3. Optimal security investment and potential loss.

information security, though effective in containing the security risk, does not always reduce overall business risks a company faces.

In the case of firms with interconnected IT systems, Ogut et al. (2005) find that the interdependency of risks reduces the firms' investment in information security to a level below optimum. However, when firms face liability in IT security—the breached firm has to pay others for collateral damages—they tend to over-invest above the optimal level. This is because each firm can only optimize its own utility, even though some part of the risks, introduced by the liability, is controlled by other interconnected firms. Consequently, the investment decision reached under such constraint is almost always less optimal than that where all interconnected firms plan and optimize their information security investment jointly. Thus, managers should be mindful of the potential overspending in the presence of possible third party liability.

3.2 Targeted attack against proprietary information

Many specialized, targeted attacks, such as those instigated by disgruntled employees to damage the company's intellectual property or the purposeful penetration into the bank's systems to transfer large amount of money by hackers, fit with the profile of high breach probability and high potential loss in our analytic framework. Such attacks, classified as the "targeted theft of proprietary information" in Fig. 2, tend to cause much higher damages to the targeted firms than do the common attacks such as virus or worms: per respondent loss from "theft of proprietary information" is three times that from virus in the 2005 CSI/FBI survey (Gordon et al., 2005). And, the probability of these attacks occurring is still relatively high, experienced by ~10% of the respondents in the aforementioned survey.

A solution for optimal investment for this type of attacks is available from traditional economic analysis, but because the optimal investment increases with potential loss (Huang et al., 2005a), when the potential loss is very high, the firm in question may not be able to afford the optimal investment due to its limited resources. While the traditional economic analysis is no longer adequate for these attacks, game theory (Friedman, 1990; von Neumann and Morgenstern, 1953) has been used to model the strategic interaction between a firm that protects its information sets and attackers intending to access or damage the proprietary information illegally. From a methodological perspective, game theoretic approach is best suited for modeling the outcome of a specific security technology with limited rounds (often two or three) of actions and reactions by a limited number of players (often the firm and the attacker). Using this approach to evaluate IDS, Cavusoglu et al. (2005) find that investing in such a technology provides a positive return to a firm only when the detection rate is higher than a critical value determined by the utility parameters of the attacker. And when the IDS is configurable, the optimal configuration is to

set the detection rate to the ratio of the benefit and cost parameter of the attacker. They further expand the result to other information security technologies and propose a model for making strategic decision in information security using a game tree approach (Cavusoglu et al., 2004a). Managers can use this methodology, with their own parameters, to determine what types, and how much, of security measures they should adopt.

Game theory as a methodology to analyze the economics of information security investment holds a lot of potential, because it takes into account the attractiveness of the information set and the cost of getting caught from the attacker's perspective, which is often ignored in the expected utility theory approach. However, the successful use of the game theory in this context depends heavily on the estimates of the attacker's utility parameters, which is a much more difficult task than estimating the utility parameters of the targeted firm. Indeed, Cavusoglu and Raghunathan (2004) show that the configuration of IDSs using game theory always results in lower cost than configuration using decision theory, as long as the firm can effectively estimate attackers' cost and benefit parameters and thus deter them from committing the intrusion with the IDS configured accordingly. This difficulty partially explains why there have been relatively few studies in this field using game theory. Nevertheless, it remains a powerful tool for analyzing the economics of information security investment, especially in the case of the targeted theft of proprietary information.

3.3 Information security disaster

When potential loss is relatively high, but breach probability is extremely low, the security risk, using Schechter's (2005) definition, can be quite small, and the investment in the security measures against this type of events may seem uneconomical and fall from the management's radar screen. But because of the relatively high potential loss that can be triggered by a breach, if one does happen (albeit highly unlikely), the consequence can be quite serious, or even devastating. A real-life analogy is observed in New Orleans of 2005. It was well know that the city's levee system could be vulnerable in a category four or five hurricane, but the possibility of one such hurricane severely affecting New Orleans was deemed small. When hurricane Katrina actually did in August 2005, the result was catastrophic.

The focus of both the expected utility theory and the game theory in information security investment is on reducing the breach probability. However, because the breach probability is already very small in this case, further reduction may be uneconomical, impractical, or outright impossible. The more appropriate approach would be to reduce the high level of potential loss caused by such a breach, and one common way of controlling risks through the reduction of potential loss is insurance (Borch, 1990; Kaas et al., 2001). Cyberinsurance, with which firms insure the most valuable corporate information assets against security breaches, offers a few advantages over

other potential loss control mechanisms. It allows firms to transfer the risk to a third party—the insurer—even when they do not *feel* adequate in their own protection. The insurer, through the use of deductibles and premium reductions, encourages insured firms to actually spend more on information security, therefore optimizing their self-protection measures (Ehrlich and Becker, 1972; Gordon et al., 2003b; Kesan et al., 2004). Further, the creation and functioning of a cyberinsurance market increases the overall societal welfare. By one calculation, the welfare gains associated with insuring worldwide security breaches and attacks could have reached $13 billion in 2000 (Kesan et al., 2005). For firms seeking protection against disaster-like information security incidents, Gordon et al. (2003b) offers a framework to employ cyberinsurance policies as a risk management instrument, something that all decision makers should consider for their information security needs.

In contrast to the benefits it brings to the insured, cyberinsurance poses significant challenges to the insurers. The most important is the difficulty in pricing. Insurance premiums are traditionally based on actuarial tables, constructed from years of data of the number and the amount of claims related to the incidents being insured. Such information clearly does not exist for information security. If the premiums are set too high, firms with lower risk of security breaches would not insure, creating the adverse selection problem, where only those with a high likelihood of an information security breach buy the cyberinsurance policies, leaving the insurer vulnerable to excessive claims (Gordon et al., 2003b). But if the premiums are too low, firms would use insurance rather than investment to manage their own information security risks, effectively transfer excessive risks to the insurers (Ogut et al., 2005). Another key issue is the structure of the IT industry, where only a few IT platforms dominate the market. Because many firms may use the same system platform, an attack targeted at some vulnerability of that platform may result in breaches experienced by many firms simultaneously. This scenario of highly correlated losses and claims prevent insurers from covering a large part of the market without premium surcharges (Böhme, 2005). Finally, the market for re-insurance—the secondary insurance insurers use to cap their losses—is not developed for cyberinsurance. In the case of an extensive information security breaches in multiple firms (similar to the damages caused by hurricanes), cyberinsurers can face extremely high level of claims and be *ruined* by such an event (Kaas et al., 2001). These are hurdles that the insurance industry has to cross in order for the cyberinsurance market to normalize and flourish in the coming years.

4 Management of information security investment

Amid the proliferation of studies on security technologies, the management and planning of information security have been identified as an important issue in information systems research more than a decade ago

(Niederman et al., 1991), and a stream of literature on this subject has since been established (Straub, 1990; Loch et al., 1992; Straub and Welke, 1998). Most of these have been based on the technical and procedural aspects of information security; ultimately, the economic analysis should add to this subject in developing a more complete framework in the management of the information security investment. There have been several such management systems and principles that invoke, implicitly or explicitly, economic principles (Bodin et al., 2005; Cavusoglu et al., 2004a; Dutta and McCrohan, 2002; Hoo, 2000). Dutta and McCrohan (2002), for instance, follow a traditional, sequential cost-benefit analysis, starting with identifying the assets and financial consequences and risks of a security breach, followed by estimating the cost of implementing proper mechanisms to enhance the security of the assets in question, and finally comparing the benefits of such mechanisms with the risks and estimated cost. Despite the various methodologies used in these studies, managing information security investment using economic principles involves three key steps: risk assessment, investment optimization and selection, and evaluation and monitoring of investment.

Risk assessment is the basis for deciding the level of needs in information security (Hoo, 2000). As discussed earlier, corporate security risk is the product of the breach probability and the potential loss (Gordon and Loeb, 2002b; Schechter, 2005). The breach probability has both internal and external components: the internal component can be derived from firm's own security audits of IT systems, procedures, and organizational structure (Dutta and McCrohan, 2002; Geer et al., 2003); and the external components would come from historic and cross-sectional data of security breaches (Geer et al., 2003) and an analysis of the adversaries' actions and reactions toward the firm's security measures (Schechter, 2005). The potential loss due to a security breach is sometimes more straightforward, because it mainly involves internal estimates of firm's information assets (Farahmand et al., 2004). However, the existence of third-party losses leads to complications such as liability and cost transfers. Alternative to this bottom-up approach of risk assessment, corporate security risks can be estimated top down by using such methods as simulation of security events (Conrad, 2005) or decision tree analysis of possible security threats (Sahinoglu, 2005). A decision maker should employ the most appropriate method—top down or bottom up—to estimate the security risk the corporation faces. The chapter by Behara and Bhattacharya in this book discusses risks and risk management further, and specifically develops a process-centric risk framework with which to understand and manage information security.

Having decided the security risk a firm is facing, the decision maker can move on to determine the appropriate selection and the optimal level of investment. One can do so with a cost-benefit analysis by laying out all the qualitative and quantitative costs and benefits associated with a particular investment (Dutta and McCrohan, 2002), a rating method of the analytic hierarchy process to determine the optimal allocation of information

security investment budget (Bodin et al., 2005), and a game-tree approach to model the three components of the IT architecture for security—prevention (such as firewall), detection (such as IDS), and response (such as manual monitoring)—that takes into account the interaction between a firm and potential hackers (Cavusoglu et al., 2004a). If the intrinsic vulnerability of the information set, the attack probability, and the potential loss in the event of actual attack can be estimated with fair accuracy, one can use the economic analysis methods developed by Gordon and Loeb (2002a) and Huang et al. (2005a,b) to determine the optimal level of security investment in different risk scenarios. There is also a procedural framework for companies to engage in cyberinsurance (Gordon et al., 2003b). Regardless of the methodology chosen by the decision maker, it is important to note that it does not make sense to invest more than the potential loss (Huang et al., 2005b), and the investment risks sometimes could outweigh the security risks (Huang et al., 2005a).

Evaluation of information security investment can be a difficult task. The main challenge comes from the fact that, contrary to other types of investment, a security investment is successful when "nothing happens." In other words, one needs to measure the *reduction* of security risks—lower breach probability and/or smaller potential loss—for evaluating the effectiveness of the information security investment. And traditional investment measures such as ROI, NPV, and IRR are really *ex ante* metrics that are of little value to evaluate the security investments *ex post* (Gordon et al., 2005). Arora et al. (2004) propose using the "bypass rate"—the percentage of adversaries capable of bypassing a security measure—as a way to measure the return on security investment. This bypass rate could be a useful metrics for security technologies, but would be a difficult metric to implement for other types of security investment such as procedures and policies. Some also have suggested that the return can be derived by measuring the business enabled by such investment, because information security is really a precursor to and enabler of e-business (Cavusoglu, 2004). To date, however, a satisfactory method of measuring and monitoring the performance of information security investment has yet to emerge.

5 Conclusion

Information security is no longer the exclusive concern of the traditionally high-risk sectors, such as federal agencies or defense companies, or the popular targets, such as banks and credit card companies. All companies, large and small, feel the importance of information security these days, and effective management of information security investment from an economic perspective becomes crucial in this environment. In this chapter, we show how the different types of economic analyses are applied to the various categories of information security cases. We also see that, as a result of these economic analyses, management systems and principles are devised to

help decision makers in determining the most effective selection and the optimal level of information security investment.

Notable progress has been made in this subject since the early 1990s, but we expect much more to come. All three major economic approaches to information security investment—expected utility theory, game theory, and insurance theory—are in their initial stage of applications. And many security investment problems—such as protection against simultaneous attacks, prioritization among common attacks and information disasters, and cyberinsurance premium pricing, to name a few—are not being addressed yet. However, with the growing use of the Internet and other information technologies, and the increased frequency and cost of security breaches, we believe that the subject of economics of information security information will grow extensively in the next few years.

6 Questions for discussion

1. What are the traditional economic measures of information systems investment? Do they apply well to investment in information security? Why and why not?
2. Compare and contrast the investment in information security and other types of information systems such as enterprise resource planning and data mining. What are the common characteristics? How are they different?
3. In this chapter, we use corporate security risk as a way to categorize different types of security breaches. Are there other possible classification schemes for analyzing the economic impacts of information security?
4. To apply the expected utility theory to the information security investment, the risk profile of the decision maker is an important factor in modeling. Would you characterize a decision maker of a firm as risk neutral, risk averse, or risk seeking, and why?
5. Utility parameters of a hacker, such as the rewards of successful breach and the penalty of being caught, are important input to the game theory analysis of information security. What are some of the ways that one can estimate the values of those parameters?
6. Security crimes committed by insiders are a major cause of information security breaches today. How would you analyze insider security crimes using the framework in Fig. 2 and apply appropriate economic analysis methodology?

Acknowledgement

This research is funded in part by a grant from the Defense Information Systems Agency of the Department of Defense.

References

Anderson, R. (2001). Why information security is hard: an economic perspective, in: *Proceedings of the 17th Annual Computer Security Applications Conference*, Purdue University, West Lafayette, Indiana, December 10–14.

Arora, A., D. Hall, C.A. Pinto, D. Ramsey, R. Telang (2004). Measuring the risk-based value of IT security solutions. *IT Professional* 6(6), 35–42.

AT&T. (July 2004). Network security: managing the risk and opportunity. *AT&T Point of View*, pp. 1–21.

Barua, A., C.H. Kriebel, T. Mukhopahhyay (1991). An economic analysis of strategic information technology investments. *MIS Quarterly* 15(3), 313–333.

Bodin, L.D., L.A. Gordon, M.P. Loeb (2005). Evaluating information security investments using the analytic hierarchy process. *Communications of the ACM* 48(2), 79–83.

Böhme, R. (2005). Cyber-insurance revisited, in: *Proceedings of the Workshop on the Economics of Information Security (WEIS05)*, Kennedy School of Government, Harvard University, Cambridge, MA, June 2–3.

Borch, K.H. (1968). *Economics of Uncertainty*. Princeton University Press, Princeton, NJ.

Borch, K.H. (1990). *Economics of Insurance*. North-Holland, Amsterdam.

Brynjolfsson, E., L.M. Hitt (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science* 42(4), 541–558.

Cambell, K., L.A. Gordon, M.P. Loeb, L. Zhou (2003). The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security* 11, 431–448.

Cavusoglu, H. (2004). Economics of IT security management, in: L.J. Camp, S. Lewis (eds.), Economics of Information Security, Kluwer Academic Publishers, Boston, MA, pp. 71–83.

Cavusoglu, H., B. Mishra, S. Raghunathan (2004a). A model for evaluating IT security investments. *Communications of the ACM* 47(7), 87–92.

Cavusoglu, H., B. Mishra, S. Raghunathan (2004b). The effect of Internet security breach announcements on market value: capital market reactions for breached firms and Internet security developers. *International Journal of Electronic Commerce* 9(1), 69–104.

Cavusoglu, H., B. Mishra, S. Raghunathan (2005). The value of intrusion detection systems in information technology security architecture. *Information Systems Research* 16(1), 28–46.

Cavusoglu, H., S. Raghunathan (2004). Configuration of intrusion detection systems: a comparison of decision and game theoretic approaches. *INFORMS Journal on Decision Analysis* 1(3), 131–148.

Conrad, J.R. (2005). Analyzing the risks of information security investments with Monte-Carlo simulations. In: *Proceedings of the Workshop on the Economics of Information Security (WEIS05)*, Kennedy School of Government, Harvard University, Cambridge, MA, June 2–3.

CSO (2005). *2005 E-Crime Watch^{TM} Survey: Summary of Findings*. CSO Magazine, Framingham, MA.

Dutta, A., K. McCrohan (2002). Management's role in information security in a cyber economy. *California Management Review* 45(1), 67–87.

Ehrlich, I., G. Becker (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economics* 80(4), 623–648.

Farahmand, F., S.B. Navathe, G.P. Sharp, P.H. Enslow (2004). Evaluating damages caused by information systems security incidents, in: L.J. Camp, S. Lewis (eds.), Economics of Information Security, Kluwer Academic Publishers, Boston, MA, pp. 85–94.

Fiegenbaum, A., H. Thomas (1988). Attitudes toward risk and the risk-return paradox: prospect theory explanations. *Academy of Management Journal* 32(1), 85–106.

Fishburn, P.C. (1989). Retrospective on the utility theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty* 2, 127–158.

Friedman, J.W. (1990). *Game Theory with Applications to Economics* 2nd ed. Oxford University Press, New York.

Friedman, M., L. Savage (1952). The expected utility hypothesis and the measurability of utility. *Journal of Political Economy* 60, 463–474.

Gal-Or, E., A. Ghose (2005). The economic incentives for sharing security information. *Information Systems Research* 16(2), 186–208.

Gerber, H.U., G. Pafumi (1998). Utility functions: from risk theory to finance. *North American Actuarial Journal* 2(3), 74–100.

Geer, D., K.S. Hoo, A. Jaquith (2003). Information security: why the future belongs to the quants. *IEEE Security and Privacy* 1(4), 24–32.

Gordon, L.A., M.P. Loeb (2002a). The economics of information security investment. *ACM Transactions on Information and Systems Security* 5(4), 438–457.

Gordon, L.A., M.P. Loeb (2002b). Return on information security investments: myths vs. *realities. Strategic Finance* 84(5), 26–31.

Gordon, L.A., M.P. Loeb, W. Lucyshyn (2003a). Sharing information on computer systems security: an economic analysis. *Journal of Accounting and Public Policy* 22, 461–485.

Gordon, L.A., M.P. Loeb, T. Sohail (2003b). A framework for using insurance for cyber-risk management. *Communications of the ACM* 46(3), 81–85.

Gordon, L.A., M.P. Loeb, W. Lucyshyn, R. Richardson (2005). *2005 CSI/FBI Computer Crime and Security Survey*. Computer Security Institute, San Francisco, CA.

Hoo, K.S. (2000). *How much is enough? A risk-management approach to computer security*. Working Paper, Consortium for Research on Information Security and Policy (CRISP), Stanford University, Palo Alto, CA.

Huang, C.D., Q. Hu, R. Behara (2005a). In search for optimal level of information security investment in risk-averse firms, in: *Proceedings of the Third Annual Security Symposium*, Tempe, AZ, September 8–9.

Huang, C.D., Q. Hu, R. Behara (2005b). Investment in information security by a risk-averse firm, in: *Proceedings of the 2005 Software Conference*, Las Vegas, NV, December 10–11.

Jegers, M. (1991). Prospect theory and the risk-return relation: some Belgian evidence. *Academy of Management Journal* 34(1), 215–225.

Kaas, R., M. Gavaerts, J. Phaene, M. Dennit (2001). *Modern Actuarial Risk Theory*. Kluwer Academic Publishers, Boston, MA.

Kesan, J.P., R.P. Majuca, W.J. Yurcik (2004). *The Economic Case for Cyberinsurance*. Law and Economics Working Paper, University of Illinois College of Law, Chicago, IL.

Kesan, J.P., R.P. Majuca, W.J. Yurcik (2005). Cyberinsurance as a market-based solution to the problem of cybersecurity—A case study, in: *Proceedings of the Workshop on the Economics of Information Security (WEIS05)*, Kennedy School of Government, Harvard University, Cambridge, MA, June 2–3.

Loch, K.D., H.C. Carr, M.E. Warketin (1992). Threats to information systems: today's reality, yesterday's understanding. *MIS Quarterly* 16(2), 173–186.

von Neumann, J., O. Morgenstern (1953). *Theory of Games and Economic Behavior* 3rd ed. Princeton University Press, Princeton, NJ.

Niederman, F., J.C. Brancheau, J.C. Wetherbe (1991). Information systems management issues for the 1990s. *MIS Quarterly* 15(4), 475–502.

Ogut, H., N. Menon, S. Raghunathan (2005). Cyber Insurance and IT security investment: impact of interdependent risk, in: *Proceedings of the Workshop on the Economics of Information Security (WEIS05)*, Kennedy School of Government, Harvard University, Cambridge, MA, June 2–3.

Pavlou, P.A., T.J. Housel, W. Rodgers, E. Jansen (2005). Measuring the return on information technology: a knowledge-based approach for revenue allocation at the process and firm level. *Journal of the Association for Information Systems* 6(7), 199–226.

Sahinoglu, M. (2005). Security meter: a practical decision-tree model to quantify risk. *IEEE Security and Privacy* 3(3), 18–24.

Schechter, S.E. (2005). Toward econometric models of the security risk from remote attacks. *IEEE Security and Privacy* 3(1), 40–44.

Straub, D.W. (1990). Effective IS security: an empirical study. *Information Systems Research* 1(3), 255–276.

Straub, D.W., R.J. Welke (1998). Coping with Systems Risk: security planning model for managerial decision making. *MIS Quarterly* 22(4), 441–469.
Varian, H. (2000). System reliability and free riding, in: *Proceedings of the Fifth International Conference on Electronic Commerce*, ACM Press, New York, pp. 305–366.

Glossary of terms

Analytic hierarchy process: the method for formalizing decision making where there are a limited number of choices but each has a number of attributes that are difficult to formalize.
Attack probability: the probability that a hacker or an adversary agent would attempt to compromise the information system in question.
Balanced scorecard: a management system that depicts and measures a firm's current operating performance, as well as the drivers of future performance, by tracking and measuring four dimensions of business: financial, customer, internal process, and innovation and learning. It was first proposed by Kaplan and Norton in 1992.
Breach probability: the probability that an attack against the information system would result in the information security being compromised.
Cost-benefit analysis: an evaluation methodology by comparing the costs of an activity to the benefits from that activity, taking into account the resource allocation among competing activities.
Cyberinsurance: insurance policy that protects a business from losses resulting from security breaches.
Denial of service attack: an attack characterized by an explicit attempt to prevent legitimate users of a network service from using that service. The most common method is to flood a network with useless traffic, overloading the network's capacity.
Expected utility theory: an economic theory for analyzing the behavior of a decision maker based on the expected value of his/her utility, a measure that represents the decision maker's preference under different combinations of risks and returns.
Game theory: a methodology concerned with how rational individuals make decisions when they are mutually interdependent.
Internal rate of return (IRR): the discount rate that equates the present value of future cash flow from a business activity to its initial cost.
Intrusion detection system (IDS): a system designed to detect security breaches after they occurred.
Net present value (NPV): a method of investment allocation based on the difference between the present value of future cash flow from a business activity to its initial cost.
Optimal investment: the level of investment in a business activity that generates the highest value (return vs. cost) of that investment.
Return on investment (ROI): an accounting measure derived by dividing the profits generated from a business activity by the cost of the investment.
Risk averse: the attitude toward risk in which one prefers a business activity with a certain return to one with an uncertain return when both activities generate the same expected return.
Risk neutral: the attitude toward risk in which one is concerned only with the expected return on a business activity.
Security breach: an event that information security is compromised.
Social welfare: the overall well being of a society (in sociology) or the utility of people considered in aggregate (in economics).
Vulnerability: weakness in the information system that make attacks on the system likely to be successful.

This page intentionally left blank

Part II:
Intelligence and Security Informatics

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 4

State of 3D Face Biometrics for Homeland Security Applications

Anshuman Razdan

Division of Computing Studies, Arizona State University at Poly Campus, AZ 85287-0180, USA

Gerald Farin

Department of Computer Science and Engineering and Director PRISM, MC8609, Arizona State University, Tempe, AZ 85287-8609, USA

Myungsoo Bae and Mahesh Chaudhari

PRISM, MC8609, Arizona State University, Tempe, AZ 85287-8609, USA

Abstract

Biometric access control focuses on measurable physiological or behavioral trait to automatically and accurately authenticate or verify the identity of that person. Biometric characteristics must ideally be distinctive to the individual, easily acquired/measurable, and able to be compared or encoded for the security validation. The characteristics should change little over time (i.e., aging) and be difficult to change, circumvent, manipulate, or reproduce by other means. The mainstream biometric technologies use morphological feature recognition such as fingerprints, hand geometry, iris and retina feature scanning, and face recognition based on extracting characteristics of the face. Each of these except face recognition is either intrusive or fails in some cases (e.g., fingerprint identification is not possible for ~10% of population with indistinct fingerprints). Current face recognition technologies use algorithms ranging from heuristics to artificial intelligence to locate facial features, such as the eye sockets, cheekbones, and sides of the mouth, in two-dimensional (2D) black/white or color space. However, even the most sophisticated algorithms in 2D domain have a high failure rate. There has been an emergence of three-dimensional (3D) technologies to combat the shortcomings of biometrics based on 2D images. In this chapter we review the current state of the art in 2D and 3D face recognition/authentication followed by in-depth description of 3D face biometrics being developed at PRISM (Arizona State University) for homeland security application.

1 The need: biometric access control

Terrorist events of September 11, 2001 have brought about the realization that the United States needs stronger national security measures and protocols. The primary goal of access control systems is to restrict access to authorized personnel. The return on the investment provided is based on the benefits to personnel safety and offsetting other risks resulting from unlawful access, fraud, and theft. Access control applications include door access, time and attendance, keep-out during off-hours, and the control of sensitive and restricted access points.

Recognition in *one-to-many* searches seeks to identify an unknown individual based on comparisons to a database of known individuals (e.g., law enforcement, surveillance, and recent driver licenses). *Authentication* involves performing verification based on a *one-to-one search* to validate the identity claim of the individual (i.e., access control for a building, room, or for making a transaction at an ATM terminal). Authentication is in one sense a simpler process: comparisons are made only to the claimed identity, and a threshold of similarity is used to accept or reject the claim. In another sense, authentication is more difficult, because of the need to determine this threshold rather than using a ''best match'' criterion as in many face recognition applications.

Several approaches have been promoted to recognize and authenticate an individual or a group of people. Access control applications authenticate by physical appearance (by guard personnel, receptionist); by something the individual knows (pins, passwords); by something the individual has (lock/key, card, badge, token); by biometric evidence (a unique physiological or behavioral characteristic of the individual); or by a combination of both ''what one has'' (i.e., a card) and ''what one knows'' (i.e., their passcode). Most workplace entry points are typically controlled by a badge/card or by physical appearance. All of these methods, except biometrics, are fallible and can be circumvented, lost, or stolen. For that reason, biometrics have been explicitly cited in several pieces of U.S. legislation, including the USA PATRIOT Act (signed in October 2001), the Aviation and Transportation Security Act (signed in November 2001), and the Enhanced Border Security and Visa Reform Act (signed in May 2002). Each calls for the implementation of biometric technology to enhance homeland security.

Interest in authentication using biometrics is therefore growing dramatically. Biometric access control uses measurable physiological or behavioral traits to automatically authenticate a person's identity. Biometric characteristics must be distinctive of an individual, easily acquired, measured, and compared for purposes of security validation. The characteristics should change little over time (i.e., with age or voluntary change in appearance) and be difficult to change, circumvent, manipulate, or

reproduce by other means. Typically, high-level computer-based algorithms and database systems analyze the acquired biometric features and compare them to features known or enrolled in the database. The mainstream biometric technologies use morphological feature recognition such as fingerprints, hand geometry, iris and retina scanning, and two-dimensional (2D) and three-dimensional (3D) face authentication. Each of these except face authentication is either intrusive or fails in some cases (e.g., ~10% of population do not have good enough fingerprints). 2D face authentication, though less intrusive, has simply not attained the degree of accuracy necessary in a security setting. Facial biometrics on one hand present opportunities but at the same time acquiring correct biometrics from facial images continues to be an open research problem. Some issues are biometric that are invariant to pose and expression, excessive facial hair and glasses that cover large part of the face, data processing time, etc.

In this chapter we will cover issues primarily related to 3D facial biometrics in the context of authentication and recognition, although we cover 2D methods for review. The chapter is divided in the following sections: 2D methods, 3D methods with detail coverage on methods developed at author's lab and some discussion on construction of 3D facial databases. We conclude with future directions and references. We start with some definitions:

- *False Acceptance Rate (FAR)*: FAR is the rate determined by number of false acceptance against different people by total number of all fraud attempts against different people.
- *False Rejection Rate (FRR)*: FRR is the rate for number of false rejection for same people by total number of all verification attempts for same people.
- *ROC Curve*: ROC or the Receiver Operating Characteristic curve is a standard method used to plot the relationship between the FRR and FAR as a function of the threshold and therefore is informative about the efficiency of a system. Ideally both FAR and FRR should be zero but it is hard to achieve that in practice.
- *Equal Error Rate (ERR)*: The point in the ROC curve when FAR and FRR are the same. This is important since some algorithms tweak to improve either the FAR or FRR and usually at the cost of the other.
- *2D*: Usually images or 2D information.
- *3D*: Three-dimensional information such as a surface representing a face. Although there are many mathematical representations for a surface, a triangle mesh representation is the most common.
- *3D Face Scanner*: A data acquisition device such as a laser scanner, multiple camera-based system, etc. that is used to acquire 3D information of a face. Usually a face scanner will also capture the 2D image

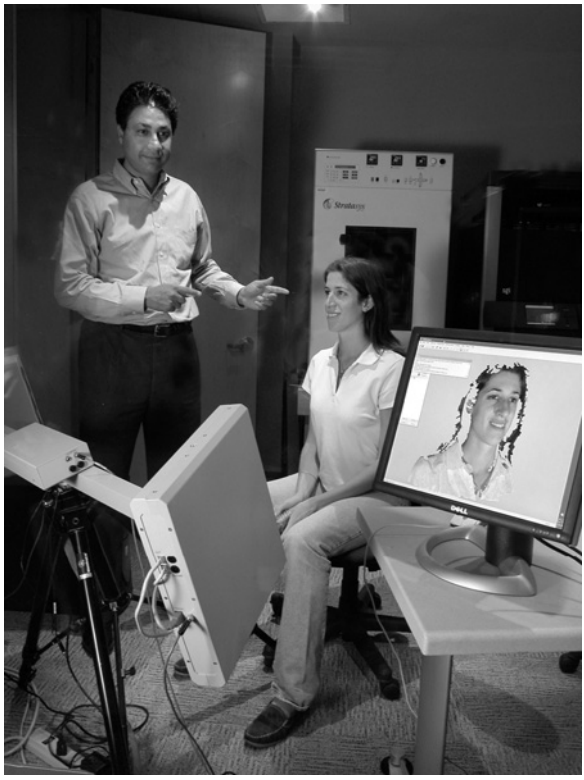


Fig. 1. A multicamera-based 3D face scanner used at the PRISM lab at ASU. The scanned data included both 3D and 2D image. Note the missing data corresponding to the hair of the subject.

or texture image and correspond the color values in the image to vertices of the 3D data acquired. Figure 1 shows a multicamera-based system being used in the author's lab at PRISM.

2 2D methods

Most 2D face biometric algorithms range from heuristics to artificial intelligence to locate facial features, such as the eye socket, cheekbones, and sides of the mouth, in 2D black/white or color space. However, even the most sophisticated algorithms in 2D domain have a failure rate that is too high for a security setting (note that identifying unknown persons—face recognition—has less stringent criteria for acceptable results). One early

study yielded false reject rates of up to 43% for images taken 18 months apart (Phillips et al., 1998, 2000a). A key limitation of 2D snapshots and the algorithms that work on them is variation in lighting conditions between images being compared. These techniques are "fooled" by the effects of lighting and orientation because they fail to take advantage of the rich variation of 3D geometric facial features. Some software and hardware have begun to emerge that claim to do 3D face recognition; however, almost all of these are merely estimating 3D surface topography from 2D information, and can still fail to provide the accuracy required by many security applications.

One of the commonly used and the best performing methods for face recognition is the approach based on principal component analysis (PCA), which was first presented by Turk and Pentland (1991). This approach transforms the initial training set of face images into eigenfaces. The vector of weights for every image in the gallery can be obtained by projecting into the eigenface components. Then, for new face image, its vector of weights is calculated, and compared with the vectors of weights in the gallery for identification. It has reported 96% correct classification averaged over lighting variation, 85% correct averaged over orientation, and 64% correct averaged over size variation on 2592 images (16 subjects, 3 orientations, 3 head sizes or scales, 3 lighting conditions). However, it is very sensitive to variations in pose, facial expression, size, or illumination. Many variations on PCA-based method have been proposed to make the method more robust to illumination, pose, and (or) expression changes.

Belhumeur et al. (1997) applied Linear Discriminant Analysis (LDA) (Fisher, 1936) to face recognition. This method also called the Fisherface method. LDA discriminates well between classes in a low-dimensional subspace, even under much variation in lighting and facial expressions. LDA maximizes the ratio of different classes while minimizing the ratio within objects in the same class. They used 330 images of 5 people (66 of each) for the experiment. The test result showed that the error rates from LDA method are much lower than those of PCA method but the experiment did not have a large dataset.

Pang et al. (2004) presented Gabor-LDA based face recognition method. From the training images, discriminant vectors are computed using LDA. From these vectors, intensity-LDA features (global features) and Gabor-LDA features (local features) are extracted and combined to produce the final classifier. Intensity-LDA features are LDA features extracted by projecting original intensity images on the discriminant vectors. Gabor-LDA features are obtained by LDA from the Gabor features which are extracted on discriminant pixels using Gabor filter (Duc et al., 1999) from the discriminant vectors. They tested 70 individuals with 6 frontal face images for each, and reported the recognition rate of 97% while other rates of 88.92, 92.5, and 95.11% in PCA, LDA, Gabor-LDA, respectively.

3 3D methods

We shall now direct our attention to methods that use 3D data as the basis for creating a biometric map of a face. Starting in the early 1990s theoretical and technical advances in 3D data capture techniques and mathematical modeling presented the opportunity to advance 3D knowledge into newer disciplines. Early scanners were laser-based and although harmless, took some time to scan an object. Many companies started developing camera-based systems (multiple cameras) to acquire 3D data in the late 1990s. The acquisition is instantaneous, although the data processing to create the 3D representation is not.

In 2002, National Science Foundation (NSF), under the ITER initiative, funded the 3D face authentication project (Face Authentication of Biometric Access Control, 2002) at PRISM[1] (Arizona Sate University). Information about the project can be seen at http://prism.asu.edu/3dface. The focus of the research is to develop intelligent and fast algorithms for representation, extraction, segmentation, query and matching of 3D facial shapes for authentication. A related goal is to build a digital library of 1000 3D human faces with multiple expressions over the course of the project and make it available to the research community.

Before we dive into the methods developed as part of this project, we present a brief summary of the literature in this area. The 3D matching problem boils down to a fundamental problem in geometry. Given two surfaces (in this case two faces), what can we say about the match (shape similarity) between the two? On one hand we are dealing with objects at the same scale (i.e., not trying to match the shape of a small pot with a large pot), but we are dealing with partial surfaces. By partial surfaces we mean that any data acquisition device may only capture part of the face, typically ear-to-ear, some parts of neck and clothing, and may not capture any hair at all. There may be facial hair (beard/moustache) that may give problem to the acquisition device. Further, two different scans of the same person may not be same due to change in pose (e.g., head tilted) and expression (smiling versus serious). Skin is not a rigid body hence any matching system must allow for that flexibility. Therefore, a large part of the matching algorithms are devoted to finding areas of the face that are more or less invariant to pose and expression, is not affected by facial hair, and still can find enough biometrics to make it unique.

3.1 Review of 3D methods

To solve the surface registration/matching problem, Besl proposed the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). This

[1] http://prism.asu.edu

algorithm transforms a target surface iteratively until it registers with the model surface. This method requires a reasonable initial guess close to the global minimum.

Medioni and Waupotitsch (2003), reported EER of better than 2%, 0% false acceptance at 3% false reject rate on 100 subjects with 7 different poses each. Beumier and Acheroy (1998, 2000, 2001) have also presented a method for 3D face comparison. This method uses profile matching, either globally or more specifically for central and lateral profiles. For global matching, they use at most 15 profiles generated by intersecting parallel planes separated by 1 cm. For comparison, an Iterative Conditional Mode (ICM) is used to minimize the global distance between the profiles of two faces. In order to speed up the facial surface comparison, the automatically extracted central and mean lateral profiles are compared in the curvature space. The EER of ~10% (with fully automatic processing) and ~4% (with manual refinement) are reported.

Chua et al. (2000) proposed the 'point signature' approach (Chua and Jarvis, 1997). Their approach registers two surfaces by searching for three pairs of corresponding points between the target surface and model surfaces using a one-dimensional "signature" for each 3D data point in the set. A point signature is a local descriptor of the shape that encodes the distances from points on a 3D contour to a reference plane. This method was extended later on to only rigid region registration and refinement by the ICP (Chua et al., 2000). This approach, however, was tested with only six people.

Some feature-based approaches based on curvature analysis have been reported. Gordon (1992) extracts a set of features on the face using Gaussian and mean curvatures, and calculates feature vectors. This method was tested on 24 faces with 8 different people (with 3 views each). Recognition rates with this approach are reported in the range of 80–100%. Moreno et al. (2003) also extracts 86 feature vectors from the segmentation of the face using Gaussian and mean curvatures. Each feature is given with a weight which is determined by Fisher coefficients (Hallinan et al. 1999). Four hundred and twenty 3D images of 60 individuals were tested, and first 35 ranked features were selected to represent faces. They report 78% recognition success rate when best match is selected and a 92% recognition success rate when the five best matches are selected. Recently Zhang et al. (2006) also reported a profile-based matching system. For individuals with normal expression, the EER and rank-one recognition rate are 0.8 and 96.9% respectively but this was only tested on 32 individuals.

Tanaka et al. (1998) presented a correlation-based face recognition approach based on the analysis of principal curvatures and their directions. They calculate the maximum and minimum principal curvatures on a face, and extract valley and ridge lines from the curvatures. Then, extended Gaussian images (EGI) of ridge and valley lines are constructed by mapping each of principal direction vectors onto two unit spheres for ridge and

valley lines. A spherical correlation coefficient is used to estimate similarity between EGI's of two faces. The algorithm was tested with 37 face range images and 100% recognition was reported.

Chang et al. (2003, 2004) have developed a method using the PCA-based algorithm with 3D and 2D images. The approach combining 2D and 3D performs better than either 2D or 3D alone. They tested on a gallery of 127 images and an accumulated time-lapse probe set of 297 images. The face recognition rates with 98.7% by PCA using 2D and 3D, 100% by PCA using 2D, 3D, and thermal images are reported. However, they select landmark points manually (the eye centers in 2D, and two eye tips and center of lower chin in 3D) for facial pose normalization.

Bronstein et al. (2003, 2005) proposed an approach for a facial expression invariant 3D face recognition. Geodesic distances of a surface are invariant under isometric surface deformation. The surface can be represented as the bending-invariant canonical form obtained by multidimensional scaling (MDS) from the geodesic distances of the surface with the corresponding texture image. Then, the aligned canonical surface and flattened texture are obtained. The PCA-based algorithm is applied to canonical texture. This approach was tested on 220 faces of 30 people (3 artificial) including identical twins. An EER of 2% was reported.

Blanz and Vetter (1999, 2003) proposed the face recognition using 3D morphable model. The morphable face model is generated from the database of 3D scans. By fitting the morphable face model to 2D images, model coefficients, which represent intrinsic shape and texture of faces, are computed and are used for face recognition by comparing the coefficients. The correct identification rates of 95% on 4488 images and 95.9% on 1940 images are reported.

3.2 Overview of our 3D matching system

Our approach is a combination of feature-based and profile-based methods. We focus on finding accurate and robust features without any user intervention. The system is divided into three parts: data acquisition, feature extraction, and authentication or recognition as shown in Fig. 2.

The 3D dataset for this project is acquired with a combination of commercial scanning technologies and research software applications developed at PRISM. The 3D Face Scanner is a 2-pod scanning system from 3Q Inc.[2] This scanning system is quick enough to take a 3D face scan in comparison with the other laser scanners. The time taken for the cameras to operate is a fraction of a second while rest of the work is done by the software provided by the same company to generate a triangle mesh of the face, which takes a few seconds to complete (Fig. 1).

[2] http://www.3q.com/

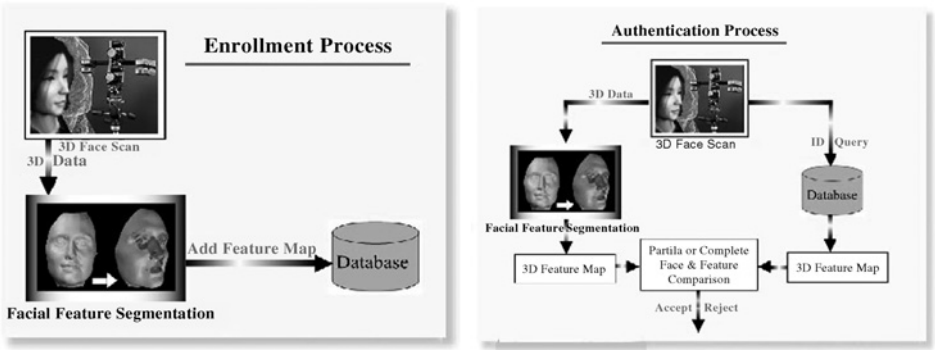


Fig. 2. General enrollment and authentication process.

The scanner does not give proper data on hair and glasses, a limitation of this and almost all other scanners in the area of facial biometrics. The triangle mesh representation is the most common way to represent 3D data from scanning systems. The feature extraction phase performs face classification using curvatures, registration of a face, finds symmetry plane, critical points, and profile curves, nose and biometrically relevant sub-face area extraction. Finally, authentication/recognition module performs comparison between two faces. We elaborate on these below.

3.2.1 Feature extraction

There are five steps in feature extraction, which include (1) surface point classification by curvatures, (2) approximated nose tip extraction and registration, (3) finding symmetry plane, (4) critical points and profiles extraction, and (5) nose and sub-face extraction.

Curvature estimation and surface classification. As a local representation, surface curvatures give us an overall picture of how smooth and uniform the object surface is. Curvature information helps to find facial features in our method.[3] Let $x(u, v)$ in terms of the curvilinear surface coordinates (u, v) be a point on a surface and let $n(u, v)$ be its normal. If any plane P passes through x which contains n, it will intersect the surface in a curve which is called the normal section of x with respect to x. At point x, we can compute the signed curvature which is called the normal curvature of the surface at point x with respect to the plane P. If we rotate the plane P around n, then we get new normal sections and normal curvatures at point x. The principal curvatures at x are the largest curvature, $k_{\max}$ and the smallest curvature, $k_{\min}$ of all the normal curvatures. Surface curvatures (Gaussian (K), mean (H), and absolute (K_{obs}) curvatures) can be computed as follows (Farin and

[3]For a deeper insight on the topic of surfaces and curvatures we recommend a classical text book on Differential Geometry.

Hansford, 2000; Farin, 2002):

$$\text{Gaussian} = K = k_{\max}\, k_{\min} = \frac{S}{F}$$

$$\text{mean} = H = \frac{1}{2}(k_{\max} + k_{\min}) = \frac{[nx_{vv}]x_u^2 - 2[nx_{uv}][x_u x_v] + [nx_{uu}]x_v^2}{F}$$

$$\text{absolute} = K_{\text{abs}} = \|k_{\max}\| + \|k_{\min}\|$$

where,

$$F = \det\begin{bmatrix} x_u & x_u & x_u & x_v \\ x_u & x_v & x_v & x_v \end{bmatrix}$$

and

$$S = \det\begin{bmatrix} nx_{u,u} & nx_{u,v} \\ nx_{u,v} & nx_{v,v} \end{bmatrix}$$

$$K_{\max} = H + \sqrt{\Delta}$$
$$K_{\min} = H - \sqrt{\Delta}$$

where, $\Delta = H^2 - K$, and the vectors$x_u$and x_v are tangent vectors respect to u and v directions respectively at the surface point x with normal vector n. Since the 3D surface representation is in discrete form (triangle mesh), estimating curvature in itself poses problems. The reader is referred to a paper by Razdan and Bae (2005) on this topic.

A point on a surface can be classified using Gaussian curvature and mean curvature (Besl and Jain, 1988; Farin, 2002; Moreno et al., 2003; Srinark and Kambhamettu, 2003). Each surface point is classified by signs of its Gaussian and mean curvatures as peak, ridge, saddle ridge, flat, saddle, pit, valley, or saddle valley. Table 1 (Srinark and Kambhamettu, 2003) shows the relationship of K and H and Fig. 3 shows examples of the surface classification.

Finding the nose point and the registration process. We do not assume any thing about the orientation (which way is up) or the tilt of the head in a

Table 1
Surface point classification by Gaussian (K) and mean (H) curvatures

	$K>0$	$K=0$	$K<0$
$H<0$	Peak	Ridge	Saddle ridge
$H=0$	None	Flat	Saddle
$H>0$	Pit	Valley	Saddle valley

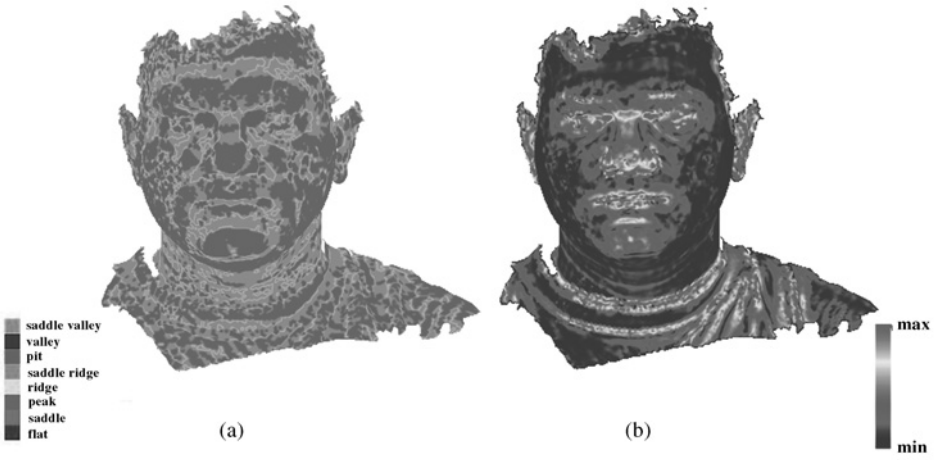


Fig. 3. Surface classification by Gaussian and mean curvatures (a) and absolute curvatures (b).

given scan or input data. Many algorithms require this as additional information before performing comparison but we feel that our methods do not need this additional information and are able to overcome issues of pose/orientation. This leads to the point that we need to find specific features such as the nose tip from a give scan. The scan may include parts of the shoulder, neck, etc. therefore it is important that any automated method should zero in on facial features without user intervention.

As mentioned before, the ICP algorithm (Besl and McKay, 1992) finds an accurate registration between two surfaces. This approach is very popular, and has been improved by researchers (Gelfand et al., 2003; Masuda and Yokoya, 1995; Rusinkiewicz and Levoy, 2001; Zhang, 1994) although it is computationally expensive and needs a reasonable initial guess close to the global minimum. Instead of registering the entire object, registering and matching surfaces by selecting a subset face has been used. Examples are: point signatures (Chua and Jarvis, 1997), harmonic map (Zhang and Hebert, 1999), spherical harmonic representation (Kazhdan et al., 2003) spin image representation (Johnson and Hebert, 1999), and surface point signature (SPS) (Yamany et al., 1999) method. We use the spin image representation as coarse registration and ICP method as a fine registration step. Spin images (Johnson and Hebert, 1999) are 2D histograms of the surface locations around a point. Let p be an oriented surface point for a spin image, and p_k be any 3D point on the surface. Let $d2(k)$ be the distance between p_k and p'_k which is the shortest distance from p_k to the tangent plane of p, and $d1(k)$ be the distance between p and p'_k. $d1$ and $d2$ form a local coordinate system $(d1, d2)$ for the 2D histogram, and $d2$ can be positive or negative while $d1$ is always positive. The term spin image is used to

refer to the result of applying the spin map (S_p at a surface point p), which is defined as follows (Johnson and Hebert, 1999):

$$S_p : R^3 \rightarrow R^2$$

$$S_p(p_k) \rightarrow (d1(k), d2(k)) = \left(\sqrt{||p_k - p||^2 - n \cdot (p_k - p)^2}, n \cdot (p_k - p) \right)$$

where n is the normal vector of p.

Spin image is a 2D array where each bin of a spin image counts total number of points in it (Fig. 4b). In our approach, the spin image of a point is generated from a neighborhood of points within a threshold (4.5 cm in our case) distance. Figure 5 shows local spin images of facial surfaces.

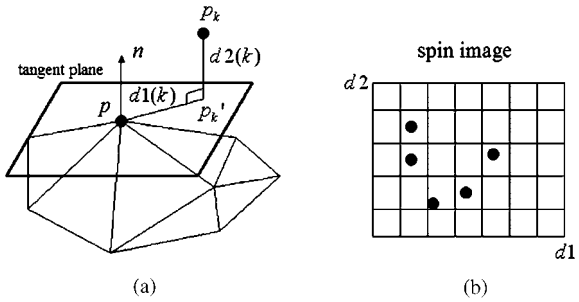


Fig. 4. (a) Oriented point; (b) spin image.

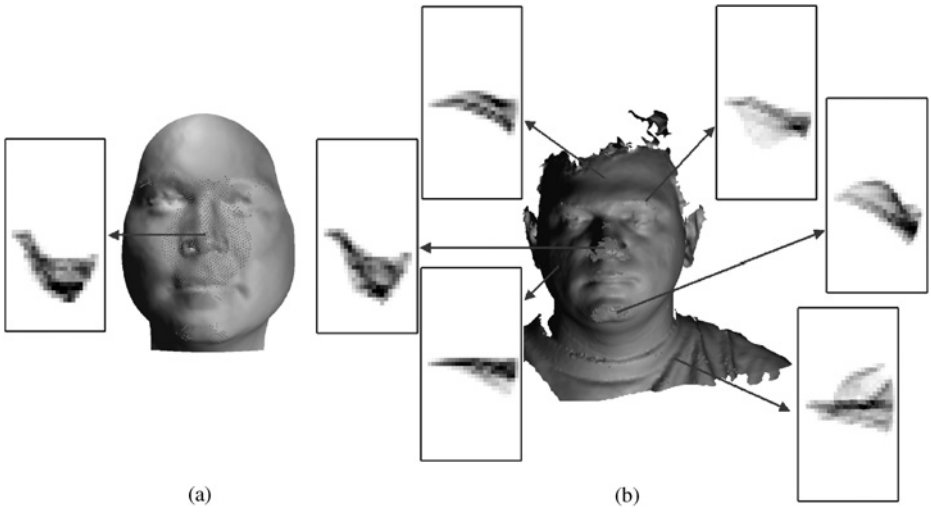


Fig. 5. Spin images. (a) Model face and the spin image on the nose tip with local area (red points); (b) test face and the spin images on surface points. The green points on (a) and (b) are the candidates for nose tips.

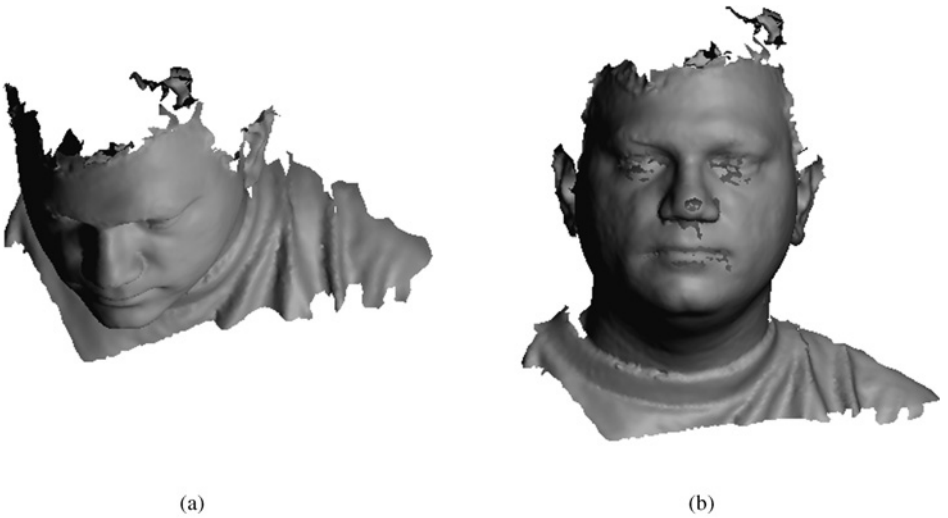


Fig. 6. Registration (a) before registration (b) after registration with the model face (purple) (coarse and fine registration).

To find the nose tip of the target face, the spin image of the nose tip of a given model face is used. We arbitrarily picked one and this suffices. The approximated nose tip of a test data is selected by taking a point with the closest spin image to the spin image of the model faces' nose tip. However, taking spin images of all points on a surface is very expensive. To reduce the number of points as candidates for spin images, we only select surface points which qualify as peak points (Table 1) and whose absolute curvatures and volume of neighboring area are over given thresholds. This is known as coarse registration. The next step is fine registration in which we can find orientation of the target face by translating the surface to the model face using the selected point and its normal obtained above. This fine registration is done using ICP (Besl and McKay, 1992). Figure 6 shows the target face before registration and after coarse and fine registration.

Finding the symmetric plane of the face using 'mirror plane' method. We then find a symmetry plane. This is the vertical plane going down the middle of the face such that it divides the face in two symmetric halves. To find a symmetry plane, we use the 'mirror' plane method by Benz and Laboureux (2002). The face mesh is reflected by an arbitrary plane in the space, and then the reflected face is aligned with the original face mesh by translating it. This coarsely aligned face is then finely aligned using the ICP algorithm. After the fine alignment of the original and the reflected meshes, the symmetry plane is determined. Figure 7 explains this method. Finally, we can find a symmetry profile curve which is a point set connected by line segments generated from intersection of the symmetry plane with the face mesh. At

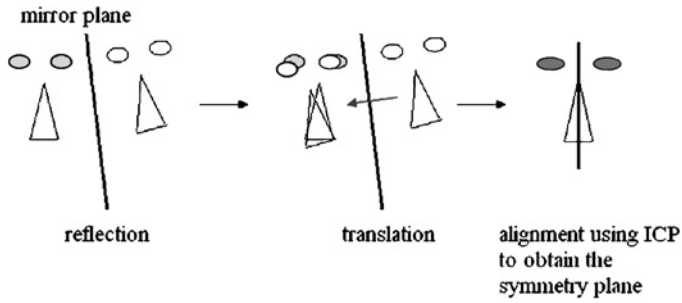


Fig. 7. Finding symmetry plane using 'mirror' plane and application of ICP algorithm and the resulting symmetry plane.

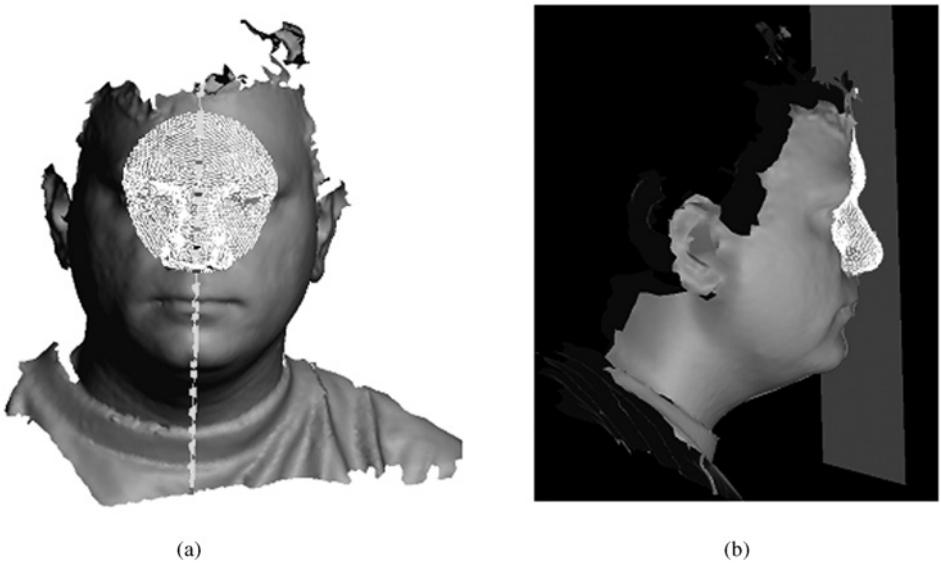


Fig. 8. Selected area for finding symmetry plane (a), facial plane by least square fitting with selected area (b).

this stage, we can find the facial plane from the selected area around the approximated nose bridge using the least square plane as shown in Fig. 8.

Finding facial features. We first approximate the profile line segments by fitting a least squares B-spline curve (Farin, 2002). We then compute curvatures along this curve. Since this is a 2D curve (in the plane of intersection), the curvatures have a sign (+/−) associated with the values where inflection points are the zero crossings of the curvature curve.

We then proceed to finding biometrically important feature points on the symmetry profile curve of a face. These are: the nose tip (NP_T), nose bridge

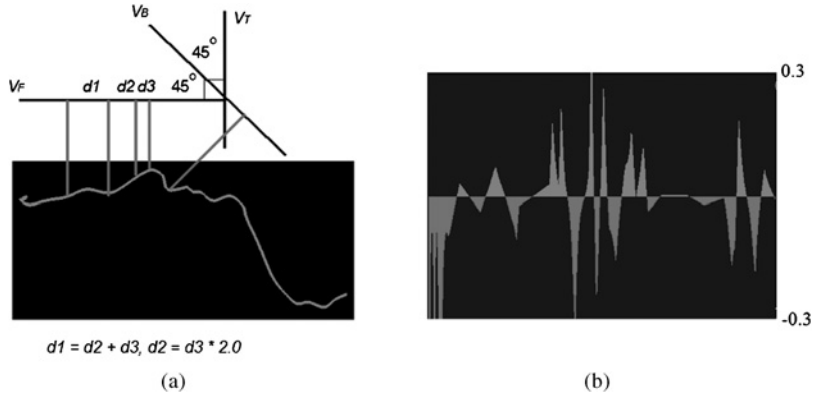


Fig. 9. Symmetry profile and its curvature plot. (a) Feature points on the symmetry profile curve; (b) signed curvature plot of the symmetry profile curve.

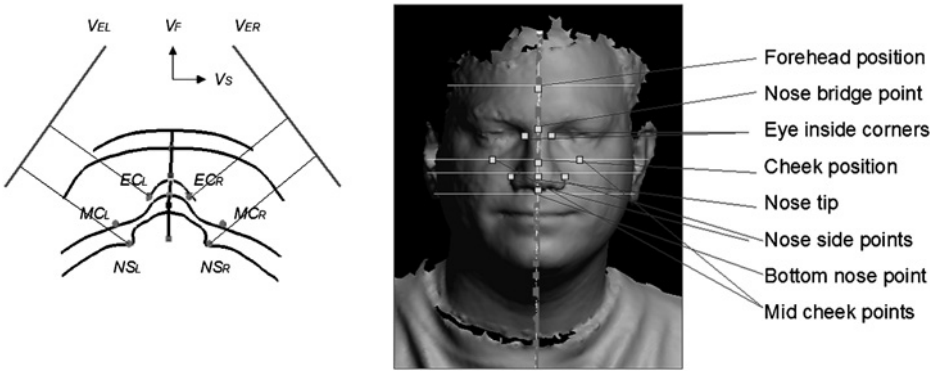


Fig. 10. Facial feature points.

(NP_B), bottom nose (NP_{BT}), forehead position point (FP), and cheek position point (CP). Figure 9 shows the feature points on the symmetry profile curve.

Next, we find eye inside corner points (EC_L, EC_R), mid-cheek points (MC_L, MC_R), and nose side points (NS_L, NS_R) using the feature points on the symmetry profile curve as shown in Fig. 10.

For mid-cheek points, we first find a planar profile curve using the plane associated with the cheek position point, CP, and the plane, V_T.

The shortest distance between the mid-cheek point and the symmetry plane is used to find facial width. We measure the heuristics of 1.5 times shortest distance between each of the mid-cheek points to the profile curve. Now, we can extract the face mask (biometrically relevant part of the face) using six planes as shown in Fig. 11.

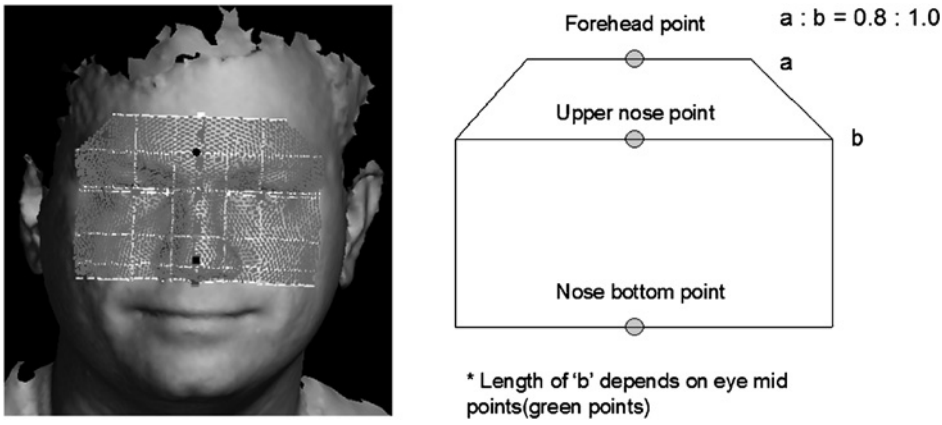


Fig. 11. Face mask extraction.

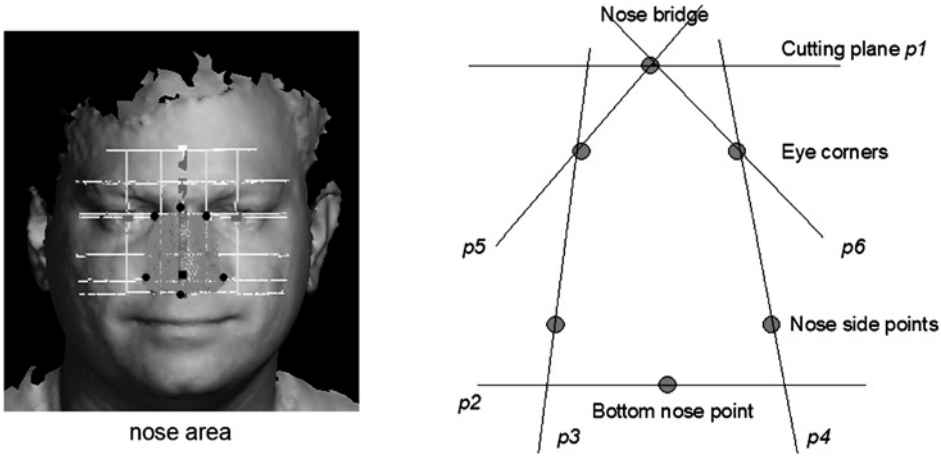


Fig. 12. Nose extraction.

The nose can be extracted by using the nose bridge, bottom point of the nose, eye inner corners, side-points of the nose, surface point type, and the relevant planes which are determined by these points. Figure 12 shows nose extraction and Fig. 13 shows some examples of feature extraction from scans with different orientations.

3.2.2 Comparison

To compare two faces, we use all facial features that have been obtained in previous section including the nose and face mask. Additionally, we use symmetry profile curve, cheek profile curve, the profile curve between eye inner corner points, and forehead profile curve. First, we align two faces using three points which are the nose bridge, nose peak, and nose bottom.

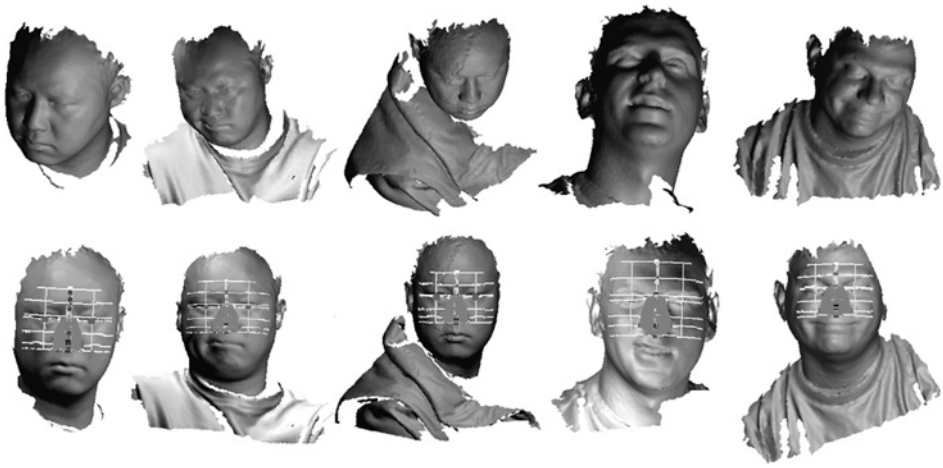


Fig. 13. Feature extraction.

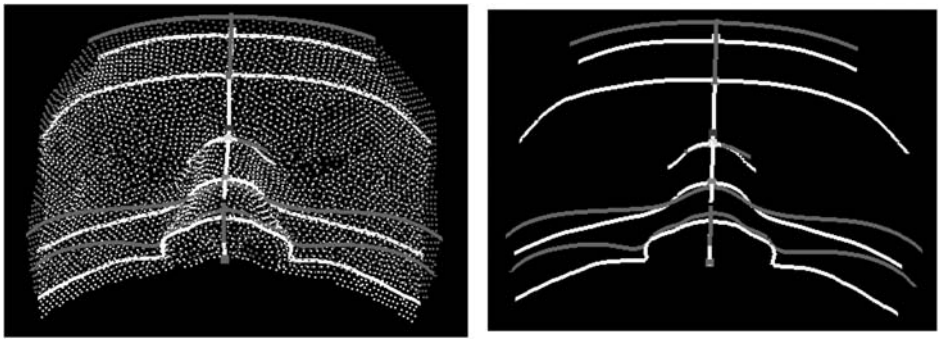


Fig. 14. Face matching examples.

Then, we can compute errors directly for symmetry profiles, noses, and selected faces as shown in Fig. 14. We have selected the following biometric features for similarity comparison:

- Curve length and point distance between nose tip and nose bridge;
- Curve length and point distance between nose tip and eye corner position on a nose profile curve;
- Curve length and point distance between eye corners;
- Length between nose tip and forehead;
- Length between nose bottom and forehead;
- Distance between eye points (inflection points on a cheek curve);
- Nose height;
- Nose area;
- Angle of the eye corner, nose tip, and another eye corner;
- Angle of the eye, nose tip, and another eye corner;

- Angle of the nose tip, bridge point, and upper nose point;
- Angle of the eye point, nose bridge, and another eye point;
- Error of approximated B-spline curves of profiles on critical points;
- Error of nose points and selected face points;
- Spin image of the nose tip from the selected face points.

These biometric features have different discriminating power. To compute the weights of these features, we use Fisher's coefficient linear discriminant criterion (Hallinan et al., 1999)

$$\frac{\sum_{i=1}^{c}(m_i - m)^2}{\sum_{i=1}^{c} 1/n_i \sum_{j=1, x\in\phi_i}^{n_i} (x_j - m_i)^2}$$

where c is the number of people, ϕ_i the set of biometric features for person i, n_i the size of ϕ_i, m_i the mean of ϕ_i, and m the global mean of the feature over all faces of the people. Higher value for a feature means that the feature has more discriminating power. After finding all weights, we exclude biometric features with very low value. We ensure that the sum of all weights is 1.0. We have used 223 individuals, 37 different people with each 2 or 3 normal expression and 3–7 different expression faces at the time of writing this chapter. First, we tested all biometric features, and then their combination with weights. Figure 15a shows ROC, which describes the relationship between the FRR and FAR as a function of the threshold, for

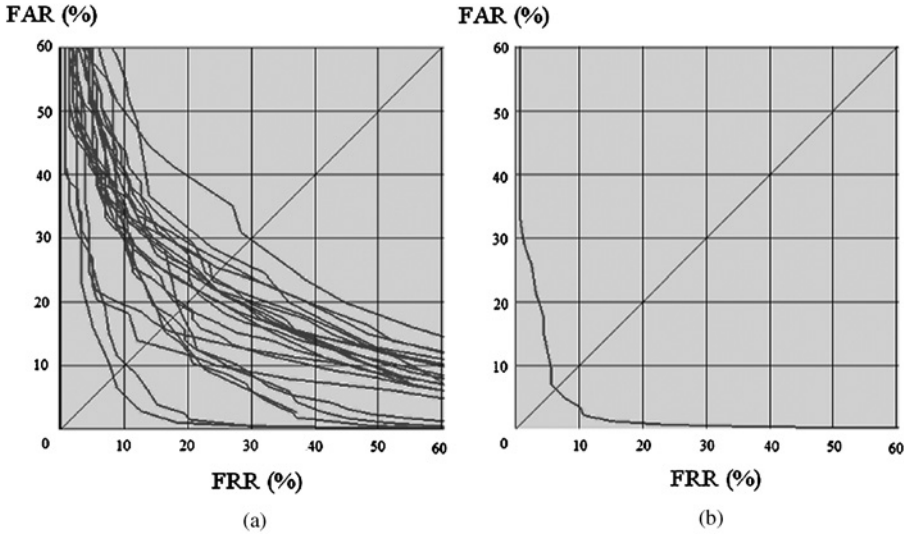


Fig. 15. ROC curves (a) ROC curves for different biometric features; (b) ROC curves for the combination with weights (EER = 6%).

all different biometric features and Fig. 15b shows ROC for the weight combination. EER with weight combination is reported with 6% which includes data with different face expressions in our experiments.

4 Biometric databases

If biometrics-based applications are to be implemented for a large set of people (such as all the visitors coming to the U.S.) then it is imperative that attention be paid to the development of a database architecture for such applications. For 3D faces, the biometric dataset is large and usually extends to gigabytes and terabytes of data for subjects numbering only a few thousand. For centralized databases, retrieving and matching large data can be an inhibiting factor in adopting the technology. Organizing and handling of such huge information has been an open research issue for a long time. The database organization of numbers and/or string information into relational or object-relational databases has already proven to be efficient and standardized via database management systems like Oracle and Microsoft SQL Server (Mhatre et al., 2005). Many researchers have been working on indexing multimedia information itself such as indexing on images or videos. There are three data retrieval mechanisms for multimedia datasets: attribute-based, text-based, and content-based retrieval. Based on each retrieval mechanism, different indexing mechanisms exist. FastMap is one of the content-based algorithms developed for indexing traditional and multimedia datasets (Faloutsos and Lin, 1995). This algorithm maps different objects into points in k-dimensional space, where k is user-defined, preserving the dissimilarities.

4.1 Use of biometric datasets

Biometric datasets can be classified as three distinctive datasets: training, target, and query dataset while working on the recognition and authentication algorithms. The training dataset is used as a basis for developing algorithms and testing the correctness and reliability of the methods implemented. This dataset is generally used to derive rules and/or assign weights to the criteria used for acceptance or rejection. The training dataset can be a subset of the actual target dataset because target dataset continues to increase in size as more and more scans are added to it. The selection of training dataset is purely randomized for a fair deduction of rules. The target dataset is the original dataset gathered over period of time and is continuously updated with more and more scans as they come in. The query dataset is a set of the scans taken at run time and fed to the recognition and/or authentication system to provide results in terms of acceptance or rejection.

There are many publicly available databases for 2D and 3D images for research purpose. A comprehensive report ~27 such databases is available at http://www.ri.cmu.edu/pubs/pub_4932.html (Gross, 2005). Some of the

common and popular databases are the Color (Facial Recognition Technology) FERET database (Phillips et al., 2000b), Face Recognition Grand Challenge (2005) FRGC database, Notre Dame Human ID database (Phillips, 2002), and the AR face database (Martinez and Benavente, 1998). Most of these databases comprise of 2D images of people under different conditions of illumination, different poses and also different angles and backgrounds. FRGC database is a probably one of the few databases that provide 3D face scans along with 2D images.

4.1.1 Architectural design of PRISM 3D face database

Figure 16 shows the architectural view of the 3D face authentication system along with different modules under development. The system has two separate interfaces, application level interface as well as World Wide Web (WWW) interface that allow public access to the system and

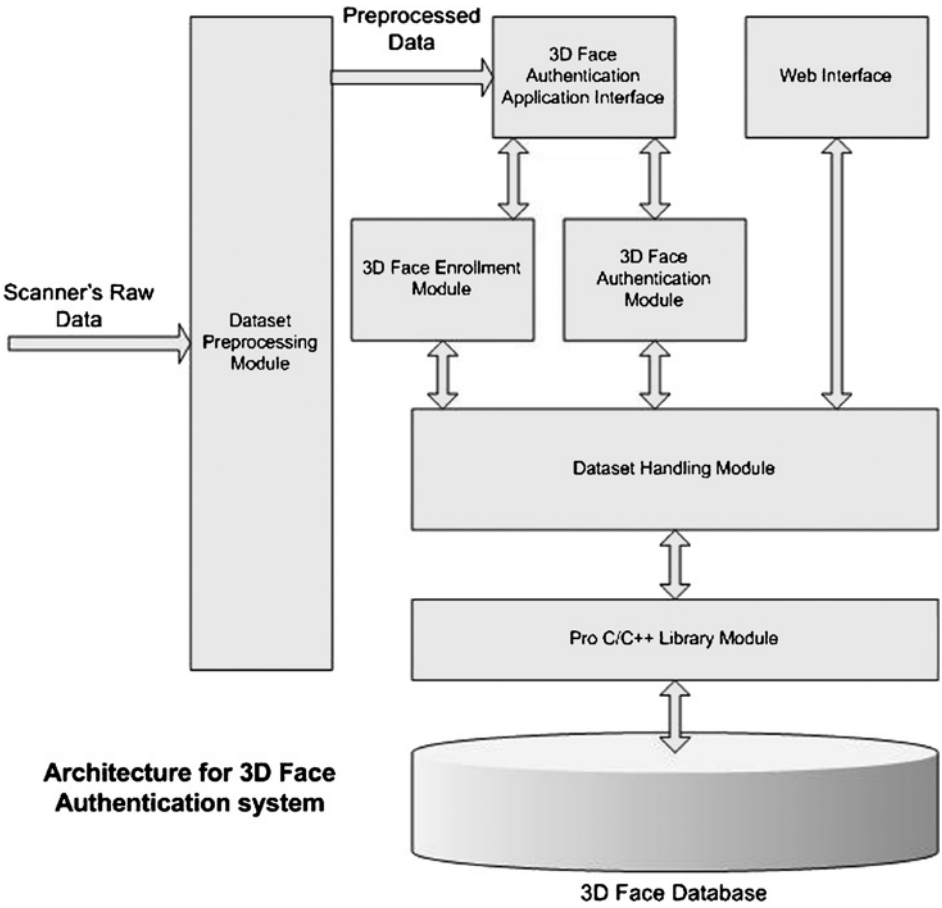


Fig. 16. Architecture for 3D face authentication system.

information. The data pre-processing module converts the raw scanned data files into proper format as need by the 3D Face enrollment and authentication modules. First time registration of a person is done by the enrollment module which interfaces with the database via Dataset handling module and Pro*C/C++ library module. The enrollment module stores the pre-processed face scan along with personal information, and necessary features and derived quantities extracted from the face scan. The dataset handling module and Pro*C/C++ library module provide database handling mechanism through which the enrollment and authentication modules communicate with the database for transporting necessary information back and forth from the database management system.

The 3D face authentication module gets a query face scan to authenticate against an individual, extracts necessary information from the query face, requests similar information about that individual from the database and matches the query information with the target information to produce the result in terms of either accept or reject. The main goal of the project to minimize the FAR and FRR.

Database design has also played a vital role in efficient functioning of the system. Most of the indexing mechanisms available for multimedia databases work on 2D images or images of small size where quick indexing takes place. Creating indexes on 3D face scans of around 20,000 points is very difficult and time consuming. The goal of authentication makes the decision here simple because we can use the personal information as the primary means of indexing the data and the subsequent relationships can also be indexed based on that.

The database design also takes advantage of latest object-relational technology, well-established relational querying and indexing algorithms for fast and efficient retrieval of information. An experimental way of providing the detailed face scan information along with the features and derived quantities in XML format is under progress. The main reason behind this is that XML has been accepted as *de facto* standard for data exchange over the WWW. Even the Biometric Experimentation Environment (BEE) infrastructure discussed in the next section has most of the communication between different applications via XML documents.

The current implementation of the 3D face authentication system is done using Visual C++ 5.0 on windows operating system with GLUT library and the database is implemented on Oracle 9*i* Personal edition. Dataset handling module is developed using Pro*C/C++ library that allows fast connectivity between Oracle and C/C++. The main features of the database design are designing of objects over the relational tables, exploitation of nested tables for modularization of data, XMLType for storing XML documents, and calling external stored procedures in dynamic linked library (DLL).

Current statistics of the database indicates that 3D Face Database contains 2100 face scans for approximately 1500 people. Out of these 1500 people, approximately 150 people are scanned multiple number of times

and still growing. The multiple scans include different facial expressions such as smile, anger, surprise, and other expressions by means of which the facial structure changes from the normal face. Multiple face scans are taken in order to test and verify whether the recognition and authentication algorithms work with different expressions or not. We will continue to report the progress of the project that is still under way.

5 Face recognition grand challenge (FRGC, 2005)

The main goal of FRGC is to promote advance face recognition technology to support existing face recognition efforts by the U.S. Government. This venture allows the growth of low error and better performance promising algorithms as compared to the results from Face Recognition Vendor Test (FRVT) 2002. It is open to all the researchers in companies, academic and research institutions.

FRGC provides two sets of challenging problems, one is the dataset of facial images and other is a defined set of experiments. FRGC provides dataset in two versions. Version 1 (Ver1) is used to introduce participants to the FRGC challenge problem format and the supporting infrastructure. Version 2 (Ver2) is designed to challenge the researchers to achieve the goals put forth by FRGC. The datasets comprises of high-resolution 2D images, 3D face scans as well as multiple images of the same person. Thus the challenge is open for 2D, 3D, and mixture of both 2D and 3D recognition and authentication systems. The defined sets of experiments are based on infrastructure provided by BEE, an XML-based framework for designing and running computational experiments. BEE allows common ground for describing, executing the experiments, recording the results, and presenting the results. FRGC has designed six experiments for both 2D and 3D images with controlled and uncontrolled environments. FRGC is sponsored by many government agencies like FBI, NIST, NIJ, and U.S. Department of Homeland Security, Science and Technology. FRGC has recorded around 100 participants worldwide including 44 universities who have shown interest in biometric research. More information can be found at the FRGC website http://www.frvt.org/FRGC/.

6 Future directions

We are not at a point that 3D facial biometrics can be adopted in the field. Rigorous field testing remains a daunting task for many researchers in this area. To make 3D face scanning unobtrusive, fast, and reliable, we have to solve problems in data processing, deal with facial hair and glasses as well distinguish genuine versus designed to defeat the system with face altering surgery.

In Table 2 there is a list of vendors and academic labs that are selling turnkey systems and conducting research respectively in this area. As is the

Table 2
List of vendors and academic labs that are selling turnkey systems and conducting research

Company	Website URL	Comments
Idteck	http://www.idteck.com/technology/w_face.jsp	FingerPrint Recognition Face Recognition (2D) Iris Recognition
CCE Software Pvt. Ltd.	http://www.ccesoft.com/Vertical_SP_Face_Recognition_System.htm	Face Recognition System (2D)
Cognitec Systems Corporation	http://www.cognitec-systems.de/index.html	FaceVACS® face recognition software (2D)
Fulcrum Strategic Partners, Inc.	http://www.fulcrumspi.com/neurotech.htm	VeriLook Software Development Kit (2D) for faces
	http://www.fulcrumspi.com/verilook.htm	VeriFinger Software Development Kit for fingerprints
Neven Vision	http://www.nevenvision.com/devtools.html	Facial Recognition SDK (2D)
Identix Incorporated	http://www.identix.com/products/pro_security_bnp_argus.html	FaceIt® ARGUS (2D)
Visiphor	http://www.imagistechnologies.com/solutions_biometrics.aspx	Face Recognition System (2D)
Viisage	http://www.viisage.com/ww/en/pub/viisage___products_new/viisage___biometrics/facefinder.htm	Viisage FaceFINDER
A4Vision, Inc.	http://www.a4vision.com/	Face Recognition System (3D)
Genex Technologies	http://www.genextech.com/pages/608/3D_Facial_Recognition.html	SureMatch 3D™ Suite (3D recognition)
BioVisec, Inc.	http://www.biovisec. com/3d-face-recognition.html	Face Recognition System (3D)
Passfaces (Real User Corporation)	http://www.realuser.com/	Face Recognition System (2D)
ImageWare Systems, Inc.	http://www.iwsinc.com/Biometrics.cfm	ImageWare Biometric Products (2D Face and Fingerprint)

Table 2. (*Continued*)

Company	Website URL	Comments
Acsys Biometrics Corp.	http://www.acsysbiometricscorp.com/product.html	Acsys Face Recognition System (2D)
Iconquest	http://www.iconquesttech.com/iconquesttech_003.htm	Face Recognition System (2D)
Research Group and University		
MIT Media Laboratory Vision and Modeling Group	http://vismod.media.mit.edu/vismod/demos/facerec/	Face Recognition System (2D)
The CSU Face Identification Evaluation System	http://www.cs.colostate.edu/evalfacerec/	Face Recognition System (2D)
IBM Exploratory Computer Vision Group	http://www.research.ibm.com/ecvg/biom/facereco.html	Face Recognition (2D)
CUBIC, Arizona State University	http://cubic.asu.edu/research/face_recognition.html	Face Recognition (2D)
Michigan State University	http://www.cse.msu.edu/~lvxiaogu/research/abstract.htm	3D Face Matching
Computer Vision Research Laboratory, University of Notre Dame	http://www.cse.nd.edu/~cvrl/	2D and 3D Face Recognition System
Drexel University	http://www.ece.drexel.edu/faculty/cohen.html	2D and 3D Face Recognition System
Computer Science and Engineering Department, University of California at San Diego	http://vision.ucsd.edu/kriegman-grp/research/9pt/index.html	Face Recognition (2D)

nature of technology the list is dynamic and therefore the readers are suggested to not limit their search to the only ones listed here.

7 Questions for classroom discussion

Here are some questions that can be used as discussion starters.

1. Under what circumstances would you need zero false positive acceptance and zero false rejection rates? Explain your answers with comments on feasibility.
2. Conduct the following experiment. First take a photograph of your face and label various features that you think are invariant to change in expression, pose, etc. Attach relative weights to these features. Then take several pictures of yourself while you try to prove the invariance of these features wrong. Tally at the end which features survive and why.
3. A practical system in operation in the field will be a hybrid system. Construct a conceptual system and its various components and the list of biometrics each module is responsible for. Explain your choices.
4. One of the biggest problems in application of any system in the field is its ease of operation and social acceptance. Conduct a survey and summarize the pros and cons of implementing a 3D face recognition system in the field (say for immigration).

Acknowledgments

We would like to thank Prof. Charles Lockwood of University College, London, for his collaboration on the project. The 3D face project was funded by NSF grant #0312849. The authors would like to thank PRISM research center at ASU for providing a stimulating environment for conducting the research. The authors would also like to thank many graduate students, especially John Femiani, for ideas and discussions on related topics.

References

Belhumeur, P., J. Hespanha, D. Kriegman (1997). Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 711–720.

Benz, M., X. Laboureux (2002). The symmetry of faces, in: *Proceedings of Vision, Modeling, and Visualization (VMV 2002)*, Nov. 20–22, Erlangen, Germany.

Besl, P., R. Jain (1988). Segmentation through variable order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10(2), 167–192.

Besl, P., H. McKay (1992). A method for registration of 3-D shapes, in: *Proceedings of the IEEE Transactions on Pattern Analysis Machine Intelligence*, Vol. 14, pp. 239–256.

Beumier, C., M. Acheroy (1998). Automatic face authentication from 3D surface, in: *British Machine Vision Conference BMVC 98*, Sept. 14–17, University of Southampton, UK, pp. 449–458.

Beumier, C., M. Acheroy (2000). Automatic 3D face authentication. *Image and Vision Computing* 18(4), 315–321.

Beumier, C., M. Acheroy (2001). Face verification from 3D and grey level clues. *Pattern Recognition Letters* 22(12), 1321–1329.

Blanz, V., T. Vetter (1999). A morphable model for the synthesis of 3D faces, in: *SIGGRAPH '99 Computer Graphics Proceedings*, Los Angeles, CA, pp. 187–194.

Blanz, V., T. Vetter (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(9), 1063–1074.

Bronstein, A., M. Bronstein, R. Kimmel (2003). Expression invariant 3D face recognition, in: *Proceedings of the Audio and Video Based Biometric Person Authentication*, Guildford, United Kingdom, pp. 62–69.

Bronstein, A., M. Bronstein, R. Kimmel (2005). Three-dimensional face recognition. *International Journal of Computer Vision* 64(1), 5–30.

Chang, K., K. Bowyer, P. Flynn (2003). Face recognition using 2D and 3D facial data, in: *Multimodal User Authentication Workshop, December*, Santa Barbara, CA, pp. 25–32.

Chang, K., K. Bowyer, P. Flynn, X. Chen (2004). Multi-biometrics using facial appearance, shape and temperature, in: *The Sixth IEEE International Conference on Automatic Face and Gesture Recognition, May*, Seoul, Korea, pp. 43–48.

Chua, C., F. Han, Y. Ho (2000, March 26–30). 3D human face recognition using point signature, in: *Fourth ICAFG*, Grenoble, France.

Chua, C., R. Jarvis (1997). Point signatures: a new representation for 3-D object recognition. *International Journal of Computer Vision* 25(1), 63–65.

Duc, B., S. Fischer, J. Bigun (1999). Face authentication with Gabor information on deformable graphs. *IEEE Transactions on Image Processing* 8(4), 504–516.

3D Face Authentication for Biometric Access Control (2002). Retrieved July 7, 2006, http://prism.asu.edu/3DFaceAuthentication.pdf

Face Recognition Grand Challenge. (2005). Retrieved July 7, 2006, http://www.frvt.org/FRGC/Default.aspx

Faloutsos, C., K.I. Lin (1995). FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets, in: *Proceedings of ACM SIGMOD* San Jose, CA, ACM Press, New York, pp. 163–174.

Farin, G.E. (2002). *Curve and Surface for CAGD: A Practical Guide* 5th ed. Morgan-Kaufmann, California.

Farin, G.E., D. Hansford (2000). *The Essentials of CAGD*. A. K. Peters, Wellesley, MA.

Fisher, R.A. (1936). The use of multiple measures in taxonomic problems. *Annals of Eugenics* 7, 179–188.

Gelfand, N., L. Ikemoto, S. Rusinkiewicz, M. Levoy (2003). Geometrically stable sampling for the ICP algorithm, in: *Proceedings of 3-D Digital Imaging and Modeling (3DIM),* Banff, Canada, October 6–10, pp. 260–267.

Gordon, G. (1992). Face recognition based on depth and curvature feature, in: Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, Illinois, pp. 808–810.

Gross, R. (2005, February). Face databases, in: S. Li, A. Jain (eds.), *Handbook of Face Recognition*, Springer, New York.

Hallinan, P., G. Gordon, A.L. Yuille, P. Giblin, D. Mumford (1999). *Two and Three-Dimensional Pattern of the Face*. A. K. Peters, Natick, MA.

Johnson, A.E., M. Hebert (1999). Using spin-images for efficient multiple model recognition in cluttered 3-D scenes. *IEEE PAMI* 21(5), 433–449.

Kazhdan, M., T. Funkhouser, S. Rusinkiewicz (2003). Rotation invariant spherical harmonic representation shape descriptors. *Eurographics Symposium on Geometry Processing*, Aachen, Germany, pp. 156–164.

Martinez, A.R., R. Benavente (1998). The AR face database. Technical Report 24, Computer Vision Center (CVC) Technical Report.

Masuda, T., N. Yokoya (1995). A robust method for registration and segmentation of multiple range images. *Computer Vision and Image Understanding* 61(3), 295–307.

Medioni, G., R. Waupotitsch (2003). Face recognition and modeling in 3D, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), Nice, France, October, pp. 232–233.

Mhatre, A., S. Palla, S. Chikkerur, V. Govindaraju (2005). Efficient search and retrieval in biometric databases. *SPIE Defense and Security Symposium*, Orlando, Florida, *March*, Vol. 5779. pp. 265–273.

Moreno, A.B., A. Sanchez, J. Fco, J. Diaz (2003). Face recognition using 3D surface-extracted descriptors, in: *Irish Machine Vision and Image Processing Conference (IMVIP 2003),* Coleraine, Ireland, *September*.

Pang, Y., L. Zhang, M. Li, Z. Liu, W. Ma (2004). A novel gabor-LDA based face recognition method, in: *The Fifth IEEE Pacific-Rim Conference on Multimedia (PCM)*, Tokyo, Japan, pp. 352–358.

Phillips, P.J. (2002). Human identification technical challenges, in: *IEEE International Conference on Image Processing*, Rochester, New York, Vol. 1, pp. 22–25.

Phillips, P.J., A. Martin, C.L. Wilson, M. Przybocki (2000a, February). An introduction to evaluating biometric systems. *Computer Magazine*, Vol. 33, No. 2, 56–63.

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J. (2000b, October). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22, 1090–1104.

Phillips, P.J., H. Wechsler, J. Huang, P.J. Rauss (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing* 16, 295–306.

Razdan, A., M. Bae (2005). Curvature estimation scheme for triangle meshes using biquadratic bezier patches. *Computer Aided Design* 37(14), 1481–1491.

Rusinkiewicz, S., M. Levoy (2001). Efficient variants of the ICP algorithm, in: *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, Quebec City, Canada, pp. 145–152.

Srinark, T., Kambhamettu, C. (2003). A novel method for 3D surface mesh segmentation, in: *Proceedings of 6th IASTED International Conference on Computers, Graphics and Imaging*, Honolulu, Hawaii, pp. 212–217.

Tanaka, H., M. Ikeda, H. Chiaki (1998). Curvature-based surface recognition using spherical correlation principal directions for curved object recognition, in: *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 372–377.

Turk, M., A. Pentland (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86.

Yamany, S.M., A.A. Fraag, A. El-Bialy (1999). Free-form surface registration and object recognition using surface signatures, in: *IEEE International Conference on Computer Vision*, Kerkyra, Greece.

Zhang, Z.Y. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision* 13(2), 119–152.

Zhang, D., M. Hebert (1999). Harmonic maps and their applications in surface matching, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '99)*, Fort, Collins, Co., Vol. 2, p. 2524.

Zhang, L., A. Razdan, G. Farin, J. Femiani, M. Bae, C. Lockwood (2006). 3D face authentication and recognition based on bilateral symmetry analysis. *The Visual Computer* 22(1), 43–55.

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 5

The Necessity of Fuzzy Logic for Identity Matching

Peter C. Went
WCC Services US, Inc., 228 Hamilton Avenue, Suite 300, Palo Alto, CA 94301, USA

Abstract

This chapter is about identity matching, which is the 'art' of finding an individual in a database using one or more descriptive criteria of the person. Such descriptive criteria can be simple biographic criteria, like name and gender, but also more complex criteria, like fingerprint and DNA. There are a number of special considerations when using biographic and/or biometric criteria to find a person in a database, because such criteria are neither 100% accurate nor 100% trustworthy nor 100% unique. We define accuracy as how accurate the infrastructure can capture a criterion as it is offered to that infrastructure. The infrastructure includes the capturing device, e.g. a fingerprint scanner, the scanning software that can read the image from the capturing device and translate it into a set of features and lastly the software algorithm that compares two sets of features. We define trustworthiness as how easy or likely a criterion can be forged by its owner. For example, the color of one's hair can very easily be changed, whereas the vein pattern in a person's hand cannot be forged at all. We define uniqueness as the percentage of individuals in a database with the same value for the given criterion, assuming it was captured 100% accurate and 100% trustworthy. Given that no criteria will ever be 100% accurate, 100% trustworthy and 100% unique, the best way to deal with that is fuzzy logic and weighted criteria. The weight given to a (search) criterion reflects its accuracy, trustworthiness and level of uniqueness.

1 Introduction

The subject of 'identity matching' can be interpreted widely. One such interpretation is obtaining access to a computer system or application, which is commonly known as authentication. A user claims an identity and supplies several credentials. Another interpretation is person lookup, commonly

used in many software applications. For example, a person who needs medical care first needs to be looked up in the system (checking insurance and medical background). We will not focus on these areas, although there is great commonality. We are discussing the subject of identity matching in the context of security, like police and border security. This space has certain specific demands that need to be dealt with, that make it complex and demanding, hence interesting from both an academic and an IT perspective.

Humans have always identified each other mainly by recognizing faces and voices. This, of course, is the reason why many passports, driving licenses and corporate ID cards now bear photographs to help officials verify the holder's identity. But when it comes to identifying people through technology—often to ensure security—still photos are just the beginning. Retinal scanning, 3D thermal imaging and spectroscopic skin analysis are just three of the exotic technologies that are already here, but that have yet to break into the mainstream. Identity matching by means of biometrics is an automated method of searching a person based on physiological and/or behavioral characteristics including fingerprints, retinal and iris scans, hand and finger geometry, signature recognition, voice patterns, facial recognition and other techniques.

1.1 How biometrics work

Different biometric systems may analyze different traits, but behind the various technologies is a common procedure. Typically the whole process starts with enrollment. The subject submits an identifiable, unprocessed image or recording—called a sample or template—of his or her biometric via an acquisition device, such as a scanner or camera. This sample is then processed to extract information about distinctive features and so enables the creation of a reference profile (essentially a large sequence of numbers). Typically the original sample image or recording cannot be reconstructed from the reference profile, which is beneficial as an extra security layer.

In case of a claimed identity, the profile of the claimed identity can be retrieved from the database and this profile is compared with a reference profile from the database. This process is also referred as verification or authentication. If no identity is claimed, all reference profiles from the database are traversed and matched against this profile being searched and if the match score is above a certain threshold it is considered a match. Multiple matches can be found in this way as multiple reference profiles may match sufficiently well.

Verification is a 1:1 comparison of profiles, where identification is a 1:N comparison of profiles. Effectively, an identification process performs N verifications. In general, due to the nature of biometrics, no two biometric profiles, even taken from the same person with the same scanner, are ever identical, and the biometric system must judge whether there is a

close-enough match. The matching score must exceed a definable threshold in order to be considered a match.

1.2 Issues with identity matching

There are several issues for identity matching which deserve special consideration. The following paragraphs describe each of these and discuss how it contributes to the challenge of developing an effective identity matching system.

Accuracy measures whether the information presented is recorded correctly. If we use automated mechanisms, like a fingerprint scanner, we have to deal with issues like scanning the finger and recognizing the patterns correctly. When data are entered manually, like an operator typing a name, no data entry mistakes should be made. Incorrectly entered data make identity matching even more challenging.

Reliability is the trustworthiness of the information recorded. In a security context, the people being identified often have an incentive to disguise their identity. So we should distrust any information presented and try to determine whether or not it is genuine. Besides intentional disguise, biometrics also tends to change over time. Changes may occur naturally, for example, a voice becomes deeper with age, but also fingerprints may wear out when people use their hands often for manual labor. Interesting tests were done by the German magazine *C't* to prove how easy it is to spoof any kind of biometrics (see references).

Uniqueness is important, because no single criterion is 100% unique. Certain biometrics, like DNA and retinal images, are close to 100% unique, but if the database with identities to search against is very large and/or frequently searched, a very high uniqueness may not be enough. A short example will clarify this further. The US government intends to build a visitor database that will grow to 500 million people. In addition, over 2 million people enter or leave the US every day. That makes $(5 \times 10^8) \times (2 \times 10^6) = 1 \times 10^{15}$ identity matches per day. A typical fingerprint has a uniqueness of 10^{-6}, and two fingerprints 10^{-10}, so that clearly is not unique enough.

Confidence is a combined concept that encompasses accuracy, reliability and uniqueness. It expresses the degree to which we trust a match. No technology currently exists that allows these factors to be expressed and taken into account when determining a match score.

Scalability is an issue, because governments are building up centralized databases. These will grow to phenomenal proportions as human populations are already very large and they are still growing.

Performance, the speed of search, is related to scalability. Extremely large scale databases that are accessed numerous times per second are a challenge to much of today's technology, especially because biometrics cannot be indexed to improve search performance. When searching on biometric criteria, a linear exhaustive search must be done (called a 'tablespace scan'

in relational databases). Going through that much data linearly results in large computations and is I/O intensive. As a reference, the FBI's IAFIS system has a typical response time of 10 min (at best)[1] and is substantially smaller than the US Visit system is planned to become. Imagine the queues at immigration. The identification process, of which the database search is part, should take no longer than 30 sec altogether (being capture, search and response).

High availability is a principal requirement for any centralized system, but if national security is relying on this system, then high availability should be genuinely 7×24 without downtime, either planned or unplanned. As a (poor) reference, the FBI's IAFIS system is 'down' 2 days per month (both planned and unplanned), which is clearly unacceptable for a border security system.

Security is fundamental to any large centralized system, but a system that holds all identifying criteria of hundreds of millions of people requires very special attention to security. Relatively many Jews were deported from The Netherlands at the beginning of WWII, because the Dutch town halls had a very good registration system of their inhabitants, including their religion, etc. Governments have a serious obligation to the people registered in the database that their data are ultimately safe! Consider that a stolen credit card can be replaced in 24 h, biometrics cannot.

1.2.1 Evolution of technology and obsolescence

Finally, biometric technology is rapidly evolving. This can quickly result in today's cutting edge biometric identity solution becoming obsolete. For example, many of today's biometric identity solutions are based on fingerprints and facial scans, with iris scans and retina scans being increasingly deployed. However, even newer systems, such as vascular pattern scans, are being developed. Such rapid evolution can present a serious obstacle for today's hardware and software that is hardwired for a particular technology.

1.2.2 Combining disparate data sources

Many government agencies have developed their own systems for storing and retrieving information on people in their databases. These may be convicted criminals, suspected terrorists, visitors entering a country, social security files, etc. None of these agencies will want to relinquish control of their data or their systems, since they are typically vital to the agency's day-to-day operations. However, identity searching is more effective when the data from multiple agencies are searched simultaneously. One likely solution is to create a database holding replicated data from disparate data

[1]This information was taken from a letter from the Under Secretary of Border and Transportation Security to Inspector General of U.S. Department of Justice, dated December 4, 2004.

sources. The issue is primarily political, and can probably be resolved by a centralized read-only database holding replicated data from each source.

1.2.3 Missing data

When data from disparate data sources are combined, there will be some form of overlap in data that are typically kept in each database, but each database will hold its unique data as well. When combining such data from different sources, not all fields will be populated in each and every record. Take, for example, the FBI capturing 10 fingerprints and US Visit (currently) capturing only 2. In a combined database, some records will then have 2 and some records will have 10 fingerprints. Searching on more than the two fingerprints captured by US Visit would prevent records from US Visit from being found.

1.2.4 Theft

Almost all traits that can (partially) identify a person can be stolen and subsequently used for spoofing an identity system. In Appendix, we have included information on how easily a trait can be copied. Most traits currently used for identification can be copied easily, without the owner's consent. A reliable identity matching solution needs to deal with that fact.

Of course, all these issues are not new to the (IT) industry at all, but the need to address each and every one of them in a single system makes it a very demanding challenge.

2 Characteristics of biometrics

In this section we will explore the characteristics of biometrics in more detail, to understand and appreciate the intricacies and issues better. We will explore uniqueness, theft and deception in detail.

2.1 Biometrics are not infallible

Biometric systems are not infallible. A search profile can be matched incorrectly against another person's reference profile, or might fail to register a match even if the user is properly enrolled. The accuracy of these systems can be measured in two ways:

- False Accept Rate (FAR);
- False Reject Rate (FRR).

But the True Accept Rate and the True Reject Rate are also relevant (Figs. 1–3).

Let us assume that we have a database with two reference profiles, of person A and person B. If person A presents him- or herself and we do a

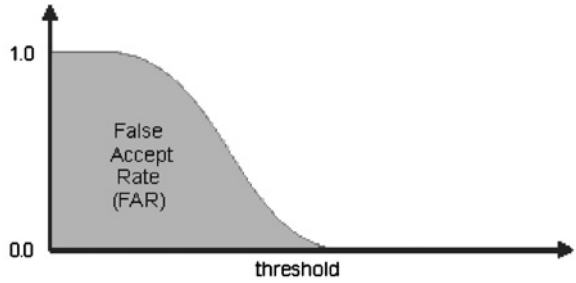


Fig. 1. False Accept Rate. The False Accept Rate (FAR) is the probability that we find people who are not the persons we are looking for. With most biometric algorithms, if we set the threshold high, the FAR will be low and with a threshold of 0 the FAR will be 100% (of course). The figure shows how the FAR is impacted by a varying threshold. In the context of checking a person against a visitor database, you do not want to find too many similar profiles, because too many matches would confuse the immigration officer (i.e. a very low FAR is required).

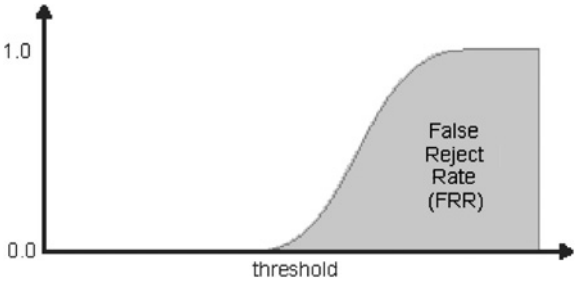


Fig. 2. False Reject Rate. The False Reject Rate (FRR) is the probability that we do not find a person in a database, while that person is actually present in that database. With most biometric algorithms, if we set the threshold high, the FRR will also be high (see figure). If the threshold is 100% (Boolean True) the probability of not finding a person as a perfect match is nearly 100% as well, which is how typical databases would work. In the context of checking a person against a watch list, you want a high probability of finding a person who is actually on the watch list, because not spotting a person on a watch list compromises security (i.e. a very low FRR is required).

search against the database, the following situations may occur and they are not necessarily mutual exclusive:

1) we find A, which is a *True Accept*;
2) we find B, which is a *False Accept*, since we find a person we should not have found;
3) we do not find A, which is a *False Reject*, since we do not find a person that we were supposed to find;

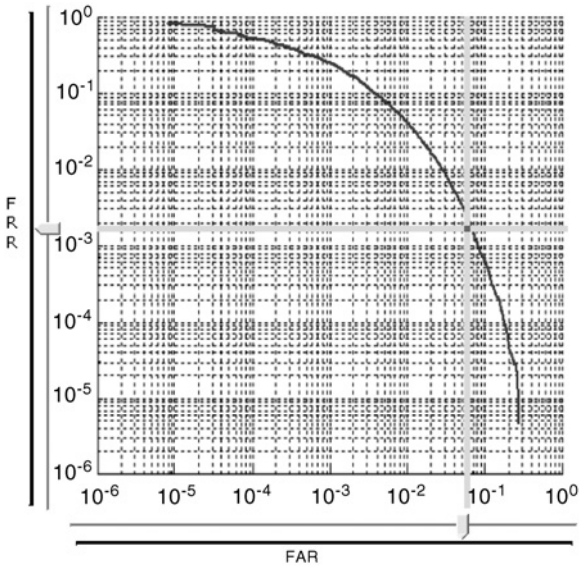


Fig. 3. FAR versus FRR. In general, the lower the FAR, the greater the FRR and vice versa. There is, therefore, a trade-off between security and ease of use. Both the FAR and the FRR reflect the system's ability to allow entry to authorized users only. For an identification system, the lower a system's FAR, the easier it is to use. The lower a system's FRR, the better its security. When building a verification system, the opposite is true: the lower a system's FRR, the easier it is to use, and the lower the FAR, the better its security. The figure shows that when the FAR improves (i.e. lower value), the FRR will worsen (i.e. get higher) and vice versa. The numbers themselves relate to the accuracy of a single finger with a common algorithm, and hence are not relevant here. Clear is the trade-off between FAR versus FRR. Exactly how the FAR and FRR relate is dependent on the biometric algorithm used and the quality of the data.

4) we do not find B, which is a *True Reject*, since we do not find the person that we were not supposed to find.

2.2 Identity theft and deception

What happens if a theft of a biometric occurs, or if a biometric is compromised? If a password is compromised it is changed to a new one, but what happens when, for instance, someone copies your fingerprint or makes a contact lens with a copy of your iris? A user's biometric cannot be changed like a password, so what happens in this case? Or, is the way in which biometrics are 'scanned' or stored able to prevent this type of compromise? A behavioral biometric such as signature or handwriting cannot be 'stolen' but someone can learn to sign or write like you to a certain extent. A physiological biometric such as fingerprint or face or iris image

can be 'stolen'—a copy of raw biometric data (or a feature template) obtained by illegal means. Ideally a biometric is what an individual possesses and another individual should not be able to possess the same.

For other systems, merely obtaining the data is not enough. The impostor will have to present the biometric to the system as well and fool the system in some way. Some systems have "liveness" tests which can reject presentations such as fingerprints copied on plastic material or faces shown as photographs. Clever ways of circumventing such checks have also been devised and there is no completely secure method. One way to prevent theft of biometrics such as iris or retinal scans (which cannot be as easily obtained as fingerprints or faces) would be not to supply them in raw form to anyone, but only in an encrypted form—what is being referred as 'cancellable' biometrics. Keys used for encryption and decryption can be changed. While not foolproof, it does make it harder to get a useful form of the data. Multimodal systems can have an advantage in this regard. It is more difficult to present a falsified face as well as a falsified fingerprint and sign like another person. Falsifying one biometric is doable; falsifying all of them is near to impossible.

It is likely that every biometric is breakable with the appropriate amount of time and money. The only thing that the biometric manufacturer can do is to increase the costs involved. The advantage for the manufacturer is that he can invest time and money to secure the system to any imaginable attack. His disadvantage is that he has to remain one step ahead of the impostor. The advantage for the impostor is that he only has to find that one feasible attack that no one anticipated. Another concept often emerging in security-related environments is security through obscurity. This means, the manufacturer of the hypothetic security system (be it cryptography or biometrics) tries to conceal the algorithmic innards of his system to some extent in order to hamper possible attacks. The reasons for this are the following:

1) The fact that the mode of operation of a system is unknown to the public probably only raises the uncertainty about its safeness and not the principal safeness itself.
2) It is hardly likely that any algorithmic part of the system will be kept concealed *in the long run*.

Security can be significantly improved through the use of a multimodal biometric system. Multimodality in this context means combining several biometric traits from possibly more than one sensor in an optimal way. Examples are combinations of fingerprint and face. This concept increases the accuracy of the system in terms of equal error rate (EER), which is the point where the plots of FAR and FRR cross, as well as the resistance to counterfeiting attempts, simply because all traits have to be counterfeited simultaneously. A multimodal solution does not depend on each trait to match, in order to still identify a person.

3 Fusion of results, the answer

As argued in previous sections, fusion is perceived as probably the best approach to tackling lack of uniqueness, possibility of theft and the chance for deception. This section explores fusion in detail from a theoretical perspective. Combining multiple biometric traits of a person increases accuracy and reliability. But how do we go about combining these biometrics into a single search? This is not as easy as it may seem.

Historically each vendor in the biometric matching space has specialized in a single biometric (see Appendix). This implies that the software of such a vendor holds all occurrences of that biometric (e.g. fingerprint) and that software can be used to search against those occurrences (called a 1:*N* match). Searching on multiple biometrics has thus been translated into 1:*N* matches for each biometric separately. The result of each separate search, a list of potential matches on a single biometric, was then combined into a single list. *Imagine a number of 'fruit baskets', each holding an 'apple' and a 'pear'. Each apple and pear is labeled with the ID of the basket it is in. Now we take all fruit from the baskets and order them sort by sort. The first task is to find the nicest apples and the second task is to find the nicest pears. The issue is that lists of nicest 'apples' and nicest 'pears' are to be combined into a list of nicest fruit baskets … not even complaining about having to extract all apples and pears from their baskets first, in order to perform this kind of scoring. Clearly this task is difficult and unnatural.*

In the literature the best approach to fusion is called score level fusion. With score level fusion, all profiles are traversed, each holding the different biometrics of a single person. Each biometric is then compared with the biometrics being searched upon (called a 1:1 match). After each profile is scored in this way, the best matching profiles are returned. *In fruit terms, we are scoring each fruit basket. One observes the apple and pear in a basket and uses those observations to score that basket. After having scored each basket, the nicest baskets can be returned. This is a more meaningful approach.*

3.1 A more formal description of fusion

Fusion, or multimodality, combines multiple biographic and/or biometric traits to improve the FRR and minimize the sensitivity for spoofing and fraud. Fig. 4 illustrates how and where fusion can take place.

In Fig. 4, fusion of the face and fingerprint asserts the veracity of the claimed identity. Techniques such as logistic regression may be used to combine both scores. These techniques attempt to lower the FRR and FAR. The Feature Extraction Module takes an unprocessed image or recording—called a sample—of a biometric. This sample is then processed to extract information about distinctive features and so enables the creation of a search profile.

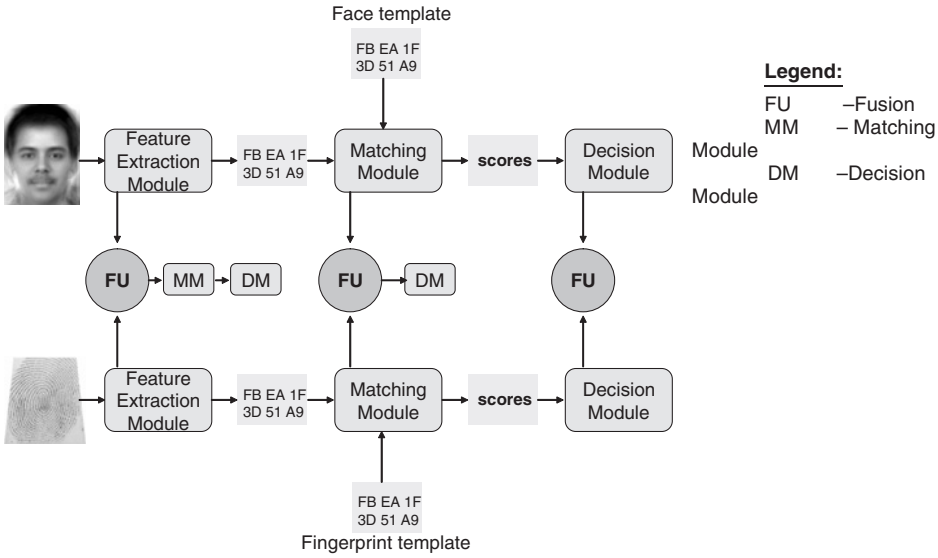


Fig. 4. Face Template Fusion. *Legend*: (FU) Fusion; (MM) Matching Module; (DM) Decision Module.

The Matching Module takes this profile, which is then compared with a reference profile from the database that was created from multiple samples when the person enrolled in the biometric system. The matching score must exceed a definable threshold. The result is a list of sufficiently matching reference profiles. In an identification situation, when a 1:*N* match is performed, the process repeats for each profile in the database. The Decision Module takes the list of all matching profiles to decide which profiles are the most likely matches. Although fusion may take place at any of the three stages, fusion at the score level, i.e. at the Matching Module level, is the generally preferred place to fuse for two main reasons:

- It is quite easy to combine different match scores using a choice of techniques. Combining at the feature level is more difficult, as that requires in-depth knowledge of the actual features used, which may not always be available when using commercial algorithms. Typically, it is possible to retrieve the match score from the biometric algorithm.
- Fusion at the score level allows more control than fusion at the decision level. A score can be said to reflect a level of certainty, i.e. confidence, of a match; this level of information is not available at the decision level, where there is only a binary yes/no verdict available.

There are two approaches for combining scores obtained from different Matching Modules. One approach is to formulate the problem as a classification problem. The different scores from the Matching Modules are combined to a single feature vector, which is then classified into one of the

two classes: "Genuine user" or "Imposter". A second approach is to directly combine the individual matching scores to generate a single scalar score, which is then used to make the final decision. Ross and Jain (2003) have shown that the combination approach performs better than some classification methods like decision tree and linear discriminant analysis.

3.2 Fusion methods

There are five recognized fusion methods, Simple-Sum (SS), Min-Score (MIS), Max-Score (MAS), Matcher-Weighting (MW) and User-Weighting (UW).

SS sums all individual scores into a summed score, which is treated as composite score. MIS takes the lowest score from all individual scores, which is then treated as the composite score. MAS takes the highest score from all individual scores, which is then treated as the composite score. MW assigns weights to the individual matchers based on their EER; thus, the weights of more accurate matchers are higher than those of less accurate matchers. UW assigns weights to individual matchers that may be different for different users, a method that is considered prohibitively expensive. The MW method can compensate for the variation in confidence of the different biometric algorithms. Therefore, for most applications this is the preferred method, although the ideal solution should be able to implement each of them if that is required.

4 Why traditional database queries do not work?

In the previous sections we have argued that biometrics are not completely unique, not always accurate and not always reliable. As a result, they pose certain demands on the database that holds the data in order for them to be searched effectively. These demands are not supported well by conventional databases. This section explains those critical deficiencies in more detail.

SQL queries, also known as SQL SELECT, are the *de facto* query language for relational databases, which is the most common type of database.

4.1 Boolean nature of SQL

SQL queries enable selection of data from one or more tables from a relational database using a SELECT statement. The WHERE condition in this statement joins rows together from different tables and limits the rows selected to those that meet certain selection criteria.

For example, if we want to select people that are ~30 years old, because the person to be identified looks ~30, we can specify:

```
WHERE age BETWEEN 29 AND 31
```

Not only will this give already a very large selection from the rows in the database, but also will not find people that are about to turn 29 shortly. Of course, we can than widen the selection, to make sure that we do not miss out the person that we are looking for, by specifying:

WHERE age BETWEEN 20 AND 40

Of course, it is quite unlikely now that we have excluded a person whom we believe is ~30, although that could still be possible, but we now have selected probably half the rows in the database. And more awkwardly, all people selected are considered equally likely to be a match. In real life, if you were given the task of identifying a person by age, would you consider a 20-year-old person to be an as likely match as a genuinely 30 years old? Probably not!

Next consider that we want to select people that have blue eyes, because the person to be identified seems to have blue eyes, we can specify:

WHERE eyecolor = 'BLUE'

How sure are we that at enrollment time and/or at identification time no observational errors were made, or that no errors in interpretation occurred? How sure are we that the person does not have gray eyes? Of course, we can widen the selection, to make sure that we do not miss out the person that we are looking for, by specifying:

WHERE eyecolor IN ('BLUE''GRAY')

But now we have widened the selection so much again that now probably we have selected half the rows in the database. And more awkwardly, all people selected are considered equally likely to be a match.

In real life, if you were given the task of identifying a person by eye color, and you believe the person in front of you has blue eyes, would you consider every person in the database with gray eyes an equally good match? Most likely not!

If we next consider biometric data, like the extracted features of a fingerprint, the 'Boolean issue' grows even bigger. Even when a single person with a single finger presses that finger twice shortly after each other against the same fingerprint scanner, neither the image nor the extracted features will be identical. They will be similar, but not identical! If normal database arithmetic would be used for that, the FRR would be close to (if not) 100%.

4.2 Dealing with missing data

In a security setting, a database being searched very often contains data (or profiles) from different sources. For example, the FBI keeps 10 fingerprints for each person, while US Visit keeps 2 (index finger) fingerprints. If we combine such data into a single database, certain profiles will then have 10 fingerprints and others will have 2 fingerprints, with remaining

fingerprint fields containing the NULL value. If we search the database with a left thumb fingerprint (LTF) and the two index finger fingerprints (LIF and RIF). We can now phrase the SQL WHERE-clause as follows:

WHERE <ltf from database> = <ltf value captured>
AND <lif from database> = <lif value captured>
AND <rif from database> = <rif value captured>

which will not find any row where one of the columns in the database is empty. So even a full match on the other two criteria would skip those rows!

We can improve the query to cope for this, by allowing NULL values as follows:

WHERE (<ltf from database> = <ltf value captured> OR <ltf from database> IS NULL)
AND (<lif from database> = <lif value captured> OR <lif from database> IS NULL)
AND (<rif from database> = <rif value captured> OR <rif from database> IS NULL)

This query will probably find too much, because rows where none of the columns is filled will return a 100% match as well. Additionally, rows where some columns have the NULL value may also match for 100%. The result for this query is that an actual match on all three criteria is equivalent to a row where all these columns contain the NULL value. This is clearly not a desirable outcome.

Databases do not handle missing data well, or phrased more accurately, they do not deal with missing data in a meaningful manner.

4.3 Dealing with partially correct data

Data can be entered into the database incorrectly for various reasons. Errors can be caused by data entry mistakes or they could be intentional (e.g. fraud or spoofing). Let us take our very first example about the eye color. If a person was incorrectly enrolled as having brown eyes, rather than blue eyes, then in a Boolean logic environment this person would never be found again if eye color were one of the search criteria. We cannot reasonably extend the search to cope with all possible errors, because each such extension would then qualify as a 100% match, i.e. Boolean True. Effectively then the whole database would match any given search query.

4.4 All criteria equally important

A fundamental issue with databases is that they operate in two-way (true/false) or three-way (true/false/absent) logic. Let us take the next example, where we search on a fingerprint (FP) and a facial image (FI). Fingerprint matching has a substantially higher accuracy than facial recognition.

So logically we would expect the match to return, in descending sequence of relevancy, the following results:

1) FP match and FI match;
2) FP match and no FI match;
3) no FP match and FI match;
4) no FP match and no FI match.

In a Boolean environment we can create a query where either the first result matches to a TRUE and the others to a FALSE (if we assume an AND condition) or the first three results match to a TRUE and last to a FALSE (if we assume an OR condition). If there is a full match on all criteria, then that clearly should be presented as the one and only result. However, that will not often be the case if some criteria are spoofed or falsified.

4.5 Too slow

Biometric criteria are never exact. A biometric trait cannot be expressed as a unique identifier, so that each time that biometric trait is presented it will produce that same unique identifier. This means a lookup using an index or comparable method is not possible. Rather, the matching algorithm must actually compare the search profile to all stored reference profiles and come up with a score that represents the likelihood of a match of each of the combinations. Given a set threshold, it is then possible to determine whether a match is genuine.

Given this process of comparing biometrics, it is impossible to index such criteria. As a result, the database must be traversed row-by-row, which is referred as a tablespace scan. Tablespace scans are extremely slow, as they do a linear exhaustive search through all of the data.

In today's servers, with a typical I/O speed of 100 Mbytes/sec, and a row size of 2 K bytes, a tablespace scan would be able to read ~50,000 rows/sec. If US Visit grows to an expected database size of 500 million profiles, each search would take at least 3 h. With a million visitors per day, this would certainly become an unmanageable challenge.

5 What is fuzzy matching?

As demonstrated in the previous section, Boolean logic has several limitations that make it unsuitable for biometric search and identity matching. Some of the functional issues can be resolved by what is referred as 'fuzzy matching'. Where Boolean logic is 'black and white', fuzzy logic recognizes shades of gray and uses 'calculated similarities' to return a ranked list of results. The best match is returned first and the least match is returned last. Fuzzy logic is not only useful for handling biometric searches, but also for

searches on biographic data. In short, fuzzy logic makes searching much more effective and meaningful.

5.1 Gliding scales

As we argued in the previous section, a Boolean approach is very rigid. Searching for a person that looks like 30–40 would not find a person of 43, while such a person of course could look very well look a few years younger. Fuzzy logic supports the concept of 'gliding scales', where values slightly out of the specified range will also be found. However, such values slightly out of range are then considered a less perfect match than values within the range and are therefore given a less than perfect match score. This concept is illustrated in Fig. 5.

In the above example, we are looking for a person between 30 and 40. If we find people between 30 and 40, they would match 100% (Boolean True). The gliding scale will allow us to find people between 20 and 60, and outside this range it is considered no match (Boolean False). From 30 down to 20, the match score declines from 100% down to 0%. This is similar to how humans would think about the problem as well. As the actual age approaches 20, it is more and more unlikely that the person could look like a person between 30 and 40 years old. The same applies to a person being a little older. The example shows a record with a person of 43 years old, that matches 85% with the given search specification.

Gliding scales make a range specification (numeric, date and time fields typically) more considerate, more forgiving than a Boolean condition would.

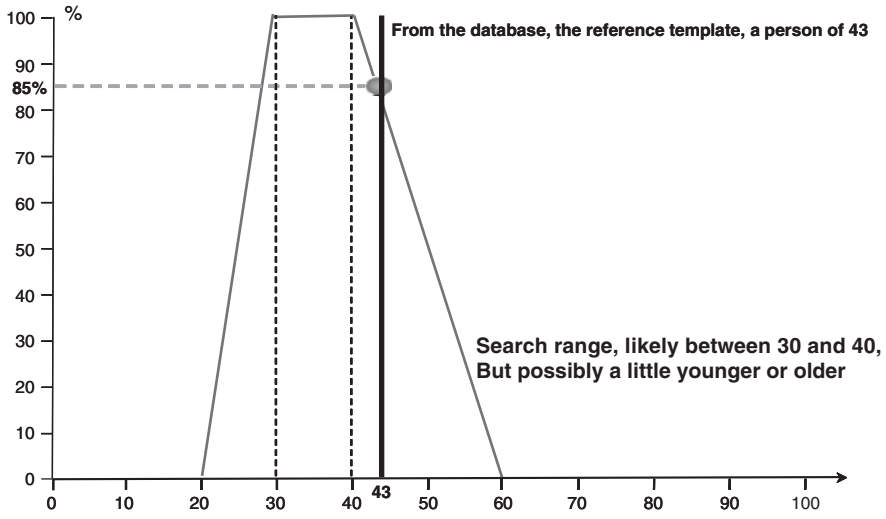


Fig. 5. Fuzzy scale for database matching.

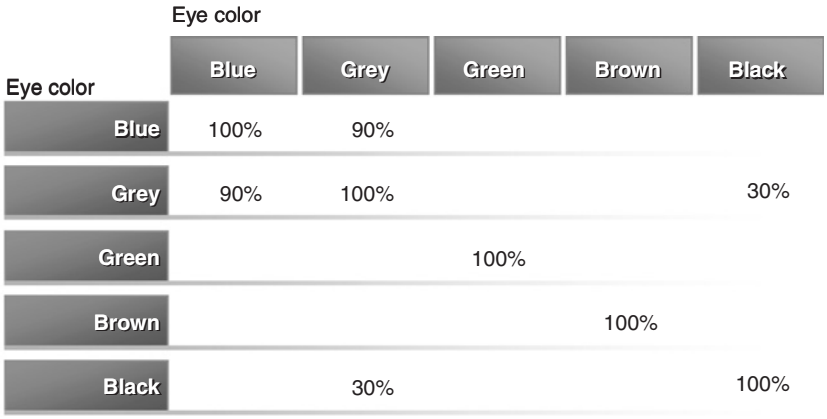


Fig. 6. Affinity matrix.

5.2 *Affinity matrices*

What gliding scales are for range specifications, affinity matrices are for pick-lists or discreet choices. A good example is eye color. Blue eyes can very easily be mistaken for gray eyes, especially when observed swiftly, from a distance or under sub-optimal light conditions. Choosing the 'wrong' color would make the person 'unfindable' in a conventional database. With fuzzy logic and affinity matrices, this can be resolved.

Figure 6 shows that certain colors have an affinity with each other, while others have no affinity whatsoever. Of course, these equivalencies are user and application dependent, and can be modified as appropriate. But their net effect is that searching for a person with blue eyes will also find persons with gray eyes, although those will be ranked lower.

5.3 *Weighted criteria*

Weighted criteria are an important aspect of fuzzy logic. When people perform a match for themselves, for example, choosing a car or a house, they usually consider certain criteria to be more important than others. These preferences and weights are thus user dependent. In an identity matching context, we have seen that certain biometrics (or even biographic criteria) are more/less reliable, accurate and/or unique. In other words, the confidence level we would give to a particular criterion varies. With fuzzy logic, this may be translated into weights. A criterion in which we have high confidence receives a high weight, whereas a criterion in which we have low confidence receives a low weight.

How does that translate in practice? We have seen that a match on facial scans is less accurate than that on fingerprints. If we assume that we have a database with two people, where person A matches on face but not on

Table 1
Assigning weights for missing criteria

Search	Database			
Search criterion	Weight on search criterion if match	Weight on search criterion if unknown	Person A	Person B
Eye color = blue	100	25	Eye color = gray (90 points)	Eye color = unknown (25 points)
Age = 30–40 (20–60)	50	5	Age = unknown (5 points)	Age = 43 (42.5 points)
		Score (maximum 150 points)	95 (63%)	67.5 (45%)

finger, and the other person B does not match on face but does match on finger. Everyone will agree that B is a more likely match than A. This can be achieved by assigning a higher weight to the finger criterion than that to the face criterion, when fusing the scores to determine the overall match score.

5.4 *Weight factors and missing data*

We saw in the previous sections that databases do not deal well with missing data. Fuzzy logic allows you to assign a particular weight when a criterion is searched, but that criterion is missing in a profile from the database. Take the example illustrated in Table 1.

In the example illustrated in Table 1, it can be seen that person A, given the weight factors defined, is the best match of the two. The penalty of a certain criterion without content can be user specified, depending on the particular situation. Do notice the 90% affinity between blue and gray eyes and the use of gliding scale on the age property, as discussed above. An alternate possibility is to assign a negative weight to an absent value, which would penalize it even more than just less points. We could also use the weight 'ignore' in case a value is absent, which would indicate that the match should ignore matching on this property. If we use 'ignore' weight for both eye color and age, then person A would score 60% (90 points out of 150), but person B would now score 85% (42.5 points out of 50).

5.5 *Custom match functions*

So far we have been talking about relatively simple criteria, like eye color and age. Generic database technology and matching engines of course support these data types. However, identity matching (in the context of security) is typically done on more complex criteria, including biometrics. Firms specialize in developing very advanced, very effective and efficient algorithms. Therefore, the vendor supplying the database infrastructure and fusion platform is likely to be a different vendor than those supplying

the biometric algorithms. There is also not a single vendor that supplies matching algorithms for each different biometric. As a consequence, it is extremely important that the fuzzy logic database and fusion platform can invoke external specialized algorithms to compare the values (search value versus value from database) for these complex criteria. It is also important that when new algorithms become available or improved versions of existing algorithms, these can be replaced on the fly without any downtime.

6 Case study—border security[2]

6.1 The case

After 9/11 most governments reacted with measures to protect their 'homeland' and hence their borders. In the past, most countries only required a passport and sometimes a visa, where the visa is based on the passport, for entry. With a new passport (expressing a new identity and costing just a few dollars to forge in some countries) a new visa could be requested. Citizens from some countries do not need a visa, the 'visa waiver countries'. Some countries require a form to be filled out and other countries just issue a stamp in the passport representing a temporary visa. As a result, the people that should not be granted access to a country had no problems getting access! Preventing unwanted people from entering a country requires two major components. First, create a list of unwanted people, a so-called watch list. Second, create a system which can easily check a person's identity. This method should not be based solely on passports and visas, since these are subject to theft and fraud.

In the case study we have been dealing with the second issue, creating a system which can easily check a person's identity. As mentioned previously, determining a person's genuine identity in a short period of time is critical due to the number of people entering and leaving the country. The US is enforcing other nations to start issuing passports with biometrics encoded; typically face and fingerprints are being used. We have seen that the accuracy by which a person can be identified on these three criteria is $\sim 10^{-10}$ (see examples later in the case study). Because of the potentially vast amounts of data and the number of visitors entering and leaving a country, this accuracy is still not enough. This is why, for example, US Visit is considering capturing up to 10 fingerprints and possibly other biographic and biometric data as well.

[2] *Disclaimer*: This case study has been written without any knowledge of classified information. All information on which this case study is based has been gathered from the Internet, public events and non-classified meetings with people involved in security agencies. The case study was developed to determine the impact of biometric fusion. While the case study deals with a fictional scenario, it is the author's strong belief that it is a realistic solution to the challenges faced by many government agencies.

6.2 Verification versus identification

Currently US Visit uses the passport nationality and number as its primary key to store information on a person, which includes the captured images of face and two index fingers. If the holder of a passport enters the US, the passport nationality and number are looked up. If the passport entry was found in the database, the biometrics of the visitor are then verified against the biometrics from the database. This check prevents people to use a stolen passport.

In 2004, ~180,000 people entered the US daily from visa waiver countries (i.e. not requiring a visa and hence are not screened upfront and thus may require further screening at the border). These visitors were required to be matched against the FBI watch list to identify a person as a potentially wanted person. The current capacity of the FBI system lacks by a factor 10–20 and the best search times against this watch list database are 10 min.[3] Clearly, searching a watch list of a few hundred thousand profiles is a relatively simple task, both in terms of performance and in terms of accuracy. When US Visit reaches its goal of building a database of 500 million profiles, which is searched a million times a day, both performance and accuracy become major issues.

6.3 Simple search—face and two fingers

The database we have used for our case study consists of 2700 profiles, holding (statistically correct) synthesized face- and fingerprint images. Each profile holds a single face image and the fingerprint images of both the left and right index fingers. Another database of 40 million profiles has been used for performance and scalability testing. The application that was developed for this case study tries to support the core process of US Visit, which is a person lookup by means of any combination of the face image and the images of the two index fingers. Different combinations are shown in Figs. 7–10, to illustrate the accuracy of the fused results and the effect that it has on the actual result.

When a search is done using only the face, the FAR is 5% and the FRR is 1%. Given a database population of 2700 persons, an FAR of 5% will on average return 135 profiles. Clearly the 144 results shown above are well within the likely range.

We have executed the same test with 25 randomly generated profiles with a single face and the result in Fig. 8 clearly illustrates the poor uniqueness of 2D facial recognition. The length of the result set from each of the 25 matches varies between 1 and 492!

When a search is done using fingerprint alone, the FAR is 10^{-7} and the FRR 2.5% with the algorithm we have used. Given a database population

[3]This information was taken from a letter from the Under Secretary of Border and Transportation Security to Inspector General of U.S. Department of Justice, dated December 4, 2004.

Fig. 7. Face searching. *Note*: Only face searching has been selected.

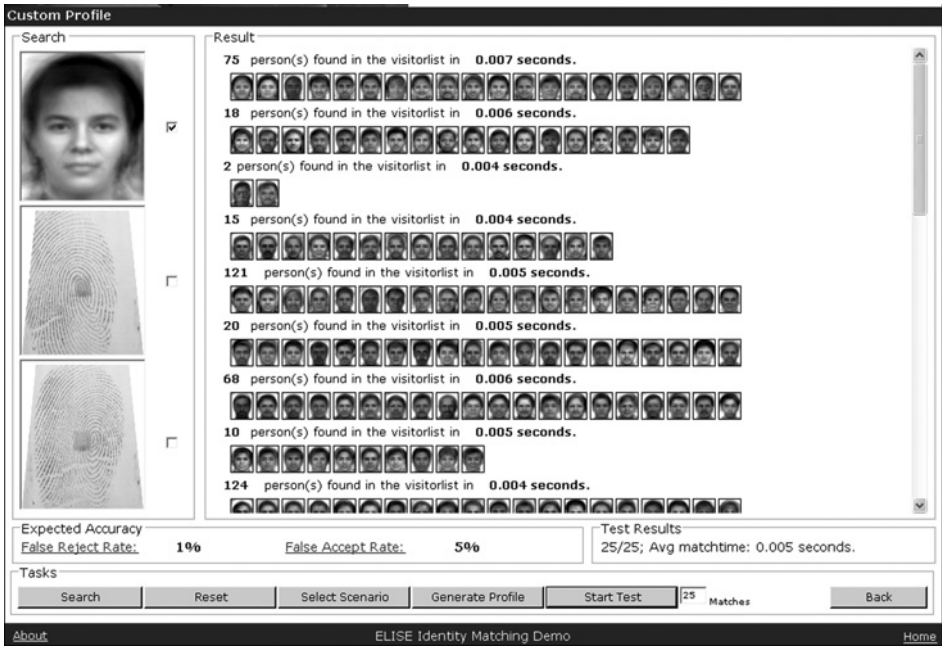


Fig. 8. Face searching (custom profile). *Note*: Only face searching has been selected.

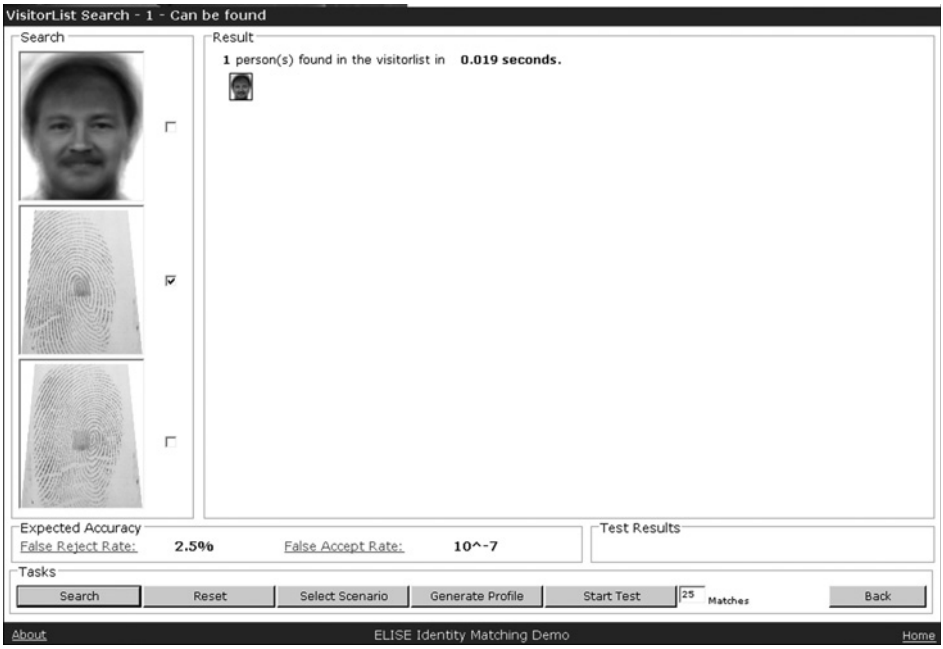


Fig. 9. Fingerprint searching. *Note*: Only fingerprint searching has been selected.

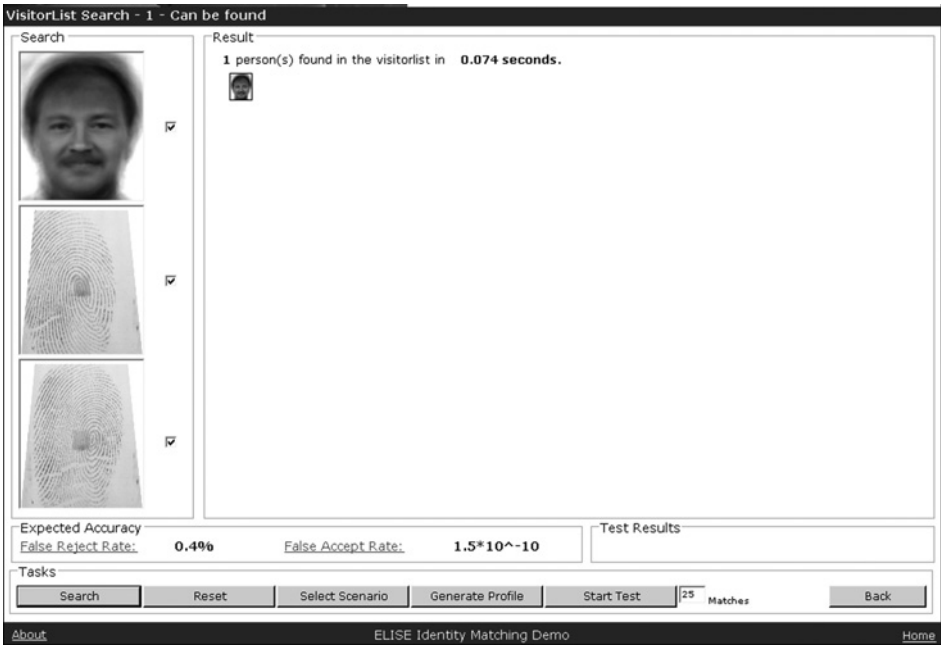


Fig. 10. Face and fingerprint searching. *Note*: Face searching and fingerprint searching on both fingers have been selected.

of 2700 persons and an FAR of 10^{-7}, we will have a 0.027% probability of a False Accept and if the person is in the database there is a 97.5% probability (100% − FRR) that we will find the person (True Accept). We know that the fingerprint that was used for this search is actually in the database. Clearly the one result shown is what one would thus expect.

When a search is done using a face and both fingerprints, the FAR is 1.5×10^{-10} and the FRR 0.4% with the algorithms used. Given a database population of 2700 persons and an FAR of 1.5×10^{-10}, we will have a 0.0000405% probability of a False Accept and if the person is in the database there is a 99.6% probability (100% − FRR) that we will find the person (True Accept). We know that the fingerprint that was used for this search is actually in the database. Clearly the one result shown is what one would thus expect. This example may look silly in the context of a 2700 profile database, but extrapolate it to the size of US Visit to understand the relevance.

6.4 Advanced search—combining biographic and biometric criteria

The goals of fusion are to increase accuracy (FAR), to be less dependent on some criteria being falsified and to reduce the post-processing issues (FRR). Capturing biometrics is both time consuming and sometimes intrusive. Although technology is improving quickly in terms of accuracy, time to capture and difficulty to forge, still choices must be weighted and made. In the next part of the case study we will focus on adding easy-to-add biographic data to the search equation. We will also introduce weighted search criteria.

With biographic data we mean things like name, nationality, gender, ethnicity, hair color, eye color, height, weight, age, etc. It takes only a swipe of the passport to obtain name, nationality, date of birth (i.e. age) and possibly some more information that is kept in the passport. In addition, a simple and swift observation gives an estimate to the other criteria mentioned. However, hair that seems blond may not be genuinely blond. It only takes US$ 5 to color your hair. Eyes may seem blue, but colored contact lenses cost less than US$ 50. But gender is more difficult to disguise. High heels might make you 5 in. taller, but a person measuring ~6 ft 6 in. is really not a genuine person of 5 ft 5 in. With weighted criteria and some fuzzy logic, we can compensate for these imperfections. Weights assigned to a particular search criterion reflect the trustworthiness of that criterion. So an observation of a person's hair color would carry little weight, while ethnicity would carry substantially more weight because it is more difficult to forge. Affinity between biometric and biographic attributes can be documented through an affinity matrix to address slightly incorrect observation.

But what do these criteria bring us, despite their poor reliability and poor uniqueness? Let us explore this via our case study (Figs. 11–13).

Figure 11 shows basically the same result as with simple search. In this example, four profiles match ~50%. How do we then decide which person is really the match we are looking for?

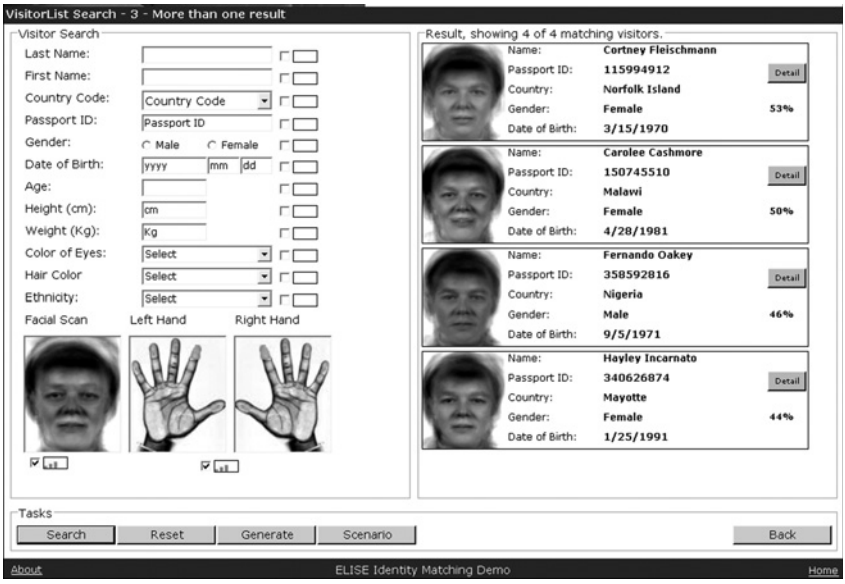


Fig. 11. Face and fingerprint searching. *Note*: Face searching and fingerprint searching on both fingers have been selected. Default weight factors have been applied for each search criterion.

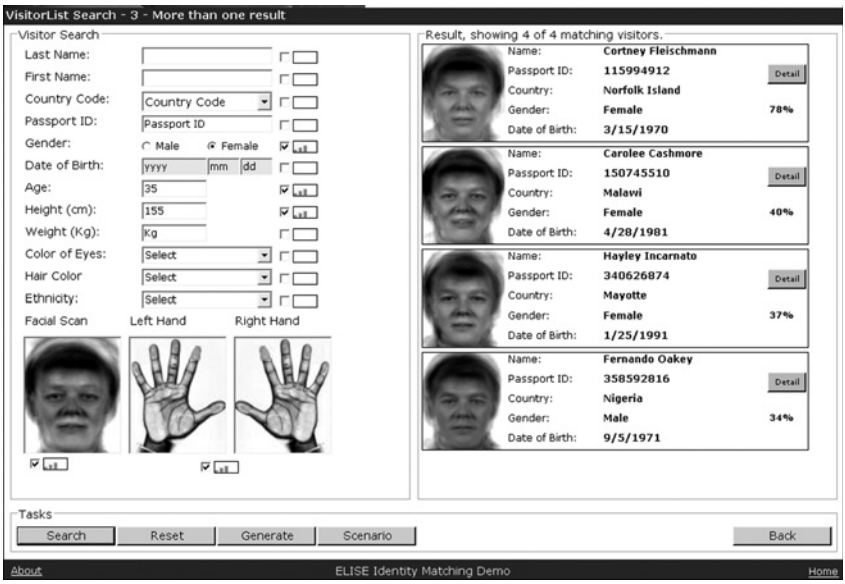


Fig. 12. Face and fingerprint searching (with biographic details). *Note*: Face searching and fingerprint searching on both fingers have been selected and also matching on gender, age and height. Default weight factors have been applied for each search criterion.

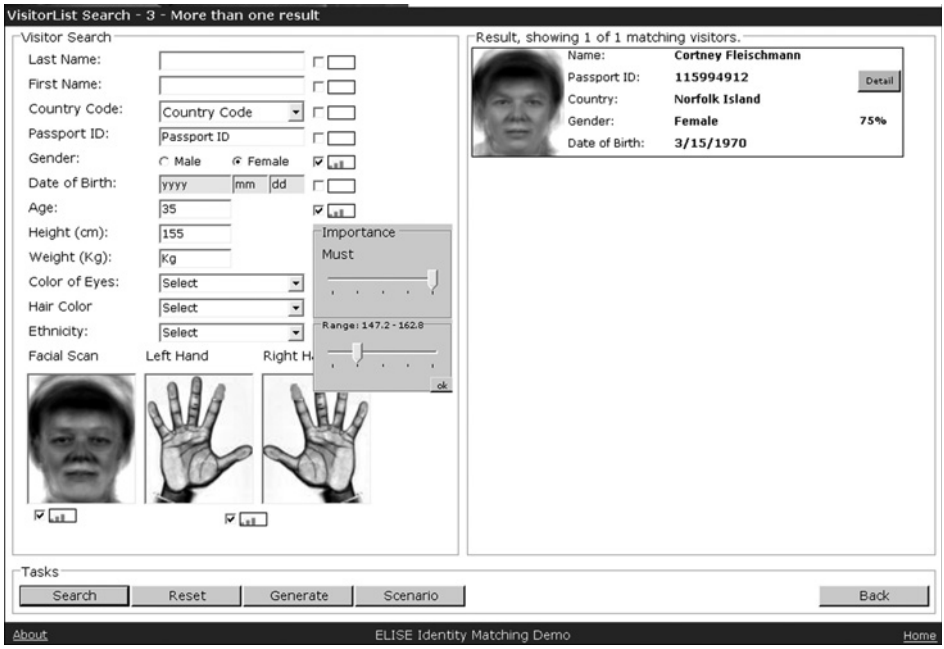


Fig. 13. Face and fingerprint searching (with biographic details/with increased weight for height). *Note*: Face searching and fingerprint searching on both fingers have been selected and also matching on gender, age and height. Default weight factors have been applied for each search criterion, and the weight of height has been raised.

Let us add some basic biographic observations. We estimate the person to be a female of ~35 with a height of 1 m 55 cm.

The result is directly apparent. The first person shown is a 78% match and the next person is at a far distance with only 40% match.

Let us explore further the effect of increasing the weight on height, if, for example, we measured the person quite precisely and as a result also reduced the gliding scale range.

A single person remains!

7 Conclusion

The hype with regards to biometrics suggests that it is the solution to a worldwide problem of security. However, when we understand the intricacies of biometrics better, we discover that there are many issues surrounding their use. New developments in this industry are constantly improving security but the improvements are also subsequently circumvented. The industry therefore will continue to undergo rapid evolutions for the foreseeable future. Much of the investment in advancing the technologies will be funded by huge governmental investments in security. The net

effect will be a tremendous innovation, in terms of both new biometric algorithms and improved algorithms for existing biometrics. The chapter's objective was to demonstrate fusion as a way to combine all these biometric algorithms to improve quality.

8 Questions for further research and debate

1) For many reasons the space of identification is 'hot'. As a result, many vendors come up with better and/or new means to identify a person. The evolution in this space is phenomenal.

   How can governments co-operate with each other to facilitate this evolution of technologies? What are some appropriate public policy strategies that governments can employ to foster continued innovations in this area?
2) With what we have been discussing, it becomes apparent that people can be identified by means of their behavior characteristics. Evolution in this space will make it more accurate and less sensitive to fraud. Criminals will have a hard time traveling in public!

   What are some implications for international espionage? Professional spies will find it extremely difficult to travel under multiple identities (i.e. multiple passports). Will spies spying in the future use their genuine identities? In short, how will advanced identity matching change the work of secret services?
3) Governments are building huge databases with people's identities. Under laboratory conditions, these are great things to have, assuming governments have no intent to misuse the information, outsiders have no access to the information and identities are not determined incorrectly. What outside threats should governments be prepared for that could jeopardize the intent of such a centralized database of people's identities?
4) The primary incentive to build a centralized database of people's identities is clearly 9/11 and the fear of more such terror. Because the sense of urgency for this argument is fading, the UK government has come up with yet another 'reason' for building such a database—preventing social security fraud.

   What other arguments can you think of that could be used as public arguments why a government should build such a database? What arguments can you make to assert that governments should *not* build such a database?
5) Governments are privatizing public assets, including roads, airports, utilities and even certain security tasks. The argument is clear, competition in the private sector leads to lower prices and better quality.

   Do you think it is conceivable that some time in the future databases holding identities of people will be operated by private companies? What are the pros and con arguments?

6) We are focusing in this book and this chapter in particular, on threats in the real world, where supposedly evil people want to do harm to the good people. Substantially less attention is given to the evil things that happen in the virtual world of the Internet. More terms are invented in this area than in any other area. Think of 'spoofing', 'phishing', 'spamming', etc.

   How long will it take before identity matching will get more attention in the context of the virtual world than in the real world? In other words, when will the identity issues on the Internet become higher priorities on the political agenda than those in the real world, if ever?

Appendix—An overview of common biometrics

This section gives an overview of the most commonly used biometrics in the identity matching space. The intent is to position each biometric, but more importantly to describe its features like 'Typical FAR', 'Discovery risk', 'Cost to copy' and 'Cause for error', as these properties of a biometric are precisely the issues that the industry experiences with biometrics. It serves as a justification of the issues with biometrics that we summarized in Section 1.

Fingerprint

Fingerprint scans recognize ridges, valleys, loops, etc., on one or more finger tips by placing the tip(s) on a scanner. In terms of revenues, the fingerprint market leads the other biometric technologies. Supported by the largest number of vendors, it is an extremely dynamic market. Note, however, that it has recently been reported to US Congress that ~2% of the population does not have a legible fingerprint!

Acquisition device	Desktop peripheral, PC card, mouse, chip or reader embedded in keyboard
Sample	Fingerprint image (optical CCD), silicon/capacitive (DC), ultrasound, RF/capacitive (AC), thermoelectric
Feature analyzed	Location and direction of ridge endings and bifurcations on fingerprint Straight pattern matching, or A combination of the two
Typical FAR	1:100,000 with an FRR of 2%
Discovery risk	Easy, forensics
Cost to copy	Very low
Cause for error	Cold finger; dry/oily finger; high or low humidity; angle of placement; pressure of placement; location of finger

	on platen (poorly placed core); cuts to fingerprint; manual activity that would mar or affect fingerprints (construction, gardening)
Key applications	Physical and logical access control for governments and banking
Key challenges	Too much choice Physiological and environmental problems Easily duped
Key vendors	Cogent, NEC, Sagem, Neurotechnologija

Facial

Facial scans, face recognition, recognize location, composition and interrelation of features of the face by means of a camera. This is a passive technology that requires no effort by the user. Face recognition is likely to grow rapidly due to surveillance and monitoring. It is also predicted to make a mark in the travel industry. The face recognition market will be driven by the fact that the technology is non-intrusive and passive, cost-effective and is good for use in government, law enforcement and casino applications. The technology is recognized by the International Civil Aviation Organization (ICAO), which develops, adopts and amends international standards to increase the safety and security of international civil aviation.

Acquisition device	Video camera, PC camera, single-image camera
Sample	2D or 3D facial image (optical or thermal). 3D image looks for complete image match
Feature analyzed	Relative position and shape of facial features (e.g. nose, cheekbones, etc.)
Typical FAR	1:100 with an FRR of 15%
Discovery risk	Easy, published image in newspaper might suffice
Cost to copy	Low
Cause for error	Change in facial hair; change in hairstyle; lighting conditions; adding/removing hat; adding/removing glasses; change in weight; change in facial aspect (angle at which facial image is captured); too much or too little movement; quality of capture device; change between enrollment and verification cameras (quality and placement)
Key applications	Physical access control, ID applications, monitoring of time and attendance
Key challenges	Cultural limitations (Muslim women, for example, may not want to reveal their faces)
Key vendor:	A4Vision, Viisage, Cognitec Systems

Voice

Voice scan technology is currently in its infancy. Voice verification will increase in use due to its utilization within the existing voice infrastructure. It is the only biometrics technology that can be used over telecommunications' networks. This market will expand due to the technology's intuitive, user-friendly, unobtrusive and cost-effective nature. Additional drivers include the existence of an infrastructure framework, high utility for wireless telephone users and opportunities driven by voice-enabled commerce.

Acquisition device	Microphone, telephone
Sample	Voice recording
Feature analyzed	Frequency, cadence and duration of vocal pattern
Typical FAR	Unknown, as yet too little objective data are available
Discovery risk	Easy, a radio or television interview may suffice
Cost to copy	Medium
Cause for error	Cold or other illness that affects the voice; different enrollment and verification capture devices; different enrollment and verification environments (inside versus outside); speaking softly; variation in background noise; poor placement of microphone/capture device; quality of capture device
Key applications	Banking, resetting corporate passwords, call centers
Key challenges	Variability of voice Poor microphone quality Susceptibility to background noise
Key vendors	Vocent Solutions, Nuance Communications, Scansoft

Hand geometry

Hand geometry and hand shape recognition are the recognition of the shape of joints, knuckles, etc., of the hand by placing the hand on a pad. A veteran in the biometric market, hand geometry is slowly losing market share to other emerging biometric technologies. Some key factors driving this market include small template size and the need for highly secured areas (such as server rooms, telecommunications' areas, etc.). Hand geometry is well-suited for such applications as access control, time and attendance, airports and border crossings.

Acquisition device	Proprietary wall-mounted unit
Sample	3D image of top and sides of hand
Feature analyzed	Height and width of bones and joints in hands and fingers

Typical FAR	1:10,000 with an FRR of 10%
Discovery risk	Medium
Cost to copy	Medium to high
Cause for error	Jewelry; change in weight; bandages; swelling of joints
Key applications	Physical access control, monitoring of time and attendance
Key challenges	Size of reader limits market High cost
Key vendors	Recognition Systems

Hand vein pattern

Hand vein patterns and hand vein recognition is the recognition of the pattern of veins in the hand by putting the hand up under a scanner. This biometric has a great future, as it is non-intrusive and can hardly be forged, due to the required ''liveness'' factor. Issues are that certain medication may impact veins.

Acquisition device	Wall-mountable unit, mouse, ATM-mountable scanner
Sample	Infrared gray-scale image of vein pattern
Feature analyzed	Subcutaneous infrared absorption patterns, vein thickness and location
Typical FAR	1:100,000 with an FRR of 2%
Discovery risk	Very difficult
Cost to copy	Very difficult
Cause for error	Unknown yet
Key applications	Governments, airports, banking
Key challenges	Full hand reader is too bulky Aging effects of vein recognition are unknown
Key vendors	Identica, Fujitsu

Handwriting and signature verification

Handwriting and signature verification is an emerging commercial market fueled by its application of paperless document management. Factors driving this market include electronic signature legislation, paperless document processing, social acceptance, cost-effectiveness and wireless devices.

Acquisition device	Signature tablet, motion-sensitive stylus
Sample	Image of signature and record of related dynamic measurements. Also signature acoustics
Feature analyzed	Speed and order of strokes, pressure and appearance of signature

Typical FAR	1:100 with an FRR of 15%
Discovery risk	Easy to medium
Cost to copy	Low to medium
Cause for error	Signing too quickly; different signing positions (e.g. sitting versus standing)
Key applications	PDA's, banking, e-government, insurance, retail sales
Key challenges	Legal-signature can convey intent (risk of 'theft')
Key vendors	Communication Intelligence Corporation (CIC), WonderNet, Sign Assured

Iris scans

Like a snowflake, the iris—the externally visible colored ring around the pupil—exhibits a distinctive pattern that forms randomly *in utero* in a process called chaotic morphogenesis that characterize the iris, by looking into a scanner. It is a fast-growing technology; iris recognition will be the second largest technology in terms of revenues by 2006. Factors driving this market include accuracy, the non-intrusive nature of the technology, financial applications network security and the introduction of new price-competitive products to the market. An issue currently is that a single vendor holds a patent on iris as a way to identify a person, and is seemingly unwilling to share this with other vendors.

Acquisition device	PC camera, video camera
Sample	Infrared iris image
Feature analyzed	Furrows and striations in iris
Typical FAR	1:1,000,000 with an FRR of 0.1%
Discovery risk	Medium, assumes small amount of cooperation
Cost to copy	Medium to high
Cause for error	Too much movement of head or eye; glasses; colored contacts
Key applications	Airports: physical access control Healthcare: monitoring time and attendance
Key challenges	Trade-off between quality and price
Key vendors	Iridian Technologies

Retina scans

Retina scans are mainly used in high-security government and military locations; retina recognition is currently seen as a highly intrusive technology. Factors that will impact this market include the need for high-security access control, small template size and the price-competitive nature of the technology. An issue is that more can be read from a retina than just identification, including illnesses, which may make people reluctant to accept this technology.

Acquisition device	Proprietary desktop or wall-mountable unit
Sample	Image of retina
Feature analyzed	Blood vessel patterns on retina
Typical FAR	1:1,000,000 with an FRR of 1%
Discovery risk	High, assumes cooperative subject
Cost to copy	High to very high
Cause for error	Too much movement of head or eye; glasses
Key applications	Airports: physical access control Healthcare: monitoring time and attendance
Key challenges	Fussy about user interaction High cost Health concerns Invasive
Key vendors	EYEdentify, Retinal Technologies

Biography

Peter Went is the CEO of WCC. In this capacity Mr. Went oversees the company operations in both Europe and North America, with a particular focus on technology. Back in 1992, Mr. Went already envisaged the constraints of RDBMS technology for matching purposes, which led him to work on a prototype matching engine. This turned out to be the foundation of the WCC product line.

Prior to WCC, Mr. Went was COO of QSD, a company that develops integral front- and back office solutions for financial institutions; CEO and founder of UniSoft, a software house and systems integration company in Prague (Czech Republic); principal consultant at James Martin Associates, a worldwide IT advisory firm; and principal consultant with Platinum Technology, a database utility vendor.

Company

WCC is a Netherlands-based software development company, which develops and markets ELISE, a high end fuzzy matching engine that is used by large multinationals and governments for matching candidates with job-openings, person lookup and identity matching. With many of its biometrics and integration partners, WCC is offering an end-to-end solution for large scale identity matching in the security space, where ELISE serves as the fusion platform.

References

Ross, A., A.K. Jain (2003). Information fusion in biometrics. *Pattern Recognition Letters* 24(13), 2115–2125.

Further Reading

BioEnable: BioEnable develops Biometrics Fingerprint recognition based products for positive identification and provides Biometrics Software and Hardware development services to companies internationally, http://www.bioenabletech.com.

Biometrics Institute: The Biometrics Institute is an independent not-for-profit membership organization, founded in July 2001. Their primary members are government and business users of biometric services and products, with other membership categories for vendors, http://www.biometricsinstitute.org.

DHS: A letter from the Under Secretary of Border and Transportation Security to Inspector General of U.S. Department of Justice, dated December 4, 2004, http://www.usdoj.gov/oig/reports/plus/e0501/app5.htm.

FindBiometrics.com: Part of the TopickZ Inc. family of sites. TopickZ Inc. is a fast-growing provider of electronic information resources. The site is a truly unique resource site; findBIOMETRICS.com can guide novice and expert professionals through every aspect of finding a suitable Biometric solution for their enterprise, http://www.findbiometrics.com/.

Gartner: Gartner is the world's leading provider of research and analysis about the global information technology industry, http://www.gartner.com.

Good reading on deception: http://www.iris-recognition.org/counterfeit.htm; http://www.heise.de/ct/english/02/11/114/.

Nandakumar, K., A.K. Jain (2005). Score normalization in multimodal biometric systems. *Pattern Recognition* 38(12), 2270–2285, http://cat.inist.fr/?aModele=afficheN&cpsidt=17129589.

NIST, National Institute of Standards and Technology, is an agency of the U.S. Commerce Department's Technology Administration, http://www.nist.gov/biometrics.

Snelick, R., U. Uludag, A. Mink, M. Indovina, A. Jain (2005). Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3), 450–455.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 6

Managing Real-Time Bioterrorism Surveillance Data

Donald J. Berndt, Alan R. Hevner and Jamie Lynn Griffiths

Information Systems and Decision Sciences, College of Business Administration, University of South Florida, Tampa, FL 33620, USA

Abstract

The development of an effective bioterrorism surveillance system requires solutions to several critical system challenges. The system must support multidimensional historical data, provide real-time surveillance of sensor data, have the capability for pattern recognition to quickly identify abnormal situations, and provide an analytic environment that accelerates investigations by epidemiologists and other responders. In this chapter, we present a framework for viewing the management of real-time data to support bioterrorism surveillance. The framework identifies four critical surveillance components: data sources, real-time data streams, historical and real-time data warehousing, and the analytic engines that perform algorithmic (automated) and exploratory (human-in-the-loop) pattern recognition. The differences between algorithmic and exploratory data analyses are demonstrated retrospectively via a case study using Florida wildfire data from the period 1998 to 2001 and their impacts on hospitalizations due to respiratory ailments.

1 The threat of bioterrorism

The threat of a premeditated biological attack on civilian populations is of real concern to all nations. Recent events, such as the sarin gas attack in Japan and the anthrax contamination of letters in the U.S. postal service, demonstrate the devastating consequences of bioterrorism both physically and psychologically (Ackelsberg et al., 2002). Biological weapons can be based on a number of different biological agents and can take many forms of distribution, making the detection and response to biological attacks very difficult to prepare for (Relman and Olson, 2001).

Analyzing the challenges of detecting a bioterrorist attack from a data-oriented perspective highlights a fundamental distinction between traditional surveillance systems (a core function of public health) and syndromic surveillance systems that are intended for early detection of threats such as bioterrorism-related epidemics. Well-established surveillance systems, such as the National Notifiable Disease Surveillance System (www.cdc.gov/epo/dphsi/phs/infdis.htm) with a lineage back to 1878, are focused on well-defined diseases and rich communication between authorities and clinicians. Therefore, healthcare data from such surveillance systems have rather unique characteristic of being validated by expert clinicians making evidence-based diagnoses. Syndromic surveillance systems are based on pre-diagnostic data collected in real-time or near real-time that relate to the prodromes (very early signs and symptoms) and syndromes (more recognizable symptoms) of specific illnesses. The goal is to develop systems that would allow detection of bioterrorism-related outbreaks before public health authorities might detect them through other means. However, the effectiveness of such systems remains unknown and the subject of considerable attention (Buehler et al., 2003).

Public anxiety over terrorist attacks is at an all-time high, and national, state, and local governments are being asked to protect citizens from these threats to community health—now, not later. Surveillance systems to assist in the detection and prevention of such bio-threats are seen as potentially important elements in our national plan (Zeng et al., 2005). Unfortunately, the data necessary to fuel such systems reside in a multitude of disparate, distributed data sources among hospitals, clinics, pharmacies, water treatment facilities, labs, emergency rooms (ERs), etc., and we cannot wait for solutions with long development and implementation times. Additionally, an effective bioterrorism surveillance system must receive, analyze, and react to these data in a real-time environment to support rapid, life-critical decision-making. There are both proponents and critics of syndromic surveillance systems, with no substantial evidence to justify abandoning current efforts or pursuing widespread adoption. Clearly, more research is needed (Sosin, 2003).

Ultimately, the success of such systems depends on providing additional benefits over traditional surveillance, which relies on astute clinicians, at a reasonable cost. Even if syndromic surveillance does not fully meet all challenges, the pursuit of such systems is likely to improve the practice of public health surveillance through better data management and analytic techniques. Since public health surveillance impacts many aspects of our healthcare systems, from monitoring of naturally occurring outbreaks to the formulation of health policy, we can be confident in a reasonable return on our investment.

Our goal in this chapter is to survey current research on the management of real-time data to support bioterrorism surveillance. We organize the discussion via a four-part framework of data sources, real-time data collection,

historical data warehousing, and analytic engines performing algorithmic (automated) and exploratory (human-in-the-loop) pattern recognition. The chapter concludes with a brief analysis of evolving bioterrorism surveillance systems.

2 Bioterrorism surveillance systems

How can a community, a county, a state, or a nation determine that it is being attacked by terrorists who are using biological or chemical agents as weapons within a time period short enough to prevent or at least ameliorate major negative health consequences? In the U.S., the Center for Disease Control and Prevention (CDC), in its public health response performance plan, emphasizes the ability of local and state health departments to respond to terrorist attacks. Central to this initiative is the establishment of sentinel networks that not only have the capacity to respond to an act of terrorism, but also have the infrastructure to anticipate and potentially prevent threats from being realized, or, at least, to minimize their epidemiological impact through early detection (Lober et al., 2002; Morbidity and Mortality Weekly Report (MMWR), 2004). The potential nexus of these sentinel networks should be an information system and dedicated data warehouse which have many associated functions, but with the primary objectives of detecting ongoing biochemical exposures and analyzing the consequences affecting the population health status.

Fundamentally, there are four major challenges required for such a surveillance system to be effective:

1. It must manage *multidimensional data*. In other words, it must include a range of appropriate indicators and sources of information in order to monitor as many types of threats and health effects as possible.
2. It must accelerate the transmission of findings and data to closely approximate *real-time surveillance* so as to provide timely attack warnings.
3. It must have the capability for *pattern recognition based on historical data* that will quickly identify a specific alarm or alert threshold value, raising the issue for further investigation or possible intervention.
4. It must provide an *integrated data analytic environment* that will allow epidemiologists to rapidly investigate unfolding events in a historical context, effectively present the information to other decision-makers, and assist in any response.

Figure 1 presents a framework that organizes these bioterrorism surveillance challenges. Bioterrorism surveillance systems, as well as many other public health surveillance tasks, often require the integration of diverse data sets, which are administered by a variety of stakeholders. Some examples of potentially useful data sources appear at the top of the diagram. Historically,

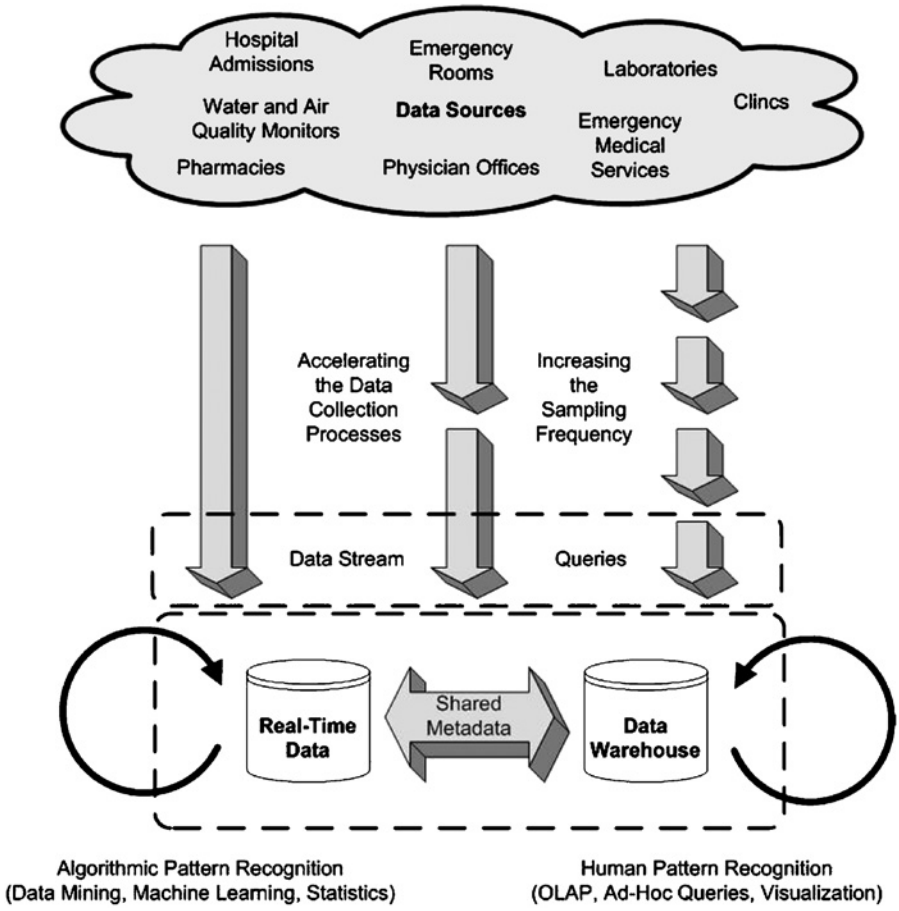


Fig. 1. Real-time data in a bioterrorism surveillance system.

many of these data sources are collected and released on long-term schedules that reflect the demands of research and regulatory activities, not the demands of real-time systems. However, the technology exists for adapting many of these data collection efforts for real-time or near real-time processing. The arrows depict the need for accelerated data collection processes. There is a growing body of research on data streams and associated query processing that seems particularly relevant in this context.

Once the real-time data are collected in a repository, there must remain a distinction between real-time and historical data. A data warehouse typically stores very large collections of historical data in (mostly) read-only environments. In surveillance systems, these data can be used to study trends and to provide baselines for interpreting unfolding situations. Real-time data may be more unstable, reflecting incomplete information, fewer

opportunities for data quality checks, with the need to extrapolate trends into an unknown future. By sharing metadata, common definitions and coding standards, the emerging real-time data can be compared with historical trends in meaningful ways. This is a critical role of data warehousing in the surveillance context.

Finally, the combination of real-time and historical data must be analyzed and interpreted by both human-driven and more automated pattern recognition techniques. Data visualization, mapping, ad hoc queries, and the drill-down operations that characterize online analytic processing (OLAP) are all examples of techniques that support the ability of people to detect patterns. More automated techniques that use data mining, machine learning, as well as more traditional statistical approaches, also attempt to discover and quantify patterns. Typically, all these techniques are used during the iterative process of discovery. One of the goals of research in this arena is to provide the infrastructure for epidemiologists and other healthcare professionals to more rapidly and thoroughly analyze situations. There is a long-term challenge to understand how to characterize abnormal events across a spectrum of potential bioterrorism-related threats.

One of the limitations of many current surveillance system efforts is the reliance on a single source of data (Buckeridge et al., 2005). Even for a single source of data, the richness of the available elements can be a limiting factor. For example, contrast the usefulness of a count of ER visits with or without additional patient characteristics such as age, gender, race, or even chief complaint. Our chance of building successful surveillance systems improves dramatically if multiple data sources are available. Just as a clinician assembles many pieces of evidence from observation, experience, or laboratory tests in order to arrive at a diagnosis, bioterrorism surveillance systems should consider many sources of evidence. The integration of multiple data sources is the hallmark of a data warehouse and a focus of this chapter. As in most data warehousing applications, the richness of the dimensions or categorizations available determine the quality of the analyses. Of particular interest are temporal and spatial characteristics that are especially relevant in the bioterrorism surveillance arena (Olson et al., 2005). Therefore, data integration and warehousing offers a key perspective from which to consider data management for bioterrorism surveillance.

3 Multidimensional data: sources of healthcare information

The multiple health effects of biological, chemical, or other types of agents or hazards require the specification and monitoring of many different information sources. Among the more obvious information sources would be hospital ERs, physician offices and primary care clinics, pharmacies, and clinical laboratories. The data taken from these sources can provide timely information about the nature of the threat, its health

consequences, and the areas and populations affected. Among the many types of information sources that have received much less attention as part of an early detection system are physical monitors which are capable of detecting hazards affecting the water supply, the air supply, or the food supply. Each of these groups of indicators, or domains, would be specified as part of an integrated, comprehensive early warning system.

The healthcare industry is characterized by complex interactions among many, largely independent, care providers and insurance organizations. A large amount of sensitive information is exchanged between these healthcare stakeholders during the delivery of care. While access to the detailed information is restricted through informal norms and explicit regulations, such as HIPAA, less sensitive information is abstracted and collected at several points in the care process. These data sets are incredibly important for understanding the processes of care, as well as for more clinically oriented research that can improve the quality of care. It is imperative that we recognize the usefulness of such data collection activities and strike a balance between privacy concerns and critical monitoring efforts. This section discusses some of the major types of healthcare data that are currently available, as well as some new initiatives, which can support a range of activities from surveillance to quality improvement.

The success of any bioterrorism surveillance system requires real-time data collection from a wide variety of sources. While this is a difficult task, it can be approached in an incremental fashion. In addition, the Internet and rapidly developing wireless networking infrastructure provide expanding opportunities to use off-the-shelf technologies. Data from many organizations can contribute to an effective surveillance system. Most database systems now incorporate many features for constructing distributed systems, thereby linking physically remote data sources. Many of the interoperability concerns have been reduced since most organizations now rely on one of the few dominant relational or object-relational database engines. The remainder of this section surveys the principal sources of data to support a comprehensive bioterrorism surveillance system.

3.1 Syndromic data

There has been growing interest in systems that support syndromic surveillance (www.syndromic.org) and even an open source initiative to foster widespread adoption (Tsui et al., 2003). As noted earlier, syndromic surveillance is focused on collecting and analyzing data in real-time (or near real-time) so that possible bioterrorist attacks can be detected before public health officials might otherwise recognize the situation. Most of these systems use syndromic or pre-diagnostic data with the goal of detection prior to a clinician making a diagnosis in an ER or hospital setting. Pre-diagnostic data includes prodromes or syndromes, the early signs and symptoms of illness that might be reported by ER clinicians. Other pre-diagnostic data

might include sales data from pharmacies and even over-the-counter drug sales from grocery stores that provide indications of self-treatment (Fienberg and Schmueli, 2005). The effectiveness of such systems remains unproven, but research efforts continue to explore new data sources and pattern discovery algorithms. Rather than focusing exclusively on single sources of pre-diagnostic data, or even post-diagnostic data, a collection of distributed systems and an integrated data warehouse that provide multiple perspectives on an emerging outbreak would seem to be the preferred infrastructure.

Standards are critical in any application that relies on the real-time digital transmission and archiving of data. This includes coding standards as well as functional standards that allow systems to interoperate. For example, there have been significant developments in the electronic reporting of laboratory results for notifiable conditions (Barthell et al., 2002; Overhage et al., 2001). Recognition of the need for more timely surveillance has been demonstrated for diarrheal disease, emergency department-based emerging infections, influenza epidemics, nosocomial infections, salmonella, meningitis, tuberculosis, and various clinical event monitors (Lazarus et al., 2002; Rotz et al., 2002). Overall, there has been increasing attention paid to rapid laboratory testing, handheld and field-deployable devices, as well as clinical decision support systems (Bravata et al., 2004). While coding standards exist for clinical diagnoses and surgical procedures, there are far fewer standards for syndromic or pre-diagnostic data (Forslund et al., 2004). There have been efforts to standardize the reporting of ER signs and symptoms. Data from pharmacies, grocery stores, and other potential sources of pre-diagnostic data are far less standardized, but offer the chance for real-time collection via point-of-sale (POS) transaction processing systems.

3.2 Administrative data

While collected mostly for reimbursement purposes, post-diagnostic administrative data has been used extensively for healthcare policy investigations, as well as for more limited clinically oriented research. There have been significant improvements in the thoroughness of collection efforts, overall data quality, and our ability to easily manipulate large-scale data sets. Though each state is free to pursue different directions, the agencies charged with collecting administrative data are an increasingly important source of healthcare information. Among the administrative data are hospital discharge records, Medicaid claims, outpatient clinic data, and even ER visits. Roughly half of the states collect hospital discharge information. Standardized summary information across states is maintained by the Agency for Healthcare Research and Quality (http://www.ahrq.gov) in the State Inpatient Database (SID).

For example, Florida hospital discharge transactions are collected by the Agency for Healthcare Administration (AHCA) from the more than 200 short-term acute care hospitals in the state. These hospitals report every

discharge transaction, regardless of payor, throughout the state. Currently, there are ten diagnostic codes and ten procedure codes that use the International Classification of Disease (ICD) coding system to express diagnoses, interventions, and co-morbidities associated with a hospital stay. There are discussions on adding coding slots for more detail. Along with the clinical codes, each hospital discharge record also includes patient characteristics, such as age, gender, race, residential area, as well as linkages to specific hospitals and providers. There are roughly two million hospital stays per year in Florida.

An interesting new AHCA initiative in Florida focuses on collecting similar administrative data with ICD-based codes for ER visits. This could be a very important resource for exploring many research avenues, including surveillance systems. The current hospital discharge records note if the admission originated in the ER, but the majority of ER visits are not captured. In 2005, the first year of data collection, there were somewhere in the neighborhood of six million emergency room visits collected.

3.3 Vital statistics data

One of the earliest public health data collection efforts involves tracking vital statistics, including the recording all births and deaths. Vital statistics data are collected by the states, but there are federal laws requiring that certain data be reported to agencies such as the National Center for Health Statistics (NCHS). Again, these data sets can be used to investigate many interesting healthcare issues. Prenatal care, teen pregnancies, and smoking habits are among the items collected by many states as part of the birth records. Of course, death certificates record the cause of death using detailed codes (such as the ICD system) allowing investigators to consider the relative impact of certain diseases, violent crimes, or even accidents.

3.4 Disease registry data

Disease registries maintained at the state and federal levels often focus on conditions that are deemed reportable under government regulation. In addition, there are disease-specific registries that may be both publicly or privately funded through major advocacy organizations. Good examples of targeted repositories are cancer registries. For instance, Florida maintains a statewide cancer registry at the University of Miami (http://fcds.med.miami.edu/) that has been collecting data since 1981. The data are available at the event level (typically without identifying information) detailing the demographic characteristics, tumor size, cancer stage, and even treatment strategies. Other disease registries compile detailed data on birth defects, sexually transmitted diseases (STDs), Alzheimer's disease, and many other health challenges. While many of these data

sets may be only indirectly useful for bioterrorism surveillance, the long-term experiences of designing and maintaining such registries are sure to be relevant.

3.5 Professional society data

The Society of Thoracic Surgeons (STS) (http://www.sts.org) and the American College of Cardiology (http://www.acc.org) are two professional societies that provide excellent examples of targeted data collection initiatives. Both of these organizations have used their clinical expertise to design data collection instruments that allow for very specific analyses, including risk adjustment, and the calculation of important outcome measures. Though participation is voluntary, many large cardiac programs across the nation actively use these systems to monitor their outcomes and implement quality improvement initiatives. For example, the items collected at the case level for the STS National Database include patient demographics, pre-operative risk factors, past treatments, medications, and the details of the current surgery. The data collection instruments are even supported by some software vendors to allow for integration with existing hospital information systems. While many surgical procedures are not emergent, and therefore would not be directly relevant to bioterrorism surveillance, these organizations demonstrate a level of sophistication possible through the coordinated actions of care providers. Other professional organizations may provide expertise across a spectrum of health issues that are more relevant for bioterrorist threats. However, any developments in specific surveillance arenas contribute to the realization of more comprehensive public health initiatives that will be useful in times of crisis and routine care provision.

3.6 Federal and state survey data

The federal government and many states regularly conduct surveys designed to support the investigation of many health-related issues. These surveys often include a mixture of very specific items, as well as more general lifestyle characteristics. The surveys may be administered to randomly selected samples that reflect the current population demographics or as panel surveys that follow a particular cohort using multiple surveys over extended periods. Though this type of data seems less relevant in a bioterrorism context than for more traditional public health surveillance, the definitions, standards, and collection methods still provide models that may influence evolving systems. Many examples of survey data can be found on the U.S. Census Bureau website (http://www.census.gov), with some particularly comprehensive surveys at http://www.factfinder.census.gov, the American Factfinder area.

3.7 Provider and insurer data

The healthcare marketplace includes a spectrum of stakeholders including direct providers, pharmaceutical companies, medical device manufacturers, benefit management organizations, and many forms of insurance companies. All of these organizations collect, analyze, and possibly supply data that may be of use in many surveillance contexts. In the short-term, most of these data collections would need to be explored through direct collaboration with specific organizations. However, through government and industry initiatives, many sectors of the healthcare market could become critical data sources.

4 Timeliness: real-time information

Timely detection is a key requirement to avoid the most serious negative consequences of any future terrorist attack involving biological or chemical agents. Timeliness, as measured by the interval separating the event from its detection, has been studied as a performance requirement for an early warning system (Wagner et al., 2001). For example, an assessment of timeliness requirements for inhalation anthrax, with scenarios ranging from treatment starting on the first of 6 days to no treatment at all, suggests detection after 3 days is nearly useless and the cost accumulation during the steepest part of the curve (between day 2 and 3) is $200 million per hour (Kaufmann et al., 1997)! A CDC study of tularemia suggests that starting treatment on the day after the attack reduces the mortality rate by two-thirds, but treatment has no effect if started as late as the fifth day.

Current systems of reporting and notification for many healthcare data sets involve reporting from the local level, to the state, and on to the federal level. This highly centralized process batches the data at several levels and 'timely' reporting can actually take from a few weeks to more than a year to traverse these organizational levels. Serving as an additional barrier to the development of the real-time systems is the fact that these necessary data elements reside in a multitude of disparate, distributed data sources as discussed above, complicating a timely implementation cycle. Many of the indicators that could be used in surveillance systems are not even identified, nor collected, and the logistical problems of capturing this information can be considerable for source organizations. Finally, the data discovery and categorization methods necessary for securing access to disparate, distributed data sources in real time requires unique and powerful technologies for identification, collection, integration, quality control, querying, reporting, and dissemination. The technologies for building distributed systems that would allow timely transmission from local archives to regional or national repositories are available in middleware components, as well as direct database linkages (Forslund et al., 2004). Initiatives focusing on bioterrorism

surveillance may provide the research and development necessary to make these federated surveillance systems a reality.

4.1 Managing streaming real-time data

Accelerating the identification, transmission, and collection of the healthcare data discussed in the previous section is essential to support effective bioterrorism surveillance. The goal is to move the relevant data into the surveillance system as soon as the data are produced in the monitored environment. The multiple sources of data would stream in parallel as shown in Fig. 1. There are basically two types of surveillance data streams: measurement data streams and transactional data streams (Chaudhry et al., 2005).

Measurement data streams consist of data from monitors (or sensors) on entities of interest. For example, sensors can report individual values of air quality, water quality, temperature, soil contamination, or other environmental qualities. Bioterrorism agents can be detected through changes in these sensor data. Real-time monitoring of these measurement data streams can be connected into active surveillance systems.

Transactional data streams typically provide structured data records of events that occur in the monitored environment. For example, an ER visit or a hospital admission would constitute an event of interest for bioterrorism surveillance. Additional events might come from pharmacy or grocery store sales. The event records are moved individually (a single ER visit) or in defined groups (all ER visits in the past hour or day). Most often transactional data are not transmitted in real-time to surveillance systems. However, we must find efficient means of accelerating the transmission of selected healthcare events in transactional data streams.

4.2 Streaming data requirements

Recent research efforts have recognized the importance of managing streaming data in such domains as traffic control, communication network management, geospatial mapping, astronomy, and many other fields with real-time data collection requirements (Babcock et al., 2002; Golab and Ozsu, 2003; IEEE Data Engineering, 2003; Chaudhry et al., 2005). Handling streaming data places a number of new requirements on systems beyond traditional data management. For example:

- *Continuous queries*: Streaming data queries are typically long running and persistent. Applications register queries on a data stream with the queries continually evaluated at periodic intervals of time. For example, we may query the mean temperature from an air sensor over the past 5 min.This query could execute every 10 sec as a moving window. A continuous query on transactional data may inquire as to the number of ER patients admitted to the hospital with high fevers over the past 24 h.

- *Timestamping of data*: Timestamping of data is critical for real-time surveillance applications. Data trends and cause/effect correlations are important analytic indicators. Thus, streaming data management requires tracking historical data values for important indicators. Research on temporal database systems has addressed many of the challenges of data timestamping (Jensen and Snodgrass, 1999).
- *Massive, unbounded data sets*: By definition streaming data are unbounded. There is no identifiable end to the data that must be monitored by the surveillance system. Thus intelligent decisions must be made on how to chunk the incoming data for analyses. In addition, the system must be able to accept and store massive amounts of data from multiple data streams.
- *Unreliable data*: Real-time streaming data carry no guarantees of quality. The surveillance system must depend upon the data sources and transmission facilities for the accuracy, timeliness, and completeness of the incoming data streams. The system must be robust enough to handle intermittent source and transmission failures, data arriving out of time order, and incorrect data resulting from either human error or transmission corruption or failure. Surveillance systems require occasional auditing of data quality at the source of the healthcare data.
- *Active capabilities*: Applications receiving streaming data typically include capabilities to react immediately to certain conditions identified in the data streams. While some of these capabilities can be automated procedures, others may require human intervention for high-level decision-making. Research on active databases is concerned with the use of triggers and rules for determining if environmental conditions require actions (Widom and Ceri, 1996).
- *Scalability*: Bioterrorism surveillance systems must scale effectively from environments with few active data streams to richer environments with many, varied data streams. As more and more information sources are added to the system, the ability of the system to manage multiple data streams, store large amounts of data, and integrate the new information into the analytical surveillance algorithms must increase accordingly.

4.3 Real-time data streaming challenges

Bioterrorism surveillance systems exhibit all of the requirements needed for managing real-time data streams as presented above. The technical advances in environmental sensors will allow wide ranging capabilities for monitoring the presence of biochemical agents in the air, water, soil, food, and surfaces via measurement data sources. As discussed in Section 3, many opportunities for retrieving healthcare information from transaction data sources are being explored. We can envision a large number of data streams

of both types potentially being fed into a bioterrorism surveillance system. Here we briefly discuss the research and development challenges that must be met before this vision becomes reality.

4.3.1 Data models and standards

The first important challenge is the use of a common data model for all data streams. This is essential to support a standard set of query languages and analytic tools that work on the integrated data. Research and development on healthcare data modeling standards has resulted in the Health Level 7 (HL7) data model based on relational data concepts (http://www.hl7.org). While the acceptance of HL7 for the modeling of healthcare data is expanding throughout the medical community, extensions to the model are needed for surveillance applications. An effective data model for streaming data must include the notion of time via either implicit or explicit timestamps (Jensen and Snodgrass, 1999). An important decision is whether a global clock exists for the entire surveillance system or if each data stream enforces its own temporal sequence. Other extensions include features that support required query operators as presented in the next subsection.

The ubiquitous standard for data exchange is the eXtensible Markup Language (XML) (http://www.w3.org/XML). The use of XML will allow healthcare data to be transmitted effectively among applications and systems. However, the overhead of XML for streaming data applications must be considered. Bruno et al. (2003) address issues of managing XML data streams for real-time systems and propose enhancements to improve performance. The CDC recognizes the requirements for standards in the transmission of healthcare data in the proposal for a Public Health Information Network (PHIN) (http://www.cdc.gov/phin). The special requirements for efficient, integrated, and interoperable healthcare surveillance systems are identified in the CDC's National Disease Surveillance System (NEDSS) (http://www.cdc.gov/nedss) initiative (NEDSS Working Group, 2001).

4.3.2 Query languages and streaming operators

The problems of querying data streams may seem similar to traditional database querying, OLAP, and data mining. However, the high-volume and unbounded nature of the data streams along with the need for real-time results call for new query concepts and optimization strategies. Research on query processing for data streams (Maier et al., 2005) focuses on the following ideas:

- *Data reduction*: In order to reduce the massive volume of data in a stream, three approaches are typically employed. *Data filtering*, or sampling, extracts a portion of the data stream with the expectation that a query on the selected data will produce a result similar to the same query on the full data set. *Data merging* attempts to combine the information in a data stream into more concise representations for

querying. Finally, *data dropping* (*aka* load shedding) simply reduces the data size by naively or blinding dropping data from the stream or eliminating data elements by use of a rule-based selection procedure. However, the use of intelligent data dropping schemes may require more overhead than is saved by reducing the volume of the data stream.

- *Punctuated data streams*: Queries that contain blocking (e.g., Group By, Sort) or stateful (e.g., Join, Set Operators) operators have problems executing on unbounded data streams. The concept of *punctuated data streams* provides a practical solution for these challenges. Data streams are divided into chunks by placing a punctuation marker into the stream that designates the end of a subset of the data. Upon receipt of the punctuation, the query result is produced based on the just-received subset of data or upon all data received before the punctuation.
- *Windows*: Another approach for handling unbounded data streams is *windowing*. Windowing uses semantic knowledge of the application domain to set clear start–end boundaries on the data to be queried in the stream. For example, a query may ask for the maximum temperature on an air sensor. This query would define a window of time, such as 30 min, over which to find the result. With a 30 min *sliding window*, this result can be produced as frequently as desired (e.g., every minute).
- *Memory-fitting joins*: Even with the use of punctuations or windows, the join operator can be tremendously expensive on data streams because of the need to maintain join values in memory. Research on streaming joins began with the traditional symmetric hash join (Wilschut and Apers, 1991) due to its memory conservation properties. New techniques are added to remove data values from memory that have expired from the window or punctuated chunk. Viglas (2005) surveys the latest research on adapting join operators and other relational query operators to streaming data.
- *Processing disordered data*: Data may arrive at the surveillance system out of temporal order. However, many window query operators may depend on ordered data for correct answers. The query engine must either resort the data in temporal order or process the data out-of-order. If disordered data is processed, there exist metrics for the degree of disorder found in the data that would allow us to evaluate the quality of the query result (Maier et al., 2005).
- *Synopses*: An approach for managing limited memory resources in the presence of data streams involves employing algorithms for summarizing the information content of the data stream in concise *synopses*. A synopsis of a data stream subset can be used to produce approximate answers to queries with some level of quality guarantee. For example, a data stream subset can be made into a synopsis when it has aged past a defined point in time. Of course, data can be deleted from memory or moved to a data warehouse once it is no longer considered useful for active query processing.

4.3.3 Quality of service

The ultimate effectiveness of a bioterrorism surveillance system depends on the quality of the services it provides. Quality can be measured in a number of important ways. For example, system developers must find the right tradeoffs among the following quality metrics:

- *Response time*: Timeliness of real-time data arrival to the surveillance system.
- *Transmission bandwidth*: Resources needed to rapidly move data from the source to the surveillance system.
- *Accuracy*: Correctness of the data upon arrival.
- *Memory*: Ability to maintain massive amounts of incoming data in the system.
- *Scalability*: Growth potential of the system as more data streams are integrated.
- *Robustness*: Ability to deal with corrupted and out-of-sequence data in intelligent ways.
- *Reliability*: Confidence that the system is ready and capable of detecting bioterrorism patterns.
- *Cost*: The economics of allocating resources to bioterrorism surveillance.

4.3.4 Managing multiple data streams

Finally, intelligent decisions must be made on the selection of which of the many available data streams as discussed in Section 3 to integrate into the surveillance system. Active research on syndromic surveillance indicators will provide needed support in this decision process. Issues of integrating the data over multiple streams into a common format for analysis are addressed in the next section.

5 Histories of indicator data: real-time data warehousing

While the transmission of data elements in real time is a necessary capability of an effective system, it is in itself insufficient without the means to determine when the signal received actually represents the existence of an alert. These alarm values, or alert thresholds, must be determined on the basis of historical pattern recognition that will enable researchers to determine both the existence and nature of the damaging event. Hospital admissions initiated in the ER, for example, vary from hospital to hospital, seasonally, and by diagnostic composition. These existing patterns must be analyzed before an alert threshold can be determined.

For example, Fig. 2 shows a high number of hospitalizations resulting from ER admissions for a hospital near many of the Florida theme parks and tourist destinations. (An OLAP tool produces the information presentation in Fig. 2 from the hospital admissions data in the underlying data

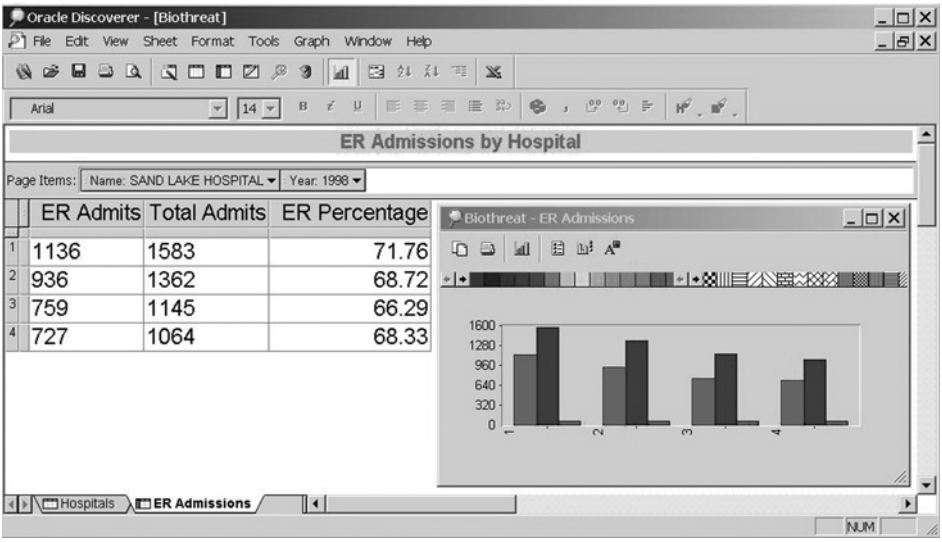


Fig. 2. High rate of hospitalizations from emergency room admissions.

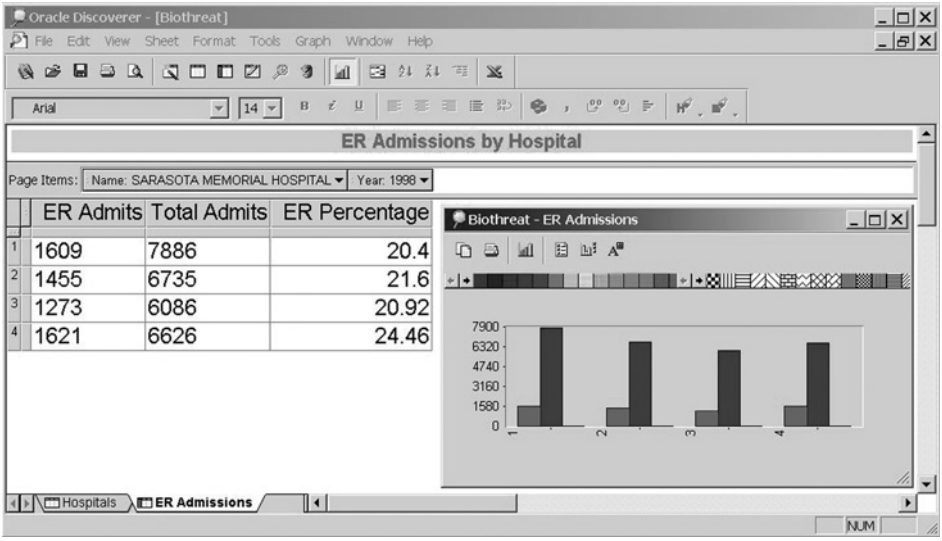


Fig. 3. Low rate of hospitalizations from emergency room admissions.

warehouse.) Using quarterly aggregations, the data shows that the ER admission rate holds steady approximately 70%. However, Fig. 3 depicts the much lower percentages that characterize a large, urban hospital located in an affluent area. For this hospital, the quarterly data show a steady ER admission rate of approximately 20%. This demonstrates the importance of

having a historical baseline of data for comparison before sounding a bioterrorism alert, as well as, the wide variation in even the simplest indicators.

5.1 Flash data warehousing

Data warehousing technologies are a natural fit for many of the surveillance system requirements. In particular, the requirement for archiving both historical and real-time data is best accomplished using a data warehouse. In addition, any pattern recognition or data mining approaches to threat detection will require a data warehouse infrastructure. Our interdisciplinary team has amassed considerable experience in using data warehousing and data mining technologies for community health status assessment (Berndt et al., 2003a). This experience has centered on supporting the Comprehensive Assessment for Tracking Community Health (CATCH) methodology with advanced data warehousing components and procedures.

The current CATCH data warehouse supports population-based health status assessments. The data warehouse serves as a historical repository of fined-grained data, such as individual births and deaths, which are used to form aggregate indicators of health status. We use a number of effective techniques to provide a thorough level of quality assurance in our healthcare data warehouse (Berndt et al., 2001). Reconstructing and analyzing historical patterns are tasks well suited to data warehousing technologies. This retrospective view is appropriate for the community-level health status reports that drove the early data warehouse development work. However, new challenges such as surveillance systems for bioterrorism require more timely data and real-time data warehousing approaches.

Corporate data warehousing efforts have followed a similar technological evolution. Data warehouses first supported the analysis of historical patterns, using the power of OLAP for queries and visualization of data extracted periodically from operational systems. Following these successful efforts, more emphasis was placed on real-time decision support activities. There is tremendous interest in moving from periodic refreshment of data warehouses toward the real-time, more incremental data loading tasks that can support up-to-the-minute decision-making (Kimball, 2002). Of course, moving from a monthly or even weekly perspective to a minute-by-minute timeframe usually means that the data being extracted may be incomplete and subject to change. There are far fewer opportunities to cleanse and transform the data in such real-time environments.

In a sense, a real-time data warehouse is like a Polaroid photograph that begins to develop. At first the image is barely recognizable, but as more of the colors darken an image begins to clarify, eventually becoming a stable snapshot of history. Early glimpses of the picture can be refined by cleverly estimating incomplete data, but the archival snapshot cannot be rushed. This process is much different than the careful extraction, transformation, and loading tasks that characterize many data warehouse staging activities.

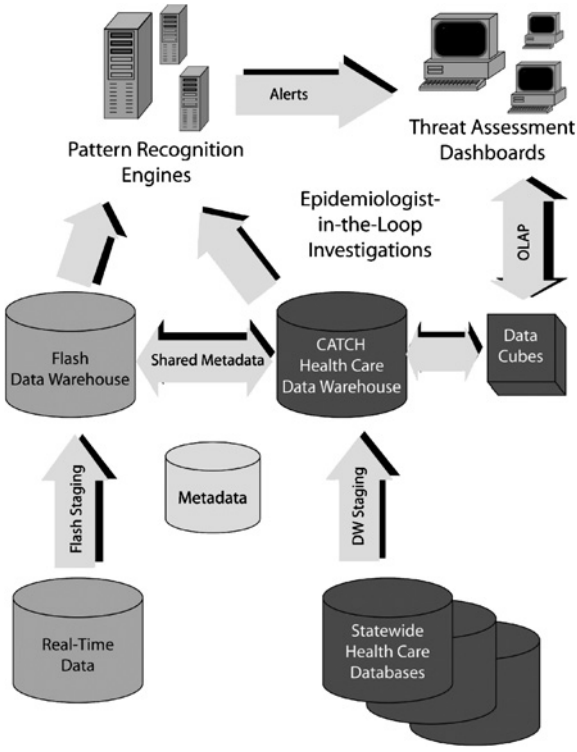


Fig. 4. Flash data warehouse architecture for bioterrorism surveillance.

Such planned staging activities are much more akin to portrait photography than Polaroid snapshots.

The challenge in bioterrorism surveillance lies in coupling a historical perspective with real-time data warehousing approaches. An overall architecture of a bioterrorism surveillance system combining historical and real-time data warehouse components is shown in Fig. 4. In this example, the existing CATCH data warehouse provides the historical information against which any new data can be compared. The new components in our prototype bioterrorism surveillance system are the real-time data feeds and associated data warehouse components (Berndt et al., 2003b). These new flash data warehouse components are used to store partially available real-time data. The components act as persistent memory for incomplete real-time data that are preprocessed for comparative queries against the archival data warehouse components, and possibly overwritten as new data become available. The flash components share common metadata with the archival data warehouse, making the important data items useful for cross queries. It is these common data items that serve as input to pattern recognition algorithms for more automated detection processes. In addition, epidemiologists can navigate

aggregated data cubes to investigate emerging situations. It is this combination of human-in-the-loop and automated pattern recognition that provides a flexible infrastructure for identifying potentially abnormal events.

6 Data analytics: algorithmic and exploratory pattern recognition

Once pre- and post-diagnostic data are integrated in a data warehouse, the challenge becomes pattern recognition. Various marker admissions can be historically monitored such as infectious and parasitic diseases, diseases of the respiratory system (e.g., those due to external agents such as chemical fumes or vapors), non-specific abnormal findings, poisonings by antibiotics, or the toxic effects of substances such as carbon monoxide or chlorine. The nature of the alarm threshold itself may vary. The actual level of the indicator value at a single hospital may serve as the alert; for example, any hospital that exceeds its own expected number of respiratory disease admissions by one standard deviation or more in any 2-h period. Similarly, an alarm threshold might be reached when a smaller increase in specified admissions is achieved (e.g., 20%) but at some number (e.g., 3 or more) of hospitals. There has been some recent work on algorithms that detect abnormal events, but much remains to be done (Siegrist and Pavlin, 2004). This is an area that will require ongoing research since it is likely that each type of threat will present unique aspects that require disease-specific adaptations. These patterns and the determination of valid alert levels can best be investigated with the creation of a healthcare data warehouse that integrates the many disparate data elements. This warehouse can then support the use of sophisticated browsing tools for either explanatory or confirmatory purposes, as well as more automated pattern recognition algorithms.

6.1 The decision-making context

The decision as to whether an unfolding epidemiological situation is in fact an act of bioterrorism is clearly a very daunting task full of uncertainty. There has been much research regarding decision-making under ambiguous and confusing circumstances. For instance, signal detection theory has been widely applied in practice and research. Signal detection theory assumes that most decision-making tasks occur under conditions of uncertainty. The basic framework proposes four outcomes for such tasks: a ''hit'' corresponds to a correctly identified event (such as a bioterrorism attack) from the signals or available information, a ''miss'' is when a decision maker fails to identify such an event, a ''false alarm,'' and finally a ''correct rejection'' of the event. This reasoning echoes the concepts of Type I and Type II errors in statistics, where the convention is to structure the hypotheses to minimize the risk of costly Type I errors, while using sample size and other design factors to control Type II errors.

The related concepts of *sensitivity*, *specificity*, and *timeliness* also describe important characteristics of the decision-making model, especially in the context of bioterrorism and disease surveillance (Wagner et al., 2001). Sensitivity relates to the level required to trigger an alarm or threshold, while specificity characterizes the accuracy or ability to correctly discriminate between outcomes. Typically, these parameters form the basis for a tradeoff, where increased sensitivity comes at the cost of reduced specificity. Lastly, timeliness is another critical issue that can often be improved by increasing sensitivity, again at the cost of other criteria. Timeliness is especially critical in the domain of disease outbreaks and bioterrorism early warning systems. Excessive delays can render effective interventions useless and dramatically reduce alternative courses of action. Since many disease or biochemical agents have unique temporal trajectories, research into the profiles of these threats is a high priority.

Many investigators have theorized and experimented in the area of decision-making under uncertain and risky conditions. Shapira (1995) offers a model of risk in managerial decision-making that further refines the relationship between possible outcomes. This model has been recast for the study of "strategic surprises" (Lampel and Shapira, 2001). While the model is more often used to characterize strategic surprises between business partners, several wartime events are used to illustrate the model, making the model very relevant for biochemical attacks. In particular, both the attack on Pearl Harbor and the Yom Kippur War provide examples where costly false alarms resulted in the upward adjustment of alarm thresholds and the resulting "surprise" attacks, despite very good intelligence. The authors suggest that it is tempting to cite an "information gap" due to less than ideal intelligence gathering activities. However, there is always incomplete information and uncertainty in such circumstances. Perfect information is usually too expensive and often simply unattainable.

These decision-making frameworks serve to highlight some important aspects of the biochemical threat detection challenge. The cost of false alarms in early warning systems for biochemical threats is extreme in terms of monetary expenditures and psychological burdens. Therefore, subsequent upward adjustments of alarm thresholds would be expected, reducing the sensitivity, timeliness, and ultimate usefulness of the system. In addition, biochemical attacks are (thankfully) exceedingly rare, providing few examples on which to refine and calibrate predictive models. Lastly, it may never be possible to definitively answer some questions regarding the origins of a particular attack, or even whether it was an intentional act or natural outbreak. In the face of such challenges, it is unlikely that a highly automated early warning system can be constructed, at least in the short term. A more appropriate goal may be to provide thorough, easily accessible, and sophisticated analytic capabilities for accelerating further investigations once early indications of a threat are identified. These types of human-in-the-loop, or more appropriately, epidemiologist-in-the-loop

systems can be supported in part by available data warehousing, data mining, and information visualization technologies. The hope would be to accelerate and enhance the epidemiological investigative processes to improve timeliness, while still controlling the specificity and associated risks of false alarms.

6.2 *Florida wildfires*

The data management framework outlined in the introductory section (see Fig. 1) draws a distinction between human pattern recognition and algorithmic pattern recognition. In order to consider examples of OLAP (human-in-the-loop) interfaces, as well as more automated pattern recognition, data on naturally occurring wildfires are used to simulate point source bio-attacks. Data on more than 100,000 wildfires spanning more than two decades have been loaded into the CATCH data warehouse. Florida wildfires show considerable variation from year to year. For instance, several years of drought conditions led to a record number of wildfires throughout Florida during the first 6 months of 2001. Over 3600 individual fires consumed nearly 320,000 acres of woodlands. Major transportation corridors, such as I-95 in the east, I-75 in the west, and I-4 across the state, were intermittently closed due to the lack of visibility from smoke and haze.

Another year of concentrated wildfires was 1998, when a string of fires burned along the Atlantic coastline. Again, the wildfires resulted in major highway closings and the widespread destruction of property. For instance, a 48-mile section of I-95, connecting Jacksonville to Cocoa Beach, was closed as a result of smoke and events such as the NASCAR Pepsi 400 in Daytona Beach Florida were rescheduled. Figure 5 depicts the year-to-year variation using a geographic visualization of the individual fires in 1998 and 1999 along with size information in number of acres burned. In 1999, we find that roughly half the acreage was burned and wildfires were spread more evenly across the state than in 1998.

Relevant to our studies of bioterrorism surveillance, the wildfires combined with the regular mix of air pollutants found in the air of urban areas, create a dangerous condition for individuals with respiratory illnesses. In the Tampa Bay area alone, the American Lung Association estimated the presence of approximately 350,000 respiratory patients who could have been affected during the spring and summer of 2001. Over the Memorial Day weekend of 2001, the Florida Department of Environmental Protection issued air pollution alerts for all of the counties in the I-4 corridor from Tampa Bay in the west to Daytona Beach in the east, including the Orlando area.

These health emergencies due to natural causes hold great similarities to a potential biological or chemical attack via airborne agents. In a similar fashion, other researchers have used wildfires to learn valuable lessons on emergency response. For instance, the study of recent California wildfires highlighted both successes and problems, such as incompatible

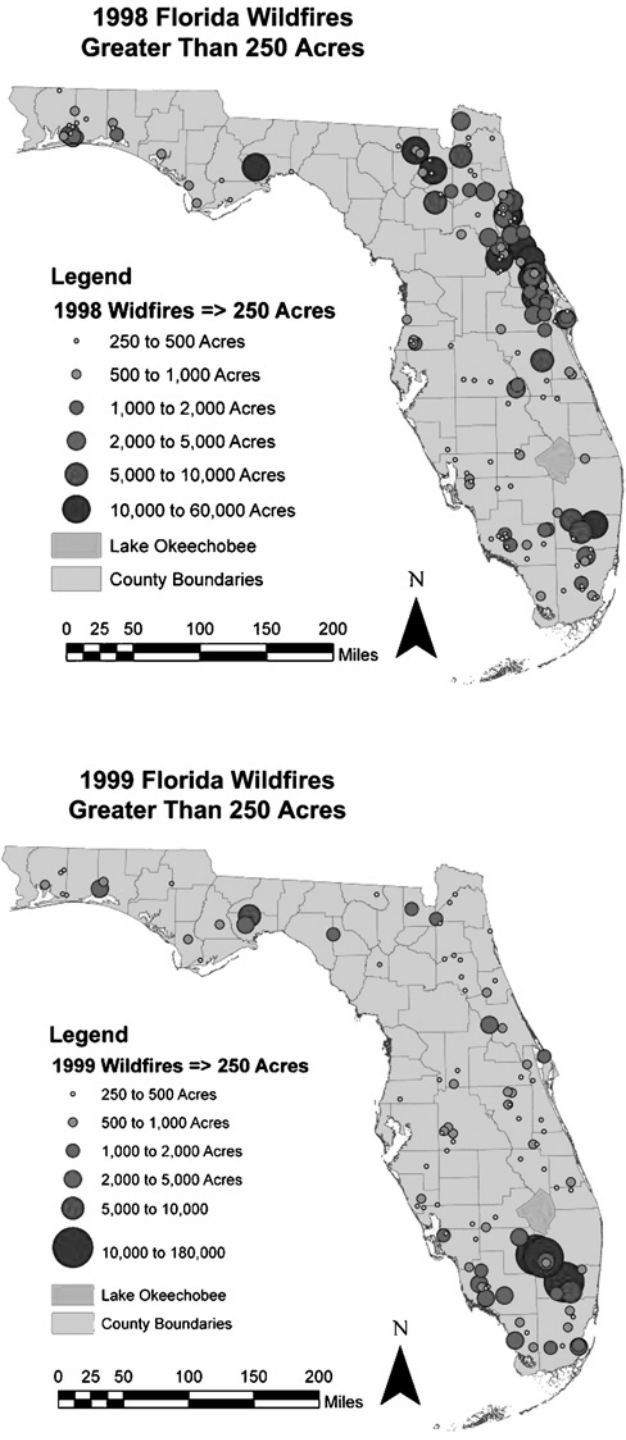


Fig. 5. Comparing 1998 and 1999 Florida wildfires.

communication systems and business continuity challenges (Ballman, 2004). In this chapter, naturally occurring wildfires are used to provide a context for understanding potential bioterrorist threats to human health. Data on hospital admissions due to respiratory illnesses are combined with wildfire events for analysis by OLAP tools and more automated algorithms. As a preliminary investigation, we use a healthcare data warehouse and associated query tools to 'slice and dice' data to investigate patterns of elevated respiratory illnesses that might have resulted from hazardous agents in the environment. A retrospective study such as this, where we know the cause of the illnesses, can inform the development of bioterrorism surveillance systems as we monitor real-time data for abnormal illness patterns and investigate any alerts or notifications that arise from automated detection systems (Berndt et al., 2006).

6.3 *OLAP and user-driven analysis*

In citing his experiences as the chief health officer for the District of Columbia, Dr. Walks notes that "the key to a successful response is the ability to communicate and share information quickly and fluidly with the appropriate people at the right time in order to make the critical decisions the situation demands" (Walks, 2003). Data warehousing and OLAP technologies are important methods for quickly analyzing and sharing information in support of critical decision-making activities. Using the data management framework, these techniques represent powerful end-user driven data exploration tools for human-in-the-loop pattern recognition. Surveillance personnel, such as epidemiologists and other healthcare professionals, can use integrated data warehousing technologies to rapidly analyze situations in ways that are currently very time consuming or simply not possible. Due to the sensitive nature of the data and potential costs of false alarms, fully automated pattern detection and alert mechanisms are unlikely to be implemented. Rather, systems that enable and accelerate analysis, feeding alerts to trained surveillance experts seem more feasible and likely to be implemented. Figure 6 presents an example OLAP screen (with tabular and graphical summaries) of selected disease data from a specific hospital in the affected wildfire areas during 2001. Quarters 1 and 2 of 2001 do seem to show elevated levels of acute respiratory infections. While this analysis is count based, OLAP tools allow users to investigate potential issues, and statistical analyses can be easily embedded into the interface. It is a combination of end-user navigation and automated pattern recognition that would seem to be the appropriate mix of technologies for bioterrorism surveillance.

6.4 *Algorithmic data analysis*

Algorithmic data analysis and pattern recognition offer the potential to automate the generation of alerts, drawing on statistical techniques, and

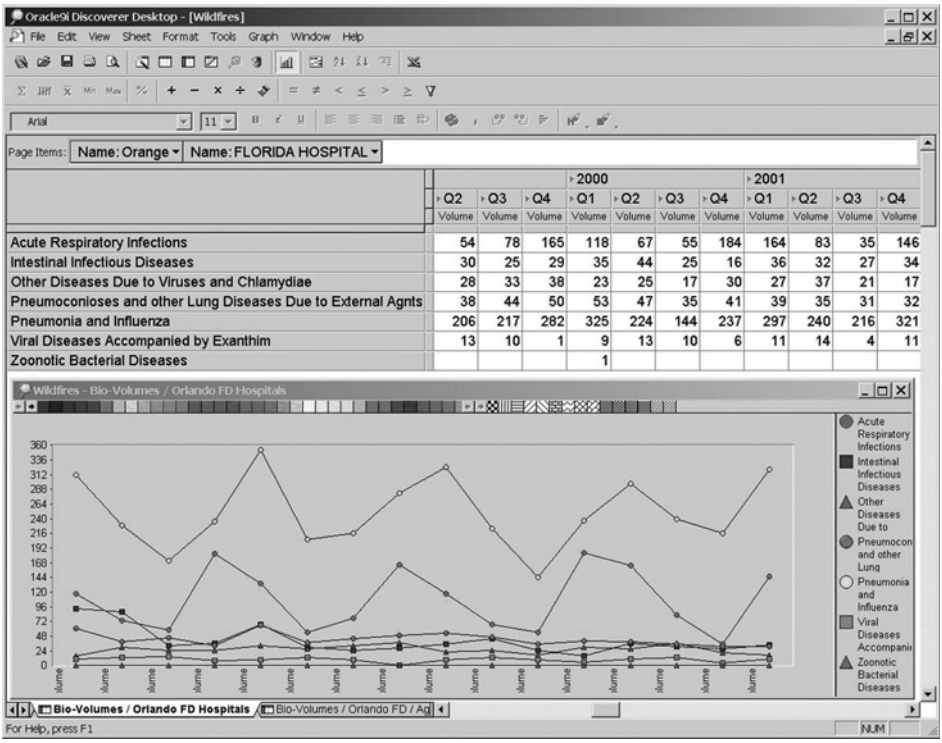


Fig. 6. Florida hospital disease data.

machine learning algorithms. While statistical approaches, neural networks, decision tree induction, and other data mining algorithms can automate part of the pattern recognition process, there is the downside risk of false alarms. As an example of a more automated approach, respiratory illnesses and wildfire data are analyzed at the yearly level from 1998 to 2000 using spatial statistics software, in this case SaTScan^{TM}. Annual data is utilized for two reasons. First, although respiratory data is available at the quarterly level, currently SaTScan only allows analysis at the daily, monthly, and yearly levels. Second, the space–time permutation statistic is sensitive to population increases, and therefore analyses at the annual level will minimize the effects of seasonal migration (a concern due to tourism in Florida).

SaTScan's retrospective space–time permutation model is utilized to detect clusters of respiratory illness hospital admissions (Kulldorff, 2005). The model was first run at the county level for all asthma cases. This resulted in no space–time clusters. The model was then restricted to patients 65+ which resulted in a primary space–time cluster in 1999 and two secondary space–time clusters in 1998 significant at 95%. Figure 7 displays the clusters along with the burn scars of all fires greater than 250 acres for 1998 and

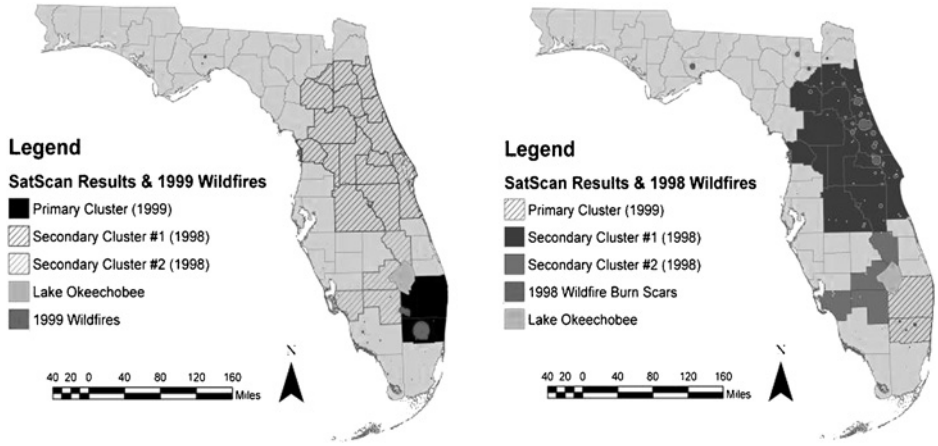


Fig. 7. SaTScan™ results and wildfire burn scars.

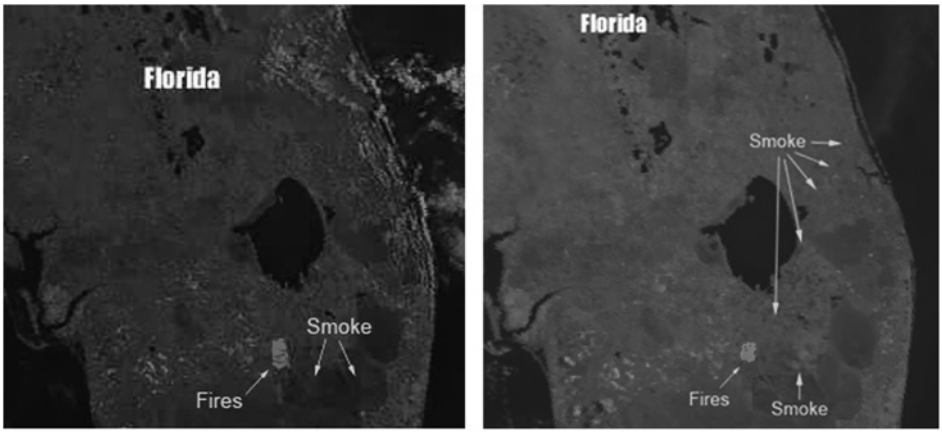


Fig. 8. Florida Wildfire #216 (1999): 173,000 acres burned.

1999. The primary cluster consists of Palm Beach and Broward Counties and is associated with three large fires along with dozens of smaller fires. The largest of the three fires burned a total of 173,000 acres, the largest single fire between 1981 and 2000 (satellite images included in Fig. 8), and the smaller two burned 25,000+ acres. The largest fire associated with 1998 burned 61,500 acres. The right side of Fig. 7 displays secondary clusters, the first of which is associated with intense fire activity throughout the fifteen counties included in the cluster.

Smoke from the 1999 wildfires in South Florida caused a 60-mile stretch of I-75 between Ft. Myers and the outskirts of Miami, known as Alligator Alley, to be closed for several days. Figure 8 contains NOAA satellite images

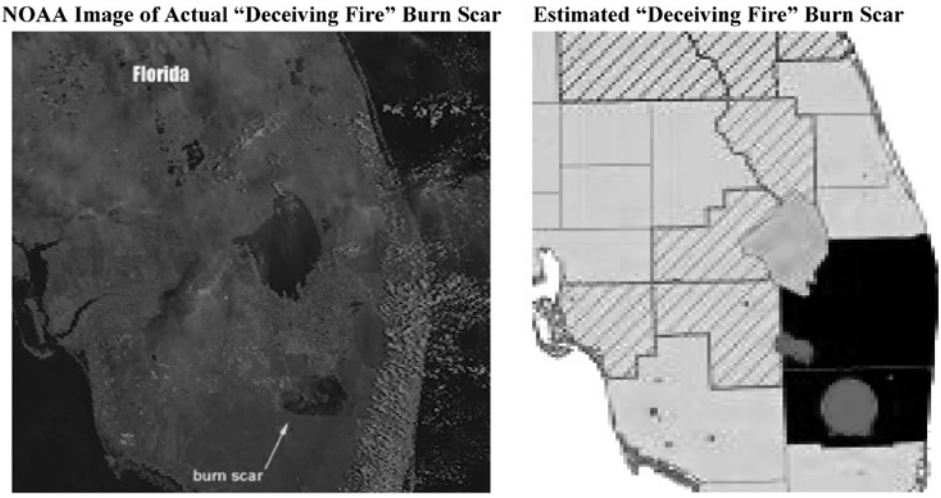


Fig. 9. Florida Wildfire #216 (1999): 173,000 acres burned.

Table 1
Asthma counts for 65+ in primary cluster counties and Miami-Dade County

	1998	1999	2000
Broward County	412	645	540
Palm Beach County	321	488	362
Miami-Dade County[a]	656	858	735

[a]Miami-Dade County is south of the primary cluster adjacent to Broward County.

of fires and smoke plumes for the 1999 ''Deceiving Fire''. The images reveal the size, extent, and rapid movement of both the wildfires and related smoke plumes. For the same fire, Figure 9 reveals the location and shape of the actual burn scar with the estimated burn scar. Firefighters dubbed this fire the ''Deceiving Fire'' because of frequently shifting wind.

Asthma counts for hospital patients age 65+ in the primary cluster counties in addition to Miami-Dade County are included in Table 1. Although Miami-Dade County was not detected for inclusion to a cluster by the retrospective space–time permutation model the authors have included asthma counts for this county which is south of the main cluster and adjacent to Broward County. We clearly see the rise in asthma counts for the year with increased wildfire activity over the previous and subsequent years.

The data available in the CATCH data warehouse do not include ER data, so hospital discharge data are used for this analysis. Since it is likely that spikes in ER admissions would be larger than changes in hospitalization rates, this analysis could be further improved with the addition of ER data. The State of Florida is currently collecting ER data, which should

be available in the CATCH data warehouse in the near future. Due to privacy concerns, the hospital data also lacks an actual date, but does include an admission quarter. Synthetic dates were added by assuming a uniform distribution across a quarter. Finally, Kulldorff (2005) points out that some outbreaks may not be clustered at residences, as the case of an exposure occurring at the workplace or other outside activities. If we assume that people go to the nearest hospital when they feel ill, then analyzing hospital location data in addition to residential ZIP code data might provide higher power to detect non-residential related outbreaks (Mostashari et al., 2003, Nordin et al., 2005).

7 Summary and conclusions

In this chapter, we have presented a framework for viewing the management of real-time data to support bioterrorism surveillance. As seen in Fig. 1, the framework identifies four critical surveillance components: data sources, real-time data streams, historical and real-time data warehousing, and the analytic engines that perform algorithmic (automated) and exploratory (human-in-the-loop) pattern recognition. We survey current research efforts in these four areas as they relate to bioterrorism surveillance. The differences between algorithmic and exploratory data analyses are demonstrated retrospectively via the use of Florida wildfire data from the period 1998 to 2001.

A majority of the research cited herein consists of isolated empirical studies or preliminary system designs for bioterrorism surveillance. Few research groups currently have the considerable system resources and access to syndromic data needed to build a fully functional surveillance system with all four components. One of the most ambitious bioterrorism surveillance efforts thus far is the Real-Time Outbreak and Disease Surveillance (RODS) project at the University of Pittsburgh (Wagner et al., 2004).

The RODS system is one of the leading research efforts in bioterrorism surveillance. From a data management perspective, the RODS system serves as an interesting example for many of the topics covered in this chapter. After preliminary tests with several hospitals, RODS was deployed to more than 75 hospitals across several states. The initial focus was on chief complaint reporting from hospital visits, a single important data stream. Data transmission relies on standards such as HL7 messaging and the XML, as well diagnostic coding schemes such as the ICD. The data are transmitted to an encapsulated database system for presentation to end-users and analysis using detection algorithms. The RODS architecture also includes a data warehousing module, which provides data integration and analysis services much like those discussed in this chapter.

The RODS system continues to be extended to handle other surveillance data types such as electronic laboratory reports, free-text chief complaints,

laboratory orders, poison control center calls, and other data sources that provide alternate perspectives on a possible bioterrorist attack. These multiple data sources can be used to build a body of evidence that supports an alert level, rather than trying to rely on a single data source and threshold. The users can navigate the integrated data through tabular and graphical interfaces, as well as by using mapping functions. Again, using the data management framework, these features are examples of human-in-the-loop analyses that allow investigators to visualize and explore emerging events. There are also RODS modules that provide for statistical analyses and algorithmic detection approaches that will automatically generate alerts.

Another interesting aspect of the RODS system is a move to open source project development (Espino et al., 2004). At first, the system was released in Java byte code form, which meant that technical support remained the responsibility of the original RODS development team. Demands for new features and customizations as basic technical support turned out to be more than could be handled by a small research-oriented development group. The developers decided to pursue an open source development model, releasing the source code and establishing a development community around the effort. This is a very interesting model and provides the foundation for affordable and extensible surveillance systems. The next few years will be a critical phase for the RODS system as open source partners adopt and extend the system. More generally, the current bioterrorism surveillance research efforts will provide the components for new integrated solutions that can be evaluated on realistic benchmark data and in the field.

8 Questions for discussion

(1) This chapter begins by outlining four fundamental challenges that must be met while implementing bioterrorism surveillance systems: (a) multidimensional data, (b) real-time surveillance, (c) pattern recognition based on historical data, and (d) an integrated data analytic environment. Briefly discuss the current solutions to each of these challenges and what research is needed to discover more effective solutions.

(2) Discuss the different sources of data that are relevant to the surveillance of bioterrorism attacks. Discuss the need for data standards in order to integrate these data for analyses.

(3) Discuss some of the technical challenges of managing multiple data streams from different healthcare and environmental sources, along with potential solutions, such as stream-oriented query languages or intelligent agents.

(4) As noted in the chapter, ''the decision as to whether an unfolding epidemiological situation is in fact an act of bioterrorism is clearly a very daunting task full of uncertainty.'' Discuss some of the potential

costs of mistakes, such as false alarms or delayed detection, and strategies to control the risks of errors.

(5) Compare and contrast more automated or algorithmic pattern recognition techniques with human-in-the-loop exploratory analyses. Do you believe that fully automated bioterrorism surveillance systems are possible?

(6) In this chapter, Florida wildfires are used as surrogates for actual bioterrorist attacks in order to experiment with different analytic approaches. What are the advantages and disadvantages of using substitute data, whether synthetic or naturally occurring, to develop bioterrorism surveillance systems?

(7) As noted by Dr. Walks (1996) , ''the key to a successful response is the ability to communicate and share information quickly and fluidly with the appropriate people at the right time in order to make the critical decision the situation demands.'' How can the information technologies and data analytic tools discussed in this chapter contribute to this key success factor?

(8) In the analyses discussed in reference to Figs. 2, 3, 6, and 7, what limitations arise if an analysis is restricted to residential locations of hospitalized patients. In particular, a state such as Florida has a large migratory population, a high percentage of elderly, and a strong tourism industry. More generally, how will data quality concerns affect bioterrorism surveillance efforts?

References

Ackelsberg, J., S. Balter, K. Bornschelgel, E. Carubis, B. Cherry, D. Das, A. Fine, A. Karpati, M. Layton, F. Mostashari, B. Nivin, V. Reddy, D. Weiss, L. Hutwagner, G.M. Seeman, J. McQuistion, T. Treadwell, J. Rhodes (September 19, 2002). Syndromic surveillance for bioterrorism following the attacks on the WTC-NYC, 2001. *Morbidity and Mortality Weekly Report*, Center for Disease Control and Prevention.

Babcock, B., S. Babu, M. Datar, R. Motawani, J. Widom (2002). Models and issues in data stream systems. *Proceedings of the PODS Conference*. Madison, WI USA.

Ballman, J. (2004). Case study: when the smoke cleared. *Disaster Recovery Journal* 17(1), 16–20.

Barthell, E., W. Cordell, J. Moorhead, J. Handler, C. Feied, M. Smith, D. Cochrane, C. Felton, M. Collins (2002). The Frontlines of Medicine Project: a Proposal for the Standardized Communication of Emergency Department Data for Public Health Uses including Syndromic Surveillance for Biological and Chemical Terrorism. *Annals of Emergency Medicine* 39(4), 422–429.

Berndt, D., J. Fisher, A. Hevner, J. Studnicki (2001). Healthcare data warehousing and quality assurance. *IEEE Computer* 34(12), 33–42.

Berndt, D., J. Fisher, J. Griffiths, A. Hevner, S. Luthur, J. Studnicki (2006). The role of data warehousing for bioterrorism surveillance. To appear in *Decision Support Systems, Special Issue on Cybersecurity for Homeland Security*.

Berndt, D., A. Hevner, J. Studnicki (2003a). The CATCH data warehouse: support for community healthcare decision making. *Decision Support Systems* 35, 367–384.

Berndt, D., A. Hevner, J. Studnicki. (June 2003b). Bioterrorism surveillance with real-time data warehousing, in: *Proceedings of First NSF/NIJ Symposium on Intelligence and Security Informatics*

(ISI 2003), Lecture Notes in Computer Science, LNCS 2665, Springer-Verlag, Berlin Heidelberg, Germany, pp. 322–335.

Bravata, D.M., V. Sundaram, K. McDonald, W. Smith, H. Szeto, M. Schleinitz, D. Owens (2004). Evaluating detection and diagnostic decision support systems for bioterrorism response. *Emerging Infectious Diseases* (www.cdc.gov/eid) 10(1), 100–108.

Bruno, N., L. Gravano, N. Koudas, D. Srivastava (2003). Navigation vs. index-based XML multi-query processing, in: *Proceedings of the IEEE International Conference on Data Engineering*, Bangalore, India, pp. 139–150.

Buckeridge, D., H. Burkom, M. Campbell, W. Hogan, A. Moore (2005). Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics* 38, 99–113.

Buehler, J., R. Berkelman, D. Hartley, C. Peters (2003). Syndromic surveillance and bioterrorism-related epidemics. *Emerging Infectious Diseases* 9(10), 1197–1204.

Chaudhry, N., N. Shaw and M. Abdelguerfi (eds.), (2005). Stream Data Management. Springer, Inc., New York, NY, USA.

Espino, J., M. Wagner, C. Szczepaniak, F.-C. Tsui, H. Su, R. Olsszewski, Z. Liu, W. Chapman, X. Zeng, L. Ma, Z. Lu, J. Dara (September 24, 2004). Removing a barrier to computer-based outbreak and disease surveillance—The RODS Open Source Project. *Morbidity and Mortality Weekly Report* 53(Suppl), 32–39.

Fienberg, S., G. Schmueli (2005). Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine* 24, 513–529.

Forslund, D., E. Joyce, T. Burr, R. Picard, D. Wokoun, E. Umland, J. Brillman, P. Froman, F. Koster (January/February, 2004). Setting standards for improved syndromic surveillance. *IEEE Engineering in Medicine and Biology Magazine* 23(1), 65–70.

Golab, L., M. Ozsu (2003). Issues in data stream management. *ACM SIGMOD Record* 32(5), 5–14.

IEEE Data Engineering Bulletin. Special Issue on Data Stream Processing, Vol. 26, No. 1, March 2003.

Jensen, C., R. Snodgrass (1999). Temporal data management. *IEEE Transactions on Knowledge and Data Engineering* 11(1), 36–44.

Kaufmann, A.F., M. Meltzer, G. Schmid (1997). The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable?. *Emerging Infectious Diseases* 3, 83–94.

Kimball, R. (February 1, 2002). Realtime partitions. Intelligent Enterprise 5(2).

Kulldorff, M. (October 2005). The SaTScan^{TM} User Guide (www.satscan.org).

Lampel, J., Z. Shapira (2001). Judgmental errors, interactive norms, and the difficulty of detecting strategic surprises. *Organization Science* 12(5), 599–611.

Lazarus, R., K. Kleinman, I. Dashevsky, C. Adams, P. Klundt, A. DeMaria Jr., R. Platt (2002). Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerging Infectious Diseases* 8(8), 753–760.

Lober, W., B. Karras, M. Wagner, J. Overhage, A. Davidson, H. Fraser, L. Trigg, K. Mandl, J. Espino, F.-C. Tsui (2002). Roundtable on bioterrorism detection: information system-based surveillance. *Journal of the American Medical Informatics Association* 9(2), 105–115.

Maier, C., P. Tucker, M. Garofalakis (2005). Filtering, punctuation, windows, and synopses, In: N. Chaudhry, N. Shaw, M. Abdelguerfi (eds.), *Stream Data Management*, Chapter 3. Springer, Inc., New York, NY, USA.

Morbidity and Mortality Weekly Report (MMWR). (September 24, 2004). Report from the Second Annual National Syndromic Surveillance Conference, Vol. 53.

Mostashari, F., M. Kulldorff, J. Hartman, J. Miller, V. Kulasekera (2003). Dead bird clusters as an early warning system for West Nile virus activity. *Emerging Infectious Diseases* 9(6), 641–646.

NEDSS Working Group (2001). National Electronic Disease Surveillance System (NEDSS). A standards-based approach to connect public health and clinical medicine. *Journal of Public Health Management and Practice* 7(6), 43–50.

Nordin, J., M. Goodman, M. Kulldorff, P. Ritzwoller, A. Abrams, K. Kleinman, M. Levitt, J. Donahue, R. Platt (2005). Simulated anthrax attacks and syndromic surveillance. *Emerging Infectious Disease* 11(9), 1394–1398.

Olson, K., M. Bonetti, M. Pagano, K. Mandl (2005). Real time spatial cluster detection using interpoint distances among precise patient locations. *BMC Medical Informatics and Decision Making* 5(19), 1–12.

Overhage, J.M., J. Suico, C. McDonald (2001). Electronic laboratory reporting: barriers, solutions and findings. *Journal of Public Health Management and Practice* 7(6), 60–66.

Relman, D., J. Olson (2001). Bioterrorism preparedness: what practitioners need to know. *Infectious Medicine* 18(11), 497–515.

Rotz, L., A. Khan, S. Lillibridge, S. Ostroff, J. Hughes (2002). Public health assessment of potential biological terrorism agents. *Emerging Infectious Diseases* 8(2), 225–230.

Shapira, Z. (1995). *Risk Taking: A Managerial Perspective*. Russell Sage Foundation, New York.

Siegrist, D., J. Pavlin (September 24, 2004). Bio-ALIRT biosurveillance detection algorithm evaluation. *Mortality and Morbidity Weekly* 53(Suppl), 152–158.

Sosin, D. (2003). Syndromic surveillance: the case for skillful investment. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 1(4), 247–253.

Tsui, F., J. Espino, V. Dato, P. Gesteland, J. Hutman, M. Wagner (2003). Technical description of RODS: a realtime public health surveillance system. *Journal of the American Medical Informatics Association (JAMIA)* 10(5), 399–408.

Viglas, S. (2005). Query execution and optimization, in: N. Chaudhry, N. Shaw, M. Abdelguerfi (eds.), *Stream Data Management*, Chapter 2. Springer, Inc., New York, NY, USA.

Wagner, M.M., F.-C. Tsui, J. Espino, V. Dato, D. Sittig, R. Caruana, L. McGinnis, D. Deerfield, M. Druzdzel, D. Fridsma (2001). The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice* 7(6), 51–59.

Wagner, M., J. Espino, F.-C. Tsui, P. Gresteland, W. Chapman, O. Ivanov, A. Moore, W. Wong, J. Dowling, J. Hutman (September 24, 2004). Syndrome and outbreak detection using chief-complaint data—removing a barrier to computer-based outbreak and disease surveillance—experience of the Real-Time Outbreak and Disease Surveillance project. *Morbidity and Mortality Weekly Report* 53(Suppl), 28–31.

Walks, I. (2003). Preparing your organization for a terrorist attack. *Disaster Recovery Journal* 16(3), 34–38.

Widom, J., S. Ceri (1996). *Active Database Systems: Triggers and Rules for Advanced Database Processing*. Morgan Kaufmann, Inc, San Francisco, CA, USA.

Wilschut, A., P. Apers (1991). *Pipelining in query execution. Conference on Databases, Parallel Architectures, and their Applications*. Miami Beach, Florida.

Zeng, D., H. Chen, M. Eidson, I. Gotham, C. Lynch (2005). Disease informatics and outbreak detection, in: H. Chen, S. Fuller, A. McCray (eds.), *Advances in Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, Chapter 13. Spring-Verlag.

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 7

Spatio-Temporal Data Analysis in Security Informatics

Daniel Zeng, Hsinchun Chen and Wei Chang

Department of Management Information Systems, University of Arizona, Tucson, AZ 85721, USA

Abstract

Spatio-temporal data analysis is an important component of security informatics since location and time are two critical aspects of most security-related events. The outputs of such analyses can provide useful information to guide the activities aimed at preventing, detecting, and responding to security problems. In this chapter, we survey recent developments in an important class of spatio-temporal data analysis task focusing on clustering or hotspot analysis. Both statistical and machine learning-based analysis techniques are discussed in offline (retrospective) and online (prospective) data analysis contexts. Computational studies carried out to evaluate the strengths and limitations of these techniques are reported. To illustrate how such techniques can be used in real-world applications, we also summarize two case studies using real datasets.

1 Introduction

Broadly defined, security informatics is the study of the development and evaluation of advanced information technologies and systems for national security-related applications (Chen et al., 2004). In most of these security-related applications, measurements of interest are made at various locations both in space and in time. For instance, incident case reports in crime databases have both occurring location and time information (Levine, 2002; Zeng et al., 2004). In the public health domain, the cases of disease reported to the Centers for Disease Control and Prevention through its National Notifiable Diseases Surveillance System are collected with timestamps at various places across the entire nation. Similar disease case reporting

practices exist at state and local jurisdictions, typically with cases identified with specific geolocations (Sonesson and Bock, 2003). As such, recent years have witnessed significant interest in spatio-temporal data analysis (Miller and Han, 2001).

In the literature, many different types of spatio-temporal data mining and knowledge discovery approaches have been proposed in the last decade (Gahegan, 2001). These approaches have been developed by researchers from a number of disciplines including biostatistics, environmental statistics, geographic information systems (GIS), data mining/machine learning, information visualization, among others. Roddick and Spiliopoulou (1999) provide an extensive bibliography of this body of literature.

In the particular context of security informatics, the following central questions of great practical importance have arisen in spatio-temporal data analysis and related predictive modeling:

(a) How to identify areas having exceptionally high or low measures (hotspots)?
(b) How to determine whether the unusual measures can be attributed to known random variations or are statistically significant? In the latter case, how to assess the explanatory factors? and
(c) How to identify any statistically significant changes (e.g., in rates of health syndromes or crime occurrences) in a timely manner in geographic areas?

Two types of approaches have been developed in the literature to address some of these questions. The first type of approach falls under the general umbrella of *retrospective* models (Kulldorff, 1997; Yao, 2003). It is aimed at testing statistically whether a disease is randomly distributed over space and time for a predefined geographical region during a predetermined time period. In many cases, however, this static perspective is inadequate as data often arrive dynamically and continuously, and in many applications there is a critical need for detecting and analyzing emerging spatial patterns on an ongoing basis. The second type of approach, *prospective* in nature, aims to meet this need with repeated time periodic analyses targeted at identification of statistically significant changes in an online context (Rogerson, 2001). Both approaches have been applied in security informatics practice. For instance, the New York City Department of Health and Mental Hygiene collected geocoded information concerning dead birds infected by West Nile Virus (WNV). Applying retrospective methods to dead bird data, they were able to detect possible WNV outbreaks before human case data became available. In another example, the Wisconsin Department of Natural Resources gathered location and time information of hunter-killed deer to study the spread of chronic wasting disease (CWD) over the eastern United States.

In this chapter, we present a tutorial and summarize recent developments in retrospective and prospective spatio-temporal data analysis. Section 2

first reviews major types of retrospective analysis and univariate surveillance approaches. Then it discusses two major types of prospective surveillance approaches. In Section 3, we introduce recently developed spatio-temporal data analysis methods based on a robust support vector machine (SVM)-based spatial clustering technique. The main technical motivation behind such methods is the lack of hotspot analysis techniques capable of detecting unusual geographical regions with *arbitrary* shapes. Section 4 reports on computational experiments based on simulated datasets. This experimental study includes a comparative component evaluating the SVM-based approaches against other methods in both retrospective and prospective scenarios. In Section 5, we summarize two case studies applying spatio-temporal data analysis methods to real-world datasets. Section 6 summarizes the chapter and concludes by discussing directions for future research.

2 Literature review

In this section, we first introduce retrospective spatio-temporal data analysis and related univariate surveillance methods. We then present the representative prospective surveillance methods, many of them developed as extensions to retrospective methods.

2.1 Retrospective spatio-temporal data analysis

Retrospective approaches determine whether observations or measures are randomly distributed over space and time for a given region. Clusters of data points or measures that are unlikely under the random distribution assumption are reported as anomalies. A key difference between retrospective analysis and standard clustering lies in the concept of ''baseline'' data. For standard clustering, data points are grouped together directly based on the distances between them. Retrospective analysis, on the other hand, is not concerned with such clusters. Rather, it aims to find out whether unusual clusters formed by the data points of interest exist *relative* to the baseline data points. These baseline data points represent how the normal data should be spatially distributed given the known factors or background information. Clusters identified in this relative sense provide clues about dynamic changes in spatial patterns and indicate the possible existence of unknown factors or emerging phenomena that may warrant further investigation. In practice, it is the data analyst's responsibility to separate the dataset into two groups: baseline data and data points of interest, typically with events corresponding to the baseline data precede those corresponding to the data points of interest. As such, retrospective analysis can be conceptualized as a spatial ''before and after'' comparison.

Below we discuss two major types of retrospective analysis methods: scan statistic-based and clustering-based. A comparative study of these two types of retrospective approaches can be found in Zeng et al. (2004).

2.1.1 Scan statistic-based retrospective analysis

Various types of scan statistics have been developed in the past four decades for surveillance and monitoring purposes in a wide range of application contexts. For spatio-temporal data analysis, a representative method is the spatial scan statistic approach developed by Kulldorff (1997). This method has become one of the most popular methods for detection of geographical disease clusters and is being widely used by public health departments and researchers. In this approach, the number of events, e.g., disease cases, may be assumed to be either Poisson or Bernoulli distributed. Algorithmically, the spatial scan statistic method imposes a circular window on the map under study and lets the center of the circle move over the area so that at different positions the window includes different sets of neighboring cases. Over the course of data analysis, the method creates a large number of distinct circular windows (other shapes such as rectangular and ellipse have also been used), each with a different set of neighboring areas within it and each a possible candidate for containing an unusual cluster of events. A likelihood ratio is defined on each circle to compute how likely the cases of interest fall into that circle not by pure chance. The circle with the maximum likelihood ratio is in turn reported as spatial anomalies or *hotspots*.

2.1.2 Clustering-based retrospective analysis

Despite the success of the spatial scan statistic and its variations in spatial anomaly detection, one of the major computational problems faced by this type of the methods is that the scanning windows are limited to simple, fixed symmetrical shapes for analytical and search efficiency reasons. As a result, when the real underlying clusters do not conform to such shapes, the identified regions are often not well localized. Another problem is that it is often difficult to customize and fine-tune the clustering results using scan statistic approaches. For different types of analysis, the users often have different needs as to the level of granularity and number of the resulting clusters and they have different degrees of tolerance regarding outliners.

These problems have motivated the use of alternative and complementary modeling approaches based on clustering. *Risk-adjusted nearest neighbor hierarchical clustering* (RNNH) is a representative of such approaches.

Developed for crime hotspot analysis, RNNH is based on the well-known nearest neighbor hierarchical clustering (NNH) method, combining the hierarchical clustering capabilities with kernel density interpolation techniques. The standard NNH approach identifies clusters of data points that are close together (based on a threshold distance). Many such clusters, however, are due to some background or baseline factors (e.g., the

population which is not evenly distributed over the entire area of interest). RNNH is primarily motivated to identify clusters of datapoints *relative* to the baseline factor. Algorithmically, it dynamically adjusts the threshold distance inversely proportional to some density measure of the baseline factor (e.g., the threshold should be shorter in regions where the population is high). Such density measures are computed using kernel density based on the distances between the location under study and other data points. We summarize below the key steps of the RNNH approach.

- Define a grid over the area of interest; calculate the kernel density of baseline points for each grid cell; rescale such density measures using the total number of cases.
- Calculate the threshold distances between data points for hierarchical clustering purposes and perform the standard nearest neighbor hierarchical clustering based on the above distance threshold.

RNNH has been shown to be a successful crime analysis tool. We argue that its built-in flexibility of incorporating any given baseline information and computational efficiency also make it a good candidate for analyzing spatial-temporal data in other security informatics applications such as infectious disease and bio-terrorism data analysis.

In Section 3, we will introduce another clustering-based method, called *risk-adjusted support vector clustering* (RSVC) (Zeng et al., in press), the result of our recent attempt to combine the risk adjustment idea of RNNH with a modern, robust clustering mechanism such as SVM to improve the quality of hotspot analysis.

2.2 *Univariate surveillance*

Univariate surveillance methods monitor one-dimensional data streams (typically time series without spatial information) and focus on quick and accurate detection and response in cases where unusual events take place. These methods vary in the alarm functions used and the procedures followed to observe the time series. A comprehensive review of univariate surveillance approaches in the context of public health surveillance can be found in Sonesson and Bock (2003). Three types of surveillance strategies are commonly used: (a) the cumulative sum (CUSUM) method monitoring the number of events in a fixed interval, (b) Chen's set method monitoring the time intervals between consecutive events (Chen, 1978), and (c) Frisen's approach monitoring the likelihood of an observed occurrence (Frisen and Mare, 1991). Among them, the CUSUM approach is the easiest to implement in practice since surveillance analysis is typically performed regularly and tracking the number of events between two consecutive surveillance runs can be easily done. For instance, the CUSUM approach has been extensively used in industry to monitor the number of defective products in a manufacturing process for the purpose of quality control.

CUSUM operates by accumulating the deviations between the observations and expectations. Formally, assume that X is the variable that we are keeping track of and X_t its value at time t. Denote by Z_t the normalized deviation of X_t, $Z_t = (X_t - \mu)/\sigma$, where u is the mean and σ the variance. The accumulated deviation at time t, denoted by S_t, can then be given as

$$S_t = \max(S_{t-1} + z_t - k, 0), \quad S_0 = 0 \tag{1}$$

where k is the normal varying range of X. When the accumulated deviation S_t is over some predefined threshold value, an alarm will be generated indicating an increase on the mean of the underlying variable of interest X. From Eq. (1), we note that S_t will be reset to 0 when the time series comes back to the normal status (i.e., the accumulated deviation is less than normal varying range).

Several evaluation metrics have been developed to measure the performance of univariate surveillance methods (Sonesson and Bock, 2003). Among them, ARL^0 and ARL^1 are two conjugated and widely accepted measures. ARL^0 is the average run length until the first alarm is triggered under the null hypothesis and ARL^1 is the average run length until the first alarm is triggered under the alternative hypothesis. In other words, ARL^0 estimates how fast a surveillance algorithm might trigger a false alarm and ARL^1 estimates how fast an algorithm can detect the anomaly if it does occur. ARL^0 and ARL^1 are conceptually similar to the type 1 and type 2 errors from statistical hypothesis testing. In addition to statistical power evaluation, a common evaluation method used in the surveillance literature is to fix ARL^0 and compare ARL^1 for different approaches. This evaluation amounts to measuring how fast a surveillance approach can detect a true abnormal event given a fixed level of false alarm rate. Another less frequently used measure is expected delay, which calculates the expected time between the time an anomaly occurs and the time an alarm is triggered. This measure is suitable in cases where the distribution of the anomaly occurring time is known and the time needed to trigger an alert after an anomaly occurs depends on the occurring time of the anomaly.

2.3 *Prospective spatio-temporal surveillance*

In security informatics, the threats of terrorist attacks, catastrophic natural disasters, and major infectious disease outbreaks, have recently generated great interests in developing and deploying prospective spatio-temporal surveillance systems for timely event detection and preemptive reactions. A major advantage that prospective approaches have over retrospective approaches is that they do not require the separation between the baseline cases and cases of interest in the input data. Such a requirement is necessary in retrospective analysis and is a major source of confusion and

difficulty to the end users. Prospective methods bypass this problem and process data points continuously in an online context.

Two types of prospective spatio-temporal data analysis approaches have been developed in the statistics literature. The first type segments the surveillance data into chunks by arrival time and then applies a spatial surveillance technique to identify abnormal changes. In essence, this type of approach reduces a spatio-temporal surveillance problem into a series of spatial surveillance problems. The second type explicitly considers the temporal dimension and clusters data points directly based on both spatial and temporal coordinates. We briefly summarize representative approaches for both types of methods including Rogerson's method and the space–time scan statistic.

2.3.1 *Rogerson's methods*

Rogerson has developed CUSUM-based surveillance methods to monitor spatial statistics such as Tango and Knox statistics, which capture spatial distribution patterns existing in the surveillance data (Rogerson, 1997, 2001). Let C_t be the spatial statistic (e.g., Tango or Knox) at time t. The surveillance variable is defined as $Z_t = C_t - E(C_t|C_{t-1})/\sigma(C_t|C_{t-1})$. Refer to Rogerson (1997, 2001) for the derivation of the conditional expected value $E(C_t|C_{t-1})$ and the corresponding variance $\sigma(C_t|C_{t-1})$. Following the CUSUM surveillance approach as shown in Eq. (1), when the accumulated deviation Z_t exceeds a threshold value, the system will trigger an alarm. Rogerson's methods have successfully detected the onset of the Burkitt's lymphoma in Uganda during 1961–1975. The variations and other applications of Rogerson's approaches can be found in Rogerson and Sun (2001) and Rogerson and Yamada (2003, 2004).

2.3.2 *Space–time scan statistic*

Kulldorff has extended his retrospective two-dimensional spatial scan statistic to a three-dimensional space–time scan statistic, which can be used as a prospective analysis method (Kulldorff, 2001). The basic intuition is as follows. Instead of using a moving circle to search the area of interest, one can use a cylindrical window in three dimensions. The base of the cylinder represents space, exactly as with the spatial scan statistic, whereas the height of the cylinder represents time. For each possible circle location and size, the algorithm considers every possible starting and ending times. The likelihood ratio test statistic for each cylinder is constructed in the same way as for the spatial scan statistic. After a computationally intensive search process, the algorithm can tell the user where the abnormal cluster is with the corresponding geolocations and time period. The space–time scan statistic has successfully detected an increased rate of male thyroid cancer in Los Alamos, New Mexico during 1989–1992 (Kulldorff, 2001).

3 Support vector clustering-based spatio-temporal data analysis

In this section, we present two recently developed robust spatio-temporal data analysis methods. The first is a retrospective hotspot analysis method called *RSVC* (Zeng et al., 2004). The second is a prospective analysis method called *prospective support vector clustering* (*PSVC*), which uses RSVC as a clustering engine (Chang et al., 2005).

3.1 Risk-adjusted support vector clustering (RSVC)

RSVC is the result of our recent attempt to combine the risk adjustment idea of RNNH with a modern, robust clustering mechanism such as SVM to improve the quality of hotspot analysis. SVMs are the most well known of a class of algorithms that use the idea of kernel substitution. Motivated by statistical learning theory, it is a systematic approach with well-defined optimization formulations, which have no local minima to complicate the learning process and can be solved using well-established computational methods. It also has a clean geometric interpretation. As a linear discriminant to separate data points with binary labels in a d-dimensional input space, an SVM-based approach finds the solution by either maximizing the margin between parallel supporting planes separating the data points of different labels or, equivalently, by bisecting closest points in the convex hulls encompassing data points of the same label. Both objective functions lead to a quadratic program that can be solved rather efficiently. Using Hilbert-Schmidt kernels, the above linear classification algorithm can be extended to handle nonlinear cases. Conceptually, a nonlinear classification problem can be converted into a linear one by adding additional attributes to the data that are nonlinear functions of the original data. This expanded attribute space is called the feature space. Computationally, however, through the use of kernels, this nonlinear mapping method can be implemented without even knowing how the mapping from the original input space to the expanded feature space is done. As a result, the same efficient and robust optimization-based training method for the linear case can be readily applied to produce a general nonlinear algorithm.

Although the above nontechnical description uses classification to illustrate the basic ideas of SVMs, they can be applied in a wide range of other types of machine learning and data mining problems. Among them, SVM-based data description and novelty detection (DDND) is particularly relevant to spatio-temporal data analysis (Ben-Hur et al., 2001). SVM-based DDND methods are aimed at identifying the *support* of a data distribution. In a simple application, for instance, these methods can estimate a binary-valued function that is 1 in those regions of input space where the data predominantly lies and clusters together and 0 elsewhere. Informally, these methods proceed as follows: first, they map implicitly the input data to a high-dimensional feature space defined by a kernel function (typically the

Gaussian kernel). Second, these methods find a hypersphere in the feature space with a minimal radius to contain most of the data. The problem of finding this hypersphere can be formulated as a quadratic or linear program depending on the distance function used. Third, the function estimating the support of the underlying data distribution is then constructed using the kernel function and the parameters learned in the second step.

SVM-based methods have several attractive modeling and computational properties, making them suitable for security informatics applications. SVM-based DDND methods can single out data clusters in complex shapes and have been well tested in complex, noisy domains (e.g., handwritten symbol recognition). Through control parameters such as the range parameter in the Gaussian kernel and the slack variables used to include the outliners, the user can easily control the behavior of the algorithm to satisfy different modeling needs. Also, the introduction of slack variables allows overlapping clusters to be generated. SVM-based density estimation methods are powerful approaches producing actual density estimations, which contain more information than what other methods can offer. As in the case for SVM-based DDND, using a limited number of parameters, the user can easily experiment with density estimations of varying properties to meet their modeling needs.

The standard version of SVM-based DDND does not take into consideration baseline data points and therefore cannot be directly used in spatio-temporal data analysis. As such, we have developed a *risk-adjusted* variation, called RSVC, based on ideas similar to those in RNNH. First, using only the baseline points, a density map is constructed using standard approaches such as the kernel density estimation method. Second, the case data points are mapped implicitly to a high-dimensional feature space defined by a kernel function (typically the Gaussian kernel). The width parameter in the Gaussian kernel function is dynamically adjusted based on the kernel density estimates obtained in the previous step. The basic intuition is as follows: when the baseline density is high, a larger width value is used to make it harder for points to be clustered together. Third, following the SVM approach, RSVC finds a hypersphere in the feature space with a minimal radius to contain most of the data. The problem of finding this hypersphere can be formulated as a quadratic or linear program depending on the distance function used. Fourth, the function estimating the support of the underlying data distribution is then constructed using the kernel function and the parameters learned in the third step. When projected back to the original data space, the identified hypersphere is mapped to (possibly multiple) clusters. These clusters are then returned as the output of RSVC.

3.2 Prospective support vector clustering

Although well grounded in theoretical development, both Rogerson's methods and the space–time scan statistic have major computational

problems. Rogerson's approaches can monitor a given target area but they cannot search for problematic areas or identify the geographic shape of these areas. The space–time scan statistic method performs poorly when the true abnormal areas do not conform to simple shapes such as circles. Furthermore, the three-dimensional search needed by the scan statistic method is very time consuming. These computational considerations provide direct motivations for our technical research on prospective spatio-temporal data analysis. Below we introduce the basic ideas behind our approach, which is called PSVC, and summarize its main algorithmic steps.

Our PSVC approach follows the design of the first type of the spatio-temporal surveillance method discussed in Section 2.3, which involves repeated spatial clusterings over time. More specifically, the time horizon is first discretized based on the specific characteristics of the data stream under study. Whenever a new batch of data arrives, PSVC treats the data collected during the previous time frame as the baseline data and runs the retrospective RSVC method.

After obtaining a potential abnormal area, PSVC tries to determine how statistically significant the identified spatial anomaly is. Many indexes have been developed to assess the significance of the results of clustering algorithms in general. Halkidi et al. (2002a, b) summarize and categorize these indexes into three categories: (a) *external criteria* which evaluate the results of a clustering algorithm based on a subjective, prespecified partition structure, (b) *internal criteria* which describe how data tend to group together using the dataset itself alone, and (c) *relative criteria* which compare different clustering results from the same algorithm but with different parameter values. However, all these criteria assess clustering in an absolute sense without considering baseline information. Thus, they are not readily suitable for prospective spatio-temporal data analysis.

Kulldorff's (1997) likelihood ratio $L(Z)$ as defined in Eq. (2) is to our best knowledge the only statistic that explicitly takes baseline information into account.

$$L(Z) = \left(\frac{c}{n}\right)^{c}\left(1-\frac{c}{n}\right)^{n-c}\left(\frac{C-c}{N-n}\right)^{C-c}\left(1-\frac{C-c}{N-n}\right)^{(N-n)-(C-c)} \tag{2}$$

In this definition, C and c are the number of the cases in the entire dataset and the number of the cases within the scanned area Z, respectively. N and n are the total number of the cases and the baseline points in the entire dataset and the total number of the cases and the baseline points within Z, respectively. Since the distribution of the statistic $L(Z)$ is unknown, the Monte Carlo simulation approach is an alternative to calculate statistical significance measured by the p-value. Using this approach, we first generate T replications of the dataset, assuming the data are randomly distributed. We then calculate the likelihood ratio $L(Z)$ on the same area Z for each

replication. Finally, we rank these likelihood ratios and if L takes the Xth position, then the p-value is set to be $X/(T+1)$.

Note that in a straightforward implementation of the above algorithmic design, alerts are triggered only when adjacent data batches have significant changes in terms of data spatial distribution. This localized myopic view, however, may lead to significant delay in alarm triggering or even false negatives because in some circumstances, unusual changes may manifest gradually. In such cases, there might not be any significant changes between adjacent data batches. However, the accumulated changes over several consecutive batches can be significant and should trigger an alarm. This observation suggests that a more "global" perspective beyond comparing adjacent data batches is needed.

It turns out that the CUSUM approach provides a suitable conceptual framework to help design a computational approach with such a global perspective. The analogy is as follows. In the CUSUM approach, accumulative deviations from the expected value are explicitly kept track of. In prospective analysis, it is difficult to design a single one-dimensional statistic to capture what the normal spatial distribution should look like and to measure the extent to which deviations occur. However, conceptually the output of a retrospective surveillance method such as RSVC can be viewed as the differences or discrepancies between two data batches, with the baseline data representing the expected data distribution. In addition, accumulative discrepancies can be computed by running RSVC with properly set baseline and case data separation. For an efficient implementation, we use a stack as a control data structure to keep track of RSVC runs which now include comparisons beyond data from adjacent single periods. The detailed control strategy is described below.

When clusters generated in two consecutive RSVC runs have overlaps, we deem that the areas covered by these clusters are risky areas. We use the stack to store the clusters along with the data batches from which these risky clusters are identified. Then we run RSVC to compare the current data batch with each element (in the form of a data batch) of the stack sequentially from the top to the bottom to examine whether significant spatial pattern changes have occurred. The objective of the stack is similar to that of variable S in Eq. (1) as part of the CUSUM approach, which accumulates the deviation Z between the observed value and expected value of the monitored variable. Stacks whose top data batch is not the current data batch under examination can be emptied since the risky areas represented by them are no longer "alive." This operation resembles one of the steps in the CUSUM calculation where the accumulated deviation is reset to 0 when the monitored variable is no longer within the risky range.

We now explain the main steps of the PSVC algorithm as shown in Fig. 1.

Each cluster stack represents a candidate abnormal area and the array *clusterstacks* holds a number of cluster stacks keeping track of all candidate areas at stake. Initially (line 1) *clusterstacks* is empty. The steps from lines 3

```
1 clusterstacks=[]
2 Whenever a new data batch arrives {
3    rsvcresult=RSVC (previousdate, currentdate)
4    For each cluster C recorded in rsvcresult {
          /*C records the identified cluster, its p-value, and the date of the associated data
          batch. */
5         If (C.p-value<threshold) {Trigger alert}
6         Else {
7              If (clusterstacks is not empty) {
8                   For each cluster stack S in clusterstacks {
9                        lastcluster=the top element of the stack S
10                       If cluster C has overlaps with lastcluster {
11                            S.append(C)
12                            For each element in the stack S from top to bottom {
13                                 tempresult=RSVC (element.date, currentdate)
14                                 For each temporal cluster TC recorded in tempresult {
15                                      if (TC.p-value<threshold) {Trigger alert}
16                                 }
17                            }
18                       }
19                  }
20                  If (current cluster C does not have any overlaps
21                       with any of the top element of the clusters in clusterstacks) {
22                        new stack NS=[C]
23                        clusterstacks.append(NS)
24                  }
25                  For each cluster stack S {
26                            If (S[top element].date!=currentdate) {delete stack S}
27                  }
28             }
29             Else {
30                  new stack NS=[C]
31                  clusterstacks.append(NS)
32             }
33        }
34   }
35}
```

Fig. 1. The main steps of the PSVC algorithm.

to 35 are run whenever a new data batch enters the system. First, the RSVC retrospective method is executed (line 3) to compare the spatial distribution of the new data batch with that of the previous data patch. The resulting abnormal clusters are saved in *rsvcresult*. Any statistically significant cluster in *rsvcresult* will immediately trigger the alert (line 5).

For those emerging candidate areas that are not yet statistically significant, they are kept in *clusterstacks*. Lines 7–32 of the PSVC algorithm describe the operations to be performed on each of these candidate clusters C. If no cluster stack exists, we simply create a new cluster stack, which contains only C as its member (line 30), and update the array *clusterstacks* accordingly (line 31). If cluster stacks already exist, for each of these cluster stack S, we determine whether the current cluster C has any overlaps with the most recent cluster (the top element) in S (line 10). If the current cluster C does overlap with an existing candidate area, further investigation beyond comparison between adjacent data batches will be warranted.

The operations described from lines 11 to 15 implement these further investigative steps. First, cluster C is added onto stack S (line 11). Then the

current data batch is compared against all remaining data batches in S in turn from the top to the bottom. Should any significant spatial distribution change be detected, an alert will be triggered (lines 13–15).

If cluster C does not overlap with any of the most recent cluster in all of the existing cluster stacks, a new cluster stack is created with C as its only element and the array *clusterstacks* is updated accordingly (lines 22 and 23). After processing the candidate cluster C, we remove all inactive cluster stacks whose top clusters are not generated at the present time (equal to the creation time of C) (line 26).

4 Experimental studies

This section reports experimental studies designed to evaluate RSVC and PSVC quantitatively and compare their performance with that of existing retrospective and prospective analysis methods.

4.1 RSVC evaluation

We have conducted a series of computational studies to evaluate the effectiveness of the three hotspot analysis techniques (SaTScan, RNNH, RSVC) (Zeng et al., in press). In the first set of experiments, we used artificially generated datasets with known underlying probability distributions to precisely and quantitatively evaluate the efficacy of these techniques. Since the true hotspots are known in these experiments based on simulated data, we use the following well-known measures from information retrieval to evaluate the performance of hotspot techniques: Precision, Recall, and F-Measure. In the spatial data analysis context, we define these measures as follows.

Let A denote the size of the hotspot(s) identified by a given algorithm, B the size of the true hotspot(s), and C the size of the overlapped area between the algorithm-identified hotspot(s) and true hotspot(s). Precision is defined as C/A. Recall is defined as C/B. F-measure is defined as the harmonic mean of precision and recall ($2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall})$). Observe that high recall indicates low false negatives and that high precision indicates low false positives. Notice that achieving high level of one measure often impacts negatively achieving high level of the other measure. F-measure represents a balance and trade-off between precision and recall.

Below we report one artificially generated scenario we have experimented with. In this scenario, as shown in Fig. 2, the true spot is a square with its circular-shaped center removed. We first randomly generated 100 baseline points in the circle located in the center. We then generated 200 case points of interest in total over the square. To make the problem more interesting, we introduced some noise—30 outlier baseline points and 40 outlier case points over the entire map. For statistical testing purposes, we repeated the

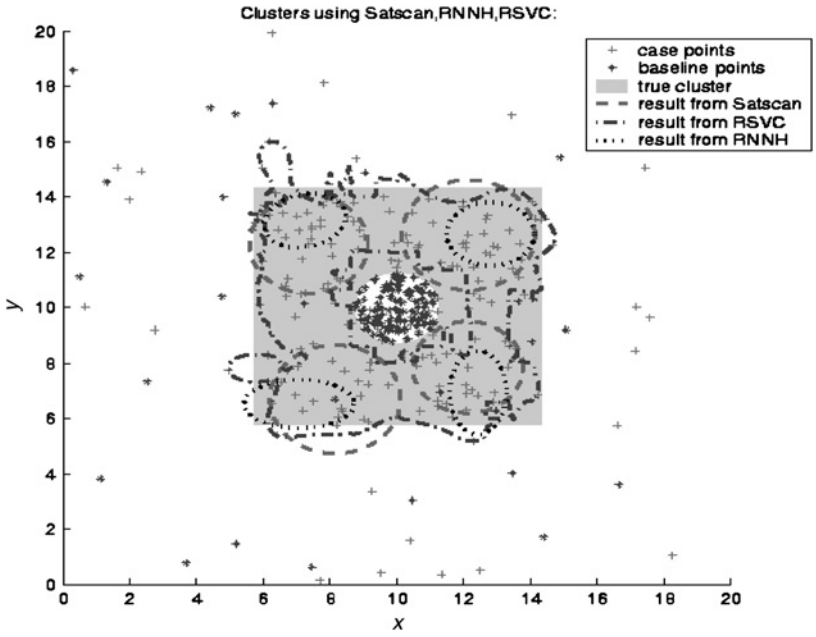


Fig. 2. Scenario 1 based on simulated data.

Table 1
Average performance of RSVC, SCAN, and RNNH on 30 instances of scenario 1

	Precision (%)	Recall (%)	*F*-measure (%)
RNNH	95	50	64
RSVC	80	92	85
SCAN	54	92	65

above data generation process for 30 times to produce 30 instances of the template scenario by moving the centers of the circle randomly across the map.

To collect the performance data, we ran all hotspot analysis methods under study on all the problem instances. Table 1 summarizes these methods' average performance across all instances.

We have also conducted additional experiments using different template scenarios. Interested readers are referred to Zeng et al. (in press). Some general observations are in order. RSVC and the spatial scan method have similar level of recall across different scenarios. However, RSVC has higher precision than the spatial scan method (confirmed by statistical tests). RNNH has the highest precision level but typically with lowest recall. When considering the combined measure, i.e., the *F*-measure, RSVC consistently

delivers the best results, suggesting it as a strong (if not the best) candidate for real-world application in security informatics for various types of hotspot identification.

4.2 PSVC evaluation

To evaluate a prospective spatio-temporal data analysis method, we need to consider both spatial and temporal evaluation measures. From a spatial perspective, the goal is to evaluate how accurate the detected clusters are geographically, relative to the location of the true clusters. When true hotspots are known, precision, recall, and *F*-measure provide appropriate performance metrics, as in the case of retrospective analysis.

As for the temporal evaluation measures, we introduced ARL^0 and ARL^1 as two widely used ones in univariate surveillance. ARL^1 reveals how timely an algorithm can detect an anomaly and ARL^0 how easily an algorithm tends to trigger a false alarm. In our study, we adopt the ARL^1 measure and rename it to "Alarm Delay," which is defined as the delay between the time an anomaly occurs and the time the algorithm triggers the corresponding alert. Using ARL^0 can be difficult in practice as it would require the system run for a long time under the normal condition to collect false alarm data. As an alternative, we have followed the following performance data collection procedure: we apply the prospective analysis method under study to a simulated data stream for a relatively long period of time. This data stream contains some anomalies generated according to known patterns. When a suspicious area reported by the method does not overlap with the true abnormal area (e.g., both precision and recall are 0) or the report date is earlier than the actual date of the abnormal occurrence, we consider it as a false alarm. In some cases, the system fails to trigger any alarms during the entire monitoring period. We count how many times an algorithm triggers false alarms and how many times it fails to detect the true anomalies as surrogate measures for ARL^0.

We have chosen the space–time scan statistic as the benchmark method since it has been widely tested and deployed, especially in public health applications, and its implementation is freely available through SaTScan. Similar to evaluation of RSVC, we have used simulated datasets with the generation of the true clusters fully under our control. Below we report one scenario used in our computational experiments. For ease of exposition, throughout this section, we use the public health application context to illustrate this scenario. This scenario corresponds to an "emerging" scenario where disease outbreaks start from some location where very few disease incidents occurred before. For this scenario, we created 30 problem instances by randomly changing the size, location, starting date, and the speed of expansion of the simulated abnormal cluster.

More specifically, for all "emerging" scenario problem instances, both x and y axes have the support of [0, 20]. The range for time is from 0 to 50 days.

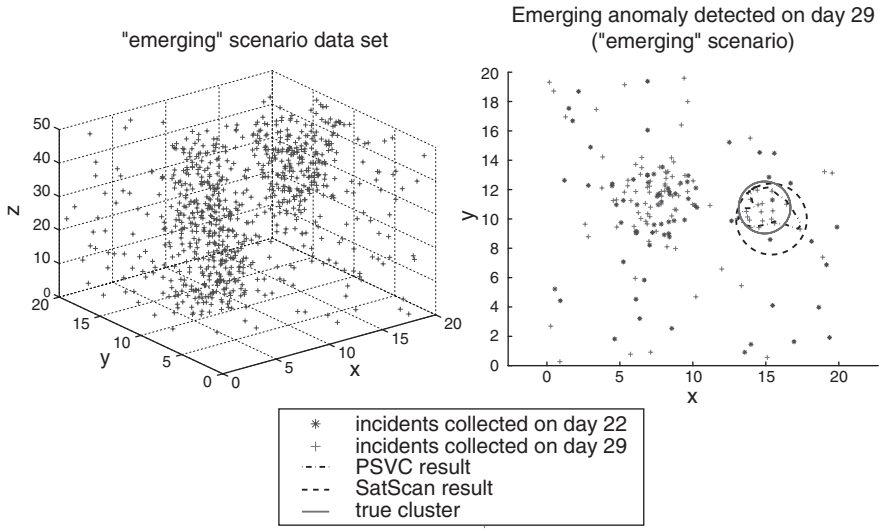


Fig. 3. A problem instance of the "emerging" scenario.

We first generated 300 data points in this three-dimensional space ([0, 20] $\times$ [0, 20] $\times$ [0, 50]) as the background. We then generated another 300 data points inside a cylinder whose bottom circle resides at center (x_l, y_l) with radius r_l. The height of this cylinder is set to 50, covering the entire time range. This cylinder is designed to test whether a prospective spatio-temporal data analysis method might identify the pure spatial cluster by mistake.

Consider the dense cone-shaped area in the left sub-figure of Fig. 3. An abnormal circular cluster that is centered at (x_r, y_r) emerges on some date *startT*. This circle starts with radius *startR* and continuously expands until the radius reaches *endR* on the last day, day 50. In contrast to the cylinder to the left, which has roughly the same number of data points every day, the cone-shaped area represents an emerging phenomenon. To approximate exponential expansion, we let the number of points inside the cone-shaped area at any given day follow the following expression:

$$a * (current\_date - start\_date + 1)\hat{}increaserate$$

where a is the number of points inside the area on the anomaly starting date and *increaserate* indicates how fast an outbreak expands. Fig. 4 shows three snapshots of an emerging scenario problem instance projected to the spatial map at three different times. The red crosses represent the new data batch for the current time frame during which the analysis is being conducted. The blue stars represent the data points from the last time frame. As shown in these snapshots, until day 22 there is no notable spatial pattern change

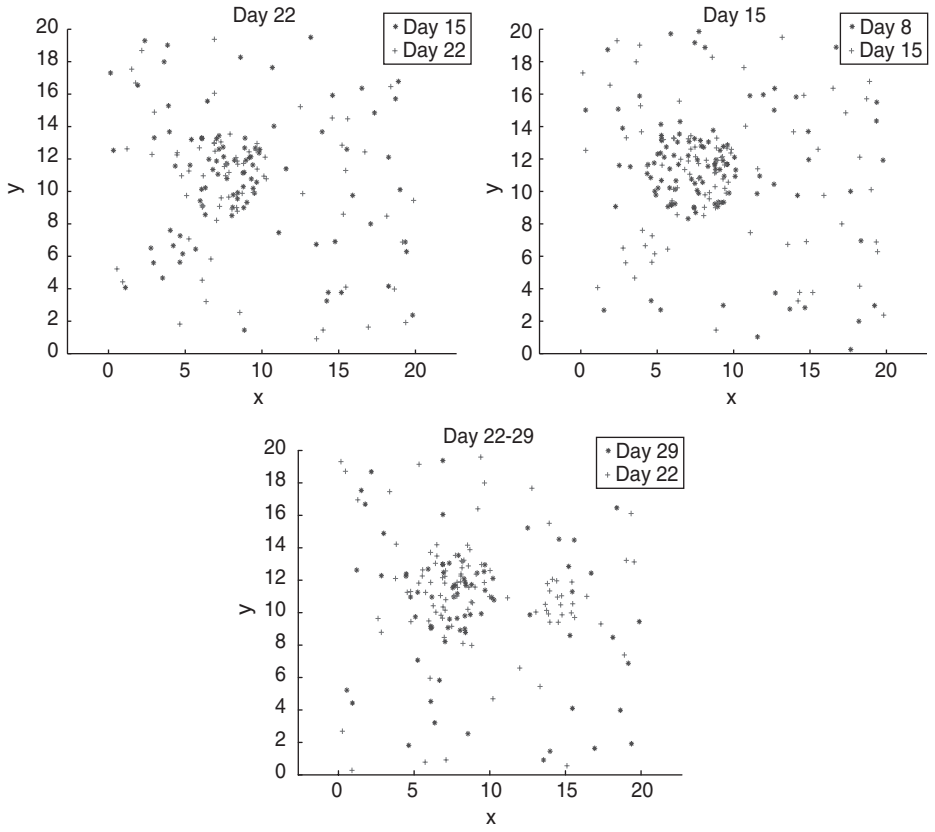


Fig. 4. Snapshots of an "emerging" scenario problem instance.

during two consecutive weeks. But during the week from day 22 to day 29, we can clearly observe an emerging circle.

When generating data points for 30 replications of the emerging scenario, we aimed to experiment with the cone-shaped area and the cylinder of varying sizes and locations under two constraints: (a) neither area is completely inside the other area, and (b) both areas are confined within the boundary of the three-dimensional space. Prospective analysis was conducted on a weekly basis with each batch containing around 80–100 data points.

The right sub-figure of Fig. 3 illustrates the results of the analyses using SaTScan and PSVC on the problem instance shown in the left sub-figure. As expected, both methods reported an emerging abnormal area. Neither reported the pure spatial cluster (cylinder), which is positive. The average performance of PSVC and SaTScan over the 30 problem instances is summarized in Table 2. We observe that for the emerging scenario, SaTScan achieves a higher level of recall and PSVC a higher level of precision. These two methods do not differ significantly with respect to the overall spatial

Table 2
Average performance of SaTScan and PSVC over 30 "emerging" scenario instances

	Precision (%)	Recall (%)	*F*-measure (%)	Alarm delay (days)	False alarm (times)	Fail to detect (times)	Computing time (sec)
SaTScan	66.2	83.6	69.5	5.4	5	2	607
PSVC	88.5	55.2	64.8	6.0	0	2	95

performance given by the *F*-measure. In general, PSVC detected anomaly as soon as SaTScan did but with less false alarms.

We now report general experimental findings based on this emerging and other scenarios. Both SaTScan and PSVC can effectively identify the abnormal areas demonstrating changes in the spatial distribution pattern over time and correctly ignore pure spatial clusters. When the abnormal area follows a simple regular shape (e.g., a circle in the emerging scenario), PSVC achieves better precision while SaTScan achieves better recall. PSVC significantly outperforms SaTScan in terms of spatial evaluation measures when detecting abnormal areas with complex, irregular shapes as in the case of the expanding and moving scenarios. PSVC detects abnormal areas as soon as SaTScan does but with less false alarms. This is particularly true when abnormal areas do not conform to simple regular shapes.

5 Case studies: public health surveillance and crime analysis

We now present two case studies using spatio-temporal data analysis approaches to analyze real-world security informatics datasets (Hu et al., 2006). We first discuss the application of retrospective methods and then that of prospective methods. Note that for real-world datasets, true hotspots are typically unknown. As such, we focus on demonstrating qualitatively the performance of these analysis techniques.

In the first case study, we used a test dataset containing dead bird sightings in the spring and summer seasons of 2002 in New York State. Each sighting is identified with its reporting time and the corresponding geo-coded location. There are 364 sightings in total. The baseline data were defined as all the sightings reported before the day when the first dead bird was diagnosed with WNV. All the sightings after that day were treated as the case data of interest. In our experiment, the baseline contains 140 sightings and the data of interest contains 224 sightings. See Fig. 5 for the results.

We make the following observations. These hotspot analysis methods can provide complementary information with each other. Sometimes using three methods at the same time can provide deeper insight for the user. Overlapped hotspot area detected by different methods means very high risk. A hotspot area detected by only one method may indicate that this

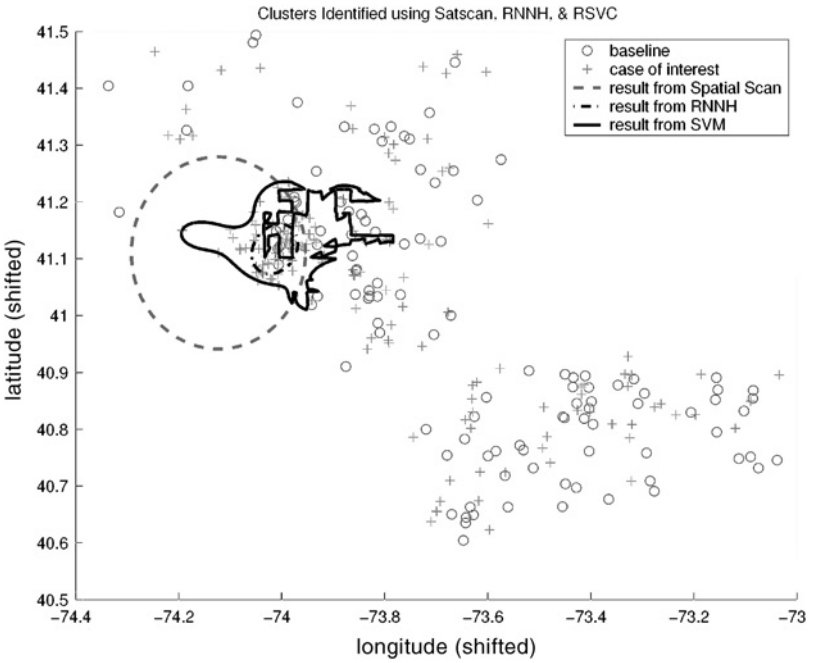


Fig. 5. Dead bird hotspots identified.

area is not significantly risky but still warrants further investigation. As far as method comparison is concerned, it seems consistent that RNNH has high precision but its recall level is typically very low. The spatial scan method has good recall but tends to identify large circles, resulting in low precision. RSVC has good balanced performance overall and is able to detect hotspots of arbitrary shape. It is also computationally robust.

We now turn our attention to the application of prospective methods. Our case study is in crime analysis using a dataset consisting of 4705 residence larceny incidents in a middle-sized city in U.S. from January 1, 2003 to March 31, 2005. While processing these 2 years and 3 months worth of data, PSVC and SaTScan triggered five and eight alarms, respectively. We report the most significant cluster identified by each method. Figure 6 shows the northwest part of the city where a lot of criminal activities take place. The left sub-figure shows the hotspot area identified by PSVC on January 3, 2004 and the right sub-figure by SaTScan on June 4, 2004. The red crosses represent the most recent larceny incidents while the blue stars represent the larceny incidents that occurred 2 weeks ago. Both methods identified a high-risk area with emerging criminal activities worth further investigation. To demonstrate the potential usefulness of such prospective analysis methods, we quote below an experienced police officer who commented on our case study. ''If this continuously monitoring system can

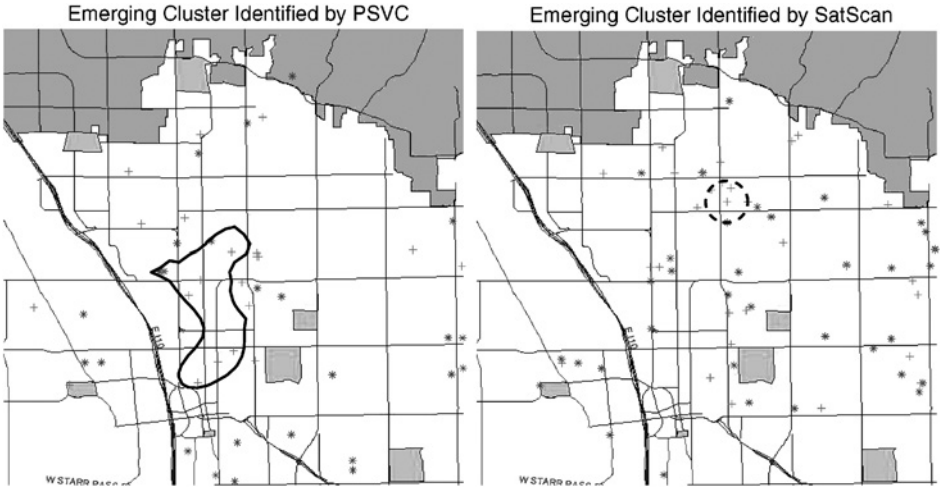


Fig. 6. Emerging residence larceny incident clusters identified by PSVC and SaTScan.

be hooked up with some easy-to-use interface, it would be very helpful for us to dispatch patrol officers and arrange the patrol routes. It will save me a lot of time to study the distribution of criminal activities. Just hope it won't trigger too many false alarms.''

6 Conclusions and future work

Spatio-temporal hotspot analysis is an important component of security informatics since location and time are two critical aspects of most security-related events. Such analyses can help guide the activities aimed at detecting and responding to security problems.

This chapter introduces both retrospective and prospective hotspot analysis methods. Compared with retrospective methods, prospective analysis methods provide a more powerful data analysis framework. Prospective methods are aimed at identifying spatio-temporal interaction patterns in an online context and do not require preprocessing data points into baseline and cases of interest. We survey major types of retrospective and prospective analysis techniques and present two case studies in public health surveillance and crime analysis to demonstrate the potential value of these techniques in real-world security informatics applications.

We conclude this chapter by discussing future research. One of the possible improvements to the PSVC algorithm is to speed up the computation through better control of the complexity of the shape of the resulting clusters. Another major area of extension is concerned with how to deal with multiple incidents occurring at exactly the same locations. In public health and many other applications including crime analysis, the events are

being recorded with spatial coordinates corresponding to the location of service centers such as hospitals as opposed to the precise location of the incidents. In addition, for privacy reasons, sometimes the records only have aggregated spatial coordinates such as the ZIP codes or the county-level identifiers associated with them. How to process such aggregated spatial information presents a technical challenge and a research opportunity of practical relevance in both retrospective and prospective contexts.

7 Questions for Discussion

1. What are the key differences between retrospective and prospective spatio-temporal data analysis frameworks? Discuss their applicability in different security informatics domain contexts.
2. Discuss the input data requirements for retrospective and prospective spatio-temporal data analysis approaches, respectively.
3. How can one interpret the findings of spatio-temporal data analysis techniques? What are the implications of false positives (e.g., hypothesized disease outbreaks that turn out to be normal background disease occurrences)? What are the implications of false negatives (e.g., missing significant crime trends)?
4. Discuss the evaluation metrics applicable to retrospective spatio-temporal data analysis approaches. Discuss the evaluation metrics applicable to prospective spatio-temporal data analysis approaches.
5. Discuss how spatio-temporal data analysis frameworks can be potentially integrated with other types of data analysis techniques (regression, correlation analysis, etc.).
6. What are the potential technology adoption issues that may hinder the wide acceptance of spatio-temporal data analysis techniques?

Acknowledgment

Research reported in this paper was supported in part by the U.S. National Science Foundation through grant no. IIS-0428241. The first author is an affiliated professor at the Institute of Automation, the Chinese Academy of Sciences, and wishes to acknowledge support from an international collaboration grant (2F05NO1) from the Chinese Academy of Sciences, and a 973 program grant (2006CB705500) from the National Natural Science Foundation of China. We wish to thank Dr. Millicent Eidson, Dr. Ivan Gotham, Ms. Jenny Schroeder, and Mr. Tim Peterson for providing the datasets used in this study and related discussions. We also thank other members of the NSF-funded BioPortal and Coplink projects for informative and constructive discussions.

References

Ben-Hur, A., D. Horn, H.T. Siegelmann, V. Vapnik (2001). Support vector clustering. *Journal of Machine Learning Research* 2, 125–137.

Chang, W., D. Zeng, H. Chen (2005). A novel spatio-temporal data analysis approach based on prospective support vector clustering, in: *Proceedings of the Workshop on Information Technologies and Systems (WITS)*, Las Vegas, NV.

Chen, R. (1978). A surveillance system for congenital malformations. *Journal of the American Statistical Association* 73, 323–327.

Chen, H., F.-W. Wang, D. Zeng (2004). Intelligence and security informatics for homeland security: information, communication, and transportation. *IEEE Transactions on Intelligent Transportation Systems* 5(4), 329–341.

Frisen, M., J.D. Mare (1991). Optimal surveillance. *Biometrika* 78, 271–280.

Gahegan, M. (2001). Data mining and knowledge discovery in the geographical domain. National Academy's white paper.

Halkidi, M., Y. Batistakis, M. Vazirgiannis (2002a). Cluster validity methods: part 1. *SIGMOD Record* 31, 40–45.

Halkidi, M., Y. Batistakis, M. Vazirgiannis (2002b). Clustering validity checking methods: part II. *SIGMOD Record* 31, 19–27.

Hu, P.J.-H., D. Zeng, H. Chen, C. Larson, W. Chang, C. Tseng, J. Ma (2006). A system for infectious disease information sharing and analysis: design, implementation, and evaluation. *IEEE Transactions on Information Technology in Biomedicine*, submitted.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26, 1481–1496.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society A* 164, 61–72.

Levine, N. (2002). *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations*. The National Institute of Justice, Washington, DC.

Miller, H.J., J. Han (2001). *Geographic Data Mining & Knowledge Discovery: An Overview*. Taylor and Francis, London.

Roddick, J.F., M. Spiliopoulou (1999). A bibliography of temporal, spatial and spatio-temporal data mining research. *SIGKDD Explorations* 1, 34–38.

Rogerson, P.A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine* 16, 2081–2093.

Rogerson, P.A. (2001). Monitoring point patterns for the development of space–time clusters. *Journal of the Royal Statistical Society A* 164, 87–96.

Rogerson, P.A., Y. Sun (2001). Spatial monitoring of geographic patterns: an application to crime analysis. *Computers, Environment and Urban Systems* 25, 538–556.

Rogerson, P., I. Yamada (2004). Approaches to Syndromic Surveillance When Data Consist of Small Regional Counts. *Morbidity and Mortality Weekly Report,* 53(Supplement), 79–85.

Rogerson, P.A., I. Yamada (2004). Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches. *Statistics in Medicine* 23, 2195–2214.

Sonesson, C., D. Bock (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society: Series A* 166, 5–12.

Yao, X. (2003). Research issues in spatio-temporal data mining, *presented at UCGIS workshop on geospatial visualization and knowldge discovery*, Lansdowne, Virginia.

Zeng, D., W. Chang, H. Chen (2004). A comparative study of spatio-temporal hotspot analysis techniques in security informatics, in: *Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems*, Washington.

Zeng, D., W. Chang, H. Chen (in press). Clustering-based spatio-temporal hotspot analysis techniques in security informatics. *IEEE Transactions on Intelligent Transportation Systems*.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 8

Deception and Intention Detection

Judee K. Burgoon, Matthew L. Jensen, John Kruse, Thomas O. Meservy and Jay F. Nunamaker, Jr.

Center for the Management of Information, University of Arizona, 1130 East Helen Street, # 427, Tucson, AZ 85719-4427, USA

Abstract

Detecting deception and hostile intentions is a difficult but critical task that security professionals across the world must perform. Numerous methods of deception detection have been created to aid security professionals in their responsibilities; however, the majority of detection methods are unwieldy in a screening scenario. We propose a model for intention detection based on deception detection that focuses on observable, behavioral cues. Observation of these cues is unobtrusive and can be performed without the cooperation of the individual being scrutinized. Our approach, spanning video, audio, and textual modalities, has yielded promising first steps. However, much more remains to be done until a real time system will be ready to be implemented and used by security professionals.

1 Introduction

In the current climate of heightened security and insistence on information assurance, catching liars represents a difficult and important task. Recent interest in security has only highlighted an ancient problem: separating deception from truth. Deception has fascinated researchers for centuries, yet understanding and reliably identifying deception are still elusive, complex problems which must be addressed.

Research has shown that the average person holds a deep-seated truth bias (Burgoon et al., 1994; Levine et al., 1999). Humans tend to believe what they are told as long as messages do not significantly violate their

expectations and preconceived notions (Feeley and deTurck, 1995). This opens a tremendous gap for those motivated to deceive. Security and law enforcement professionals simply do not have the wherewithal to rigorously challenge the veracity of everything they hear or see. The vigilance required by the typical border agent or airport security screener is truly daunting given the multitude of communication exchanges that takes place during a shift. Moreover, those entrusted with these types of tasks often suffer from what is known as the Othello error, wherein they tend to misjudge truthful exchanges as deceptive (Ekman, 1985).

Through the years, researchers and others have proposed various methods to identify deception. These methods have ranged from discredited physiognomic evaluations to cutting edge brain scans. Over the last several years, the Center for the Management of Information (CMI) at the University of Arizona has conducted over a dozen experiments to study deception with over 2,100 subjects (Burgoon et al., 2003a,b; Zhou et al., 2003, in press). These experiments have been instrumental in creating an understanding of the factors influencing deception, and have guided the building of automated tools for enhancing deception detection and the creation of training for security personnel (George et al., 2003; Zhou et al., in press). Furthermore, research into deception has led to the examination of the relationship between deception and human intention. In this chapter, we present relevant findings concerning deception and deception detection. We begin by illustrating the relationships between deception, intentions, and behavior. Next we review current methods to identify deception. We present a model that allows the creation of an unobtrusive system which distinguishes between truth and deception through analysis of multiple communication channels. Finally, we present descriptions of our automated methods and offer results from testing these methods.

2 Deception, intentions, and behavior

Deception is a message knowingly and intentionally transmitted with the intent to foster false beliefs or conclusions (Buller and Burgoon, 1996). This definition encompasses strategies to mislead such as equivocation, ambiguity, evasiveness, and outright falsification. Deception has been linked to a number of observable behaviors and physiological reactions (see Section 3). Deception cannot be directly measured; however, the arousal, cognitive load, and self-monitoring that may accompany deception can be observed and measured. It is through these cues that lie-catchers ferret out deception.

Determining possible human intention from observable cues is a more difficult endeavor. The variety of intentions is limitless and an individual may possess a single intention or multiple intentions at any point in time. Further, there are no known links between intention and all possible behavior manifestations. For example, a simple smile could represent some

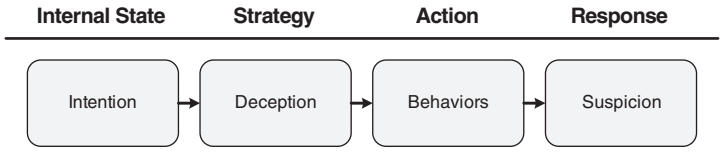


Fig. 1. Role of deception in determining intention from behavior.

harmless intention or alternatively it could indicate "duping delight" or masked criminal objectives.

To narrow the focus of investigation, we have limited our attention to differentiating between hostile and benign intentions. Hostile intention is defined as any intention to do something which is criminal in nature. The judgment between hostile and benign intentions can only be known and validated in hindsight; therefore, we have adopted suspicion level as a way to compare intentions. The connection between intention and suspicion is not direct. As shown in Fig. 1, intentions may be inferred by detecting the presence of deception in behavior. The premise of inferring intention from deception rests on the idea that individuals with hostile intentions will be deceptive about their true intentions in order to avoid detection. We recognize that deception is only one strategy a person who harbors hostile intent may employ, and not all people who deceive harbor hostile intentions. However, we believe that identifying deception is a productive first step in recognizing criminal intent.

Timing is key to correctly identifying hostile intentions. In order to be able to detect hostile intentions, individuals who harbor hostile intentions must conceal or deceive before they have a chance to enact their intentions. If deception is present in their communication, deceptive behavioral cues may be displayed that may arouse suspicion.

There are several different circumstances in which deception and hostile intentions can be observed. Our research program primarily focuses on three main scenarios—monitoring of people queuing in lines or standing in common areas, a standing interaction between two individuals, and a seated interview.

The monitoring of people queuing in lines and standing in common areas does not elicit deception *per se*, as there is no direct interaction between individuals where overt deception can take place. Instead, humans and systems can focus on identifying arousal cues that may indicate concealment and avoidance that accompanies deception. For instance, a person in a line with hostile intent may behave differently than others around him or her. Such a person may exhibit behaviors consistent with agitation, or more likely, overcontrol.

The second scenario occurs when the subject interacts with security personnel. In some cases this is as simple as moving through a metal detector. In others, it may involve detailed communication about where one is going

and why. The confrontation may increase subject arousal and subsequent behavior and also may broaden the scope of potential detection.

In the third scenario, if the suspicions of security personnel are raised significantly, the subject can be diverted to a structured interview with more controlled conditions. Under these conditions, the subject may experience significant arousal. Also, in such controlled conditions, the opportunity arises to use more intrusive methods of deception detection such as the polygraph.

3 Existing methods of deception detection

Many methods of deception detection currently exist and some are regularly utilized by professional lie-catchers such as police and security professionals. Additionally, many new methods of deception detection have been proposed, and their levels of effectiveness and reliability are currently being studied. The methods of deception detection can be divided into two broad categories: physiological and behavioral. Methods belonging to each category are listed in Table 1 and brief descriptions of each method are given below.

3.1 Physiological methods

These methods are premised on the assumption that arousal, emotions, and cognitive changes associated with deception generate systematic, physiological changes in blood flow, hemo-oxygenation, neuronal activity, and the like (Vrij, 2000). Technologies that tap into one or more physiological processes may accurately discriminate truthful from deceptive communication. Out of all the deception detection approaches, perhaps the most recognized is the polygraph or "lie-detector." The basic assumption behind the polygraph is that deception causes an increase in arousal stemming from feelings of guilt or fear of the consequences of being caught in a lie. The polygraph detects arousal via sensors attached to the body that measure heart rate, palmar sweat, and respiratory features. There are two main methods of interviewing which use the polygraph: the Control Question

Table 1
Methods of deception detection

Physiological	Behavioral
Polygraph	Statement validity assessment
Brain activity analysis	Linguistic analysis
Voice stress analysis	Micro-momentary expression analysis
Thermal scanning	Behavioral analysis

Test (CQT) and the Guilty Knowledge Test (GKT). The CQT uses three types of questions for comparison to crime-specific questions to ferret out possible deception. Neutral questions are innocuous questions which are not intended to increase arousal. Relevant questions relate directly to the crime. Finally, control questions are vague and cover long periods of time. They are meant to evoke embarrassment and interviewees typically respond deceptively. The levels of arousal from the control and neutral questions are used as a comparison to the crime-related questions (Vrij, 2000). Although the CQT test is commonly used in the United States, it has often been criticized as subjective, non-scientific, and unreliable (Ben-Shakhar and Elaad, 2003; Faigman et al., 2003).

The GKT determines whether an interviewee has knowledge about a crime that would only be known to the perpetrator. A series of questions presents both irrelevant and crime-related information to the interviewee and the polygraph records any heightened reaction. The GKT enjoys a more objective, scientific footing (Ben-Shakhar and Elaad, 2003); however, specific and confidential details about a crime must be obtained for its use. These details may be difficult to obtain and the guilty party may not recognize the specific details. Both the CQT and the GKT require interpretation by a skilled examiner and the decisions of these examiners are subjective and therefore may vary.

Other emergent methods intended to augment the polygraph rely on analysis of brain activity. One method utilizes an electroencephalogram (EEG) to measure event-related brain potentials (ERPs). Experimentally and in criminal applications, this method has yielded high accuracy (Johnson, Barhardt, Zhu, 2003). In addition to the EEG, functional magnetic resonance imaging (fMRI) has been used to differentiate between real and imagined events (Ganis et al., 2003). Currently the reliability and accuracy of deception detection based on the fMRI is being debated. As with the polygraph, analysis of brain activity requires the use of sensors attached to the interviewee's body. A non-invasive alternative that is now being investigated is near infrared spectroscopy (NIRS) (Villringer, 1993), which uses optical technology to measure neuronal, metabolic, and hemodynamic changes. NIRS does not require people to remain stationary and can be used while they are ambulatory, thus making it more suitable for field environments.

Another alternative that measures peripheral cardiac activity is high-definition thermal imaging (Pavlidis and Levine, 2002; Pavlidis et al., 2002). This method is also based on the premise that deception increases the level of arousal in interviewees. Proponents of the thermal imaging method suggest that arousal is shown by an instantaneous warming pattern around the eyes. This method has been tested in conjunction with the polygraph; however, it may be used independently of the polygraph, thus loosening the requirement for intrusive equipment to be attached to the interviewee.

Voice stress analysis (VSA) has been proposed as a method to automatically detect deception. Proponents of this method believe that deception causes psychological stress and indicators of this stress are identified from multiple cues in the voice. Numerous studies have concluded that deceivers exhibit an elevation in voice pitch when deceiving (Vrij, 2000). However, the validity of the VSA has been challenged in recent tests (Hollien and Harnsberger, 2006). Other vocal features may prove to be more diagnostic.

3.2 Behavioral methods

Statement validity assessment (SVA) is a general category of deception detection methods which focuses on verbal content. Two common methods of SVA are Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM). CBCA is based on the Undeutsch hypothesis which states that 'a statement derived from a memory of an actual experience differs in content and quality from a statement based on invention or fantasy' (Undeutsch, 1989). CBCA takes place during a structured interview where the interviewer scores responses according to predefined criteria such as general characteristics, specific contents, motivation-related contents, and offense-related elements. CBCA has been used successfully in judging the validity of statements given by children and it has been used in criminal cases where children are involved (Vrij, 2000).

RM also uses a scoring mechanism to judge potential deception; however, it is based on the hypothesis that verbal recall of actual events will contain more perceptual, contextual, and affective information than recall of fabricated events. Reality monitoring requires the interviewer to judge levels of clarity, perceptual information, spatial information, temporal information, affect, reconstruction ability of the story, realism, and cognitive operations (Sporer, 1997). CBCA and RM require trained interviewers to conduct interviews and probe where needed. Although common criteria exist for judging validity, interviews are not standardized and may be subjective. Further, these methods require careful review of verbal content and do not provide immediate feedback.

By building on the hypothesis that descriptions of fabricated events differ from descriptions of actual events, researchers have experimented with the structural nature of statements to determine if linguistic characteristics differ between deceptive and truthful communication. In contrast to CBCA and RM, this method operates independently of message meaning and can be automatically conducted. This method has been utilized in the analysis of written statements, emails, and chat conversations (Zhou et al., 2003, 2004). Two areas in linguistic analysis are message feature mining and speech act profiling. Message feature mining begins with the identification of potential deceptive features from text. Sample features include average sentence length, passive voice ratio, emotiveness, and word diversity (Zhou et al., 2003, 2004). These features are then used to classify

potential deceptive communication. Speech act profiling is a method of conversation classification and visualization. This method is useful in transcribed conversations or interactions via chat and it compares the communicative style of the interactants (Burgoon et al., under review). Deceptive strategies such as equivocation and indecisiveness can be seen using speech act profiling.

Finally, observation of behavioral cues is also used as a method of deception detection. Many people suspect that deceivers act differently than truth tellers. However, most people are mistaken in their beliefs about which cues are associated with deception. Even trained professionals are fooled by over-reliance on misleading cues (Vrij, 2000). Despite the reliance on mistaken deceptive cues, numerous studies have shown that deceivers behave differently than truth tellers (DePaulo et al., 2003; Ekman, 1985; Zuckerman and Driver, 1985). Differences include lack of head movement (Buller et al., 1994) and lack of illustrating gestures that accompany speech (Vrij et al., 2000). Other studies have indicated that deceivers may not be able to control micro-momentary facial expressions when lying about emotions (Ekman, 1985). Based on the changes in behavior and facial expressions observed in deceivers and truth tellers, researchers have investigated the possibility of determining truth through automatic analysis of the body and face. One attempt at automatically identifying emotions from the face includes computational methods of compositing and caricature (Benson et al., 1999). However, many other methods also exist.

4 Application of methods in screening scenarios

The applications of a general-purpose deception detection tool or system are many and varied. We believe that some of the greatest value could be realized in the three typical security monitoring and screening scenarios which we previously described. For a method of deception detection to be feasible in screening it must be usable in a natural setting. This necessitates unobtrusive instrumentation, a prompt judgment of suspicion, and robustness under varying conditions. For example, the use of a polygraph in each screening scenario would be infeasible as it requires each interviewee to be attached to body sensors for an extended amount of time.

Additionally, the tool or system must scale well to match the potentially large numbers of people who pass through screening measures. Proper scalability precludes methods that require extensive human intervention and analysis. Methods which can take advantage of computer processing and are partially automatable are preferred.

Finally, to gain widespread adoption, the method must be reasonably inexpensive. It may not be reasonable to equip all screening checkpoints with expensive thermal imaging cameras which must be maintained at a constant temperature. Such cameras may be useful in highly sensitive areas

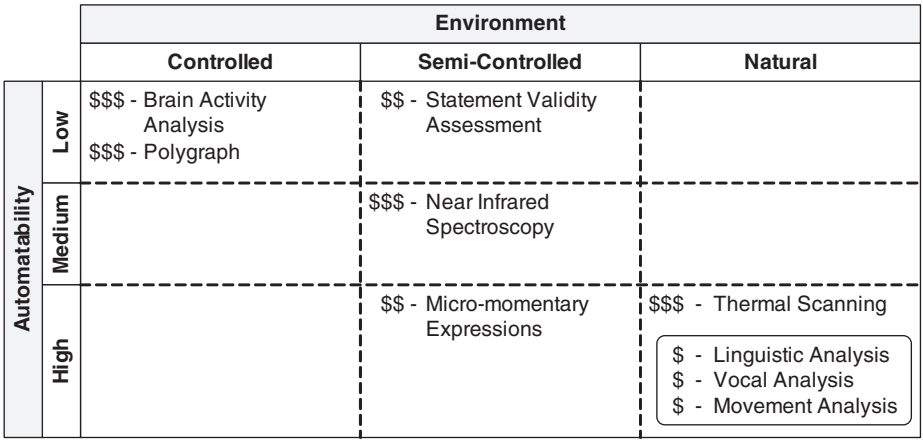


Fig. 2. Characterization of each method of deception detection.

where improved detection is necessary; however, the cost may be prohibitive for other screening areas. Figure 2 displays a characterization of each deception detection method according to requirements on the environment, potential automatability, and possible expense. CMI has chosen to focus its efforts on methods of deception and intention detection which are highly automatable, low in cost, and can be used in a natural environment. These methods are linguistic analysis, vocal analysis, and movement (kinesic and proxemic) analysis.

5 Model for unobtrusive deception and intention detection

Our framework for identifying deception and hostile intentions stems from three multi-disciplinary theories. Interpersonal Deception Theory (IDT) (Buller and Burgoon, 1996) models the dynamic, interactive nature of interpersonal deception. It specifies that receivers and deceivers alter their tactics based on perceived successes or failures during the interaction. IDT is the key for mapping behavioral cues into behavioral signatures that may be used for deception detection. IDT depicts the process-oriented nature of interpersonal deception and multiplicity of pre-interactional, interactional, and outcome factors that are thought to influence it. Among its relevant precepts is the supposition that deception is a strategic activity subject to a variety of tactics for evading detection. It also recognizes the influence of receiver behaviors on sender displays, and it views deception as a dynamic and iterative process, a sequence of moves and countermoves that enables senders to make ongoing adaptations that further hamper detection.

Expectancy Violations Theory (EVT) (Burgoon, 1978) is concerned with what nonverbal and verbal behavior patterns are considered normal or

expected, what behaviors constitute violations of expectations, and what consequences violations create. Its proponents contend that specific behavioral cues are less diagnostic than whether a sender's behavior conforms to or violates expected behavioral patterns and that receivers are more likely to attune to such violations. In other words, it is more useful to type information according to whether it includes behavioral anomalies, deviations from baseline, and discrepancies among indicators.

Finally, Signal Detection Theory (Green and Swets, 1966) provides a classification technique which can be used to identify deception. Signal Detection Theory specifies that a threshold should be set while classifying potentially deceptive behavior. If a sender displays deceptive cues which exceed the threshold, then that sender would be classified as deceptive. The threshold is set so as to maximize correct classifications. An updated model of our approach based on these three theories is shown in Fig. 3.

In this model, deception is manifested by various types of behavior. These behaviors contain numerous cues which are good candidates for analysis. All of these cues can be remotely monitored and have the potential to be automatically analyzed. Linguistic cues include features like word selection, phrasing, and sentence structure. Content/theme cues are taken from the meaning of the sender's words. Meta-content cues are derived from the types of topics the content addresses. Kinesic cues are found in the way a person moves. Proxemic cues are determined from how people use spacing and distancing to communicate nonverbally. Chronemic cues concern a person's use of time. For example, a person might establish dominance by arriving late to a meeting. Finally, vocalic cues refer to properties of the voice. For example, tension may be evident in the tone of a person's voice.

Multiple behavioral cues extracted from various communication channels are fused together to derive a behavioral profile. This profile is then compared with past individual or general norms and any deviation between the profiles is used to create a judgment of deception or truth. Alternatively, behavioral cues could be first compared with behavioral norms and the results of this comparison could then be fused into a judgment of truth or deception.

6 Automatic detection of deception

In following our general approach for identifying deception, we have taken initial steps in creating a system that can be used in a natural environment, is inclined to automation, and is relatively inexpensive. Our general approach for automatically detecting deception is based on a typical pattern classification approach. The input stream is segmented into meaningful units, low- and high-level features are extracted from these units, and then a variety of classification methods are used to identify segments that are likely deceptive. Table 2 provides a general overview of the steps in context of the text, audio, and video channels.

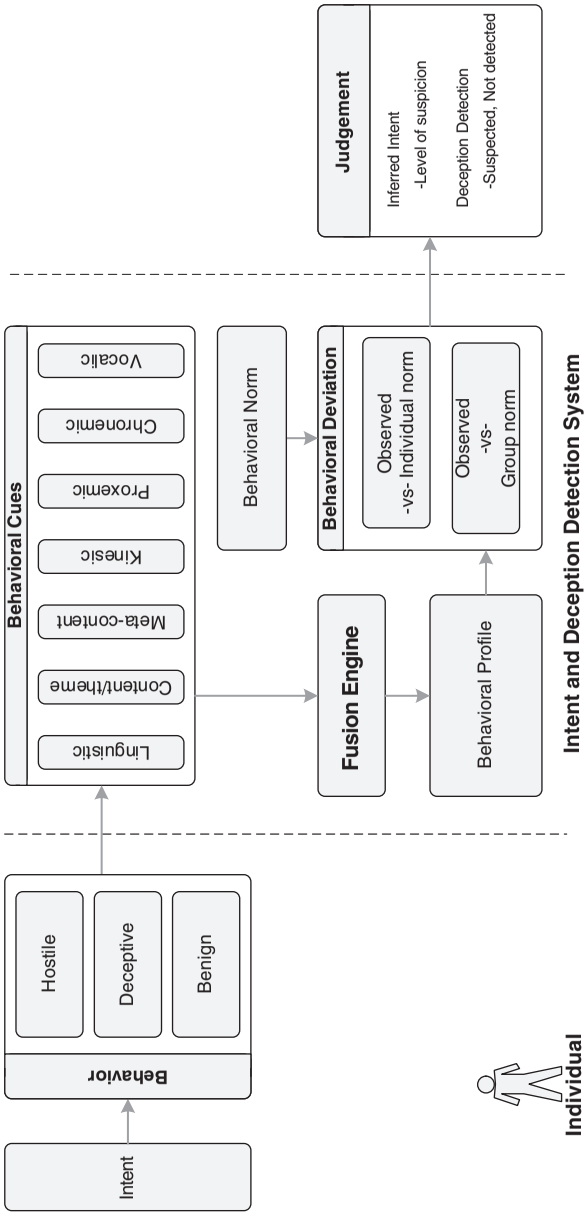


Fig. 3. Model for deception and intention detection.

Table 2
Basic approach for deception detection in text, audio, and video

	Text	Audio	Video
Low-level feature extraction	Word count, sentence length, part of speech tagging	Fundamental frequency, gain/energy, low-pass filter	Blob analysis to extract information about head and hands
High-level feature extraction	Speech act extraction	Response latency, turns, interruptions, tension	Movement feature extraction
Classification methods	Discriminant analysis	Discriminant analysis	Discriminant analysis
	Logistic regression	Logistic regression	Logistic regression
	Decision trees	Decision trees	Decision trees
	Neural networks	Neural networks	Neural networks
	Support vector machines	Support vector machines	Support vector machines

6.1 *Automatic deception detection in text*

First, textual interactions are manually segmented into logical and meaningful units. For example, an interviewee's response to an interrogator's question might be considered a meaningful unit.

A number of low-level features are then identified from the text. Simple low-level features include word quantity, average sentence length, and so forth. Other features depend on the results of a part of speech tagger. For instance, the passive voice ratio divides the total number of passive verbs in a segment by the total number of verbs. More sophisticated features include emotiveness (total number of adjectives and adverbs divided by total number of nouns and verbs) and content word diversity (total number of unique content words divided by total number of content words).

A number of high-level features are also extracted from each segment. These high-level features often involve multiple words or phrases and many of the features attempt to understand the semantics of the text. For instance, a number of speech acts are extracted from the text in an attempt to build a speech act profile. Speech act profiling is a method of analyzing and visualizing conversations and participants' behavior according to how they converse rather than the subject of the conversation (Burgoon et al., under review). Some example speech acts include simple acknowledgements, an agreeable opinion, a non-opinion statement, and WH-questions. Using speech acts, we can gain an understanding of deceptive indicators. For example, by using speech acts we can gauge the uncertainty present in a

given segment which may be indicative of deceivers' messages (DePaulo et al., 2003).

In past research we have used a variety of classification methods for deception detection in text. Typically, supervised learning methods that utilize manually coded results are used to train a model. Once obtained, the results can be used as a feedback tool for modifying the features, the granularity, and/or the classification methods in an effort to improve the results (Burgoon et al., under review). There are numerous methods that could be used to classify, but often we use decision trees, neural networks, and support vector machines to determine deception from an original segment. The results of these methods based on our datasets are promising (approaching 90% correct cross-validated classification).

6.2 Automatic deception detection in audio

Recorded voices in the form of digitized audio files serve as input to this method. These files are segmented into logical and meaningful units. The audio signal for both the subject and the interviewer needs to be identified. In this chapter we do not address automatic identification and segmentation of speech segments spoken by each individual in a conversation, though several automated methods do exist (Adami et al., 2002). In our data set, subject and interviewer voices were recorded on separate channels, and thus features for each individual were extracted using audio in the relevant channels.

Next, low-level audio features are extracted from these segments. Some basic low-level features include fundamental frequency and gain/energy. These features can be directly extracted from the audio signal. In our research we use proprietary toolkits, though basic low-level features could be extracted with existing, publicly available toolkits. More sophisticated low-level audio features, such as low-pass filter output, response latency, and audio sample speech/silence segments are also created. These features often require additional calculations. For example, audio samples that had a signal-to-noise ratio (SNR) less than or equal to 9 dB were declared silence frames. Additionally, some features build on other low-level calculations. For example, the response latency feature captures the length of time of silence between the end of an interviewer's question and the beginning of the subject's response and necessarily relies on the speech/silence measures.

From the low-level features, we compute additional high-level features that may help to distinguish deception from truth. Some of these features include identification of interviewer/subject turns and general speech disfluencies (interruptions, unfilled pause length). Figure 4 identifies some of these high-level features that were extracted for an interviewer and subject.

Both statistical methods, including discriminant analysis and logistic regression, and machine learning methods, including alternating decision trees, neural networks, and support vector machines have been used as

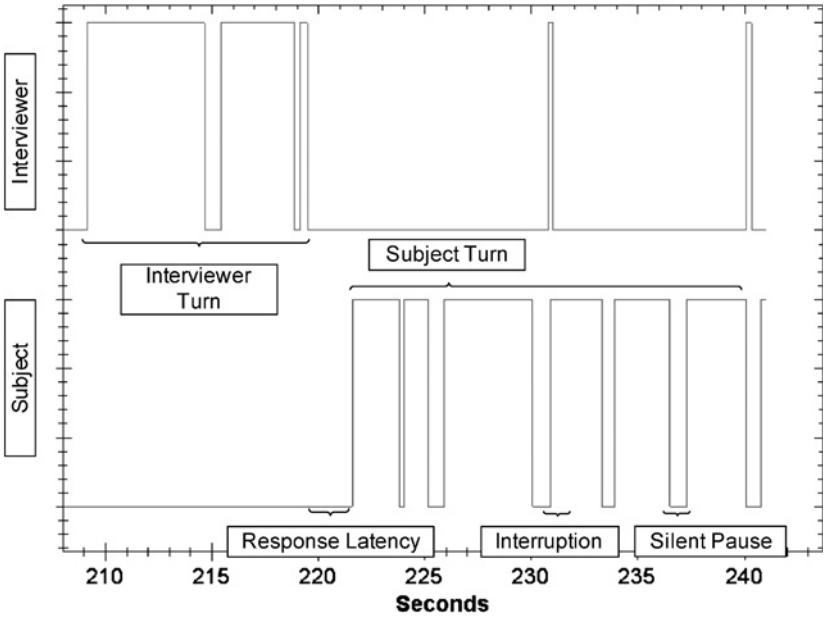


Fig. 4. High-level vocalic features in an interaction.

methods in classifying deception in audio. The results of these methods show promising cross-validation rates (approaching 78% correct classification).

6.3 *Automatic deception detection in video*

Videos are also manually segmented into logical and meaningful units. Similar to the text condition, an interviewee's response to an interrogator's question might be considered a meaningful unit. Although we have investigated automatic segmentation of videos in a lecture environment, we have not yet applied those techniques to deception detection.

General metrics are extracted from the video segments using a refined method of ''blob'' analysis developed by the Computational Biomedicine Imaging and Modeling Center (Lu et al., 2005). The algorithm uses color analysis, shape approximation, and filtering to track the head and the hands in the video clip. Figure 5 displays a video frame subjected to blob analysis. General metrics extracted from each blob include x and y coordinates for the center of each blob, a major and minor axis length, and the angle of tilt of the major axis.

From the extracted general metrics additional features are computed. Previously, we proposed a taxonomy of nonverbal, movement-based features for use in deception detection (Meservy et al., 2005a). The taxonomy contains both features that trained human observers might recognize easily and accurately (e.g., a hand touching the head) and other features that

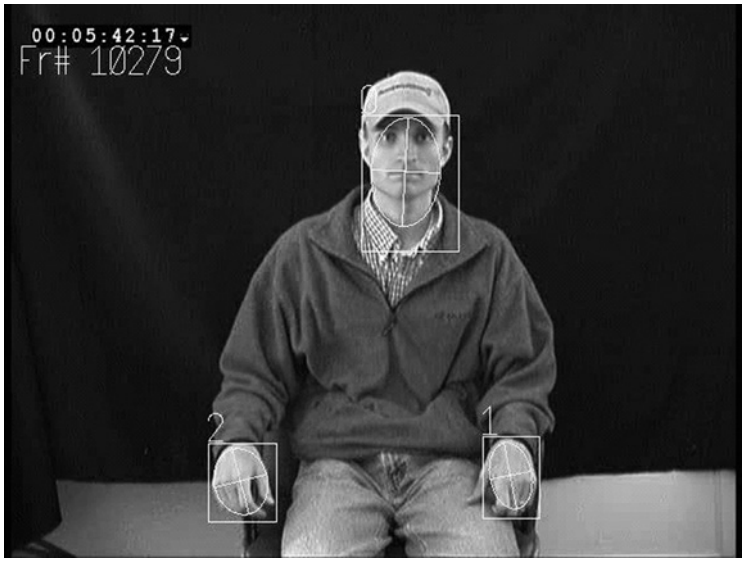


Fig. 5. Video frame subjected to blob analysis (Meservy et al., 2005b).

trained human observers would have difficulty precisely tracking (e.g., the speed of movement of a particular blob over a single video frame). Figure 6 illustrates (a) the low-level metrics extracted as part of blob analysis, (b) a single-frame distance feature that is computed from the general metrics using a simple Euclidean distance between the center coordinates of two blobs, and (c) a multiple frame feature that captures the distance a blob has moved over time.

From the extracted features, general patterns in movements can be detected as shown in Fig. 7. Past research has utilized a number of different classification methods to discriminate between truth and deception. Like other deception detection methods, both statistical and machine learning methods have been used in classifying deception. The results of these methods show promising cross-validation rates (approaching 88% correct classification).

7 Lessons learned

We have learned a number of lessons from the approaches we have taken to classify deception. While some of these lessons are specific to a particular modality, most of the lessons are generalizable across modalities. Below, we discuss some general lessons learned for each step in our approach.

The extraction of low-level features is particularly challenging in the video modality. In audio, many of the low-level features are easily extracted with existing toolkits. In the text modality low-level features such as word counts

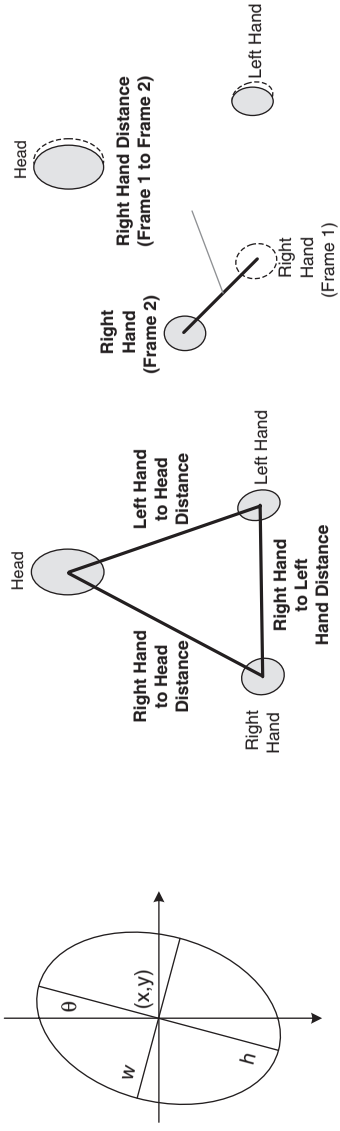


Fig. 6. Sample extracted features (Meservy et al., 2005b).

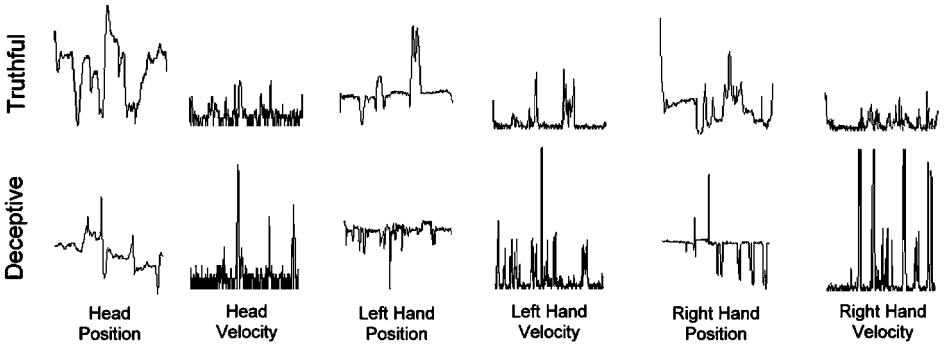


Fig. 7. Deceptive and truthful feature signatures (Meservy et al., 2005b).

and sentence length are easily computed. Part of speech tagging is a well-studied phenomenon and we primarily rely on existing tools. The extraction of low-level features from video, as previously mentioned, relies on blob analysis. We have found that our implementation of blob analysis handles a wide range of skin tones and once initialized can run unattended. It returns precise measurements as it tracks the blobs over time. However, it does require a number of skin samples to be manually extracted from the video in order to identify the head and hands. We have found that results may vary depending on the clothing that an interviewee wears. For example, low neck lines and sleeveless shirts can dramatically change the shape of the blobs. When hands are unidentifiable, either due to occlusion or movement out of the video frame, blob analysis attempts to find another matching region.

Lessons learned from the extraction of high-level features are generally applicable to all modalities. Many of these high-level features are based on existing deception theory and thus have been studied for many years. Automatic extraction of these features provides a level of consistency and overcomes some of the limitations of humans. For example, in the video modality we extract more than a hundred features. Many of these features, such as the distance that a hand moves in 1/30th of a second, capitalize on the precision that a computer provides. Existing deception theories may need to be expanded to include these precise features.

Many cues from all categories can be automatically extracted. However, the difficulty of extraction varies greatly. For example, audio cues extracted for frequency and energy/intensity are fairly straightforward as they can be directly extracted from the audio signal without any additional contextual information. Additionally, many features contain similar information and are highly correlated. High-level features are computed from low-level features and thus, the level of unique information provided by each new feature is limited.

We have found that in the data sets we have used, the classification methods that we employ to identify deception in the text, audio, and video

modalities provide higher accuracy rates than a typical human. Although accuracy rates exceed typical human ability, we argue that a human agent should not be replaced. The human agent should incorporate the prediction of the automated system with additional information the system is unable to detect. We have also found that after we train a model, the prediction is often a quick calculation. Finally, feature selection greatly influences accuracy rates. We have to pay particular attention to feature selection when we use machine-learning techniques in order to maximize cross-validation rates.

In addition to the lessons that we have learned from our approach, we have also gained some general insight about deception from studying different modalities. For example, we have found that deceptive messages vary depending on the modality. The analysis of our data sets indicates that in face-to-face interactions, deceivers convey fewer details and shorter messages compared with deceivers in the text modality. The analysis of our video data sets indicates that deceivers typically display rigid head posture, limited hand movement, and general over control. Additional significant features that we can automatically extract are listed in Table 3. The features are described from the reference point of the deceiver.

Table 3
Summarized deceptive cues

Text	Audio	Video
More quantity	Higher pitch	Less change in head angle
More ellipses in text (pauses)	More pitch variance	Less change in head position
Less possessive pronoun usage	Decreased intensity	Less change in hand positions
Less complexity	Increased response latency	Less distance between the right and left hands
Less diversity	Increased subject turns	Less distance between hands and the body
More non-immediacy	Less fluency	Less amount of time hands are away from body
More time communicating	Increased overlap in speaking	
Simpler sentences	Less talk time	
Shorter sentences	Increased unfilled pause length	
Less informality		
Less redundancy		
Fewer modal verbs		
Fewer modifiers		
More subjectivity		
More uncertainty		

8 Challenges and future steps

Although progress has been made in the creation of a system that can assist humans in detecting deception, many challenges remain. We are currently working to resolve these issues by employing the strategies described below.

8.1 Nature of deception

The very nature of deception is a challenge as we try to identify lies. Much deceptive communication is mixed with an element of truth. Deception levels can vary from outright falsification to simple concealment. Thus one person's type of deception may differ from another person's. We have attempted to address this challenge by asking subjects in experiments how honest they were (measured on a 0 to 10 scale) in their communication. Further, we have begun to implement an individual level of deception detection where interactions from the same subject are analyzed to look for indicators that distinguish deceptive responses from truthful ones.

8.2 Real time

There are many technical difficulties that we must confront to create a real time system. First, initialization of many of the analysis methods is time consuming. One example is video analysis where skin color samples must be manually collected in order to track body movements (recent tracking methods have moved away from color sampling). Second, real time analysis of the most granular level of detail is currently not feasible given computing and cost limitations.

We are experimenting with different initialization techniques that can decrease start-up time. We are testing sampling methods to discover what feasible granularity provides the most accurate detection ability and we are strengthening our processes in order to handle the load of real time data.

8.3 Fusion of multiple cues

We recently started an investigation into the fusion of multiple types of cues in order to provide a more accurate prediction of deception. We believe that groups of cues rather than single cues or a handful of cues is more effective in detecting deception. Also, fusing cues together provides larger coverage of all possible indicators as a skilled deceiver may try to hide certain deceptive cues. Fusion will provide flexibility in various conditions as each communication channel may be weighted for reliability in a given environment.

8.4 Real data sets

Ecologically valid data sets represent the most significant challenge because they are so difficult to obtain. Using a real data set is the fundamental test of the system and the results are a true measure of how the system may be useful. The subjects are under real stress and engage in consequential deception. Future work on automating behavioral analysis will need to be subjected to the litmus test of accuracy and minimized false alarms in real-world contexts.

9 Conclusion

Detecting deception and hostile intentions is a basic task that security professionals across the world face. While intentions of an individual remain difficult to fully recognize, we believe that the study of deception is a productive first step toward identifying intentions that may be harmful or criminal in nature. Much effort has been exerted to create deception detection techniques; however, many of the detection methods are unwieldy in a screening scenario. Efforts at the CMI at the University of Arizona have focused on building a system that is useful in a natural environment, unobtrusive, and inexpensive. Such a system could augment security professionals' ability to detect deception. Our methods, which include video, audio, and text analysis methods, have yielded promising first steps. However, much remains to be done before a real time system will be ready for implementation with security professionals.

For many decades, researchers have attempted to teach a computer how to understand human communication. With constantly expanding technological capabilities, this goal is becoming more attainable. We believe that the future of this research will see a shift from focusing on physiological cues of deception to investigating strategic interactions manifested in behavior. We believe that computers will increasingly be able to understand and interpret human communication of which only a small part will be deceptive communication. As the abilities of computers increase, researchers will face serious ethical and moral issues as these technologies are implemented and used.

10 Questions for discussion

1. What are the benefits and drawbacks of unobtrusive deception detection methods compared with more intrusive methods?
2. Interpersonal deception theory models deception as a strategic interaction. What are some strategies that people use to conceal their true intentions? How might these strategies affect detection?

3. There are benefits and limitations in analyzing video, audio, and textual data streams. What are they? How easily can noise be introduced into each of these streams?
4. How can analysis of the video, audio, and text channels be combined to create a more robust method of deception detection?
5. There are numerous methods for fusing information to arrive at a final judgment of deception or truth (e.g., fusing features from each individual channel to generate a judgment; merging judgments from individual channels; etc.). Discuss the advantages and disadvantages of each fusion approach.
6. A major obstacle in this area of research is obtaining high-stakes deceptive interactions that can be shared among researchers. Where could these interactions be collected and how could they be shared while maintaining an individual's right to privacy?

Acknowledgments

Portions of this research were supported by funding from the U.S. Air Force Office of Scientific Research under the U. S. Department of Defense University Research Initiative (Grant #F49620-01-1-0394) and Department of Homeland Security—Science and Technology Directorate under cooperative agreement NBC2030003. The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official U.S. Government position, policy, or decision.

References

Adami, A., S. Kajarekar, H. Hermansky (2002). A new speaker change detection method for two-speaker segmentation. *Paper presented at the IEEE International Conference on Acoustics, Speech*, Orlando, FL.

Ben-Shakhar, G., E. Elaad (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a meta-analytic review. *Journal of Applied Psychology* 88(1), 131–151.

Benson, P.J., R. Campbell, T. Harris, M.G. Frank, M.J. Tovee (1999). Enhancing images of facial expressions. *Perception and Psychophysics* 61(2), 259–274.

Buller, D., J. Burgoon (1996). Interpersonal deception theory. *Communication Theory* 6, 203–242.

Buller, D., J. Burgoon, C. White, A. Ebesu (1994). Interpersonal deception: VII. Behavioral profiles of falsification, equivocation and concealment. *Journal of Language and Social Psychology* 13(4), 366–395.

Burgoon, J.K. (1978). A communication model of personal space violations: explication and an initial test. *Human Communication Research* 4(2), 129–142.

Burgoon, J.K., J.P. Blair, E. Moyer (2003a). Effects of communication modality on arousal, cognitive complexity, behavioral control and deception detection during deceptive episodes. *Paper presented at the Annual Meeting of the National Communication Association*, Miami Beach, FL.

Burgoon, J.K., J.P. Blair, T. Qin, J.F. Nunamaker, Jr. (2003b). Detecting deception through linguistic analysis. *Paper presented at the NSF/NIJ Symposium on Intelligence and Security Informatics, Proceedings Lecture Notes in Computer Science*, Vol. 2655, Tucson, AZ, pp. 91–101.

Burgoon, J.K., D.B. Buller, A.S. Ebesu, P. Rockwell (1994). Interpersonal deception: V. Accuracy in deception detection. *Communication Monographs* 61(4), 303–325.

Burgoon, J.K., D.P. Twitchell, M.L. Jensen, M. Adkins, J. Kruse, A. Deokar, et al. (Under Review). Detecting concealment in transportation screening.

DePaulo, B.M., B.E. Malone, J.J. Lindsay, L. Muhlenbruck, K. Charlton, H. Cooper (2003). Cues to deception. *Psychological Bulletin* 129, 74–118.

Ekman, P. (1985). *Telling Lies*. W. W. Norton & Company, New York.

Faigman, D.L., S.E. Fienberg, P.C. Stern (2003). Limits of the polygraph. *Issues in Science and Technology* 20(1), 40–46.

Feeley, T.H., M.A. deTurck (1995). Global cue usage in behavioral lie detection. *Communication Quarterly* 43, 420–430.

Ganis, G., S.M. Kosslyn, S. Stose, W.L. Thompson, D.A. Yurgelun-Todd (2003). Neural correlates of different types of deception: an fMRI investigation. *Cerebral Cortex* 13(8), 830–836.

George, J., D.P. Biros, J.K. Burgoon, J.F. Nunamaker, Jr. (2003). Training professionals to detect deception. *Paper presented at the NSF/NIJ Symposium on "Intelligence and Security Informatics"*, Tucson, AZ.

Green, D.M., J.A. Swets (1966). *Signal Detection Theory and Psychophysics*. Wiley, New York, NY.

Hollien, H., J.D. Harnsberger, (2006). Voice stress analyzer instrumentation evaluation. Final report, Counter-Intelligence Field Activity (Contract FA 4814-04-0011).

Johnson, R. Jr., J. Barnhardt, J. Zhu (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology* 64(3), 217–253.

Levine, T.R., H.S. Park, S.A. McCornack (1999). Accuracy in detecting truths and lies: documenting the "veracity effect". *Communication Monographs* 66(2), 125–144.

Lu, S., G. Tsechpenakis, D.N. Metaxas, M.L. Jensen, J. Kruse (2005). Blob analysis of the head and hands: a method for deception detection. *Paper presented at the Hawaii International Conference on System Science (HICSS'05)*, Hawaii.

Meservy, T.O., M.L. Jensen, J. Kruse, J.K. Burgoon, J.F. Nunamaker, Jr. (2005a). Automatic extraction of deceptive behavioral cues from video, in: P. Kantor, G. Muresan, F. Roberts, D. Zeng, F.-Y. Wang, H. Chen, R. Merkle (eds.), *Intelligence and Security Informatics: Proceedings of the IEEE International Conference on Intelligence and Security Informatics ISI 2005*, Atlanta, GA, USA, Vol. 3495. Springer-Verlag, Berlin, pp. 198–208.

Meservy, T.O., M.L. Jensen, J. Kruse, D.P. Twitchell, G. Tsechpenakis, J.K. Burgoon, et al. (Sept/Oct, 2005b). Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems* 20(5), 36–43.

Pavlidis, I., N.L. Eberhardt, J. Levine (2002). Seeing through the face of deception: thermal imaging offers a promising hands-off approach to mass security screening. *Nature* 415(3), 35.

Pavlidis, I., J. Levine (2002). Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine* 21(6), 56–64.

Sporer, S.L. (1997). The less traveled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology* 11, 373–397.

Undeutsch, U. (1989). The development of statement reality analysis, in: U. Undeutsch (ed.), Credibility Assessment, Kluwer, Dordrecht, The Netherlands, pp. 101–121.

Villringer, A. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience Letters* 154, 101.

Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and Implications for Professional Practice*. Wiley, Chichester, UK.

Vrij, A., K. Edward, K.P. Roberts, R. Bull (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior* 24(4), 239–263.

Zhou, L., J.K. Burgoon, J.F. Nunamaker, Jr., D.P. Twitchell (2004). Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: an empirical investigation. *Group Decision and Negotiation* 13(1), 81–106.

Zhou, L., J.K. Burgoon, D. Twitchell, J.F. Nunamaker, Jr. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation* 13(1), 81–106.

Zhou, L., D.P. Twitchell, T. Qin, J.K. Burgoon, J.F. Nunamaker, Jr. (2003, January 6–9, 2003). An exploratory study into deception detection in text-based computer-mediated communication. *Paper presented at the Thirty-Sixth Annual Hawaii International Conference on System Sciences (CD/ROM)*, Big Island, Hawaii.

Zhou, L., D.P. Twitchell, T. Qin, J.K. Burgoon, J.F. Nunamaker, Jr. (2004). Toward the automatic prediction of deception—an empirical comparison of classification methods. *Journal of Management Information Systems* 20(4), 139–166.

Zuckerman, M., R.E. Driver (1985). Telling lies: verbal and nonverbal correlates of deception, in: A.W. Siegman, S. Feldstein (eds.), Multichannel Integrations of Nonverbal Behavior, Lawrence Erlbaum Associates, Hillsdale, NJ.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 9

Identification of Hidden Groups in Communications[1]

J. Baumes, M. Goldberg, M. Magdon-Ismail and W. Wallace
Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA

Abstract

This chapter presents statistical and algorithmic approaches to discover groups of actors that hide their communications within the myriad of background communications in a large communication network. Our approach to discovering hidden groups is based on the observation that a pattern of communications exhibited by actors in a social group pursuing a common objective is different from that of a randomly selected set of actors. We distinguish two types of hidden groups: *temporal*, which exhibits repeated communication patterns, and *spatial* which exhibits correlations within a snapshot of communications aggregated over some time interval. We present models and algorithms, together with experiments showing the performance of our algorithms on simulated and real data inputs.

1 Introduction

1.1 Motivation

Modern communication networks (telephone, e-mail, Internet chatroom, etc.) facilitate rapid information exchange among millions of users around the world. This vast communication activity provides the ideal environment for groups to plan their activity undetected: the related communications are embedded (hidden) within the myriad of random background communications, making them difficult to discover. When a number of individuals in a network exchange communications related to a common goal, or a common activity, they form a group; usually, the presence of the coherent

[1]This research was partially supported by NSF grants 0324947 and 0346341.

communication activity imposes a certain structure of the communications on the set of actors, as a group. A group of actors may communicate in a structured way while not being forthright in exposing its existence and membership. This chapter develops statistical and algorithmic approaches to discover such hidden groups.

Finding hidden groups on the Internet has become especially important since the September 11, 2001 attacks. The tragic event underlines the need for a tool (a software system) which facilitates the discovery of hidden (malicious) groups during their *planning* stage, before they move to implement their plans. A generic way of discovering such groups is based on discovering *correlations* among the communications of the actors of the communication network. The *communication graph* of the network is defined by the set of its actors, as the vertices of the graph, and the set of communications, as the graph's edges. Note that the content of the communications is not used in the definition of the graph. Although the content of the messages can be informative and natural language processing may be brought to bear in its analysis, such an analysis is generally time consuming and intractable for large datasets. The research presented in this chapter makes use of only three properties of a message: its time, the name of the sender, and the name of the recipient of the message.

Our approach to discovering hidden groups is based on the observation that a pattern of communications exhibited by actors in a social group pursuing a common objective is different from that of a randomly selected set of actors. Thus, we focus on the discovery of such groups of actors whose communications during the observation time period exhibit statistical *correlations*. We will differentiate between *spatial* and *temporal* correlations, which, as we shall see, lead to two different notions of hidden groups.

1.2 Temporal correlation

One possible instance of temporal correlation is an occurrence of a *repeated communication pattern*. Temporal correlation may emerge as a group of actors are planning some future activity. This planning stage may last for a number of time cycles, and, during each of them, the members of the group need to exchange messages related to the future activity. These message exchanges imply that, with high probability, the subgraph of the communication graph formed by the vertices corresponding to the active members of the group is connected. If this *connectivity* property of the subgraph is repeated during a sufficiently long sequence of cycles, longer than is *expected* for a randomly formed subgraph of the same size, then one can discover this *higher-than-average* temporal correlation, and hence identify the hidden group.

Thus, in order to detect hidden groups exhibiting temporal correlations, we exploit the non-random nature of their communications as contrasted

with the general background communications. We describe efficient algorithms, first appearing in Baumes et al. (2004, 2005c), which, under certain conditions on the density of the background communications, can efficiently detect such hidden groups. We differentiate between two types of temporally correlated hidden groups: a *trusting*, or *non-secretive* hidden group, whose members are willing to convey their messages to other hidden group members via actors that are not hidden group members, using these non-hidden group members as "messengers"; and a *non-trusting*, or *secretive* hidden group, where all the "sensitive" information that needs to be conveyed among hidden group members uses only other hidden group members as messengers.

Our results reveal those properties of the background network activity and hidden group communication dynamics that make detection of the hidden group easy, as well as those that make it difficult. We find that if the background communications are dense or more structured, then the hidden group is harder to detect. Surprisingly, we also find that when the hidden group is non-trusting (secretive), it is easier to detect than if it is trusting (non-secretive). Thus, a hidden group which tries to prevent the content of its messages from reaching third parties undermines its operations by becoming easier to detect!

1.3 Spatial correlation

We use spatial correlation to refer to correlations in the communications of a single communication graph, which represents a snapshot of the communications aggregated over some time interval (in contrast to temporal correlation which refers correlation in the communications over multiple communication graphs which represent successive snapshots of the communications). Spatial correlation of messages initiated by a group of actors in a social network can be identified by a *higher-than-average* total communication level *within* this group. This property does not rely on the content of the messages and is adequately described by the communication graph: the *edge density* of the corresponding set of vertices of the graph is higher than that of the average set. To be able to address a wide variety of applications, we consider a general notion of edge density, which compares the intensity of communications between the actors within a particular set and that between the set and the "outside world." The edge density may be defined in numerous ways depending on the desired characteristics of the discovered groups; our algorithms for discovering groups of higher density (potential hidden groups) are generic with respect to the definition of density. Furthermore, we find only groups which are more dense than any group sufficiently close, which reflects the principle of locality in a social network.

For our numerical experiments, we use two main ideas in defining density: one is the proportion of the number of actual communications to the

total number of possible communications and the other is the ratio of the number of communications within the group to the total number of group communications, including messages to individuals outside the group.

In graph-theoretical terminology, the problem we study is *clustering*. An important implication of our approach is that our algorithms construct clusters that may overlap, i.e., some actors may be assigned to more than one group. While there is much literature in the area of graph clustering, up until very recent work it has mainly focused on a specific subcase of the problem: graph-partitioning. As opposed to partitioning algorithms, which decompose the network into *disjoint* groups of actors, general clustering allows groups to extend to their natural boundaries by allowing overlap. We discuss prior work in the area of *partitioning*, and present three general *clustering* heuristics, originally described in Baumes et al. (2005a,b). We refer to these procedures by the names Iterative Scan (IS), Rank Removal (RaRe), and Link Aggregate (LA). We present experimental data that illustrate the efficiency, flexibility, and accuracy of these heuristics.

Searching for both spatial and temporal correlations may be combined to produce a more effective algorithm for the identification of hidden groups. The temporal algorithms may indicate that a large group of individuals are involved in planning some activity. The spatially correlated algorithm may then be used to cluster this large group into overlapping subgroups, which would correspond to smaller working groups within the larger group.

We present results from the testing of our spatial-hidden group algorithms on a number of real-world graphs, such as newsgroups and e-mail. We analyze the quality of the groups produced by the clustering algorithms. We also test the algorithms on random graph models in order to determine trends in both runtime and accuracy. One of the interesting experimental discoveries is that different implementations of the IS algorithm are optimized for different domains of application based on the sparseness (density) of the communication network.

2 Discovering temporal correlation

2.1 Literature review

Identifying temporally correlated hidden groups was initiated in Magdon-Ismail et al. (2003) using Hidden Markov models. Here, our underlying methodology is based on the theory of random graphs (Bollobás, 2001; Janson et al., 2000). We also incorporate some of the prevailing social science theories, such as homophily (Monge and Contractor, 2002), by incorporating group structure into our model. A more comprehensive model of societal evolution can be found in Goldberg et al. (2003) and Siebecker (2003). Other simulation work in the field of computational analysis of social and organizational systems (Carley and Prietula, 2001;

Carley and Wallace, 2001; Sanil et al., 1996) primarily deals with dynamic models for social network infrastructure, rather than the dynamics of the actual communication behavior, which is the focus of this chapter.

One of the first works analyzing hidden groups is found in Erickson (1981). Here, the author studies a number of secret societies, such as a resistance that was formed among prisoners at Auschwitz during the Second World War. The focus, as it is in this paper, was on the structure of such societies, and not on the content of communications. An understanding of a hidden network comes through determining its general pattern and not the details of its specific ties.

The September 11, 2001 terrorist plot spurred much research in the area of discovering hidden groups. Specifically, the research was aimed at understanding the terrorist cells that organized the hijacking. Work has been done to recreate the structure of that network and analyze it to provide insights on general properties that terrorist groups have. Analyzing their communication structure provides evidence that Mohamed Atta was central to the planning, but that a large percentage of the individuals would have needed to be removed in order to render the network inoperable (Stewart, 2001). In ''Uncloaking Terrorist Networks'', Krebs (2002) uses social network measures such as betweenness to identify which individuals were most central to the planning and coordination of the attacks. Krebs has also observed that the network that planned September 11 attempted to hide by making its communications sparse. While these articles provide interesting information on the history of a hidden group, our research uses properties of hidden groups to discover their structure before a planned attack can occur.

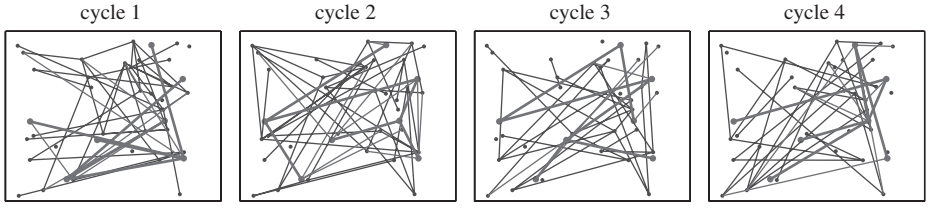
There is also work being done in analyzing the theory of networked groups, and how technology is enabling them to become more flexible and challenging to deal with. The hierarchical structure of terrorist groups in the past is giving way to more effective and less organized network structure (Ronfeldt and Arquilla, 2001). Of course, the first step to understanding decentralized groups is to discover them. What follows are some strategies to solve this problem.

2.2 Methodology

A temporally hidden group is a different kind of group from the normally functioning social groups in the society that engage in ''random'' communications. We define a *temporally hidden group*, or in this section labeled a *hidden group*, as some subset of the actors who are planning or coordinating some activity over time; the hidden group members may also be engaging in other non-planning related communications. The hidden group may be malicious (e.g., some kind of terrorist group planning a terror attack) or benign (e.g., a foursome planning their Sunday afternoon golf game). So in this sense, a hidden group is not assumed to be intentionally hiding, but the

group activity is initially unknown and masked by the background communications. The hidden group is attempting to coordinate some activity, using the communication network to facilitate the communications between its members. Our task now is to (1) discover specific properties that can be used to find hidden groups and (2) construct efficient algorithms that utilize those properties. The next steps such as formulation of empirically precise models and further investigation of the properties of hidden groups are beyond the scope of this methodology.

Whether intentional or not, in a normal society, communications will, in general, camouflage the planning related activity of the hidden group. This could occur in any public forum such as a newsgroup or chatrooms, or in private communications such as e-mail messages or phone conversations. However, the planning related activity is exactly the Achilles heel that we will exploit to discover the hidden group: on account of the planning activity, the hidden group members need to stay "connected" with each other during each "communication cycle." To illustrate the general idea, consider the following time evolution of a communication graph for a hypothetical society; here, communications among the hidden group are in bold, and each communication cycle graph represents the communications that took place during an entire time interval. We assume that information must be communicated among *all* hidden group members during one communication cycle.



Note that the hidden group is connected in each of the communication cycle figures above. We interpret this requirement that the communication subgraph for the hidden group be connected as the requirement that during a single communication cycle, information must have passed (directly or indirectly) from some hidden group member to all the others. If the hidden group subgraph is disconnected, then there is no way that information could have been passed from a member in one of the components to a member in the other, which makes the planning impossible during that cycle. The information need not pass from one hidden group member to every other directly: a message could be passed from A to C via B, where A, B, and C are all hidden group members. Strictly speaking, A and C are hidden group members; however, B need not be one. We will address this issue more formally in the next section. A hidden group may try to hide its existence by changing its connectivity pattern, or by throwing in "random" communications to non-hidden group members. For example, at some

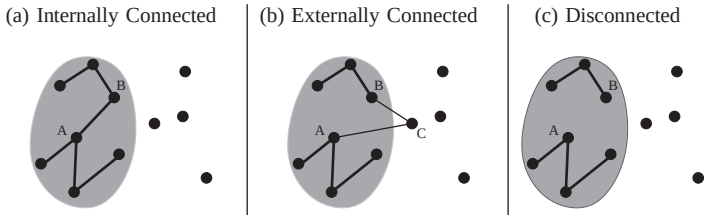
times the hidden group may be connected by a tree, and at other times by a cycle. None of these disguises changes the fact that the hidden group is connected, a property we will exploit in our algorithms.

We make the assumption here that the hidden group remains static over the time period when communications are collected. The algorithms described here would still be useful, however, as long as a significant subset of the group remains the same. The algorithms would likely not detect members that joined or left the group, but would discover a "core" group of members.

2.2.1 *Trusting versus non-trusting hidden groups*

Hidden group members may have to pass information to each other indirectly. Suppose that A needs to communicate with B. They may use a number of third parties to do this: $A \to C_1 \to \cdots \to C_k \to B$. *Trusting* hidden groups are distinguished from *non-trusting* ones by who the third parties C_i may be. In a trusting (or non-secretive) hidden group, the third parties used in a communication may be any actor in the society; thus, the hidden group members (A, B) trust some third-party couriers to deliver a message for them. In doing so, the hidden group is taking the risk that the non-hidden group members C_i have access to the information. For a malicious hidden group, such as a terrorist group, this could be too large a risk, and so we expect that malicious hidden groups will tend to be non-trusting (or secretive). In a non-trusting (secretive) hidden group, *all* the third parties used to deliver a communication *must* themselves be members of the hidden group, i.e., no one else is trusted. The more malicious a hidden group is, the more likely it is to be non-trusting.

Hidden groups that are non-trusting (versus trusting) need to maintain a higher level of connectivity. We define three notions of connectivity as illustrated by the shaded groups in the following figure:



A group is *internally connected* if a message may be passed between any two group members without the use of outside third parties. In the terminology of graph theory, this means that the subgraph induced by the group is connected. A group is *externally connected* if a message may be passed between any two group members, perhaps with the use of outside third parties. In graph theory terminology, this means that the group is a subset of a connected set of vertices in the communication graph. For

example, in figure (b) above, a message from A to B would have to use the outside third party C. A group is *disconnected* if it is not externally connected. The following observations are the basis for our algorithms for detecting hidden groups:

(i) Trusting hidden groups are *externally connected* in every communication cycle.
(ii) Non-trusting hidden groups are *internally connected* in every communication cycle.

We can now state the idea behind our algorithm for detecting a hidden group: a group of actors is *persistent* over communication cycles 1, …, T if it is connected in each of the communication graphs corresponding to each cycle. The two variations of the connectivity notion, internal or external, depend on whether we are looking for a non-trusting or a trusting hidden group. Our algorithm is intended to discover potential hidden groups by detecting groups that are persistent over a long time period. An example is illustrated in Fig. 1.

A hidden group can be hidden from view if, by chance, there are many other persistent subgroups in the society. In fact, it is likely that there will be many persistent subgroups in the society *during any given short time period*. However, these groups will be short-lived on account of the randomness of the society communication graph. Thus, we expect our algorithm performance to improve as the observation period increases.

2.2.2 *Detecting the hidden group*

Our ability to detect the hidden group hinges on two things. First, we need an efficient algorithm for identifying maximally persistent components over a time period II. Second, we need to ensure, with high probability, that over this time period there are no persistent components that arise, by

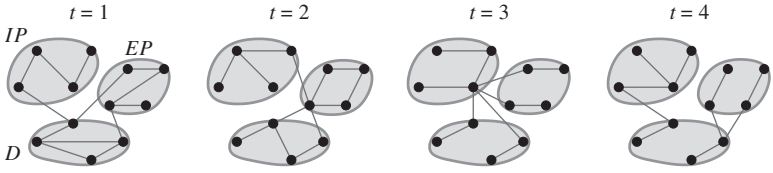


Fig. 1. Internally persistent, externally persistent, and non-persistent groups. The communication graph during four communication cycles is shown. Three groups are highlighted, *IP*, *EP*, and *D*. One can easily verify that *IP* is internally persistent during these four communication cycles, and so is a candidate for a non-trusting hidden group. *EP* is only internally persistent only for time periods 1 and 2. If we only observed data during this time period, then *EP* would also be a candidate for a non-trusting hidden group. However, *EP* is only externally persistent for all the communication cycles, and hence can only be a candidate for a trusting hidden group. *D* becomes disconnected during communication cycle 4, and hence is not a candidate for a hidden group.

chance, due to the background societal communications. We will construct algorithms to efficiently identify maximal components that are persistent over a time period II. Given a model for the random background communications, we can determine (through simulation) how long a time period of a group of a particular size must be persistent in order to ensure that, with high probability, this persistent component did not arise by chance, due to background communications.

2.3 *Algorithms*

Select $\mathbf{\Delta}$ to be the smallest time interval during which it is expected that information is passed among *all* group members. Index the communication cycles (which are consecutive time periods of duration $\mathbf{\Delta}$) by $t = 1, 2, \ldots, T$. Thus, the duration over which data are collected is $\Pi = \mathbf{\Delta} \cdot T$. The communication data are represented by a series of communication graphs, G_t, for $t = 1, 2, \ldots, T$. The vertex set for each communication graph is the set V of all actors.

The input to the algorithm is the collection of communication graphs $\{G_t\}$ with a common set of actors V. The algorithm splits V into persistent components, i.e., components that are connected in every G_t. The notion of connected could be either external or internal, and so we develop two algorithms, Ext_Persistent and Int_Persistent.

Each algorithm develops the partition in an iterative way. If we have only one communication graph G_1, then both the externally and the internally persistent components are simply the connected components of G_1. Suppose now that we have one more graph, G_2. The key observation is that two vertices, i and j, are in the same external component if and only if they are connected in both G_1 and G_2, i.e., they are in the same component in both G_1 and G_2. Thus, the externally persistent components for the pair G_1 and G_2 are exactly the intersections of the connected components in G_1 with those in G_2. This argument clearly generalizes to more than two graphs, and relies on the fundamental property that any subset of an externally connected set is also externally connected. Unfortunately, the same property does not hold for internal connectivity, i.e, a subset of an internally connected set is not guaranteed to be internally connected. However, a minor modification of the externally connected algorithm where one goes back and checks any set that gets decomposed leads to the algorithm for detecting internally persistent components (Fig. 2b). The formal details of the algorithms are given in Baumes et al. (2005c).

2.3.1 *Analysis*

The correctness and computational complexity results of the algorithms given in Fig. 2 are stated here. For full detail see Baumes et al. (2005c). We say that a set A is a *maximal* persistent set (internal or external) if it is persistent, and any other persistent set that contains at least one element of

(a) Externally persistent components

```
Ext_Persistent(\{G_t\}_{t=1}^T, V)
//Input: Graphs \{G_t = (E_t, V)\}_{t=1}^T.
//Output: A partition \mathcal{P} = \{V_j\} of V.
Use DFS to get the connected components \mathcal{C}_t of every G_t;
Set \mathcal{P}_1 = \mathcal{C}_1 and \mathcal{P}_t = \{\} for t > 1;
for t = 2 to T do
    for Every set A \in \mathcal{P}_{t-1} do
        Obtain a partition P' of A by intersecting A with every set in \mathcal{C}_t;
        Place P' into \mathcal{P}_t;
    end for
end for
return \mathcal{P}_T;
```

(b) Internally persistent components

```
Int_Persistent(\{G_t\}_{t=1}^T, V)
//Input: Graphs \{G_t = (E_t, V)\}_{t=1}^T.
//Output: A partition \mathcal{P} = \{V_j\} of V.
\{V_i\}_{i=1}^K = Ext_Persistent(\{G_t\}_{t=1}^T, V)
if K = 1, then
    \mathcal{P} = \{V_1\};
else
    \mathcal{P} = \cup_{k=1}^K Int_Persistent(\{G_t(U_k)\}_{t=1}^T, V_k);
end if
return \mathcal{P};
```

Fig. 2. Algorithms for detecting persistent components.

A is a subset of A. Clearly, any two maximal persistent sets must be disjoint, which also follows from the following lemma:

Lemma 1. *If A and B are non-disjoint externally (respectively, internally) persistent sets, then A∪B is also externally (respectively, internally) persistent.*

Theorem 1. (Correctness of Ext_Persistent). *Algorithm Ext_Persistent correctly partitions the vertex set V into maximal externally connected components for the input graphs* $\{G_t\}_{t=1}^T$.

Let E_t denote the number of edges in G_t, and let E denote the total number of edges in the input, $E = \sum_{t=1}^T E_t$. The size of the input is then given by $E + V \cdot T$.

Theorem 2. (Complexity of Ext_Persistent). *The computational complexity of algorithm Ext_Persistent is in* $O(E + VT)$ *(linear in the input size).*

Theorem 3. (Correctness of Int_Persistent). *Algorithm Int_Persistent correctly partitions the vertex set V into maximal internally connected components.*

Theorem 4. (Complexity of Int_Persistent). *The computational complexity of algorithm Int_Persistent is in* $O(V \cdot E + V^2 \cdot T)$.

2.3.2 Statistical significance of persistent components

Let h be the size of the hidden group we wish to detect. Suppose that we find a persistent component of size $\geq h$ over T communication cycles. A natural question is to ask how sure we can be that this is really a hidden group versus a persistent component that happened to arise by chance due to the random background communications.

Let $X(t)$ denote the size of the largest persistent component over the communication cycles 1, …, t that arises due to normal societal communications. $X(t)$ is a random variable with some probability distribution, since the communication graph of the society follows a random process. Given a confidence threshold, ε, we define the detection time $\tau_\varepsilon(h)$ as the time at which, with probability $1 - \varepsilon$, the largest persistent component arising by chance in the background is smaller than h, i.e.,

$$\tau_\varepsilon(h) = \min\{t : P[X(t) < h] \geq 1 - \varepsilon\}.$$

Then, if after $\tau_\varepsilon(h)$ cycles, we observe a persistent component of size $\geq h$, we can claim, with a confidence $1 - \varepsilon$, that this did not arise due to the normal functioning of the society, and hence must contain a hidden group. $\tau_\varepsilon(h)$ indicates how long we have to wait in order to detect hidden groups of size h. Another useful function is $h_\varepsilon(t)$, which is an upper bound for $X(t)$, with high probability $(1 - \varepsilon)$, i.e.,

$$h_\varepsilon(t) = \min\{h : P[X(t) < h] \geq 1 - \varepsilon\}.$$

If, after a given time t, we observe a persistent component with size $\geq h_\varepsilon(t)$, then with confidence at least $1 - \varepsilon$, we can claim it to contain a hidden group. $h_\varepsilon(t)$ indicates what sizes hidden group we can detect with only t cycles of observation. The previous approaches to detect a hidden group assume that we know h or fix a time t at which to make a determination. By slightly modifying the definition of $h_\varepsilon(t)$, we can get an even stronger hypothesis test for a hidden group. For any fixed $\delta > 0$, define

$$H_\varepsilon(t) = \min\left\{h : P[X(t) < h] \geq 1 - \frac{\delta}{t^{1+\delta}}\varepsilon\right\}.$$

Then one can show that if $X(t) \geq H_\varepsilon(t)$ at any time, we have a hidden group with confidence $1 - \varepsilon$.

Note that the computation of $\tau_\varepsilon(h)$ and $h_\varepsilon(t)$ constitutes a pre-processing of the *society's communication* dynamics. This can be done either from a model (such as the random graph models we have described) or from the true, observed communications over some time period. More importantly, this can be done off-line. For a given realization of the society dynamics, let $T(h) = \min\{t : X(t) < h\}$. Some useful heuristics that aid in the computation of $\tau_\varepsilon(h)$ and $h_\varepsilon(t)$ by simulation can be obtained by assuming that $T(h)$ and $X(t)$ are approximately normally distributed, in which case:

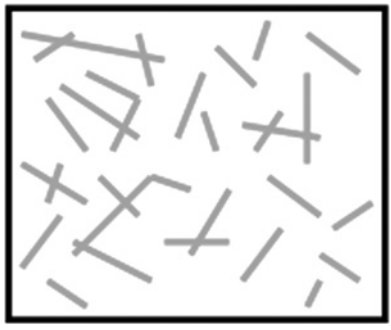
Confidence level (%)	$\tau_\varepsilon(h)$	$h_\varepsilon(t)$
50	$E[T(h)]$	$E[X(t)]$
84.13	$E[T(h)] + \sqrt{\mathrm{Var}[T(h)]}$	$E[X(t)] + \sqrt{\mathrm{Var}[X(t)]}$
97.72	$E[T(h)] + 2\sqrt{\mathrm{Var}[T(h)]}$	$E[X(t)] + 2\sqrt{\mathrm{Var}[X(t)]}$

2.4 Random graphs as communication models

Social and information communication networks, e.g., the Internet and WWW, are usually modeled by graphs (Newman, 2003; Carley and Prietula, 2001; Carley and Wallace, 2001; Sanil et al., 1996), where the actors of the networks (people, IP-addresses, etc.) are represented by the vertices of the graph, and the connections between the actors are represented by the graph edges. Since we have no *a priori* knowledge regarding who communicates with whom, i.e., how the edges are distributed, it is appropriate to model the communications using a random graph. In this paper, we study hidden group detection in the context of two random graph models for the communication network: uniform random graphs and random graphs with embedded groups. In describing these models, we will use standard graph theory terminology (West, 2001), and its extension to *hypergraphs* (Berge, 1978). In a hypergraph, the concept of an edge is generalized to a *hyperedge* which may join more than two vertices. In addition to these two models, there are other models of random networks, such as the small world model and the preferential attachment model (Albert and Barabasi, 2002). However, in this work we limited our experiments to the following models.

2.4.1 Random model

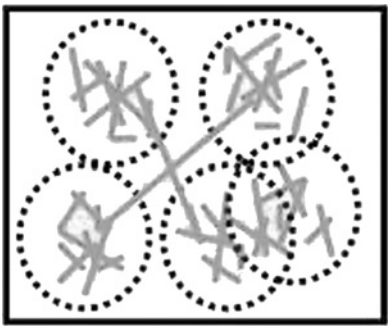
A simple communication model is one where communications happen at random uniformly among all pairs of actors. Such a communication model can be represented by the random graph model developed and extensively studied by Erdős and Rényi (1959, 1960, 1961) and Bollobás (2001). In this model, the graph is generated by a random process in which an edge between every pair of vertices is generated independently with a given probability p. The probability space of graphs generated by such a random process is denoted $G(n, p)$, or sometimes are called the Bernoulli graphs. We will use the $G(n, p)$ notation throughout this paper.



2.4.2 Group model

The $G(n, p)$ random graph model may not be a suitable model for large communication networks. Actors tend to communicate more often with

certain actors and less frequently with others. In a more realistic model, actors will belong to one or more social groups where communication among group members is more frequent than communication among actors that do not belong to the same group. This leads us to the hypergraph model of the communication network, in which the actors associate themselves into groups. In this paper, we assume that each group is static and contains m actors. While this is a simplification, it serves to illustrate all the essential ideas and results without undue complication. A group of actors is represented by a hyperedge in the graph, and an actor may belong to zero or more hyperedges. The set of all hyperedges represents the structure of the communication network. Since groups tend to be small, it is appropriate to model the communications within a group as a $G(m, p_g)$, where p_g is the probability within the group. We also allow communication between two actors that do not share a group in common; we denote such communications as external. The probability of an external communication is p_e; we further assume that $p_e \ll p_g$ because intra-group communications are much more likely than extra-group communications.



2.4.3 Connectivity of random graphs

The key idea of our algorithms is based on the following observation. For any subset of actors in a random model network, it is very unlikely that this subset is connected during a "long" consecutive period of time cycles, while a hidden group must stay connected (for its operations) as long as it functions as a group. Thus, we summarize here some results from random graph theory regarding how the connectivity of a $G(n, p)$ depends on n and p (Erdős and Rényi, 1959, 1960, 1961; Bollobás, 2001). These results are mostly asymptotic in nature (with respect to n); however, we use them as a guide that remains accurate even for moderately sized n.

Given a graph $G = \{V, E\}$, a subset $S \subseteq V$ of the vertices is connected if there exists a path in G between every pair of vertices in S. G can be partitioned into disjoint *connected components* such that every pair of vertices from the same connected component is connected and every pair of vertices in different connected components is not connected. The size of a

component is the number of its vertices; the size of the largest connected component is denoted by $L(G)$.

The remarkable discovery by Erdős and Rényi, usually termed *The Double Jump*, deals with the size of the largest component, and essentially states that $L(G)$ goes through two phase transitions as p increases beyond a critical threshold value. All the results hold asymptotically, with high probability, i.e., with probability tending to 1 when $n \to \infty$:

$p = c/n$		$p = ((\ln n)/n) + (x/n), \quad x > 0$
$L(G(n,p)) = \begin{cases} O(\ln n), & 0 < c < 1 \\ O(n^{2/3}), & c = 1 \\ \beta(c)n, & c > 1, \beta(c) < 1 \end{cases}$		$L(G(n,\ p)) = n$ with probability $e^{-e^{-x}}$

Note that when $x \to \infty$, the graph is connected with probability 1. Since our approach is based on the tenet that a hidden group will display a higher level of connectivity than the background communications, we will only be able to detect the hidden group if the background is not maximally connected, i.e., if $L(G) \neq n$. Thus, we expect our ability to detect the hidden group to undergo a phase transition exactly when the background connectivity undergoes a phase transition. For $p =$ constant or $p = d \ln n/n$ with $d > 1$, the graph is asymptotically connected which will make it hard to detect the hidden group. However, when $p =$ constant, connectivity is exponentially more probable than when $p = d \ln n/n$, which will have implications on our algorithms.

2.5 *Experiments and results*

In these tests, we simulate societies of sizes $n = 1000$ and 2000. The results for both the random background communication model and the group background communication model are presented in parallel. For each model, multiple time series of graphs are generated for communication cycles $t = 1, 2, \ldots, T$, where $T = 200$. Experiments were run on multiple time series (between 5 and 30), and averaged in order to obtain more statistically reliable results. In order to estimate $h_\varepsilon(t)$, we estimate $E[X(t)]$ by taking the sample average of the largest persistent component over communication cycles $1, \ldots, t$. Given h, the time at which the plot of $E[X(t)]$ drops below h indicates the time at which we can identify a hidden group of size $\geq h$.

We first describe the experiments with the random model $G(n, p)$. The presence of persistently connected components depends on the connectivity of the communication graphs over periods $1, 2, \ldots, T$. When the societal communication graph is connected for almost all cycles, we expect the society

to generate many large persistent components. By the results of Erdős and Rényi described in Section 2.4, a phase transition from short-lived to long-lived persistent components will occur at $p = 1/n$ and $p = \ln n/n$. Accordingly, we present the results of the simulations with $p = 1/n$, $p = c \ln n/n$ for $c = 0.9$, 1.0, and 1.1, and $p = 1/10$ for $n = 1000$. The rate of decrease of $E[X(t)]$ is shown in Fig. 3. For $p = 1/n$, we expect exponential or

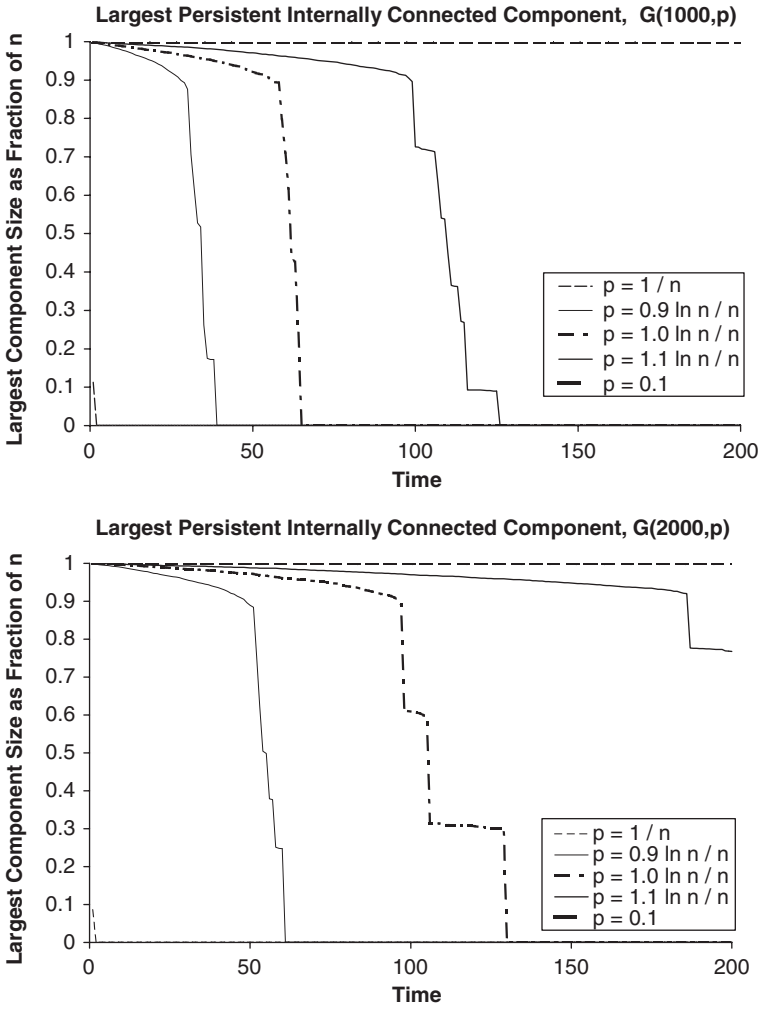


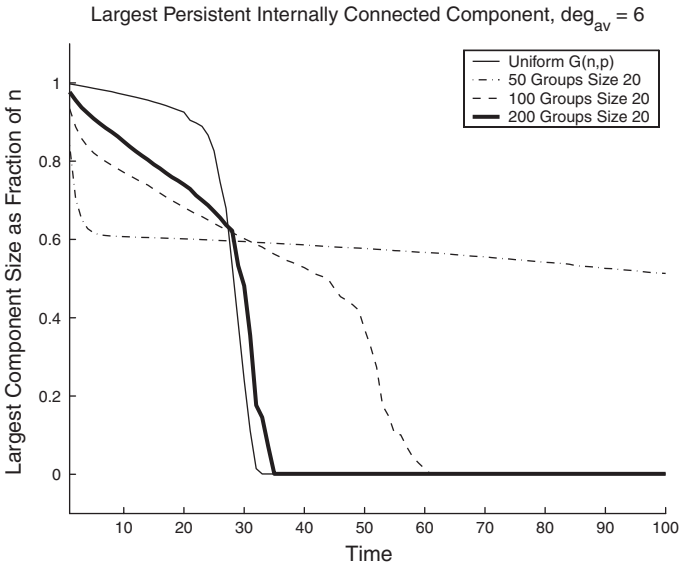
Fig. 3. The largest internally persistent component $E[X(t)]$ for the $G(n, p)$ model with $n = 1000$ and 2000. The five lines represent $p = 1/n$, $p = c \ln n/n$ for $c = 0.9$, 1.0, and 1.1, and $p = 1/10$. Note the transition at $p = \ln n/n$. This transition becomes more apparent at $n = 2000$. When p is a constant (i.e., does not depend on n; here we used 1/10), the graph is almost always connected. The results were averaged over a number of runs. The sharp jumps indicate where the largest component quickly jumped from $\sim 0.9n$ to 0 in different runs.

super-exponential decay in $E[X(t)]$ (Fig. 3, thin dashed line). This is expected because $L(G)$ is at most a fraction of n. An abrupt phase transition occurs at $p = \ln n/n$ (Fig. 3, dot-dashed line). At this point the detection time begins to become large. For constant p (where p does not depend on n, in this case 1/10), the graph is connected with probability tending to 1, and it becomes essentially impossible to detect a hidden group using our approach without any additional information (Fig. 3, thick dashed line). This will occur for any choice of a constant as n becomes large. That is, for any constant $p > 0$, there is an integer N such that if $n > N$, then $G(n, p)$ is connected with high probability, tending to 1.

The parameters of the experiments with the group model are similar to those of the $G(n, p)$ model. We pick the group size m to be equal to 20. Each group is selected independently and uniformly from the entire set of actors; the groups may overlap; and each actor may be a member of zero or more groups. If two members are in the same group together, the probability that they communicate during a cycle is p_g, otherwise the probability equals p_e. It is intuitive that p_g is significantly bigger than p_e; we picked $p_e = 1/n$, so each actor has about one external communication per time cycle. The values of p_g that we use for the experiments are chosen to achieve a certain average number of communications per actor; thus, the effect of a change in the structure of the communication graph may be investigated while keeping the average density of communications constant. The average number of communications per actor (the degree of the actor in the communication graph) is set to 6 in the experiments. The results do change qualitatively for different choices of average degree. The number of groups g is chosen from $\{50, 100, 200\}$. These cases are compared to the $G(n, p)$ structure with an average of six communications per actor. For the selected values of g, each actor is, on average, in one, two, and four groups, respectively. When g is 50, an actor is, on average, in approximately one group, and the overlaps of groups are small. However, when g is 200, each actor, on average, is in about four groups, so there is a significant amount of overlap between the groups. The goal of our experiments is to see the impact of g on finding hidden groups. Note that as g increases, any given pair of actors tends to belong to at least one group together, so the communication graph tends toward a $G(n, p_g)$ graph.

We give a detailed comparison between the society with structure (group model) and the one without (random model) in Fig. 4. The table shows $T(1)$, which is the time after which the size of the largest internally persistent component has dropped to 1. This is the time at which any hidden group would be noticed, since the group would persist beyond the time expected in our model.

We have also run similar experiments for detecting trusting groups. The results are shown in Fig. 5. As the table shows, for the corresponding non-trusting communication model, the trusting group is much harder to detect.



	Group Model			Random Model
g (# of groups)	50	100	200	$G(n, \frac{6}{n})$
$T(1)$	> 100	63	36	32

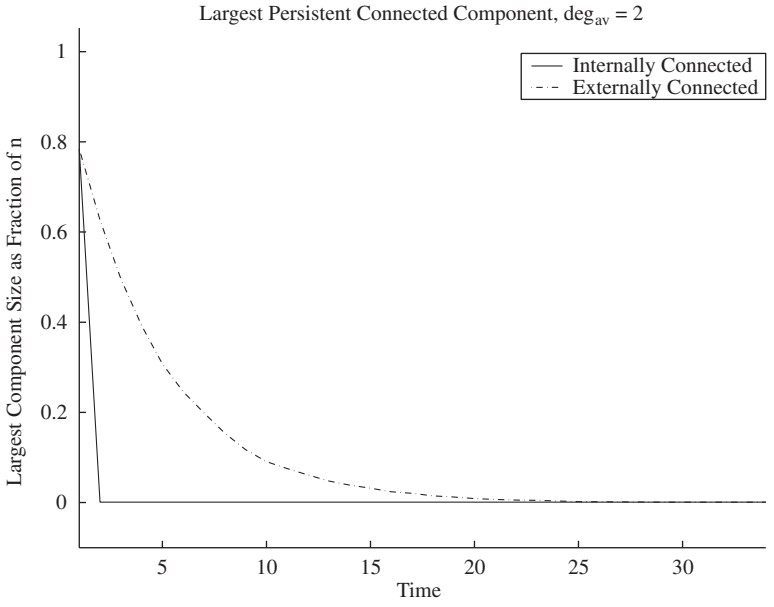
Fig. 4. Times of hidden group discovery for various amounts of group structure; each group is independently generated at random and has 20 actors. In all cases, $n = 1000$, $\deg_{av} = 6$, and the group size $m = 20$. Note how, as the number of groups becomes large, the behavior tends toward the $G(n, p)$ case.

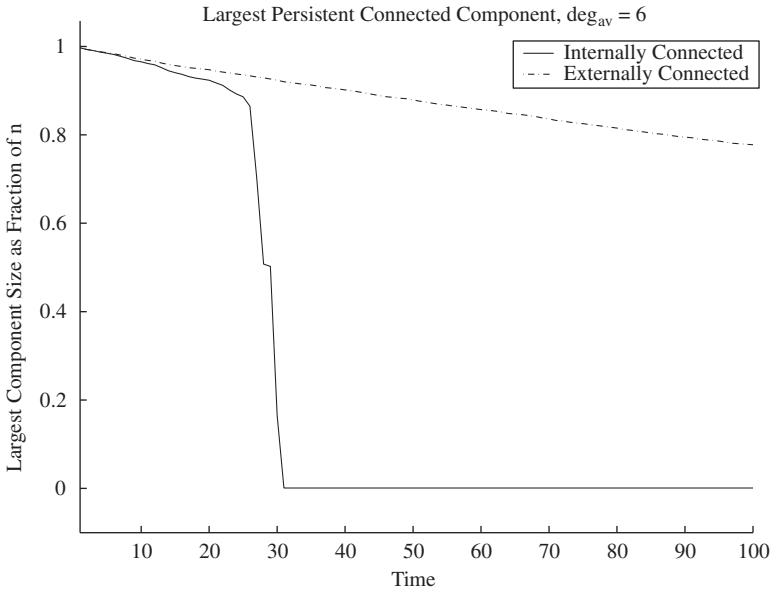
3 Discovering spatial correlation

3.1 Literature review

While an informal definition of the goal of clustering algorithms is straightforward, difficulty arises when formalizing this goal. There are two main approaches to clustering: partitioning and general clustering.

Partitioning, or hierarchical clustering, is the traditional method of performing clustering. In some circles, clustering and partitioning are synonymous. For example, Kannan et al. (2004) define clustering as "partitioning into dissimilar groups of similar items." However, the partitioning approach forces every cluster to be either entirely contained within or entirely disjoint from every other cluster. Partitioning algorithms are useful when the set of objects needs to be broken down into disjoint categories. These categories simplify the network and may be treated as separate entities. Partitioning is used in the fields of VLSI design, finite element methods, and transportation (Karypis and Kumar, 1998).





deg_{av}	T (1) for trusting groups	T (1) for non-trusting groups
2	28	2
6	> 100	32

Fig. 5. Times of hidden group discovery for non-trusting (internally connected) hidden groups and trusting (externally connected) hidden groups. In all cases the communication graphs are $G(n, p)$ with $n = 1000$.

Many partitioning algorithms attempt to minimize the number of connections between clusters, also called the *cut size* of the partition (Kheyfits, 2003; Kernighan and Lin, 1970; Hendrickson and Leland, 1995; Karypis and Kumar, 1998, 1999). The *ρ-separator* metric attempts to balance the sizes of the clusters, while also minimizing the cut size (Even et al., 1999). The *betweenness* metric is used to find a small cut by removing edges that are likely to split the network into components (Freeman, 1977; Girvan and Newman, 2002). In addition to trying to minimize the cut size, some algorithms attempt to maximize the quality of each cluster. Two metrics used in partitioning which define cluster quality are *expansion* and *conductance* (Flake et al., 2002; Kannan et al., 2004). A final metric relates to how well the members of the same cluster are related to the values in eigenvectors of the adjacency matrix of the network (Capocci et al., 2004).

Groups in social networks do not conform to this partitioning approach. For example, in a social network, an individual may belong to numerous groups (e.g., occupational, religious, political activity). A general clustering algorithm may put the individual into all these clusters, while a partitioning algorithm will only place the individual into one cluster. Classifying an individual as belonging to a single cluster or social group will often miss the full picture of the societal structure. As opposed to partitioning, general clustering allows individuals to belong to many groups at once. General clustering algorithms determine the zero, one, or more groups that each actor belongs to, without enforcing a partition structure on the clusters. This complex structure may more directly correspond to real-world clusters. However, permitting overlapping groups is a more complex problem, since each cluster may not be treated as a separate entity. See Fig. 6 for a comparison of partitioning and general clustering.

When clustering a network, there needs to be a definition of what constitutes a "good" cluster. In some sense, members of a cluster need to be "close" to each other, and "far" from the other objects in the network. There are many ways to define the criterion for a valid cluster.

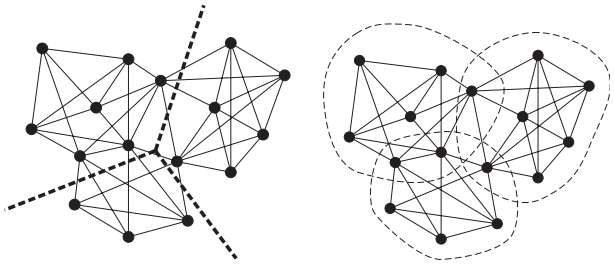


Fig. 6. A comparison of a partitioning (a) and a general clustering (b) of the same network.

The general clustering problem has been less widely studied than the partitioning problem; however, there are some algorithms that exist for discovering a general clustering of a network.

Some algorithms of this type are well suited for web networks (Kleinberg et al., 1999; Gibson et al., 1998; Mishra et al., 2002). These algorithms all attempt to find clusters by optimizing a metric referred to as *bicliqueness*. Though often used for partitioning, eigenvector correlation may also be used to discover overlapping clusters (Skillicorn, 2004). *Local optimality* is a generic technique which can optimize many of the previously mentioned metrics, even metrics originally developed for partitioning. These algorithms have been applied to social networks (Baumes et al., 2005a,b). We present these algorithms in more detail in the following sections.

3.2 Methodology

Let $G = (V, E)$ be a graph whose nodes represent individuals, web pages, etc., and whose edges represent communications, links, etc. The graph may be directed or undirected. We present the definitions for directed graphs, the undirected case being similar. A *graph cluster* C is a set of vertices which can be viewed as a binary vector of length $|V|$ that indicates which nodes are members of C. The set of all graph clusters, $\mathscr{C}$, is the power set of V.

A *weight* function, or *metric*, is a function $W : \mathscr{C} \to \mathbb{R}$ that assigns a weight to a graph cluster. Associated to cluster C, we define three edge sets: $E(C)$, the edges induced by C; $E(C, \bar{C})$, the edges in E *from* C to its complement; $E(C, \bar{C})$, the edges in E to C from its complement. Let $E_{\text{out}}(C) = E(C, \bar{C}) + E(\bar{C}, C)$. We define the *internal and external edge intensities*,

$$p_{\text{in}}(C) = \frac{E(C)}{|C| \cdot (|C| - 1)}, \qquad p_{\text{ex}}(C) = \frac{E_{\text{out}}(C)}{2|C| \cdot (n - |C|)}$$

($p_{\text{ex}} = 1$ when $|C| = |V|$). We will consider three weight functions: the *internal edge-probability* W_{p}; the *edge ratio* W_{e}; and, the *intensity ratio* W_{i},

$$W_{\text{p}}(C) = p_{\text{in}}(C), \quad W_{\text{e}}(C) = \frac{E(C)}{E(C) + E_{\text{out}}(C)},$$

$$W_{\text{i}}(C) = \frac{p_{\text{in}}(C)}{p_{\text{in}}(C) + p_{\text{ex}}(C)}.$$

These metrics are measures of how intense the communication within the cluster is, relative to that outside the cluster; they can be efficiently updated locally, i.e., the metric may be updated by knowing only the connectivity of the one node that is added or removed (which improves the efficiency of the algorithms). A *set-difference* function δ is a metric that measures the difference between two clusters C_1 and C_2. Two useful set-difference functions are the *Hamming or edit distance* δ_{h} and the

percentage non-overlap δ_p:

$$\delta_h(C_1, C_2) = |(C_1 \cap \bar{C}_2) \cup (\bar{C}_1 \cap C_2)|,$$
$$\delta_p(C_1, C_2) = 1 - \frac{|C_1 \cap C_2|}{|C_1 \cup C_2|}.$$

The ε-*neighborhood* of a cluster $B_\varepsilon^\delta(C)$ is the set of clusters that are within ε of C with respect to δ, i.e., $B_\varepsilon^\delta(C) = \{C' | \delta(C, C') \leq \varepsilon\}$. For weight function W, we say that a cluster C^* is ε-*locally optimal* if $W(C^*) \geq W(C)$ for all $C \in B_\varepsilon^\delta(C^*)$.

We are now ready to formally state our abstraction of the problem of finding overlapping communities in a communication network. The input is a graph G, the communication graph, along with the functions W, δ, and ε. The output is a set of clusters $\mathcal{O} \subseteq \mathcal{C}$ such that $C \in \mathcal{O}$ iff C is ε-locally optimal. While our heuristic approaches are easily adapted to different weight and set-difference functions, we will focus on the choices $W = W_e$, $\delta = \delta_h$, and $\varepsilon = 1$, referring to the output clusters as locally optimal.

As stated, the problem is NP-hard. In fact, the restriction to $\delta = \delta_h$ and $\varepsilon = |V|$asks to find all the globally optimal clusters according to an arbitrary weight function W, which is well known to be NP-hard. Thus, we present heuristic, efficient (low-order polynomial time) algorithms that output candidate (overlapping) clusters, and then evaluate the quality of the output.

3.3 Algorithms

3.3.1 k-Neighborhood (k-N)

k-N is a trivial algorithm that yields overlapping clusters. The clusters are simply the k-Ns of a randomly selected set S of cluster centers. The inputs to this algorithm are k and$|S|$.

3.3.2 Rank removal

Algorithm RaRe is based on the assumption that within a communication network, there is a subset of "important" or high-ranking nodes, which do a significant amount of communication. RaRe attempts to identify these nodes and remove them from the graph, in order to disconnect the graph into smaller connected components. The removed node(s) are added to a set R. This process is repeated, until the sizes of the resulting connected components are within a specified range. These connected components can be considered the *core* of each cluster. Next, the vertices in R are considered for addition into one or more of these cores. If a vertex from R is added to more than one cluster, then these clusters now overlap. Note, however, that the cores of each cluster are disjoint, and only communicate with each other through vertices in R.

"Important" or high-ranking nodes are determined by a ranking function φ. These are the nodes which are removed at each iteration. We wish to remove nodes that will result in disconnecting the graph as much as possible. One choice is to remove vertices with high degree, corresponding to the choice $\varphi_{\mathrm{d}}(v) = \deg(v)$. Another approach that we have found to be experimentally better is to rank nodes according to their page rank, $\varphi_{\mathrm{p}}(v)$ (Page et al., 1998). The page rank of a node is defined implicitly as the solution to the following equation:

$$\phi_{\mathrm{p}}(v) = c \sum_{u,v} \frac{\phi_{\mathrm{p}}(u)}{\deg^{-}(v)} + \frac{1-c}{n} \tag{1}$$

where n is the number of nodes in the graph, $\deg^{-}(v)$ the out degree of vertex v, and c a decay factor between 0 and 1. An iterative algorithm to compute $\varphi_{\mathrm{p}}(v)$ for all the nodes converges rapidly to the correct value.

Once we have obtained the cores, we must add the vertices in R back into the cores to build up the clusters. Intuitively, a vertex $v \in R$ should be part of any cluster to which it is immediately adjacent, as it would have been part of the core if it were not removed at some step. Also, if we do not take this approach, we run the risk of v not being added to any cluster, which seems counter-intuitive, as v was deemed "important" by the fact that it was at one time added to R. This is therefore the approach which we take. We also add vertices in R to any cluster for which doing so increases the metric W. The algorithm is summarized in Fig. 7, and all the user specified inputs are summarized in Table 1.

It is important to note that the initial procedure of removing vertices, though not explicitly attempting to optimize any single metric, does produce somewhat intuitive clusters. The cores that result are mutually disjoint and non-adjacent. Consider a connected component C at iteration i. If C has more vertices than our maximum desired core size *max*, we remove a set R_i of vertices, where $|R_i| = t$. If the removal of R_i results in disconnecting C into two or more connected components $C_1, C_2, \ldots, C_k$, we have decreased the diameter of $C_1, C_2, \ldots, C_k$ with respect to C, resulting in more compact connected components. If the removal of R_i does not disconnect the graph, we simply repeat the procedure on the remaining graph until it either becomes disconnected or its size is less than *max*.

As an added performance boost, the ranks may be computed initially, but not recomputed after each iteration. The idea is that if the set R' is being removed, the rank of a vertex v in G will be close to the rank of v in $G - R'$.

3.3.3 *The link aggregate algorithm*

The IS algorithm performs well at discovering communities given a good initial guess, for example, when its initial "guesses" are the outputs of another clustering algorithm such as RaRe as opposed to random edges in

```
procedure RaRe(G, W)
global R ← ∅;
{H_i} are connected components in G;
for all H_i do
  ClusterComponent(H_i);
end for
Initial clusters {C_i} are cluster cores;
for all v ∈ R do
  for all Clusters C_i do
    Add v to cluster C_i if v is adjacent
    to C_i or W(v ∪ C_i) > W(C_i);
  end for
end for
```

```
procedure ClusterComponent(H)
if |V(H)| > max then
  {v_i} are t highest rank nodes in H;
  R ← R ∪ {v_i}; H ← H \ {v_i};
  {F_i} are connected components in H;
  for all F_i do
    ClusterComponent(F_i);
  end for
else if min ≤ |V(H)| ≤ max then
  mark H as a cluster core;
end if
```

```
procedure LA(G, W)
C ← ∅;
Order the vertices v_1, v_2, ..., v_|V|;
for i = 1 to |V| do
  added ← false;
  for all D_j ∈ C do
    if W(D_j ∪ v_i) > W(D_j) then
      D_j ← D_j ∪ v_i; added ← true;
    end if
  end for
  if added = false then
    C ← C ∪ {{v_i}};
  end if
end for
return C;
```

```
procedure IS(seed, G, W)
C ← seed; w ← W(C);
increased ← true;
while increased do
  if G is dense then
    N ← All nodes adjacent to C;
  else
    N ← All nodes in G;
  end if
  for all v ∈ N do
    if v ∈ C then
      C' ← C \ {v};
    else
      C' ← C ∪ {v};
    end if
    if W(C') > W(C) then
      C ← C';
    end if
  end for
  if W(C) = w then
    increased ← false;
  else
    w ← W(C);
  end if
end while
return C;
```

Fig. 7. Algorithms Rank Removal (RaRe), Link Aggregate (LA), and Iterative Scan (IS).

the communication network. We discuss a different, efficient initialization algorithm here.

RaRe begins by ranking all nodes according to some criterion, such as page rank (Page et al., 1998). Highly ranked nodes are then removed in groups until small connected components are formed (called the cluster cores). These cores are then expanded by adding each removed node to any cluster whose density is improved by adding it.

Table 1
User specified inputs for algorithm RaRe

Input	Description
W	Weight function
ϕ	Ranking function
min, *max*	Minimum and maximum core sizes
t	Number of high-ranking vertices to be removed

While this approach was successful in discovering clusters, its main disadvantage was its inefficiency. This was due in part to the fact that the ranks and connected components need to be recomputed each time a portion of the nodes is removed. The runtime of RaRe is significantly improved when the ranks are computed only once. For the remainder of this paper, RaRe refers to the RaRe algorithm with this improvement, unless otherwise stated.

Since the clusters are to be refined by IS, the seed algorithm needs only to find approximate clusters. The IS algorithm will "clean up" the clusters. With this in mind, the new seed algorithm LA focuses on efficiency, while still capturing good initial clusters. The pseudocode is given in Fig. 7. The nodes are ordered according to some criterion, for example, decreasing page rank, and then processed sequentially according to this ordering. A node is added to any cluster if adding it improves the cluster density. If the node is not added to any cluster, it creates a new cluster. Note that every node is in at least one cluster. Clusters that are too small to be relevant to the particular application can now be dropped. The runtime may be bounded in terms of the number of output clusters C as follows:

Theorem 5. *The runtime of LA is* $O(|C||E| + |V|)$.

Proof. Let C_i be the set of clusters just before the ith iteration of the loop. The time it takes for the ith iteration is $\mathrm{O}(|C_i|\deg(v_i))$, where $\deg(v_i)$ is the number of edges adjacent to v_i. Each edge adjacent to v_i must be put into two classes for every cluster in C_i: the other endpoint of the edge is either in the cluster or outside it. With this information, the density of the cluster with v_i added may be computed quickly ($O(1)$) and compared to the current density. If $\deg(v_i)$ is zero, the iteration takes $O(1)$ time. Therefore, the total runtime is asymptotically on the order of

$$\sum_{\deg(v_i)>0} |C_i|\deg(v_i) + \sum_{deg(v_i)=0} 1 \leq \sum_{i=1}^{|V|} |C_i|\deg(v_i) + \sum_{i=1}^{|V|} 1$$

$$\leq \sum_{i=1}^{|V|} |C|deg(v_i) + |V| = 2|C||E| + |V| = O(|C||E| + |V|).$$

□

3.3.4 Iterative scan (IS)

Algorithm IS explicitly constructs a cluster that is a local maximum with respect to a density metric by starting at a "seed" candidate cluster and updating it by adding or deleting one vertex at a time as long as the metric strictly improves. The algorithm stops when no further improvement can be obtained with a single change. This algorithm is given in pseudocode format in Fig. 7. Different local maxima can be obtained by restarting the algorithm at a different seed, or changing the order in which vertices are examined for cluster updating. The algorithm terminates if the addition to C or deletion from C of a single vertex does not increase the weight. During the course of the algorithm, the cluster C follows some sequence, C_1, C_2, …, with the property that $W(C_1) < W(C_2) < \cdots$, where all the inequalities are strict. Since the number of possible clusters is finite, the algorithm must terminate when started on *any* seed, and the cluster output will be a locally optimal cluster.

The cluster size may be enforced heuristically by incorporating this criterion into the weight function. This is done by adding a penalty for clusters with size outside the desired range. Such an approach will not impose hard boundaries on the cluster size. If the desired range is $[C_{\min}, C_{\max}]$, then a simple penalty function $\mathrm{Pen}(C)$ that linearly penalizes deviations from this range is

$$\mathrm{Pen}(C) = \max\left\{0, h_1 \cdot \frac{C_{\min} - |C|}{C_{\min} - 1}, h_2 \cdot \frac{|C| - C_{\max}}{|V| - C_{\max}}\right\},$$

where $C_{\min}$, $C_{\max}$, h_1, and h_2 are user specified parameters. All the user specified inputs are summarized in Table 2.

We emphasize that algorithm IS can be used to improve any seed cluster to a locally optimal one. Instead of building clusters from random edges as a starting point, we can refine clusters, that are output by some other algorithm—these input clusters might be good "starting points," but they may not be locally optimal. IS then refines them to a set of locally optimal clusters.

The original process for IS consisted of iterating through the entire list of nodes over and over until the cluster density cannot be improved. In order

Table 2
User specified inputs to algorithm IS

Parameter	Description
W	Weight function
δ	Set-difference function ($\delta = \delta_{\mathrm{h}}$ in our implementation)
ε	Size of set neighborhood ($\varepsilon = 1$ in our implementation)
max_fail	Number of unsuccessful restarts to satisfy stopping condition
$[C_{\min}, C_{\max}]$	Desired range for cluster size
h_1, h_2	Penalty for a cluster of size 1 and $\lvert V\rvert$

to decrease the runtime of IS, we make the following observation. The only nodes capable of increasing the cluster's density are the members of the cluster itself (which could be removed) or members of the cluster's immediate neighborhood, defined by those nodes connected to a node inside the cluster. Thus, rather than visiting each node on every iteration, we may skip over all nodes except for those belonging to one of these two groups. If the neighborhood of a cluster is much smaller than the entire graph, this could significantly improve the runtime of the algorithm.

This algorithm provides both a potential decrease and an increase in runtime. A decrease occurs when the cluster and its neighborhood are small compared to the number of nodes in the graph. This is the likely case in a sparse graph. In this case, building the neighborhood set N takes a relatively short time compared to the time savings of skipping nodes outside the neighborhood. An increase in runtime may occur when the cluster neighborhood is large. Here, finding the neighborhood is expensive, plus the time savings could be small since few nodes are absent from N. A large cluster in a dense graph could have this property. In this case, placing all nodes in N is preferable.

Taking into account the density of the graph, we may construct N in either of the two methods described here, in order to maximize efficiency in all cases. If the graph is dense, all nodes are placed in N, but if the graph is sparse, the algorithm computes N as the neighborhood of the cluster.

In the experiments that follow, the behavior of IS for sparse graphs is denoted IS, and the behavior for dense graphs is denoted IS^2.

3.4 Experiments and results

A series of experiments were run in order to compare both the runtime and the performance of the new algorithm with its predecessor. In all cases, a seed algorithm was run to obtain initial clusters, and then a refinement algorithm was run to obtain the final clusters. The baseline was the seed algorithm RaRe followed by IS. The proposed improvement consists of the seed algorithm LA followed by IS^2. The algorithms were first run on a series of random graphs with average degrees 5, 10, and 15, where the number of nodes ranges from 1000 to 45,000. In this simple model, all pairs of communication are equally likely.

All the algorithms take as input a density metric W, and attempt to optimize that metric. In these experiments, the density metric was chosen as W_{ad}, called the *average degree*, which is defined for a set of nodes C as

$$W_{\text{ad}}(C) = \frac{2|E(C)|}{|C|},$$

where $E(C)$ is the set of edges with both endpoints in C.

The runtime for the algorithms is presented in Fig. 8. The new algorithm remains quadratic, but both the seed algorithm and the refinement

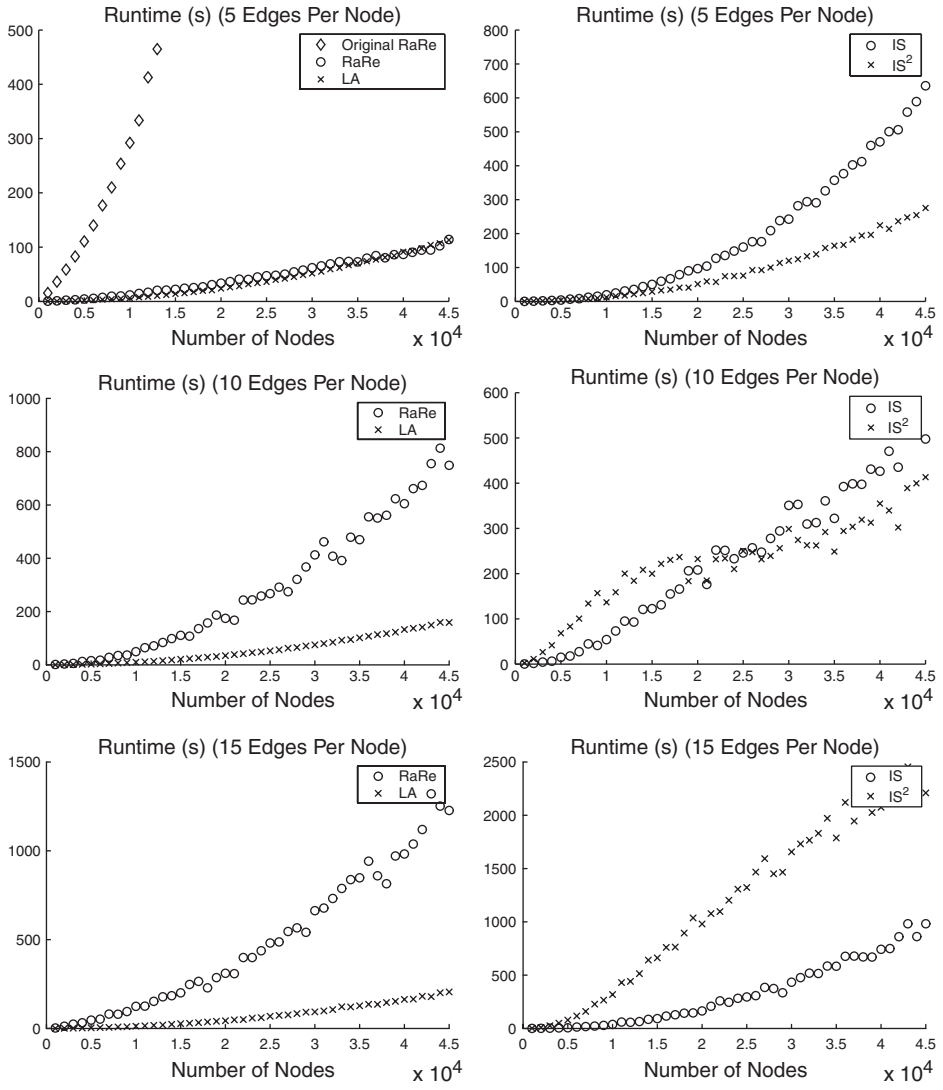


Fig. 8. Runtime of the previous algorithm procedures (RaRe and IS) compared to the current procedures (LA and IS^2) with increasing edge density. On the left is a comparison of the initialization procedures RaRe and LA, where LA improves as the edge density increases. On the right is a comparison of the refinement procedures IS and IS^2. As expected, IS^2 results in a decreased runtime for sparse graphs, but its benefits decrease as the number of edges becomes large.

algorithm runtimes are improved significantly for sparse graphs. In the upper left plot in Fig. 8, the original version of RaRe is also plotted, which recalculates the node ranks a number of times, instead of pre-computing the ranks a single time. LA is 35 times faster than the original RaRe algorithm

and IS^2 is about twice as fast as IS for graphs with 5 edges per node. The plots on the right demonstrate the tradeoff in IS^2 between the time spent computing the cluster neighborhood and the time saved by not needing to examine every node. It appears that the tradeoff is balanced at $\sim$10 edges per node. For graphs that are more dense, the original IS algorithm runs faster, but for less dense graphs, IS^2 is preferable.

Figure 9 shows that the quadratic nature of the algorithm is based on the number of clusters found. When the runtime per cluster found is plotted, the resulting curves are linear.

Runtime is not the only consideration when examining this new algorithm. It is also important that the quality of the clustering is not hindered by these runtime improvements. Figure 10 compares the average density of the clusters found for both the old and the improved algorithms. A higher

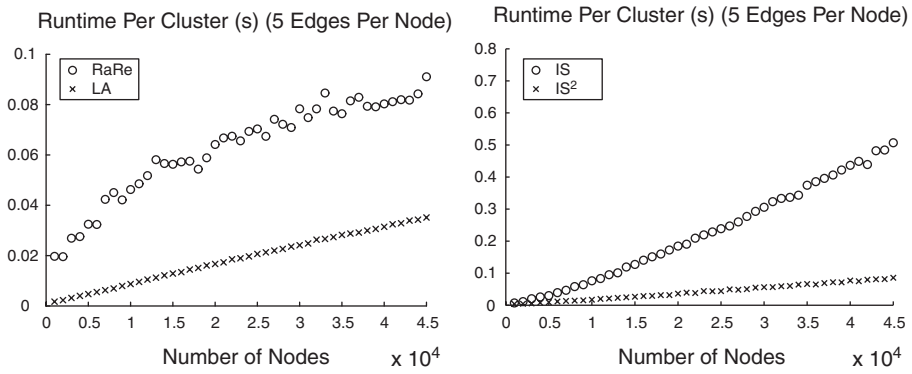


Fig. 9. Runtime per cluster of the previous algorithm (RaRe followed by IS) and the current algorithms (LA followed by IS^2). These plots show the algorithms are linear for each cluster found.

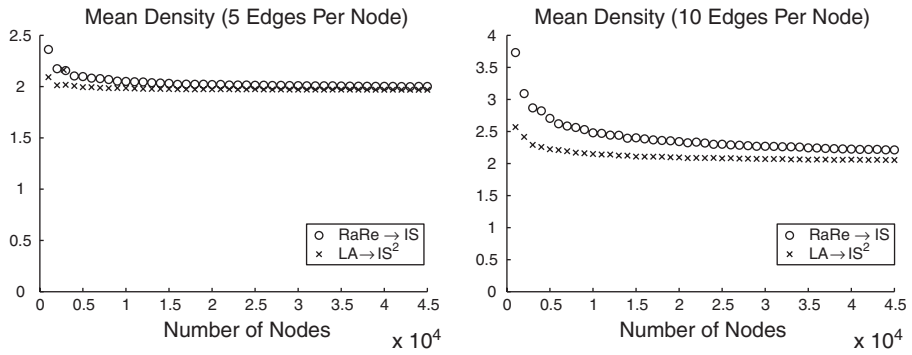


Fig. 10. Performance (average density) of the algorithm compared to the previous algorithm.

average density indicates a clustering of higher quality. Especially for sparse graphs, the average density is approximately equal in the old and new algorithms, although the older algorithms do show a slightly higher quality in these random graph cases.

Another graph model more relevant to communication networks is the preferential attachment model. This model simulates a network growing in a natural way. Nodes are added one at a time, linking to other nodes in proportion to the degree of the nodes. Therefore, popular nodes get more attention (edges), which is a common phenomenon on the web and in other real-world networks. The resulting graph has many edges concentrated on a few nodes. The algorithms were run on graphs using this model with 5 links per node, and the number of nodes ranging from 2000 to 16,000. Figure 11 demonstrates a surprising change in the algorithm RaRe when run on this type of graph. RaRe removes high-ranking nodes, which correspond to the few nodes with very large degree. When these nodes are added back into clusters, they tend to be adjacent to most clusters, and it takes a considerable amount of time to iterate through all edges to determine which connect to a given cluster. The algorithm LA, on the other hand, starts by considering high-ranking nodes before many clusters have formed, saving a large amount of time. The plot on the right of Fig. 11 shows that the quality of the clusters is not compromised by using the significantly faster new algorithm LA→IS^2.

Figure 12 confirms that constructing the clusters in order of a ranking such as page rank yields better results than a random ordering. LA performs better in terms of both runtime and quality. This is a surprising result since the random ordering is obtained much more quickly than the ranking process. However, the first nodes in a random ordering are not likely to be well connected. This will cause many single-node clusters to be formed in the early stages of LA. When high degree nodes are examined, there are many clusters to check whether adding the node will increase the cluster

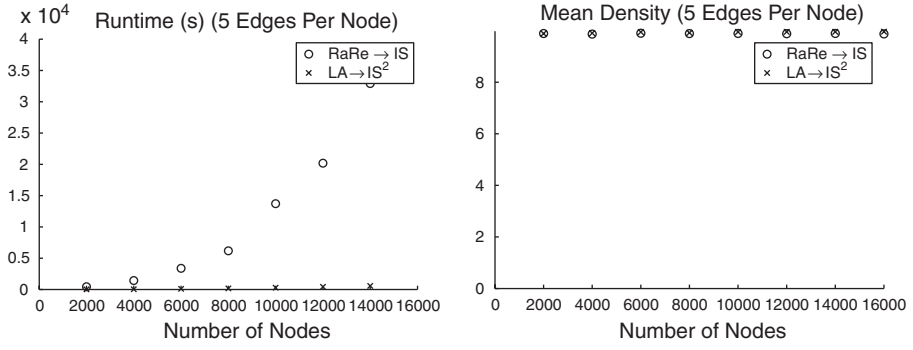


Fig. 11. Runtime and performance of the previous algorithm (RaRe followed by IS) and the current algorithm (LA followed by IS^2) for preferential attachment graphs.

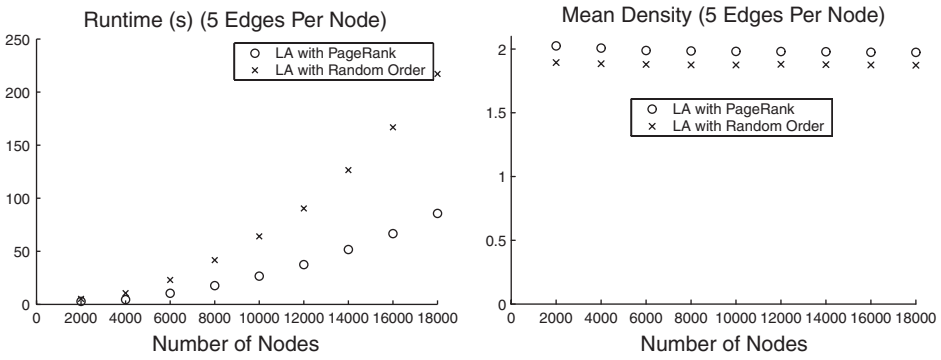


Fig. 12. Runtime and performance of LA with two different ordering types.

Table 3
Algorithm performance on real-world graphs

Algorithm	E-mail	Web
RaRe→IS	1.96 (234, 9); 148	6.10 (5, 8); 0.14
LA→IS^2	2.94 (19, 25); 305	5.41 (6, 19); 0.24
	Newsgroup	Fortune 500
RaRe→IS	12.39 (5, 33); 213	2.30 (104, 23); 4.8
LA→IS^2	17.94 (6, 40); 28	2.37 (288, 27); 4.4

Note: The first entry in each cell is the average value of W_{ad}. The two entries in parentheses are the average number of clusters found and the average number of nodes per cluster. The fourth entry is the runtime of the algorithm in seconds. The e-mail graph represents e-mails among the RPI community on a single day (16,355 nodes). The web graph is a network representing the domain www.cs.rpi.edu/~magdon (701 nodes). In the newsgroup graph, edges represent responses to posts on the alt.conspiracy newsgroup (4526 nodes). The Fortune 500 graph is the network connecting companies to members of their board of directors (4262 nodes).

density. This is time consuming. If the nodes are ranked, the high degree nodes will be examined first, when few clusters have been created. These few clusters are likely to attract many nodes without starting a number of new clusters, resulting in the algorithm completing more quickly.

The algorithms were also tested on real-world data. The results are shown in Table 3. For all cases other than the web graph, the new algorithm produced a clustering of higher quality.

4 Conclusion

In this paper, we described methods for discovering hidden groups based only on communication data, without the use of communication contents. The algorithms rely on the fact that such groups display correlations in

their communication patterns (temporal or spatial). We refer to such groups as hidden because they have not declared themselves as a social entity. Because our algorithms detect hidden groups without analyzing the contents of the messages, they can be viewed as an additional, separate toolkit, different from approaches that are based on interpreting the meaning of the messages. Our algorithms extract structure in the communication network formed by the log of messages; the output groups can further be studied in more detail by an analyst who might take into account the particular form and content of each communication, to get a better overall result. The main advantage is that our algorithms greatly reduce the search space of groups that the analyst will have to look at.

The spatial and temporal correlation algorithms target different types of hidden groups. The temporal hidden group algorithms identify those groups which communicate periodically and are engaged in planning an activity. Our algorithms have been shown to be effective at correctly identifying hidden groups artificially embedded into the background of random communications. Experiments show that as the background communications become more dense, it takes longer to discover the hidden group. A phase transition occurs if the background gets too dense, and the hidden group becomes impossible to discover. However, as the hidden group becomes more structured, the group is easier to detect. In particular, if a hidden group is secretive (non-trusting), and communicates key information only among its members, then the group is actually more readily detectable.

Our approach to the discovery of spatial correlation in communications data is based on the observation that social groups often overlap. This fact rules out the traditional techniques of partitioning and calls for novel procedures for clustering actors into overlapping groups. The families of clustering algorithms described here are able to discover a wide variety of group types based on the clustering metric provided. The algorithms have been shown to be both efficient and accurate at discovering clusters and retaining meaningful overlap between clusters.

5 Questions for discussion

1. The temporal hidden group algorithms are designed based on the assumption that the group communicates during an interval of a fixed length Δ. However, the value of Δ is generally not known. Strategies for calculating Δ may include trying different values and comparing the results of applying the algorithms for every choice. However, this may lead to inefficient and perhaps not very accurate algorithms. Are there better strategies for determining Δ?
2. The algorithms for detecting hidden groups presented in this chapter assume that the communications exchanged between the members of

the group form a connected subgraph in the total communication graph. Are there other structural properties that do not depend on the contents of the communications and are typical for a set of actors that make them a group?

3. The clusters of a communication network are defined as locally optimal subsets with respect to a given definition of density of a set. Different definitions of the density function may lead to quite different sets of clusters. Are there general rules for an appropriate choice of the density function or functions?
4. The experiments with our algorithms for clustering networks show their tendency to produce relatively small clusters. It appears that the main reason for this lies in the seed-procedure (Link Aggregate and Rank removal), which tends to produce small seed clusters. We have also found that the quality of the group of seed clusters affects the quality of the group of final clusters. What might be good approaches to obtain large seed clusters that lead to good large clusters?
5. Our algorithms work from a hypothesis that the planning cycles of a temporal hidden group are disjoint. How would temporal hidden groups be handled whose planning cycles overlap? Such hidden groups are called streaming hidden groups in Baumes et al. (2006).

References

Albert, R., A. Barabasi (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, 47–97.

Baumes, J., M.K. Goldberg, M. Hayvanovich, M. Magdon-Ismail, W. Wallace (2006). Finding hidden groups in streaming communication data, in: *Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI)*, Springer, Berlin, *Lecture Notes in Computer Science* 3975, 201–212.

Baumes, J., M. Goldberg, M. Krishnamoorthy, M. Magdon-Ismail, N. Preston (2005a). Finding communities by clustering a graph into overlapping subgraphs, in: *Proceedings of IADIS Applied Computing*, Algarve, Portugal, pp. 97–104.

Baumes, J., M. Goldberg, M. Magdon-Ismail (2005b). Efficient identification of overlapping communities, in: *Intelligence and Security Informatics (ISI)*, Atlanta, GA, pp. 27–36.

Baumes, J., M. Goldberg, M. Magdon-Ismail, W. Wallace (2005c). On hidden groups in communication networks. Technical Report, TR 05-15, Computer Science Department, Rensselaer Polytechnic Institute.

Baumes, J., M. Goldberg, M. Magdon-Ismail, W. Wallace (2004). Discovering hidden groups in communication networks, in: *Intelligence and Security Informatics (ISI)*, Tuscon, AZ, pp. 378–389.

Berge, C. (1978). *Hypergraphs*. North-Holland, New York.

Bollobás, B. (2001). *Random Graphs* 2nd ed. Cambridge University Press, New York.

Capocci, A., V.D.P. Servedio, G. Caldarelli, F. Colaiori (2004). Detecting communities in large networks, in: *Workshop on Algorithms and Models for the Web-Graph (WAW)*, Rome, Italy, pp. 181–188.

Carley, K. and M. Prietula (eds.), (2001). Computational Organization Theory. Lawrence Erlbaum Associates, Hillsdale, NJ.

Carley, K., A. Wallace (2001). Computational organization theory: a new perspective, in: S. Gass, C. Harris (eds.), Encyclopedia of Operations Research and Management Science, Kluwer Academic Publishers, Norwell, MA.

Erdős, P., A. Rényi (1959). On random graphs. *Publicationes Mathematicae Debrecen* 6, 290–297.

Erdős, P., A. Rényi (1960). On the evolution of random graphs. *A Magyar Tudomanyos Akademia Matematikai es Fizikai Tudomanyok Osztalyanak Kozlemenyei* 5, 17–61.

Erdős, P., A. Rényi (1961). On the strength of connectedness of a random graph. *Acta Mathematica Academiae Scientiarum Hungaricae* 12, 261–267.

Erickson, B.H. (1981). Secret societies and social structure. *Social Forces* 60, 188–211.

Even, G., J. Naor, S. Rao, B. Schieber (1999). Fast approximate graph partitioning algorithms. *SIAM Journal on Computing* 28(6), 2187–2214.

Flake, G.W., R.E. Tarjan, K. Tsioutsiouliklis (2002). Clustering methods basen on minimum-cut trees. Technical Report 2002-06, NEC, Princeton, NJ.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry* 40, 35–41.

Gibson, D., J. Kleinberg, P. Raghavan (1998). Inferring web communities from link topology, in: *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*. Pittsburgh, PA.

Girvan, M., M.E.J. Newman (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7821–7826.

Goldberg, M., P. Horn, M. Magdon-Ismail, J. Riposo, D. Siebecker, W. Wallace, B. Yener (2003). Statistical modeling of social groups on communication networks, in: *First Conference of the North American Association for Computational Social and Organizational Science (NAACSOS)*, PA (electronic proceedings).

Hendrickson, B., R.W. Leland (1995). A multi-level algorithm for partitioning graphs, in: *Supercomputing*. San Diego, CA.

Janson, S., T. Luczak, A. Rucinski (2000). Random graphs, in: *Series in Discrete Mathematics and Optimization*. Wiley, New York.

Kannan, R., S. Vempala, A. Vetta (2004). On clusterings: good, bad, and spectral. *Journal of the ACM* 51(3), 497–515.

Karypis, G., V. Kumar (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific computing* 20(1), 359–392.

Karypis, G., V. Kumar (1999). Multilevel k-way partitioning scheme for irregular graphs. *Journal of Parallel and Disturbuted Computing* 48(1), 96–129.

Kernighan, B.W., S. Lin (1970). An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal* 49(2), 291–307.

Kheyfits, A. (March 17, 2003). Introduction to clustering algorithms: hierarchical clustering, in: *DIMACS Educational Module Series* 03-1.

Kleinberg, J.M., R. Kumar, P. Raghavan, S. Rajagopalan, A.S. Tomkins (1999). The web as a graph: measurements, models, and methods. *Lecture Notes in Computer Science* 1627, Springer, Berlin, pp. 1–17.

Krebs, V.E. (2002). Uncloaking terrorist networks. *First Monday* 7 (4).

Magdon-Ismail, M., M. Goldberg, W. Wallace, D. Siebecker (2003). Locating hidden groups in communication networks using hidden Markov models, in: *International Conference on Intelligence and Security Informatics (ISI)*, Tucson, AZ, pp. 126–137.

Mishra, N D., Ron, R. Swaminathan (2002). Large clusters of web pages, in: *Workshop on Algorithms and Models for the Web Graph (WAW)*. Vancouver, Canada.

Monge, P., N. Contractor (2002). *Theories of Communication Networks*. Oxford University Press, USA.

Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Reviews* 45(2), 167–256.

Page, L., S. Brin, R. Motwani, T. Winograd (1998). The pagerank citation ranking: bringing order to the web. Stanford Digital Libraries Working Paper.

Ronfeldt, D., Arquilla, J. (2001). Networks, netwars, and the fight for the future. *First Monday* 6 (10).

Sanil, A., D. Banks, K. Carley (1996). Models for evolving fixed node networks: model fitting and model testing. *Journal of Mathematical Sociology* 21(1–2), 173–196.

Siebecker, D. (2003). *A hidden Markov model for describing the statistical evolution of social groups over communication networks*. Master's thesis (advisor: Malik Magdon-Ismail), Rensselaer Polytechnic Institute, Troy, NY.
Skillicorn, D.B. (2004). Social network analysis via matrix decompositions: al Qaeda.
Stewart, T.A. (2001). Six degrees of Mohamed Atta, in: *Business 2.0*, Vol. 2 issue 10, p. 63.
West, D.B. (2001). *Introduction to Graph Theory*. Prentice Hall, Upper Saddle River, NJ, USA.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 10

Social Network Analysis for Terrorism Research

Edna Reid and Hsinchun Chen

Department of Library Science, Clarion University, 840 Wood Street, Clarion, PA 16214, USA

Jennifer Xu

Department of Computer Information Systems, Bentley College, 175 Forest Street, Waltham, MA 02452, USA

Abstract

This chapter presents evidence-based approaches for analyzing and visualizing the relations and network structures of social movement organizations such as terrorist/extremist groups. There is a dearth of research in this area because of problems with collecting and accessing reliable data and the lack of advanced methodologies in the field. To address these problems, we present two case studies in which we used a range of advanced social network methodologies to analyze the Jihad terrorist groups' networks in the real world and cyberspace.

For the initial case study, we employed social network analysis (SNA) techniques to analyze the patterns of relations between and among individuals, groups, and organizations of the Global Salafi Jihad (GSJ) network. We also introduced the Web structural mining technique (which uses hyperlink analysis) into the terrorist network analysis field, which to the best of our knowledge, has never been used in this domain.

To better understand the terrorist groups' Web network structures, cluster affinity, and relations in cyberspace, the second case study analyzes their Web hyperlink structures and content. The results from both case studies were reviewed and validated by domain experts. They provide empirical insights and validation of the application of advanced terrorist network methodologies for the security informatics, social network, and law enforcement communities. Research communities stand to gain from the lessons learned from these case studies because the methodologies we employed can be applied to other social movement organizations and other types of data.

1 Introduction

Since terrorism has far reaching economic, political, psychological, and social impacts, a thorough understanding of the terrorism phenomena from different disciplinary perspectives is warranted if we are to succeed in a collective quest to analyze acts of global extremism. Because terrorist organizations often operate in network forms in which individual perpetrators cooperate with each other and exploit information technology to plan and implement their attacks (Arquilla and Ronfeldt, 2001), we could gain valuable knowledge about the terrorist organizations by studying various structural properties of terrorist network operations in the real world and cyberspace. Such knowledge may help researchers understand the root causes, agenda setting, information operations of groups (Earl and Emery, 2003), and the emergence of new types of ''virtual'' social movements (Diani, 1999). It can also provide an overview of the ways that relations are formed, maintained, and changed between social movement actors. However, key challenges in conducting systematic research on terrorist/extremist groups are the difficulty in collecting and accessing comprehensive and reliable data, and the lack of advanced methodologies (Qin et al., 2005; Sageman, 2004).

1.1 Case study 1: SNA of the GSJ network

Although the terrorism-related research domain has experienced tremendous growth since September 11th (Zahn and Strom, 2004), more studies of terrorist network structures are needed to answer pressing research questions: Do terrorist networks share the same topological properties with other types of networks such as ordinary organization networks and social networks? Do they follow the same organizing principle? How do they achieve efficiency under constant surveillance and threats from authorities?

To answer these questions, we report in this chapter a case study of the structure of a very large global terrorist network, the Global Salafi Jihad (GSJ) network, using methods and techniques from several relevant areas such as social network analysis (SNA) and Web structural mining (Qin et al., 2005). We consider this case study as unique and beneficial from three different perspectives. First, unlike most previous studies which used unreliable data sources such as news stories and media-generated incident databases, our study was based on a reliable dataset of sociological data (e.g., geographical origins, occupations, kinships) on 366 members of the GSJ network compiled by Sageman (2004) from court transcripts and legal proceedings, corroborated information from people with direct access to the information provided, scholarly articles, the press, and Internet articles.

Second, our first case study introduced advanced network analysis methodologies (e.g., small-world and scale-free models) into the study of terrorist networks. For example, we conducted in-depth SNA (e.g., centrality measures and blockmodeling) on specific terrorism attacks to better understand how terrorist groups planned and coordinated the 9/11 attacks (National Commission on Terrorist Attacks upon the United States, 2004). In order to discern the communication patterns in the GSJ network, such as members who have stronger social influences or higher social status than the others, we adopted the hyperlink analysis methodology from the Web structural mining area. Third, our results provide empirical insights because they have some interesting similarities to other underground networks, such as drug dealers' networks, and can be used, in part or as a whole, by the research, law enforcement, and intelligence communities.

1.2 Case study 2: network analysis of the Dark Web

Since it is now widely recognized that many terrorist/extremist groups use the Internet to communicate, disseminate propaganda, and recruit/train their members (SITE Institute, 2003; Thomas, 2003; Weimann, 2004, 2006), our second case study focuses on the development of an integrated approach for harvesting Jihad terrorist groups' websites to construct a high quality collection of website artifacts that can be used for analyzing and visualizing how Jihad terrorist groups use the Internet, especially the Web, in their terror campaigns (Reid et al., 2005). Terrorist groups and their sympathizers' multi-lingual Web artifacts are considered to be the alternate side of the Internet and referred to as the Dark Web. Although the Dark Web has recently received government and media attention (Coll and Glasser, 2005; Noguchi and Goo, 2006; Weimann, 2004), our systematic understanding of how terrorists use the Internet for their web of terror is limited (Conway, 2005; ISTS, 2003).

Hyperlink and content analyses were used to analyze the nature of relationships and communication channels between the Jihad terrorist groups' Arabic language websites. Hyperlink analysis is also known as hyperlink network analysis (HNA) because hyperlinks between websites (or webpages) can be viewed as social and communicational ties so much so that the standard techniques from SNA can be applied to this new data source: hyperlinks (Park and Thelwall, 2003). With the extensive link structure of the Web, hyperlinks are highly promising because they can be mined for previously hidden patterns of information that can shed light on the network structures of groups of websites. As a result, hyperlink analysis was applied to the Jihad terrorist groups' websites and similarity measures between all pairs of websites were calculated to form a similarity matrix. The matrix serves as a tool for showing the relationships between various

hyperlinked communities and helps to foretell likely relationships between groups in the real world (Reid et al., 2005). The websites were further analyzed using content analysis to compare the content of the hyperlinked communities.

The second case answers the following research questions: What are the most appropriate techniques for collecting high-quality Jihad terrorism webpages? What are systematic approaches for analyzing and visualizing Jihad terrorist Web artifacts so as to identify usage, relations, and network structures among groups? The second case study is beneficial because it develops and validates systematic methodologies for collecting, analyzing, and visualizing terrorist groups' Web artifacts. The results provide an integrated approach to the study of the Jihad terrorist groups' Web structure.

The remainder of the chapter is organized as follows. Section 2 reviews related works in different domains in relation to terrorist network analysis and background information on the Jihad groups including the GSJ network. In Section 3 we present our methodologies and report our findings from the analysis. Section 4 concludes this chapter with discussions, implications, and future directions.

2 Related works

In this section, we review network analysis methodologies employed in other domains: SNA, statistical analysis of network topology, and Web hyperlink analysis. These techniques can be used to analyze terrorist networks. Different techniques reveal different perspectives of terrorist networks.

2.1 Social network analysis

SNA is used in sociology research to analyze patterns of relationships and interactions between social actors in order to discover an underlying social structure (Scott, 1991, 2001; Wasserman and Faust, 1994). It has recently been recognized as a promising technology for studying criminal organizations and enterprises (McAndrew, 1999; Sparrow, 1991; Xu and Chen, 2003). Studies involving evidence mapping in fraud, terrorism, and conspiracy cases have recently been added to this list (Bollobas, 1985; Carley et al., 2001; Krebs, 2001; Sageman, 2004).

In SNA studies, a network is usually represented as a graph, which contains a number of nodes (network members) connected by links (relationships). SNA can be used to identify key members and interaction patterns between sub-groups in terrorist networks. Several centrality measures can be used to identify key members who play important roles in a network. Freeman (1979) provided definitions of the three most popular centrality measures: degree, betweenness, and closeness.

Degree measures how active a particular node is. It is defined as the number of direct links a node a has:

$$C_D(a) = \sum_{i=1}^{n} c(i, a)$$

where n is the total number of nodes in a network and $c(i, a)$ a binary variable indicating whether a link exists between nodes i and a. A network member with a high degree could be the leader or "hub" in a network.

Betweenness measures the extent to which a particular node lies between other nodes in a network. The betweenness of a node a is defined as the number of geodesics (shortest paths between two nodes) passing through it:

$$C_B(a) = \sum_{i<}^{n} \sum_{j}^{n} g_{ij}(a)$$

where $g_{ij}(a)$ indicates whether the shortest path between two other nodes i and j passes through node a. A member with high betweenness may act as a gatekeeper or "broker" in a network for smooth communication or flow of goods (e.g., drugs).

Closeness is the sum of the length of geodesics between a particular node a and all the other nodes in a network. It actually measures how far away one node is from other nodes and sometimes it is called "farness" (Bollobas, 1985; Freeman, 1979; Gibson et al., 1998):

$$C_C(a) = \sum_{i=1}^{n} l(i, a)$$

where $l(i, a)$ is the length of the shortest path connecting nodes i and a.

Blockmodeling is another technique used in SNA to model interaction between clusters of network members. Blockmodeling could reduce a complex network to a simpler structure by summarizing individual interaction details into relationship patterns between positions (Watts and Strogatz, 1998). As a result, the overall structure of the network becomes more evident. A blockmodel of a network is thus constructed by comparing the density of the links between each pair of positions, d_{ij}, with d: a between-position interaction is present if $d_{ij} \geqslant d$, and absent otherwise. In this first case study, however, we used blockmodeling to extract interaction patterns between sub-groups rather than positions. A sub-group is defined as a cluster of nodes that have stronger and denser links within the group than with outside members.

2.2 *Statistical analysis of network topology*

Statistical topological analysis has been widely applied in capturing and modeling the key structural features of various real-world networks such as

scientific collaboration networks, the Internet, metabolic networks, etc. Three models have been employed to characterize these complex networks: random graph model (Bollobas, 1985; Erdos and Renyi, 1960), small-world model (Watts and Strogatz, 1998), and scale-free model (Barabasi and Albert, 1999). In random networks, two arbitrary nodes are connected with a probability p and as a result each node has roughly the same number of links. The degree distribution of a random graph follows the Poisson distribution (Bollobas, 1985), peaking at the average degree. A random network usually has a small average path length, which scales logarithmically with the size of the network so that an arbitrary node can reach any other node in a few steps. However, most complex systems are not random but are governed by certain organizing principles encoded in the topology of the networks (Albert and Barabasi, 2002). The small-world model and scale-free model are significantly deviant from the random graph model (Albert and Barabasi, 2002; Newman, 2003).

2.3 *Web structural analysis*

The Web, as a network of webpages connected by hyperlinks, bears some similarities with social networks because previous studies have shown that the link structure of the Web represents a considerable amount of latent human annotation (Gibson et al., 1998; Park and Thelwall, 2003). For example, when there is a direct link from page A to page B, it often means that the author of page A recommends page B because of its relevant contents. Moreover, similarly to *citation analysis* in which frequently cited articles are considered to be more important, webpages with more incoming links are often considered to be better than those with fewer incoming links. Co-citation is another concept borrowed from the citation analysis field that has been used in link-based analysis algorithms. Webpages are co-cited when they are linked to by the same set of parent webpages and heavily co-cited pages are often relevant to the same topic.

Researchers have developed many algorithms to judge the importance and quality of webpages using the criteria mentioned above. PageRank is one of the most popular algorithms. The PageRank algorithm is computed by weighting each incoming-link to a page proportionally to the quality of the page containing that incoming-link (Cho et al., 1998). The quality of these referring pages is also determined by PageRank. Thus, the PageRank of a page p is calculated recursively as follows:

$$\text{PageRank}(p) = 1 - d + d \times \sum_{\text{all } q \text{ link top}} \frac{\text{PageRank}(q)}{c(q)}$$

where d is a damping factor between 0 and 1 and $c(q)$ the number of outgoing links in q. PageRank was originally designed to calculate the importance of webpages based on the Web link structure and is used in the

commercial search engine Google to rank the search results. However, it can also be used to determine the importance of social actors in a proper social network where links imply similar ''recommendation'' or ''endorsement'' relationships as the hyperlinks in Web graph. We believe that PageRank can also be used to rank the importance of terrorists within a properly constructed GSJ terrorist network (Qin et al., 2005).

2.4 GSJ network

The GSJ is a worldwide religious revivalist movement with the objective of re-establishing past Muslim glory in a great Islamist state stretching from Morocco to the Philippines (Sageman, 2004). With Al Qaeda as its vanguard, the GSJ includes many terrorist groups who collaborate with members from different countries and forms a large global terrorist network. It has successfully planned and launched large-scale terrorist attacks such as the Strasbourg cathedral bombing in France in 2000, the 9/11 tragedy in 2001, and the Bali bombing in 2002 (Qin et al., 2005; Sageman, 2004).

Collecting data on the GSJ terrorist network presents many challenges (Krebs, 2001; Sageman, 2004), mostly because of a general lack of information. The GSJ data used in this study was collected by Sageman (2004) in his long-term empirical study on the GSJ members. In decreasing degrees of reliability, the information sources include transcripts of court proceedings involving GSJ terrorists and their organizations; followed by reports of court proceedings; then corroborated information from people with direct access to the information provided; uncorroborated statements from people with the access; and finally statements from people who had heard the information secondhand. In contrast, Krebs (2001) network analysis of the 19 hijackers associated with the September 11 terrorism attacks only relied on news articles from the major newspapers.

The final dataset consists of the profile information of 366 GSJ terrorists roughly divided into 4 clumps based on their geographical origins: central member, core Arab, Maghreb Arab, and Southeast Asian (Sageman, 2004). The central member clump mainly consists of the key Al Qaeda members who take the leading position in the whole GSJ network. The core Arab clump consists of GSJ terrorists from core Arabic countries such as Saudi Arabia and Egypt. The Maghreb Arab clump consists of GSJ terrorists from North African countries such as Morocco and Algeria. Finally, the Southeast Asian clump consists of terrorists from Jemaah Islamiyah centered in Indonesia and Malaysia.

The data for each of the 366 terrorists includes a set of sociological features (e.g., geographical origins, original socio-economic status, education, occupation, etc.) and individual psychological features (e.g., mental illness, personality, pathological narcissism, etc.) that could be the explanations for why these people became terrorists. More importantly, the data also captures all known relationships and interactions between these 366

GSJ terrorists. These relationships and interactions include personal relationships (e.g., acquaintance, friend, relative, and family member), religious relationships (following the same religious leader), operational interactions (participating in the same attacks), and other relationships.

2.5 Content and hyperlink analysis

With the current trends of Jihad terrorist groups using the Internet to support their terror campaigns, analyses have been expanded to include other datasets such as groups' websites and online discussion forums as well as Web mining methodologies (Chen et al., 2004; Kelley, 2002; SITE, 2003; Tsfati and Weimann, 2002; Wilson, 2002). Table 1 provides a summary of the research on Jihad terrorist groups and other Middle Eastern terrorist/extremist groups' use of the Web. The studies used terrorists' and their sympathizers' websites as their primary data sources and several methodologies such as observation, content analysis, and Web hyperlink analysis.

Because of the messiness of Web data, the need for systematic collection building methodologies, and extensive data cleaning procedures (Park and

Table 1
Summary of research on Middle Eastern extremist groups Internet usage

Methodologies	Findings
Observation	(a) As early as 1999, Denning (1999) found that terrorist groups use email, websites, chat rooms, and bulletin boards to plan operations and coordinate activities. (b) Knight and Ubayasiri's (2002) findings indicate that websites have become technically sophisticated with multimedia that encourages interaction through online polls and games. (c) Bunt (2003) analyzed Sunni and Shiite websites and found that the Internet forms part of a religious conceptual framework, incorporating sacred texts and the power to motivate one to act. (d) ISTS (2003) provided illustrations of five ways that the Internet has been used to support terrorist groups.
Content analysis	(a) Tzfati and Weimann (2002) identified patterns of groups' use of the Internet and provided detailed examples. (b) Weimann (2004) updated his content analysis study and found that most sites contained external links to other extremist websites and had about seven Internet usage patterns.
Content and hyperlink analysis	(a) Chen et al. (2004) found that analysis of the Middle Eastern groups' hyperlink structures provided insights into their inter-organizational Web networks. (b) Reid et al. (2005) provided further validation of how extremists groups are using the net.

Thelwall, 2003), there are few substantive research studies of terrorist groups' use of the Internet (Conway, 2005). In Table 1, the studies are mainly observation and content analysis with limited descriptions of their methodologies, samples, and validation procedures. Exceptions include studies of Chen et al. (2004) and Reid et al. (2005) conducted at the Artificial Intelligence Lab, University of Arizona. Both studies validate systematic approaches to harvest and analyze terrorist groups' website content.

Web harvesting is the process of gathering and organizing unstructured information from pages and data on the Web (Kay, 2004). Albertsen (2003) uses an interesting approach in the "Paradigma" project. The goal of Paradigma is to archive Norwegian legal deposit documents on the Web. It employs a Web crawler that discovers neighboring websites by following Web links found in the HTML pages of a starting set of webpages. Metadata is then extracted and used to rank the websites in terms of relevance.

For the second case study, we use a Web spider to discover new Jihad terrorist websites and use them as seeds to perform backlink searches (i.e., Google's backlink search tool). However, we do not use metadata but rely instead on judgment calls by human experts because there are so many fake Jihad terrorism websites. The "Political Communications Web Archiving" group also employs a semiautomatic approach to harvesting websites (Reilly et al., 2003). Domain experts provide seed URLs as well as topologies for constructing metadata that can be used in the spidering process.

After the webpages have been harvested, the extensive Web hyperlink structure can be mined for previously hidden patterns of information that can shed light on the network structures of the terrorist groups' websites. Borgman and Furner (2002) define two classes of Web hyperlink analysis studies, also known as HNA: relational and evaluative. Relational analysis gives insight into the strength of relations between Web entities, in particular websites, while evaluative analysis reveals the popularity or quality level of a Web entity. In this study, we are more concerned with relational analysis as it may bring us insight into the nature of relations between websites and, possibly, terrorist organizations.

To reach an understanding of the various facets of Jihad terrorism Web usage, a systematic analysis of the websites' contents is required. Researchers in the terrorism domain have traditionally used qualitative methodologies for analyzing terrorist groups' website data (ISTS, 2003; Tsfati and Weimann, 2002). We believe Jihad terrorism content on the Web falls under the category of communicative contents and a quantitative analysis is critical for a study to be objective.

The work of Demchak et al. (2001) focused on measuring "openness" of government websites using a Website Attribute System (WAS) tool that is basically composed of a set of high-level attributes such as transparency and interactivity. Each high-level attribute is associated with a second layer of attributes at a more refined level of granularity. This system is an

example of a well-structured and systematic content analysis exercise and provides guidance for the present study.

3 Results

3.1 Case study 1: SNA of the GSJ network

In this section, we describe our analytical procedures and report our findings. For the dataset of 366 GSJ terrorists, we calculated the "distance" between each pair of terrorists in the network based on the number of relationships between them and visualized the network using the multi-dimensional scaling (MDS) technique. Our visualization provides an intuitive and clear view of the overall GSJ network (see Fig. 1).

Figure 1 is the visualization of the GSJ network with all types of relations. Each node represents a terrorist/extremist. A link represents a social relation. The four terrorist clumps are color-coded: red for central member clump, yellow for core Arab clump, blue for Maghreb Arab, and green for Southeast Asian clump.

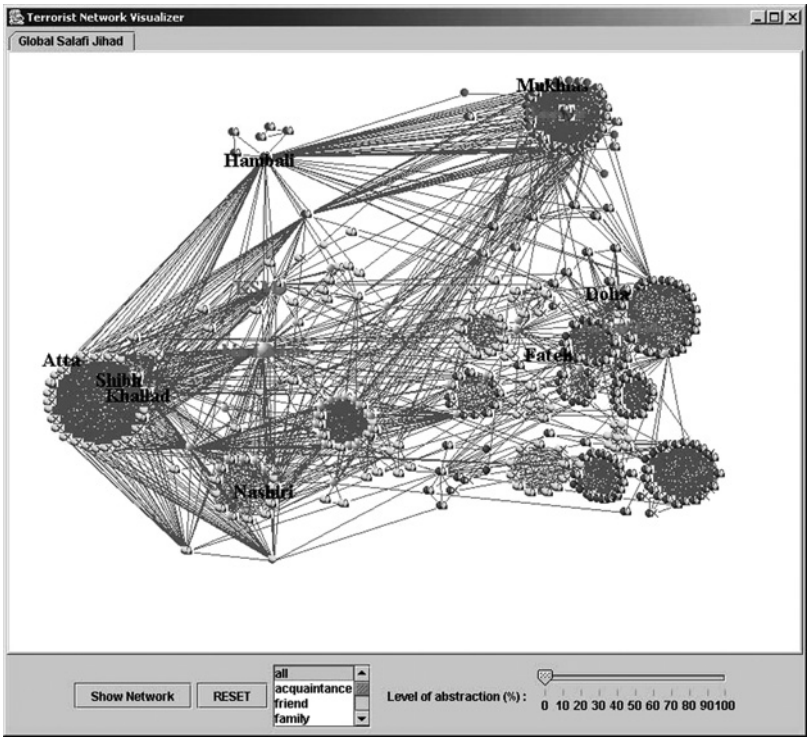


Fig. 1. Visualization of the full GSJ network.

To answer the research questions, our initial analysis conducted on the GSJ dataset was a SNA in which we used the centrality measures and blockmodeling to identify key members and sub-groups in the GSJ network. For each terrorist, three centrality measures were calculated: degree, betweenness, and closeness. Degree measure was used to identify the leaders of each clump in the GSJ network. High degrees indicate high-levels of activity and wide social influence, which means the members with high degrees are likely to be the leaders of their local networks. Gatekeepers, members with high betweenness, hold special interest for terrorist experts because gatekeepers are usually the contact between several terrorist groups and play important roles in coordinating terrorist attacks. The closeness measure was used differently from the previous two centrality measures. Instead of terrorists with high closeness, we identified those with low closeness whom are usually called outliers in SNA literatures. Outliers are of special interest because previous literature showed that, in illegal networks, outliers could be the true leaders. They appear to be outliers because they often direct the whole network from behind the scene, which prevents authorities from getting enough intelligence on them. Table 2 summarizes the top terrorists ranked by the three centrality measures in each of the four clumps.

After showing our SNA results to the domain experts, we confirmed that the key members identified by our algorithm matched the experts' knowledge on the terrorism organization. Members with high degree measures are also known by the experts as the leaders of the clumps in the real world. For example, Osama bin Laden, the leader of the central member clump, had 72 links to other terrorists and ranked the second in degree. Moreover, the experts mentioned that each clump has a Lieutenant who acts as an important connector between the clumps. For example, Zawahiri, Lieutenant of the central member clump, connects the central member clump and the core Arab clump together. Hambali, Lieutenant of the Southeast Asian clump, connects the Southeast Asian clump and the central member clump. These Lieutenants were also correctly identified by the algorithm for their high betweenness.

3.1.1 Analysis of the September 11th attack

To get a better understanding of how terrorists plan and coordinate attacks, we conducted in-depth SNA on a terrorist attack case. The September 11th terrorist attacks were selected. Centrality analysis and blockmodeling analysis were employed on the networks to identify key members who may have led or coordinated these attacks as well as to identify the connections between different attacks.

The September 11th incident was a series of coordinated attacks against the United States and thousands of innocent people were killed. Such a large-scale attack would require a high-level of planning and coordination to carry out. Based on the information in our dataset, we constructed the

Table 2
Terrorists with top centrality ranks within each clump

Ranking	Leader (degree)	Gatekeeper (betweenness)	Outlier (closeness)
Central member			
1	Zawahiri	bin Laden	Khalifah
2	Makkawi	Zawahiri	bin Laden
3	Islambuli	Khadr	Ghayth
4	bin Laden	Sirri	M. Atef
5	Attar	Zubaydah	Sheikh Omar
Core Arab			
1	Khallad	Harithi	Elbaneh
2	Shibh	Nashiri	Khadr4
3	Jarrah	Khallad	Janjalani
4	Atta	Johani	Dahab
5	Mihdhar	ZaMihd	Mehdi
Maghreb Arab			
1	Hambali	Baasyir	Siliwangi
2	Baasyir	Hambali	Fathi
3	Mukhlas	Gungun	Naharudin
4	Iqbal	Muhajir	Yunos2
5	Azahari	Setiono	Maidin
Southeast Asian			
1	Doha	Yarkas	Mujati
2	Benyaich2	Zaoui	Parlin
3	Fateh	Chaib	Mahdjoub
4	Chaib	DavidC	Zinedine
5	Benyaich1	Maaroufi	Ziyad

1-hop network of the September 11th attack which contained 161 members and covered nearly half of the whole GSJ network (see Fig. 2).

SNA identified Osama bin Laden (the yellow node) as the leader and gatekeeper of the September 11th attacks because he had the highest degree and betweenness in the 1-hop network. Furthermore, four major lieutenants (bin Laden, Zawahiri, Hambali, and KSM) who have the highest betweenness values among all GSJ members appeared in the September 11th network. They linked the 19 hijackers who directly participated in the attacks to all four clumps of the GSJ network, which indicates a worldwide cooperation in the planning of the attacks.

In the blockmodeling result of the September 11th network (see Fig. 3), the circle surrounding bin Laden contains all the hijackers. This result also confirmed that bin Laden was leading the attacks. Two lieutenants, Zawahiri and Hambali, connected the September 11th attack group to members from the Maghreb Arab and Southeast Asian clumps. Another lieutenant, KSM, served within the central member clump as the major planner of the attacks. Intelligence showed that KSM kept advocating the

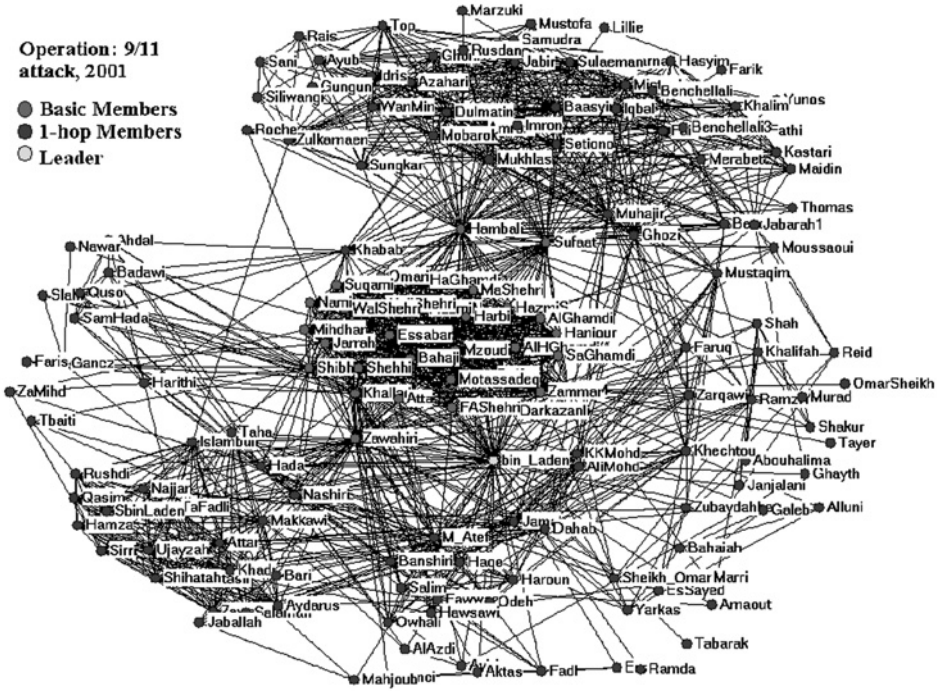


Fig. 2. The 1-hop network of the September 11th attacks.

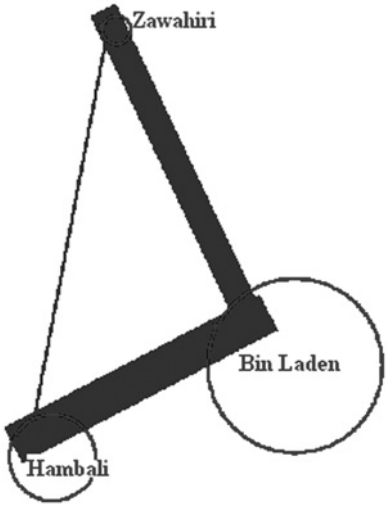


Fig. 3. Blockmodeling of the September 11th attacks network.

use of airliners as suicide weapons against specific targets. He later supervised bin al-Shibh, coordinating the September 11th operations. KSM was in overall control: bin al Shibh was the link between KSM and the field as well as the general coordinator of the operation.

3.1.2 Topological properties of the GSJ network

While SNA could provide important information about the individual members in the terrorist networks, we also need to study the overall topological properties of the GSJ network to understand how the terrorist organizations function. To address this problem, we conducted a statistical analysis on the GSJ network.

Several important statistical properties of the GSJ network were examined, including the link density, the average degree of the nodes, the degree distribution, etc. These properties then were checked against the small-world and scale-free models. Table 3 presents the small-world and scale-free properties of the GSJ network. The network contains a few small components and a single giant component. The giant component contains 356 or 97.3% of all members in the GSJ network. The separation between the 356 terrorists in the GSJ network and the remaining 10 terrorists is because no valid evidence has been found to connect the 10 terrorists to the giant component of the network. We focused only on the giant component in these networks and performed topology analysis. We found that this network is a small-world (see Table 3). The average path length and diameter (Wasserman and Faust, 1994) of the GSJ are small with respect to its size. Thus, a terrorist can connect with any other member in a network through just a few mediators. In addition, the GSJ network is quite sparse with a very low link density of 0.02. These two properties have important implications for the efficiency of the covert network function–transmission of goods and information. Because the risk of being detected by authorities increases as more people are involved, the small path length and link sparseness can help lower risks and enhance efficiency.

The other small-world topology, high clustering coefficient, is also present in the GSJ network (see Table 3). The clustering coefficient of the GSJ network is significantly higher than its random graph counterpart. Previous studies have also shown the evidence of groups and teams inside this kind of illegal network (Chen et al., 2004; Sageman, 2004; Xu and

Table 3
Small-world properties of the GSJ network

	GSJ network	Random graph
Average path length	4.20	3.23
Diameter	9	6.00
Clustering coefficient	0.55	0.2×10^{-1}

Chen, 2003). In these groups and teams, members tend to have denser and stronger relations with one another. The communication between group members becomes more efficient, making an attack easier to plan, organize, and execute (McAndrew, 1999).

Moreover, the GSJ network is also a scale-free system. The network follows an exponentially truncated power–law degree distribution (Newman, 2003), $P(k) \sim k^{-\gamma} e^{-(k/\kappa)}$, with exponent $\gamma = 0.67$ and cutoff $\kappa = 15.35$. Different from other types of networks (Albertsen, 2003; Qin et al., 2005; Wasserman and Faust, 1994) whose exponents usually are between 2.0 and 3.0, the exponent of the GSJ network is fairly small. The degree distribution decays much more slowly for small degrees than for that of other types of networks, indicating a higher frequency for small degrees (see Fig. 4). At the same time, the exponential cutoff implies that the distribution for large degrees decays faster than is expected for a power–law distribution, preventing the emergence of large hubs which have many links. Two possible reasons have been suggested that may attenuate the effect of growth and preferential attachment (Albert and Barabasi, 2002): (a) the aging effect: as time progresses some older nodes may stop receiving new links and (b) the cost effect: as maintaining links induces costs, there is a constraint on the maximum number of links a node can have. We believe that the cost effect does exist in the GSJ networks. Under constant threats from authorities, terrorists may avoid attaching to too many people, limiting the effects of preferential attachment. Another possible constraint on preferential attachment is trust (Krebs, 2001). This constraint is especially common in the

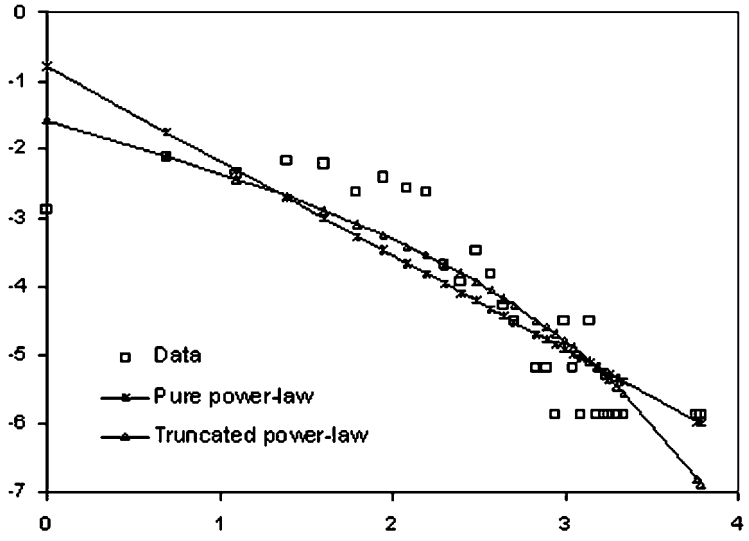


Fig. 4. Degree distribution of the GSJ network.

GSJ network where the terrorists preferred to attach to those who were their relatives, friends, or religious partners (Sageman, 2004).

3.1.3 *Social influences of members*

In a social network, not all the members play equal roles. Instead, some members may have stronger social influence or higher social status than the others. In a terrorist network context, a terrorist may act in a leading role and pass directions and orders to a group of terrorists who have lower status than him and at the same time he is also receiving directions and orders from someone who has higher status. Such unequalized social relationships between the terrorists may hold special interest for experts who study terrorist organization behavior. However, neither of our previous analyses allowed us to study the communication patterns in the GSJ network with such unequalized social relationships. To address this issue, we borrowed link analysis methodology from the Web structural mining area.

The core link algorithm we employed was the PageRank algorithm because it was used in previous studies to calculate the "importance" of authors within an authorship network. The link analysis we conducted on the GSJ network is described as follows. First, we used the PageRank algorithm reviewed in Section 2 to calculate a "social importance" score for each of the terrorists in the network. In this process, the PageRank algorithm will rank a terrorist higher if (1) he links to more other members in the network and (2) he links to other members with high importance scores in the network. Similarly to the degree measure, high importance scores given by the PageRank algorithm are also indications of leading roles in the terrorist network. However, the PageRank algorithm determines the importance of a specific member based on the structure of the whole network, while the degree measure makes the same judgment based only on very limited, local structural information.

After the importance scores for all the members in the GSJ network were calculated, for each member in the network the neighboring member with the highest importance score was identified. The assumption here is that the most important neighboring member for a terrorist may well be the local leader that the terrorist directly reports to. We then draw a directional link from each of the terrorists to their local leaders to visualize the terrorist social hierarchy. This graph is called an Authority Derivation Graph (ADG) (Toyoda and Kitsuregawa, 2001). Figure 5 shows the ADG of the GSJ network.

In the ADG, each node represents a terrorist in the GSJ network. A link pointing from terrorist A to terrorist B means that B has the highest rank among all members who have direct relationships with A and it is likely that, in their interaction, B acts in the role of "leader" and A acts in the role of "follower." The color of a node indicates which clump the member belongs to and the shape of a node indicates how many attacks the member has been involved in. The thickness of the links between nodes indicates the

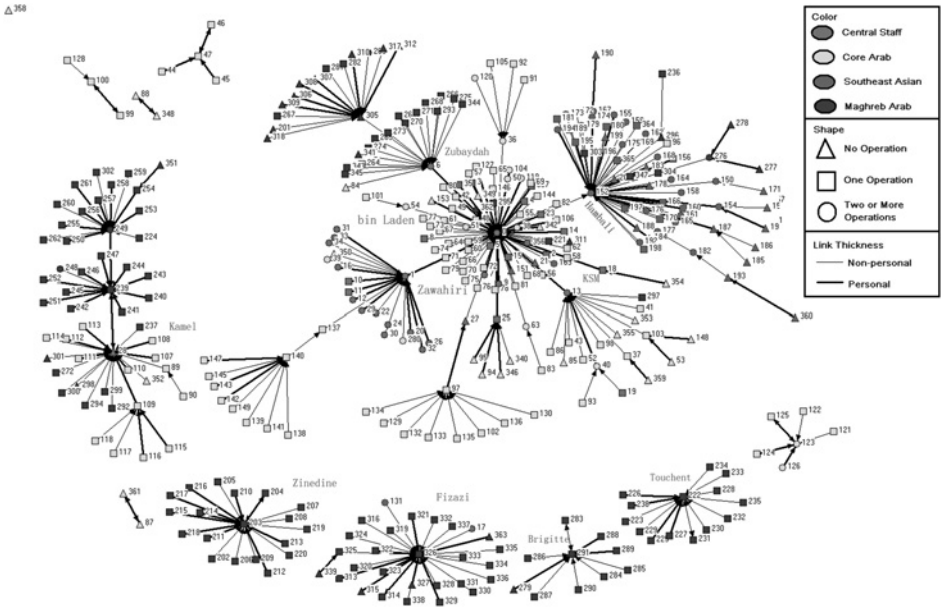


Fig. 5. ADG of the GSJ network.

type of relationship between the members. A thick link means there are personal relationships (e.g., kinship, family, friends, acquaintance) between two members while a thin link means there are only operational relationships (involved in the same attack) between two members.

The ADG of the GSJ terrorist network contains a large central component and several small and relatively autonomous components. The central component, consisting of key Al Qaeda members, has a more traditional hierarchy or ''corporate structure.'' We can clearly see that bin Laden has the highest status or, in other words, he is the leader of the whole GSJ network. Several major Lieutenants serve as the first-level underlings and the middle-person between bin Laden and key members of the other three clumps. More specifically, Hambali is the middle-person between bin Laden and the Southeast Asian clump; Zubaydah serves as the middle-person between bin Laden and the Maghreb Arab clump; Zawahiri connects bin Laden to the remainder of the central member clump; and KSM acts as the middle-person between bin Laden and the core Arab clump.

Except for the central component, the other components in the ADG have smaller size and shorter average shortest paths. This overall structure of the ADG suggests that the GSJ network may function as a ''holding company'' model, with Al Qaeda as the ''umbrella organization'' in charge of planning and many small independent groups as ''operating divisions.'' Such a model allows effective planning of attacks by having Al Qaeda as the ''master brain'' of the whole network and reduces the risk of being

disrupted by leaving the operations to the smaller groups that have minimum interactions with the central members.

Another interesting observation we made from the ADG is the difference in the link types between different types of members in the network. We found that 65% of the links between the leaders (members with incoming links) are personal links (acquaintances, friends, relatives, and family members), while only 38% of the links between the leaders and the followers (members with no incoming links) are personal links. Such differences in the link types between different members were also demonstrated in some other illegal networks such as drug dealer networks. The high percentage of personal relationships between the leaders forms the trustworthy ''backbone'' of the GSJ network and the low percentage of personal relationships between other members and the core members helps keep the network decentralized, covert, and less vulnerable.

3.2 Case study 2: network analysis of the Dark Web

In this second case study, we develop a systematic methodology for collecting the Jihad terrorism websites. The methodology ensures that our collection, which is the cornerstone of the study, is comprehensive and representative. To answer the research questions, we use a three step, systematic approach to construct the Dark Web collection:

(1) *Identify seed URLs of Jihad terrorism groups and perform backlink expansion*: we first identified a set of Jihad terrorist groups from the US Department of State's list of foreign terrorist organizations. Then, we manually searched major search engines (google.com, yahoo.com, etc.) using information such as the group names as queries to find their websites. Three Jihad terrorist websites were identified: www.qudsway.com of the Palestinian Islamic Jihad, www.hizbollah.com of Hizbollah, and www.ezzedine.netwhich is a website of the Izzedine-Al-Qassam, the military wing of Hamas. Then, we used Google's backlink search service to find all the websites that link to the three terrorist websites mentioned above and obtained a total of 88 websites.

(2) *Filtering the collection*: because bogus or unrelated terrorist sites can make their way into our collection, we have developed a robust filtering process based on evidence and clues from the websites. We constructed a short lexicon of Jihad terrorism with the help of Arabic language speakers. Examples of highly relevant keywords included in the lexicon are: "حرب صليبية" (''Crusader's War''), "المجاهدين" (''Moujahedin''), "الكفار" (''Infidels''), etc. The 88 websites were checked against the lexicon. Only those websites which explicitly identify themselves as the official sites of a terrorist organization and the websites that contain praise of or adopt ideologies espoused by a terrorist group

are included in our collection. After the filtering, 26 out of the 88 websites remained in our collection.

(3) *Extend the search manually*: to ensure the comprehensiveness of our collection we augment it by manually searching large search engines using the lexicon constructed in the previous step. The websites that are found are then filtered using the same rules used for filtering the backlink search results. As a result, 13 more websites were identified and our final Jihad collection contains 39 terrorist websites.

After identifying the Jihad terrorist websites, we download all the webpages within the identified sites. Our final collection contains more than 300,000 high-quality webpages created by Jihad terrorists.

3.2.1 Web communities

Next, we shed light on the infrastructure of Jihad websites. We believe the exploration of hidden Jihad communities over the Web can give insight into the nature of relationships and communication channels between the Jihad terrorist groups.

Uncovering hidden Web communities involves calculating a similarity measure between all pairs of websites. We define similarity to be a real-valued multi-variable function of the number of hyperlinks between website "A" and website "B." In addition, a hyperlink is weighted proportionally to how deep it appears in the website hierarchy. For instance, a hyperlink appearing at the homepage of a website is given a higher weight than hyperlinks appearing at a deeper level. We calculated the similarity between each pair of websites to form a similarity matrix. Then, this matrix was fed to a MDS algorithm which generated a two-dimensional graph of the website link structure. The proximity of nodes (websites) in the graph reflects the similarity level. Figure 6 shows the visualization of the Jihad website link structure.

Interestingly, domain experts recognized the existence of six clusters representing hyperlinked communities in the network. On the left side of the network resides the Hizbollah cluster. Hizbollah is a Lebanese militant organization. Established in 1982 during the Israeli invasion of Lebanon, the group routinely attacked Israeli military personnel until their pullout from south Lebanon in 2000. A cluster of websites of Palestinian organizations inhabits the bottom left corner of the network: Hamas, Al-Aqsa Martyr's Brigades, and the Palestinian Islamic Jihad. The Hizbollah community and the Palestinian militant groups' community were connected through hyperlink. Hizbollah has traditionally sympathized with and supported the Palestinian cause. Hence, it is not surprising at all to see a link between the two virtual communities.

On the top left corner sits the Hizb-ut-Tahrir cluster which is a political party with branches in many countries across the Middle East and Europe.

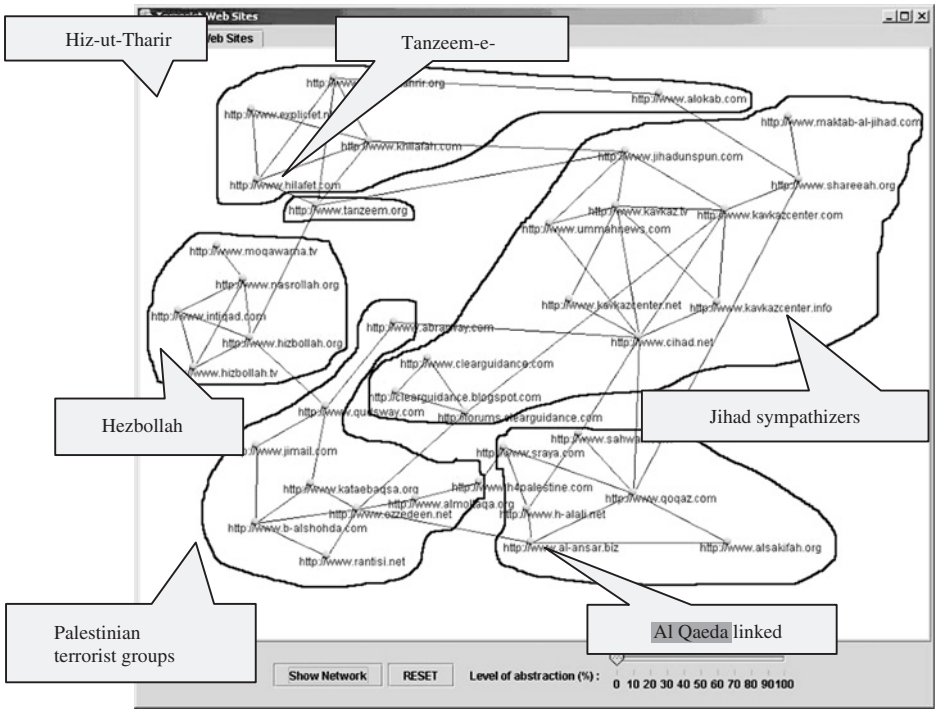


Fig. 6. The Jihad terrorism network with automatically generated hyperlinked communities.

Although groups in this cluster are not officially recognized as terrorist groups, they do have links pointing to the Hezbollah cluster.

Looking at the bottom right corner one can see a cluster of Al Qaeda affiliated sites. This cluster has links to two websites of the radical Palestinian group Hamas. Al Qaeda sympathizes with Palestinian groups. As well, some Palestinian Islamist groups like Hamas and Islamic Jihad share the same Salafi ideology with Al Qaeda. In the top right hand corner, the Jihad Sympathizers Web community gathers websites maintained by sympathizers of the Global Salafi movement. This community of Salafi sympathizers and supporters has links to three other major Sunni Web communities: the Al Qaeda community, Palestinian extremists, and Hizb-ut-Tahrir communities. As expected the sympathizers' community does not have any links to the Hezbollah community as they follow radically different ideologies.

Visualizing hyperlinked communities can lead to a better understanding of the underlying Jihad terrorism Web infrastructure. In addition, the visualization serves as a tool for showing the relationships between various hyperlinked communities. Furthermore, it helps foretell likely relationships between terrorist groups in the real world.

3.2.2 Usages of the web

To further analyze Jihad terrorism on the Web, we propose a framework for content analysis of the websites. The framework consists of high-level attributes, each of which is composed of multiple fine-grained low-level attributes (see Table 4). This approach is similar to what is presented in Demchak et al. (2001) study of ''openness'' of government websites. For this case study, the high-level attributes were identified from research on Middle Eastern terrorist groups (see Table 4) and consultations with domain experts. Conway (2005) confirms that there is considerable overlap amongst terrorist groups' usages of the Internet as identified by other authors. Middle Eastern terrorist/extremist groups' usages of the Internet are organized into high-level (e.g., communications) and low-level attributes (e.g., email, telephone, multimedia).

Currently we only consider the presence of an attribute in a website. In other words, the attribute for a given website is assigned a ''0'' if it does not appear in a website and a ''1'' if it does appear. However, this binary scheme does not capture the true contribution of the attributes. Hence, we assigned weights to each attribute such that the results reflect the content in a more realistic manner.

We asked our domain expert to go through each website in our collection and confirm the presence of low-level attributes. After completing the coding scheme for 39 websites in the collection, we then compared the content of the clusters or hyperlinked communities in the network. We aggregated data from all websites belonging to a cluster and displayed the results in snowflake diagrams (see Fig. 7).

The ''sharing ideology'' attribute seems to have the highest frequency of occurrences which indicates that more clusters use the Web for spreading their ideological messages. This supports assertion by Diani (1999) that the sharing of collective identification and solidarity can bond movement actors together and secure the persistence of their campaigns. In Fig. 7, an interesting observation in these snowflake diagrams is the ''propaganda towards insiders'' attribute. Militant groups, in this case Palestinian groups, tend to use the Web for disseminating their ideas in their own communities. They utilize propaganda as a tool for influencing youth and possibly recruiting new members.

Conversely, the Jihad supporters (sympathizers) try to explain their views to outsiders (e.g., Westerners, uncommitted) and justify terrorist actions. Terrorist groups use stratification strategies to influence different audiences and achieve their tactical goals which have been described by Weimann (2004, 2006) and Earl and Emery's (2003). According to Earl and Emery's, Al Qaeda has developed information strategies to influence four different audiences: opposing, uncommitted, sympathetic, and active. They use propaganda such as fatwas, media releases, multimedia, and websites to support this effort.

Table 4
Previous research and attributes used in the case study

Middle Eastern terrorist groups' uses of the net (research)	High-level attribute (uses of the net)	Low-level attribute (specific examples)
Arquilla and Ronfeldt (2001); Bunt (2003); Denning (1999); Earl and Emery (2003); ISTS (2003); SITE (2003); Thomas (2003); Tzfati and Weimann (2002); Weimann (2004, 2006)	Communications	Email
		Telephone
		Multimedia
		Online feedback form
		Documentation
	Fundraising	External aid mentioned
		Fund transfer
		Donation
		Charity
		Support groups
	Sharing ideology	Mission
		Doctrine
		Justification of the use of violence
		Pin-pointing enemies
	Propaganda (insiders)	Slogans
		Dates
		Martyrs
		Leaders
		Banners and seals
		Narratives of operations and events
	Propaganda (outsiders)	References to Western media coverage
		News reporting
	Virtual community	Listserv
		Text chat room
		Message board
		E-conferencing
		Web ring

4 Conclusion and discussion

In this chapter we present two case studies of the application of a range of SNA to the Jihad terrorist/extremist groups' networks in the real world and

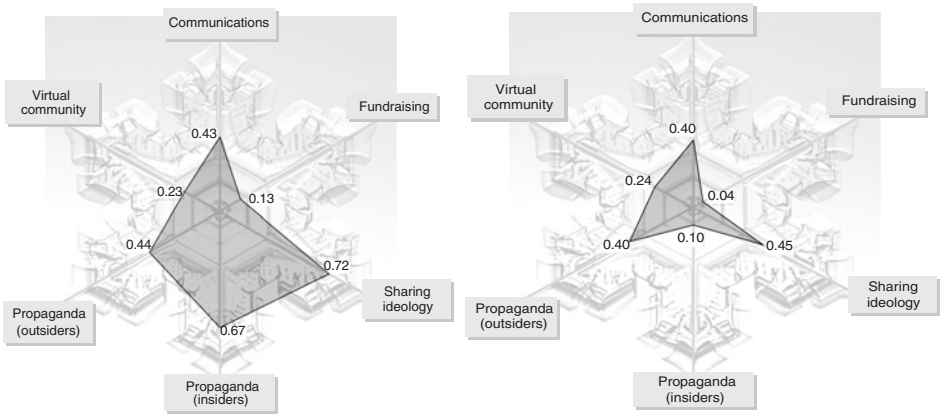


Fig. 7. Snowflake diagrams for Palestinian terrorist groups and Jihad supporters (sympathizers) Web communities.

cyberspace. Although this area has not been subjected to widespread empirical analysis, we needed to circumvent major challenges such as getting access to reliable data and lack of advanced methodologies in the field. Table 5 provides a summary of how we approached the challenges in the case studies.

For the first case, we employed several advanced network analysis techniques on a GSJ network dataset collected through a large-scale empirical study. Our results showed that centrality measures from the SNA field are effective tools to identify key members in a terrorist network. Our statistical analysis on the GSJ network revealed that the GSJ network is a small-world as well as a scale-free system. The results provide empirical insight because they have some interesting similarities to other underground networks such as drug dealers' networks. The topological features may help experts better understand how the GSJ network functions.

To the best of our knowledge, the Web structural mining (*aka* hyperlink analysis) methodology has never been applied in this domain. It was used to analyze the terrorist groups' organization structures under a social hierarchy assumption. This may provide insights into better understanding of terrorist/extremist organization behaviors, information operations, and recruitment strategies.

We have several future research directions to pursue. First, we are working with terrorism experts to fine tune our algorithms to increase the accuracy of our results. Second, we plan to extend the scope of our project to other types of illegal networks such as crime networks. Third, we want to add time-series analysis to get a more comprehensive understanding of the evolution and dynamics of terrorism networks.

For the second case study, we developed an integrated approach to the study of the Jihad terrorism Web Infrastructure. Hyperlinked communities'

Table 5
Summaries of the case studies

	SNA of the GSJ network (case study 1)	Network analysis of the Dark Web (Jihad terrorist groups on the Web) (case study 2)
Research question	Do terrorist/extremist networks share the same topological properties with other types of networks?	What are systematic approaches for analyzing and visualizing Jihad terrorist Web artifacts so as to identify network structures, relations, and usages?
Data	Sociological data about groups (e.g., kinship, educational background, etc.)	Sociological data about groups (cyberspace data from their websites such as messages, slogans, images)
Methodologies	Centrality measures (communities)	Similarity measures (communities)
	Hyperlink analysis (relationships and communication)	Hyperlink analysis (relationships and communication)
	Statistical analysis (topological properties)	Content analysis (Internet usages)
Findings	GSJ network is a small-world network similar to drug dealers' networks; high percentage of personal relationships between the leaders forms the trustworthy "backbone" of the GSJ network and the low percentage of personal relationships between other members and the core members helps keep the network decentralized, covert, and less vulnerable	Using systematic Web mining approaches, six clusters representing Web hyperlinked communities in the Jihad network were identified; some clusters used the net to share their propaganda with insiders while others focus on trying to influence people outside their active communities; with the increased use of Internet, the clusters can help foretell likely relationships among terrorist/extremist groups in the real world

analysis brings an overall view of the terror Web infrastructure. Visualizing hyperlinked communities facilitates the analysis of Web infrastructures. We then conducted a systematic content analysis of the websites and compared the content of various clusters. As part of our future work, we envisage implementing feature extraction algorithms for automatically detecting attributes in webpages. We believe that our methodology can be an effective tool for analyzing Jihad terrorism on the Web. Moreover, it can be easily extended to analyze other Web contents.

The Jihad websites collection as well as the methodologies will be integrated into a research portal called the Dark Web Portal. The Portal is being designed as a testbed that will eventually support searching, post-retrieval analysis (e.g., summarization, categorization, and visualization), content analysis, and hyperlink analysis of terrorist groups' multi-lingual and multimedia Web data. We also plan to implement feature extraction algorithms for automatically detecting attributes in webpages. The integrated approach and the Dark Web Portal can be effective tools for the security research and law enforcement communities.

The results from both case studies were reviewed and validated by domain experts. They provide empirical insights and validation of the application of advanced terrorist network methodologies for the security research and law enforcement communities. The social network, security, and information systems communities stand to gain from the lessons learned from these two case studies because the methodologies we employed can be applied to other social movement organizations and other types of Web data.

5 Questions for discussions

5.1 Case study 1: SNA of the Global Salafi Jihad (GSJ) network

1. What are some challenges in analyzing terrorist/extremist groups?
2. What are the topological properties of the GSJ terrorist network?
3. What kinds of relationships did the GSJ maintain?
4. How did they operate under constant surveillance?

5.2 Case study 2: network analysis of the Dark Web

1. In analyzing Jihad extremist/terrorist groups' Web structure, what types of research approaches are normally used?
2. How can researchers use the Web hyperlink structures to explore online networks (communities) of terrorist/extremist groups?
3. What are some barriers involved in harvesting a collection of terrorist/extremist groups' multi-lingual and multimedia websites? How can we overcome the barriers?
4. Describe the features of the integrated methodologies for collecting and analyzing terrorist/extremist groups' websites.
5. How can the methodologies be applied to other research domains?

Acknowledgments

This research has been supported in part by the following grant: NSF/ITR, "COPLINK Center for Intelligence and Security Informatics—A

Crime Data Mining Approach to Developing Border Safe Research," EIA-0326348, September 2003–August 2005.

We would like to thank Dr. Marc Sageman, Al Qaeda Expert, Professor Gabriel Weimann at the University of Haifa, Israel, Dr. Joshua Sinai from the Department of Homeland Security, and James Ellis from the Memorial Institute for the Prevention of Terrorism (MIPT) for their insightful comments and suggestions on our research. We would also like to thank all members of the Artificial Intelligence Lab at the University of Arizona who have contributed to the project, in particular Jialun Qin, Homa Atabakhsh, Cathy Larson, Chun-Ju Tseng, Ying Liu, Wei Xi, Charles Zhi-Kai Chen, Guanpi Lai, and Shing Ka Wu.

References

Albert, R., A.L. Barabasi (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* 74(1), 47–97.

Albertsen, K. (2003). Paradigma web harvesting environment. *Paper presented at the Third ECDL Workshop on Web Archives*, Trondheim, Norway.

Arquilla, J., D. Ronfeldt (2001). *Networks and Netwars: The Future of Terror, Crime, and Militancy*. Rand, California.

Barabasi, A.L., R. Albert (1999). Emergence of scaling in random networks. *Science* 286(5439), 509–512.

Bollobas, B. (1985). *Random Graphs*. Academic, London.

Borgman, C.L., J. Furner (2002). Scholarly communication and bibliometrics, in: B. Cronin (ed.), The Annual Review of Information Science and Technology (ARIST), ASIST, Washington, DC.

Bunt, G.R. (2003). *Islam in the Digital Age E-Jihad, Online Fatwas and Cyber Islamic Environments*. Pluto Press, London.

Carley, K.M., J.-S. Lee, D. Krackhardt (2001). Destabilizing networks. *Connections* 24(3), 31–34.

Chen, H., J. Qin, E. Reid, W. Chung, Y. Zhou, W. Xi (2004). Dark Web Portal: collecting and analyzing the presence of domestic and international terrorist groups on the web. *Paper presented at the IEEE Intelligence Transportation Conference*, Washington, DC.

Cho, J., H. Garcia-Molina, L. Page (1998). Efficient crawling through URL ordering. *Paper presented at the Proceedings of the Seventh International WWW Conference*, Brisbane, Australia.

Coll, S., S.B. Glasser (2005). Terrorists turn to the web as base of operations. *Washington Post*, p. A01.

Conway, M. (2005). Terrorist use of the internet and fighting back. Paper presented at the Conference Cybersafety: Safety and Security in a Networked World: Balancing Cyber-Rights and Responsibilities, Oxford University, UK.

Demchak, C.C., C. Friis, T.M. LaPorte (2001). Webbing governance: national differences in constructing the face of public organizations, in: G.D. Garson (ed.), Handbook of Public Information Systems, Marcel Dekker, New York City, NY.

Denning, D.E. (1999). Activism, hacktivism, and cyberterrorism: the internet as a tool for influencing foreign policy. *Paper presented at the Internet and International Systems: IT and American Foreign Policy Decisionmaking Workshop*, California.

Diani, M. (1999). Social movement networks virtual and real. Paper for the Conference "A New Politics? CCSS, University of Birmingham, 16-17, September 1999. Retrieved January 6, 2005, from http://www.nd.edu/~dmyers/cbsm/vol2/bgham99.pdf

Earl, R.S., Emery, N.E. (2003). Terrorist Approach to Information Operations. Master's Thesis, Naval Postgraduate Institute.

Erdos, P., A. Renyi (1960). On the evolution of random graphs. Publication of the Mathematical Institute of the Hungarian Academy of Sciences 5, 17–61.

Freeman, L.C. (1979). Centrality in social networks: conceptual clarification. *Social Networks* 1, 215–240.

Gibson, D., J. Kleinberg, P. Raghavan (1998). Inferring web communities from link typology. *Paper presented at the Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*, Pittsburgh, PA.

ISTS. (2003). Examining the cyber capabilities of Islamic terrorist groups. Institute for Security Technology Studies (ISTS), Dartmouth College, Hanover. Retrieved March 2005, http://www.ists.dartmouth.edu

Kay, R. (2004). Web harvesting. *Computerworld*. Retrieved March 2005, from http://www.computerworld.com

Kelley, J. (2002). Militants wire web with links to Jihad. *USA Today*. Retrieved March 10, 2005, from http://www.usatoday.com/news/world/2002/07/10/web-terror-cover.htm

Knight, A., K. Ubayasiri (2002). eTerror: journalism, terrorism and the internet. Retrieved February 2005, http://www.ejournalism.au.com/ejournalist/alkas.pdf

Krebs, V.E. (2001). Mapping network of terrorist cells. *Connections* 24(3), 43–52.

McAndrew, D. (1999). *Structural Analysis of Criminal Networks. The Social Psychology of Crime: Groups, Teams and Networks*. Aldershot, Dartmouth, MA.

National Commission on Terrorist Attacks upon the United States. (2004). 9-11 Commission Report. Washington, DC.

Newman, M.E.J. (2003). Structure and function of complex networks. *SIAM Review* 45(2), 167–256.

Noguchi, Y., S.K. Goo (2006). Terrorists' web chatter shows concern about internet privacy. *Washington Post*, p. A14.

Park, H.W., M. Thelwall (2003). Hyperlink analyses of the World Wide Web: a review. *Journal of Computer-Mediated Communication* 8(4). Retrieved March 10, 2005, from http://jcmc.indiana.edu/vol8/issue4/park.html

Qin, J., J.J. Xu, D. Hu, M. Sageman, H. Chen (2005). Analyzing terrorist networks: a case study of the Global Salafi Jihad network. *Paper presented at the IEEE International Conference on Intelligence and Security Informatics, ISI 2005*, Atlanta, GA.

Reid, E., J. Qin, Y. Zhou, G. Lai, M. Sageman, G. Weimann, H. Chen (2005). Collecting and analyzing the presence of terrorists on the web: a case study of Jihad websites. *Paper presented at the IEEE International Conference on Intelligence and Security Informatics, ISI 2005*, Atlanta, GA.

Reilly, B., G. Tuchel, J. Simon, C. Palaima, K. Norsworthy, L. Myrick (2003). Political communications web archiving: addressing typology and timing for selection, preservation and access. *Paper presented at the Third ECDL Workshop on Web Archives*, Trondheim, Norway.

Sageman, M. (2004). *Understanding Terror Networks*. University of Pennsylvania Press, Philadelphia.

Scott, J. (1991). *Social Network Analysis*. Sage, London.

Scott, M. (2001). War's new front. *CIO Insight* 82–83. http://www.cioinsight.com/article2/0,1397,37741,00.asp

SITE Institute. (2003). Report. Retrieved January 2005, http://www.siteinstitute.org/mission.html

Sparrow, M.K. (1991). Application of network analysis to criminal intelligence: an assessment of the prospects. *Social Networks* 13, 251–274.

Thomas, T.L. (2003). Al Qaeda and the internet: the danger of cyberplanning. *Parameters,* 112–123. Retrieved February 9, 2004, from http://www.carlisle.army.mil/usawc/Parameters/03spring/thomas.htm

Toyoda, M., M. Kitsuregawa (2001). Creating a web community chart for navigating related communities. *Paper presented at the Proceedings of ACM Conference on Hypertext and Hypermedia*, Arhus, Denmark.

Tzfati, Y., G. Weimann (2002). www.terrorism.com: terror on the internet. *Studies in Conflict and Terrorism* 25, 317–332.

Wasserman, S., K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Watts, D.J., S.H. Strogatz (1998). Collective dynamics of small-world networks. *Nature* 393, 440–442.

Weimann, G. (2004). www.terror.net: How Modern Terrorism Uses the Internet (Special Report). U.S. Institute of Peace, Washington, DC.

Weimann, G. (2006). *Terror on the Internet: the New Arena, the New Challenges*. U.S. Institute of Peace, Washington, DC.

Wilson, M. (2002). Considering the Net as an Intelligence Tool: Open-Source Intelligence. Decision Support Systems, Inc. Retrieved February 4, 2004, from http://www.metatempo.com/NetIntelligence.pdf

Xu, J., H. Chen (2003). Untangling criminal networks: a case study. Paper presented at the Proceedings of the First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI'03), Tucson, AZ.

Zahn, M.A., K.J. Strom (2004). Terrorism and the federal social science research agenda, in: M. Deflem (ed.), *Terrorism and Counter-Terrorism: Criminological Perspectives*, Vol. 5. Elsevier, Amsterdam, pp. 111–128.

Online resources

Conference on Safety and Security in a Networked World: Balancing Cyber-Rights & Responsibilities, Oxford Internet Institute, Sept. 8–10, 2005. Full text conference papers. www.oii.ox.ac.uk/research/cybersafety

Database of Terrorist Websites and eGroups, Anti-Terrorism Coalition (ATC), July 2005. http://news.atcoalition.net/

International Terrorism: Attributes of Terrorist Events (ITERATE) 1978–2002. http://www.columbia.edu/cgi-bin/eds/datagate.pl?C1385-3

Memorial Institute for Prevention of Terrorism (MIPT) Knowledge Base Portal: Datasets of cases, groups, and indictment. http://www.tkb.org/IndictmentDownload.jsp

Next Step in Social Network Analysis, Portal for software and research papers on the use of SNA for Visualizing and Diagramming Relationships. http://www.ire.org/sna/

Social Network Analysis of the 9-11 Terrorist Network http://orgnet.com/hijackers.html

Part III:
Emergency Preparedness and Infrastructure Protection

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 11

Disaster Response and the Local Public Health Department

Jonathan B. Weisbuch
Arizona State University, Tempe, AZ 85287, USA

Abstract

Every community emergency, disaster, or catastrophe can be dissected into several phases: the non-disaster (inter-disaster phase), the pre-disaster (warning phase), the impact phase, the emergency (relief) phase, and the reconstruction (rehabilitation or recovery) phase. Planning and preparation is realistic in the first two phases and the impact phase is reactionary. A successful outcome in a life threatening community crisis is a function of thorough preparation, the efficient and effective utilization of resources (during the impact and emergency phase), and the aggregate skills in managing the aftermath. In a biologic disaster, where human health and disease are major factors the local health agency will play a lead role in event management. This chapter will delineate basic public health functions; and describe the importance of local public health leadership in preparation, response, and management of the recovery phase.

1 Introduction

Communities are subject to disasters of natural and human origin. In his comprehensive text "The Public Health Consequences of Disasters," Eric Noji describes the several phases that define community emergency, disaster or catastrophes: the non-disaster (inter-disaster phase), the pre-disaster (warning phase), the impact phase, the emergency (relief) phase, and the reconstruction (rehabilitation or recovery) phase (Noji, 1997). Appropriate human activity during these phases can diminish the burden of pain, suffering, destruction, and death that follows cataclysmic events. One only needs to review the last decade to witness overwhelming events. Earthquakes, tsunamis, hurricanes, urban unrest, terrorism, genocide, war, and epidemics

(natural and manmade) have afflicted large populations worldwide. Injuries and deaths resulting from disasters are products of the magnitude of the event, but the number of deaths, injuries and their severity can be diminished by planning, preparation and an appropriate response to the event, all of which are within human control. Natural disasters occur by chance; they are usually explosive, imparting their energy in minutes or hours. Biologic events, on the other hand, do not occur explosively. These events, whether infectious or due to chronic toxic exposure develop slowly over days, weeks, or years. An infectious disease may take many days to manifest, tuberculosis may take years, and HIV may take a decade to produce symptoms of disease. The results of chronic human exposure to air or water pollution may not manifest for decades as described by Jonathan Harr, in ''Civil Action,'' the story of several children with leukemia, presumably caused by chemical pollution of their city's water supply (Harr, 1995).

Tip O'Neal, Speaker of the House of Representatives in Congress (1977–1987), argued, ''All politics are local'' (O'Neil and Hymel, 1994). It may also be said, ''All health is local.'' Human misery, pain, destruction, disability, and death due to catastrophe have their primary impact on the local community. Reducing these ills with careful preparation, competent event management, and a well-organized plan to meet community needs during the recovery phase is the direct responsibility of local agencies. *Local outcome will be determined by local action before, during, and after the disaster*. These activities may be part of a wider process managed by the state, by federal authorities under national policy, or by the international community. However, at ground zero the burden of response throughout recovery will fall on local responders managing local resources under local leadership. Preparation will define the outcome (Holden, 2005; Ryan and Montgomery, 2005).

With the exception of tsunamis that ravage thousands of miles of coastline, large floods, fires, and earthquakes that lay waste tens of thousands of square miles, pandemics that rage across one or more continents, and the occasional food borne epidemic carried across a nation or a continent (Boase, 1999), disasters initially impact relatively small regions with limited populations. Inadequate response to the initial situation can cause a local event (disease outbreak, a small urban disturbance, or minor conflagration) to engulf a much wider region. The 1972 smallpox outbreak in Yugoslavia (Henderson, 1998), the 1992 riots in Los Angeles (Schnaubelt, 1997) and Hurricane Katrina in 2005 (Cooper & Block, 2006) are all examples of major catastrophes, the impact of which might have been greatly reduced had better planning and preparation preceded the events, and more appropriate efforts brought to the event in their initial stages.

2 Functions and responsibilities of local public health departments (LPHDs)

In the United States, LPHDs serve populations as small as a few thousand or as large as several million people in metropolitan regions. Services

provided by LPHDs vary widely but are all created under state laws and local ordinance to provide a few simple functions. They provide assessment of health hazards, planning and intervention to reduce or eliminate risk, and the assurance that individuals in the community are protected against infectious disease, toxins, pollution, and unsanitary or pathogenic conditions.

Epidemics occur frequently in most communities. The number of food borne outbreaks in the United States reported to the CDC averaged 1330 events between 1998 and 2002 with approximately 10 individuals infected for each event (Lynch et al., 2006). Water-borne outbreaks occur occasionally, and may be associated with small private deep aquifer water systems (Amann et al., 2003) or with large municipal networks, such as the 1993 cryptosporidium outbreak in Milwaukee, WI (Mac Kenzie et al., 1994). School tuberculosis outbreaks, hepatitis A and B, whooping cough, and annual influenza epidemics all occur with regularity. The public health department responds to disease outbreaks by coordinating the actions of infectious disease experts, epidemiologists (local, state, or federal), and physicians. Small outbreaks become the training ground on which public health professionals build skills to respond to larger events. Small outbreaks are the foundation for large-scale disaster management (Olsen, MacKinon, & Bean, 2000).

Infectious diseases that can harm the community are identified and actions are taken to prevent the spread of these pathogens. Many LPHDs assume responsibility for educating the community about behavior that causes disease, disability, and death (smoking, drug use, poor nutrition, domestic violence, etc.). Under state public health authority, clinical practitioners are regulated by State Boards of professional practice, hospitals and other health facilities are licensed by state, and occasionally local, health departments to assure safe clinical environments for the population. Health departments work to preserve the current health status of the community, protect the region from dangerous or pathogenic phenomenon, and promote good health practices by individuals, groups, and businesses. They encourage the passage of legislation that prevents disease and promotes health. Protection of the public is a legal mandate for all health departments. Planning for public health emergencies, disasters, and catastrophes are essential parts of this mandate.

The LPHD must have sufficient resources to perform its daily functions if it is to protect the community in an epidemic, natural disaster, or intentional biologic attack. One essential element is a surveillance system or systems that will assess the community health status on a regular basis and detect untoward events as quickly as possible. A health department must also have sufficient resources to train and *retrain* its work force so they may do their jobs flawlessly; and, during an emergent event, understand and meet expectations. The department must have adequate communication resources and be skilled in presenting the facts to the public, its partners in response, and the political structure responsible for overall management of

the catastrophe. Preparation also requires the LPHD to be linked to other resources at the state and federal levels and to non-governmental agencies that respond to crisis (e.g., Red Cross, religious charities, suppliers of food, water, and shelter).

The LPHD is essential when confronting an emergency since it has many feet on the ground able to respond immediately to the situation. The local health department has eyes and ears in the community; it knows the local leadership, elected, appointed, and natural. The LPHD, having worked other emergencies, knows its local partners, colleagues, and fellow responders. It has resources in the community that can be immediately applied to the situation. An adequate response, however, can only occur if the LPHD is fully staffed, resourced, and trained. A department that has insufficient resources to perform the daily functions required by legislative mandate is an organization that will fail when stressed by a major public health event. The failure will cause excess community morbidity and mortality.

Figure 1 shows the several major activities of a well-resourced LPHD, including the ring of training that encircles it.

Clinical services vary from caring for those with sexually transmitted diseases (STDs) and tuberculosis, or the uninsured needing immunizations and family planning, up to the operation of giant public hospitals that provide care to hundreds of thousands of patients who are poor, destitute, traumatized, and chronically ill. Regardless of the amount of clinical care provided within a local health department, some public providers will be required to respond to the clinical demands of a local, regional, or national bio-emergency.

Community outreach may include home visits, nutrition education in schools, and health status improvement programs in special communities or large-scale health education programs across an entire region. Outreach workers will all be available for redeployment during a crisis.

Environmental services include the monitoring of food services and other businesses that serve the public. They evaluate and improve the quality of air, water, and environment including the suppression of vectors of disease (mosquitoes, rats, rabid feral animals, and other pathogens). Many public health emergencies occur because the environmental protective infrastructure

Fig. 1. The service distribution of most local public health departments.

has broken down. Environmental health professionals determine the reasons behind system failure, and how to rebuild it.

Surveillance, epidemiology, and the maintenance of vital records are at the core of the public health system. They form the basis for the continuing assessment of the health status of the community. As a clinician evaluates the daily progress of hospital patients with blood pressure, temperature, respiration, pulse, and frequent laboratory values; so the public health system constantly monitors its patient (the community), to assure that no strange pathogens have entered the community. One or two unusual cases can identify a rare illness or a new disease that may be overwhelming. This was well told in Shilt's chronicle of the early days of HIV in "And The Band Played On" (Shilts, 1987; Gottlieb et al., 1981), and similarly with the initial cases of Rickettsialpox and trichinosis described by Roueché (1953) in "Eleven Blue Men".

A sound surveillance system draws information from many sources, two of which are of critical importance. Vital records, births, and deaths, are the front line in public health surveillance. Births are monitored for trends in symptoms, birth weight, congenital abnormalities, and other abnormal findings. Death certificates are scanned for frequency and occurrence of rare illnesses, unusual infections, potential toxic hazards, poisonings, medical misadventures, and a variety of other factors that may indicate a problem in the community.[1]

Disease reports are the second critical source of information. All states require by law that physicians, nurses, hospitals, laboratories, schools, pharmacies, and other medical providers report illnesses of public health significance to the state or local health department within 24 hours for some illnesses and 3 days for others.[2] The list varies by state, but is essential in monitoring the health status of the public. Most reportable illnesses are infectious diseases that may be fatal and cause widespread community illness, but some are poisonings, environmental toxicity, or trauma. The list of reportable infectious diseases for Arizona, which is similar in every other state, is shown in Fig. 2.

An effective surveillance system will have sufficient resources to respond to unusual birth or death findings or reports of disease with an initial call to the attending physician or hospital, and further review if necessary. When the reported disease indicates a possible outbreak, all victims, family members, and others in the community may be interviewed. The process of case review and follow-up helps to determine the cause of the untoward event. It

[1]In 2002, the deaths of two 5-year-old from a meningococcal infection within 4 h of each other, initiated an investigation of Naeglaria fowleri in Maricopa County, AZ, which has wide implications for the safety of the local water supply (Marciano-Cabral et al., 2003).

[2]In 1999, a pathologist in Seattle identified two unusual salmonella organisms in one afternoon. Realizing the rarity of the event, he reported the two cases to the local Seattle-King County Health Department, and within 48 h a nationwide food contamination epidemic resulting in thousands of cases had been recognized (Boase, 1999).

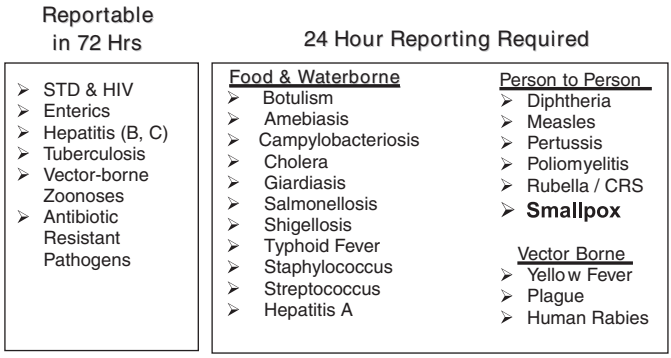


Fig. 2. Reportable disease in Arizona, by time for reporting to LPHD.

also serves as excellent training in preparation for larger outbreaks. A public health system unable to follow up on significant disease reports and deaths will be unable to respond when a major event occurs. Preparedness for disaster begins with sufficient resources to respond to small outbreaks and unusual cases. Practice leads to perfection.

The surveillance system must maintain disease databases for every reportable illness so that new cases may be reviewed against previous trends and all diseases can be monitored chronologically. This latter capacity allows the health department epidemiologist to determine the community frequency and trends for each illness when it exceeds the expectation, national standard, or pre-set level. These trends help to determine disparities between groups regarding childbirth and perinatal mortality, disease treatment, public health interventions, changes in group behavior resulting in illness, and the occurrence of a new or intentionally introduced disease.

A quality community health surveillance system will also maintain hospital and emergency room discharge reports, trauma data, ambulance runs, etc. Changes in any of these variables can adumbrate an emerging situation. A secondary value of hospital data is its use in managing patient flow when an epidemic or other catastrophe requires immediate knowledge of bed capacity and hospital utilization. In a community emergency, the Incident Command may need access to these data in order to allocate bed space and dictate the type of patients admitted to specific hospitals.[3]

In some states cancer registries and those for other chronic illnesses provide long-term evaluation of the changing nature of the community and the potential causes of disease. This last group of records may not be as relevant in an acute emergency situation as are infectious disease reports or discharge data; but when examining long-term pollution trends, especially

[3]During the 2003 SARS outbreak in Toronto, the epidemic command finally required specific hospitals to only care for SARS patients, and other hospitals to limit their intake of elective cases (Campbell, 2004).

for groundwater contamination and air pollution (around point sources of pollution), the long-term trends may be all that is available for determining the possible cause of a problem. Long-term environmental contamination can be as disastrous to a community as an infectious disease epidemic and is often far more difficult to resolve. The Love Canal (Beck, 1979; Whalen, 1978) and groundwater contamination by heavy metals in Woburn, MA (Harr, 1995), are clear examples of this point (Harr, 1995).

A surveillance system must be able to carry out four distinct functions if it is to serve the community well during a biologic emergency. It must be able to record and maintain a large database of disparate disease and medical utilization data. It must be able to initiate a rapid response by trained, knowledgeable, and skilled professionals to examine a variety of rare or unusual diseases or death reports when they occur. The surveillance system helps to identify community factors that may place individuals at risk to illness. The system must be able to evaluate current information against long-term trends and be able to integrate information from multiple databases. A health department without these capacities will be unable to respond when unusual cases appear in physician offices and regional emergency rooms.

A sound surveillance system turns birth, death, and disease reports into action to prevent or mitigate any pathologic processes afflicting the community. These data are the foundation for annual budgetary planning and the foundation on which an emergency response can be built.

Training is a critical element in any organization involved in crisis management. Productive businesses invest time and money to improve their employee's skills, knowledge and productivity through training and continuing education. Fire departments and police agencies provide continuous training to their work force. No current standard, however, exists for the number of hours a LPHD should invests in training and upgrading the skills of its staff. Public health workers must be trained to respond, to have the tools to perform their daily jobs efficiently and adapt to new responsibilities. If the agency is unable to carry out its daily mandated functions, if large numbers of positions are unfilled, and if training programs have been ineffectual or non-existent, the agency will be unable to mount an appropriate response to any situation.

Communication skills are another essential element in maintaining control during a major public health event. An LPHD of any size must employ a qualified *Public Information Officer* (PIO) trained in communication. This individual understands medical and public health jargon, the public health culture, and the need to protect individual privacy while giving the public adequate information. The PIO knows how to get accurate information to the community through the media, and how to inform other members of the action team quickly and cogently. The PIO understands the principles of simple communication, using simple linked concepts to describe the facts and the implications of those facts to convey information. The goal is to

give facts and information while preventing panic, terror, or distress. The PIO must know how to prepare the Public Health Officer (PHO) and/or political leadership for formal press conferences with open questions and answers. He or she must be prepared to organize other forms of information-sharing events for the public when the need arises. During an emergency, the communication skills of the PIO and the PH leadership are often the defining factors in how the community responds (Holden, 2005).

Links to other community resources are essential during normal periods and critical in emergencies. A major event is not the time to become acquainted with leaders of other agencies, nor to determine who in the public safety realm will cordon off a quarantine area or transport sick patients to hospitals.

Each functional area of the LPHD has its own set of partners. Clinical services must have good relations with local medical providers, hospitals, nursing groups, ambulance systems, and pharmacists. Community outreach coordinates with churches, schools, ethnic minorities, and NGOs with whom the agency will work during a crisis. Environmental health is tied to political leadership and departments of environmental quality in cities and at the state level. It is linked to the businesses it regulates and the legal services required when executing those regulations. Surveillance is tied to the medical community, schools, and businesses, as well as the state health department and CDC. The LPHD administration is tied to political leadership and emergency planning agencies in the cities, counties, and state. The public health leader works with the Sheriff, local fire chief, and directors of city and county government and is familiar with the NGO leadership in the region. Legal council attached to the department exclusive of other facets of government must also support the Director of the LPHD. When difficult decisions relating to the health of the community are to be made, they must be based on what is best for the health of the community, and consistent with its laws and regulations (Goodman, Rothstein, Hoffman, Lopez, & Matthews, 2003).

Three examples of situations that occurred in Maricopa County, AZ illustrate the development of a LPHD becoming a key member of the emergency response team.

1997 Sky Harbor Airport: On a Saturday afternoon in early March 1997, an American Air West 737 aircraft from Acapulco, Mexico, landed at the Phoenix Sky Harbor Airport. Over half of the 125 passengers were suffering from acute diarrhea. Prior to landing, the pilot informed the tower of the situation; the tower called the Phoenix Fire Department emergency response team. After landing, the aircraft stopped on a taxiway where it was met by several ambulances. Paramedics assessed the situation, judged 25 passengers to be sufficiently ill to warrant transfer to local hospital emergency departments (EDs), where they were seen and treated. Six patients were admitted for dehydration, staying 24 h under medical supervision.

After the very sick were removed, the plane proceeded to the gate, disembarking the remaining passengers. The aircraft was cleaned, and then continued its trip to Detroit.

In spite of the legal requirement noted above that certain gastrointestinal diseases must be reported to the local health department, the Maricopa County Health Department was not informed. None of the patients were interviewed as to a possible cause for their illness; no central record of names, hospital admissions, or disposition was kept. Stool samples were not obtained. The infectious disease investigators had no way to determine the cause of the illness. The comment of the Fire Department Medical Officer was "*I didn't know we had a county health department.*" That mistake would not be made again.

The new county health officer initiated a process to build an emergency response system that linked the ambulance systems, hospital emergency rooms, local health department, and the county emergency management system so that in any future event all parties would be informed immediately. In 1999, following an anthrax letter scare in Phoenix, the Maricopa County Department of Public Health (MCDPH) was called, its medical director went to the site, and worked with police and fire first responders to assure that the (potentially) exposed were adequately managed, decontaminated, and sent to appropriate emergency rooms for observation. The developing relationship between the first responders in the local fire departments, the emergency planners in the county Emergency Services Agency, and the public health department were important community assets 2 years later when four games of the 2001 World Series were played in Phoenix.

2001 World Series between the Arizona Diamondbacks and the NY Yankees: The Maricopa County public health and emergency management system was tested severely in 2001 when the local baseball team won the National League Pennant. In less than a week, hundreds of thousands of sports fans would descend on Phoenix for the World Series with the NY Yankees. What event would be more attractive to a terrorist? With the destruction of the World Trade Center and an intentional anthrax exposure still under investigation, and the possibility of smallpox bioterrorism a very real threat, the county emergency management, public safety, and public health agencies joined to protect the public.

Public Health and Environmental Services were responsible for preventing disease. This meant protecting food and water in the baseball stadium, reducing the chance that someone with a contagious disease like smallpox could enter the stadium. And, since prevention might fail, public health monitored regional hospital emergency rooms for 5 weeks after the start of the Series to identify anyone with symptoms reflecting one of the diseases of terrorism: anthrax, plague, tularemia, hemorrhagic fever, botulism, or smallpox.

The "Syndromic Surveillance" system was introduced into 11 Maricopa County hospital EDs, prior to the first game of the Series and maintained

for 3 weeks after the final out. The system, developed by CDC for the Winter Olympics (February, 2002), had been used in New York City following the anthrax outbreak in October 2001 (Fleischaur, 2004). Every patient entering an ED was entered into the database with relevant information about their chief complaint and whether they had attended the World Series. The data was sent electronically to Atlanta, evaluated, and returned to the MCDPH Division of Epidemiology within 24 h. The county's infectious disease medical staff interviewed patients with symptoms associated with the diseases of bio-terrorism. In 5 weeks over 37,000 emergency room visits were entered into the system. The health department was in contact with each ED on a daily basis, interviewing all patients whose symptoms were consistent with one of the diseases of biologic concern. None, however, were found to have an illness associated with terrorism (Schumacher, Nohre, & Santana, 2003).

By October 2001, after several years of preparation and many emergency simulations, the MCDPH had developed a complete network of partners and collaborators in community-based emergency management. It had worked closely with the County Sheriff's Department on several events in the previous years and with the local Fire Departments in both simulated and actual events. Contacts with local police and the FBI were a more recent connection, but were very important in preparation for the World Series. Long-term associations with the CDC were strengthened as a result of the event. The public health agency educated its non-health partners in disease, forced changes in the delivery of food services in the ball park, coordinated the efforts of emergency room staff during the syndrome surveillance phase, and followed suspicious patients to assure that no one had been infected during the World Series. No one became ill; bioterrorism did not occur, nor did any other destructive event, but uniting diverse agencies to preserve public safety was significant because it established public health as a key member of the county emergency-planning group. Three years later, in the spring of 2004, these connections became important when West Nile virus infected Maricopa County.

2004 West Nile virus outbreak in Maricopa County, AZ: The advance of West Nile viral infection across the United States, which began in 1999 in New York City, finally arrived in Arizona in the fall of 2003 with one confirmed case.[4] The first human infection in 2004 occurred in April, a blood donor showed a positive test for the virus. Over the next 25 weeks 391 confirmed cases were reported, 16 of whom ultimately died. The epidemic peaked in late June and early July as the county began rapid expansion of ground fogging for mosquitoes.

[4]In October of 2003, a patient was treated in a local hospital with encephalitis like symptoms. Blood and cerebral spinal fluid were submitted to CDC for confirmation. Maricopa County was not informed of the positive results, however, until Spring 2004, at the outset of the largest encephalitis outbreak in the county's history.

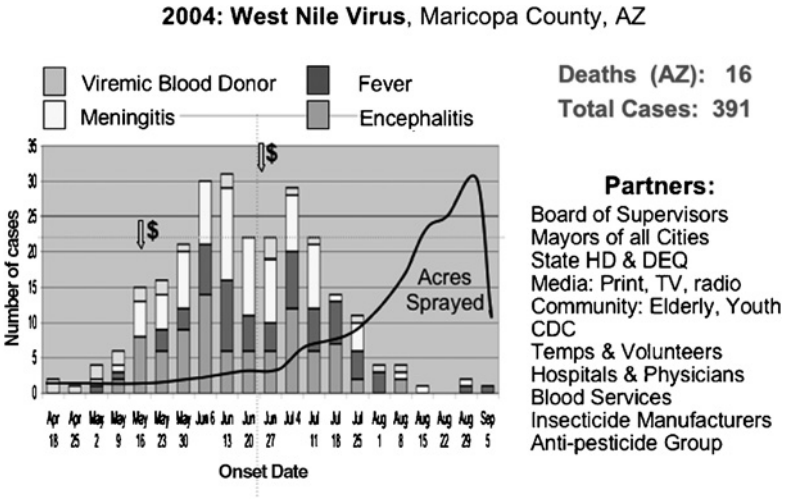


Fig. 3. 2004: West Nile Virus, Maricopa County, AZ.

In this outbreak partnerships that had developed over the past 6 years, made it possible to use the media to saturate the community with information, ask elected officials to direct local public employees to identify mosquito harborage sites (including non-chlorinated swimming pools), and gain support from youth groups and the elderly to work in local neighborhoods to reduce the risk of mosquito development. The hospital system, medical community, blood banks, and laboratories were all part of an aggressive reporting system that tracked the epidemic so that appropriate applications of larvicides and pesticides could be applied. The political support provided to the MCDPH and MCESD was critical, adding resources at two important times in the epidemic, once after the first death, and second when CDC recommended a massive increase is pesticide spraying. (These events are outlined in Fig. 3, The 2004 Maricopa County West Nile Virus Epidemic.)

3 When an emergency occurs

Depending on the nature of a biologic event, the public health response will vary. One case of smallpox in the community will require immediate mobilization; whereas a case of anthrax may result in a more measured response until the nature of the situation is clarified. Regardless of the speed of the response, the shift from daily operations to emergency mode has the same elements. The LPHD will open its Emergency Operations Center (EOC) staffed by members of the emergency response team and other leaders of the organization. The Director or Chief Health Officer will assume *Command*, notify the political leadership to which he or she reports,

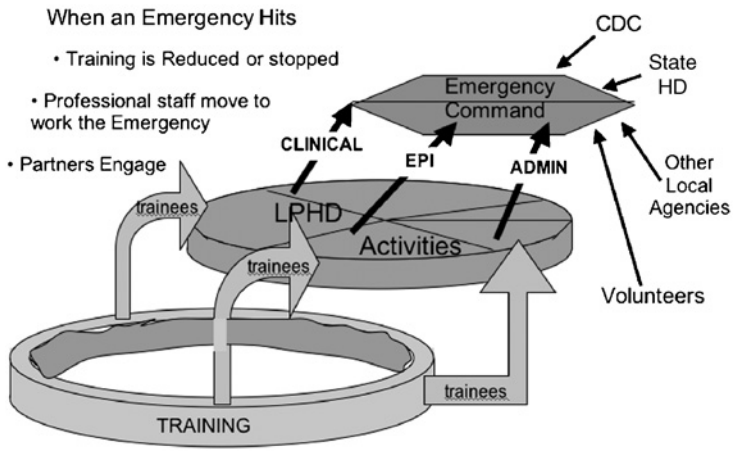


Fig. 4. LPHD response to an emergency situation.

and begins the process of gaining *control* over the situation. The PIO begins the *communication* process by preparing appropriate messages for release and establishing the mechanism for distribution. The organization shifts into emergency operation mode. Those in training return to their normal jobs or take on responsibilities pertinent to the situation. Other professionals within the organization stop their daily activities to assume work on the situation. The emergency management team evaluates the emergency, marshal's resources, and plans the response. Partners from local agencies or NGOs move into the process as needed. State public health professionals and those from the CDC are informed.

Figure 4 shows the process in graphic form. *The first priority is to evaluate the nature of the emergency, assign the appropriate PH professionals to the task, notify political leaders at all levels of government, and build the communication messages for release. The second priority is to maintain as many of the normal PH functions as possible. Clinical care, surveillance, environmental health activities do not cease when an event occurs. The objective is to protect the health of the community*.

In most situations the local health department, in concert with local partners, will be able to resolve the emergency through the investment of local resources coupled with moderate outside assistance (State resources or those from CDC). However, when an overwhelming emergency occurs such as a major earthquake, flood, massive civil disorder, the outbreak of a regional epidemic spreading beyond the jurisdiction of the LPHD, or the intentional introduction of a deadly disease with both health and criminal implications, the entire structure is elevated to another level.

At this point, with the declaration of a state of emergency (usually by the Governor of the state), a regional emergency *Command and Control Center*

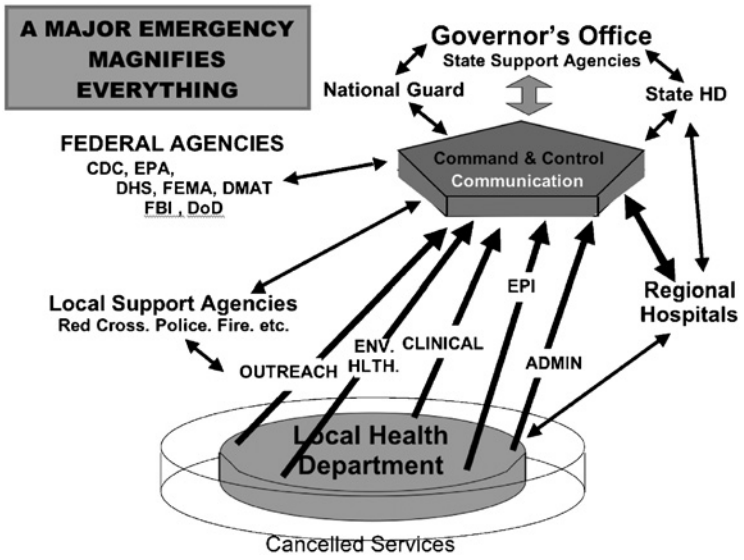


Fig. 5. Local and regional response to a major emergency.

is established with direct links to the Governor's office, to other state and local agencies (National Guard, state heath department), and to appropriate Federal agencies. Professional resources from the LPHD are recruited into the process, forcing a decision to reduce the volume of local services being provided. In this situation the original Incident Commander, presumably the local Health Officer, will relinquish control to someone of higher authority (The state Health Officer, Director, of Emergency Services, etc.). Local professionals will provide resource support and direct links to the prime responders in the community. (Figure 5 depicts the changing structure as the emergency as it expands beyond a local level.)

A critical element in the ''Major Event'' scenario is the role of Communications within the Command and Control Center. Regular and reliable messages to the media about the event must flow from one communication center. When a news conference or other public announcement is required, a single spokesperson that is in authority, has *gravitas*, is skilled in communication, and is knowledgeable about the disease in question must become the face for the emergency. In the absence of a central source of information and an appropriate spokesperson, the process may degenerate into chaos. During a PH emergency, the single communication center must give accurate, comprehensive, and timely information. The individual giving the briefing should not speculate about the outcome, and be assiduous in protecting the privacy of those victims who have succumbed to the disease or threat. In a crisis, quality communication may be as important as quality clinical care to the victims and their community.

4 Conclusion

In all community disasters, the local public health department is an essential member of the emergency team. In a biological emergency, however, the local public health department becomes the essential player. Its local connections, resources, knowledge of disease, and ability to conduct rapid epidemiologic investigations into the causal relationships associated with the outbreak are critical when an infectious or toxic disease is spreading through the population. In a biologic event, the health department can provide the timely, accurate, and appropriate responses that prevent major problems that occur when an infectious disease is uncontrolled in its earliest phase.

The resources available to the LPHD at the outset of the event determine the success of the response. An agency with sufficient resources to carry out all its mandated activities on a daily basis, with resources to train and retrain its work force to perform their normal tasks with precision and be able to manage local emergencies skillfully will be an effective respondent during a major event. It will be an effective member of the community emergency team, having collaborated with other local, state, and federal agencies within an Incident Command Structure under emergency conditions.

A community that is prepared to respond to significant emergency events will have agency directors, including the local Health Officer, who know when to assume *Command and Control* and when to relinquish *Command and Control*. They all must know who is in *Command* at all times. Each local agency PIO will have worked with other PIOs and be able to identify who will be the spokesperson for any given situation. In a crisis, a single *Communicator* is as essential as having a single commander. The community will listen to all that the media has to communicate; multiple stories about the same event can lead to confusion and chaos.

No community can know when it will be hit by an emergency, but every community can allocate sufficient resources to prepare its leadership and their agencies to work together. The initial reaction to the event will be at the local level, if biologic in nature, the reaction will be by the health department. Its preparation and response will directly impact the nature of death, disability, and disease that the region will suffer.

5 Study questions

1. What public health problems are associated with the Sky Harbor event?
2. What should have been the public health response to the Sky Harbor event?
3. What public health functions were utilized during the World Series in 2001?

4. What would be the appropriate reaction were someone apprehended entering the World Series with what appeared to be the markings of smallpox?
5. What actions might have been taken to mitigate the 2004 West Nile outbreak in Maricopa County?
6. How does the Director of Public Health determine which public health functions are to be shut down during a major event requiring large numbers of public health personnel?
7. What factors during a major biologic event dictate the shift in Command and Control?

References

Amann, J., Berisha, V., Visvesvara, G., Arrowood, M., Santana, S., Kolman, J., Shafer, M., Sriram, R., Reese, D., Waddell, V., Vaz, V., Waldbilling, T., Juranek, D., Maguire, J., Brown, A., England, R., Weisbuch, J., Beach, M. (2003). *Naegleria fowleri in a Drinking Water System: Two Fatal Cases of Primary Amebic Meningoencephalitis – Arizona, 2002,* CDC EIS Conference on Parasitic and Vector-Borne Diseases, Abstracts for Session H-2, Atlanta, GA, April 2, Session Archives, www.cdc.gov/eis/conference/archives/2003ProgramAbstracts.pdf

Boase, J. (1999). *Outbreak of Salmonella Serotype Muenchen Infections Associated with Unpasteurized Orange Juice — United States and Canada, June 1999*. Accessed May 14, 2006. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm4827a2.htm

Beck, E.C. (1979). The Love Canal Tragedy. *Environmental Protection Agency Journal*, January. http://www.epa.gov/history/topics/lovecanal/01.htm

Campbell, A. (2004). *The SARS Commission Interim Report, SARS and Public Health in Onterio*. Executive Sumary, Province of Ontario, Canada, http://www.sarscommission.ca/report/Interim_Report.pdf

Cooper, C., R. Block (2006). *Disaster: Huricane Katrina and the Failure of Homeland Security*. Times Books Henry Holt and Co., New York.

Fleischauer, A.T., B.J. Silk, M. Schumacher, K. Komatsu, S. Santana, V. Vaz, M. Wolfe, L. Hutwagner, J. Cono, R. Berkelman, T. Treadwell (2004). The validity of chief complaint and discharge diagnosis in emergency department-based syndromic surveillance. *Academic Emergency Medicine* 11(12), 1262–1267.

Goodman, R.A., M.A. Rothstein, R.E. Hoffman, W. Lopez, G.W. Matthews (2003). *Law in Public Health Practice*. Chapter 10, Sections II and III, Oxford University Press, New York, p. 195.

Gottleib, M.S., H.M. Schanker, P.T. Fan, A. Saxon, J.D. Weisman, I. Pozalski (1981). Pneumocystic pneumonia – Los Angeles. *CDC MMWR* 30(21), 1–3.

Harr, J. (1995). *A Civil Action*. Random House, NY.

Henderson, D.A. (1998). *Smallpox: Clinical and Epidemiologic Features*. Johns Hopkins Center for Civilian Biodefense Studies, Baltimore, M.D. http://www.cdc.gov/ncidod/EID/vol5no4/henderson.htm

Lynch, M., J. Painter, R. Woodruff, C. Braden (2006). Surveillance for Gooddborne-disease outbreaks – United States, 1998–2002. *CDC MMWR* 55(SS10), 1–34.

Mac Kenzie, W.R., N.J. Hoxie, M.E. Proctor, M.S. Gradus, K.A. Blair, D.E. Peterson, J.J. Kazmierczak, D.G. Addiss, K.R. Fox, J.B. Rose, J.P. Davis (1994). A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. *New England Journal of Medicine* 331(3), 161–167.

Marciano-Cabral, F., R. MacLean, A. Mensay, L. LaPat-Polasko (2003). Identification of *Naegleria fowleri* in domestic water sources by nested PCR. *Applied and Environmental Microbiology* 69(10), 5864–5869.

Noji, E.K. (1997). *The Public Health Consequences of Disasters*. Oxford University Press, Oxford, UK, p. 13.

Olsen, S., G. MacKinon, S. Bean (2000). *Surveillance for Foodborne Disease Outbreaks—United States 1993–1997*. http://www.cdc.gov/mmwR/preview/mmwrhtml/ss4901a1.htm

O'Neil, T. P., G. Hymel (1994). *All Politics is Local: and other rules of the game.* Bob Adams/Random House, Holbrook, MA, p. xv (Introduction).

Ryan, M., H. Montgomery (2005). Terrorism and the medical response. *New England Journal of Medicine.* 353(6), Massachusetts Medical Society, Weston, MA. pp. 543–545.

Roueche', B. (1953). *Eleven Blue Men. The Medical Detectives.* Times Books, New York, NY, pp. 1–10.

Schumacher, M., L. Nohre, S. Santana (2003). Syndromic surveillance using emergency department data. Division of Epidemiology and Bioterrorism Preparedness and Response. Maricopa County Department of Public Health. Partial Evaluation of a Drop-in Bioterrorism Surveillance System in Phoenix, AZ. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 80, (2, Suppl 1), p. 118. New York Academy of Medicine, New York, NY.

Shilts, R. (1987). *And the Band Played On*, Chapter 7, St. Martin's Press, New York, NY, pp. 61–69.

Whalen, R.P. (1978). *Governor's Love Canal Inter-agency Task Force Report*, State of New York, Department of Public Health, September, http://www.health.state.ny.us/nysdoh/lcanal/lctimbmb.pdf

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 12

Challenges of Bioterrorism Preparedness for Organizational Processes and Resources

Orneita Burton and Minu Ipe

Center for Advancing Business Through Information Technology, Arizona State University, Tempe, AZ 85287-4606, USA

Abstract

The threat of bioterrorism has prompted systemic change in the business processes of organizations. The need for change is particularly significant in response mechanisms used to acquire and utilize information for decision making in government agencies such as departments of public health. Because information and human resource requirements increase exponentially in a bioterrorism environment, there is a need to better understand the impact of bioterrorism preparedness efforts on organizational processes. This chapter provides an overview of the impact of this change on the structural aspects of surveillance as organizations reorient to deal with intentionally in disease surveillance.

The threat of bioterrorism has prompted systemic change in the business processes of organizations. The need for change is particularly significant in response mechanisms used to acquire and utilize information for decision-making processes in government agencies such as departments of public health. Public health departments are at the forefront of the response to a bioterrorism event, not only as critical responders to identify and contain the biological agent, but also to manage resulting long-term health issues within the community. In the event of a bioterrorism attack, public health surveillance mechanisms should lead to decision outcomes that are both expeditious and accurate. However, bioterrorism adds complexity to surveillance processes as traditional scientific investigations of natural outbreaks gravitate to criminal investigations of malicious intent. Unlike

natural outbreaks, investigations of intentional acts of bioterrorism require additional information and resources, not only to treat the affected population but also to identify and apprehend a perpetrator to prevent reoccurrence of the event. This change moves the investigation process from collaboration between agencies within the internal structure of public health to a cross-functional investigation involving agencies with dissimilar decision-making processes and investigation goals.

Because information acquisition and investigation resource requirements increase exponentially in a bioterrorism environment, there is a need to better understand the impact of bioterrorism preparedness efforts on organizational processes as departments of public health experience the forced convergence of disease surveillance, law enforcement, and risk management. In this chapter, we provide an overview of the impact of this change on the structural aspects of surveillance as organizations gear up to deal with intentional outbreak conditions while maintaining a level of preparedness commensurate with more traditional approaches to disease surveillance, vis-à-vis, public health in the reporting, investigation, and control of natural outbreak conditions. The overall impact is viewed from a systems perspective, as organizational processes adapt to spatio-temporal and knowledge fluctuations in the development of information structures and thus experience changes in informational and human resource requirements.

The objectives of this chapter are to:

- illustrate and examine the conditions that impact how information is structured in surveillance systems and the relationships between processes used to gather and structure information.
- discuss the impact of intentionality introduced by bioterrorism preparedness on the development of information structures used in decision making.
- examine the human resource investments necessary for a dynamic surveillance process.

In doing so, we evaluate the complexity of decision making in public health surveillance by considering changes in environmental factors such as information, people, and technology. In an area of explosive population growth, the combined effects of these factors could greatly influence the efficiency and effectiveness of resource allocations and thus alter decision outcomes that result from existing surveillance processes.

1 Organizational adaptations in public health: a systems perspective

Public health is an excellent example of an organization experiencing a major transition in operations to support preparedness efforts for bioterrorism. Departments of public health have traditionally operated to control the spread of natural disease through surveillance systems that employ

reactive case-based surveillance processes. Most cases or incidents of disease are reported by community agents (e.g., medical units, school nurses, nursing home attendants, concerned individuals) to the local public health department, where interventions are planned over time to guide response strategies. Although interventions are also planned by analyzing historical data to identify patterns of disease development, response is typically reactive vs. predictive in nature.

In reaction to changes in the environment, organizations and associated systems also experience change through natural growth processes. Public health and its surveillance systems have undoubtedly experienced natural periods of growth as advances in information technology have introduced more sophisticated, patient-based, and syndromic surveillance systems to capture real-time data from multiple information sources to signal the probability of an outbreak and initiate public health response (Lober, 2003). However, organic growth occurs naturally only to a certain point without disrupting the normal flow of standard decision and work processes. Exponential growth or a significant reorientation of any system would require a full evaluation of business processes to understand the organizational perturbations that occur through systemic change. With the October 2001 anthrax attacks and severe acute respiratory syndrome (SARS) outbreaks in 2003, the increased awareness that terrorist groups may be capable of releasing life-threatening biological agents within the population has added credibility to the threat of bioterrorism. Preparing for a bioterrorism attack represents the need for a reorientation of the public health system and, oddly enough, offers a timely opportunity to reassess business processes in public health organizations.

The threat of bioterrorism has created a major shift in orientation from the need for patient-based surveillance operations to event-centric surveillance strategies. This change in perspective has established the need to reassess surveillance goals and the resources needed to acquire information whose value is determined by time and knowledge restrictions that occur as an event progresses (see Fig. 1).

In public health, surveillance is viewed as a decision-making process involving an allocation of resources, which in turn impacts the effectiveness and efficiency of decision outcomes. The key resource components in surveillance processes are the human resources utilized in public health investigations and informational sources acquired during the investigation process. In a bioterrorism environment, a dynamic mechanism of resource allocation is needed to support surveillance activities. Pre-event surveillance strategies, once deemed too costly to justify the expense of evaluating environmental conditions prior to the occurrence of an event, are now regarded as a necessary component of risk assessment to acquire the information needed to determine the probability and intentionality of an act of bioterrorism. When an outbreak condition is positively identified, surveillance resources are expended exponentially when a traditional

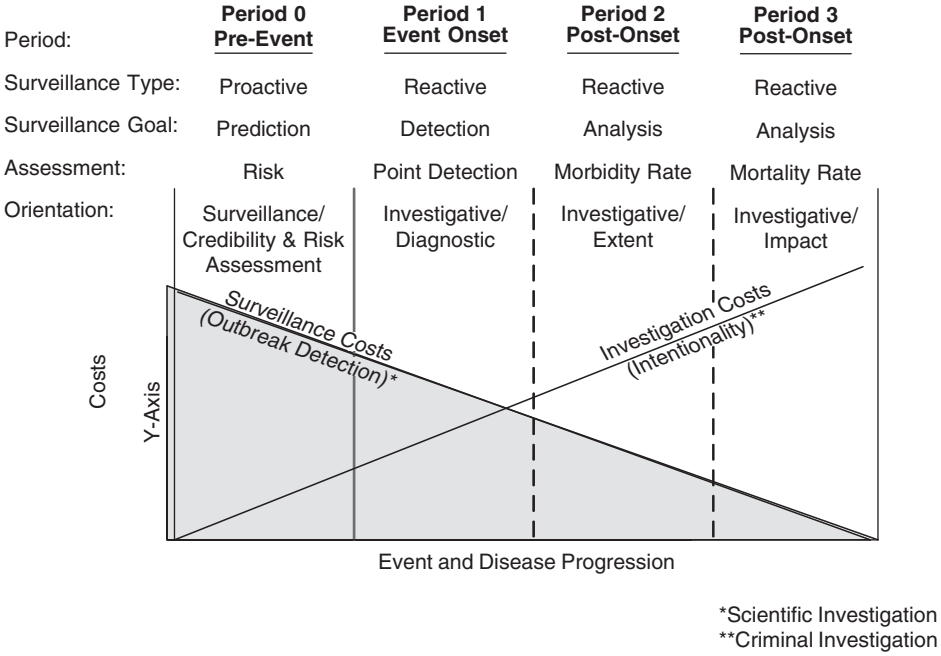


Fig. 1. Event progression in bioterrorism: time value of resources expended in outbreak surveillance and investigation.

scientific investigation becomes a criminal investigation as additional informational sources are required as defined by the outbreak type. With this in mind, a key concern in public health is the efficient allocation of human and informational resources to coincide with the severity and intentionality of an event. In an environment of intentionality, the heart of this problem lies in making a distinction between a natural and an intentional outbreak. The complexity of differentiating between outbreaks can readily overwhelm the effectiveness and efficiency of existing information gathering and distribution networks. Therefore, the availability of relevant, outbreak specific information and the resulting information structures can significantly impact decision outcomes and influence how human resources are allocated with changing decision-making goals.

With these considerations in mind, the following sections discuss how information structures and investments in human capital are impacted by unique aspects of bioterrorism.

2 Information structure

Surveillance processes in public health acquire information from a number of sources throughout the surveillance environment. As a result,

a large quantity and variety of information is gathered that must be organized or structured for subsequent analysis and reporting. The ability to gather and structure relevant information in surveillance activities is often associated with problems of fragmentation in information structure. Fragmentation or suboptimality in information structure has been associated with a lack of existence or adherence to strategy (Weill and Broadbent, 1998; Monteiro and Hepso, 2002; CDC, 2001). Because relevant information is not always distributed efficiently and distortions in information are normal occurrences, information quality and the resulting development of information structures is invariably compromised.

In an environment with bioterrorism, such inconsistencies in information structure can significantly impact decision outcomes and the allocation of resources used in decision processes. With the need for additional and nontraditional information sources to make a distinction between a natural and intentional outbreak and to identify the true cause (i.e., assess intentionality), conditions such as information overload and distortions in information can lead to delays and inconsistencies in strategies used to respond to outbreak conditions. Because these conditions are intrinsic to the surveillance environment, a challenge in surveillance is to better understand the relationships that exist between organizational processes and the development of information structures in order to generate the best possible decision outcomes. In an environment impacted by threats of bioterrorism, three important conditions impact the development of information structures: (1) information relevance; (2) information value; and (3) information availability. The relative importance of each condition is discussed in greater detail in the following sections.

3 Information relevance

The science of surveillance is well defined in the public health literature (Beaglehole and Kjellström, 1993; Teutsch and Churchill, 2000; Bhopal, 2002). In traditional public health surveillance, the acquisition, analysis, and interpretation of disease information is performed through the science of epidemiology. Epidemiology is the scientific discipline of studying the incidence, distribution, and control of disease in a population, the factors affecting the progression of an illness, and, in the case of many chronic diseases, their natural history (Beaglehole and Kjellström, 1993). In public health epidemiology, information structures used to plan interventions are formed through the compilation of information acquired during the investigation process. The quality or "completeness" of the resulting information structure impacts decisions made regarding the origin and intentionality of an outbreak. Table 1 is an example of an information structure from a public health surveillance environment with relevant information highlighted. With each form of bioterrorism, the information structure can

Table 1
Relevant information structure

Relevant question	Who	What	Where	When	Why	How	Extent
Source ID	*Name*	*Disease/agent*	*Location*	*Timing*	*Intentionality*	*Method of delivery*	Range of impact
Label	*N*	A	*L*	*T*	*I*	*M*	*E*
Intelligence	Possible perpetrator or victim	Possible biological agent	Intended target, location of perpetrator	Possible time of attack	H-bioterror L-accident or nature	Possible method of attack	Target percentage of population
Law enforcement	Perpetrator and victim	Bio-agent used	Where incident occurred	When occurred	H, M, L	Method of attack	No. of people attacked
Medical community	Patient	Disease or symptoms	Where patient became ill	When patient became ill	Why person became ill (e.g., age)	How person became ill	No. of people with similar symptoms
Public health epidemiology	Infected person	Disease or pathogen	Source of exposure	Disease onset	H, M, L	How pathogen got into source	No. of people initially exposed (prevalence)
Environmental	Host	Bio-agent	Source	Introduced in environment	Man or nature	Method of introduction	Percentage population affected
Animal surveillance	Animal type	Disease or bioagent	Where animal was exposed	When infected	Why animal became ill	How was animal infected	How many same type/location
Medical examiner	Victim	Disease or bioagent	Where deceased	Time of death	Cause of death	Method of delivery	No. of people dead from same cause

differ and will be defined by the outbreak type. The resulting information state and thus the soundness of the information structure govern the quality of decision outcomes and responses to the event.

In this information structure, public health is regarded as an information hub, gathering information from partners in the surveillance environment and using the resulting information structure in decision processes. What the structure suggests is that public health cannot operate as an independent organization in a bioterrorism environment. With the introduction of intentionality, there is a need to view the decision structure in terms of an "information supply chain." Table 1 shows the individual contributions to this information supply chain by partner agencies, with information categorized by type and relevancy in detecting an outbreak condition. With the notion of intentionality added to the equation of public health surveillance, this information structure emphasizes the need to leverage relevant contributions from surveillance partners.

4 Differentiation between natural and intentional outbreaks

The need to understand the formation of information structures is particularly important in public health due to the decision-making options available in outbreak detection. In public health surveillance, strategies to acquire information are used in the surveillance and detection of both natural and intentional outbreaks. However, investigations of intentional outbreaks require additional information not only to treat the affected population, but also to identify and prosecute the terrorist. Because resources are expended exponentially in a criminal investigation, falsely identifying a natural outbreak as a bioterrorism event comes with significant cost as valuable surveillance resources are misallocated and community trust is compromised.

To properly identify the causal agent in a disease outbreak, relevant information is needed to distinguish between a natural outbreak and an outbreak with intentional origin. Although such causal information is required to develop more complete information structures to optimize decision outcomes, investigators typically accept less than optimal structures to minimize information acquisition costs. In real world situations, conditions of optimality are costly and difficult to attain. In fact, research suggests that acquiring additional information has diminishing returns (Balakrishnan and Whinston, 1991). In public health, conditions often exist under which exact information is too costly to acquire, where suboptimality becomes a satisficing solution (Simon, 1981).

A satisficing solution implies the decision to accept a choice or judgment as one that is good enough, one that satisfies. In public health surveillance, incomplete structures that develop based on partial information might be more appropriate than using detailed, accurate structures that confirm

causality. Often, plans for intervention can be made without knowing the true cause of an event. On the other hand, not identifying a true cause or not searching for causal information carries a substantial cost if the development of suboptimal information structures results in process inefficiencies and detrimental decision outcomes.

Prior research has associated distorted or incomplete information structures with poor decision outcomes. The cause of distortions can be by default, as in the choice to accept satisficing solutions, intentional, as in having a cognitive bias for certain decision outcomes (Lichtenstein et al., 1978), or unintentional, due to inconsistencies in the acquisition process (Nelson et al., 2005). In this view, decision makers are seen as susceptible to biases due to the goal of the surveillance process.

Prior research in decision-making processes has also confirmed the fact that organizational decision makers can have strong preferences for certain decision outcomes (Hayek, 1945; Pfeffer and Salancik, 1977, 1978; Pfeffer, 1981; Staw, 1981). An argument has been made that decision makers may develop strong preferences for certain outcomes through organizational goals, incentives, and control systems (O'Reilly, 1990). Once committed, decision makers are seen as susceptible to biases in both the acquisition and processing of information. For example, information that supports desired outcomes is sought out, while information that supports an opposite view is challenged, either by questioning the accuracy or credibility of the source (O'Reilly, 1990). In this instance, information is not fixed in use but may be selectively perceived and processed.

5 Information value

The value of information is based largely on its ability to affect decisions. The threat of bioterrorism adds complexity in understanding relationships between information value and decision outcomes in the surveillance environment. Unlike acts of terrorism, bioterrorism creates unique conditions in time and people that hinder the effectiveness of decision makers when identifying and communicating the occurrence of a bioterror event. These conditions create disparities in information states that are bounded in time and space by specific individuals. This is referred to as a condition of information specificity.

The concept of information specificity is derived from the concept of asset specificity (Coase, 1937; Williamson, 1975, 1985). Just as asset specificity refers to the extent to which the value of an asset is restricted to specific transactions, information specificity is defined as the extent to which the value of information is restricted to certain uses or acquisition modes by specific individuals or during specific periods of time. In the context of information, two forms of specificity are important in surveillance:

knowledge specificity and time specificity (Choudhury and Sampler, 1997). Information that has high knowledge specificity can be acquired and used only by individuals or groups with the expertise, skills, or knowledge necessary to make the information available and valuable in the surveillance process. Information that has high time specificity loses its value when not acquired soon after it first originates and decreases in value unless used shortly after it becomes available. Information has low time and knowledge specificity when its value is not affected even if it is acquired or used later in the investigation process.

The concept of information specificity explains some of the complexity posed by bioterrorism. First, the time that information is made available is critical in bioterrorism surveillance. For example, the information needed to identify and apprehend a perpetrator may be available only temporarily in the investigation process. Pre-event information that can be used to predict the occurrence of an event also has no value once an event has been confirmed. Likewise, certain types of information may be available in forms or languages that are understood only by individuals with specific skills or knowledge. The result is that, although information that is critical to the success of the surveillance process actually exists, it may not be in the possession of individuals who can use it to add value to the investigation process. Table 2 provides examples of information that could be missing from information structures due to time and knowledge specificities.

As an additional consideration, reactive approaches to surveillance often use secondary data sources to detect the occurrence of a disease outbreak. However, secondary data that is high in time specificity would have limited value in an intentional outbreak, as disease progression could occur rapidly due to the deliberate introduction or spread of the disease. The result would be abnormally high morbidity and mortality rates over shorter periods of time than would be experienced with a natural disease outbreak. In developing an information structure for timely decision-making, secondary data that is high in time specificity would optimize the information structure if it could be acquired earlier in the investigation process.

Although the effect of time on surveillance processes is well known in epidemiology, employing the concept of time and knowledge specificity in the acquisition and use of information to develop relevant information structures is not commonly practiced in disease surveillance. Time specificity adds value to surveillance processes by defining information that is relevant based on its value over time. This information is particularly useful in justifying the cost of proactive or pre-event surveillance strategies. Knowledge specificity adds value by defining knowledge and skill level requirements for epidemiologists and other investigators who are responsible for acquiring and using information in event detection. Both concepts can be useful in supporting the need for additional funding to support prevention and detection strategies in bioterrorism surveillance.

Table 2
Missing information categorized by time and knowledge specificity

Time specificity		Knowledge specificity	
Acquisition: Information that must be acquired immediately or shortly after it originates or becomes available	Use: Information that decreases in value unless used immediately or shortly after it becomes available	Acquisition: Information that can be acquired only by someone with the required specific knowledge	Use: Information that can be effectively used only by someone with the required specific knowledge
Not captured in a timely manner; not available	Not available; no longer of value	Lack of skills; models or algorithms not robust	Lack of training, skills
Information not relevant to the surveillance phase	Not recognized as relevant or important in this phase of surveillance	Relevant information not identified or miscommunicated	Lack of capacity or capability to use information
Not shared or disseminated in a timely manner	Not received in time	Difference in job structure, divisions, system design	Difference in skill levels; not shared with appropriate decision makers
Not acquired because certainty (timing) of threat not known	Not used because not available	Not acquired because certainty (likelihood) of threat not known	Not used due to behavioral effects, disbelief

6 Information availability

The concept of information specificity is useful in explaining why some information is not included when information structures are formed. Information can be missing from information structures because it was not acquired in a timely manner. Also, valuable information may exist but may not be acquired or used because an investigator lacked the skills or knowledge needed to properly utilize the information in the surveillance process.

Environments where information is absent or missing in structures or processes that acquire information for decision making have been studied extensively in economics and IS research (Kmietowicz and Pearman, 1983; Weber, 1987; Moore et al., 1994). Conditions under which information is not available for decision making have been defined as *incomplete* (Weber, 1987; Moore et al., 1994; Rasmusen, 2001), *imperfect* (March and Simon, 1958; Rothschild and Stiglitz, 1976; Carley and Zhiang, 1997; Rasmusen, 2001), *asymmetric* (Grossman, 1981; Keen, 1986; Fulk and Steinfield, 1990; Smith et al., 1992), and *uncertain* (Conrath, 1967; General Accounting

Office (GAO), 2004). These states define conditions where critical information is missing from the surveillance environment. An incomplete information state occurs when information needed to define the event is not available mainly because it has not been collected or captured by the surveillance system. The condition of imperfect information occurs when information relevance is miscommunicated or not regarded in the same manner by all parties within the surveillance environment. In this context, the problem is one of knowledge specificity, where information judgments are influenced by factors such as the prior knowledge of the individual.

Conditions of asymmetry occur when the same information is not known at the same time by all parties. Information asymmetries can exist internally or external to the surveillance environment, depending on how the environment is defined. Information asymmetries caused by external sources can result from willful acts to hide information to gain time and knowledge advantages. Both internal and external asymmetries can exist through data integration and distribution issues that affect the timeliness and effectiveness of gathering and sharing information.

Uncertainty in information can result from variable or questionable information quality. However, in bioterrorism, conditions of uncertainty can exist within the environment when an assessment of the likelihood or plausibility of the occurrence of the event cannot be made, i.e., if a credible threat exists. Likelihood must be assessed based on the ability and willingness of an agent to carry out the threat as well as the likelihood of when and where a threat will be carried out (GAO, 2004).

Although these conditions have been defined separately in different contexts in research, prior studies have not collectively considered the impact of these conditions on the development of information structures used in public health. In bioterrorism, the surveillance environment consists of dynamic network effects and conditions of imperfect, asymmetric, incomplete information under uncertainty. This condition represents the development of information structures with fuzzy, indistinct boundaries, making it difficult to determine what information is needed to detect the likelihood or occurrence of an event. Definitive information structures could provide the boundaries needed to define the surveillance environment for specific outbreak types and improve the degree of accuracy in event detection.

Decisional and organizational changes, including changes in information structure, do not exist or occur in a vacuum. The ''human'' impact and contribution to these changes as discussed in the following sections is considered as a key influence when examining organizational adaptations to changing external environments.

7 Organizational adaptations in public health: a human capital perspective

As organizations evolve to meet the needs of a rapidly changing external environment, the resulting transformation impacts not just information

structures, but also the organizational structure, social structures, and the composition and configuration of the entity's human resources. Indeed, it can be argued that the particular capabilities of an organization's human resources is the key to ensure that reorientation efforts truly transform the organization to be better prepared to meet the challenges posited by a changing environment. Having the right people, in the right place, at the right time is crucial to an organization's ability to successfully reconfigure itself during the change process.

The knowledge that resides within an individual significantly impacts their ability to successfully acquire and process information, and this knowledge contributes to the human capital of the organization. Human capital refers to an individual's or employee's knowledge, skills, and expertise (Youndt and Snell, 2004). The human capital of an organization comes from its ability to strategically select, develop, and use its human resources (Koch and McGrath, 1996). In the context of public health, the changes in information structure imposed by bioterrorism threats also pose challenges to the human capital of local organizations.

As previously discussed, information relevance, value and availability impact the development of information structures within organizations. It is important to recognize that many of the factors that influence information structure are often rooted in the placement, efficacy, and the infrastructures that support the human resources of an organization. An organization is not a static storehouse of knowledge. Prior research has investigated the development of information structures as information networks. Information networks consist of nodes and the connections between each node, where a node represents the stock of knowledge that resides in individuals or organizations. Multiple knowledge nodes of the organizati on interact with each other and combine in unique ways so as to create value in the form of new knowledge (Bontis and Fritz-enz, 2002). Knowledge in organizations exists in varied forms and is stored and maintained in different places. However, it is only people who can learn, and so human resources become the primary repository of both explicit and tacit knowledge (Lado and Wilson, 1994). The following section extends this general idea of human capital and examines investments needed in public health to maximize and sustain the value of its human resources in order to satisfy the formidable task of bioterrorism preparedness.

8 Human capital investments

Epidemiological surveillance, considered the foundation of public health,[1] necessitates appropriate information technology infrastructures

[1]Institute of Medicine, National Academy of Science. (1988). *The Future of Public Health*. National Academy Press, Washington, DC.

as well as institutionalized information sharing and decision-making processes within public health departments. At the core of the surveillance function are the trained professionals who serve as the brains of the system. The ability of epidemiologists to analyze and interpret data allows surveillance information to be translated into public health intelligence that can then be used for immediate action or for monitoring and study over a period of time. Epidemiological surveillance thus represents a human capital intensive function within public health.

In the context of bioterrorism preparedness, it is clear that the traditional information structures surrounding surveillance systems are inadequate. Along with redefining information structure, the human capital issues that are crucial to the success of public health surveillance have to be understood and addressed. The three most important aspects of human capital investments in the surveillance arena are (1) adequate numbers of well-qualified professionals, (2) maintaining and enhancing skill sets of existing staff, and (3) long-term programs to develop and retain employees. Table 3 provides examples of investments in human capital and suggests implications of each for departments of public health.

9 Adequate qualified professionals

The knowledge pool created by well-trained and experienced professionals serves as the basis for everyday operations and the foundation for future learning in organizations. Such professionals serve as the strategic assets of the organization—capabilities and resources that are specialized, scarce, hard to imitate and find (Amit and Shoemaker, 1993). Sustaining and enhancing this knowledge pool requires the organization to have adequate numbers of well-qualified professionals.

Shortages in the public health workforce have been well documented over the years. Since 2001, when the public health arena came under new

Table 3
Human capital investments and implications

Human capital investments	Implications for public health
Address shortages in the public health workforce	Better capabilities within the organization to adapt to external changes
Increase numbers of highly trained epidemiologists	Higher quality of decision making during a health emergency
Ongoing training for surveillance staff	More robust surveillance systems for the long term
Develop surge capacities for daily operations	Greater levels of institutionalized knowledge and experience
Long-term programs to develop and retain employees	

scrutiny, the shortage of skilled professionals has been recognized as an acute problem. Shortages extend across the spectrum of public health to include nurses, laboratory workers, environmental health specialists, public health managers, and microbiologists. This shortage is particularly prominent with epidemiologists who are critical to the surveillance function. Epidemiologists fall behind nurses as the occupational class most affected by worker shortages. A survey by the Council for State Governments (CSG) revealed that the number of full-time equivalent epidemiology positions engaged in surveillance nationwide dropped from 1700 in 1992 to 1400 by 2002.[2] The problem of workforce shortages has already been compounded with a growing population in most counties across the country, stretching thin existing epidemiological capacities within local public health departments. With the current focus on bioterrorism, the inadequacy of skilled resources stands as a major roadblock to public health departments achieving required levels of preparedness against these new threats.

While most public health departments are adequately prepared to deal with commonly prevalent diseases, the threat of bioterrorism embodies a change in the roles and skill sets of those associated with the surveillance function. Adequate preparedness for bioterrorism requires the public health department, specifically the epidemiology disease surveillance group, to be well connected with an entire host of agencies within the community. Apart from the traditional connections with hospitals, laboratories, schools, and day care centers, epidemiologists now need to be closely connected to first responders such as the police and fire departments, investigative agencies such as the FBI and others in the public domain such as retail and commercial centers. Building relationships with these entities and creating and establishing information-sharing protocols involve intensive preparation and effort, taking away resources from traditional surveillance-related responsibilities. Managing relationships with external entities and institutionalizing information-sharing processes would require reinforcements to the existing resources of public health departments.

Qualifications of epidemiologists in public health range from doctoral level training to undergraduate and other types of educational qualifications. The differences in levels of training point to some of the challenges in bioterrorism surveillance because of the differing levels of knowledge among these professionals. The need for, and the increasing numbers of terrorism preparedness programs across the country, necessitates the availability of a large pool of well-trained professionals. While terrorism-related programs had one of the highest concentrations of professionals with degrees in epidemiology compared to other programs by 2004,[3] it still does

[2]The Council of State Governments. *Public Health Worker Shortages*. Available at http://www.njpha.org/Events/Public+Health+Worker+Shortages.pdf

[3]CDC. (2005). Brief report: terrorism and emergency preparedness in state and territorial public health departments—United States, 2004. *Morbidity and Mortality Weekly Report* 54(18), 459–460.

not mitigate long-term concerns of adequately maintaining staff strength in the surveillance arena.

The changing information structures around bioterrorism surveillance require changes not only in the relationship with external entities and epidemiological skills, but also with processes within public health departments as well. Creating and monitoring processes related to information sharing and decision making within public health departments require intensive efforts and dedicated resources in the short term. If the new processes are institutionalized within the organization, these resources may be deployed elsewhere. However, in the current situation, with barely enough professionals with skills needed to fulfill everyday functional requirements, public health departments are not in the best position to develop robust preparedness programs. The escalating shortage of workers and change in knowledge requirements is thus seen as a serious threat to the capacity and capability of public health departments to respond to bioterrorist events.[4]

A series of federal grants have helped to address some of the shortages of public health staff. With respect to epidemiologists, the number of those working in terrorism preparedness nationwide increased 103%, from 115 epidemiologists in 2001 to 234 in 2004.[5] While this increase in numbers represents a significant growth in epidemiological strength in public health, adequate surveillance for bioterrorism preparedness demands a greater increase in the number of professionals with experience in outbreak detection coupled with training in the intricacies of biological warfare. Public health organizations nationwide will soon have a much younger workforce than exists currently. According to a CSG/ASHTO survey in 2004, the average age of employees in public health is 47 and ~25% of public health employees are currently eligible for retirement. Public health organizations need to recruit and retain greater numbers of professionals in order to sustain their current surveillance programs and to develop more sophisticated support systems for terrorism preparedness and to moderate inherent workforce limitations for the future.

10 Maintaining and enhancing skill sets

Human capital is a depleting resource and is unique in that it is the only asset that can be developed (Fritz-enz, 2000). Hiring well-qualified professionals does not suffice to serve the learning needs of an organization if the knowledge and skills of individuals are not maintained and continually refined. While enhancing the numbers of well-qualified professionals

[4]Association of State and Territorial Health Officials. *State Public Health Employee Worker Shortage Report: A Civil Service Recruitment and Retention Crisis*. Available at http://www.astho.org/pubs/Workforce-Survey-Report-2.pdf

[5]Council of State and Territorial Epidemiologists. *National Assessment of Epidemiologic Capacity in Public Health: Findings and Recommendations*. Available at http://www.cste.org/pdffiles/ecacover1.pdf

through hiring is definitely one strategy in dealing with human capital concerns, it can be but one part of the solution. There is considerable agreement in the human capital literature regarding the importance of developing knowledge within the individual (Seetharaman et al., 2004; Youndt and Snell, 2004). Enhancing the skill sets of existing employees has to be a critical element of the strategy as public health looks to the future.

Effective bioterrorism surveillance and response requires surveillance professionals to acquire new knowledge and skills to complement their existing competencies. Bioterrorism surveillance requires knowledge of biological agents that could potentially be weaponized along with remedial measure to address the effects of each of these agents. Additionally, it requires the surveillance team to acquire information on a real-time basis from a variety of sources, analyze the information, and then prepare a response strategy in a timely manner. They also need to have skills related to interacting with the media and the public in the event of an attack. Bioterrorism preparedness, thus, requires public health professionals to not only acquire new knowledge but also to reorient their skills and business processes to deal with unconventional health emergencies. There is overwhelming evidence for the relationship between training and the value of human resources and subsequent performance of organizations (e.g., Bassi et al., 2002; Hatch and Dyer, 2004). While greater investments in training are critical for enhancing the everyday operations of public health organizations, it becomes absolutely imperative in the context of bioterrorism preparedness.

Bioterrorism preparedness requires surveillance professionals to operate and make decisions in highly unstable environments where uncertainty prevails and the available information could be imperfect, incomplete, or asymmetric. The skills and knowledge needed to be effective in such highly volatile situations are acquired either through experience with real events or through practice in simulations or other types of specific training programs. Training thus becomes critical to ensuring that individuals in decision-making capacities are able to rapidly interpret available information to contain and prevent catastrophes. The uncertainties and inadequacies of information that are likely to result from a bioterrorism event require individuals to make critical decisions without a comprehensive review of the situation (i.e., suboptimal information and decision-making structures). Additionally, information that is available may be in forms that can be understood and interpreted only by individuals with specific knowledge and skills. Also, valuable information may be overlooked because individuals did not have knowledge that allowed them to use it in a timely manner. Adequate preparation for a bioterrorism event through training should therefore extend not just to the key decision makers, but also to individuals across the information supply chain that would exist in the case of a real event.

The shortage of qualified professionals is a key factor that affects training and skill enhancement efforts in local public health departments. Public health organizations that are forced to operate with staff whose capacities

are stretched extensively do not have resources to deploy in training and education programs. Professionals who have extensive educational qualifications and field experience often do not have the time to develop and conduct training programs and exercises for others within the department. Even if training programs are available, there is often no surge capacity to sustain daily operations when staff members attend training sessions. An additional issue in many public health departments is the shortage of ancillary professionals such as trainers. Bioterrorism preparedness demands a new focus on maintaining and enhancing employee skills and knowledge. It means providing monetary and nonmonetary incentives to individuals who want to pursue higher education and investing in professional training for skill enhancement conducted either within or outside the organization.

11 Long-term programs for employee development and retention

Evidence from management literature suggests that organizations that are better at acquiring, developing, and deploying its human resources benefit from increased learning, reduced costs, and have a competitive advantage in the market (Hatch and Dyer, 2004). Therefore, a third and equally important aspect of human capital investment in public health is the creation of long-term programs for career development and employee retention. Human capital tends to be most valuable when it is specific to the organization and it continues to reside in the environment where it was originally developed (Hitt et al., 2001; Lepak and Snell, 1999). High employee turnover, especially of individuals in key positions results in poor institutionalized knowledge within the organization. While it is possible for organizations to enhance human capital by hiring qualified employees, there are adjustment costs involved in discovering the full potential of new employees and deploying them in positions where they can contribute significantly to the organization (Hatch and Dyer, 2004).

Part of the human capital of any organization that is particularly inimitable is the tacit knowledge held by individuals and groups. Tacit knowledge, acquired through experience and problem solving (Nelson and Winter, 1982), allows individuals to react to situations rapidly as compared to reviewing codified material and then preparing a response to the same situation. The availability of a large pool of tacit knowledge through employees with significant organizational and functional experience is likely to make a significant difference in response time and efficiency in a bioterrorism event. Asymmetries of information and rapidly evolving situations need decision makers that have significant expertise in dealing with large-scale emergencies. Additionally, because a paucity of research exists concerning the development of information structures formed in bioterrorism, such structures may include information that is high in knowledge specificity that can be interpreted and used only by individuals or groups

that have the appropriate expertise and skills, which often develop over a period of time through extensive on-the-job learning and experience. High turnover of employees with such experience constitutes *dysfunctional turnover* that leads to degeneration in the organization's productivity (Abbasi and Hollman, 2000).

Investing in the development and retention of human resources also means investing in the social capital of the organization represented through information channels and social networks that are built over time. Organizations can enhance their level of human and social capital by developing a pool of knowledgeable individuals, thus gaining complementary knowledge through the acquisition of information from relationships within and between organizations (Laszlo, 1996). As in all network types, strong social webs and reliable information channels play a significant role in developing preparedness and actually responding to a bioterrorism event. In an ideal situation, responding to bioterrorism or any other large-scale emergency would involve relying on highly institutionalized organizational processes. However, in many real situations, social connections between individuals are often the most reliable channel through which information is transmitted. These connections often serve to address gaps in information caused by technology glitches or other communication problems in an emergency.

Retaining highly qualified epidemiologists has been a challenge for public health organizations as their skills have been in increasingly high demand since 2001 with the heightened awareness of potential bio- and chemical terrorism. Public health recruits trained professionals from the same resource pool as private sector organizations that are able to offer higher salaries and benefits and potentially better career opportunities. According to the GAO (2004), barriers to recruiting and retaining public health professionals, especially epidemiologists, include noncompetitive salaries and a shortage of professionals. Research proves that investing in pay and benefits contribute significantly to reducing voluntary turnover in organizations (Shaw et al., 1998). Additionally, attention to hiring practices, managerial styles, recognition plans, and the workplace environment as a whole also contributes to the retention of employees (Abbasi and Hollman, 2000). Long-term investments in human resources are necessary for public health if it is to consistently maintain high levels of surveillance for both bioterrorism as well as the more common and prevalent health issues.

A proactive stance towards bioterrorism preparedness involves both short- and long-term investments in human capital. Federal funding has served to address immediate concerns of worker shortages and infrastructure issues. But it will be difficult to sustain these efforts in the long run if funding strategies are based solely on grants that support programmatic surveillance and intervention efforts. What is needed is a more comprehensive strategy of employee hiring, development, and retention that, over time, will not just enhance the efficacy of bioterrorism surveillance but recreate business processes and organizational structures to systemically

address the complexities of bioterrorism. A long-term strategy is needed at the federal level that would be percolated down to the levels of local public health departments if there is to be a nationwide change in the resource structure and allocation within public health. None of the approaches discussed in this chapter represent radically new ideas—they have all been successfully adopted by business organizations over the years and merit consideration in the public health context. Public health organizations would do well to adopt human resource best practices from business organizations in order to enhance the efficacy of their current and future operations, and be better prepared to deal successfully with a large-scale emergency such as a bioterrorism event.

12 Conclusion

Preparedness for bioterrorism reflects an environment that demands a significant change in organizational information and resource structures. Existing information structures vary in the relevance, value, and availability of information acquired during investigation processes. Variability in surveillance and investigation processes also occurs with inconsistencies in human capital investments in personnel, training, and human resource development. As a result, social systems vary in their ability to effect decisions that optimize surveillances outcomes. Although traditional surveillance systems in departments of public health are currently being upgraded to improve the likelihood of detecting and responding to a bioterrorism event, a clear strategy does not exist for designing surveillance systems that are robust to conditions that create suboptimal information structures and misappropriate investments in human capital.

The threat of bioterrorism offers public health a new opportunity to deploy advanced information technologies in surveillance processes. However, in developing and implementing surveillance systems, critical issues related to the features and intended use of the system should be addressed in advance. Because surveillance strategies in epidemiology are heavily influenced by the expertise of individual professionals, the resulting system can be designed with a range of functionality, either as resource systems to support the social and informational needs in existing processes or as advanced governance systems that introduce change in organizational processes and govern the operation of integrated systems. The emphasis in design is the need to anticipate the intended goal of introducing technology into existing processes.

An analysis of organizational processes and structural relationships in the environment is also needed to effectively evaluate real costs against realized benefits in surveillance. This analysis should include an assessment of losses that can be offset through preventive measures. With this knowledge, the cost of acquiring pre-event information can be justified in the design of

predictive surveillance systems. In addition, an assessment of the capacities needed to optimize surveillance strategies can reveal the gaps in organizational structures and processes that support the human element in the surveillance process. Addressing these gaps by expanding and enriching the knowledge workforce necessary for surveillance and using information systems to overcome natural limitations introduced through bounded rationality and judgmental errors in the human interface will significantly contribute to an environment that is adequately prepared for the dynamics unleashed by acts of bioterrorism.

13 Discussion questions

This chapter has analyzed changes in information structures in public health surveillance and the human capital investments that are necessary to develop and sustain enhanced surveillance processes in the context of bioterrorism preparedness. If public health is to adequately meet the challenges produced by the threat of bioterrorism, important questions have to be answered and critical concerns addressed such as:

- How can processes in public health be implemented that are robust to the dynamic surveillance conditions present in bioterrorism preparedness?
- What types of advanced information technologies can be designed that mediate unfavorable appropriations of information and surveillance resource systems?
- How can advanced technologies be used to address changing information structures and to address the greater need for better distribution and dissemination of relevant information?
- What short- and long-term investments need to be made to expand and maintain the public health workforce?
- What long-term strategies need to be developed that deploy the full potential of information technologies in the surveillance process?

References

Abbasi, S.M., K.W. Hollman (2000). Turnover: the real bottom line. *Public Personnel Management* 2(3), 333–342.

Amit, R., P. Shoemaker (1993). Strategic assets and organizational rent. *Strategic Management Journal* 14(1), 33–46.

Balakrishnan, A., A. Whinston (1991). Information issues in model specification. *Information Systems Research* 2(4), 263–286.

Bassi, L.J., J. Ludwig, D.P. McMurrer, M. Van Buren (2002). Profiting from learning: firm-level effects of training investments and market implications. *Singapore Management Review* 24(3), 61–78.

Beaglehole, R., B.T. Kjellström (1993). *Basic Epidemiology*. World Health Organization, Geneva.

Bhopal, R.S. (2002). *Concepts of Epidemiology*. Oxford University Press, New York, NY.
Bontis, N., J. Fritz-enz (2002). Intellectual capital ROI: a causal map of human capital antecedents and consequences. *Journal of Intellectual Capital* 3(2), 223–247.
Carley, K.M., L. Zhiang (1997). A theoretical study of organizational performance under information distortion. *Management Science* 43(7), 976–997.
CDC. (2001). The Public Health Response to Biological and Chemical Terrorism: Interim Planning Guidance for State Public Health Officials. United States Department of Health and Human Services, Washington, DC, pp. 1–106.
Choudhury, V., J.L. Sampler (1997). Information specificity and environmental scanning: an economic perspective. *MIS Quarterly* 21(1), 25–53.
Coase, R. (1937). The nature of the firm. *Economica* 4(November), 386–405.
Conrath, D.W. (1967). Organizational decision making behavior under varying conditions of uncertainty. *Management Science* 13(8), 487–500.
Fritz-enz, J. (2000). *The ROI of Human Capital*. Amacom Books, New York, NY.
Fulk, J., C. Steinfield (1990). The theory imperative, in: C. Steinfield, Janet Fulk (eds.), Organizations and Communication Technology, Sage Publications, Newbury Park, CA, pp. 13–25.
General Accounting Office (2004). *Bioterrorism: Information Technology Strategy Could Strengthen Federal Agencies' Abilities to Respond to Public Health Emergencies*. United States Government Printing Office, Washington, DC.
Grossman, S.J. (1981). An introduction to the theory of rational expectations under asymmetric information. *The Review of Economic Studies* 48(4), 541–559.
Hatch, N.W., J.H. Dyer (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal* 25, 1155–1178.
Hayek, F.A. (1945). The use of knowledge in society. *The American Economic Review* 35(4), 519–530.
Hitt, M.A., L. Bierman, K.M. Shimizu, R. Kochhar (2001). Direct and moderating effects of human capital on strategy and performance in professional firms: a resource-based perspective. *Academy of Management Journal* 44(1), 13–28.
Keen, P.G.W. (1986). *Competing in Time: Using Telecommunications for Competitive Advantage*. Ballinger Publishing Company, Cambridge.
Kmietowicz, Z.W., A.D. Pearman (1983). *Decision Theory and Incomplete Knowledge*. Gower Publishing, Ltd, Aldershot, Hampshire, UK.
Koch, M.J., R.G. McGrath (1996). Improving labor productivity: human resource management policies do matter. *Strategic Management Journal* 17(5), 335–354.
Lado, A.A., M.C. Wilson (1994). Human resource systems and sustained competitive advantage: a competency based perspective. *Academy of Management Review* 19(4), 699–727.
Laszlo, E. (1996). *The Systems View of the World: A Holistic Vision for Our Time*. Hampton Press, Cresskill, NJ.
Lepak, D., S. Snell (1999). The human resource architecture: toward a theory of human capital allocation and development. *Academy of Management Review* 24, 31–48.
Lichtenstein, S., P. Slovic, B. Fischhoff (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory* 4(6), 551–578.
March, J.G., H.A. Simon (1958). *Organizations*. John Wiley, New York, NY.
Monteiro, E., V. Hepso (2002). Purity and danger of an information infrastructure. *Systemic Practice and Action Research* 15(2), 145–166.
Moore, J.C., H.R. Rao, A.B. Whinston (1994). Multi-agent resource allocation: an incomplete information perspective. *IEEE Transactions on Systems, Man, and Cybernetics* 24(8), 1208–1219.
Nelson, K.E., C.M. Williams, N.M.H. Graham (2005). *Infectious Disease Epidemiology: Theory and Practice*. Jones and Bartlett, Boston, MA.
Nelson, R., S. Winter (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA.
C.A. O'Reilly, III (1990). The use of information in organizational decision making: a model and some propositions, in: B.M. Staw, L.L. Cummings (eds.), Information and Cognition in Organizations, JAI Press Inc, Greenwich, CT, pp. 89–125.

Pfeffer, J. (1981). *Power in Organizations*. Pitman, Marshfield, MA.
Pfeffer, J., G. Salancik (1977). Administrator effectiveness: the effects of advocacy and information on achieving outcomes in an organizational context. *Human Relations* 30, 641–656.
Pfeffer, J., G. Salancik (1978). *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, New York, NY.
Rasmusen, E. (2001). *Games and information: an introduction to game theory*. Blackwell Publishers, Malden, MA.
Rothschild, M., J. Stiglitz (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *The Quarterly Journal of Economics* 90(4), 629–649.
Seetharaman, A., K.L.T. Low, A.S. Saravanan (2004). Comparative justification on intellectual capital. *Journal of Intellectual Capital* 5(4), 522–539.
Shaw, J.D., J. Delery, N. Gupta (1998). An organizational-level of analysis of voluntary and involuntary turnover. *Academy of Management Journal* 41(5), 511–525.
Simon, H.A. (1981). *The Sciences of the Artificial*. The MIT Press, Cambridge, MA.
Smith, K.G., C.M. Grimm, M.J. Gannon (1992). *Dynamics of Competitive Strategy*. Sage Publications, Newbury Park, CA.
Staw, B.M. (1981). Threat-rigidity effects in organizational behavior: a multilevel analysis. *Administrative Science Quarterly* 26(4), 501–524.
Teutsch, S.M., R.E. Churchill (2000). Considerations in planning a surveillance system, in: S.M. Teutsch, R.E. Churchill, R. Elliott (eds.), Principles and Practice of Public Health Surveillance, Oxford University Press, New York, NY, pp. 17–29.
Weber, M. (1987). Decision making with incomplete information. *European Journal of Operational Research* 28, 44–57.
Weill, P., M. Broadbent (1998). *Leveraging the New IT infrastructure*. Harvard Business School Press, Cambridge, MA.
Williamson, O. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. Free Press, New York, NY.
Williamson, O. (1985). *The Economic Institutions of Capitalism*. Free Press, New York, NY.
Youndt, M.A., S.A. Snell (2004). Human resource configurations, intellectual capital and organizational capital. *Journal of Managerial Issues* 16(3), 337–360.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 13

PulseNet Provides Early Warning for Foodborne Disease Outbreaks[1]

Bala Swaminathan, Brenda L. Brown, Robert Long and Peter Gerner-Smidt

Foodborne and Diarrheal Diseases Branch, Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Mail Stop CO3, Atlanta, GA 30333, USA

Abstract

The Centers for Disease Control and Prevention (CDC), in partnership with the Association of Public Health Laboratories, USA, established PulseNet USA in 1996, as an early warning system for foodborne disease outbreaks in the USA. Four bacterial pathogens (*E. coli* O157:H7, *Salmonella*, *Shigella*, and *Listeria monocytogenes*) are routinely monitored by PulseNet by their DNA "fingerprints." The national library of these DNA fingerprints is maintained at CDC in SQL databases; a customized version of BioNumerics (Applied Maths, Sint-Martens-Latem, Belgium) software enables rapid normalization and comparison of DNA fingerprint patterns. Information security procedures are continually evaluated and updated to make sure that time-sensitive information is not accessed by unauthorized personnel. As soon as a cluster of clinical isolates of a pathogen with indistinguishable DNA fingerprints is identified, the patients in the cluster are rapidly interviewed. If preliminary findings indicate potential epidemiologic links between patients, that cluster is designated as an outbreak and a detailed investigation is initiated. This strategy has enabled the U.S. public health system to identify foodborne disease outbreaks that would not have been previously identified and facilitated outbreak identification and recall of a food product with as few as two cases in a cluster. The PulseNet network is now being replicated in Canada, Europe, Latin America, and the Asia Pacific Region. A Memorandum of Understanding formally recognizing the collaboration between PulseNet Canada and

[1]Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or the U.S. Department of Health and Human Services.

PulseNet USA was signed in Winnipeg, Canada, by the Canadian Health Minister and the U.S. Ambassador to Canada on August 12, 2005.

1 Introduction

Foodborne infections continue to pose major threats to public health in the USA with an estimated 76 million cases annually that result in 325,000 hospitalizations and 5000 deaths (Mead et al., 1999). Most foodborne infections have no known connection to each other, i.e., they are sporadic. However, when two or more people become ill with the same infection or intoxication after consuming the same contaminated food, that incident is termed a foodborne disease outbreak. Some outbreaks that are caused by contamination of widely distributed processed foods may affect very large number of people (e.g., more than 250,000 sickened by pasteurized milk contaminated with *Salmonella* in the Midwestern USA in 1984). Most of our understanding of the microorganisms (bacteria, viruses, and parasites) that are most likely to cause foodborne disease, the most common food vehicles of infection, and the way in which foods become contaminated comes from investigations of foodborne disease outbreaks (Batz et al. 2005).

The Centers for Disease Control and Prevention (CDC), an agency of the U.S. Department of Health Human Services, is charged with protecting public health in the USA. Among its many activities, CDC and its public health partners in the states, counties, and cities detect and investigate foodborne disease outbreaks. Once public health officials identify the food vehicle of an outbreak, they work closely with the federal and state food regulatory agencies to trace the contamination to its source so that it can be removed from distribution, thus preventing illness in more people from the same contaminated food. The food regulatory agencies and the food industry further explore ways to prevent recurrence of the same type of outbreak. Where appropriate, the food regulatory agency may promulgate new regulations or amend existing regulations to prevent a specific type of outbreak from happening again, e.g., specifying higher cooking temperatures for hamburgers at fast-food restaurants, treatment of alfalfa seeds with chlorine disinfectant before using the seeds for sprouting, or requiring pasteurization of fruit juices. The food industry also may perform a critical review of its food production, handling, processing, storage, and distribution procedures to identify critical steps where more rigorous control need to be exercised.

2 DNA fingerprinting-based foodborne disease surveillance—PulseNet

As may be apparent from the previous description, the early detection of outbreaks allows pubic health officials to begin early investigation of

outbreaks so that the magnitude of the outbreak is kept as small as possible. Because most foodborne infections are sporadic, public health officials need a reliable means of identifying those cases that may be part of a common source outbreak and distinguishing them from sporadic cases that may occur in the same geographic locations and during the same time period. PulseNet USA, the national molecular subtyping network for foodborne disease surveillance, was established in 1996 to facilitate early outbreak recognition, discriminate outbreak-related cases of human illness from temporally and geographically related sporadic cases, and provide independent confirmation of the epidemiologically implicated food source (Swaminathan et al., 2001). The network includes all 50 state public health laboratories, county, and city public health laboratories from large metropolitan areas and counties (e.g., New York city, Los Angeles county), laboratories of the U.S. Food and Drug Administration (FDA), and the Food Safety and Inspection Service (USDA/FSIS) of the U.S. Department of Agriculture. PulseNet USA accomplishes its objective by routinely performing DNA fingerprinting of foodborne and diarrheal disease-causing bacteria isolated from ill persons in the USA. Clinical diagnostic laboratories that recover disease-causing bacteria from patient specimens forward them to their local (in the case of large cities and counties) or state public health laboratory. At the state or local public health laboratory, the bacteria are "fingerprinted" by a method called macrorestriction analysis. In this method, the DNA of the bacteria is allowed to react with a special enzyme called restriction endonuclease (RE). Many REs have been discovered in the past 25 years. Each RE recognizes a very specific four to eight base sequence on the target DNA and cuts within the target sequence at a very specific location. In the macrorestriction process, the restriction enzyme is selected such that the target bacterial DNA has only 10–40 recognition sites for that enzyme. After reaction with the enzyme, the bacterial DNA is cut into 11–41 large pieces of DNA fragments. The total length of DNA in bacteria such as *Escherichia coli* O157:H7 or *Salmonella* is ~5 million nucleotides. On treatment of *E. coli* O157:H7 or *Salmonella* DNA with an appropriately chosen restriction enzyme, DNA fragments in the range of 20,000–1,000,000 nucleotides are obtained. The DNA fragment mixture is then separated by a process called pulsed-field gel electrophoresis (PFGE), in which the DNA mixture is placed in an electric field and the direction of current flow is changed (pulsed) every few seconds to allow even the large DNA fragments to migrate through an agarose gel matrix at a speed inversely proportional to their size. Larger DNA fragments move slower than the smaller DNA fragments. After the smallest DNA fragments have traversed through 90% of the length of the agarose matrix (18–28 h), the separated DNA fragments are stained with a dye to render them fluorescent when viewed under ultraviolet light. The image of the separated DNA fragments fluorescing under ultraviolet light is captured digitally. Each bacterial strain of a species (*E. coli* O157:H7, *Salmonella*, or *Listeria*)

has a unique DNA fingerprint that can be used to discriminate it from another strain of the same species in a manner similar to the DNA fingerprinting used on humans for forensic purposes (Fig. 1).

When several people become infected with the same strain of a disease-causing organism (as determined by PFGE) over the same period of time (called a PulseNet disease cluster), there is a good reason to suspect that they may be acquiring the infection from the same source. Epidemiologists mobilize their resources and interview the case-patients in the PulseNet cluster to determine if there are any common exposures between them. If the epidemiologic investigation reveals evidence that the cases have a common exposure, the cluster is elevated to an outbreak status and a thorough investigation is launched to identify the source of the infection. When a food or environmental source of the outbreak is identified by epidemiologic analysis, appropriate samples of the implicated source are examined for the presence of the bacteria that is causing the outbreak. When the bacteria are isolated, they are subjected to macrorestriction analysis to determine if their DNA fingerprint matches that of the bacteria from the patients.

Since its inception in 1996, PulseNet USA has identified many foodborne disease outbreaks that would have gone undetected before, served as an effective early warning system for foodborne disease outbreaks, and facilitated accelerated epidemiologic investigation of outbreaks by providing critical laboratory support for ongoing investigations. Since 1996, the

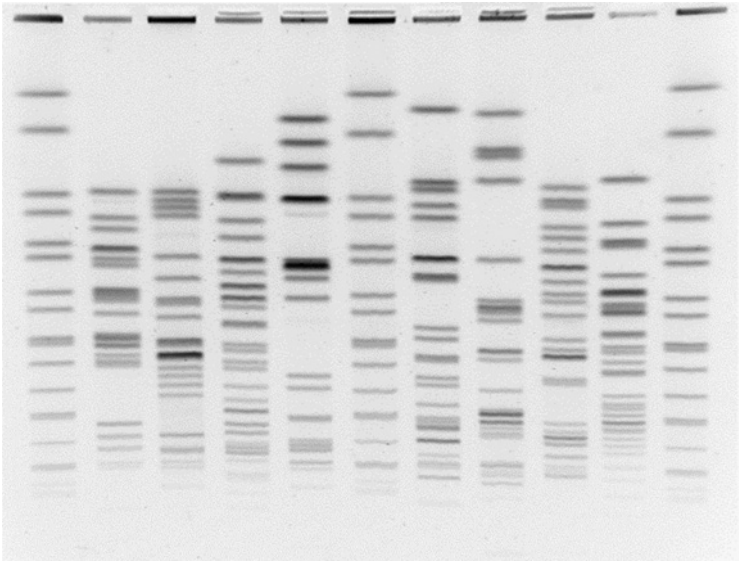


Fig. 1. PFGE profiles of Shiga-toxin producing *E. coli* O157:H7 digested with the restriction enzyme *Xba*I (lanes 1, 6, and 11 contain the PulseNet reference standard *Salmonella* Brenderup H9812).

network has grown steadily from 5 participating laboratories to 70 participants in 2005. Full national participation of all 50 states was achieved in 2001. Also, the number of patterns submitted to the national database on an annual basis has increased exponentially from 196 STEC O157 in 1996 to almost 40,000 of 5 organisms in 2004.

PulseNet relies on a decentralized network of public health and food regulatory laboratories nationwide for its disease cluster detection capabilities. Therefore, the participating laboratories must perform the DNA fingerprinting of foodborne disease-causing bacteria in a timely manner and submit the patterns and associated patient information to the national databases. Also, because PulseNet relies on comparison of DNA fingerprint patterns from many laboratories, it is imperative for the participating laboratories to follow the detailed standardized protocols developed by CDC without any deviation. Extensive training, certification, and proficiency testing procedures have been developed and implemented for PulseNet to assure inter-laboratory comparability of DNA fingerprint patterns with a high degree of confidence.

PulseNet USA uses integrated networking technology to provide a system of information and data sharing, information security procedures, cutting-edge application software, web servers, application server, and relational database engines (Fig. 2).

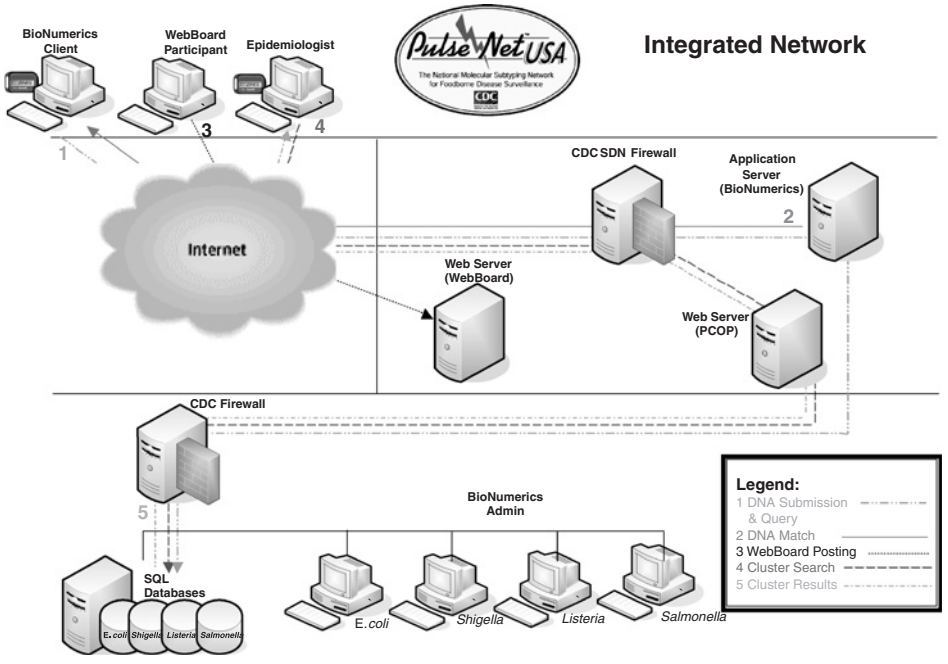


Fig. 2. PulseNet electronic infrastructure and dataflow.

3 PulseNet information security

Implementation of the PulseNet USA system is governed by information security policies and procedures from the National Institute of Standards and Technologies (NIST), a federal technology agency that works with industry to develop and apply technology, measurements, and standards. Specifically, publications NIST SP 800-53 and Federal Information Processing Standard (FIPS) Publication 200 are used. The Department of Health and Human Services and CDC interpret and diligently enforce a high standard of information security through a certification and accreditation (C&A) process for all systems on the network. As new security procedures are developed, they are validated and implemented on the systems. We currently have a C&A process that re-evaluates all systems every 3 years for re-certification.

4 Databases

PulseNet USA uses an industry-standard relational database engine, standard query language (SQL), for its national databases. The DNA fingerprint and demographic data are formatted using XML to allow sharing data between the PulseNet participants via the PulseNet national database.

PulseNet uses the BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium) to normalize the DNA fingerprint patterns against a global reference PulseNet standard (Hunter et al., 2005) and compare them. The normalization of PFGE patterns generated by PulseNet participating laboratories against a global reference standard pattern minimizes and eliminates any difference between DNA fingerprint patterns due to routine experimental variations and differences between different pieces of equipment used to do the PFGE analysis. BioNumerics is a commercial off-the-shelf client–server application that has been customized for PulseNet USA. Customized client, server, and administrator versions of BioNumerics are used by PulseNet.

The client version of BioNumerics is used by state and county public health laboratories and federal regulatory agencies to analyze the DNA fingerprints that are submitted to the laboratory, and uploaded to the national databases at CDC in Atlanta. Each client uses a proprietary database, located in their local computer, or a Microsoft Access database located in their local network. After a technologist at a participating PulseNet laboratory has been certified as competent in both standardized DNA fingerprinting and computer-assisted analysis of the PFGE patterns, he/she may connect to the national database and upload the DNA fingerprint, along with demographic information using a set of established security procedures. Submitters are able to query the national database and compare patterns from their local database against the patterns in the national database. If the

comparison yields a match to existing patterns in the national database and suggests a disease cluster, the submitter from the client laboratory notifies the PulseNet USA database team at CDC and posts the information to the PulseNet listserv. The PulseNet listserv is the central communication tool used to notify the PulseNet community of a potential outbreak.

The administrator version of BioNumerics is used by the national PulseNet database team located at CDC in Atlanta. It allows each PulseNet database administrator to connect directly to the database via open database connection (ODBC). Also, it provides additional functionalities for modifying the information that has been uploaded to the national database. It allows compilation of statistical information on the data in the national database, which can be used for reports for management and disseminated to the PulseNet participants.

The BioNumerics server is the front end of the central collection of all analyzed DNA fingerprint and demographic information from the public health laboratories and regulatory agencies. The server uses ODBC to link to the SQL databases and allows sharing database information for querying and comparisons.

5 The PulseNet listserv

The central communication tool used by PulseNet USA is the listserv. It is a threaded discussion board with e-mail capability, and it allows rapid information exchange within the PulseNet community, which includes the PulseNet database team, CDC epidemiologists, laboratorians in participating laboratories, and state and local health department epidemiologists. Currently, the PulseNet listserv is based on the WebBoard application (Akiva Corporation, Carlsbad, CA). It is organized into conferences or message topics for each organism and other topics of interest to the PulseNet community. Postings to the listserv 1/4 include information that has been gathered after a PulseNet participant has uploaded DNA fingerprint and demographic information to the PulseNet national database and he has executed queries and comparisons against existing data in the database. For example, a PulseNet-certified laboratorian may upload data to the national database and run a query that yields a potential match to existing data. He/she will post these results in an appropriate conference on the listserv, informing other listserv participants of the newly discovered disease cluster (Fig. 3). The listserv posting generates an e-mail to subscribers to the conference. Upon reading the new posting, the other participants review their recent DNA fingerprinting information and post responses to the initial posting. Also, such postings and responses encourage laboratories that may have a temporary backlog of isolates for DNA fingerprinting to assign high priority to the pathogen implicated in the disease cluster reported on the listserv. Currently, PulseNet USA has over 350 subscribers to the listserv.

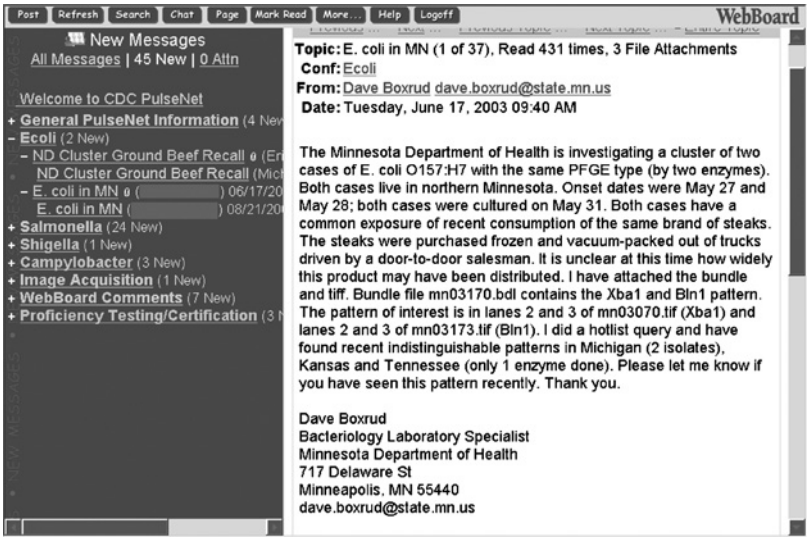


Fig. 3. An example of a posting on the PulseNet listserv that informs participants about a foodborne disease cluster in Minnesota.

6 The PulseNet web portal

In the beginning of 2006 the PulseNet Central Online Portal (PCOP) or PulseNet web portal, another information sharing tool, will be integrated in the current PulseNet configuration. The PulseNet web portal is a secured web server that will allow a real-time view to the PulseNet community. It will enable the user to view current cluster information on a specific pathogen from the PulseNet database in a table or a graphic format without accessing the BioNumerics software. The listserv is intended to notify the subscribers of PulseNet activity, whereas the web portal will allow them to get a quick and simplified view of the underlying information.

7 Certification and proficiency

PulseNet USA has an organism-specific certification and proficiency process. The PulseNet quality assurance and quality control (QA/QC) program was developed to ensure the quality and integrity of the results obtained with the standardized PFGE techniques used to subtype foodborne bacterial pathogens. It includes performing PFGE gel analysis in the laboratory and DNA fingerprint analysis using BioNumerics client software, initially for certification, and then annually for proficiency. Certification is open to laboratorians working at state and county public health laboratories

and regulatory agencies. After the laboratorians have been certified, they participate in proficiency testing once a year.

Currently, PulseNet USA has over 135 certified laboratorians. They are certified for one or more of the pathogens that are under surveillance in PulseNet.

8 Successes of PulseNet

From its inception, PulseNet has proved to be an effective early warning system for foodborne disease outbreaks. Among the first successes of PulseNet was the discovery of a cluster of *E. coli* O157 infections in Colorado. The Colorado State Public Health Laboratory, which had just started routine DNA fingerprinting of foodborne disease-causing bacteria using PulseNet protocols in July 1997, identified a cluster of eight case-patients all of whom had been infected with *E. coli* O157 with the same DNA fingerprint. This information was forwarded to the epidemiologists in the Colorado State Health Department who opened an investigation, interviewed the patients who had been identified as part of the cluster, and administered a food history questionnaire. The epidemiologic investigation revealed an association of the illnesses with frozen ground beef patties. Further investigation and traceback by the officials of the USDA/FSIS identified a meat processing factory in Nebraska as the source of the outbreak. The investigation led to the voluntary recall of 11 million kg of possibly contaminated ground beef, the largest recall in USA history (Table 1). Other investigations in which PulseNet played a prominent role and large volumes of food were recalled are also shown in the table. From a public health standpoint, these recalls enabled the public health and regulatory authorities to remove contaminated food with potential to cause serious, possibly life-threatening, disease from the distribution channels. Elbasha et al. (2000) analyzed the costs and benefits of Colorado's PulseNet surveillance system using the previously mentioned outbreak for their case study. They concluded that if 15 cases of *E. coli* infections were averted by the recall of the 11 million kg of potentially contaminated beef in this outbreak, the

Table 1
Large USA food recalls in which PulseNet played a prominent role

Year	Pathogen	Food	Recall (million kg)
1998	*Listeria monocytogenes*	Hot dogs, deli meat	16.0
2002	*L. monocytogenes*	Ready-to-eat poultry products	12.4
1997	STEC O157:H7	Frozen ground beef	11.3
2002	STEC O157:H7	Ground beef	8.4
1998	*Salmonella*	Toasted oats cereal	1.4

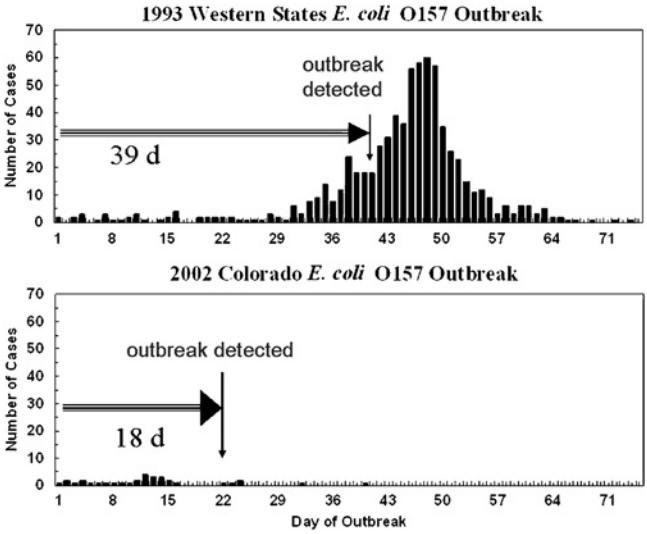


Fig. 4. Epidemic curves for two outbreaks of Shiga-toxin producing *E. coli* O157:H7.

Colorado system would have recovered all costs for the 5 years of start-up and operation.

By detecting disease clusters when they are small, PulseNet frequently permits early detection of outbreaks when just a few people have been infected. Epidemiologists can then begin their work early to identify the outbreak source so that it can be removed from the distribution. PulseNet's early detection capability is best illustrated in Fig. 4 in which the epidemic curves of the 1993 western states' fast-food hamburger-associated outbreak (before PulseNet was implemented) and a 2002 ground beef-associated outbreak in Colorado are compared (Gerner-Smidt et al., 2005). PulseNet has greatly increased the sensitivity of public health surveillance for foodborne diseases. By identifying a disease cluster early, morbidity and mortality in the population due to foodborne disease outbreaks can be significantly reduced.

A sample PulseNet listserv communication illustrates the small number of cases that may suffice to detect a cluster through PulseNet (Fig. 3). In this instance, only two case-patients were identified as part of a cluster, and concurrent interviews of those patients revealed a possible association with vacuum-packed steaks sold door-to-door. The listserv posting on this cluster and the search of the national PulseNet database for *E. coli* O157:H7 led to the identification of a few additional case-patients infected with the bacteria with the outbreak DNA fingerprint. The information obtained from these case-patients allowed public health officials to positively link the vacuum-packed steaks with the outbreak. A recall of the contaminated product was initiated within 12 days of the first listserv posting on the disease cluster.

9 International

The success of PulseNet USA has generated a great deal of excitement and enthusiasm in the international public health community for adopting the PulseNet approach in their countries and regions. PulseNet Canada, which began operations in 1999, routinely shares DNA fingerprints of human disease isolates of foodborne disease-causing bacteria in real-time with PulseNet USA. In addition, public health officials in both countries have access to each other's threaded listserv. This has enabled both countries to identify outbreaks that have affected patients in Canada and the USA and jointly identify the source of infection. A formal memorandum of understanding authorizing the sharing of DNA fingerprints of foodborne pathogens (clinical isolates) and associated information in real-time between the two countries was signed in August 2005.

PulseNet Europe (30 countries), PulseNet Asia Pacific (13 countries and areas), and PulseNet Latin America (13 countries) are in various stages of development and implementation. The various international PulseNet networks are coordinated through PulseNet International, a virtual network of networks. A steering committee composed of representatives of each of the regional networks coordinates activities and communication between the regional networks. All participants of the different PulseNet networks have agreed to use standardized protocols for DNA fingerprinting and analysis of the fingerprints to ensure comparability of the DNA fingerprint patterns.

10 Summary

PulseNet is an effective early warning system for foodborne disease outbreaks in the USA. PulseNet uses multiple channels of secure communication systems to enable real-time sharing of information and DNA fingerprints of foodborne disease-bacteria between foodborne disease epidmiologists, laboratory scientists, and food regulatory agency officials. Early detection of diffuse outbreaks of foodborne disease through PulseNet has enabled health officials to investigate outbreaks early to identify the source of contamination and, with the help of food regulatory officials, remove the contaminated foods from distribution channels, thus preventing additional illness and possible deaths. The PulseNet concept is now being replicated in different parts of the world to enable coordinated international foodborne disease surveillance and outbreak detection in the future.

11 Questions

1. What is the difference between a sporadic case of infection and an outbreak of infections?

2. Why do the CDC emphasize the investigation of foodborne disease outbreaks as part of their disease prevention strategy?
3. Describe the role of PulseNet in foodborne disease surveillance and outbreak investigations.
4. List some information security concerns for a network like PulseNet. What strategies should be adapted to address the concerns you have identified?

References

Batz, M.B., M.P. Doyle, G. Morris, Jr., J. Painter, R. Singh, R.V. Tauxe, M.R. Taylor, D.M. Wongthe Food Attribution Working Group (2005). Attributing illness to food. *Emerging Infectious Diseases* 11(7), 993–999.

Elbasha, E.H., T.D. Fitzsimmons, M.I. Meltzer (2000). Costs and benefits of a subtype-specific surveillance system for identifying *E. coli* O157:H7 outbreaks. *Emerging Infectious Diseases* 6(3), 293–297.

Gerner-Smidt, P., J. Kincaid, K. Kubota, K. Hise, S.B. Hunter, M.A. Fair, D. Norton, A. Woo-Ming, T. Kurzynski, M.J. Sotir, M. Head, K. Holt, B. Swaminathan (2005). Molecular surveillance of Shiga toxigenic *E. coli* O157 by PulseNet USA. *Journal of Food Protection* 68(9), 1926–1931.

Hunter, S.B., P. Vauterin, M.A. Lambert-Fair, M.S. Van Duyne, K. Kubota, L. Graves, D. Wrigley, T.J. Barrett, E. Ribot (2005). Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard. *Journal of Clinical Microbiology* 43(3), 1045–1050.

Mead, P.S., L. Slutsker, V. Dietz, L.F. McCaig, J.S. Bresee, C. Shapiro, P.M. Griffin, R.V. Tauxe (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases* 5(5), 607–625.

Swaminathan, B., T.J. Barrett, S.B. Hunter, R.V. Tauxe (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerging Infectious Diseases* 7(3), 382–389.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 14

Government Agency Interoperation in Security Applications

Nabil R. Adam, Aabhas V. Paliwal and Vijay Atluri
RUTGERS University, CIMIC, Ackerson Hall, Newark, NJ 07102, USA

Soon Ae Chun
City University of New York, Staten Island, NY 10314, USA

Jim Cooper and John Paczkowski
Operations and Emergency Management, Port Authority of New York and New Jersey, USA

Christof Bornhövd, Ike Nassi, Joachim Schaper and John Ellenberger
SAP Labs, LLC, Palo Alto Research Center, 3475 Deer Creek Road, Palo Alto, CA 94304, USA

Abstract

Incident management for homeland security requires the accurate up-to-date situational awareness for rapid engagement of first responders at state and local levels. Major challenges for agile and effective responses include: (1) identifying and visualizing the right type of information that is relevant to the incident to see the coherent picture of the incident, (2) identifying resources to handle the incidents, including agencies, specialists, other personal and resources specific to a given alert type (e.g., fire, hazmat spills), (3) disseminating appropriate information and tasks to the right level of responders and to the public in an appropriate format to their available devices. We present a semantic incident management framework which uses a common incident ontology that captures the concepts of different incident types and their relationships among different incidents. The concepts are tied to the information resources such as textual description of incidents, audio and video clips from the incident scene. This framework allows: (1) dynamic composition of customized information, relevant resources, reports, and models

based on the nature and location of the alert; (2) automated manifestation of the modality and format of the information based on the recipient's role and device, and (3) automated composition of alert components and models through Semantic Web and Semantic Web Services (SWS). The composition and dissemination adheres to the National Incident Management System (NIMS) and the National Response Plan (NRP) protocols and has been implemented using the Common Alerting Protocol (CAP) and the Ontology Web Language for Services (OWL-S).

1 Introduction

Government agencies need to form an ad hoc global virtual organization and collaborate in order to handle homeland security-related incident management. This virtual security team consists of people from executive level, management level, and responder level from different organizations that are geographically and functionally distributed. The executive people may include political and government leaders, agency and organization administrators and department heads, incident commanders for either a specific area and single incident or multi-agency coordination. The management level personnel often includes unit leaders, technical specialists, strike team and task force leaders, single resource leaders, and field supervisors. Finally, the responder level emergency response providers and disaster workers include emergency medical service personnel, firefighters, medical personnel, police officers, public health personnel, public works/utility personnel, and others.

This team-oriented virtual "agency" needs to make accurate decisions in a timely manner for an effective incident management to reduce the severity and damage. Decision on facilities may include the selection of facilities and sites for command post, evacuation, casualty collection sites, and transportation sites (e.g., heliports). Decisions on the resources usually rely on resource needs and available resources in the incident areas and the resource capacities of local agencies that are specified in the emergency operations procedures. The incident management also needs to consider current incident situations, objectives, hazard types, and hazard severities. These decision-making tasks related to homeland security are highly decentralized given the apparent diversity of agencies and information sources. A key challenge for the virtual government entity for effective decision-making for rapid responses to threats to homeland security is to consider data from diverse sources in different formats. Effective assimilation, exchange, and dissemination of information are vital for homeland security wherein it is important for agencies to communicate in a way where information can be fused and exchanged in a more efficient manner.

Second, command and control of incident management is based on incident commander's situational observations, the situational reports by different agencies, or the incoming data from sensors, if any. The data volume from various sources can be overwhelming and the interpretation of data and information needs to be done efficiently. Thus, another challenge is a rapid analysis and interpretation of voluminous real-time data from different sources. In order to achieve this, the data from different sources needs to be shared, fused, and analyzed as it becomes relevant without a prior data-sharing agreement among different agencies.

Finally, the decisions made may call for some actions (tasks) by various team members. The proper tasks, agencies (representative person) that can accomplish the tasks, and the information needed for tasks have to be identified and disseminated to each agency. Putting the information and tasks together manually jeopardizes the timely response to an incident. Thus, there should be an automated way to identify and compose tasks and to disseminate these tasks and relevant information to be executed by different team members.

Therefore, information interoperation in homeland security application is required to support complex decision-making process that involves multi-agencies (horizontal coordination) that may encompass several jurisdictions, multi-layer (vertical coordination) that may involve the executive team, objectives, special forces team, resources, and services within an agency. In order for each individual agency to work toward the common goal of stabilizing the incident and protect life, property and the environment, it needs to have the right type of information and services at their hands. To illustrate these challenges, consider the following scenario of an incident.

> Scenario: *At 9:30 am, it was reported that a truck carrying a chemical substance on route to NY Lincoln tunnel is missing. At 10:01 am, a truck accident on highway was reported and an unknown chemical spill was reported. A hazmat team needed to be brought in. At 10:17 am, the chemical spill is identified as toxic chlorine and immediate area residents needed to be alerted for evacuation. The police department needed risk assessment information and how the wind may carry the chemical to identify immediate risk areas, evacuation facilities, and hospitals. Also, volunteer information is needed.*

As seen in this scenario, the incident management has the following characteristics:

- The incident characterization is continuous and dynamic as more information is available throughout the management of the incident.
- As the situations change constantly, the information needs and resource requirements are changing as well. The agency and participants to handle the incident change as time progresses, and the resource requirements also change as the situation of the incident develops.
- There should be an efficient dissemination of incident messages that include incident-related information, services, and tasks.
- The devices of responders may be diverse and heterogeneous. Thus, the incident information and services need to be portable and sharable.

The key to achieving success and breakthroughs in homeland security lies in effective team communication, resourceful knowledge management, consistent coordination of team processes, and timely dissemination of relevant information; all ensured within a structured collaborative decision environment.

In this paper, a collaborative framework for information interoperation is proposed as a critical provider of a distributed information and decision-making backbone for homeland security. We identify data mining based information filtering as the first key step for effective decision-making. The objective here is to filter the vast information base so that relevant and important situational awareness information is accessible quickly to key decision makers. Most of the current filtering systems provide minimal means to classify documents and data. A common criticism of these systems is their extreme focus on information storage, and their failure to capture the underlying metadata. As a consequence, our proposed approach employs an ontology framework, which allows specifying domain-level context that enables users to attach rich domain-specific semantic information and additional annotations to situational awareness information and services and to employ the meta-information for effective response analysis and execution.

Second, our approach uses Semantic Web Services (SWS) to achieve complex tasks to automate the discovery of the necessary information services (tasks) and compose these services for a particular incident situation in accordance with the national, regional, or local incident management protocols (policies). The final step involves disseminating appropriate information and tasks to the right level of responders and to the public in an appropriate format to their available devices.

This chapter is organized as follows. Section 2 provides the overview of current incident management efforts. Section 3 describes an overall framework of our approach and its components. Section 4 presents our semantic incident management approach using incident ontology and how it is utilized for semantic filtering of information and the identification of Web Services. It shows the SWS and how these are automatically discovered and composed to achieve desired functionalities to added value services. Section 5 presents the dissemination of the information and services customized to fit to the agency and responder's roles, and devices. Section 6 addresses the prototype implementation followed by the description of our conclusion and on-going and future work in Section 7.

2 Incident management

An Incident of National Significance (INS) is "an actual or potential high-impact event that requires robust coordination of the federal response in order to save lives and minimize damage, and provide the basis for long-term

community and economic recovery'' (NRP, 2004). According to Homeland Security Presidential Directive-5 (HSPD-5) issued by President Bush in 2003, the Secretary of Homeland Security is directed to develop and administer a National Incident Management System (NIMS) to prevent, prepare for, respond to, and recover from terrorist attacks, major disasters, and other emergencies. NIMS serves as a single, consistent nationwide template to enable all government, private sector, and non-governmental organizations to work together during domestic incidents (NIMS, 2004).

Six major components of the NIMS are:

1. Command and management that defines standard incident command systems, multi-agency coordination systems, and public information system. These standard incident command structures define the operating characteristics, interactive management components, and structure of incident management and emergency response organizations, and processes for communicating timely and accurate information to the public during crisis or emergency situations.
2. Preparedness that involves an integrated combination of planning, training, exercises, personnel qualification and certification standards, equipment acquisition and certification standards, and publication management processes and activities.
3. Resource management that defines standardized mechanisms and establishes requirements for processes to describe, inventory, mobilize, dispatch, track, and recover resources over the life cycle of an incident.
4. Communications and information management that identifies the requirement for a standardized framework for communications, information management (collection, analysis, and dissemination), and information-sharing at all levels of incident management.
5. Supporting technologies that include voice and data communications systems, information management systems (i.e., record keeping and resource tracking), and data display systems. Also included are specialized technologies that facilitate ongoing operations and incident management activities in situations that call for unique technology-based capabilities.
6. Ongoing management and maintenance component that establishes an activity to provide strategic direction for and oversight of the NIMS, supporting both routine review and the continuous refinement of the system and its components over the long term.

HSPD-5 requires all federal departments and agencies to adopt the NIMS and to use it in their individual domestic incident management and emergency prevention, preparedness, response, recovery, and mitigation programs and activities, as well as in support of all actions taken to assist state, local, or tribal entities. The directive also requires federal departments and agencies to make adoption of the NIMS by state and local organizations a condition for federal preparedness assistance (through grants, contracts,

and other activities) beginning in FY 2005. The participation and integration of all state, territorial, and community-based organizations, including public, non-governmental, and private organizations, such as private sector emergency medical and hospital providers, transportation systems, utilities, and special facilities such as industrial plants, nuclear power plants, factories, military facilities, stadiums, and arenas. Full NIMS implementation is a dynamic and multi-year phase-in process with important linkages to the National Response Plan (NRP), the HSPD-8 (i.e., the "National Preparedness Goal"), and the National Infrastructure Protection Plan (NIPP).

The NIMS provides a comprehensive national framework to incident management by representing a core set of doctrine, concepts, principles, terminology, and organizational processes to enable effective, efficient, and collaborative incident management at all levels. On the other hand, the NRP is an operational incident management or resource allocation plan. The NRP specifies how the resources of the federal government will work in concert with state, local, and tribal governments and the private sector to respond to Incidents of National Significance. It specifies various centers and officers in charge of incident management, e.g., joint field office, joint information center, federal coordinating officer, resource officer, and defines supporting resources such as transportation, communication infrastructure as well as interact with the state, county, and local Emergency Operations Centers and Incident Command Post that provides tactical level incident management operations.

The NRP

1. Describes the structure and processes comprising a national approach to domestic incident management designed to integrate the efforts and resources of federal, state, local, tribal, private sector, and non-governmental organizations. It includes planning assumptions, roles and responsibilities, concept of operations, incident management actions, and plan maintenance instructions.
2. Provides detailed supporting information, including terms, definitions, acronyms, authorities, and a compendium of national interagency plans.
3. Details the missions, policies, structures, and responsibilities of federal agencies for coordinating resource and programmatic support to states, tribes, and other federal agencies or other jurisdictions and entities during Incidents of National Significance.
4. Provides guidance and describes the functional processes and administrative requirements necessary to ensure efficient and effective implementation of NRP incident management objectives.

In order to have a successful implementation of these national directives, guidelines, and operational procedures, the information technology can be utilized to enhance incident management capabilities for different levels of incident responders from different organizations.

The high-quality, accurate, and timely information affects the quality of decisions and more effective incident management tasks. The information technology research in incident management focuses on gathering, processing, managing, using, and disseminating information as well as training incident-related personals.

The 2002 National Research council recommended development of a threat-based 3-D simulation models and visualization tools for Emergency Operation Center training. To achieve this, a network of consortium called Homeland Defense Center Network was formed to develop reusable and standard-based simulation and modeling tools that can be easily shared and integrated to another systems, and to support federal, state, and local response teams, including decision makers and first responders (Corley and Lejerskar, 2003; NIST, 2003). These tools include graphical display of the unfolding of the simulated disaster event and response actions for decision makers to increase the situation awareness, and to make high-level decisions such as deploying and coordinating multi-organizational units of first responders in different areas impacted by the disaster. The first responder training tools include immersive virtual reality, stereo displays, 3-D sound, hand and tracking control to make the disaster responders to feel the disaster zone.

These simulation tools are used not only for training of emergency management teams but also for planning such as location of police, fire, and hospital facilities, defining evacuation procedures and designing communication infrastructure. These are also used to assess vulnerabilities in action plans and strategies such as city emergency plans. The simulation tools are also useful for determining the likelihood of disaster event and identifying potential targets, and for real-time situation updates to project current and future impact of the incidents. The simulation-based incident training and management research areas within the HDC organizations include urban assessment, surveillance, sensor simulation, critical infrastructure, firefighting, and HAZMAT dispersal predictions. Similar 3-D visualization and human-interaction reasoning with artificial intelligence tool (DEFACTO) is developed for training the incident commanders to gain the experience and evaluate tactics in real disaster incidents (Schurr et al., 2005).

Crisis management utilizes the geospatial data that can be located on a map. GIS is a useful tool in all aspects of emergency management from planning to mitigation to response by combining hazards data with other geospatial data. For example, GIS facilitates planning the mitigation and response needs, and identifying and assessing the real and potential damage levels on lives, property, and environment from potential disasters. GIS can provide real-time monitoring for emergency tasks and early warnings, and it can identify resource selection and routing for quick responses for crisis management (Johnson, 2000). A series of research work in Rauschert et al. (2002), Fuhrmann et al. (2003), and Cai et al. (2004) recognize the limitation of the current GIS technologies to support group collaboration as required by the crisis management and propose GeoCollaborative Crisis

Management (GCCM) where the maps play a role as visual mediator of communication and collaboration among distributed team players. This group collaborative GIS allows the natural, multi-modal (i.e., two or more combined user input modes such as speech, gesture, gaze, or body movements), multi-user dialog-enabled interfaces using large screen displays. The geo-collaborative crisis management is based on a distributed multi-agent system that captures the mental states of participants and reasons about the role of maps in order to determine its contents, presentation format, and sharing requirements.

An interdisciplinary RESCUE project (RESCUE, 2004, 2005) assumes humans as sensors collecting data from the incident scenes, in addition of other device driven sensors. Four major research areas to enhance the ability of emergency response focus on information collection, information analysis, information sharing, and information dissemination. Research areas in information collection include speech recognition and event extraction from voice signals, video analysis to track multiple people and recognition of license plates, sensing technologies including remote sensing and bridge sensing, robust networking systems to support information gathering in unpredictable situations, adaptable data collection including cell phone sensors and cellular-based location tracking, privacy protecting data collection.

Topics in information analysis cover information and event extractions from text, video, and multi-modal speech, event awareness such as event database systems or event reasoning, spatial awareness to locate the events from reports, people awareness such as vehicle tracking and peoples location, decision support tools such as loss estimation, emergency vehicle routing and Bayesian analysis of informant reports, etc.

In information sharing area, research goal is to share the information across different organizational boundaries with trust building and controlled access to preserve privacy. Thus, topics include network analysis of the incident command system, trust management in crisis network such as trust negotiation schemes, storing and managing credentials in a mobile environment, encryption of sensitive credentials for limiting access, security, and privacy management such as secure XML publishing or secure processing of queries, distributed peer-based data sharing in crisis-response organizations using overlay networks, peer-based sharing and searching, GIS-based search, etc. Information dissemination areas includes research topics such as establishing responder networks and data sets by analyzing communications among responders in disaster scenes, modeling information flow through networks, customizable dissemination, navigation support for users with disabilities, targeted dissemination, flash dissemination of critical information to a large number of recipients in a very short period time, audio-cued location-based orientation and maintenance to safe paths, etc.

While RESCUE project is comprehensive in terms of research topic areas, the focus is to improve the situation awareness of the first responders by collecting and analyzing communication data from the first responders, and

providing distributed network infrastructure. It does not address important issue of filtering data for incident commanders to identify the real threats from superfluous ones. The incoming sensor data, either from surveillance devices or humans, should be integrated and direct the command officers for a set of information and response tasks. To this end, our approach focuses on the semantics of data to identify real threats, identify tasks and resources for composite tasks as SWS to manage the incidents. Like other dissemination research projects mentioned, our approach focuses on location-, device-, security credential-aware customization of information.

3 Our approach and architecture

With the support of technology, we provide a framework for semantic incident management in homeland security applications to support the automatic data and information filtering, identification of relevant information and resources, dissemination of customized information to the needs of agencies and responders. This provides the responding virtual government team with the right information relevant to the incident type and location, adhering to communication and other collaborative protocols among participating agencies as well as adhering to the individual agencies business policies. Challenges in developing this framework include: (1) identifying the right type of information that is relevant to the incident and visualize them to see the coherent picture of the incident, (2) identifying resources to handle the incidents, including agencies, specialists, other personal and resources specific to a given alert type (e.g., fire, hazmat spills), (3) disseminating appropriate information and tasks to the right level of responders and to the public in an appropriate format to their available devices. Our approach achieves automatic filtering of alert information and data using an incident knowledge base represented as an semantic graph (ontology). The semantic graph captures the concepts of different incident types and their relationships among different incidents. The information resources such as textual description of incidents, audio and video clips from the incident scene are tied to the concepts. Our approach for semantic incident management is to construct a knowledge base (ontology and relational knowledge base) and to utilize it to automatically identify the relevant incident management multimedia component information and services related to an incident type. The knowledge base is constructed based on each agency's Emergency Operations Plan (EOP). Figure 1 shows the overall framework and its components. A brief description of the components is given below.

3.1 Alert profile update and alert generation data mining

These two modules capture the incident/alert profile from the situation reports or by automatic data mining components from the data in the

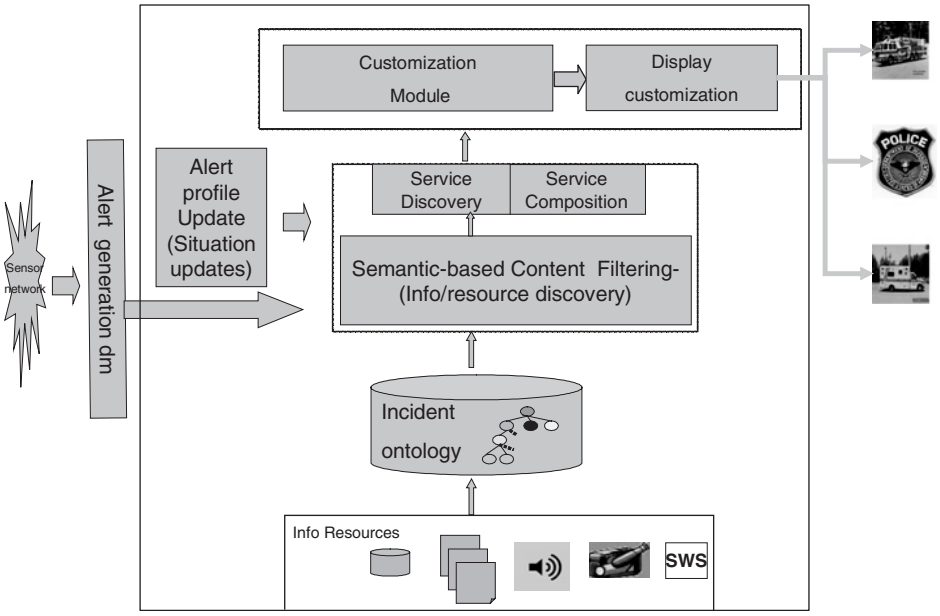


Fig. 1. Semantic incident management for homeland security.

sensor network (Adam et al., 2005). Alert profiles include alert types, location, severity, casualties, etc. (Janeja et al., 2005a).

3.2 *Semantic-based content filtering*

The necessary resources, services, and information for handling the incident are automatically discovered. This module uses a knowledge base of incident ontology (concepts, types) and incident relationships (incident RDF/semantic graphs) to discover resources (Sheth et al., 2002). Each of the resources is described using concepts and relationships defined in the ontology. For instance, the semantic description of the resources takes into account the type of incidents (fire, radiological, or chemical, etc.), severity, and human casualty levels. Individual agencies are also considered as resources, and they are described with the concepts from the incident ontology. Thus, the types of the incidents may determine which agencies need to be involved and which resources are required.

3.3 *Service discovery and service composition*

Some of the information resources are not static, for instance, the map of coordinate (x, y) needs to be generated using a map generation software. Thus the map generation Web services are described as a part of information resource. In order to plan an evacuation, a host of information is

required. First the evacuation plan should identify a hazardous material spread modeling tool to assess where may be most severely affected, and a map of potential evacuation sites and hospitals are needed around the incident site (Chandrasekaran et al., 2002). Thus decision on the evacuation plan may require a complex set of information and service resources composed together, e.g., hazard spread modeling with wind directions from the weather forecasting service. Then the plan will require determining the number of volunteers based on the size of the evacuation. This requires the volunteer lookup service. The SWS and Service Composition modules are to discover available services and compose them for complex information needs. These services are also described in terms of the incident ontology to be discovered automatically via semantic concepts (McIlraith et al., 2001; McIlraith and Martin, 2003).

3.4 Customization

The alert information and services discovered from the semantic filtering stage now can be disseminated to each agency and group of responders. However, not all the semantically related incident information is needed for every agency. The fire department and medical organization's information needs are different and a particular responder's role may further restrict information based on the need-to-know access authorization, thus further role-based filtering is conducted to customize the alert-related information for each agency and responder.

3.5 Display customization

This module determines the device-specific content filtering and spatial and temporal display preferences and constraints are considered to provide the customized display of alert/incident information. A PDA display and a PC display may contain similar information, but the PDA may not be able to play video and only text or audio can be selected while a PC may display the audio, video, and text components. The spatial arrangement and temporal synchronization of information from different sources are to be considered.

4 Incident ontology

An ontology can be understood as a graph whose nodes and edges represent concepts and the relationships between those concepts. Ontologies are used for the conceptualization of the application domain in a human understandable and machine-readable form (Gómez-Pérez et al., 2003). We have developed an incident ontology to represent different incident types as shown in Fig. 2. National Research Council (NRC, 2002) identified nine

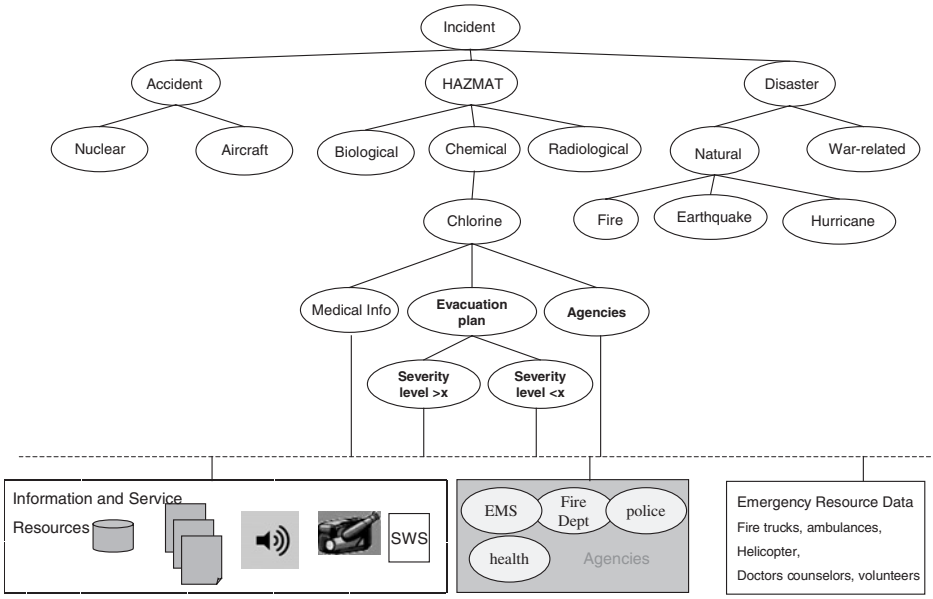


Fig. 2. Incident ontology.

critical areas of terrorism-related threat and incident areas: nuclear and radiological incidents; human and agricultural health threats; incidents on toxic chemicals and explosive materials; information technology and telecommunication attacks; incidents on energy systems; incidents on transportation systems; threats on cities and fixed infrastructure; human response-related incidents; and complex and independent systems incidents.

Similarly, we have organized incidents with several situational types, such as accidents, natural disaster types, or hazardous materials. Each of these subtypes has finer incident types. For instance, the hazardous material-related incidents can be radiological, chemical, or biological incidents. These hierarchical type and subtype incidents are related via "is–a" relationships (Kokla and Kavouras, 2001). The information resources (database, Web pages, audios, videos) and Web services as well as IM resources (agencies, equipment, facilities) are described using these semantic incident concepts from the ontology showing its relevance to an incident/alert type. These incident concepts augment the NIMS National Incident Management Resource Typing Protocol where resources (personnel, teams, facilities, supplies, and major items of equipment) are described by common category, kind, components, metrics, and type data in order to avoid confusion in crisis management.

4.1 Semantic filtering for incident information discovery

Our approach to identify and discover relevant information and resources uses semantic filtering. Alert/incident profiles (alert type, location,

severity, etc.) are generated from either situation reports or a sensor alert generation module (Janeja et al., 2005b). They are specified in CAP (Common Alerting Protocol) compliant format. An alert profile is matched against the concepts in the ontology using either exact match or approximate matching techniques. Then the resources that are described with the same semantic labels or the subcategories of the concept label are searched, and put into the candidate information pool. For example, the chlorine spill incidents may require medical information on chlorine effects on health. The following step involves looking for medical information described with chlorine health effects. The resources may be either in textual, audio, or visual format. The agencies to treat the chlorine hazards can be discovered. The evacuation plan depends on the severity levels. There are different resource requirements for different levels of severity. The severity level is also specified in the resources. These will be discovered. Equipment and facilities to manage the spill are also described. Thus the semantic labels can be used for identifying these.

Our approach allows the incident-specific information and resources to be discovered in *ad hoc* manner dynamically as needed, rather than in a predefined and static manner, providing more flexible incident information discovery and management. The information can be optionally composed into complex multi-media objects. Similarly, services also need to be automatically discovered and composed to provide required functionalities. This is discussed in the following sub-section.

4.2 Semantic web services and web service composition

In incident management, often not only information and resources but also services (e.g., to support the adequate response to an incident) are needed. These services can be made available via the Web, thus called Web Services (Harris et al., 2003). Using the incident ontology, we describe the semantics of the Web Services to achieve their automated discovery (Kulvatunyou and Ivezic, 2002). We call such services SWS. Unlike other information, SWS can be described not only regarding their semantics (behavior) but also their operational (syntactic) characteristics such as input, output or service bindings (Sollazzo et al., 2002). We use WSDL and OWL-S [OWL-S, 2004] to describe properties and capabilities of SWS which support the automated discovery and Composition of Web Services. Using the Process Ontology, for parts of the OWL-S specifications, Web Services are considered to provide simple or complex actions with pre-conditions and effects. To provide information rich resources that form a part of the overall alert, we consider both simple services that are independent, self-reliant services implementing the task functionality and composite services that are a combination of services providing the task functionality. We discover relevant simple Web Services using concepts

from our Incident ontology. For composite Web services, we use Semantic Web Service Composition and selection based on service pre-conditions and post-effects.

4.2.1 Semantic web service composition

WS Composition involves the process of selecting, combining, and executing WS to achieve the purpose of a user request (Wu et al., 2003). This involves match making of constraints between Web Service inputs, outputs, preconditions, and effects (IOPEs) along with the outputs and effects (OEs) of a user request. In addition to matching IOPEs, the automated WS Composition problem also can involve the selection from alternative Web Services that match the IOPE constraints of the composition problem (Martin et al., 2004).

In our approach, we first require an OWL-S description of that service that more fully represents the inputs and outputs of the service, i.e., constructing a composite process model that links the various operations provided by the Web Service into semantically meaningful message patterns, e.g., checking responder identification before displaying traffic details (Chun et al., 2005). We make use of capability matching as described in (Martin et al., 2004) that compares the capabilities provided by any of the advertised services in UDDI with those needed by the requester. The goal is to find the service provider that produces the results required for the requester. In general, it is unrealistic to expect that the capabilities offered by a service will exactly match the request. For example, the request may be for traffic information based on location, and the task of the matching engine is to decide whether it can be accomplished by a service that accepts zip codes (Martin et al., 2004; Akkiraju et al., 2003). Our matchmaking algorithm determines how likely it is that each capability advertisement indicates that the service will accomplish the particular function specified in the request.

The matchmaking algorithm initially maps the composite Web Service operations to a set of specialized UDDI TModels that store the corresponding OWL-S information. Next, each calling operation of the composite service is mapped to one or more operations of the existing service. The algorithm looks for the composite Web Services description so that the purpose and category are compatible with that of the available services. Then the algorithm verifies that interacting services are binding composable.

For example, Web services for a chemical HAZMAT incident related to chlorine may include Traffic Status and Plume Modeling. A typical scenario would be of a responder, at the executive level or at the site of the incident, using the associated services for better decision-making. The responder would feed in the environment parameters to the chosen services to obtain specific information. For example, the executive responder may want to view the traffic conditions for the site including the surrounding areas (based

on the address in terms of the city and state) whereas the on-site responder may want to view the traffic status at a specific geo-coordinate. One of the (pre) conditions for the TrafficStatus operation of TC is validLocation and service constraints are the access control privileges accorded to the user to view generateTrafficReport. Preconditions are logical formula that need to be satisfied by a service requestor prior to the execution of the service, e.g., check for the validity of responder identification to view the details of the requested information. Effects are logical formula that state what will be true on the successful execution of the service, e.g., displayTrafficEventTag, for the retrieval of a related traffic event. OWL-S effects are the side effects of the execution of the service (Martin et al., 2004).

The following steps describe the mapping of the submitTrafficArea operation of the TrafficStatus service according to our matchmaking algorithm (Paliwal et al., 2004).

- Map the composite Web Service operations to a set of specialized UDDI TModels that store the corresponding OWL-S information of the Traffic Service.
- Identify the component services (e.g., Traffic Service) supporting the SOAP protocol so that submitTrafficArea's purpose and category are compatible with the service purpose and category.
- Determine operations of the Traffic Service that are composable with submitTrafficArea. Since submitTrafficArea is a solicit-response type of operation it shall map to a corresponding request–response operation getLocationInfo.
- Test the operations for message composability. The input of submitTrafficArea is compared with the output of getLocationInfo. All of getLocationInfo output's parameters are mapped to the corresponding parameters of submitTrafficArea's. Since we can determine such a mapping, the two operations are message composable.
- Insert a ''plug-in'' between the operations in the layout, since both operations are syntactically and semantically composable.
- Perform iterations for all other service operations required by the composite service.

5 Customization and dissemination

For alert message filtering and delivery, especially in an emergency situation, we have to consider the inter-organizational relationships across entities such as responding agencies. Some information distributed along with the alert is situation-specific while other information is agency-specific. The alerts should be disseminated based on the recipient's different credentials on the agency level as well as the hierarchical level within the same agency. For example, the information required by the police department is

different from the information required by the fire department; in addition the information required by the chief of the fire department is different from the information needed by a fireman.

The customization module receives the alert information from the semantic filtering and service composition module. It then starts filtering the alert based on (1) the recipients' role and (2) the recipients' devices and preferences. The role filtering of the alert is to select the relevant components to the related agencies as well as the role of the recipient's role in each agency.

After selecting the related components for each agency, we adapt the components' spatial layout and rendering format based on the recipient's devices characteristics (e.g., monitor size, Operating System), and the recipient's preferences (e.g., audio format instead of text) (Atluri et al., 2003).

For the underlying protocol in the customization module, we adhere to the NIMS and the NRP protocols. NIMS has been developed and administered by the Department of Homeland Security to provide a consistent nationwide template to enable all government, private sector, and nongovernmental organizations to work together during domestic incidents. The NRP focuses on prevention, preparedness, response, and recovery within the life cycle of an incident by establishing incident monitoring and reporting protocols. One way to allow different agency's information systems to communicate is CAP. CAP is an XML non-proprietary data interchange format that can simultaneously transmit emergency alerts through different communication networks. The Organization for the Advancement of Structured Information Standards, an international standards body, has adopted CAP as a standard. Once the policy for manifesting the alerts is determined, the alert format is presented using CAP that provides a digital message format for all types of alerts and notifications (Common Alerting Protocol v 1.0, 2003).

CAP defines the alert message structure which includes four main segments. (1) The <alert> segment, which provides the message identifier, purpose, source, and status. It may contain one or more segments. (2) The <info> segments which describe the urgency, severity, certainty, actions to be taken, and related parameters of an anticipated or actual event. Each <info> segment may include one or more resource segments. (3) The <resource> segment provides a reference to additional information such as video, image, text, or audio file and one or more area segments. (4) The <area> segment describes one or more geographic areas related to the <info> segment.

5.1 *Role filtering*

The role filtering of the alert message is based on the jurisdiction policies, agency policies, and the role policies. Based on the NIMS and NRP, certain

protocols need to be followed in an emergency scenario. A jurisdiction policy determines which agencies should coordinate the emergency management based on the alert magnitude and area. An agency policy determines the access to certain components based on the access rights of the agency. It is used to identify which information resources should be accessed from an alert message by which individual. A role policy determines which resources to be accessed based on the role within an agency.

5.2 Personal preference and device filtering

Based on the individual accessing the alert, we start our second stage filtering to best convey the alert to the recipient. In this filtering stage, instead of creating separate style sheets to layout the XML information for each role in each agency based on each device, and individual preference, we automatically select the components modalities that match the recipients' device characteristics and then reconstruct the layout of the alert accordingly (Gomaa et al., 2005).

6 Prototype implementation

We have developed a prototype based on our semantic incident management framework for Emergency Management Office of New York and New Jersey Port Authority with multi-media contents and interfaces that includes maps with basic location information and other thematic layers (e.g., available evacuation facilities), video feeds from the incident site, video-conference interface for communication with various partner agencies in incident management, status of situation report, etc. The multimedia information gives a comprehensive view of the incident situation and interaction interface. Figure 3 shows the alert/incident information view where a truck with radiological material is missing on route to the normal course described in the bill of lading. Continuous situation reports come in and are summarized in the form of text headlines (in the lower left corner), a map and a video feed from the highway with suspicious truck (upper-right corner) are displayed, and a set of Web services to assess the risk model (lower-right side) are shown with links. The video-conference interface is shown as well (upper-right corner).

Alert messages from situation reports and information components (such as video feeds) identified from the semantic filtering are encapsulated in CAP compliant XML format. The use of XML facilitates the portability and sharing of information among different agencies and responders. XML

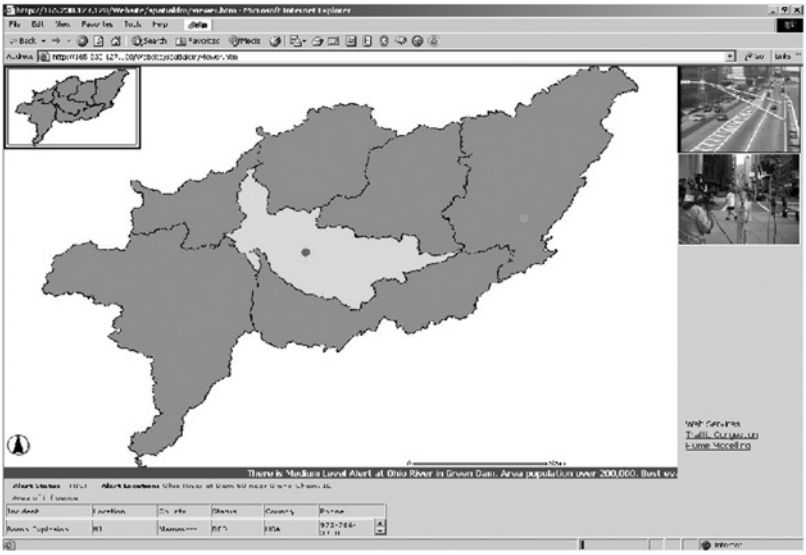


Fig. 3. Multi-media incident/alert information presentation.

messages are customized according to the roles and policies, user device properties as well as user preferences. We implement the role filtering by using jurisdiction, agency, and role policies.

On user access, the system retrieves the user role, preferences, and used device properties (Gomaa et al., 2005). The prototype then selects the relevant formats for the alert components. It then decides the spatio-temporal layout of the alert components to finally send the alert to the user based on his role, device, and preferences. For the customized view implementation we use a web interface using Java Servlets. The alert is associated with location coordinates to be presented on a map in a GIS interface, along with the information to be presented to all the relative agencies. When a member of a related agency logs in to the system, he receives the alert related to his department. The received alert is associated with a generated XSL style sheet based on the alert and the receiving device. The general alert is represented as shown in Fig. 4. It is then customized for each agency so that only relevant information is displayed. Figure 5 shows the customized XML for a specific agency.

The adjustment of the layout changes according to the receiving device. For example, Fig. 6 shows the view for a normal PC monitor and Fig. 7 shows the corresponding view on a PDA. We detect the monitor resolution and send it to our server to apply the changes to the layout.

Web services discovery and composition are implemented with WSDL and OWL-S based on SAP's NetWeaver and Auto-ID Infrastructure (AII) (Bornhövd et al., 2004).

Fig. 4. Alert with all related agencies.

7 Conclusion

In this chapter, we have presented an approach for semantic incident management for homeland security that supports the provisioning of incident-related information, resources and service discovery, composition, customization, and dissemination. We have presented incident/alert ontology to capture the semantics and situations of the different incidents and threats including those for homeland security. Ontology concepts are used to describe information and service resources. The incident management-related resources and information are discovered through a semantic filtering process where the alert profile information is used to match the semantic descriptions of the information and services. Web services are also described with concepts from the incident ontology and are discovered

Fig. 5. Customized alert for one agency.

similarly. The added functionalities can be achieved through Web Service Composition. The discovered information and services are further customized according to the roles and preferences of responders and agencies. Then the dissemination and display of information is customized according to the device and display preferences such as spatial and temporal layout constraints. Our prototype system is implemented in the domain of NJ–NY Port Authority Emergency Management Office where the situation reports are used to capture the alert profiles in an XML-based CAP format. Information and service discovery and composition is implemented using

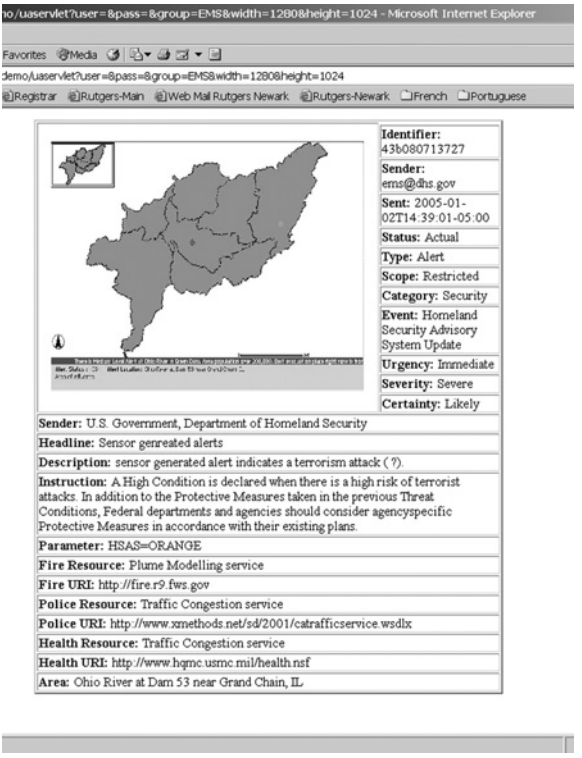


Fig. 6. PC monitor view.

SAP's NetWeaver platform, and customization and dissemination are implemented for various devices, like Laptops, PCs, and PDAs. We are in the process of developing comprehensive incident ontology and we are planning to incorporate the consideration of dynamic situations in the incident information and process management. From the recent experiences, incident management requires quantum transformation.

To achieve this, quantum or rapid advances in information technologies is necessary. Data gathering from static surveillance devices are changing to data streaming from ad hoc networks of devices and humans that move around. The key research challenges include to measure the quality and relevance of the heterogeneous data coming from a particular situation and context, interpret, process, and construct the emergency situation in rapid manner to be useful in decision-making and response task execution. The manual textual interface to access data is changing to multi-modal interface. The combined speech, gesture, and visual interface will be a normal rather than an exception. We foresee IT advances in the intuitive multimodal data access and interface. Another critical area to advance is "flexible" information protection and controlled access. The ad hoc coalition of

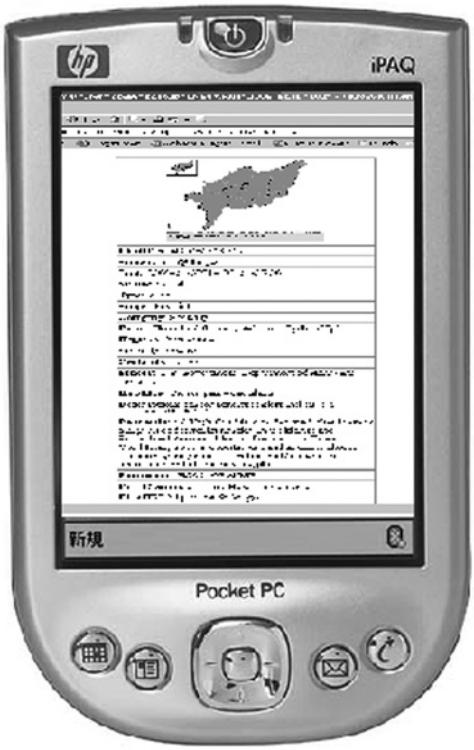


Fig. 7. PDA view.

multi-agency and multi-organization incident management network generates temporary roles to play. The information access needs to be determined according to not only the roles, but also contexts, including time, location of incidents, and time and location of the responders. The accountability of temporary roles in dynamic incident management still needs to be investigated. The protection of sensitive data and privacy protection of data access are critical. The dynamic data fusion is required rather than static one for seamless access to data. To this end, the service-oriented approach has potential as seen in our approach. The mobile collaborative tools have to be devised. The incident management opens up these opportunities and many other challenges for emerging technology areas.

8 Questions for discussions

1. Discuss the major characteristics and differences between incident management for man-made such as 9/11 terrorist attacks and natural

disasters such as Tsunami incidents or Hurricane Katrina. What are some of available information technology tools to prevent and prepare for these incidents, and what are the challenges facing information technology research?

2. Illustrate an example of resource identification process in a multi-agency incident management, and that of a single/local agency incident management. Discuss the differences and commonalities.
3. The incident management involves dynamic "coalition" of multiple organizations ranging from federal, state, and local agencies to non-governmental organizations, depending on the needs and the jurisdictions involved in the incident. Discuss the NIMS specifications on the different roles for incident command, how the chain of commands are established, and what conflicts may arise?
4. Discuss available evaluation criteria and schemes to measure effectiveness of incident management.
5. Often incident management requires decision-making with real-time data streaming from sensor data including human (responders as well as observers). Discuss the challenges that the current decision-making technologies may face. Specifically, discuss the issues of data quality and available tools and approaches to identify and prioritize the critical data in decision-making.
6. Information sharing in multi-organizational incident management requires responders of different levels to access data owned by different organizations. Discuss the possible approaches to render the temporary access to data, and how to protect the sensitive data?
7. Mobile devices are used by responders form a distributed network among responders. The peer-to-peer communication among responders is used for situation awareness. Discuss different ways mobile devices are used in incident management, and what the challenges are?
8. Discuss how the location and context information can be used in information collection, sharing, and dissemination. What are the research challenges to consider the context in each stage of incident management?

References

N. Adam, V. Atluri, S. Chun, A. Gomaa, A. Paliwal, J. Vaidya, M. Youssef, A. Suenbuel, C. Bornhoevd, S. Raiyani, T. Lin, J. Cooper and J. Paczkowski (2005). Semantic-based Incident Management System, *Proceedings of Inaugural conference on Working Together: Research & Development Partnerships in Homeland Security, Department of Homeland Security Science & Technology,* Boston, MA.

Akkiraju, R., R. Goodwin, P. Doshi, S. Roeder (2003). A method for semantically enhancing the service discovery capabilities of UDDI, in: *Proceeding of IJCAI-03 Workshop on Information Integration on the Web (IIWeb-03)*, Acapulco, Mexico, August 9–10, 2003.

Atluri, V., N. Adam, A. Gomaa, I. Adiwijaya (2003). Self-manifestation of composite multimedia objects to satisfy security constraints, in: *Proceedings of the 2003 ACM SAC*, pp. 927–934.

Bornhövd, C., T. Lin, S. Haller, J. Schaper (2004). Integrating automatic data acquisition with business processes, experiences with SAP's auto-ID infrastructure, in: *Proceeding of the 30th VLDB Conference*, Toronto, Canada, August 29/September 3, 2004.

Cai, G., A.M. MacEachren, L. Bolelli, GCCM. (2004). Map-mediated collaboration among emergency operation centers and mobile teams, in: *Proceedings of GIScience 2004*, Adelphi, MD, USA.

Chandrasekaran, S., G. Silver, J. Miller, J. Cardoso, A. Sheth (2002). Web Service technologies and their synergy with simulation. *Winter Simulation Conference (WSC'02)*, *December*, San Diego, CA, USA.

Chun, S.A., V. Atluri, N.R. Adam (2005). Using semantics for policy-based Web Services composition, a special issue on Web Services. *Journal of Distributed and Parallel Databases* 18(1), 37–64.

Common Alerting Protocol v 1.0, OASIS Emergency Management TC [OASIS 200402], 12 August 2003.

Corley, J., D. Lejerskar (2003). Homeland defense center network—captializing on simulation, modeling and visualization for emergency preparedness, response and mitigation, in: *Proceedings of the 2003 Winter Simulation Conference*, New Orleans, LA, pp. 1061–1067.

Fuhrmann, S., I. Brewer, I. Rauschert, A. MacEachren, G. Cai, R. Sharma, H. Wang, L. Bolelli, B. Shaparenko (2003) Collaborative emergency management with multimodal GIS, in: *Proceedings of ESRI User Conference*, San Diego, CA, USA.

Gomaa, A., N.R. Adam, V. Atluri (2005). Adapting spatial constraints of composite multimedia objects to achieve universal access, in: *IEEE International Workshop on Multimedia Systems and Networking (WMSN'05)*, Phoenix, AZ, USA.

Harris, S., N. Gibbins, N. Shadbol (2003). Agent-based semantic Web Services, in: *World Wide Web Conference (WWW2003)*, Budapest, Hungary.

Janeja, V.P., V. Atluri, J.S. Vaidya, N. Adam (2005a). Collusion set detection through outlier discovery, in: *IEEE Intelligence and Security Informatics*, LNCS 3495, Springer, Berlin, Germany.

Janeja, V.P., V. Atluri, A. Gomaa, N. Adam, C. Bornhoevd, T. Lin DM-AMS (2005b). Employing data mining techniques alert management, in: *NSF National Conference on Digital Government*, Atlanta, GA, USA.

Johnson, R. (2000). *GIS Technology for Disasters and Emergency Management: An ESRI White Paper*, http://www.esri.com/library/whitepapers/pdfs/disastermgmt.pdf

Kokla, M., M. Kavouras (2001). Fusion of top-level and geographical domain ontologies based on context formation and complementarity. *International Journal of GIS* 15(7), 679–687.

Kulvatunyou, B., N. Ivezic (2002). Semantic web for manufacturing web services, in: *World Automation Congress Eight International Symposium on Manufacturing with Applications, June,* Orlando, FL, USA.

Martin, D., M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, K. Sycara (2004). Bringing semantics to Web Services: The OWL-S approach, in: *Proceeding of the First International Workshop on Semantic Web Services and Web Process Composition*, San Diego, CA, USA, July 6–9, 2004.

McIlraith, S., D. Martin (Jan/Feb, 2003). Bringing semantics to web services. *IEEE Intelligent Systems*, 18(1), 90–93.

McIlraith, S., T. Son, H. Zeng (2001). Semantic web services. *IEEE Intelligent Systems* 16(2), 46–53.

NIMS National Incident Management Resource Typing System (2004). http://www.nimsonline.com/nims_3_04/national_incident_management_resource_typing_system.htm#purpose

NIST. (2003). *Conference on Modeling and Simulation for Emergency Response*, http://www.mel.nist.gov/div826/msid/sima/simconf/mns4er.htm

NRC. (2002). *Making the Nation Safer — The Role of Science and Technology in Countering Terrorism*, Committee on Science and Technology for Countering Terrorism, Division on Engineering and Physical Sciences, National Research Council, National Academy Press, Washington, DC, 2002.

NRP. (National Response Plan) Homeland Security, http://www.dhs.gov/dhspublic/interweb/assetlibrary/NRPbaseplan.pdf, December 2004.

OWL-S (2004). At http://www.daml.org/services/owl-s/1.0/

Paliwal, A.V., N. Adam, C. Bornhövd, J. Schaper (2004). Semantic discovery and composition of Web Services for RFID applications in border control, in: *Proceeding 1st. Intl. Workshop SWS'2004 at ISWC 2004, Hiroshima*, Japan, November 8, 2004, CEUR Workshop Proceedings, ISSN 1613-0073.

Rauschert, I., P. Agrawal, S. Fuhrmann, I. Brewer, H. Wang, R. Sharma, G. Cai, A. MacEachren (2002). Designing a human-centered, multimodal GIS interface to support emergency management, in: *Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems*, McLean, Virginia, pp. 119–124.

RESCUE, First Year RESCUE Progress Report (2004). *Responding to Crises and Unexpected Events*, http://www.itr-rescue.org/bin/pubdocs/rescue%20docs/2004%20RESCUE%20annual%20report.pdf

RESCUE, Second Year RESCUE Progress Report (2005). *Responding to Crises and Unexpected Events*, http://www.itr-rescue.org/bin/pubdocs/rescue%20docs/2005%20RESCUE%20annual%20report.pdf

Schurr, N., J. Marecki, M. Tambe, P. Scerri (2005). Demonstration of DEFACTO: training tool for incident commanders, in: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, Utrecht, The Netherlands.

Sheth, A., C. Bertram, D. Avant, B. Hammond, K. Kochut, Y. Warke (2002). Managing semantic content for the web. *IEEE Internet Computing* 6(4), 80–87.

Sollazzo, T., S. Handschuh, S. Staab, M. Frank, N. Stojanovic (2002). Semantic web service architecture—evolving web service standards toward the semantic web, in: *FLAIRS 2002*, Pensacola, FL.

Wu, D., B. Parsia, E. Sirin, J. Hendler, D. Nau (2003). Automating DAML-S Web Services composition using SHOP2, in: *Proceeding of 2nd International Semantic Web Conference (ISWC2003), October*, Sanibel Island, FL, USA.

This page intentionally left blank

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 15

Process-Centric Risk Management Framework for Information Security

Ravi S. Behara

IT & Operations Management Department, College of Business, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA

Somnath Bhattacharya

School of Accounting, College of Business, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA

Abstract

This chapter discusses the development of a process-centric risk management framework for information security. On the basis of an initial review of existing network risk management standards as well as IT governance approaches, a three-stage effort is undertaken to develop an integrated framework for risk management. This effort highlights the need to extend the defense-in-depth philosophy to a managerial level so as to emphasize the importance of technology, business process, human effort, and financial management in developing a comprehensive approach to network security risk management. Then, using this framework and by consciously stepping away from the traditional technology-oriented view of network risk management, this chapter develops a process-centric framework within which to understand and manage network security. Contemporary security breach cases are used to illustrate this process-centric framework.

1 Introduction

"The *National (US) Strategy for the Physical Protection of Critical Infrastructures and Key Assets* identifies a clear set of national goals and objectives and outlines the guiding principles that will underpin government efforts to secure the infrastructures and assets vital to our national security, governance, public health and safety, economy, and public confidence"

(DHS, 2003a). These critical infrastructures include information, telecommunications, energy, transportation, healthcare, and banking and financial services. The United States Federal Government places high priority on the consistent application of security across the infrastructure. An agreement on a sustainable security threshold and corresponding security requirements also remains elusive. Efforts to address these challenges are not just the responsibility of the United States Federal Government but also those of concerned citizens. Academic research efforts in this domain should be seen in the context of that larger responsibility. Further, information infrastructure is at the core of all other infrastructures in our knowledge-intensive society today. As such, this chapter focuses on risk management for information security, thereby contributing to the growing understanding of securing all critical infrastructures.

Formalized approaches to assessing and analyzing information security risks are being attempted today. Management typically uses qualitative methods that are characterized by subjective risk measures such as ordinal ranking (low risk or value, medium risk or value, and high risk or value) in a risk-to-value matrix. The qualitative methods emerged in part from a persistent belief that it was simply too difficult to get the real numbers. Also, qualitative approaches appeal to management, which could be looking for the "least-effort" way to prove they had "assessed their risks." Quantitative approaches to risk, though limited in their use, have also typically limited themselves to quantifying a loss and its associated probability of occurrence. Such an approach requires the availability of suitable data. But this is lacking in many organizations that cite difficulties in definition and measurement. As such, little attention has been paid to a more comprehensive evaluation of risk analysis and assessment until recent years in which security-related events have caused severe disruptions. Hence, it is time to establish and formalize the framework of risk management necessary to support a more comprehensive approach in the context of information security.

One approach to address information security is to consider the root causes of any security breach that may be classified into four categories: technology, people, information, and work-process related causes. The traditional "solutions" that dominate the current security debate involves addressing the *technology causes* of hardware, software, or network components. Studies are beginning to recognize the need to address the *people causes*, especially as the scope and scale of internal security breaches have increased in organizations. *Information causes* are related to the intrinsic vulnerabilities associated with the attractiveness of the information that is being protected. By redefining the information structure and requirements, this vulnerability can be mitigated or resolved by reducing its attractiveness to attack. A typical example is the use of employee identification numbers, instead of social security numbers, to identify employees in organizational databases. It is only after that, that solutions such as encryption may be

considered. However, the framework developed in this chapter specifically draws attention to an additional component—work-processes—which are typically ignored. *Work-process causes* are inherent in the way work is accomplished in any domain. Additionally, the way information is handled in the work-process could be of significant interest. However, the distinction between work-process and information is increasingly not discernable as much of the work today is information-intensive. Hence, information security is not something that has to be done in addition to doing work (as it is now conceived), but an integral part of how work is designed and done. A useful process to benchmark would be the nuclear material handling processes in the nuclear power and nuclear reprocessing industries. Considering information to be "radioactive" would be a useful metaphor that can highlight the importance of integrating safe practices into work routines in a systematic and ubiquitous manner.

2 Work-process driven security failures

Despite the implementation of various technical advances to enhance network security, there continue to be a number of security breaches. The causes for many of them can be traced back to the work-processes due to which they primarily occur. Some of these (Wailgum, 2005) are briefly discussed below to give a context to the discussion in the previous section.

2.1 Information management

The University of Kansas' Office of Student Financial Aid sent out an e-mail to 119 students, including the names of all students in the e-mail, informing them that their failing grades could result in them losing financial aid. University administrators may have violated the Department of Education's Family Education Rights and Privacy Act, which protects the privacy of students' grades and financial situations. A local government agency in Florida sent out an e-mail listing of names of patients with AIDS, in clear violation of healthcare and other privacy laws. They later sent out an e-mail requesting that all copies be deleted and not printed. These examples calls into question the common workflow practice where information is attached to e-mails instead of informing the e-mail recipient to log into a secure source for that information.

2.2 Equipment management

A former Morgan Stanley executive sold his old BlackBerry on eBay for $15.50, along with hundreds of confidential company e-mails. Morgan Stanley did have a policy that stated that mobile devices should be returned to IS for "data cleansing" which was obviously not followed. In a somewhat

similar situation, MCI said that a company financial analyst's laptop had been stolen from his car while parked at his home garage. That laptop contained the names and social security numbers of 16,500 current and former employees. The data on the machine was not encrypted. Those with an intention to steal are beginning to realize that the information on portable devices could be much more valuable than the device itself. This calls for a closer examination of the management of mobile devices with respect to security, and the protocols in place to transfer information between such devices and the more secure organizational systems.

2.3 Technology management

A typical work-process cause of security breach can be traced to the fact that the work boundaries between office and home do not exist for many employees. Many employees tend to setup wireless home networks to make working from home more convenient. In doing so, as non-experts, they tend to use the default password and user ID, and expose themselves to potential hackers. Once the employee authenticates on these open networks, the organization's systems will be at risk as well. Training of employees to recognize the vulnerabilities of working remotely, development of robust remote-login protocols, and an audit of remote access practices by employees is now a necessity.

2.4 Supplier management

CardSystems Solutions, a third-party processor of credit card transactions for MasterCard, Visa, American Express, and Discover, had an unauthorized infiltration of its network and put up to 40 million cardholders' information at risk. The primary cause was not that of inadequate technical security of their network, but the fact that CardSystems violated its agreement with the credit card companies by storing cardholders' account information on its own systems. This obviously calls for stricter training and auditing of third-party service providers and others in an organization's information supply chain.

2.5 Materials management

Iron Mountain, a company that specializes in handling data tapes, has lost customer tapes on more than one occasion. They lost 40 backup tapes that had the names and social security numbers for 600,000 of current and former US-based Time Warner employees and some of their dependents and beneficiaries. If this happens at a specialized transportation service, then companies that use traditional business shipping services are at significantly higher risk. For instance, CitiFinancial's shipment of computer tapes (through UPS) was lost while in transit to a credit bureau, putting

3.9 million CitiFinancial customers' data at risk. This data included their names, social security numbers, account numbers, and payment histories. The underlying cause of the breach was the business process of transporting unencrypted data over unsecured channels. This calls for a re-evaluation of what is being transported and how it is being transported. For example, lessons from the transportation of hazardous materials may provide some useful insights.

2.6 Employee training

A company found employees maintained 40 files on its Exchange servers for documents called "passwords.doc" placing company applications, business information, and personnel information at risk. This can be traced back to a lack of security-related training for employees. A security-conscious culture shift is called for in information security, similar to the quality-conscious culture that gradually evolved over the past two decades in many organizations and societies.

2.7 Customer verification

Criminals posing as small-business owners accessed names, addresses, and social security numbers of 145,000 ChoicePoint customers. Such fraudulent activity can occur when basic customer verification process are inadequate or non-existent. More robust work-processes have to be designed and implemented to mitigate such problems.

3 Risk management

Before developing a framework, it is useful to review some of the perspectives in risk management. Early concepts of risk can be seen in the writings of Pascal, the 17th century French Mathematician, Physicist, and Philosopher. He suggested that it was less risky to believe in God than to be a nonbeliever. This can also be seen as an early application of the laws of probability for predicting the likelihood of unwanted consequences. His famous work in philosophy, *Pensées*, was a collection of personal thoughts on human suffering and faith in God, which he worked on from 1656 to 1658. It contains, 'Pascal's wager', in which he uses probabilistic and mathematical arguments to prove that belief in God is rational with the following argument: "*If God does not exist, one will lose nothing by believing in him, while if he does exist, one will lose everything by not believing.*"… concluding that "…*we are compelled to gamble*…"(O'Connor and Robertson, 2005).

While the insurance industry can be seen as the early users of formal risk management principles in the early part of the 20th century, modern risk

management can be traced back to the 1950s with Gallagher's (1956) article in the Harvard Business Review calling for the formalization of the efforts to mange an organization's risk. Risks encountered in organizations are varied, challenging, and complex in nature and include financial, political, technology, and operational, to name a few categories. Two distinct pathways to manage risk have emerged. One approach to risk is the insurance-approach to protect against catastrophic and unaffordable loss, and the other a systematic approach to manage and mitigate risk through planning. The latter approach has a fundamental belief that planning and implementation (of those plans) can help reduce both the probability of unfavorable events and their associated loss. This approach grew in significance through the second half of the 20th century and is epitomized by the efforts at National Aeronautics and Space Administration (NASA) in space exploration. It is deeply rooted in contemporary management thought. As such it is fundamental to all infrastructure security management in the current national effort, and is exemplified by the National Infrastructure Protection Center (NIPC) approach to protecting critical assets (NIPC, 2002). The National Strategy to Secure Cyberspace (DHS, 2003b) also highlights the need to conduct integrated risk modeling of cyber and physical threats, vulnerabilities, and consequences.

Some basic terms and definitions related to risk are briefly discussed below. *Risk* is the combination of the likelihood and the consequence of a specified hazard being realized. While this is the traditional "downside" view of risk, it must not be forgotten that the consequence may also be a positive one leading to the concept of "upside" risk. The latter is however very rarely thought of when "risk" is being evaluated. Hence, risk is usually represented by $Risk = consequence \times threat \times vulnerability$; i.e., $Risk = consequence\ of\ event \times probability\ of\ event$.

Risk assessment is the systematic approach to organize and analyze scientific knowledge and information about potentially hazardous activities; simply stated, the analysis of risk; generally includes problem formulation, hazard assessment, exposure analysis, and risk characterization.

Risk management is the systematic application of policies, practices, and resources to the assessment and control of risk affecting human and system safety. It involves an analytical process to determine the likelihood that a threat will harm an asset or resource, and the identification of actions to reduce the risk and mitigate the consequences of an attack or event. Risk management principles acknowledge that risk generally cannot be eliminated but enhancing protection from known or potential threats can reduce it. A typical *Risk Management Model* would consist of the following steps:

- Asset assessment
- Threat assessment
- Vulnerability assessment

- Risk assessment
- Identification and implementation of cost-effective countermeasure options
- A continuous assessment of the above.

Further, successful risk management organizations have senior management who support and are involved in the process, employ the concept of *Risk Acceptance Authority*, and create procedures for establishing and tracking accountability. Another important element that contributes to the successful implementation of a risk management approach is the availability of an acceptable implementation framework or set of guidelines. The following sections provide an overview of such guidelines from the security management and IT governance perspectives.

4 Security management guidelines

The National Institute for Standards and Technology's (NIST) Risk Management Guide for Information Technology Systems provide recommendations for one of the basic frameworks for risk management (NIST SP800-30, 2002). It adopts the traditional definition of risk as a combination of likelihood and impact. It considers risk management as comprising of three processes: risk assessment, risk mitigation, and evaluation. Assessment includes identification and evaluation of risks and risk impacts, and recommendation of risk-reducing measures, mitigation involves prioritizing, implementing, and maintaining the appropriate risk-reducing measures recommended from the risk assessment process, and evaluation being a continual process of monitoring. The NIST Special Publication on Recommended Security Controls for Federal Information Systems (NIST SP800-53, 2005) builds on this further by providing guidelines in three general classes of security controls: management, operational and technical. All these efforts have however been mainly technically focused, both in terms of defining the problem space as well as the recommended solutions.

CERT[1] is a major research center for current thought on risk management in the context of information security (www.cert.org). It provides an ongoing array of guidelines in this rapidly changing domain. As such, it is a very important resource for managers and academics that have an interest in information security. The CERT Coordination Center (CERT/CC) is a federally funded research and development center specializing in security of networked systems at the Software Engineering Institute. It was established in 1988 and is the first computer security incident response team.

On the basis of federal and other emerging guidelines, it appears that it would be more insightful if we begin to look at information security more

[1]CERT is not an acronym, but a name and a registered service mark of Carnegie Mellon University.

from the perspective of its purpose and context, rather than one that is solely dominated by the physical arrangement of technology. By doing so, basic framing issues that emerge on the basis of the above guidelines and standards are:

- A secure technical infrastructure is a necessary, but not sufficient, condition for information and organizational security
- Security is an organizational problem, not just a technical problem, though much of the current work is techno-centric
- An emerging risk perspective of security is an effort to move to a mission-centric or organization-centric perspective
- Security is an organization-wide problem similar to what quality was in the 1980s, but it involves a more complex and dynamic context
- Security delivery is a process that has to be continually improved
- Typical security efforts are limited in their organizational approaches and limited by their asset focus
- Regulatory impact has to be studied and should include the risk of limiting efforts only to meet compliance requirements

These issues would lead us to consider risk, not just from a technology perspective, but also from an organizational perspective.

5 IT governance

Risk management frameworks for organizational controls are not new efforts. The two leading frameworks of Committee of Sponsoring Organizations of the Treadway Commission (COSO), and the more recent Control Objectives for Information and related Technology (COBIT), are well established in the business environment. The latter takes a broader IT governance perspective, wherein risk management is a subset. In addition to government standards and other domain expertise mentioned above, these frameworks provide an additional context within which a risk management framework is developed in this chapter.

The COSO issues Internal Control—Integrated Framework (Control Framework) has been widely accepted as the internal control standard for organizations. COSO requires five interrelated components: the control environment, risk assessment, control activities, information and communication, and monitoring. They have recently extended this by creating an *Enterprise Risk Management—Integrated Framework*. This views uncertainty both as a risk and an opportunity, with the potential to erode or enhance value. It enables management to effectively deal with uncertainty and the associated risk (downside risk) and opportunity (upside risk), and enhances the capacity to build value for an organization.

The essential elements of the COSO Enterprise Risk Management framework are (COSO, 2004):

- Aligning risk appetite and strategic alternatives
- Enhancing risk response decisions by providing the rigor to identify and select among alternative risk responses of risk avoidance, reduction, sharing, and acceptance
- Reducing operational surprises and losses by identifying potential events and establish responses
- Identifying and managing multiple and interrelated cross-enterprise risks
- Seizing opportunities to identify and proactively realize opportunities
- Improving deployment of capital.

COSO also recognizes that such enterprise risk management efforts help ensure compliance with laws and regulations, and helps avoid damage to the entity's reputation and associated consequences.

The Information Systems Audit and Control Association (ISACA) is globally recognized as the major provider of standards and controls for the general IT environment. In 1996, ISACA's affiliated foundation published the first version of COBIT® as a framework within which IT governance could be managed. Effective IT governance helps ensure that IT supports business goals, maximizes business investment in IT, and appropriately manages IT-related risks and opportunities. The framework and supporting toolset allows managers to bridge the gap between control requirements, technical issues, and business risks in an organization. COBIT, now in its fourth edition, is generally considered to be a leading governance, security, control, and assurance framework across the world (COBIT, 2005). Also of significance is the fact that the COBIT framework, along with the COSO framework, is considered to be critical in regulatory compliance with the US Sarbanes–Oxley Act of 2002.

The following IT processes, identified within the COBIT Framework, may be of particular interest in connection with deployment of technologies:

- Define a strategic IT plan
- Manage the IT investment
- Ensure compliance with external requirements
- Assess risks
- Identify automated solutions
- Ensure systems security
- Manage the configuration
- Manage problems and incidents.

COBIT also provides maturity models for control over IT processes, so management can map where the organization is today, where it stands in relation to the best-in-class in its industry and to international standards

and where the organization wants to be; critical success factors, which define the most important management-oriented implementation guidelines to achieve control over and within its IT processes; key goal indicators, which define measures that tell management whether IT processes have achieved their business requirements; and key performance indicators, which define measures of how well the IT processes are performing in enabling the goal to be reached.

6 Developing the risk management framework

The established view to risk management that is embedded in various standards forms the starting point for developing a more refined and comprehensive approach. The resulting risk management framework discussed below was developed in three distinct stages. While some of the following discussion is based on an analysis of existing frameworks, it also includes alternative approaches that are further developed in the next section.

6.1 Risk management framework stage 1

Defense-in-depth has been the traditional philosophic basis for most security initiatives. The same underlying philosophy can be seen in most information security efforts to-date. However, it is being applied in a very limited fashion by implementing it through multiple technology solution layers. NIST guidelines extend this by taking a limited socio-technical perspective and address issues related to human behavior (such as employee behaviors) in additional to technical issues. While this is a step in the right direction, it is still limited. As such a more comprehensive basis has to be developed.

Reason's (2000) framework from safety science provides a useful basis to better understand network security risk management. It also highlights the problem with the defense-in-depth approach as each defensive layer can have ''holes'' of vulnerabilities which can become ''aligned'' and thereby allow a specific attack ''through'' the multiple defensive layers. We adapt the Reason model to identify four distinct layers of defense for network risk management. These layers are:

- Technology risk management layer
- Process risk management layer
- Human risk management layer
- Financial risk management layer.

These layers comprise the managerial defense-in-depth that should be the foundation of any risk management framework developed.

The technology risk management approach is the typical mechanism that is adopted in all current standards and frameworks. Such frameworks are

limited to external threats and attacks. However, recent studies show that there is a significant internal threat. These are in the form of a few individuals with malicious intent but a majority of those who make unintended errors. Current efforts at understanding these issues and developing responses have been focused on personnel and behavioral issues. This comprises the human risk management layer. Gradually, there is a growing realization that the cost of providing network security can become prohibitive. More important is the realization that there is a need to understand the security investment portfolio better. For instance, with reduced security incidents comes the inertia in continued security spending. Also, in an increasingly resource constrained environment, an inadequate understanding of the security investment-returns relationship could very well result in an unbalanced security investment portfolio that could result in placing the network at significant risk. Hence, the financial risk management layer is very significant as it guides the rest of the risk management effort. But our framework focuses on the fourth layer, which is the process risk management layer. Technology is usually implemented within the context of the operational process of an organization. Hence, understanding and managing process risk plays a key role. This is further elaborated in the next stage of the framework development.

6.2 Risk management framework stage 2

A systems-theoretic approach was used as a first step to define an intuitively appealing *Security Management System* with the following main elements because organizations are open systems interacting with a complex operational, technical, and regulatory environment:

- Environmental context: Security regulation, security threats, security-related performance of competitors, and security solutions
- Organizational context: Security strategy, policies, and culture
- Security delivery process: Evaluate and promote awareness, assess and analyze risk, apply or implement security controls, and audit and monitor effectiveness
- Individual actions: Actions of individual participants.

Each of these subsystems, the elements there-in, and their interactions together form the security management system for communications networks. It is in this context that a risk management framework has to be developed to avoid a limited technology-perspective of the problem domain.

The security delivery process is the focus of the framework being developed here and can be identified as the *Continuous Security Improvement Process* comprising *Awareness, Assessment, Application, and Auditing*. This is elaborated in the next stage of the framework development.

6.3 Risk management framework stage 3

An integrated risk management framework is presented in Fig. 1 and discussed below.

6.3.1 Awareness

Consists of an *Awareness* of the environmental and the organizational contexts. This includes understanding related regulatory issues, adequate levels of enterprise security, and operational risk as shown in Fig. 1. For instance, some of the critical regulatory issues today for firms are the Sarbanes–Oxley (SOX) Act of 2002, the GLB Act, and HIPAA. But these regulations do not mention information security per se. However, internal controls and privacy—their focal point involve aspects of identity management, security access controls, document management, learning management, business intelligence, business process management, data and transaction monitoring, e-mail archiving, and records retention—and the mortar between all of these bricks is information security. SOX, for instance, makes no mention of information security, but it is overwhelmingly about corporate risk reduction. This includes regulatory risk, operational risk, and financial risk. In this context, companies have to demonstrate that appropriate IT and business process controls are in place and are working effectively. An integral part of this *Awareness* is an understanding and definition of an *Adequate Level of Security*. At this stage, the security strategies to protect an organization's assets and processes should be appropriate for the firm's risk appetite and tolerances (CERT, 2005a).

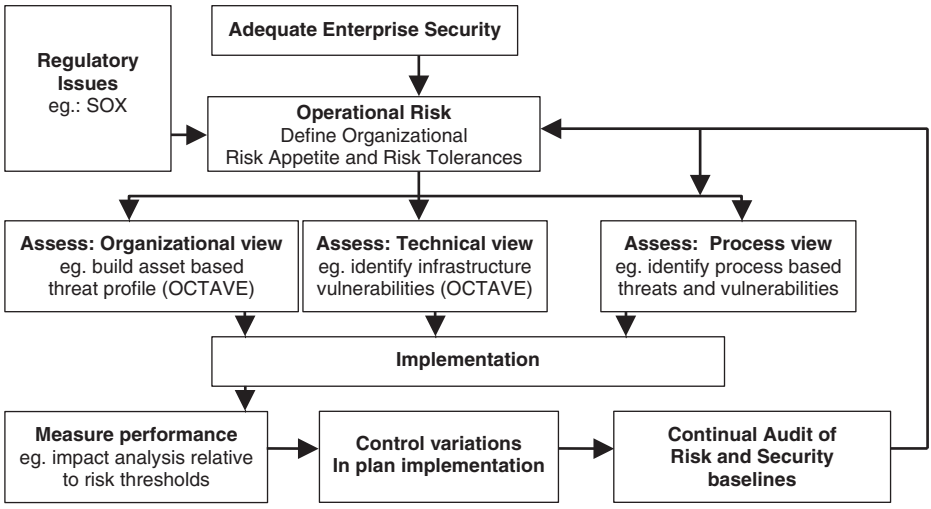


Fig. 1. Integrated framework for risk management.

6.3.2 Assessment

There are three distinct perspectives that have to be adopted here. The *Organizational view*, an example being the building of an asset based threat profile, and the *Technical view* that identifies infrastructure vulnerabilities, are both based on the OCTAVE® (Operationally Critical Threat, Asset, and Vulnerability Evaluation℠) approach to assessment that is gaining recognition and usage (CERT, 2005b). OCTAVE is a risk-based strategic assessment and planning technique for security developed by CERT as a standard approach for a risk driven, asset and practice based information security evaluation. However, we also specifically introduce the *Process view* that addressed identifying process based threats and vulnerabilities. This involves a process analysis from a security perspective rather than the typical productivity or quality perspectives that are more common to many organizations. This is discussed further in the next stage of development of the framework in the following section.

6.3.3 Application

The avoidance, reduction, and mitigation of the risks assessed through awareness and assessment methods comprise the implementation area shown in Fig. 1. It is easily seen here that technology solutions provide only part of the solution, while managerial solutions provide the rest. Such managerial solutions primarily address organizational and process issues in an organization. Raghu and Vinze (2005) provide a suitable framework where they present knowledge management in an operational context. Adopting their approach, we could address the security issues in organizations by analyzing information flow and workflow structures in a business process. In addition, decision-making processes that are embedded in business processes need to be evaluated from an information security perspective. The protocols for making information classified and de-classified from a security perspective in government is an example of how decision making processes are deeply embedded in business process, and need to be surfaced and analyzed in any security improvement effort.

6.3.4 Auditing

The two specific areas of interest that this framework highlights are the need to control the variation in implementation plans, and the need for a continual auditing mechanism. Both these are consistent with quality management philosophies that have been widely adopted over the past three decades. It truly reinforces the "security as a process" perspective. What is distinct in this framework is the call for continual auditing. The best approach to implementing such an idea is to consider high reliability organizations or HROs (Weick and Sutcliffe, 2001). HROs refer to organizations or systems that operate in hazardous conditions but have fewer than their fair share of adverse outcomes. These organizations are characterized by a preoccupation with failure that comes with an acknowledgement of the

high-risk error-prone activities in the organization, a commitment to resilience rather than perfection, a sensitivity to frontline operations, and a culture of safety in which attention is drawn to failure without fear of censure. The information security domain at large is nowhere near this level of performance, but can fruitfully aspire to it.

Current and anticipated industry benchmarks and legislation are on a trajectory that will increasingly mandate information security management practices. These include the need to evaluate and promote risk awareness, assess risk, implement security controls, and monitor/audit effectiveness of those controls. It can be expected that in the not-too-distant-future organizations will have to document their adherence to such industry standards and legislative initiatives in order to conduct commerce with other business partners.

7 Process-centric risk management framework

Figure 1 provides a general risk management framework with specific enhancements such as regulatory awareness, adequate security, assessing risk from a process view, controlling variation in implementation, and continual audit. This framework is now developed further. The following four significant gaps are identified upon studying existing risk management frameworks, and corresponding improvements are highlighted.

- *Gap 1*: While risk management follows a process in current frameworks, it takes an event-centric approach to analyze risk.
  Proposed Improvement 1: Efforts to ensure information security is an ongoing process, therefore the approaches to risk management should be process-centric.
- *Gap 2*: Organizations utilize assets and work processes to deliver on their missions, but current security efforts typically focus on assets only.
  Proposed Improvement 2: We specifically include work processes as targets of a risk analysis framework.
- *Gap 3*: An organizational view of risk appears to typically begin with assessing and analyzing threats to, and vulnerabilities of, assets.
  Proposed Improvement 3: We specifically begin addressing risk at the enterprise level where operational risk analysis begins with loss threshold and risk tolerance; and these are addressed before any specific risk analysis is conducted.
- *Gap 4*: Impact analysis usually involves analyzing loss magnitude, with the probability of occurrence used in certain cases.
  Proposed Improvement 4: The use of probability is not easily operationalized for rare events and magnitude of loss is relative; so we

specifically suggest use of ''possibility'' loss appetite and risk tolerance using a ''threshold'' model.

Based on these, a process-centric risk management framework is now presented in Fig. 2. This is in direct contrast to the event-centric risk management approach that is the dominant paradigm today.

A security management process-centric framework is built upon the process management concepts that are established in quality management. Here ''delivering security'' is considered a process. As such, since all processes

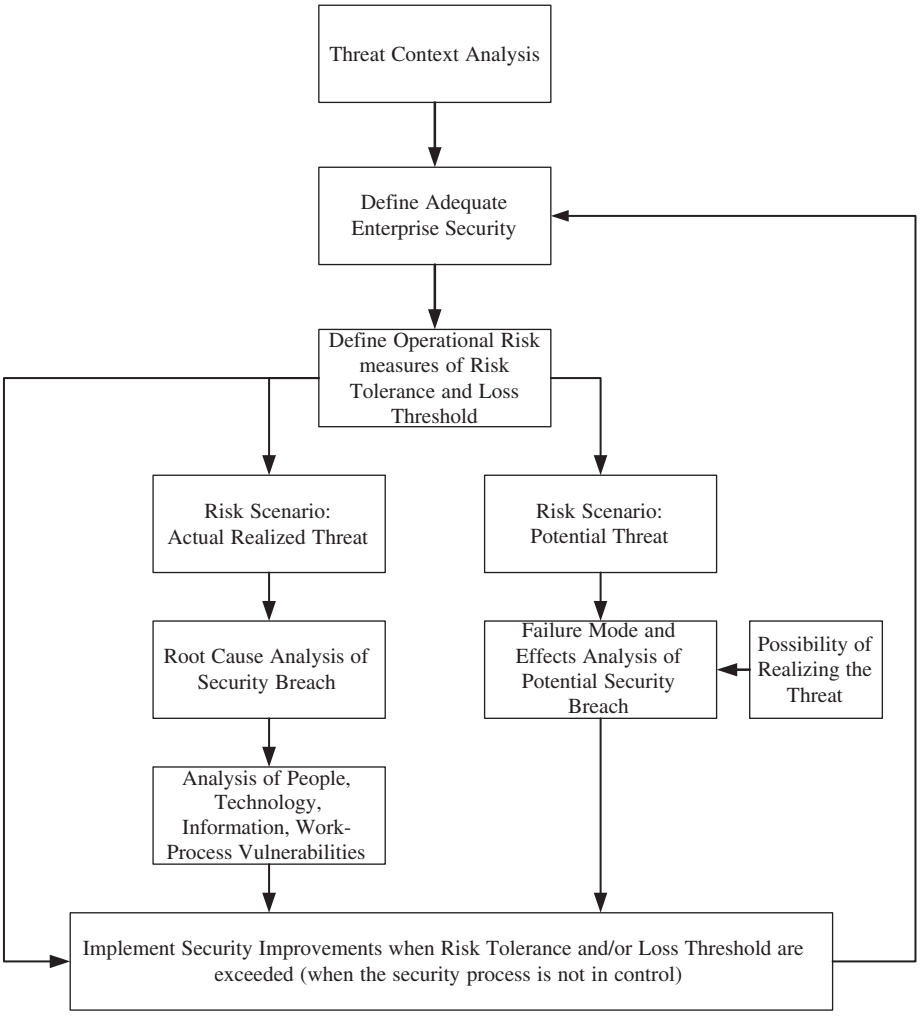


Fig. 2. Process-centric risk management framework.

have variation, it is a foregone conclusion that there will always be some "variation" in the extent of security afforded to users. Such variation in security can be due to assignable causes (that have to be identified and removed) or common causes (that have to be designed out). This framework focuses on the identification and elimination/reduction of causes for poor security. This, in turn, can be achieved by analyzing when a security breach has occurred by using a root-cause analysis, or proactively using failure mode and effects analysis.

Realized breaches in security analyzed using the above approach would produce targeted solutions addressing all categories of causes. Specifically, the many work-process related security incidents discussed at the beginning of this chapter can be analyzed within this framework to develop solutions, define the risk tolerances, and define loss thresholds that a firm can accept. Security can also be addressed by taking a pro-active stance and using a failure mode and effects analysis approach. It is only in this context that the probability of an event occurring is relevant. But probabilities cannot be calculated when events have not occurred in the past but are likely to occur in the future due to changes in the threat environment. As such it is more useful to consider "possibilities" and not probabilities. This is a major departure from the current mindset in security that is based on probability of an event occurring and its associated losses.

The new framework is process-centric and is focused on preventing repeat breaches by closing gaps identified through a root-cause analysis, and by closing potential gaps identified through a possibility-based failure mode and effects analysis. In doing so, the framework is action-oriented while at the same time addressing areas of technical and organizational issues. Concepts of adequacy, thresholds and tolerances are byproducts of managing this process within acceptable "control limits". It significantly alters the traditional framework (as shown in Fig. 1) by understanding the causes first, followed by solution options, finally control mechanisms to ensure consistent or reduced variability in the delivery of security.

8 Conclusion

The development of risk management frameworks is an ongoing effort by various organizations and government agencies as a part of their guidelines for information security. This chapter initially extends the defense-in-depth philosophy to a managerial level so as to highlight the importance of technology, business process, human effort, and financial management in developing a risk management framework. Then a systemic perspective is adopted to generalize it. Finally, by consciously stepping away from the traditional technology-oriented event-centric view of risk management, this chapter develops a process-centric framework with which to understand and manage information security.

We can see parallels in the development of various risk management frameworks to the developments of quality management frameworks that occurred during the early part of the quality management movement more than two decades ago. As in quality management efforts of those days, we find many different risk management frameworks being put forward as part of the current broader information security efforts. We can expect that these risk frameworks will eventually evolve to a well-defined generalized framework, as it did in quality management with the development of the Malcolm Baldrige National Quality Award criteria by the NIST. This chapter is a contribution to that ongoing evolution of a comprehensive risk management framework for information security.

9 Discussion questions

1. What is the risk management framework adopted by your organization (or one you are familiar with) for information security? Do they subscribe to any specific set of guidelines such as those from NIST, CERT, COSO, or COBIT?
2. Discuss how your organization (or one you are familiar with) implements a "defense-in-depth" strategy for information security.
3. Undertake a root cause analysis of a security breach that you are familiar.
4. Analyze a business process you are familiar with to determine possible causes of a security breach maintaining an emphasis on work-process causes.
5. Discuss the security protocols for portable information and mobile communication and information devices in your organization (or one you are familiar with)? What is the extent of their actual use by employees?

Acknowledgment

This research was partly funded by the Defense Information Systems Agency of the U.S. Department of Defense.

References

CERT. (2005a). Computer Emergency Readiness Team (CERT), Governing for Enterprise Security, How Much Security Is Enough? http://www.cert.org/governance/adequate.html accessed on 3rd December 2005.

CERT. (2005b). OCTAVE® (Operationally Critical Threat, Asset, and Vulnerability Evaluation℠), http://www.cert.org/octave/accessed on 3rd December 2005.

COBIT. (2005). Control Objectives for Information and related Technology COBIT 4.0, IT Governance Institute, http://www.isaca.org/ accessed on 3rd December 2005.

COSO. (2004). *Enterprise Risk Management: Integrated Framework*, September 2004, http://www.coso.org/publications.htm accessed on 3rd December 2005.

DHS. (2003a). *The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets*, The Whitehouse, Washington, D.C., February 2003, http://www.dhs.gov/dhspublic/display?theme=31&content=463 accessed on 3rd December 2005.

DHS. (2003b). *The National Strategy to Secure Cyberspace*, The Whitehouse, Washington, D.C., February 2003, http://www.dhs.gov/dhspublic/display?theme=31&content=935 accessed on 3rd December 2005.

Gallagher, R. B. (1956). *Risk Management: A New Phase of Cost Control*, Harvard Business Review, September–October 1956.

NIPC. (2002). *Risk Management: An Essential Guide to Protecting Critical Assets, National Infrastructure Protection Center*, November 2002, http://www.securitymanagement.com/library/NIPC_Risk0203.pdf accessed on 3rd December 2005.

NIST SP800-30. (2002). *Risk Management Guide for Information Technology Systems*, Recommendations of the National Institute of Standards and Technology, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, July 2002, http://csrc.nist.gov/publications/nistpubs/index.html accessed on 3rd December 2005.

NIST SP800-53. (2005). *Information Security: Recommended Security Controls for Federal Information Systems*, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, February 2005, http://csrc.nist.gov/publications/nistpubs/index.html accessed on 3rd December 2005.

O'Connor, J.J., E.F. Robertson (2005). *Biography of Blaise Pascal* (1623–1662), http://www-groups.dcs.st-andrews.ac.uk/~history/Mathematicians/Pascal.html accessed on 3rd December 2005.

Raghu, T.S., A. Vinze (2005). *A Business Process Context for Knowledge Management*, Decision Support Systems (in press), available online at http://www.sciencedirect.com/science/journal/01679236 from 14 July 2005.

Reason, J. (2000). *Managing the Risks of Organizational Accidents*, Ashgate, Aldershot, England.

Wailgum, T. (2005). *Security: 50-Cent Holes*, CIO Magazine, 15 October 2005, http://www.cio.com/archive/101505/security.html accessed on 3rd December 2005.

Weick, K.E., K.M. Sutcliffe (2001). *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. Jossey-Bass, San Francisco, CA.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 16

Intrusion Detection and Information Infrastructure Protection

Xiangyang Li
University of Michigan, Dearborn, MI 48128, USA

Nong Ye
Arizona State University, Tempe, AZ 85287, USA

Abstract

Intrusions into computer and network systems have presented significant threats to these critical infrastructures in providing continued service. Intrusions exploit the vulnerabilities in computer systems and take different forms of attack scheme, compromising the confidentiality, integrity, and availability of victim systems. Intrusion detection has become an important part of assuring the service quality of computer system. An intrusion detection system (IDS) provides service through the on-line monitoring of system activities and the detection of any attempts to break into computer and network systems and to compromise normal services. This article serves as a comprehensive introduction into recent research efforts in intrusion detection and a deep analysis for insights of latest trends in intrusion detection technique development, challenges still faced by modern IDSs, and future research directions. Compared with the existing literature on intrusion detection and its application in information infrastructure protection, we employ a novel angle from information and knowledge engineering to the important topics relevant to intrusion detection. This approach is centered on organizing and understanding intrusion detection techniques within the whole lifecycle of the data and knowledge essential to this task, from data collection to processing algorithms and models to information fusion in complex and distributed environments. Therefore we place intrusion detection in the big picture of information-centered and network centric information age warfare, an essential view that today's information technology researchers should hold. The balance between research and practice is kept while writing this article, aiming to provide helpful information for practitioners also.

1 Introduction to intrusion detection

Three metrics generally measure the security of services provided by a computer system or network (Bishop, 2005). Confidentiality ensures that only authorized users have access to the information content or the available service. Integrity pledges that correct information, not tampered during processing and transmittance, is created and issued by the right party. Availability assures the promised capability in terms of speed and time when an information system performs services to customers, an important part of quality of service (QoS) measures for a computer system. Therefore any activities violating or degrading these three security measures of a computer system are intrusions into the computer infrastructure. When these activities occur, we can directly or indirectly observe that the quality of an information service degenerates in terms of interruptions, delays, transaction errors and discrepancies, and lost identity.

Over decades of evolvement, intrusions take various forms of physical exploitation, malicious code, virus, worms, coordinated service requests, or social engineering means as in more and more phishing and identity thefts. The most common examples are virus and worms that can spread to thousands of computers in just a couple of hours, via various media such as email and network connections (Denning, 2004). They can make damages to various resources in a computer system, including loss of data and disclosure of critical information. While the past studies largely examine attacks that unlawfully expose confidential content and illegitimately manipulate data and identity during information processing and transmission, violations against availability of services receive more and more attention. One such type of attack is the distributed denial of service (DDoS) (Mirkovic and Reiher, 2004). In this attack, a set of computers coordinate to send out large volume of malicious network traffic in a short period of time in order to deprive the victim computer of resources required for the normal service. Normally attackers start from one or several computer hosts at different locations of the network to take the control over more computers at other locations. Then those computers (called ''zombies'') launch denial of service (DoS) attacks to the victim computer.

1.1 Intrusion detection data and models

User requests and service responses change the state of a computer system and incur new events. Such state and event changes can be captured by recording facilities such as the very commonly used log systems in many computer platforms. Several examples of these facilities include the Solaris Basic Security Module (BSM) and the Windows log program for auditing a single computer host, the Tcpdump and Windump typically used to capture network traffic packets, and the available administration

functions to monitor and analyze network connections in most firewall products. One data record example from processing the binary audit data generated by the Solaris BSM module is given in Table 1, showing a set of attributes such data can contain. This type of BSM data was used in the DARPA Intrusion Detection Evaluation Program in 2000 (http://ideval.ll.mit.edu/).

Considering the above computer audit data or network traffic data as observable signals emitted from underlying computing processes, an intrusion detection system (IDS) tries to determine the intention of a network connection, a computer access session, or a running program to be benign or malicious. The category of intention can be "normal" or a specific type of attack if "malicious". An alarm is generated if it is not normal. A function to determine such a classified category is called an intrusion detection algorithm or model. Several examples of intrusion detection models are association rules analysis (Lee et al., 1999), statistical test (Ye et al., 2001), decision tree (Li and Ye, 2003), clustering (Li and Ye, 2002, 2006), and Bayesian networks (BNs) (Valdes and Skinner, 2000). These techniques and algorithm are from a variety of knowledge fields such as machine learning, pattern recognition, statistics and probability, artificial intelligence, and so on.

Traditionally we consider that intrusion detection essentially performs a classification or pattern recognition task, i.e., recognize attacks from normal activities and in a better effort distinguish among specific attack types. Thus an analysis engine residing in the IDS, or mathematically a classification algorithm or model, reports the current status in terms of a specific intrusive level based on input of audit data or network traffic data. Existing efforts in developing and building intrusion detection models have

Table 1
Some attributes provided in the BSM audit data of Solaris operating system

Attribute	Type	Description
Event	Nominal	Audit event type
Auid	Nominal	Audit user id
euid	Nominal	Effective user id
egid	Nominal	Effective group id
ruid	Nominal	Real user id
rgid	Nominal	Real group id
pid	Nominal	Process id
sid	Nominal	Session id
RemoteIP	Nominal	Remote host IP address
time	Numeric	Occurrence time stamp
error_message	Nominal	Error message
process_error	Nominal	Process error status

considered mainly patterns in the following attributes of activity data in information systems:

- occurrence of individual events, e.g., audit events, system calls, commands, error messages, IP source address, and so on;
- frequency of individual events, e.g., number of consecutive password failures;
- duration of individual events, e.g., CPU time of a command, and duration of a connection;
- occurrence of multiple events combined through logical operators such as AND, OR, and NOT;
- frequency histogram (distribution) of multiple events, and sequence or transition of events;
- sequence or transition of events.

In a formal notation, a set of attributes or predictor variables, $X = (X_1, X_2, \ldots, X_p)$, as in audit data or network traffic data, represent the data record collected for an intrusion detection task. A target variable, Y, represents the nature (normal/intrusive or finer categories such as DoS attack, worm, and so on) of the activity generating this data record. Therefore this intrusion detection task is to assign a class to the target variable of this data record, a classification into normal or intrusive categories, relying on a certain function, f: $X \rightarrow Y$. As in typical classification applications, the parameters of such a function can be learned in a training stage on some carefully collected data, called training data, before this function is used in classification. The nature (ground-truth) of training data is known to this function in the training stage.

1.2 Challenges to intrusion detection systems

Unfortunately an IDS faces enormous challenges inherent in complex computer systems. First complex dynamics is attributed to complicated information flow, heterogeneous structure, constantly changing topology, and rapid growth of a computer system. Various protocols, standards, and platforms complicate the situation. Modern computing infrastructure spreads extensively both functionally and geographically. Intrusion detection algorithms themselves have different characteristics of accuracy, speed, and reliability. Second enormous uncertainty and noise exist in this system, accumulated through different network and application layers. Hackers often use disguise by generating ambiguous and misleading traffic. New types of attack emerge constantly each year. Until now no clear characterization exists for the underlying data distribution for normal and malicious activities in such computer systems. And lastly performance requirements and constraints make IDSs subject to severe stress. Computing cost and processing speed can be seriously limited, especially in wireless networks with serious constraint on power consumption and computation

capacity. Within such an extremely complex environment, it is a consensus that intrusion detection has a long way to go to achieve the desired functionality in practical deployment.

1.3 Structure of this chapter

After the above overview of intrusion detection tasks and techniques, detailed discussion on intrusion detection techniques starts with a generic classification of intrusion detection models. Categorizing existing techniques into anomaly detection and misuse detection helps readers gain insights of the basic problems faced by intrusion detection, with detailed description and discussion of each category and the hybrid IDS that integrates these two categories of techniques. A new paradigm called cyber attack-norm separation is briefly described with pointers given after that, representing the latest research advance on intrusion detection models. Then the discussion shifts to tackling intrusion detection tasks from the perspective of knowledge engineering and information fusion in distributed environments, with topics on sensor management and information fusion, and coordination and optimization of distributed IDSs. This chapter is concluded by further discussion on the relationship of intrusion detection with other information assurance tasks for information infrastructure protection, namely intrusion prevention and intrusion response.

2 A generic classification—anomaly detection and misuse detection techniques

Since the seminal rule-based pattern-matching model by Denning (1987), great demand has witnessed a variety of intrusion detection techniques and systems applying diverse theories, computational algorithms, and models of computer systems and computer users. An appropriate classification of these IDSs can help identify important assumptions behind various techniques and models, leading to integration of existing techniques and development of new solutions. For example, a system can be classified into host and network IDS, based on the type of used data that is attributable to host machine or network traffic. Using different criteria, more than one classification can exist for the same algorithm and model applied in an academic or commercial IDS. In this section a classic classification system is discussed that organizes existing techniques into two categories, anomaly detection and misuse detection (also called signature recognition), according to the assumption for the definition of "intrusions" (Axelsson, 2000; Debar et al., 1999).

A basic assumption for IDSs to work is that there is identifiable difference between the signals from the normal operations and the intrusive operations. Further from this assumption, assumptions in anomaly detection and misuse

detection techniques follow different perspectives to define an "intrusive" operation. Anomaly detection considers any deviation or difference from the normal patterns or profiles present in the available historic data (used as training data) as an anomaly and thus an intrusion. Therefore an anomaly detection system requires a training dataset representing normal activities of a computer system, ideally capturing enough aspects of normal usage and behavior. On the other hand, patterns summarized from the past intrusion records or the mechanism of known attack types can function as intrusion "profiles" and can be used to match patterns in misuse detection or signature recognition. The occurrence of a match between the monitored activity and the existing intrusion pattern leads to an intrusion alarm. Consequently a misuse detection system requires a set of training data covering as much as possible the behaviors of targeted types of intrusion in order to learn "intrusive" patterns or signatures. In misuse detection systems that build up such signatures manually, human experts identify the required conditions and steps for intrusions to succeed, and describe them using certain suitable languages such as a state transition diagram using state nodes and transitions (Ilgun et al., 1995).

We can use a Venn diagram as shown in Fig. 1 to show the relationship among known (in training stage) and unknown normal and intrusive activities in a cyber space for IDS to work in. In anomaly detection activities outside the known normal activity area are classified as possible intrusions, causing the normal activities unknown in training stage to be misclassified. The trouble with the misuse detection comes from the unknown intrusive activities that elude detection because their signatures are not captured in the training stage.

2.1 *Anomaly detection*

Various algorithms and techniques can be applied for the objective to build up a suitable profile of normal operations in a computer system and estimate

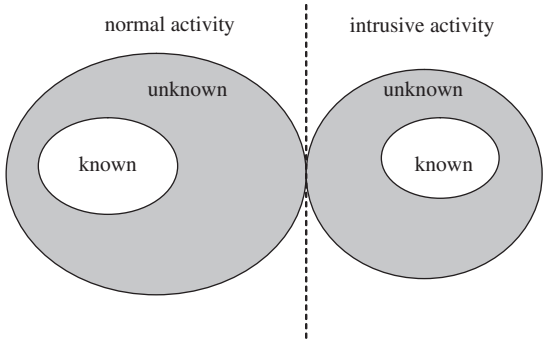


Fig. 1. Un/known normal and intrusive activities in a computer system shown in a Venn diagram.

the deviation of new activity data from this profile. Such a normal profile can be about the different objects or entities in the targeted system, e.g., users, sessions, processes, services, files, network resources, and even the computer system as a whole. Therefore it is very natural to create and manage a set of normal entity profiles at different physical layers and abstraction levels of the target computer system. An incomplete list of algorithms and models used in existing anomaly detection systems includes logic-based profiling, artificial neural networks, regression, computer immunology, Markov chains, BNs, hidden Markov models, and statistics-based profiling. Theoretically any algorithms that are able to extract and represent behavioral patterns from observable data are potential candidates for anomaly detection.

The norm profile can be captured and coded manually or automatically. For example, a specification-based anomaly detection system can manually encode the security policies into the detection engine, such as the required authority to access certain files (Ko et al., 1997). An automatic profiling technique is desirable if we consider the growing complexity of computer systems and activities. For example, anomaly detection techniques based on artificial neural networks (Debar et al., 1992) learn from historic data to predict the next event from a series of the past events. Anomaly detection techniques based on immunology capture a large set of event sequences as the norm profile from historic data of a subject's normal activities, and use either negative selection or positive selection algorithms to detect the difference of incoming event sequences from event sequences in the norm profile (Forrest et al., 1997). There are also anomaly detection techniques that use a first-order or high-order Markov model of event transitions to represent a norm profile.

We discuss the technique to use statistical process control (SPC) technique to build up profiles of normal activity data and to detect the deviation in classification. In this example we use two statistics, namely Hotelling's T^2 and Chi-squared χ^2 test metrics, as defined below on a multivariate data:

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \tag{1}$$

$$\chi^2 = \sum_{i=1}^{p} \frac{(Xi - \bar{X}_i)^2}{\bar{X}_i} \tag{2}$$

where S is the variance–covariance matrix and $\bar{X}$ the sample mean vector.

In training, a sample of normal behavior data is used to calculate the mean vector and the variance–covariance matrix. In classification, a large T^2 or χ^2 score calculated for a new data record represents the deviation from the in-control population, here the normal data distribution space. We can choose different control thresholds to decide the boundary of the desirable control (norm) area. If the calculated deviation score is larger than the threshold, we will signal the associated data record as intrusive. Therefore the mean vector, the variance–covariance matrix, and the threshold together

stand for the profile of normal behavior. We call the norm profile (or attack signature) together with the algorithm used in training and detection of a specific intrusion detection model.

Same as a general classification problem, the performance of intrusion detection can be characterized using a confusion matrix of accuracy or recall scores. However here we are more interested in two performance measures, the hit rate and the false alarm rate. If there is a signal on a data record from an intrusive event in the testing data, this is a hit. If a signal is generated on a data point for a normal event, this is a false alarm. The detection rate or hit rate is computed from dividing the total number of hits by the total number of intrusive events in the testing data. The false alarm rate is computed from dividing the total number of false alarms by the total number of normal events in the testing data. Therefore a receiver operator characteristic (ROC) curve is usually employed to plot these two measures, associated with different control thresholds as used in the T^2 or χ^2 techniques. Each point in the ROC curve indicates a pair of the hit rate and the false alarm rate for a specific control threshold. The closer the ROC is to the top-left corner (representing 100% hit rate and 0% false alarm rate) of the chart, the better detection performance the intrusion detection technique yields.

The T^2 or χ^2 techniques were evaluated using some training and testing datasets from the Intrusion Detection Evaluation Program of the Defense Advanced Research Programs Agency (DARPA) (http://ideval.ll.mit.edu/). More details about this study on the T^2 or χ^2 techniques and the datasets are available in the paper by Ye et al. (2001). The test results show the χ^2 test can yield better detection performance in terms of false alarm and hit rates for certain control thresholds. The χ^2 multivariate test detects mainly mean shifts. Hotelling's T^2 test detects both mean shifts and counter-relationships because the T^2 test statistic is determined largely by the correlated structure of variables (variance–covariance matrix). Hence, this means shifts may be more important to intrusion detection than counter-relationships for this data.

We should also compare different intrusion detection models about their computational cost and scalability in dealing with large volumes of training or detection data. The χ^2 metric is better here since Hotelling's T^2 test is computationally intensive, requiring a large memory to store the variance–covariance matrix and much computation time to compute the matrix and its inverse. Further from this discussion we have to pay attention to the underlying assumption on data distribution in order for these statistical tests to work. For example, you can not expect good performance from either of T^2 and χ^2 tests if the normal data has several centers distributed in the multivariate space.

2.2 *Misuse detection*

The misuse detection or signature recognition technique takes a different view in identifying possible attacks. The signatures of known computer

attacks, i.e., patterns learned in an automatic training stage or extracted manually, are used to match new attack instances in detection. Various techniques from statistics and machine learning have been used for misuse detection, including string matching, state transition, Petri nets, rule-based systems, expert systems, decision trees, association rules, artificial neural networks, genetic algorithms, BNs, and so on. For example, a decision tree can be built up by recursively splitting important predictors or attributes on a training dataset of both intrusive and normal activities, with each path from the root node to a leaf node representing the signature of certain type of activity, e.g., a specific kind of computer attack (Ye et al., 2000). Association rule analysis, also called frequent itemset analysis, can capture the concurrence patterns of itemsets observed in the data from intrusive activities (Lee et al., 1999). Those itemsets can be any system values, such as a specific port number or IP address. Such concurrence patterns can evolve along time or other important system variables. Other than the inherent problem in dealing with new attacks, misuse detection techniques also face challenges of learning signature patterns from large amounts of activity data in a scalable and incremental manner, due to the constant evolvement of computer attacks.

A misuse detection technique based on clustering has been developed in response to the above challenges. This Clustering and Classification Algorithm—Supervised (CCAS) uses supervised clustering for learning patterns of normal and intrusive activities, and instance-based learning to classify observed activities, combining the advantage of clustering to deal with arbitrary data distributions and the simplicity and power of instance-based learning (Li and Ye, 2002, 2005).

In the core of this algorithm lies a grid-based incremental supervised clustering in the training stage that considers the distance among data points as well as the class in target variable, given in Fig. 2. Each dimension of the data space is divided into a set of intervals within the range defined by the minimum and maximum values of data points, separating the space into "cubic" cells. This clustering is incremental in that each time only one new data record is considered within the existing cluster structure. Either this new data point is added into the cluster structure or a new cluster is created to contain it.

At any instance during the above supervised clustering, the cluster structure considers only the data points processed so far, reflecting a local view on training data. Thus several post-processing steps are applied to strengthen the CCAS algorithm, which can be flexibly arranged in certain workflow. Redistribution of data points is a common way to remedy the localization in incremental clustering. This redistribution process can be repeated many times. The supervised hierarchical grouping procedure re-groups the clusters that may be split into too small clusters because of the existence of grids. This algorithm is different from the traditional hierarchical clustering in that it combines a pair of closest clusters only when they

```
For each data point D in training data
    Find the grid cell of D
    For each existing cluster L_i
        If D and L_i are in the same grid cell
            determine whether L_i is the nearest cluster to D
    End
    If there is a nearest cluster L_n and its class label is same as D
        incorporate D into L_n
        update the centroid coordinates
        update the number of points in L_n
    Else create a new cluster containing only D
        the label of this cluster is same as the class of D
End
```

Fig. 2. The grid-based incremental supervised clustering in the training stage.

have the same class. Clusters that have few data points may represent noises in data samples and can be removed in the outlier removal procedure. Each of these steps in the workflow functions independently and can be arranged flexibly, linked by the clusters in input and output.

The resulting cluster structure represents the patterns (signatures) of normal and intrusive activities. In classification, we classify a new data point by comparing new data points with the nearest clusters to it. The classical instance-based learning methods can be used to assign the class label to each new data point.

A network traffic dataset is used to evaluate this CCAS technique, with some data attributes shown in Table 2. It contains features extracted from the network connections for the 1998 DARPA Intrusion Detection Evaluation Program. The training data of both normal and intrusive activities includes ~5 million connection records of 7 weeks of network traffic. A connection contains a sequence of TCP packets within a defined time period for data flows from a source IP address to a destination IP address using an application protocol. For each connection features are extracted from raw connection records. Each connection has a data record consisting of the extracted features for this connection. For example, "serror_rate" is a feature that represents the percentage of connections showing "SYN" errors. Among these features, there are 34 numeric attribute variables and 7 nominal attribute variables. Attacks fall into four main categories of Probe (surveillance and other probing), DoS, R2L (unauthorized access from a remote machine), and U2R (unauthorized access to local root privileges). More details about this dataset are available at the website of the third International Knowledge Discovery and Data Mining Tools Competition (KDD Cup'99) (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html).

Table 2
Sample attributes in the network traffic data record

Attributes	Data type	Description
Duration	Numeric	Length of the connection
Service	Nominal	Network service on the destination
Flag	Nominal	Normal or error status of the connection
src_bytes	Numeric	Number of data bytes from source to destination
dst_bytes	Numeric	Number of data bytes from destination to source
num_failed_logins	Numeric	Number of failed login attempts
su_attempted	Numeric	1 if "su root" command attempted; 0 otherwise
count	Numeric	Number of connections to the same host as the current connection in the past 2 sec
srv_count	Numeric	Number of connections to the same service as the current connection in the past 2 sec
serror_rate	Numeric	Percentage of connections that have "SYN" errors

According to the confusion matrix for CCAS on the testing dataset that contains ~300,000 connection records, the CCAS algorithm is comparable to the winning algorithms in the KDD Cup'99, with enhanced capability of incremental and scalable learning important to real-world deployment. Please refer to the reports by Li and Ye (2001, 2005) for more details on this CCAS algorithm and the experimental results.

2.3 *Hybrid intrusion detection systems*

Intrusion detection systems still face great challenges in real-world deployment after rigorous research efforts in the past years. Strict requirements of almost 100% detection rate and near-zero false alarm rate fail many intrusion detection models in practical environments. False alarms especially cause big nuisance in a practical task setting. Consider that benign computer operations normally dominate the data generated in a computer system. Let us assume that 99.9% of all data records are from normal activities, which is not uncommon in real world. An IDS of 99.9% detection rate and 0.1% false alarm rate has seemingly very good performance. Counterintuitively, according to the Bayes theorem we will only be 50% confident that an alarm raised by this system is indeed caused by an attack. Such detection quality is not acceptable at all to those administrators who deal with large computer networks where millions of events happen in each day.

As we compared before (refer to Fig. 1), the two main categories of intrusion detection techniques, i.e., anomaly detection and misuse detection, have advantages and disadvantages respectively. In fact these two categories of technique complement each other in terms of detection ability

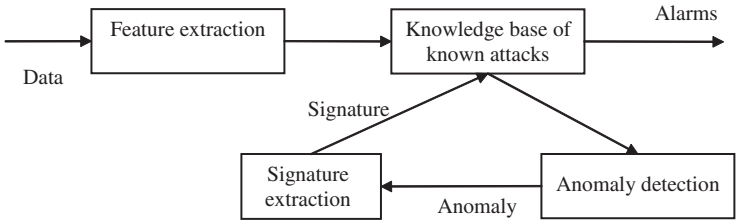


Fig. 3. The structure of a hybrid intrusion detection system.

of new types of attack and false alarm rate. It is natural to design a hybrid IDS that applies these two techniques together to improve the overall performance. For example, we can design a hybrid system as shown in Fig. 3. This design is very similar to the so-called MINDS system developed at the University of Minnesota (Ertoz et al., 2004). In this structure the known attack detection module detects attacks that have signatures in the knowledge base in this system after the feature extraction module preprocesses audit or network traffic data. After removing those attack components, the remaining data is fed into the anomaly detection module for connections with high anomalous scores. Then these anomalous connections are examined automatically or by human analysts. Further analysis based on algorithms such as association pattern analysis can be carried out to extract signatures of emerging attacks, which will then be stored in the known attack knowledge base for use in the future.

A hybrid IDS can function as the technical foundation for efforts to develop an artificial immune architecture. Such a system tries to beat the epidemic pace of viruses in order to take the situation in control before a serious "infection" happens over computer networks. Some exploratory studies are done by the researchers at IBM (Kephart, 1994; Kephart et al., 1997). Besides the capability to recognize known intrusions and provide elimination/neutralization of intrusions, this system has to learn previously unknown intrusions and provide quick recognition and response. An anomaly detection engine can provide more general functions than the suggested heuristics in those studies, such as the checksum to detect unknown file infectors. The algorithms that can automatically extract the signatures of unknown attacks are the key in the design of such hybrid systems.

The integration of anomaly detection and misuse detection can be very flexible in the design of a hybrid IDS. For example, if the anomaly detection engine is much faster than the misuse detection engine, it can function as the information filter at the data entry. Only the signaled anomalous data will be fed into the intrusion signature database to reduce false alarms. Thus the misuse detection module processes less data and has less chance to become the system bottleneck.

3 A new paradigm for intrusion detection

As we discussed before, the two existing approaches to detect cyber attacks on computers and networks have shortcomings related to detection accuracy and efficiency. A new approach, called attack-norm separation (Ye and Farley, 2005), is proposed recently. The new paradigm is based on the observation that a natural mapping exists between cyber intrusion detection and signal detection in the physical space (e.g., radar and sound signal detection), which often employs models to deal with both signal and noise simultaneously, i.e., all elements that exist and are mixed together in data. An example is the cuscore model that considers both noise and signal models for detecting a sine wave signal buried in random noise (Box and Luceno, 1997). Box and Luceno provide cuscore models to detect other types of signal and noise, such as parameter change signals with the noise of a first-order autoregressive time series model or the nonstationary disturbance noise of an Integrated Moving Average (IMA) time series model.

A signal detection model is sensitive to low signal-to-noise ratios, a situation often true in cyber attack detection since there are many more normal users than attackers. This attack-norm separation approach engages in the scientific discovery of data, features, and characteristics for cyber signal (attack data) and noise (normal data). Equipped with well-established signal detection models in the physical space (e.g., radar signal detection), the approach builds attack-norm separation models that incorporate the characteristics of both cyber signals and noise. This new paradigm considers not only activity data, but also state and performance data along the cause–effect chains of cyber attacks on computers and networks.

The cyber attack-norm separation allows for more accurate results and can replace signature recognition. The coverage on types of attack of the cyber signal-noise separation approach will expand with increasing knowledge of signal and noise characteristics, just as signal detection knowledge and technologies in the physical world evolved. Ultimately, the scientific knowledge of cyber signal and noise characteristics will grow to a sufficient level to replace anomaly detection as well. This will lead to the confidence in detection accuracy, efficiency, and adequacy, paving the way for practical applicability. More details can be found in Ye and Farley (2005).

4 Information fusion in distributed intrusion detection environments

A view to put intrusion detection in parallel with the sensor and information processing concepts can help design and implementation of IDSs in a distributed environment. Fig. 4 illustrates a typical distributed intrusion detection system (DIDS) structure, where a "report-control" cycle occurs between local IDS sensors and a regional IDS manager, which is the central manager for this sub-network. At any time instant the local IDSs report

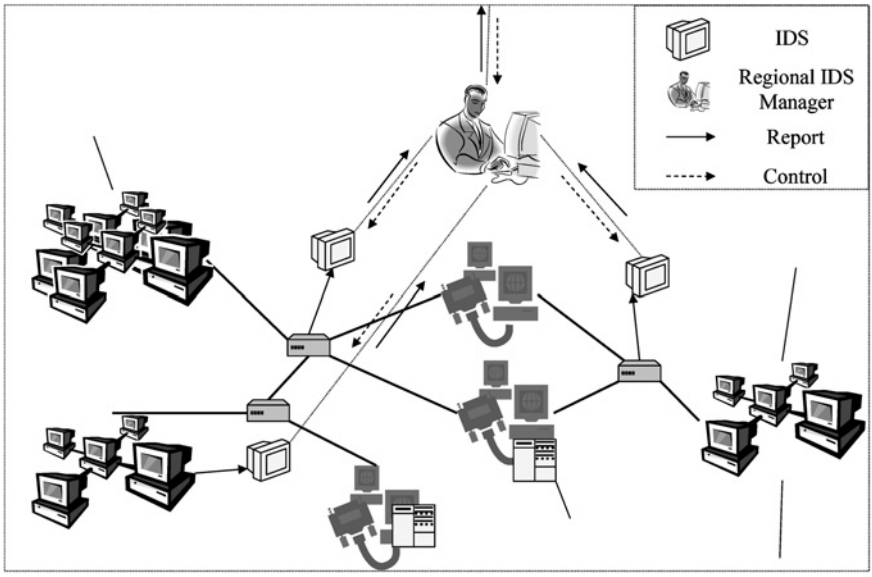


Fig. 4. An illustration of a distributed intrusion detection system, where local IDSs collaborate with a regional IDS manager.

findings including the intrusive level, as well as collected information required for further analysis at the regional level. The information from a set of sensors is integrated at the regional IDS manager. The feedback from this manager controls further information collection and intrusion detection actions of the local IDS sensors until the suspicious hypothesis about this region is either rejected or confirmed. This scheme could be easily extended into a hierarchy incorporating high-level IDS managers. In such an architecture, every IDS functions as not only a sensor but also an information monitor for the manager at a higher level.

4.1 Challenges in a distributed intrusion detection sensor network

Recently there are a few research efforts toward developing distributed intrusion detection architecture. The Common Intrusion Detection Framework (CIDF) protocol is developed to exchange attack information between IDSs that are distributed in a computer network (Staniford-Chen et al., 1998). This protocol defines a set of IDS components with different functions as event generator, analysis engine, etc., and the communication protocol among them. Lee et al. (2000) give an implementation based on CIDF, which focuses on the communication mechanism of several IDSs. A DIDS architecture using autonomous agents to monitor security-related activity within a network is described in Barrus and Rowe (1998). The EMERALD system tries to deal with large-scale network environment

with a hierarchy of monitors, only focusing on the component structures and handling messages (Porras and Neumann, 1997). To detect DDoS attacks, a method based on calculating the global Kolmogorov complexity value on information pieces collected from different locations of a computer network, is employed in Kulkarni et al. (2001). Similar to the plan recognition method used in Geib and Goldman (2001), Huang and Wicks (1998) consider the importance of attack strategy analysis in distributed environment, but without an actual design to search for evidence needed in this analysis. In wireless *ad-hoc* networks, researchers have realized the need of sensors to monitor every network node. Zhang et al. (2003) install one individual IDS at each computer node and then use voting to determine the final suspicion score. A packet snooping method is developed to allow wireless node to monitor neighboring nodes and to vote to determine a node's attack state. There are plenty of other IDSs that either claim to handle distributed attacks or are distributed over a computer network. In summary, all existing systems concentrate on only two aspects in a narrow view of DIDS management: collection of information distributed in the network or information exchange format and structure.

A typical DIDS consists of a variety of intrusion detection facilities that are deployed at different network nodes and generate reports of alarms, warnings, and normal logs over time. These reports from individual IDS sensors are then fused at certain functional module normally centralized at a network location, in the hope of yielding a better composite score or a clearer picture of the entire network of concern. Various questions arise regarding the desired working procedures, such as: where should we put these IDS sensors and of what type? After we receive findings from several sensors that reside at different locations, can we fuse these alarms to get a better composite score? After we have the initial fusion outcome in this system, should we report this immediately or should we gain more information to confirm or reject the initial hypothesis? What can we do to improve the individual estimates and how? How to define the quality of the answer to each question so that this quality score can be used in high-level decision support systems? Relevant considerations hold true not only in detecting intrusions, but also in diagnosing the source and their root cause.

In distributed and complex environments, existing distributed intrusion detection techniques face several main problems to respond to the above questions: (1) one-time fusion. Current research focuses on fixed physical architecture (or static hierarchy) and specific algorithms. They are not process-oriented and lack the systematic strategy to manage the entire detection task. The outcomes from individual IDS components are at best fused once and then provided to other decision-making parties. This is natural since intrusion detection problem has originated primarily in the computer engineering discipline. This is the main reason behind the

discrepancy between many available algorithms/systems, and the performance required in practical deployment in terms of accuracy and efficiency. (2) Lack of reconfigurability. Existing DIDS structures are rigid in their ability to accommodate the required information fusion requirements. The system structure is fixed for a static computer system assuming no change over time. However, a DIDS functions in a dynamic environment and possesses dynamically changing physical systems. This DIDS should be designed and configured flexibly, and upgraded and reconfigured dynamically. With a reconfigurable system, improved and required DIDS can supposedly be introduced with considerably less expense and ramp-up time in response to the change and evolvement of the target environment and the decision-making requirements. (3) Lack of quality assurance. Existing systems can only generate certain quantitative classification scores without any quality measurements regarding confidence and certainty. Such score leads to distrustful decisions and cannot be used as feedback to direct the improvement and optimization of the knowledge fusion effort. The potential of dynamically seeking the best configuration of DIDS is not fully explored. On the other hand, determining such quality score is not even possible for these existing IDSs since they are not able to provide suitable quality measures in large-scale systems. They cannot deal with the complexity in such knowledge fusion, represented by the huge set of members and the uncertainty associated with these members.

4.2 An adaptive information fusion framework for distributed intrusion detection

We want to seek a paradigm shift from the rigid intrusion detection employing "one-time" fusion, to a process and system oriented adaptive fusion that is well-controlled with predictable quality assurance. The adaptive knowledge fusion in our research aims to exploit the synergistic power of multiple IDS components to improve the overall performance of the system as a whole. This holistic view of DIDS offers a uniquely different angle and course to understand the network attack correlation problem. An integrated framework is proposed to combine three organic components seamlessly in fulfilling the above target, i.e., a dependency DIDS model, a dynamic feedback control mechanism, and a quality assurance model based on information and utility theories.

As the focus of opening discussion on this important subject, the generic tasks in consideration include: (1) correlation of individual IDS outcomes to provide better intrusion scores and (2) quality assurance to decide when to engage and stop information integration, what information and where to collect from (e.g., the portion of network, group of sensory components, and type of information), and how good the correlation outcome is.

4.3 A dependency model for distributed intrusion detection systems

Dependency modeling and information theory are employed in the proposed solution. Dependency models use nodes to represent variables in a system and links to represent dependency relationships among these variables (Spirtes et al., 1993). BNs are probabilistic dependency models representing joint probabilities of a set of random variables and their acyclic conditional independence (Pearl, 1988). The nodes characterize the hypothesis, hidden state, and evidence variables in the physical system, while the arcs linking these nodes represent the causal dependency among these variables.

Dependency and probabilistic models are suitable techniques to describe the intercorrelation and uncertainty common in complex systems, regarding the structural features as well as the relationships among different constructs, including but not limited to network topology, correlation of information flow, functional type of the IDS sensor (e.g., the type of model and algorithm), and its quality and reliability. In Fig. 5, a hierarchical dependency structure and associated probability parameters show a conceptual illustration of such a dependency model for different components in DIDS. A DIDS system is a dynamic system with the information flow pattern changing over time constantly, and the control scheme changes the configuration in terms of IDS sensor placements and parameters in working cycles. The dependency model is updated constantly to reflect the latest situation.

(1) The links between two node-layers and their associated probabilities describe the interconnection among the physical nodes and the inter-correlation within their information flow in the information

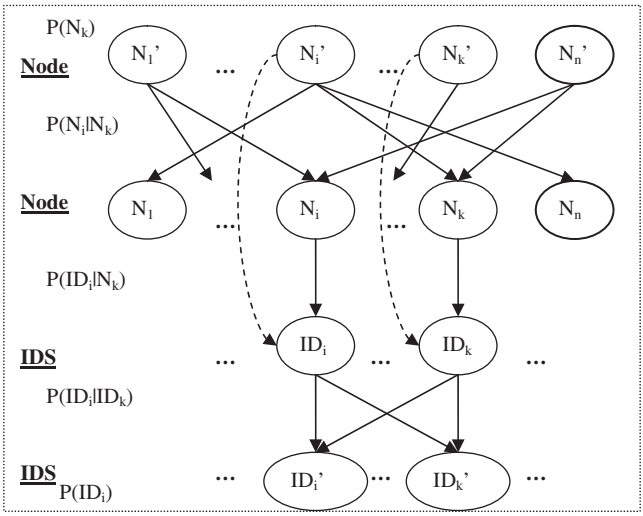
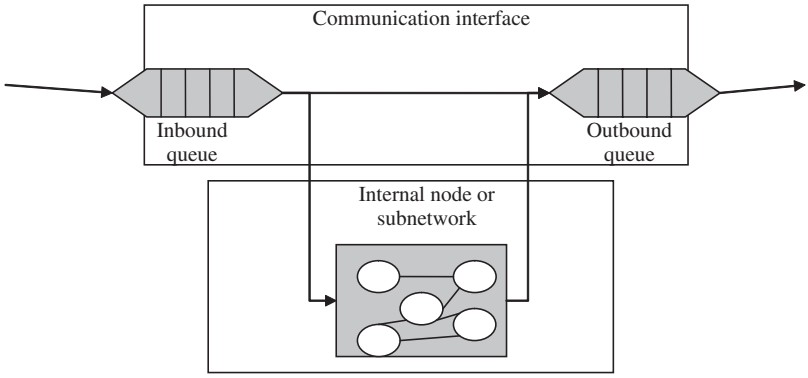
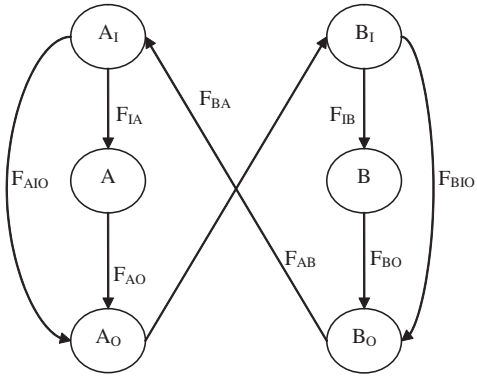


Fig. 5. The conceptual dependency model of a DIDS.

infrastructure. We focus on the interdependency resulting from the information flow among these nodes. Taking a simple situation as example, we can define the dependency of node B on A to be $I(AB)/I(A)$, where $I(A)$ and $I(AB)$ are the traffic throughputs through A and through B from A, respectively. The dynamic behavior in short-term along with the statistics over the long-term historical data is used in determining the structure and parameters. Thus a very important task is to model the information flow path and throughput in the target system. Figure 6 shows some simplified examples illustrating information traffic within and between network nodes. Obviously more research efforts are needed in building up accurate and high-fidelity dependency models, and representing data, components, services, and protocols such as different network layers in a TCP/IP network.



(a) Information traffic within a node



(b) A two-node computer network

Fig. 6. Modeling the information flow within and between computer network nodes.

(2) The links between the node layer and the IDS sensor layer, and the associated probability parameters describe the deployment of IDS components and their information collection capacity. The configuration of the IDS components regarding the information boundary is part of the configuration decision. For example, the sampling frequency of an IDS component can change, and a suspicious node may be worth a close look. We could also remove or add IDS components. Thus the dependency structure between these two layers can dynamically change.
(3) The links between the two IDS sensor layers and the associated probabilities describe the reliability and quality of IDS components due to the employed detection functions or other factors. Even if we put an IDS sensor at a computer node where an attack occurs, it may not always be able to detect this attack. For example, IDS sensors using the same algorithm together may be good at detecting certain type of attack while not good at other types.

In BN representation, a set of random variables, $V = \{V_1, \ldots V_M\}$, represent the nodes in the above model, and the prior and conditional probabilities represent the dependency among the states of these nodes. In the task of detecting computer attacks in a distributed network, the states of concern to us are the normal state and various attack states. In this proposal, we simplify the description by discussing the most basic hypothesis about these nodes, i.e., binary states being normal (or state '0') or intrusive (or state '1'). In the probabilistic dependency model, the knowledge fusion includes two inference tasks. These two tasks are the belief inference and the most probable explanation (MPE), given the observed reports (evidence $E = e$) from individual IDS components. The belief inference is to assess the posterior probability (belief) for the states of certain nodes (variables), i.e., $p(V_i = 0 \text{ or } 1) = p(V_i = 0 \text{ or } 1 \,|\, E = e)$. The MPE task is to find out an assignment of states for all the nodes that best explains the observed evidence (finding reports of individual IDS sensor). The assignment of states to the physical nodes represented by node Ns tells us the current picture of the systems, or in another word, the fused detection/diagnosis result. The beliefs are the scores characterizing confidence or certainty of hypothesis states.

Inference in the causal models, supported by causal relationships represented into its structure, can be used in confirmation and alleviation of intrusion suspicions represented by the beliefs of intrusive state at those network nodes. Thus, the evidences of alarms, warning, and other information from different intrusion detectors may strengthen or alleviate each other by propagating to those network nodes in the dependency structure, and generating a better picture about the underlying activities.

4.4 Configuration, planning, and scheduling

Decisions should be made dynamically in the overall working stages of deployment, detection, and diagnosis. We seek an active and dynamic working strategy based on belief update and utility calculation. The utility of each IDS component is defined and calculated and updated over time within this probabilistic and causal model, regarding the potential of clarifying a hypothesis and associated cost, reliability, etc. Thus we are able to compare alternative configurations, i.e., where to invest and to install IDS sensors, and which IDS components to engage and fine-tune. The planning and scheduling decision can extend to complicated aspects about the volume, level, and boundary of information to choose and collect, in terms of dimensions, locations, details, time granularity, etc.

We elaborate the control scheme by a very important decision, i.e., which IDS to turn on and when. First we define the benefit of each IDS sensor in terms of mutual information of this IDS sensor to the hypothesis variable, representing those physical network nodes. According to Shannon's entropy theory, mutual information measures the dependency between two random variables as:

$$I(N_j; ID_i) = \sum_{ki} \sum_{lj} p(ID_i = e_{ki}, N_j = h_{lj}) \log\left(\frac{p(ID_i = e_{ki}, N_j = h_{lj})}{p(ID_i = e_{ki})P(N_j = h_{lj})}\right) \tag{3}$$

where the ith IDS sensor provides as evidence the kth state e_{ki}, and the jth node has the lth state h_{lj}. The higher this mutual information score is, the better this IDS sensor. When there are several (s) network (hypothesis) nodes in consideration and several (n) IDS sensors to activate at one time, the above equation is extended to the following:

$$I(N; ID) = \frac{1}{s} \sum_{j} \sum_{k1} \ldots \sum_{kn} \sum_{lj} \left(p(ID_1 = e_{k1}, \ldots, ID_n = e_{kn}, N_j = h_{lj}) * \log\left(\frac{p(ID_1 = e_{k1}, \ldots, ID_n = e_{kn}, N_j = h_{lj})}{p(ID_1 = e_{k1}, \ldots, ID_n = e_{kn})P(N_j = h_{lj})}\right)\right) \tag{4}$$

The extended formula rewards a higher mutual information value to the IDS sensor that is tightly correlated to all hypothesis nodes. During the inference in this Bayesian model, the change of the positions of IDS sensor and the parameters in terms of the conditional probabilities between random variables will consequently change the mutual information values.

To calculate the utility index, we also have to define the cost that negates the benefit from activating this IDS sensor. The cost may include the cost of reconfiguration and information collection, the computation time for data processing, and the hardware execution time. Combining the mutual

information, $I(N;ID)$ and a cost term, $C(ID)$, we form the expected utility for a subset of IDS sensors, ID:

$$EU(ID) = \alpha I(N; ID) - (1 - \alpha)C(ID) \tag{5}$$

where α is the balance coefficient between the two terms. The optimal action is found by the following decision rule:

$$ID^* = \arg_{ID} \quad \max EU(ID) \tag{6}$$

Examining the utilities for all combinations of IDS sensors allows us to dynamically select a subset of IDS sensors of the highest utility in order to adaptively assess the underlying situation. The challenge here is the complexity due to the large number of nodes in the dependency model, representing a NP-complete problem of enumerating all combinations. Techniques that try to provide faster inference capability in dependency models and explore the special structure features of DIDS can be helpful in dealing with such complexity.

4.5 *Quality assurance and adaptive fusion*

So far we have answered the questions about how to fuse the distributed knowledge and information and how to make best decisions. The description until now lacks a very important part, i.e., when to start and stop this fusion process with feedback control. For this sake, we rely on a quality model to characterize the certainty or confidence about the composite fusion score.

As in the SPC technique used in quality control and improvement, basically we also need to continuously monitor certain quality scores in the fusion procedures. The challenge lies in the fact that the situation of the system at any time instance involves the states for a large number of nodes, associated with different belief values. We need to define quality measures over the states (beliefs) summarized across all these variables in this complex system, consisting of a large number of variables representing the system status.

The first requirement of this quality model is to characterize the *confidence* of the fusion result. A global quality *score* could be based on the complexity of current state beliefs. Confidence represents the certainty level of the fusion outcome. The less uncertainty and complexity, the more confidence do we have. The candidates include the information entropy by Shannon and the Kolmogorov complexity. Here we can use the information entropy form:

$$\begin{aligned} CF(N) &= 1 - H(N) = 1 - \frac{1}{s}\sum_j H(N_j) \\ &= 1 + \frac{1}{s}\sum_j \sum_{lj} (p(N_j = h_{lj}) \log p(N_j = h_{lj})) \end{aligned} \tag{7}$$

Table 3
Four different state combinations of the same confidence scores, for nodes N_1 and N_2

N_1	N_2	
	Normal (0)	Intrusive (1)
Normal (0)	00	01
Intrusive (1)	10	11

The value for this confidence score becomes smaller when these nodes have less uncertainty, i.e., one state of each variable has a high belief (probability) while other states of this node have very low beliefs. This small value indicates that we have a clear picture about the states that these nodes should have. In other words, we are more positive about the locations where attacks may exist.

Besides confidence, we also need to distinguish between different situations of consistency or complexity. Take as example two nodes N_1 and N_2. Four different situations are identified in Table 3, although all of them yield the same confidence score.

Moreover, we are interested in characterizing the dynamics of these states in terms of the change in their associated beliefs, not the stable states at one time instance. In other words, we want to capture the change point when the system undergoes external disturbance. The disturbance can be signaled by the change of complexity in the system. For example, the state transition of N_1 and N_2 from '00' to '11' demands more attention to increase the computation power or information collection. Thus we calculate the relative entropy between two time instances of fusion:

$$RE(T|t-1) = D(t|t-1) = \frac{1}{sT}\sum_{j}\sum_{lj} p^{t}(N_j = h_{lj})\log\left(\frac{p^{t}(N_j = h_{lj})}{p^{t-1}(N_j = h_{lj})}\right) \quad (8)$$

where p^t and p^{t-1} are the state beliefs at current and last time instances respectively, and T is the time interval. This relative entropy is nonnegative and equals zero only when there is no change in such beliefs between two time instances.

Based on these quality scores, we can set different thresholds to control the adaptive knowledge fusion process. For example, one threshold T_{H} can be the minimum confidence of fusion outcome, and another threshold T_{D} can be the minimum relative entropy between two consecutive knowledge fusion repetitions.

In this section we conceptually illustrate a methodology that integrates adaptive fusion with quality metrics in a well-controlled framework. However many important research issues are needed to be addressed. These topics can include building comprehensive and high-fidelity dependency

models, addressing inference degradation due to repeated IDS sensor selection, and developing efficient approximation algorithms dealing with complex network structures.

5 Intrusion detection and information assurance

The computing infrastructure is a part of critical national network infrastructures that consist of computer networks, wireless networks, sensor networks, power grid, supply chain/network, financial networks, disaster surveillance and response systems, transportation infrastructures, healthcare systems, social networks, and so on. These critical network infrastructures play essential roles in supporting the modern society and the global world economy. Great demand exists to strengthen the quality and reliability of these infrastructures, in terms of various performance metrics in availability, integrity, and confidentiality. With the development of new theories, latest computing technologies and ever-growing computational capacities, novel solutions to describe and model such networks, to understand the problems in terms of the above performance metrics, and to optimize their performance are constantly emerging in knowledge management, data mining, information fusion, evolutionary computing, and sensor techniques.

To protect the confidentiality, integrity, and availability of these computing infrastructures, three important tasks should be carried out cooperatively as shown in Fig. 7. Prevention deploys a set of security mechanism such as encryption, authorization, and authentication against various intrusive violations from happening. Intrusion detection is necessary since no prevention mechanism can defeat all types of attack. Intrusion detection needs auditing/logging services in order to fully function. Once a computer attack is detected, a sequence of reaction procedures should be applied to diagnose this attack, document the trace, and recover the computer infrastructure to normal status. The vulnerability exploited by the attack should

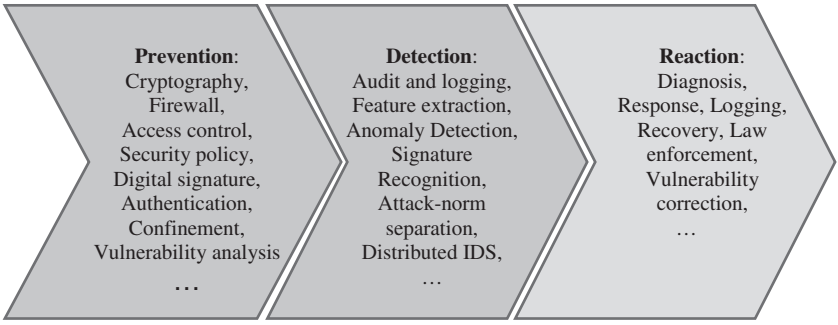


Fig. 7. Major tasks in information assurance.

be immediately fixed to fortify the system against this type of attack in the future. Law enforcement and legal action could also be considered if necessary and possible.

An efficient and accurate IDS is vital in the above information assurance agenda that fulfills the threefold mission of prevention, detection, and response. An IDS can provide decision support to the prevention for mechanisms such as admission control and authentication, by monitoring information throughput and analyzing audit and log data in the protected computer system. Some security technologies can help both intrusion prevention and intrusion detection tasks. For example, a "honeypot" decoy can be used for early warning for new attacks in intrusion detection and vulnerability analysis for intrusion prevention.

Collaboration between intrusion detection and intrusion response is already shown in the example of an artificial immune system given in Section 2 about the application of hybrid IDSs. Such an artificial immune system requires the capability to rapidly identify new attacks and extract attack signatures for dissemination. An IDS applying anomaly detection models can efficiently identify the abnormal events for further study.

More importantly, intrusion detection can greatly support intrusion response task forces with automatic and efficient diagnosis at the first time of raising an alarm. The diagnosis can yield auxiliary information including root causes and attack paths for the detected intrusion. Network intrusion diagnosis study is in its infancy with a few studies on IP tracing such as in Goodrich (2002), which adds special tags to network packets in order to find out the source and the passing path for the message. The other important task is called "attack traceback" that tries to find out the master computer from which the attack is launched. The literature for this type of diagnosis is much less but it could use the IP tracing as the starting point (Wang and Schulzrinne, 2004).

The intrusion diagnosis function, e.g., can be efficiently supported by the information fusion framework for DIDS in Section 4. Knowledge about root causes and attack paths can be identified during the adaptive information fusion process. In a diagnosis mode, this system can highlight the network nodes that have significantly high intrusive state beliefs. By monitoring and plotting the highlighted nodes over time in a network topology graph, potential spreading patterns of attacks can be revealed. Consequently this system can also predict potential epidemic paths of computer attacks in the future by analyzing the connection patterns of the network graph.

6 Conclusions and discussion

Intrusion detection systems use the data collected by auditing activities in computer hosts and monitoring traffic in networks. Intrusion detection recognizes the normal and intrusive patterns in such data. Because of the

complex operational environment and the requirement on their functionality, intrusion detection faces great challenges summarized as follows: huge amounts of complex data with constantly changing pattern and unclear distribution, mixed and unstructured variable types in data, achieving high detection rate while maintaining the false alarm rate at a very low level, and the complication of huge and distributed computer systems in the typical implementation environment. In this chapter the assumptions underlying these challenges and their demands on intrusion detection are articulated.

Both computer audit data and network traffic data are given in examples and discussed. Raw data collected for intrusion detection in general are time series data with noise and uncertainty. Within the procedural view of knowledge discovery and data mining, important features are extracted after data preparation and cleaning. The general statistical (temporal transform) and domain knowledge-based (computer engineering expertise) feature extraction methods are commonly used.

Existing IDSs employ algorithms and models from various fields of statistics and probability, artificial intelligence, machine learning, and biology. These algorithms fall into the classic categories of anomaly detection and misuse detection. Classic stochastic process control is used to illustrate the mechanism of anomaly detection. A clustering and classification technique demonstrates the stages of pattern capturing and recognition in misuse detection. The usage and assumptions of different algorithms are discussed with associated advantage and disadvantage in performance, cost, and applicability. Hybrid structures to integrate anomaly detection and misuse detection represent a significant means to reduce the false alarms while improving the detection accuracy. Latest advances such as the research based on signal-processing techniques are introduced. Future directions in improving individual IDSs according to the view of the author include emerging evolutionary computing and self-regenerative systems within the vision of the next generation of computing, which are robust and capable of adaptive adjustment to handle complex changes.

Distributed IDSs are especially important in wireless and *ad-hoc* architectures. Heterogeneous computer and networking resources raise issues affecting deployment and coordination of a large set of intrusion detection facilities. To optimize the performance of such systems under the constraints of resources and time, coordinating and planning strategies become the key issue to gain benefits from multiple detection facilities in addition to strengthening individual IDSs, like any other sensor networks. A few works toward such distributed intrusion detection networks are discussed. This section examines the generic tasks in distributed IDSs including deployment, modeling, scheduling, and planning of a set of intrusion detection sensors. An integrated modeling architecture is described based on probabilistic dependency network models. A quality model combining the above modeling architecture and the information and utility theory can be used to define quality measures and to direct the decision-making in implementation.

Some miscellaneous issues play a nontrivial role with the implementation of IDS especially in practice. Among them there are legal and privacy that raise more and more concerns in the public. The security of IDSs themselves is also an issue worthy of notice particularly in distributed environments. Another issue important to the success of IDSs is the human factor aspect in such systems where findings in human computer interaction and cognitive science can greatly improve the performance.

7 Questions for discussions

Exercise 1. As an example several data attributes in Solaris operating systems are given in the chapter that can be used as input to an intrusion detection system (IDS). Can you find out some similar data attributes existing in Windows and Macintosh operating systems for the same purpose? Do such useful data only come from the logging facility of a computer operating system?

Exercise 2. To build up a mathematical model for solving a physical problem such as in an IDS, assumption about the physical phenomenon plays a key role to rationalizing the underlying model and guiding the system design. After reading the introduction on misuse detection and anomaly detection techniques, discuss the assumption behind each of these two types of intrusion detection model and their associated advantages and disadvantages.

Exercise 3. This chapter discusses the dilemma facing IDSs due to dominance of normal behaviors in computer systems, using a set of numbers about the normal activity, the detection rate, and the false alarm rate. In this exercise, for the same example please use the Bayes theorem to induce the confidence as given in the chapter for an alarm raised by an IDS to indicate a real attack.

Exercise 4. We have discussed the classical categories of IDSs, i.e., anomaly detection or misuse detection, and the various hybrid IDSs that combine different categories detection engines in one system. Are there other different structures to integrate individual intrusion detection engines than those discussed in the chapter? What are the advantages and disadvantages for these different system structures?

Exercise 5. In this exercise, let us do a practice about formulating an intrusion detection model. Instance-based classification, a classical learning model, finds the class/category that a new data point belongs to by looking at similar instances (near to the new data point) collected in the history. Read about the instance-based classification by yourself if you are not very familiar with it. Your task here is to design an anomaly detection system and also a misuse detection system applying this instance-based classification model. Specifically answer the following questions for anomaly detection and misuse detection, respectively.

(1) How do you generate the required patterns/profiles in training?
(2) How do you classify a new data point/event to be normal or intrusive in detection?

Exercise 6. As discussed in the chapter, distributed IDSs represent a solution to the increasing system complexity of network security. Several different classifications of existing IDSs are discussed in the chapter. Can you think of several ways to categorize distributed IDSs? Here you should be able to identify the different criteria in categorization.

References

Axelsson, S. (2000). Intrusion detection systems: a survey and taxonomy. Report, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden.

Barrus, J., N.C. Rowe (1998). A distributed autonomous-agent network-intrusion detection and response system. *1998 Command and Control Research and Technology Symposium*, Montery, CA, June–July.

Bishop, M. (2005). *Introduction to Computer Security*. Addison Wesley, Boston, MA.

Box, G., A. Luceno (1997). *Statistical Control by Monitoring and Feedback Adjustment*. Wiley, New York NY.

Debar, H., M. Becker, D. Siboni (1992). A neural network component for an intrusion detection system, in: *Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy*, *May*, Oakland, CA, pp. 240–250.

Debar, H., M. Dacier, A. Wespi (1999). Towards a taxonomy of intrusion-detection systems. *Computer Networks* 31, 805–822.

Denning, D. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering* 13(2), 222–232.

Denning, D. (2004). *Information Warfare and Security*. Addison Wesley, Boston, MA.

Ertoz, L., E. Eilertson, A. Lazarevic, P. Tan, J. Srivastava, V. Kumar, P. Dokas (2004). The MINDS—Minnesota Intrusion Detection System, in: H. Kargupta, A. Joshi, K. Sivakumar, Y. Yesha (eds.), Data Mining: Next Generation Challenges and Future Directions, MIT/AAAI Press, Menlo Park, CA.

Forrest, S., S.A. Hofmeyr, A. Somayaji (1997). Computer immunology. *Communications of the ACM* 40(10), 88–96.

Geib, C., R. Goldman (2001). Plan recognition in intrusion detection systems, in: *DARPA Information Survivability Conference and Exposition (DISCEX)*, June, Anaheim, CA.

Goodrich, M.T. (2002). Efficient packet marking for large-scale IP traceback, in: *Ninth ACM Conference on Computer and Communications Security (CCS)*, Washington, DC, pp. 117–126.

Huang, M.-Y., T.M. Wicks (1998). A large-scale distributed intrusion detection framework based on attack strategy analysis, in: *Web Proceedings of the First International Workshop on Recent Advances in Intrusion Detection (RAID'98)*.

Ilgun, K., R.A. Kemmerer, P.A. Porras (1995). State transition analysis: a rule-based intrusion detection approach. *IEEE Transactions on Software Engineering* 21(3), 181–199.

Kephart, J.O. (1994). A biologically inspired immune system for computers. Artificial Life IV, in: *Proceedings of the Fourth International Workshop on Synthesis and Simulation of Living Systems*, MIT Press, Cambridge, MA, pp. 130–139.

Kephart, J.O., G.B. Sorkin, M. Swimmer, S.R. White (1997). Blueprint for a computer immune system, in: *Virus Bulletin International Conference*, October 1–3, San Francisco, CA.

Ko, C., G. Fink, K. Levitt (1997). Execution monitoring of security-critical programs in distributed systems: a specification-based approach, in: *Proceedings of IEEE Symposium on Security and Privacy*, Oakland, CA, pp. 134–144.

Kulkarni, A.B., S.F. Bush, S.C. Evans (2001). Detecting distributed denial of service attacks using Kolmogorov complexity metrics. Report 2001CRD176, GE R&D Center, December.

Lee, W., R. Nimbalkar, K. Yee, S. Patil, P. Desai, T. Tran, S. Stolfo (2000). A data mining and CIDF based approach for detecting novel and distributed intrusions, in: *Proceedings of The Third International Workshop on Recent Advances in Intrusion Detection (RAID 2000), October*, Toulouse, France.

Lee, W., S.J. Stolfo, K. Mok (1999). A data mining framework for building intrusion detection models, in: *Proceedings of 1999 IEEE Symposium on Security and Privacy*, Oakland, CA, pp. 120–132.

Li, X., N. Ye (2002). Grid- and dummy-cluster-based learning of normal and intrusive clusters for computer intrusion detection. *Journal of Quality and Reliability Engineering International* 18(3), 231–242.

Li, X., N. Ye (2003). Decision tree classifiers for computer intrusion detection. *Journal of Parallel and Distributed Computing Practices* 4(2), 77–93.

Li, X., N. Ye (2006). A supervised clustering and classification algorithm for mining data with mixed variables. *IEEE Transactions on Systems, Man, and Cybernetics-Part A* 36(2), 396–406.

Mirkovic, J., P. Reiher (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communications Review* 34(2), 39–53.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo CA.

Porras, P.A., P.G. Neumann (1997). EMERALD: event monitoring enabling responses to anomalous live disturbances, in: *Proceedings of the Nineteenth National Computer Security Conference*, 22–25 October, Baltimore, MA, pp. 353–365. NIST/NCSC.

Spirtes, P., C. Glymour, R.R. Schienes (1993). *Causation, Prediction, and Search*. Springer-Verlag, New York.

Staniford-Chen, S., B. Tung, D. Schnackenberg (1998). The Common Intrusion Detection Framework (CIDF). *Information Survivability Workshop*, *October*, Orlando, FL.

Valdes, A., K. Skinner (2000). Adaptive, model-based monitoring for cyber attack detection, *Recent Advances in Intrusion Detection (RAID 2000)*, Toulouse, France.

Wang, B.T., H. Schulzrinne (2004). A denial-of-service-resistant IP traceback approach, *IEEE Symposium on Computers and Communications (ISCC)*, Alexandria, Egypt, June.

Ye, N., T. Farley (2005). Attack-norm separation—a new approach to cyber attack detection. *IEEE Computer* 38(11), 71–77.

Ye, N., X. Li, S.M. Emra (2000). Decision tree for signature recognition and state classification, in: *Information Assurance and Security Workshop of IEEE Systems, Man, and Cybernetics,* West Point, New York, pp. 189–194.

Ye, N., X. Li, Q. Chen, S.M. Emran, M. Xu (2001). Probabilistic techniques for intrusion detection based on computer audit data. *IEEE Transactions on Systems, Man, and Cybernetics—Part A* 31(4), 266–274.

Zhang, Y., W. Lee, Y. Huang (2003). Intrusion detection techniques for mobile wireless networks. *ACM/Kluwer Wireless Networks Journal (ACM WINET)* 9(5), 545–556.

Chen, Raghu, Ramesh, Vinze & Zeng, Eds., *Handbooks in Information Systems, Vol. 2*

Chapter 17

Anomaly-Based Security Framework for Network Centric Systems

Guangzhi Qu and Salim Hariri

Electrical and Computer Engineering Department, The University of Arizona, 1230 E Speedway Blvd., Tucson, AZ 85721, USA

Abstract

With the rapid growth of Internet systems and applications, the vulnerability of Internet is increasingly threatening national economics and security. Most of existing methodologies for analyzing and preventing network attacks rely on an offline analysis, and lack a proactive self-protection capability. Moreover, most of them focus only on known attacks and do address the impact of network attacks on the overall network operations and services.

The goal of our research is to develop a theoretical framework and a general methodology for anomaly analysis and protection against network attacks to achieve online monitoring and analysis of network attacks, and automatically identifying critical vulnerable resources whose failures can severely impact the overall operations of the network and its services.

1 Introduction

The Internet has grown exponentially over the last few years and has expanded commensurately in both scope and variety. In addition to the increasing number and dependence on Internet resources and services, there are also increasing interconnectedness and interdependence among large and complex systems; a failure in one sector can easily affect other sectors. Disruption of critical Internet services can be very expensive to businesses, life threatening for emergency services, and ultimately threaten the defense and economic security nationwide. In fact, the Commission on Critical Infrastructure reports that the potential for disaster in the U.S. as a result of network attacks is catastrophic (Report to the President's Commission on Critical Infrastructure Protection, 1997). This raises an important question

given the fragility of Internet infrastructures to small random failures (such as hardware failures, bad software design, innocent human errors, and environmental events). According to a recent study, Internet is robust to random failures; however, the Internet's reliance on a few key nodes makes it vulnerable to an organized attack and it has been shown that the average performance of the Internet would be reduced by a factor of 2 if only 1% of the most connected nodes are disabled. If 4% of them were shut down the network would become fragmented and unusable (National Research Council, 2003).

The existing techniques to respond to these attacks such as intrusion detection systems (IDSs) and firewall hardware/software systems are not capable of handling these complex interacting organized network attacks. The development of countermeasures (e.g., signatures for IDSs) is manually intensive activity and cannot detect future unknown attacks. Furthermore, the users are required to manually install/update the signature databases in order to protect themselves against the existing or new attacks.

In this chapter, we introduce advanced methodologies and effective technologies that can effectively detect network attacks in real time and configure the network and system resources and services to proactively recover from network attacks and prevent their propagations. However, automatic modeling and online analysis of the anomaly of the Internet infrastructure and services is a challenging research problem due to the *continuous changes in network topology*, the *variety of services and software modules* being offered and deployed, and the *extreme complexity of the asynchronous behaviors of attacks*. In general, the Internet infrastructure can be considered as a structure varying system operating under a complex environment with a variety of unknown attacks. In order to develop an effective system for the attack detection, identification, and protection, it becomes highly essential for the system to have the functionality of online monitoring, adaptively modeling and analysis tailored for real-time processing, and proactive self-protection mechanisms.

Anomaly analysis involves detailed analysis of network faults and/or attacks, identifying accurate network metrics, and using these metrics for quantifying the impact of network faults and/or attacks on various network traffic and network system components. Here the network system components consist of logical components and physical components. Logical components include network protocols and network services. Physical components could be computers (clients, servers) and network devices (routers, switches). Vulnerability analysis identifies system holes that can be exploited by the attackers to launch attacks. A lot of work has been done in studying various network faults and attacks on the network services and components; however, very little has been done to quantify the impact of network faults and attacks and use this information to provide proactive protection from network attacks. As the uses of the Internet expand, the stakes and thus the visibility of vulnerability and fault tolerance concerns

will rise (Albert et al., 2000; Hou et al., 1998). This identifies the need for robust self-protection architectures that integrate vulnerability analysis and network survivability.

The capabilities of the Internet protocol such as type of service (ToS) and quality of service (QoS) facilities are utilized for priority-based and policy-based routing. However, due to the increasing vulnerability, it is necessary to prioritize traffic based on security and vulnerability in addition to the traditional classification of applications based on their QoS. Furthermore, there are no standard protocols that can prohibit an attacker from consuming all available network bandwidth and resources. Due to lack of such protocols, packet-based attacks are not only possible, but they have also become quite common. The vulnerability of the Internet could be primarily accounted to TCP/IP, because security was not a main design consideration when it was initially designed. Traffic engineering principles and real-time network measurement are the two fields that show promising application in analyzing network infrastructure vulnerability and provide measures to improve network survivability under various attack and fault scenarios (Phillips and Swiler, 1998; Liu et al., 2000).

This rest of this chapter is organized as follows. In Section 2, we review network security and network attacks, and current network security techniques with a focus on IDSs.

In Section 3, we introduce our anomaly analysis methodology and framework. Anomaly distance (AD) is developed to quantify the operational states of network components in terms of measurement attributes (MAs). Based on the anomaly distance metric multivariate statistical analysis and information theory-based anomaly analysis are proposed.

In Section 4, based on the novel decision dependent correlation metric optimal subset of features can be effectively chosen from large amount of features to increase the detection rate and reduce the false alarm.

In Section 5, we conclude the works in this chapter and give future research directions.

2 Background and related works

2.1 Classification of network attacks

Network attacks can be grouped into five major categories: DoS, user to root (U2R), remote to local (R2L), probe, and worm/virus attacks (Kendall, 1998).

2.1.1 Denial of service attacks

The different types of denial of service attacks can be broadly classified into *software exploits* and *flooding* attacks. In software exploit-based attacks, the attacker sends a few malformed packets to exercise specific

known software bugs within the target's OS or application in order to disable the victim. Examples of attacks in this category include Ping of Death (Insecure Inc., 1996), Teardrop (CERT, 1997), Land (CERT, 1997), and ICMP Nukes. In flooding attacks, one or more attackers are sending continuous streams of packets aimed at overwhelming network bandwidth or computing resources at the victim. There are numerous DoS attacks in the flooding attacks category such as TCP SYN attack, Smurf IP attack, UDP flood attack, and ICMP flood attacks (Webopedia, 2004).

2.1.2 Distributed denial of service attacks

Flooding attacks can be classified based on the location of the observation point as single source and multiple sources attack (also known as distributed denial of service—DDoS). There are two common scenarios for DDoS attacks: *DDoS* attack (Dittrich, 2007) and the *distributed reflector denial of service* (DRDoS) attack (Paxson, 2001; Gibson, 2002).

2.1.3 U2R and R2L attacks

U2R and R2L attacks exploit bugs or misconfiguration in the operating system to control the target system. For example, a buffer overflow or incorrectly setting file permissions in a set user ID (SUID) script or program can often be deployed by this kind of attacks.

2.1.4 Probe attacks

Probe attacks involve testing a potential target resource to gather information. These are usually harmless (and common) unless vulnerability is discovered and later exploited.

2.1.5 Worm/virus attacks

Modern virus and worm attacks inherit U2R and R2L characteristics to deploy vulnerability in the vulnerable hosts. After being infected, the resources will degrade the overall network performance and their services as experienced in the CodeRed (CERT, 2001a), Nimda (CERT, 2001b), SQL Slammer (CERT, 2003a), RPC DCOM (Microsoft, 2003), W32/Blaster (CERT, 2003b), SoBig (CERT, 2003c), and other typical worm/virus attacks (Gaudin, 2003). Their self-replicating and propagation characteristics make the security of information infrastructure a challenging research problem.

2.2 Existing network security techniques

As the number of network attacks increases significantly (Moore et al., 2001), many router-based defense mechanisms have been proposed, including ingress filtering (Ferguson and Senie, 1998; Killalea, 2000), egress filtering and route filtering (Killalea, 2000), router throttling (Yau et al., 2002),

Pushback (Ioannidis and Bellovin, 2002; Manajan et al., 2001), Traceback (Bellovin, 2000; Savage et al., 2000; Snoren et al., 2001; Song and Perrig, 2001; Stone, 2000), and various intrusion detection mechanisms (Crosbie et al., 1996; Ilgun et al., 1995; Smaha and Winslow, 1994; Tjaden et al., 2000; Botha et al., 2002; Anderson et al., 1995; Hochberg et al., 1993). Besides the reactive techniques discussed above, some systems take proactive measures against DoS attacks. For example, distributed packet filtering (Park and Lee, 2001) blocks spoofed packets using local routing information and SOS (Keromytis et al., 2002) uses overlay techniques with selective re-routing to prevent large flooding attacks.

2.2.1 Modern intrusion detection systems

IDS is a critical component of network security in addition to user authentication, authorization, encryption, and the techniques discussed in this section. Furthermore, IDS is an important component of the defense-in-depth or layered network security mechanisms. It aims at detecting early signs of attack attempts so that the proper response can be evoked to mitigate the impact on the overall network behavior. An IDS collects system and network activity data (e.g., TCP dump data and system logs) and analyzes the information to determine whether there is an attack occurring with least false alarms and false negative alarms.

IDS have been an active area of research for quite some time. Kemmerer and Vigna (2002) give a brief overview of IDS techniques since the original IDS that was proposed by Denning (1987). Typically, based on the general detection strategy, IDSs can be classified into two categories—signature-based detection and anomaly-based detection.

Signature-based detection system finds intrusion by looking for activities corresponding to previously known intrusions. It is the most widely used commercial IDS model that is designed to detect and defend systems from malicious and intrusive events depending on ''signatures'' or ''thumbprints'' developed by human experts or by semi-automated means from prior known intrusion experiences. It is deployed in most of the current commercial security products/approaches. IDIOT (Crosbie et al., 1996), STAT (Ilgun et al., 1995), Sophos anti-virus application, and Microsoft patch remote update system are typical examples of the misuse detection system. A disadvantage of this approach is that signature-based IDS can only detect intrusions that follow pre-defined patterns. This approach becomes ineffective if the attack follows different patterns from those registered in the knowledge repository.

The second model is the anomaly-based IDSs, such as IDES (Lunt et al., 1992), Stalker (Smaha and Winslow, 1994), Haystack (Smaha and Winslow, 1994), and INBOUNDS (Tjaden et al., 2000). This approach uses a reference model of normal behavior and flags any significant deviation from normal behaviors. The normal behavior with respect to one metric (response time, buffer utilization, etc.) is defined based on rigorous

analysis of the selected metrics under normal and abnormal conditions. This approach can offer unparalleled effectiveness and protection against unknown or novel attacks since no a priori knowledge about specific intrusions is required. For example, the normal behavior of a web browsing could be profiled in terms of average frequency of requests, average connection duration, and so on. If these MAs change significantly, an anomaly alarm will be raised. However, it may cause more false-positives than signature-based IDS because anomaly can be caused due to a new normal network behavior. The limitation of this approach is because of the difficulty in performing online monitoring and adaptive analysis of newly network data. This limitation can be addressed to some extent by using a hybrid approach (Botha et al., 2002). IDES, NIDES (Anderson et al., 1995), and NADIR (Hochberg et al., 1993) use both statistical methods and built-in signatures to detect network attacks.

3 Anomaly analysis system

In this section, we present a theoretical framework for the construction of global metrics that can be used to analyze, manage, and control the operation of network systems under faults and/or attacks. The *l* conceptual model of our approach is illustrated in Fig. 1. Any network or system's behavior can be characterized by using a set of anomaly functions (e.g., AD). AD quantifies the behavior of individual network components and network services as being normal, uncertain, or abnormal (see Fig. 1). In fact, the concept of using index functions to quantify and characterize the behavior of networked systems and services is analogous to biological metrics such as temperature or blood pressure to determine the state of an organism (Parashar, 1994). The AD value will be used as the mathematical basis to proactively respond to network attacks and/or faults in real-time mode. In our approach, any application and resource (network or pervasive system) is assumed to be in one of the following three states: normal, uncertain, and abnormal states.

Our approach is based on real-time monitoring to perform online anomaly analysis of network attacks and then deploy the appropriate proactive self-protection mechanisms. Our framework shown in Fig. 2 consists of three main activities: (1) online monitoring and analysis of network anomaly; (2) identifying critical network components/services (Hariri et al., 2003); and (3) proactive self-protection from network attacks and their propagations. The data mining and statistical engine is designed to automatically select effective MAs for profile modeling of operational states and attacks. The anomaly analysis engine calculates the anomaly distance $AD_{MA}(t)$ for some MA from the normal value (MA_{normal}). When $AD_{MA}(t)$ deviates from the threshold that determines the normal operational region of the overall network or for a network device, an event will be generated to

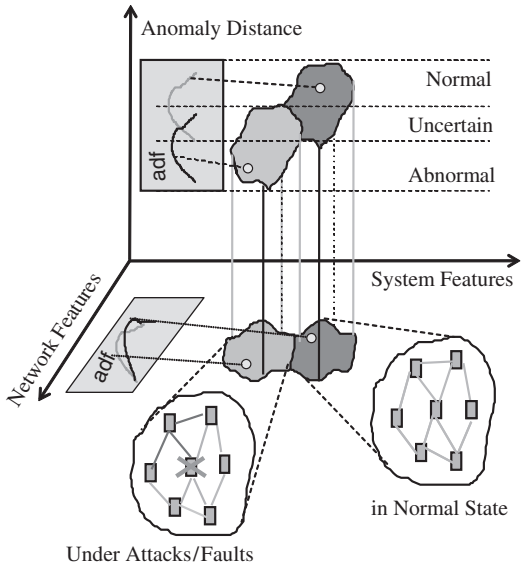


Fig. 1. Anomaly analysis conceptual model.

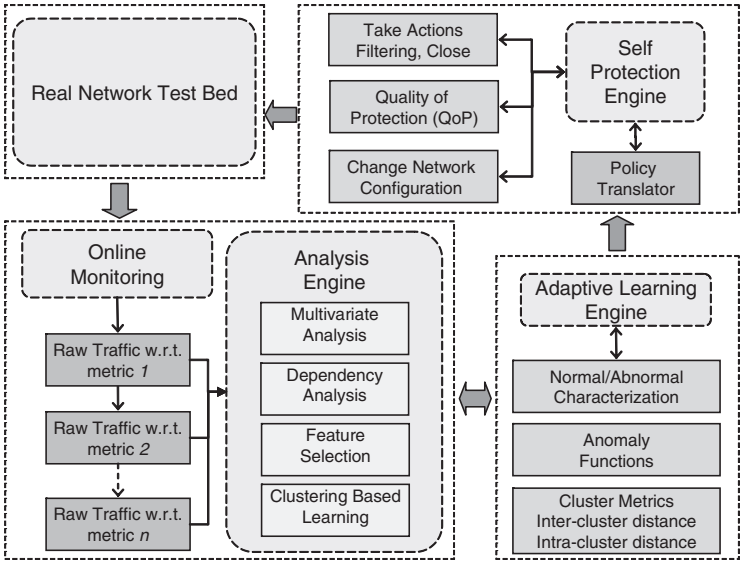


Fig. 2. Anomaly analysis framework.

inform the self-protection module. The self-protection engine will then take the appropriate actions such as shutting down node interface, closing network service ports, dropping network packets, or even shutting down the node itself to prevent the propagation of the attacks.

3.1 Online monitoring

The main goal of the online monitoring and analysis is to compute anomaly metrics that characterize and quantify the operational state of actual network systems at runtime. Hence, it is very important to identify effectively the appropriate MAs that can be used to reflect the operational state of the network components.

3.1.1 Measurement attributes

An MA denotes the value of some attributes that can be measured online such as the rate of outgoing TCP SYN packets of a network node ($TCP_{syn,out}$), the total number of outgoing UDP packets for the network system (UDP_{out}), and the CPU utilization for a computer (either client or server) (CPU_{util}).

We have identified different MAs for all network and system components that can be used to characterize the operational states of all software and hardware layers starting from the applications level (FTP, Telnet, web surfing, e-mails, etc.) down to the physical device level (CPU, memory).

The online monitoring engine is capable of monitoring the multiple MAs for a single node or the whole network system. The online monitoring module includes two basic functional modules—*host monitor*, which sits on network node (e.g., PCs with windows OS, Linux, etc.) that can report host resource usage statistics, for example, CPU utilization, memory usage; *network monitor*, which can report connection-based information in network-wide range. This flow related information (e.g., packet rate, packet size) can be indexed by source/destination IP address, source/destination port number, and protocol type.

3.2 Anomaly analysis techniques

3.2.1 Anomaly distance function

The anomaly distance function is used to quantify the operational states of network services and network itself. The AD_{MA} can be calculated as in Eq. (1). $D(\cdot,\cdot)$ is a function to compute the distance of two variables. This distance function could be a Euclidean (Batchelor, 1978), Manhanttan (Batchelor, 1978), Mahalanobis (Nadler and Smith, 1993), Camberra, Chebychev, quadratic, correlation, and χ^2 distance metrics (Michalski et al., 1981; Diday, 1974). The $MA(t)$ represents the current value of some MAs at time t. MA_{normal} denotes nominal value of MA when the network node operates normally.

$$AD_{MA}(t) = D(MA(t),\ MA_{normal}) \tag{1}$$

Anomaly distance function (ADF) is used to quantify the component/resource operational states (e.g., *normal*, *uncertain*, and *abnormal*) with

respect to one or more MAs (AD_{MA}) as shown in Fig. 3. The ADF_{MA} can be calculated as the ratio of the current distance from normal level with respect to one MA divided by the normal value for the MA as shown in Eq. (2).

$$ADF(MA, t) = \begin{cases} 1, & \text{if } AD_{MA}(t) \geq \Delta_{MA} \\ \frac{AD_{MA}(t)}{\Delta_{MA}}, & \text{Othewise} \end{cases} \quad (2)$$

$AD_{MA}(t)$ is the online calculated operational index according to Eq. (1) and the Δ_{MA} denotes the minimal distance from a normal operational state to the abnormal operational with respect to an MA. Figure 4 shows that when the network operates in normal state, the value of ADF is ~0.1. When the node endures an attack, the ADF value approaches 1.

3.2.2 *Multivariate anomaly analysis*

To improve the detection accuracy, we use multivariate analysis because single network attribute may not accurately capture an abnormal behavior.

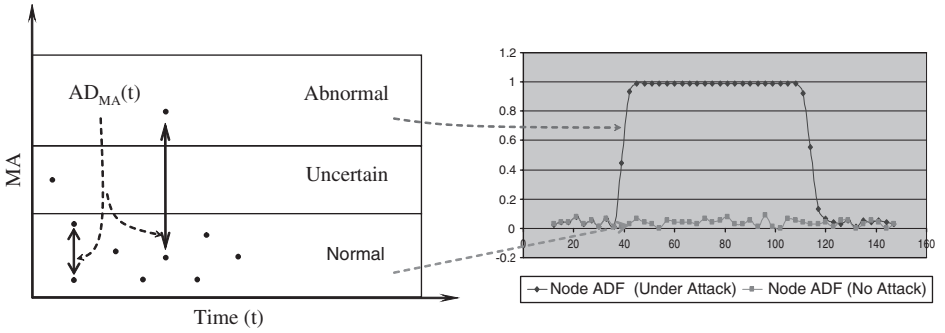


Fig. 3. Anomaly distance with respect to measurement attribute.

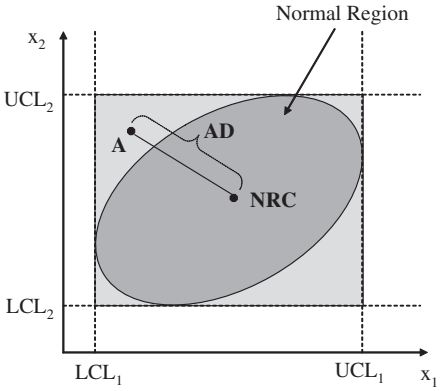


Fig. 4. Multivariate analysis model.

For example, in a two-dimensional attributes space as shown in Fig. 4, where UCL_i and LCL_i ($i = 1$ or 2) are the upper and lower thresholds for attributes x_1 and x_2, respectively, assume that the state of a node or application is represented by point A of Fig. 4, then for any single attribute x_1 or x_2 the node/application operates normally. But, when we combine the two metrics, the state A is outside the shaded oval area (that designates the normal state with respect to the two metrics) and thus it is abnormal.

The multivariate analysis can be applied to study the behavior of network applications and the correlation between the measured attributes to determine whether or not they are operating normally. Ye et al. have used Hotelling's T^2 and χ^2 multivariate analysis methods on the system logs in offline host intrusion detection (Ye et al., 2000; Ye and Chen, 2001). Our approach for online multivariate analysis focuses on network infrastructure related information and builds a new metric to measure the abnormal behavior.

Hotelling's T^2 control chart is applied in the implementation (Qu et al., 2005a). Hotelling's T^2 control chart (Montgomery, 2001) has been used to perform anomaly detection and root cause analysis in manufacturing systems (Jin and Shi, 2001) and it has the capability of generating a normal region from the normal profile of multiple MAs. First, we need to determine the baseline profile. According to the Hotelling's T^2 method, we need to have M ($M > 20$) observations. Suppose we have P MAs. Then the normal behavior of the MA can be represented as:

$$\mathrm{MA} = \left\{ \begin{pmatrix} \mathrm{MA}_1(t) \\ \mathrm{MA}_2(t) \\ \vdots \\ \mathrm{MA}_p(t) \end{pmatrix} \begin{pmatrix} \mathrm{MA}_1(t+1) \\ \mathrm{MA}_2(t+1) \\ \vdots \\ \mathrm{MA}_p(t+1) \end{pmatrix} \cdots \begin{pmatrix} \mathrm{MA}_1(t+\mathrm{M}) \\ \mathrm{MA}_2(t+\mathrm{M}) \\ \vdots \\ \mathrm{MA}_p(t+\mathrm{M}) \end{pmatrix} \right\} \tag{3}$$

Based on these normal MAs, two control limits—upper control limit (UCL) and lower control limit (LCL)—can be determined using the M preliminary blocks to obtain in-control data to determine the mean $\overline{\mathrm{MA}}$ and covariance matrix S. From these values, we can then determine a normal region with respect to the P MAs. The sample mean $\overline{\mathrm{MA}}$ determines the normal region center (NRC) and the sample covariance matrix S determines the shape of the normal region as shown in the shaded part of Fig. 4. For a pre-defined Type-I error$^{\alpha}$, the UCLs and LCLs are computed:

$$\mathrm{UCL} = \frac{(M-1)^2}{M} B_{1-\alpha/2}\left[P, \frac{M-P-1}{2}\right] \tag{4}$$

$$\mathrm{LCL} = \frac{(M-1)^2}{M} B_{\alpha/2}\left[P, \frac{M-P-1}{2}\right] \tag{5}$$

where $B_{\alpha/2}[P,(M-P-1)/2]$ is the $1-\alpha/2$ percentile of the β distribution with P and $(M-P-1)/2$ denotes the degrees of freedom.

The AD metric is used to quantify how far the current operational state of a component is from the normal state for a given attack scenario based on one or more MAs. Our initial research results show that the mean shifts of the well-selected network attributes can be effectively used for attack detection. The normalized anomaly distance AD_k with respect to an attribute MA_k is defined as

$$\mathrm{AD}_k = \left[\frac{\mathrm{MA}_k(t) - \mu_{\mathrm{MA}_k}}{\sigma_{\mathrm{MA}_k}}\right]^2 \tag{6}$$

where μ_{MA_k} and $\sigma^2_{\mathrm{MA}_k}$ are the mean and variance under the normal operation condition corresponding to the measurement attribute k. $\mathrm{MA}_k(t)$ is the current value of a network attribute k.

A general definition of the AD metric with respect to a subgroup J of multiple correlated attributes can be defined using the statistic T^2 distance as:

$$\mathrm{AD}_J = (\mathrm{MA}_J(t) - \mu_{\mathrm{MA}_J})^T \sum_{\mathrm{MA}_J}^{-1} (\mathrm{MA}_J(t) - \mu_{\mathrm{MA}_J}) \tag{7}$$

where μ_{MA_j} and $\sum_{\mathrm{MA}_J}$ are the mean vector and variance matrix under the normal operation condition corresponding to the grouped attributes J. $\mathrm{MA}_J(t)$ is the current measurement of attribute group J.

3.2.3 Information-based anomaly analysis

We have developed the AD metric based on Hotelling's T^2 method as described in section 'Multivariate anomaly analysis'. When multivariate analysis method is used to calculate the AD, the variance matrix has to be computed. Consequently, we need to compute the correlation between any pair of features in the whole feature set. There are two important limitations of the statistical multivariate analysis method. The first one is that all features are assumed to have the same level of importance. In fact, they are not in most scenarios. Second, when the number of features is large, the calculation will have high computational overhead and thus it will make online monitoring and analysis infeasible, too long to be useful for online computations. Hence, we developed an alternative approach to compute the anomaly distance metric. In this approach, we choose the MAs based on information theory in order to improve the prediction and accuracy of the desired data mining task, i.e., detection of an abnormal behavior in a

network service due to network attacks. After computing the optimal feature set, a linear classification function is trained by using a genetic algorithm (GA) in order to determine the appropriate weights that will produce the appropriate detection rates for any type of network attacks (*DoS*, *U2R*, *R2L*, *probe*). The linear function represents the weight summation of discrete and/or discretized feature values. We have validated our approach using the DARPA KDD99 benchmark dataset (KDD, 1999). More information on feature selection algorithm, learning classification algorithm, and the results will be discussed in further detail in the next section.

4 Information theory-based anomaly analysis

4.1 Introduction

In this section, we present a new approach that effectively removes irrelevant features from the ranked feature list based on the mutual information between each feature and the decision variable. We obtain the ranked lists of features by using a simple forward selection hill climbing search starting with an empty set and evaluating each feature individually and forcing it to continue to the far side of the search space. Redundant features are removed through the pair wise decision dependent correlation analysis. The evaluation process of subset features is done in the abridged ranked lists of features after reducing irrelevant features.

4.2 Feature selection techniques

Feature selection techniques can be categorized according to a number of criteria. One popular categorization consists of "filter" and "wrapper" to quantify the worth of features (Das, 2001; Kohavi and John, 1997). Filters use general characteristics of the training data to evaluate attributes and operate independently of any learning algorithm. Wrappers, on the other hand, evaluate attributes by using accuracy estimates provided by the actual target learning algorithm. Due to the fact that wrapper model is computationally expensive (Langley, 1994) the filter model is usually a good choice when the number of features becomes very large.

Das et al. (Das, 2001) combined both models into a hybrid one to improve the performance of a particular learning algorithm. In this section, we focus on the filter model and present a novel feature selection algorithm, which can effectively remove both irrelevant and redundant information.

Evaluation of individual feature emphasizes the relevance of the feature to the final decision. There are two typical individual feature-based evaluation approaches. The first one is information-based feature ranking. In this approach, the mutual information between decision and feature is used

to evaluate the importance of the feature with respect to the decision under consideration. This method is independent of the underlying distribution and especially efficient when the datasets have a sheer dimensionality. The second type of algorithms relies on the relevance evaluation of features such as Relief which is an instance-based feature ranking scheme introduced by Kira and Rendell (1992) and ReliefF which can handle multiple class data enhanced by Kononenko (1994) from Relief. The rationale of Relief and ReliefF is that a useful feature should differentiate between instances from different classes and have the same value for instances from the same class. Relief approach is based on randomly sampling a number (m) of instance from the training dataset and then locating each feature's nearest neighbor from the same and opposite class. The values of the features of the nearest neighbors are compared to the sampled instance and used to update relevant scores for each feature.

Although feature selection techniques that focus only on relevance can reduce the number of features to be considered significantly, they could not help remove the redundant information existing among multiple features. Hall (2000) and Kohavi and John (1997) show that redundant features along with irrelevant features affect severely the accuracy of the learning algorithms. The reason is that if we do not consider the dependency among features, the feature selection algorithm will select multiple highly correlated features. Our results show that the linear summation of the individual mutual information values with respect to a particular decision will not linearly decrease the uncertainty in the decision because of the dependency that exists between features.

Subset searching algorithms search through candidate feature subsets guided by a certain evaluation measure, which captures the goodness of each subset. Some evaluation measures that have been effective in removing both irrelevance and redundancy include consistency measure (Das, 2001; Almuallim and Dietterich, 1991; Liu and Setiono, 1996) and correlation measure (Hall, 2000). The consistency method looks for the minimum combinations of features that could divide the training data into subsets containing a strong single class majority. This separation is hoped to be as consistent as the whole set of features. Correlation-based feature selection evaluates subsets of features rather than individual features. The ideal subsets should contain features that are highly correlated with the decision and have low-level inter-correlation with each other.

4.3 Correlation analysis measure

It has been shown that dependency measures or correlation measures qualify the accuracy of decision to predict the value of one variable (Dash and Liu, 1997). The main shortcomings of classical linear correlations are the assumption of linear correlation between the features and the requirement

that all features contain numerical values. To overcome these shortcomings, several information theory-based measures of association were introduced for the feature–class correlations and features' inter-correlations such as the gain ratio (Quinlan, 1993) and information gain (Al-Ani and Deriche, 2002), the symmetric uncertainty coefficient (Press et al., 2005), and several others based on the minimum description length principle (Kononenko, 1995). Good results were acquired through using the gain ratio for feature–class correlations and symmetric uncertainty for features' inter-correlations (Battiti, 1994; Hall, 2000; Yu and Liu, 2004).

However, the symmetric uncertainty measure is not accurate enough to quantify the dependency among features with respect to a given decision. A critical point was neglected that the correlation or redundancy between features is strongly related with the decision variable under consideration. On the other hand, the symmetric uncertainty may provide false or incomplete information. Hence, to accurately quantify the dependency or correlation among features we have defined new dependency analysis metric and theorems used in our approach (Qu et al., 2005b).

4.4 Experimental methodology

We analyzed the benchmark KDD99 dataset (KDD, 1999) used in the Third International Knowledge Discovery and Data Mining Tools Competition to validate our approach. Lincoln Labs set up an environment to acquire 9 weeks of raw TCP dump data for a local-area network (LAN) simulating a typical U.S. Air Force LAN. A connection is a sequence of TCP packets starting and ending at some well defined times, between which data flow to and from a source IP address to a target IP address. Each connection is labeled as either normal or as an attack, with exactly one specific attack type. It is important to note that the testing data are not from the same probability distribution as the training data. There are 494,021 records in the training dataset and the number of records in the testing dataset are $\sim$5 millions. The datasets contain a total of 22 different attack types. They fall into the following four main categories: DoS, R2L, U2R, and probe attacks.

In the process of feature selection and linear model building, we evaluate three different feature filtering methods: (1) using discrete features only; (2) using continuous and discrete features; (3) using dependency among features in the selection.

4.4.1 Using discrete features

We analyze all the nine discrete features from the original dataset with respect to each category of attacks. Let variable Y represent the decision variable about the behavior of records (normal or attack) and let variable X represent a single discrete feature.

We compute mutual information of discrete features with respect to the decision variable (to detect each type of attack) and find *service*, *logged_in*, *protocol_type*, and *flag* are the four important features because they have the largest mutual information with respect to the decision variable. The other features are irrelevant in terms of their mutual information with respect to the decision variable. Hence, using the mutual information we can reduce the total 41 features to only the 4 features mentioned above.

4.4.2 *Using both discrete and continuous features*

Usually, there are few distinct values for the discrete features, and therefore, it is straightforward to apply information theory. However, the continuous features have a wide range of distinct values. For example, the *src_bytes* feature has 3300 distinct values. It is obvious that different processing methods should be used to handle discrete and continuous features. For the discrete features, we assign a value to each nominal value. But for continuous features, it is not feasible to process each distinct value. There are two concerns. The first one is both overhead in memory and computation if we consider all possible values. The second one is regarding the accuracy of the learning algorithm. If the testing dataset has different distributions in the training dataset there will be some distinct values that do not appear from the training dataset. That will reduce the accuracy of the analysis.

Features are ranked in descending order according to their relevance to the final decision. Tables 1 and 2 show the chosen features reflecting the removal of the irrelevant features. Without feature–feature correlation analysis, we were able to get good detection rates for DoS and probe attacks as shown in Table 3. However, the results are not good for U2R and R2L attacks. Similarly, other algorithms failed to achieve good detection rates for these attacks.

Table 1
Ranked features for DoS and probe attacks

DoS		Probe	
Feature	$I(Y; X)/H(Y)$	Feature	$I(Y; X)/H(Y)$
count	0.89973	src_bytes	0.617323
service	0.823221	service	0.508163
dst_bytes	0.711719	dst_host_diff_srv_rate	0.45012
logged_in	0.545972	dst_bytes	0.425978
dst_host_same_src_port_rate	0.531105	rerror_rate	0.343698
srv_count	0.475077	count	0.341383
protocol_type	0.43273	flag	0.329652
dst_host_count	0.428534	dst_host_srv_diff_host_rate	0.313887
src_bytes	0.403728	same_srv_rate	0.313486

Table 2
Ranked features for U2R and R2L attacks

U2R		R2L	
Feature	$I(Y; X)/H(Y)$	Feature	$I(Y; X)/H(Y)$
service	0.481635	service	0.559618
root_shell	0.37445	dst_host_srv_count	0.328744
dst_host_srv_count	0.281805	dst_host_same_src_port_rate	0.205641
duration	0.26578	dst_host_src_diff_host_rate	0.183676
num_file_creations	0.255163	is_guest_login	0.159472
dst_host_count	0.177618	srv_count	0.149472
dst_host_same_src_port_rate	0.134272	dst_bytes	0.136806
srv_count	0.113392	dst_host_count	0.131907
dst_host_srv_diff_host_rate	0.091564	count	0.131043
src_bytes	0.086327	src_bytes	0.088246

Table 3
Results comparison of different approaches

Class	Our approach using continuous and discrete features (%)	Our approach using discrete features only (%)	Winner entry using C5.0 (%)	CTree (%)
Normal	98.45	98.34	99.5	92.78
DoS	99.93	99.33	97.1	98.91
U2R	75.34	63.64	13.2	88.13
R2L	41.34	5.86	8.4	7.41
Probe	99.91	93.95	83.3	50.35

4.4.3 *Using dependency among features in the selection*

We calculated the decision dependent correlation among the features and obtained two correlation matrices for U2R and R2L attacks. By applying feature selection algorithm shown in Qu et al. (2005b) the features chosen for U2R and R2L attacks are shown in Tables 4 and 5. When using features *service*, *dst_host_srv_count*, *num_file_creations*, *dst_host_count*, and *dst_host_same_src_port_rate* to detect U2R attacks, we get a detection rate of 92.5% with a 0.7587% false alarm. We also note that using the feature set {*service*, *dst_host_srv_count*, *num_file_creations*, *dst_host_count*} can lead to a detection rate of 96.2% with a 1.43% false alarm. These results are significantly better than those obtained using the sequential feature selection approach.

For R2L attacks' detection, the feature selection algorithm yields a feature subset that consists of *service*, *dst_host_same_src_port_rate*, *dst_host_srv_diff_host_rate*, and *dst_host_count*. Using these features in

Table 4
Different feature subsets and their prediction for U2R attacks

Subset (S)	False alarm	Detection rate (%)
{service, dst_host_srv_count, num_file_creations, dst_host_count, dst_host_src_same_port_rate}	0.007587	92.55
{service, dst_host_srv_count, num_file_creations, dst_host_count}	0.01431	96.227
{dst_host_srv_count, duration, num_file_creations, dst_host_count}	0.01531	91.5
{service, dst_host_srv_count, dst_host_count}	0.019583	94.34
{service, dst_host_srv_count, num_file_creations}	0.067961	90.06

Table 5
Different feature subsets and their prediction for R2L attacks

Subset (S)	False alarm	Detection rate (%)
{service, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_count}	0.092581	91.14
{service, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate}	0.09476	92.7
{service, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate}	0.09476	92.37
{service, dst_host_same_src_port_rate}	0.083524	92.46

detection R2L attacks, we get a 91.13% detection rate with a 9.258% false alarm. When training on features *service* and *dst_host_same_src_port_rate*, we obtained a detection rate of 92.47% with 8.35% false alarm. These results give us more meaningful information that the small number of features will result in a much faster learning process and it will reduce the overhead in collecting data when used in real network environment.

5 Conclusion and future works

5.1 Summary

The research presented is part of a large effort to develop an autonomic control and management environment (AUTONOMIA) that provides self-configuring, self-optimizing, self-healing, and self-protecting services.

In this work, we developed a theoretical framework and general analysis methodology to achieve the following: (a) analyzing anomaly operations of networks and applications; (b) automatically identifying critical vulnerable Internet infrastructure resources; and (c) proactively self-protecting networks from a wide range of organized network and/or host attacks. The main research activities presented in this chapter can be highlighted in the following points:

(1) We have studied the behavior of networks and hosts under normal and abnormal conditions that might be caused by attacks. We have also studied techniques to characterize the current states of networks/hosts based on new metrics anomaly distance (AD).

(2) We have studied techniques to monitor and obtain MAs (features) that can be used to characterize the current state of networks and their services. We have studied filtering techniques based on information theory to identify the most important features to detect abnormal behaviors. We validated our approach using the DARPA KDD99 benchmark dataset and the results show that using the new decision dependent correlation metric we can detect efficiently the difficulty to detect network attacks such as U2R attack and R2L attack. The best reported detection rates for U2R and R2L on the KDD99 datasets were 13.2% and 8.4% with 0.5% false alarm, respectively. For U2R attacks, our approach can achieve a 92.5% detection rate with false alarm of 0.7587%. For R2L attacks, this approach can achieve a 92.47% detection rate with false alarm of 8.35%. These feature importance analyses also help identify the optimal feature set that must be monitored and analyzed to determine whether the target system is under attack or not.

5.2 Future direction

Our future research activities will include using the theoretical framework discussed in this chapter to develop a self-protection engine that can proactively detect any anomaly caused by network attacks and protect against them. We need to monitor not only network traffic, but also packet payload information as well as host activities and correlate these activities in order to accurately detect anomaly caused by known or unknown network attacks with minimal false alarms. Most of the current research has focused on external attacks and very little research addressed insider attacks. Our future research will also expand the current self-protection framework to include insider attacks as well as external attacks.

Anomaly-based analysis of network attacks is still immature research area and many researches must be carried out in order to reduce the false alarms that are triggered when the current normal behavior changes due to changes in topology, user tasks or adding new services. All these will lead to new profiles that were not known before and most of anomaly-based

techniques will characterize them as attacks rather than being normal. They might also fail to recognize a new attack behavior that is very close to normal behavior. More research is needed in the following areas:

1. How to accurately define what we mean by normal user/network behavior and how this can be adopted in real-time as network and user behaviors change dynamically?
2. How to filter the huge amount of information that can be generated from high-speed networks without severely impacting the accuracy of detection and our ability to proactively react to network attacks in a timely manner?
3. How to automatically generate anomaly signatures that can be used to detect a wide range of the network attacks?
4. How to scale the monitoring, detection, and protection in large scale networks that involve thousands or even hundred of thousands of resources?

6 Questions and discussions

The following questions propose more interests and challenges to the network security research communities:

1. Discuss new approaches to detect attacks in the presence of code obfuscation and highly variable benign traffic, through existing statistical methods and data mining techniques that can help profile the network behavior.
2. Discuss the scalability of attack detection and protection in large scale high-speed networks in conjunction with the real-time response to some fast attacks. Some very fast propagating attacks (viruses, worms, etc.) can infect large amount of vulnerable machines within minutes or even seconds. So it is critical for developing the approaches that are scalable to the line speeds.
3. Discuss the use of supervised and unsupervised learning algorithms to network security prevention systems. During the run time, the validation of the intermediate data directly affects the accuracy of the training results.
4. Discuss the compatibility and inter-operability issues among multiple security systems. Provided each existing network security system can protect the network system against some specific categories of network attacks, it will be effective to share the knowledge among the network security system to achieve the maximum extent of security.
5. Discuss root cause analysis and how it can be used to mitigate the impact of network attacks and to accurately pinpoint the type of attacks and the location information (source, target, channels, etc.) in order to apply the appropriate protection schemes.

References

Al-Ani, A., M. Deriche (2002). Feature selection using a mutual information based measure, in: *Proceedings of 16th International Conference on Pattern Recognition*, Vol. 4, pp. 82–85.

Albert, R., H. Jeong, A.-L. Barabási (2000). The Internet's Achilles' heel: error and attack tolerance of complex networks. *Nature* 406, 378–382.

Almuallim, H., T.G. Dietterich (1991). Learning with many irrelevant features, in: *Proceedings of the Ninth National Conference on Artificial Intelligence*, AAAI Press, Menlo Park, CA, pp. 547–552.

Anderson, D., T. Frivold, A. Valdes (May 1995). Next-generation intrusion detection expert system (NIDES): a summary. Technical Report SRI-CSL-95-07, Computer Science Laboratory, SRI International, Menlo Park, CA.

Batchelor, B.G. (1978). *Pattern Recognition: Ideas in Practice*. Plenum Press, New York, pp. 71–72.

Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks* 5, 537–550.

Bellovin, S.M. (March 2000). ICMP Traceback Message, Internet Draft: draft-bellovin-itrace-00.txt.

Botha, M., R.V. Solms, K. Perry, E. Loubser, G. Yamoyany (2002). *The Utilization of Artificial Intelligence in a Hybrid Intrusion Detection System*. South African Institute for Computer Scientists and Information Technologists, Republic of South Africa, pp. 149–155.

CERT (1997). CERT® Advisory CA-1997-28 IP Denial-of-Service Attacks, retrieved from http://www.cert.org/advisories/CA-1997-28.html on April 2004.

CERT (2001a). CERT® Advisory CA-2001-19 "Code Red" Worm Exploiting Buffer Overflow in IIS Indexing Service DLL, retrieved from http://www.cert.org/advisories/CA-2001-19.html on April 2004.

CERT (2001b). CERT® Advisory CA-2001-26 Nimda Worm, http://www.cert.org/advisories/CA-2001-26.html

CERT (2003a). CERT® Advisory CA-2003-04 MS-SQL Server Worm, http://www.cert.org/advisories/CA-2003-04.html

CERT (2003b). CERT® Advisory CA-2003-20 W32/Blaster worm, http://www.cert.org/advisories/CA-2003-20.html

CERT (2003c). CERT® Incident Note IN-2003-03, http://www.cert.org/incident_notes/IN-2003-03.html

Crosbie, M., B. Dole, T. Ellis, I. Krsul, E. Spafford (September 1996). IDIOT—users guide. Technical Report TR-96-050, Purdue University, COAST Laboratory.

Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection, in: *Proceedings of the 18th International Conference on Machine Learning*, pp. 74–81.

Dash, M., H. Liu (1997). Feature selection methods for classifications. *Intelligent Data Analysis: An International Journal* 1(3), 131–156 http://www-east.elsevier.com/ida/free.htm

Denning, D.E. (1987). An intrusion detection model. *IEEE Transactions on Software Engineering* SE-13, 222–232.

Diday, E. (1974). Recent progress in distance and similarity measures in pattern recognition, in: *Second International Joint Conference on Pattern Recognition*, pp. 534–539.

Dittrich, D. (2007). Distributed denial of service (DDoS) Attacks/Tools Page, from http://staff.washington.edu/dittrich/ddos/

Ferguson, P., D. Senie (January 1998). Network Ingress Filtering: Defeating Denial of Service Attacks which Employ IP Source Address Spoofing, RFC 2267.

Gaudin, S. (2003). 2003 Worst Year Ever for Viruses, Worms, http://www.internetnews.com/infra/article.php/3292461

Gibson, S. (February 2002). Distributed Reflection Denial of Service, retrieved from http://grc.com/dos/drdos.htm

Hall, M.A. (2000). Correlation-based feature selection for discrete and numeric class machine learning, in: *Proceedings of the 17th International Conference on Machine Learning*, pp. 359–366.

Hariri, S., G. Qu, T. Dharmagadda, M. Ramkishore, C.S. Raghavendra (2003). Impact analysis of faults and attacks in large-scale networks. *IEEE Security and Privacy* 1(5), 49–54.

Hochberg, J., K. Jackson, C. Stallings, J.F. McClary, D. DuBois, J. Ford (1993). Nadir: an automated system for detecting network intrusion and misuse. *Computers and Security* 12(3), 235–248.

Hou, Y.T., Y. Dong, Z. Zhang (1998). Network Performance Measurement and Analysis Part 1: A Server-Based Measurement Infrastructure, retrieved on April 20, 2003, from http://citeseer.nj.nec.com/249251.html

Ilgun, K., A.R. Kemmerer, A.P. Porras (1995). State transition analysis: a rule-based intrusion detection approach. *IEEE Transactions on Software Engineering* 21(3), 181–199.

Insecure Inc. (1996). Ping of Death, http://www.insecure.org/sploits/ping-o-death.html

Ioannidis, J., S.M. Bellovin (2002). Implementing pushback: router-based defense against DDoS attacks, in: *Proceedings of NDSS'2002, February*, San Diego, CA.

Jin, J., J. Shi (2001). Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *Journal of Intelligent Manufacturing* 12, 257–268.

KDD (1999). http://kdd.ics.uci.edu/databases/kddcup99/task.html

Kemmerer, R., G. Vigna (2002). Intrusion detection: a brief history and overview. *IEEE Computer* 35, 27–30.

Kendall, K. (1998). A database of computer attacks for the evaluation of intrusion detection systems. Master's thesis, Massachusetts Institute of Technology.

Keromytis, A.D., V. Misra, D. Rubenstein (2002). SOS: secure overlay services, in: *Proceedings of ACM SIGCOMM 2002, August*.

Killalea, T. (November 2000). Recommended Internet Service Provider Security Services and Procedures, RFC 3013.

Kira, K., L.A. Rendell (1992). A practical approach to feature selection, in: *Ninth International Workshop on Machine Intelligence*, Morgan Kaufmann, Aberdeen, Scotland.

Kohavi, R., G.H. John (1997). Wrappers for feature subset selection. *Artificial Intelligence* 97(1–2), 273–324.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of relief, in: *Proceedings of the Seventh European Conference on Machine Learning*, Springer-Verlag, Catania, pp. 171–182.

Kononenko, I. (1995). On biases in estimating multi-valued attributes, *IJCAI*, pp. 1034–1040.

Langley, P. (1994). Selection of relevant features in machine learning, in: *Proceedings of the AAAI Fall Symposium on Relevance*, AAAI Press, Menlo Park, CA.

Liu, H., R. Setiono (1996). A probabilistic approach to feature selection: a filter solution, in: *Proceedings of the 13th International Conference on Machine Learning*, Morgan Kaufmann, Menlo Park, CA, pp. 319–327.

Liu, Y., D. Tipper, D. Medhi, A. Srikitja (2000). Self-configuring Survivable Techniques for Quality of Service Enabled Internet.

Lunt, T., A. Tamaru, F. Gilham, R. Jagannathan, C. Jalali, P.G. Neumann, H.S. Javitz, A. Valdes, T.D. Garvey (February 1992). A Real Time Intrusion Detection Expert System (IDES)—Final Report, SRI International, Menlo Park, CA.

Manjan, R., S.M. Bellovin, S. Floyd, J. Ioannidis, V. Paxson, S. Shenker (July 2001). Controlling high bandwidth aggregates in the network. ICSI Technical Report.

Michalski, R.S., R.E. Stepp, E. Diday (1981). A recent advance in data analysis: clustering objects into classes characterized by conjunctive concepts, in: L.N. Kanal, A. Rosenfeld (eds.), *Progress in Pattern Recognition*, Vol. 1. North-Holland, New York, pp. 33–56.

Montgomery, D.C. (2001). *Design and Analysis of Experiments* 5th ed. Wiley, New York.

Moore, D., G. Voelker, S. Savage (2001). Inferring Internet denial of service activity, in: *Proceedings of USENIX Security Symposium'2001, August*.

Microsoft. (2003). Microsoft Security Bulletin MS03-026 Buffer Overrun in RPC Interface could allow Code Execution (823980), http://www.microsoft.com/technet/security/bulletin/MS03-026.mspx

Nadler, M., E.P. Smith (1993). *Pattern Recognition Engineering*. Wiley, New York pp, 293–294.

National Research Council. (2003). Committee on the Internet under Crisis Conditions: Learning from the Impact of September11, The National Academies Press, Washington, DC, retrieved from http://www.nap.edu/books/0309087023/html/.

Park, K., H. Lee (2001). On the effectiveness of route-based packet filtering for DDoS attack prevention in power-low Internets, in: *Proceedings of the ACM SIGCOMM, August*, ACM Press, New York, pp. 15–26.

Parashar, M. (1994). *Interpretive performance prediction for high performance computing*. Ph.D. thesis, Department of Computer Engineering, Syracuse University.

Paxson, V. (2001). An analysis of using reflectors for distributed denial-of-service attacks. *Computer Communication Review* 31(3), 38–47.

Phillips, C., L.P. Swiler (1998). A graph-based system for network-vulnerability analysis, in: *Proceedings of the 1998 Workshop on New Security Paradigms*, Charlottesville, Virginia, United States, pp. 71–79.

Press, W.H., B.P. Flannery, S.A. Teukolski, W.T. Vetterling (2005). Numerical Recipes in C. Cambridge University Press, http://www.library.cornell.edu/nr/bookcpdf.html

Qu, G., S. Hariri, X. Zhu, J. Jin, M. Yousif (2005a). Multivariate Statistical Online Analysis for Self Protection against Network Attacks, AICSSA'05.

Qu, G., S. Hariri, M. Yousif (2005b). A new dependency and correlation analysis for features. *IEEE Transactions on Knowledge and Data Engineering* 17(9), 1199–1207.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Report to the President's Commission on Critical Infrastructure Protection. (1997). Threat and Vulnerability Model for Information Security.

Smaha, S.E., J. Winslow (1994). Misuse detection tools. *Computer Security Journal* 10(1, Spring), 39–49.

Savage, S., D. Wetherall, A. Karlin, T. Anderson (2000). *Practical network support for IP traceback, in: Proceedings of ACM SIGCOMM'2000, August*.

Snoren, A.C., C. Partridge, L.A. Sanchez, C.E. Jones, F. Tchakountio, S.T. Kent, W.T. Strayer (2001). Hash-based IP traceback, in: *Proceedings of ACM SIGCOMM'2001, March*.

Song, D., A. Perrig (2001). Advanced and authenticated marking schemes for IP traceback, in: *Proceedings of ACM SIGCOMM'2001, March*.

Stone, R. (2000). CenterTrack: an IP overlay network for tracking DoS floods, in: *Proceedings of Ninth USENIX Security Symposium, August*, Denver, CO.

Tjaden, B., L. Welch, S. Ostermann, D. Chelberg, R. Balupari, M. Bykova, A. Mitchell, D. Lissitsyn, L. Tong (2000). INBOUNDS: The Integrated Network-Based Ohio University Network Detective Service, retrieved from http://www.mts.jhu.edu/~marchette/ID04/Papers/SCI2000.pdf

Webopedia (2004). DoS Attack, retrieved on April 2004 from http://www.webopedia.com/TERM/D/DoS_attack.html

Yau, D., J. Liu, F. Liang (2002). Defending against distributed denial-of-service attacks with max–min fair server-centric router throttles, in: *Proceedings of IWQoS'2002, May*, Miami Beach, FL.

Ye, N., Q. Chen, S.M. Emran, S. Vilbert (2000). Hotelling's T^2 multivariate profiling for anomaly detection, in: *Proceedings of the 2000 IEEE Workshop on Information Assurance and Security, June*, West Point, NY.

Ye, N., Q. Chen (2001). An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. *Quality and Reliability Engineering Journal* 17, 105–112.

Yu, L., H. Liu (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research* 5(October), 1205–1224.

Subject Index

This page intentionally left blank