Paul Kantor Gheorghe Muresan
Fred Roberts Daniel D. Zeng
Fei-Yue Wang Hsinchun Chen
Ralph C. Merkle (Eds.)

LNCS 3495

Intelligence and Security Informatics

IEEE International Conference
on Intelligence and Security Informatics, ISI 2005
Atlanta, GA, USA, May 2005, Proceedings

Springer

Lecture Notes in Computer Science 3495

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Paul Kantor Gheorghe Muresan
Fred Roberts Daniel D. Zeng
Fei-Yue Wang Hsinchun Chen
Ralph C. Merkle (Eds.)

Intelligence and Security Informatics

IEEE International Conference
on Intelligence and Security Informatics, ISI 2005
Atlanta, GA, USA, May 19-20, 2005
Proceedings

Springer

Volume Editors

Paul Kantor
Gheorghe Muresan
Rutgers University
School of Communication, Information and Library Studies
4 Huntington Street, New Brunswick, NJ 08901-1071, USA
E-mail: {kantor,muresan}@scils.rutgers.edu

Fred Roberts
Rutgers University
Department of Mathematics
Center for Discrete Mathematics and Theoretical Computer Science
96 Frelinghuysen Road, Piscataway, NJ 08854-8018, USA
E-mail: froberts@dimacs.rutgers.edu

Daniel D. Zeng
Hsinchun Chen
University of Arizona
Department of Management Information Systems
1130 East Helen Street, Tucson, AZ 85721-0108, USA
E-mail: {zeng,hchen}@eller.arizona.edu

Fei-Yue Wang
University of Arizona
Department of Systems and Industrial Engineering
1127 East North Campus Drive, Tucson, AZ 85721-0020, USA
E-mail: feiyue@sie.arizona.edu

Ralph C. Merkle
Georgia Institute of Technology
College of Computing, Georgia Tech Information Security Center
801 Atlantic Drive, Atlanta, GA 30332-0280, USA
E-mail: merkle@cc.gatech.edu

Library of Congress Control Number: 2005925606

CR Subject Classification (1998): H.4, H.3, C.2, H.2, D.4.6, K.4.1, K.5, K.6

ISSN 0302-9743
ISBN-10 3-540-25999-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25999-2 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11427995 06/3142 5 4 3 2 1 0

Preface

Intelligence and security informatics (ISI) can be broadly defined as the study of the development and use of advanced information technologies and systems for national and international security-related applications, through an integrated technological, organizational, and policy-based approach. In the past few years, ISI research has experienced tremendous growth and attracted substantial interest from academic researchers in related fields as well as practitioners from both government agencies and industry.

The first two meetings (ISI 2003 and ISI 2004) in the ISI symposium and conference series were held in Tucson, Arizona, in 2003 and 2004, respectively. They provided a stimulating intellectual forum for discussion among previously disparate communities: academic researchers in information technologies, computer science, public policy, and social studies; local, state, and federal law enforcement and intelligence experts; and information technology industry consultants and practitioners.

Building on the momentum of these ISI meetings and with sponsorship by the IEEE, we held the IEEE International Conference on Intelligence and Security Informatics (ISI 2005) in May 2005 in Atlanta, Georgia. In addition to the established and emerging ISI research topics covered at past ISI meetings, ISI 2005 included a new track on Terrorism Informatics, which is a new stream of terrorism research leveraging the latest advances in social science methodologies, and information technologies and tools.

ISI 2005 was jointly hosted by Rutgers, the State University of New Jersey; the University of Arizona (UA); and the Georgia Institute of Technology (GATECH). The two-day program included one plenary panel discussion session focusing on the perspectives and future research directions of the government funding agencies, several invited panel sessions, 62 regular papers, and 38 posters. In addition to the main sponsorship from the National Science Foundation, the Intelligence Technology Innovation Center, and the US Department of Homeland Security, the conference was also co-sponsored by several units within the three hosting universities including: the School of Communication, Information and Library Studies at Rutgers; the Center for Discrete Mathematics and Theoretical Computer Science at Rutgers; the Eller College of Management and the Management Information Systems Department at UA; the NSF COPLINK Center of Excellence at UA; the Mark and Susan Hoffman E-Commerce Laboratory at UA; the Artificial Intelligence Laboratory at UA; the Program for Advanced Research in Complex Systems at UA; the College of Computing at GATECH; and the Georgia Tech Information Security Center.

We wish to express our gratitude to all members of the ISI 2005 Program Committee and additional reviewers who provided high-quality, constructive review comments under an unreasonably short lead-time. Our special thanks go to Dr. Joshua Sinai and Dr. Edna Reid who recruited high-caliber contributors from the terrorism

informatics research community and helped process submissions in the terrorism informatics area. We wish to express our gratitude to Ms. Catherine Larson, Ms. Priscilla Rasmussen, Mr. Shing Ka Wu, Ms. Shelee Saal, and Ms. Sarah Donnelly for providing excellent conference logistics support. We also would like to thank Mr. Jialun Qin, Mr. Daning Hu, Mr. Tao Wang, Mr. Jian Ma, and Mr. Xiaopeng Zhong, all graduates students at UA, for their great help with compiling the proceedings.

ISI 2005 was co-located with the 6th Annual National Conference on Digital Government Research (DG.O). We wish to thank the DG.O organizers and support staff for their cooperation and assistance. We also would like to thank the Springer LNCS editorial and production staff for their professionalism and continuous support for the ISI symposium and conference series.

Our sincere gratitude goes to all of the sponsors. Last, but not least, we thank Dr. Joshua Sinai, Dr. Art Becker, Dr. Michael Pazzani, Dr. Valerie Gregg, and Dr. Larry Brandt for their strong and continuous support of the ISI series and other related ISI research.

May 2005

Paul Kantor
Gheorghe Muresan
Fred Roberts
Daniel Zeng
Fei-Yue Wang
Hsinchun Chen
Ralph Merkle

Organization

ISI 2005 Organizing Committee

Conference Co-chairs:

Ralph Merkle	Georgia Institute of Technology
Hsinchun Chen	University of Arizona

Program Co-chairs:

Paul Kantor	Rutgers University
Fred Roberts	Rutgers University
Fei-Yue Wang	University of Arizona
Gheorghe Muresan	Rutgers University
Daniel Zeng	University of Arizona

Government Liaisons:

Valerie Gregg	National Science Foundation
Art Becker	Intelligence Technology Innovation Center
Joshua Sinai	Department of Homeland Security

ISI 2005 Program Committee

Yigal Arens	University of Southern California
Antonio Badia	University of Louisville
Art Becker	Intelligence Technology Innovation Center
Don Berndt	University of South Florida
Larry Brandt	National Science Foundation
Donald Brown	University of Virginia
Judee Burgoon	University of Arizona
Guoray Cai	Pennsylvania State University
Kathleen Carley	Carnegie Mellon University
Michael Chau	University of Hong Kong
Sudarshan S. Chawathe	University of Maryland
Peter Chen	Louisiana State University
Lee-Feng Chien	Academia Sinica, Taiwan
Wingyan Chung	University of Texas at El Paso
Greg Conti	United States Military Academy
Ram Dantu	University of North Texas
Boyd Davis	University of North Carolina at Charlotte
Chris Demchak	University of Arizona

James Ellis	National Memorial Institute for the Prevention of Terrorism
Mark Frank	Rutgers University
Susan Gauch	University of Kansas
Johannes Gehrke	Cornell University
Joey George	Florida State University
Mark Goldberg	Rensselaer Polytechnic Institute
Valerie Gregg	National Science Foundation
Bob Grossman	University of Illinois at Chicago
Jiawei Han	University of Illinois at Urbana-Champaign
Alan R. Hevner	University of South Florida
Eduard Hovy	University of Southern California
Paul Hu	University of Utah
Judith Klavans	University of Maryland
Moshe Koppel	Bar-Ilan University, Israel
Don Kraft	Louisiana State University
Taekyoung Kwon	Sejong University, Korea
Seok-Won Lee	University of North Carolina at Charlotte
Gondy Leroy	Claremont Graduate University
Haifeng Li	Jilin University, China
Ee-peng Lim	Nanyang Technological University, Singapore
Chienting Lin	Pace University
Cecil Lynch	University of California at Davis
Therani Madhusudan	University of Arizona
Ofer Melnik	Rutgers University
Brinton Milward	University of Arizona
Pitu Mirchandani	University of Arizona
Clifford Neuman	University of Southern California
Greg Newby	University of Alaska, Fairbanks
Kwong Bor Ng	City University of New York
Jay Nunamaker	University of Arizona
Joon S. Park	Syracuse University
Ganapati P. Patil	Pennsylvania State University
Neal Pollard	Science Applications International Corporation
Peter Probst	Institute for the Study of Terrorism and Political Violence
Karin Quinones	University of Arizona
Yael Radlauer	International Policy Institute for Counter-Terrorism
Ram Ramesh	State University of New York at Buffalo
H. Raghav Rao	State University of New York at Buffalo
Edna Reid	University of Arizona
Mirek Riedewald	Cornell University
Dmitri Roussinov	Arizona State University
Marc Sageman	University of Pennsylvania
Raghu Santanam	Arizona State University

Reid Sawyer	United States Military Academy
Olivia Sheng	University of Utah
Amit Sheth	University of Georgia
Elizabeth Shriberg	Stanford Research Institute
Andrew Silke	University of East London, UK
Joshua Sinai	Department of Homeland Security
David Skillicorn	Queen's University, Canada
Cole Smith	University of Arizona
Sal Stolfo	Columbia University
Gary Strong	MITRE
Katia Sycara	Carnegie Mellon University
Paul Thompson	Dartmouth College
Ajay Vinze	Arizona State University
Ke Wang	Simon Fraser University, Canada
Gabriel Weimann	University of Haifa, Israel
Andrew Whinston	University of Texas at Austin
Chris Yang	Chinese University of Hong Kong
Mohammed Zaki	Rensselaer Polytechnic Institute
Xiangmin Zhang	Rutgers University
Huimin Zhao	University of Wisconsin, Milwaukee
Lina Zhou	University of Maryland, Baltimore County
David Zimmermann	Federal Bureau of Investigation

Additional Reviewers

Boanerges Aleman-Meza	University of Georgia
Homa Atabakhsh	University of Arizona
Jinwei Cao	University of Arizona
Wei Chang	University of Arizona
Dmitriy Fradkin	Rutgers University
Benjamin Fung	Simon Fraser University
Wei Gao	University of Arizona
Zan Huang	University of Arizona
Eul Gyu Im	Hanyang University, Korea
Vandana Janeja	Rutgers University
Dae-Ki Kang	Iowa State University
Gunes Kayacik	Dalhousie University, Canada
Siddharth Kaza	University of Arizona
Kameswari Kotapati	Pennsylvania State University
Guanpi Lai	University of Arizona
Jian Ma	University of Arizona
Anirban Majumdar	University of Auckland, New Zealand
Thomas Meservy	University of Arizona
Mark Patton	University of Arizona
Jialun Qin	University of Arizona
Rob Schumaker	University of Arizona

Yael Shahar	International Policy Institute for Counter Terrorism
Ambareen Siraj	Mississippi State University
Catherine L. Smith	Rutgers University
Shuang Sun	Pennsylvania State University
Wei Sun L.	Harbin Engineering University, China
Alan Wang	University of Arizona
Jennifer Xu	University of Arizona
Bulent Yener	Rensselaer Polytechnic Institute
Ozgur Yilmazel	Syracuse University
Yilu Zhou	University of Arizona
William Zhu	University of Auckland, New Zealand

Table of Contents

Part I: Long Papers

Data and Text Mining

Collusion Set Detection Through Outlier Discovery
Vandana P. Janeja, Vijayalakshmi Atluri, Jaideep Vaidya, Nabil R. Adam 1

Digging in the Details: A Case Study in Network Data Mining
John Galloway, Simeon J. Simoff 14

Efficient Identification of Overlapping Communities
Jeffrey Baumes, Mark Goldberg, Malik Magdon-Ismail 27

Event-Driven Document Selection for Terrorism Information Extraction
Zhen Sun, Ee-Peng Lim, Kuiyu Chang, Teng-Kwee Ong, Rohan K. Gunaratna 37

Link Analysis Tools for Intelligence and Counterterrorism
Antonio Badia, Mehmed Kantardzic 49

Mining Candidate Viruses as Potential Bio-terrorism Weapons from Biomedical Literature
Xiaohua Hu, Illhoi Yoo, Peter Rumm, Michael Atwood 60

Private Mining of Association Rules
Justin Zhan, Stan Matwin, LiWu Chang 72

Infrastructure Protection and Emergency Response

Design Principles of Coordinated Multi-incident Emergency Response Systems
Rui Chen, Raj Sharman, H. Raghav Rao, Shambhu J. Upadhyaya 81

Multi-modal Biometrics with PKI Technologies for Border Control Applications
Taekyoung Kwon, Hyeonjoon Moon 99

Risk Management Using Behavior Based Bayesian Networks
Ram Dantu, Prakash Kolan 115

Sensitivity Analysis of an Attack Containment Model
Ram Dantu, João Cangussu, Janos Turi 127

Toward a Target-Specific Method of Threat Assessment
Yael Shahar 139

Information Management

Incident and Casualty Databases as a Tool for Understanding Low-Intensity Conflicts
Don Radlauer 153

Integrating Private Databases for Data Analysis
Ke Wang, Benjamin C.M. Fung, Guozhu Dong 171

Deception Detection and Authorship Analysis

Applying Authorship Analysis to Arabic Web Content
Ahmed Abbasi, Hsinchun Chen 183

Automatic Extraction of Deceptive Behavioral Cues from Video
Thomas O. Meservy, Matthew L. Jensen, John Kruse, Judee K. Burgoon, Jay F. Nunamaker 198

Automatically Determining an Anonymous Author's Native Language
Moshe Koppel, Jonathan Schler, Kfir Zigdon 209

Monitoring and Surveillance

A Cognitive Model for Alert Correlation in a Distributed Environment
Ambareen Siraj, Rayford B. Vaughn 218

Beyond Keyword Filtering for Message and Conversation Detection
David B. Skillicorn 231

Content-Based Detection of Terrorists Browsing the Web Using an Advanced Terror Detection System (ATDS)
Yuval Elovici, Bracha Shapira, Mark Last, Omer Zaafrany Menahem Friedman, Moti Schneider, Abraham Kandel 244

Modeling and Multiway Analysis of Chatroom Tensors
Evrim Acar, Seyit A. Çamtepe, Mukkai S. Krishnamoorthy, Bülent Yener 256

Selective Fusion for Speaker Verification in Surveillance
Yosef A. Solewicz, Moshe Koppel 269

Terrorism Informatics

A New Conceptual Framework to Resolve Terrorism's Root Causes
Joshua Sinai 280

Analyzing Terrorist Networks: A Case Study of the Global Salafi Jihad Network
Jialun Qin, Jennifer J. Xu, Daning Hu, Marc Sageman, Hsinchun Chen 287

A Conceptual Model of Counterterrorist Operations
David Davis, Allison Frendak-Blume, Jennifer Wheeler, Alexander E.R. Woodcock, Clarence Worrell III 305

Measuring Success in Countering Terrorism: Problems and Pitfalls
Peter S. Probst 316

Mapping the Contemporary Terrorism Research Domain: Researchers, Publications, and Institutions Analysis
Edna Reid, Hsinchun Chen 322

Testing a Rational Choice Model of Airline Hijackings
Laura Dugan, Gary LaFree, Alex R. Piquero 340

Part II: Short Papers

Data and Text Mining

Analysis of Three Intrusion Detection System Benchmark Datasets Using Machine Learning Algorithms
H. Güneş Kayacik, Nur Zincir-Heywood 362

Discovering Identity Problems: A Case Study
Alan G. Wang, Homa Atabakhsh, Tim Petersen, Hsinchun Chen 368

Efficient Discovery of New Information in Large Text Databases
R.B. Bradford 374

Leveraging One-Class SVM and Semantic Analysis to Detect Anomalous Content
Ozgur Yilmazel, Svetlana Symonenko, Niranjan Balasubramanian, Elizabeth D. Liddy 381

LSI-Based Taxonomy Generation: The Taxonomist System
Janusz Wnek 389

Some Marginal Learning Algorithms for Unsupervised Problems
Qing Tao, Gao-Wei Wu, Fei-Yue Wang, Jue Wang 395

Information Management and Sharing

Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites
Edna Reid, Jialun Qin, Yilu Zhou, Guanpi Lai, Marc Sageman, Gabriel Weimann, Hsinchun Chen 402

Evaluating an Infectious Disease Information Sharing and Analysis System
Paul J.-H. Hu, Daniel Zeng, Hsinchun Chen, Catherine Larson, Wei Chang, Chunju Tseng 412

How Question Answering Technology Helps to Locate Malevolent Online Content
Dmitri Roussinov, Jose Antonio Robles-Flores 418

Information Supply Chain: A Unified Framework for Information-Sharing
Shuang Sun, John Yen 422

Map-Mediated GeoCollaborative Crisis Management
Guoray Cai, Alan M. MacEachren, Isaac Brewer, Mike McNeese, Rajeev Sharma, Sven Fuhrmann 429

Thematic Indicators Derived from World News Reports
Clive Best, Erik Van der Goot, Monica de Paola 436

Copyright and Privacy Protection

A Novel Watermarking Algorithm Based on SVD and Zernike Moments
Haifeng Li, Shuxun Wang, Weiwei Song, Quan Wen 448

A Survey of Software Watermarking
William Zhu, Clark Thomborson, Fei-Yue Wang 454

Data Distortion for Privacy Protection in a Terrorist Analysis System
Shuting Xu, Jun Zhang, Dianwei Han, Jie Wang 459

Deception Detection

Deception Across Cultures: Bottom-Up and Top-Down Approaches
Lina Zhou, Simon Lutterbie 465

Detecting Deception in Synchronous Computer-Mediated Communication Using Speech Act Profiling
Douglas P. Twitchell, Nicole Forsgren, Karl Wiers, Judee K. Burgoon, Jay F. Nunamaker Jr. 471

Information Security and Intrusion Detection

Active Automation of the DITSCAP
Seok Won Lee, Robin A. Gandhi, Gail-Joon Ahn, Deepak S. Yavagal 479

An Ontological Approach to the Document Access Problem of Insider Threat
Boanerges Aleman-Meza, Phillip Burns, Matthew Eavenson, Devanand Palaniswami, Amit Sheth 486

Filtering, Fusion and Dynamic Information Presentation: Towards a General Information Firewall
Gregory Conti, Mustaque Ahamad, Robert Norback 492

Intrusion Detection System Using Sequence and Set Preserving Metric
Pradeep Kumar, M. Venkateswara Rao, P. Radha Krishna, Raju S. Bapi, Arijit Laha 498

The Multi-fractal Nature of Worm and Normal Traffic at Individual Source Level
Yufeng Chen, Yabo Dong, Dongming Lu, Yunhe Pan 505

Learning Classifiers for Misuse Detection Using a Bag of System Calls Representation
Dae-Ki Kang, Doug Fuller, Vasant Honavar 511

Infrastructure Protection and Emergency Response

A Jackson Network-Based Model for Quantitative Analysis of Network Security
Zhengtao Xiang, Yufeng Chen, Wei Jian, Fei Yan 517

Biomonitoring, Phylogenetics and Anomaly Aggregation Systems
David R.B. Stockwell, Jason T.L. Wang 523

CODESSEAL: Compiler/FPGA Approach to Secure Applications
Olga Gelbart, Paul Ott, Bhagirath Narahari, Rahul Simha, Alok Choudhary, Joseph Zambreno 530

Computational Tool in Infrastructure Emergency Total Evacuation Analysis
Kelvin H.L. Wong, Mingchun Luo 536

Performance Study of a Compiler/Hardware Approach to Embedded Systems Security
Kripashankar Mohan, Bhagirath Narahari, Rahul Simha, Paul Ott, Alok Choudhary, Joseph Zambreno 543

A Secured Mobile Phone Based on Embedded Fingerprint Recognition Systems
Xinjian Chen, Jie Tian, Qi Su, Xin Yang, Fei-Yue Wang 549

Terrorism Informatics

Connections in the World of International Terrorism
Yael Shahar 554

Forecasting Terrorism: Indicators and Proven Analytic Techniques
Sundri K. Khalsa 561

Forecasting Terrorist Groups' Warfare: 'Conventional' to CBRN
Joshua Sinai 567

The Qualitative Challenge of Insurgency Informatics
Scott Tousley 571

The Application of PROACT® RCA to Terrorism/Counter Terrorism Related Events
Robert J. Latino 579

Part III: Extended Abstracts for Posters and Demos

Data and Text Mining

A Group Decision-Support Method for Search and Rescue Based on Markov Chain
Huizhang Shen, Jidi Zhao, Ying Peng 590

A New Relationship Form in Data Mining
Suwimon Kooptiwoot, Muhammad Abdus Salam 593

A Study of "Root Causes of Conflict" Using Latent Semantic Analysis
Mihaela Bobeica, Jean-Paul Jéral, Teofilo Garcia, Clive Best 595

An Empirical Study on Dynamic Effects on Deception Detection
Tiantian Qin, Judee K. Burgoon 597

Anti Money Laundering Reporting and Investigation – Sorting the Wheat from the Chaff
Ana Isabel Canhoto, James Backhouse 600

Application of Latent Semantic Indexing to Processing of Noisy Text
Robert J. Price, Anthony E. Zukas 602

Detecting Misuse of Information Retrieval Systems Using Data Mining Techniques
Nazli Goharian, Ling Ma, Chris Meyers 604

Mining Schemas in Semistructured Data Using Fuzzy Decision Trees
Wei Sun, Da-xin Liu 606

More Than a Summary: Stance-Shift Analysis
Boyd Davis, Vivian Lord, Peyton Mason 608

Principal Component Analysis (PCA) for Data Fusion and Navigation of Mobile Robots
Zeng-Guang Hou 610

Information Management and Sharing

BioPortal: Sharing and Analyzing Infectious Disease Information
Daniel Zeng, Hsinchun Chen, Chunju Tseng, Catherine Larson, Wei Chang, Millicent Eidson, Ivan Gotham, Cecil Lynch, Michael Ascher 612

DIANE: Revolutionizing the Way We Collect, Analyze, and Share Information
Jin Zhu 614

Processing High-Speed Intelligence Feeds in Real-Time
Alan Demers, Johannes Gehrke, Mingsheng Hong, Mirek Riedewald 617

Question Answer TARA: A Terrorism Activity Resource Application
Rob Schumaker, Hsinchun Chen 619

Template Based Semantic Similarity for Security Applications
Boanerges Aleman-Meza, Christian Halaschek-Wiener, Satya Sanket Sahoo, Amit Sheth, I. Budak Arpinar 621

The Dark Web Portal Project: Collecting and Analyzing the Presence of Terrorist Groups on the Web
Jialun Qin, Yilu Zhou, Guanpi Lai, Edna Reid, Marc Sageman, Hsinchun Chen …….. 623

Toward an ITS Specific Knowledge Engine
Guanpi Lai, Fei-Yue Wang 625

Information Security and Intrusion Detection

A Blind Image Watermarking Using for Copyright Protection and Tracing
Haifeng Li, Shuxun Wang, Weiwei Song, Quan Wen 627

Towards an Effective Wireless Security Policy for Sensitive Organizations
Michael Manley, Cheri McEntee, Anthony Molet, Joon S. Park 629

A Taxonomy of Cyber Attacks on 3G Networks
Kameswari Kotapati, Peng Liu, Yan Sun, Thomas F. LaPorta 631

An Adaptive Approach to Handle DoS Attack for Web Services
Eul Gyu Im, Yong Ho Song 634

An Architecture for Network Security Using Feedback Control
Ram Dantu, João W. Cangussu 636

Defending a Web Browser Against Spying with Browser Helper Objects
Beomsoo Park, Sungjin Hong, Jaewook Oh, Heejo Lee 638

Dynamic Security Service Negotiation to Ensure Security for Information Sharing on the Internet
ZhengYou Xia, YiChuan Jiang, Jian Wang 640

Enhancing Spatial Database Access Control by Eliminating the Covert Topology Channel
Young-Hwan Oh, Hae-Young Bae 642

Gathering Digital Evidence in Response to Information Security Incidents
Shiuh-Jeng Wang, Cheng-Hsing Yang 644

On the QP Algorithm in Software Watermarking
William Zhu, Clark Thomborson 646

On the Use of Opaque Predicates in Mobile Agent Code Obfuscation
Anirban Majumdar, Clark Thomborson 648

Secure Contents Distribution Using Flash Memory Technology
Yong Ho Song, Eul Gyu Im 650

Infrastructure Protection and Emergency Response

Background Use of Sensitive Information to Aid in Analysis of Non-sensitive Data on Threats and Vulnerabilities
Richard A. Smith 652

Securing Grid-Based Critical Infrastructures
Syed Naqvi, Michel Riguidel 654

The Safety Alliance of Cushing – A Model Example of Cooperation as an Effective Counterterrorism Tool
David Zimmermann 656

Surveillance, Border Protection, and Transportation Systems

A Framework for Global Monitoring and Security Assistance Based on IPv6 and Multimedia Data Mining Techniques
Xiaoyan Gong, Haijun Gao 658

An Agent-Based Framework for a Traffic Security Management System
Shuming Tang, Haijun Gao 660

Application of a Decomposed Support Vector Machine Algorithm in Pedestrian Detection from a Moving Vehicle
Hong Qiao, Fei-Yue Wang, Xianbin Cao 662

Application of Cooperative Co-evolution in Pedestrian Detection Systems
Xianbin Cao, Hong Qiao, Fei-Yue Wang, Xinzheng Zhang 664

Biometric Fingerprints Based Radio Frequency Identification
Sundaram Jayakumar, Chandramohan Senthilkumar 666

BorderSafe: Cross-Jurisdictional Information Sharing, Analysis, and Visualization
Siddharth Kaza, Byron Marshall, Jennifer Xu, Alan G. Wang, Hemanth Gowda, Homa Atabakhsh, Tim Petersen, Chuck Violette, Hsinchun Chen 669

Author Index 671

Collusion Set Detection Through Outlier Discovery*

Vandana P. Janeja, Vijayalakshmi Atluri, Jaideep Vaidya, and Nabil R. Adam

MSIS Department and CIMIC,
Rutgers University
{janeja, atluri, jsvaidya, adam}@cimic.rutgers.edu

Abstract. The ability to identify collusive malicious behavior is critical in today's security environment. We pose the general problem of *Collusion Set Detection* (CSD): identifying sets of behavior that together satisfy some notion of "interesting behavior". For this paper, we focus on a subset of the problem (called CSD′), by restricting our attention only to outliers. In the process of proposing the solution, we make the following novel research contributions: First, we propose a suitable distance metric, called the *collusion distance metric*, and formally prove that it indeed is a distance metric. We propose a *collusion distance based outlier detection* (CDB) algorithm that is capable of identifying the *causal dimensions* (n) responsible for the outlierness, and demonstrate that it improves both precision and recall, when compared to the Euclidean based outlier detection. Second, we propose a solution to the CSD′ problem, which relies on the semantic relationships among the causal dimensions.

1 Introduction

Malfeasants are now increasingly aware of the successful use of information technology against them. Now, lawbreakers (both terrorists as well as organized crime) try to come up with hidden associations that may stay undetected from the law agencies. One common method is to use multiple unconnected people to perform parts of the required activities. Each activity, in itself, is not suspicious. But seen as a whole, it would cause alarm flags to go up. This does not get detected by traditional classification unless the underlying link between the people performing the activities is found.

This is even more complicated when the data collection is distributed. Data from various domains may not be simply integrated due to a number of reasons, primarily due to the privacy and security policies of the different agencies owning the data. Even if the data were shared, merging would tremendously increase the number of dimensions to be mined, leaving many dimensions blank for objects that do not have a representation in the other domain. As a result, many traditional data mining techniques cannot be adopted as they do not scale

* This work is supported in part by the National Science Foundation under grant IIS-0306838.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 1–13, 2005.

well. Another reason where integration does not prove beneficial is that mining across multiple domains will need to incorporate some level of domain knowledge such that significance can be attached to the results discovered. Semantic knowledge can and, indeed, must be incorporated in some way to help reveal possible hidden associations. Thus, the general problem we pose is the following: how to detect independent sets of behavior, that taken together would be classified as "interesting behavior" ("interesting behavior" is application dependent). We designate our problem as *Collusion Set Detection* (CSD). Following is a *real* motivating example:

Example 1. Importers shipping into the United States fill out a manifest form that the U.S Customs currently gathers in the Automated manifest system (AMS). Typically, the AMS data for a shipment consists of about 104 attributes, and the profile of each company involved in the shipment consists of about 40 attributes. This data may be gathered by multiple entities, including different ports such as the port of entry, port of departure, point of arrival, or different agencies such as the U.S. Department of transportation (DOT), U.S. Coast Guard, U.S. Customs and Border Protection (CBP), etc. U.S.Customs is currently undergoing modernization of the current systems to produce an integrated trade and border protection system, Automated commercial environment/The International Trade Data System (ACE/ITDS). As a result, additional agencies such as Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF), CIA, FBI would be tapped for validating shipments in terms of their compliances. Each of these entities collecting data can be treated as a different domain. Let us consider a simple scenario comprising of the following two shipments:

Shipment 1: *A shipment to MerPro, Inc. entering through the eastern border via Newark Port comprising of Liquid Urea, which is used as a chemical fertilizer. This shipment will be recorded in the Newark port database (say domain D_1). Assume Liquid Urea is recorded as one of its dimensions (say d_{11}).*

Shipment 2: *A shipment to a consignee, ChemCo, carrying Ammonium Nitrate, whose destination is Phoenix, Arizona, which is entering through a port in the west coast, Los Angeles. This shipment will be recorded in the Los Angeles port database (say domain D_2) with Ammonium Nitrate as one of its dimensions (say d_{21}).*

While these two shipments are considered in isolation within their own domains (i.e., Newark port data and Los Angeles port data) they appear to be benign. However, upon a closer observation of these two shipments, one may discover that the two materials shipped into the country can potentially be used to make an explosive (which in fact can be determined by a chemical expert based on the reactivity of chemicals [11]). Moreover, the fact that the destination of these two shipments are geographically close, it increases the likelihood of such a thing materializing. Such results may trigger a further analysis on the companies' behavior, which may lead to the discovery (by a financial analyst) of an unusual financial transaction with respect to the amount of funds transferred.

This could be a third domain (say D_3) and the "amount of funds transferred" could be one of its dimensions (say d_{31}). In essence, our goal is to identify such potential "collusions" among the entities responsible for these two shipments (the CSD problem). Discovering such information can be valuable to a terrorist task force agent. □

Formally, we define the CSD problem in terms of classification. Given a collection of records, the goal of classification is to derive a model that can assign a record to a class as accurately as possible. Given a model, the goal of CSD is to identify sets of records whose behavior, taken together will satisfy the model. For this paper, instead of the complete general problem (CSD), we focus on an interesting subset of the problem (CSD′) – by restricting our search to outlier objects. Thus, the problem we solve in this paper is to detect sets of outlier objects whose behavior when taken together satisfies some notion of "interesting behavior". Instead of considering all the shipments as in example 1, we consider only the outliers in identifying the collusion sets (the CSD′ problem). We assume that the semantic relationships among the different dimensions are known in advance (typically specified by domain experts), and propose a data mining approach to discover potential collusion sets among the outliers.

Outlier detection has been extensively studied in the statistical community [3]. Knorr and Ng [8] have proposed a distance-based outlier detection approach that is both simple and intuitive, with several extensions proposed [9, 12]. The concept of local outlier is explored in [4, 5]. To address the problem of dimensionality curse, Aggrawal and Yu [1] propose an approach that considers projections of the data and considers the outlier detection problem in subspace. He et al. [6] propose a semantics based outlier detection. However, none of these approaches identify the causal dimensions or other causal knowledge of the outliers, which is essential in addressing the problem being tackled in this paper. Xu et al. [16] focus on detecting and specifying changes in criminal organizations. Wang et al. [14] address identification of record linkages. Snowball method [15] works well in sampling ties from people, however this type of information cannot be gleaned from large data sets. Therefore the results obtained through these approaches can be complimentary to those obtained using data mining approaches. Kubica et al. [10] propose a graph-based approach for link analysis and collaboration queries. However, this approach considers co-occurrences and recent links, which may not provide the most intuitive knowledge. Moreover, the links may originate from multiple data sources and the weighting functions would need to incorporate such knowledge dynamically.

In this paper, section 2 provides an overview of the approach. Section 3 presents the new distance metric that we propose and the approach to identify outliers and causal dimensions. Section 4 describes the CSD′ approach in detail. Section 5 presents our experimental results. Finally, section 6 concludes the paper and provides an insight into ongoing and future research.

2 Basic Approach

In order to solve the problem of CSD′, we take a two step process:

1. *identification of outlier objects* from the different datasets under consideration and determining the *causal dimension* contributing to the outlierness.
2. *identification of collusion sets* by appropriately linking the outlier objects based on the *semantic relationships* among their causal dimensions.

To detect outliers from high dimensional datasets, we adopt a distance based outlier detection approach, similar to that proposed by Knorr and Ng [8]. However, since metrics used in the existing distance based methods do not lend themselves to identify the dimensions causing the outlierness, we propose a new distance metric, called the *collusion distance metric*. We formally prove that this indeed is a distance metric and demonstrate that our *collusion distance based outlier detection* (CDB) approach consistently performs better than the traditional Euclidean distance based outlier detection (EDB). We also experimentally show that the causal dimensions identified using our approach are indeed the correct causal dimensions.

In order to detect the collusion sets, we exploit the *semantic relationships* among the causal dimensions. These relationships are either *intra-domain*, in which case are specified by an expert in that domain, or *inter-domain*, in which case are specified by an expert that has the knowledge across the domains. For example, the semantic relationship among dimensions d_{11} and d_{12} in example 1 can be specified based on the knowledge drawn from a chemist, and the inter-domain semantic relationship between d_{12} and d_{31} can be specified based on the expertise of a terrorist task force agent. A network of outliers is then constructed by conservatively applying the transitivity among the semantic relationships. The network essentially results in a forest, where each graph comprises of outliers as nodes and undirected weighted edges. Cliques identified in the graphs indicate potential collusion sets. We show that our approach significantly reduces the cost of computation to 2^n from 2^N, where n is the number of causal dimensions and N is the total number of dimensions, and $n \ll N$. The following two sections present these two steps of our approach.

3 Identification of Causal Outlier Dimensions

An *outlier* is a point, which varies sufficiently from other points such that it appears to be generated by a different process from the one governing the other points. We begin by identifying outliers and the dimension causing the outlying behavior. We adopt the definition of the outlier from [8].

Definition 1 (Outlier[8]): *An object O in a dataset T is a distance based, $DB(p, D)$ outlier if at least a fraction p of the objects in T are at a greater distance than D from O.*

Here distance based outliers rely on the computation of distance values based on a *distance metric*. Euclidean distance metric is the most commonly used, which can be measured as follows.

Definition 2 (Euclidean Distance): *Given two N dimensional objects $X = (x_1, x_2 \ldots x_N)$ and $Y = (y_1, y_2 \ldots y_N)$, the Euclidean distance $E(X, Y) = \sqrt{((x_1 - y_1)^2 + (x_2 - y_2)^2 \ldots (x_N - y_N)^2)}$.*

Although a number of distance-based outlier detection approaches have employed Euclidean as a distance metric, it can mitigate the outlierness, specifically in high dimensional datasets. In this paper, we propose a new distance metric, called the *collusion distance metric*, which is motivated by the Hausdorff distance metric [13]. The need for a new distance metric is driven by the following two primary reasons. First, in high dimensional data sets, the collusion distance performs better in identifying outliers when compared to Euclidean as a distance metric; we demonstrate this empirically as indicated by the results in section 5. Second, to identify collusion sets, we need to first identify the dimensions causing the outlierness for each outlier. In the following, we first define our distance metric and then prove that the function we propose is, indeed, a metric.

Definition 3 (Collusion Distance): *Given two N dimensional objects $X = (x_1, \ldots, x_N)$ and $Y = (y_1, \ldots, y_N)$, the Collusion Distance $CD(X, Y) = \max_i |x_i - y_i|$. Along with the distance, the dimension i is identified as the causal dimension.*

Essentially, given two points X and Y, CD computes the maximum of the distances between each of the dimensions of X and Y. In doing so, it also identifies in which dimension(s) this maximum difference is occurring. Thus along with the distance, the dimension i is identified as the causal dimension. These causal dimensions will later (in section 4) be used in the identification of collusion sets.

Theorem 1. *$CD(X, Y)$ is a distance metric.*

Proof. A *non-negative* function $g(x, y)$ is a metric if it satisfies:

- the identity property, $g(x, x) = 0$,
- the symmetry property, $g(x, y) = g(y, x)$, and
- the triangle inequality, $g(x, y) + g(y, z) \geq g(x, z)$

CD is clearly a nonnegative function. $CD(X, X) = \max_i |x_i - x_i| = \max_i 0, \ldots, 0 = 0$. Therefore, CD satisfies the identity property. Since, we take the maximum of absolute differences, CD also satisfies the symmetry property. We now prove that CD satisfies the triangle inequality as well. Assume that,

- $CD(X, Y) = \max_i |x_i - y_i| = |x_k - y_k|$
- $CD(Y, Z) = \max_i |y_i - z_i| = |y_l - z_l|$
- $CD(X, Z) = \max_i |x_i - z_i| = |x_m - z_m|$

Now,

$$CD(X,Z) = |x_m - z_m| = |x_m - y_m + y_m - z_m|$$
$$\leq |x_m - y_m| + |y_m - z_m|$$

Since $|x_k - y_k| = \max_i |x_i - y_i|$, therefore, $|x_k - y_k| \geq |x_m - y_m|$. Similarly, since $|y_l - z_l| = \max_i |y_i - z_i|$, therefore, $|y_l - z_l| \geq |y_m - z_m|$. Putting it all together,

$$CD(X,Z) \leq |x_k - y_k| + |y_l - z_l|$$
$$\leq CD(X,Y) + CD(Y,Z)$$

Thus, CD also satisfies the triangle inequality and is a distance metric. □

In order to accurately measure the collusion distance, we first normalize the data using the min-max normalization, which preserves all original relationships and does not introduce any bias in the data. This is defined as follows: $C' = (C - min) * (new\_max - new\_min)/(max - min) + new\_min$, where C is the existing value and C' is the normalized value, max and min are the maximum and minimum values of each dimension, new_min is set to 0 and new_max is set to 1. Unlike traditional distance based outlier detection methods that use a single value as a threshold, one needs to use a *threshold vector* in collusion distance based outlier detection. This is because, each dimension is accounted for in collusion distance. Therefore, we need to set a threshold for each dimension such that the object is marked as an outlier due to the extreme value in the specific dimension. We compute the threshold vector v for the data set T based on standard deviation. Let st_d be the standard deviation for each dimension d. Given a set of dimensions $\{d_1, \ldots d_n\}$, a threshold vector $v = \{c*st_{d_1}, \ldots c*st_{d_n}\}$, where c is a constant. (In our experimental results, we consider different threshold vectors with c ranging from .25 to 3.) We define the *collusion distance metric* based outlier as follows:

Definition 4 (Outlier) : *An object O in a dataset T is $CDB(p, D)$ outlier if at least a fraction p of the objects in T are at a greater distance than $dist_i$ from O where i is the dimension and $dist_i = c * st_{d_i}$ in the threshold vector $v = \{c * st_{d_1}, \ldots c * st_{d_n}\}$, such that c is a constant.*

Algorithm 1 presents the steps to identify outliers and the dimensions contributing to the outlierness. Essentially, given a domain D_i such that the dimensions in $D_i = \{d_{i1} \ldots d_{in}\}$, an outlier o_j from domain D_i is first identified and then is associated with a set of causal dimensions $cd(o_j) \subseteq \{d_{i1} \ldots d_{in}\}$. In this paper, we identify one casual dimension.

This algorithm 1 is based on the Nested Loop (NL) algorithm outlined in [8] except that the distance metric is replaced by the collusion distance. The parameters to the algorithm are the standard parameters provided in the NL algorithm. If any object is at a greater distance than the threshold in any dimension from more than $p\%$ (5%) of points then it is marked as an outlier and the causal dimension causing the maximum distance is noted as the causal dimension for this outlier. Thus algorithm 1 returns the distance along with the causal dimension.

Algorithm 1. Outliers and causal dimension detection in high dimensional datasets

Require: k partitions, $S_1, \ldots, S_k$
  for each partition S_l **do**
    for all objects $i \in s_l$ **do**
      for all objects $j \in s_l, j \neq i$ **do**
        {Invoke the Collusion Computation Function}
        Compute $Dist, CD \leftarrow cd(i, j)$ {$DistCD$ is the actual Collusion distance, cd is the Causal Dimension, which is the dimension with the maximum difference between the two objects}
        Increment $count_{cd}$ if i is an outlier wrt j in the causal dimension
      end for
      Compute $globalcount = \sum_{CausalDim_r} count_r$
      if $globalcount > p\%of|s_l|$ **then**
        Mark i, cd as an outlier and its causal dimension
      end if
    end for
  end for

Algorithm 2. Identification of collusion sets

Require: s, domains $M_1, \ldots, M_s$
  G, the set of local graphs built from intra-domain semantic tuples, $G_1, \ldots, G_{|G|}$ Y, the set of inter-domain semantic tuples, $y_1, \ldots, y_{|Y|}$ n outliers $O_1, \ldots, O_n$, with annotated causal dimension $cd(O_i)$ (Output of Algorithm 1)
  {Prune inter-domain tuples based on relevance}
  for each inter-domain tuple y_i **do**
    if any dimension of tuple y_1 belongs to $\bigcup_i cd(O_i)$ **then**
      $actionable\_tuples \leftarrow actionable\_tuples \bigcup y_i$
    end if
  end for

  Generate global semantic graphs by applying transitivity
  Sort outliers into groups on the basis of causal dimensions
  for each possible combination c of causal dimensions in $\bigcup_i CD(O_i)$ **do**
    if c exists in any global graph **then**
      Generate outlier collusion sets (every possible combination of outliers from different groups)
    end if
  end for

4 Identification of Collusion Sets

In this section, we present our approach to solve the CSD′ problem by utilizing the outliers and causal dimensions detected in section 3 and exploiting the semantic relationships among these causal dimensions. We describe *collusion sets*, for which we first describe *semantic relationships*. Each semantic relationship is comprised of a set of dimensions that have semantic associations, where the weight indicates the significance of the relationship.

Semantic Relationship: A domain contains a set of semantically meaningful dimensions. Thus given a domain D_i such that the dimensions in $D_i = \{d_{i1} \ldots d_{in}\}$. The *semantic relationships* among the dimensions, both intra- and inter-domain are specified as follows: Given a set of dimensions $M = \{d_1, d_2 \ldots d_n\}$, a semantic relationship $r = M_r, w_r$, where $M_r \subseteq M$ such that $|M_r| \geq 2$, and w_r is the weight of r. In general, semantic relationships may exist among any type of objects, however in this paper, we limit it to dimensions.

Collusion Set: A collusion set $C=\{C_1, \ldots, C_t\}$, where each collusion $C_i \in C$ is nothing but a set of objects $O=\{O_1, \ldots, O_u\}$ such that each $O_j \in O$ is an outlier object belonging to a causal dimension. The number of objects in a collusion is obviously ≥ 2. These collusions are identified based on the explicitly specified semantic relationships as well as those inferred.

Our approach to generating collusion sets begins with the construction of *static semantic graphs*. These graphs are constructed statically in advance, by considering only the intra-domain semantic relationships. This construction can be done offline, immediately once the intra-domain relationships have been identified by the domain expert. Causal dimensions are identified by the CDB outlier detection process. These are used to prune the static graphs to include only relevant links. Now, the inter-domain semantic links (for dimensions) are used to appropriately link the graphs. Transitivity is conservatively applied across only the intra-domain links to generate the complete graphs. A forest may also result depending on the linkages occurring. Finally, using this, we identify potential collusion sets among outliers. In the following, we explain these steps in more detail. Algorithm 2 outlines these steps. We also run through a simple example to clarify the presentation of the algorithm.

1. Generation of the intra-domain static semantic graphs: We assume the set of domains $D = (D_1, \ldots, D_s)$, the size of the dataset is $|T| = \bigcup_s |T_s|$ and each domain D_i has N_i dimensions. The total number of dimensions, across all domains is $N = \sum_{i=1}^{s} N_i$. For each domain D_i, we consider each intra-domain semantic relationship $r_j : (d_{ik}, d_{il}, w_j)$, and include an undirected edge d_{ik}, d_{il} in the static semantic graph of this domain, if it is not included as an edge d_{il}, d_{ik}. The weight is also attached to this edge. For example, consider a simple example in figure 1(a) comprising of three domains D_1, D_2 and D_3 with semantic relationships r_1 to r_7 being the intra-domain of the domains indicated, and r_8 to r_{10} being the inter-domain relationships. These graphs may be sorted based on the dimensions.

2. Pruning of the intra-domain semantic graphs based on outlier causal dimensions: The pruning is performed by essentially eliminating any graph if it does not contain any node that is detected as a causal dimension. This step is to improve the efficiency of the latter process. In figure 1(c), assume the set of causal dimensions is $\{d_{11}, d_{12}, d_{13}, d_{14}, d_{21}\}$. Based on this set, the static semantic graphs $(d_{31}, d_{32}, 0.8; d_{31}, d_{33}, 0.7)$ of D_3 and $(d_{22}, d_{23}, 0.7)$ of D_2 can be eliminated since they do not contain any causal dimension.

3. Pruning of intra-domain relationships: Similar to the above pruning process, the intra-domain relationships that do not contain any causal dimension are also eliminated. For example, $r_{10} : (d_{22}, d_{31}, 0.5)$ has been removed as shown in figure 1(c).

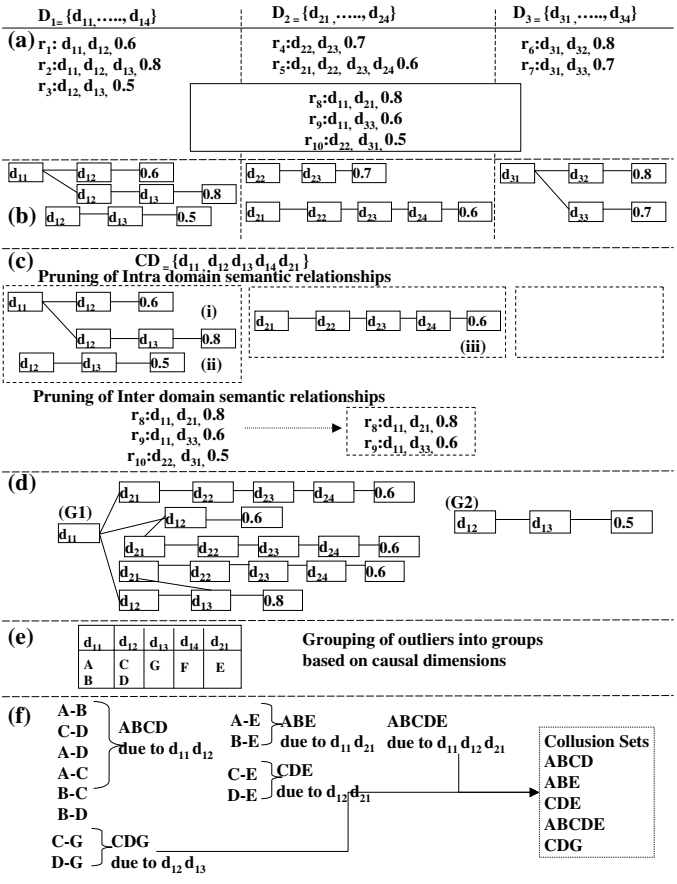


Fig. 1. An example depicting the collusion set detection steps

4. Extension of the pruned graphs by applying transitivity using inter-domain semantic relationships: We now connect the different inter-domain semantic graphs using the intra-domain relationships. At this point, we also consider the transitivity among the edges. Two graphs G_1 and G_2 will be generated as shown in figure 1(d).

5. Grouping of outliers: The outliers are grouped based on their causal dimensions and these groups are sorted based again on the causal dimensions, as shown in figure 1(e).

6. Identification of collusion sets: Based on the semantic relationships of the causal dimensions the non relevant collusion sets are eliminated resulting in the interesting collusion sets. Now, for every interesting collusion set of dimensions, we can use the outlier list for those dimensions to generate all possible combinations of the outliers identified. All of these are added into the final

collusion set output. Thus, in our example in figure 1(f), from all possible collusions, the relevant collusion sets are $(AB, AC, BC, CD, ABC, ABD, ABCD)$ due to relationship between the causal dimensions of the outliers namely d_{11} and d_{12}, (AB, AE, ABE) and so on. Thus, the potential collusion sets identified are $(ABCD, ABE, CDE, ABCDE, AGE)$. This step might seem computationally inefficient. Unfortunately, detection of a collusion set does not guarantee that all its subsets are also collusion sets. Thus, an APRIORI [2] kind of frequent itemset mining approach cannot be applied.

Analysis: In the following, we discuss the cost savings that can be potentially achieved with our CSD′ approach. Let the size of the global dataset be $|T|$ and the total number of dimensions across all domains be N. The dominant cost of the algorithm is in identifying the collusion sets. In the worst case, the cost of identifying the collusion sets using the CSD′ approach is $O(2^N)$, if all the dimensions are detected as causal dimensions.

Recall from section 3, our CDB outlier detection approach identifies outliers and the causal dimensions, which are the data objects and the dimensions to be considered in our CSD′ approach, respectively. Let the size of the outlier objects be $|t|$ and size of the number of causal dimensions be n. Due to obvious reasons, it is always true that $n \leq N$ and $|t| < |T|$, and more frequently, $n \ll N$ and $|t| \ll |T|$. As a result, the cost of identifying collusion sets using the CSD′ is reduced to 2^n. This is a reasonable assumption because the number of outliers ($|t|$) and therefore the number of causal dimensions (n) is bounded by $p\%$ of $|T|$, and moreover, in many cases $N < |T|$.

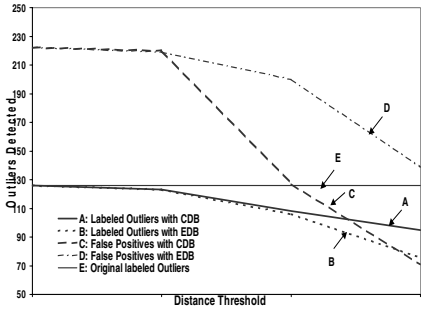
5 Experimental Results

We have examined $CDB(p, D)$ outlier detection using collusion distance metric on various datasets available from: `www.ics.uci.edu/~mlearn/MLSummary.html`, However, due to space limitation we only discuss some sample results here. Complete results can be found at `http://cimic.rutgers.edu/~vandana/CDBResults.pdf`. We discuss the results obtained with and Ionosphere dataset with 351 cases, 35 dimensions and 1 domain. For multi domain data we discuss Simulated Data with 50-130 cases, 19-20 dimensions and 3 domains. The following two observations can be made from the results in this section:

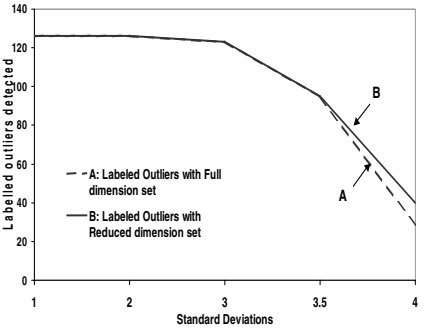
1. CDB shows an improvement over Euclidean distance based (EDB) outlier detection.
2. The causal dimensions identified using collusion distance are indeed the correct causal dimensions.

Results of CDB: To demonstrate the first observation, for each data set, we show a comparison of the results from the EDB and those from the CDB outlier detection. The latter is based on algorithm 1. We essentially show the recall (number of labeled outliers detected) and precision (the number of false positives) for both EDB and CDB, for varying threshold values. (Note that the

threshold values are different for EDB and CDB). Specifically, in case of EDB, it is a single distance measure, whereas in case of CDB, it is a variation of c in the threshold vector.



(a) Performance EDB and CDB



(b) Full vs. reduced dimension

Among the traditionally labeled data sets ionosphere data shows that CDB fares better when compared to EDB in both precision and recall (figure 2(a)). In the simulated dataset there is a significant improvement in both precision and recall with 100% of the labeled outliers being detected. Further, some unlabeled outliers are introduced randomly in the simulated datasets. The CDB detects these outliers along with the correct causal dimensions.

To demonstrate that the causal dimensions identified by the CDB are indeed the correct dimensions causing outlierness, we first take the top q of n dimensions with the most outliers detected in them. Then we demonstrate that the CDB outlier detection by considering only these q dimensions, is same or better than the case where all dimensions are considered. Since collusion distance uses extreme distances between the dimensions for identifying outliers, if the non outlier causing dimensions are eliminated, the results of the outlier detection should not get affected. For the ionosphere dataset, we reduce the number of dimensions to 20 (it was initially 34), based on the result of collusion distance based outlier detection.

As can be seen in figure 2(b), the reduced dimension dataset initially produces similar number of labeled outliers and for some thresholds more labeled outliers are detected with reduced dimensions as compared to the full dimension set.

Results of CSD′: In the following, we present the results of our CSD′ approach for the simulated data comprising of three domains, where $D_1 = \{d_{11} \ldots d_{121}\}$, $D_2 = \{d_{21} \ldots d_{220}\}$, and $D_3 = \{d_{31} \ldots d_{119}\}$. We have defined 18 intra-domain and 4 inter-domain semantic relationships. Overall, these semantic relationships are across 38 dimensions in all three domains. We have introduced outliers in 18 dimensions appearing in the semantic relationships. Thus, 20 dimensions are not appearing in the semantic relationships. Overall, we are able to identify all

the outliers that are inserted randomly in other dimensions as well. If an object is an outlier in more than one dimensions, all its relevant collusions are also detected. Note that these results validate that our approach indeed identifies all collusion sets. We are currently working with domain experts (in the Department of Border Protection, which is part of US Customs and the Terrorist Task Force) to identify the outliers and Collusion sets within the PIERS[7] data.

6 Conclusions and Future Research

In this paper, we have identified the general problem of Collusion Set Detection (CSD) as detecting sets of behavior that together satisfy some notion of "interesting behavior". Specifically, we have proposed a solution to an interesting subset of the problem (called CSD′) and restrict our attention only to outliers. We have proposed a novel technique using semantics to identify collusive behavior among outliers. In particular, we have made the following novel research contributions: First, we have proposed a new distance metric, called the *collusion distance metric*, when used in an outlier detection algorithm is capable of identifying the *causal dimensions* responsible for the outlierness. Second, we have demonstrated that our outlier detection algorithm based on the new distance metric improves both precision and recall when compared to the Euclidean distance based outlier detection. We proposed an algorithm for solving the CSD′ problem, which relies on the semantic relationships among the causal dimensions and significantly reduces the cost of identifying the collusion sets.

We have performed experiments with real life and simulated datasets. We propose to extend the experiments to social network and link discovery datasets. While in this paper, we offer a solution to the CSD′ problem, as part of our future research, we focus on addressing the CSD problem itself. Moreover, under our approach, we assume that the semantic relationships among data are specified by domain experts. We intend to investigate data mining techniques to extract such relationships from the data itself.

References

1. Charu C. Aggarwal and Philip S. Yu. Outlier detection for high dimensional data. In *Proceedings of the ACM SIGMOD*, pages 37–46, 2001.
2. Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases*, pages 487–499, Santiago, Chile, September 12-15 1994.
3. Vic Barnett and Toby Lewis. *Outliers in Statistical Data*. John Wiley and Sons, 3rd edition, 1994.
4. Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and J. Sander. Optics-of: Identifying local outliers. In *Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery*, pages 262–270, 1999.
5. Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the ACM SIGMOD*, 2000.

6. Zengyou He, Shengchun Deng, and Xiaofei Xu. Outlier detection integrating semantic knowledge. In *Proceedings of the International Conference on Advances in Web-Age Information Management*, pages 126–131. Springer-Verlag, 2002.
7. http://www.piers.com/default2.asp. Piers global intelligence solutions.
8. Edwin M. Knorr and Raymond T. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proceedings of the International Conference on Very Large Data Bases (VLDB 1998)*, pages 392–403, August 1998.
9. Edwin M. Knorr and Raymond T. Ng. Finding intensional knowledge of distance-based outliers. In *Proceedings of 25th International Conference on Very Large Data Bases*, pages 211–222, 1999.
10. J. Kubica, A. Moore, D. Cohn, and J. Schneider. Finding underlying connections: A fast graph-based method for link analysis and collaboration queries. In *Proceedings of the International Conference on Machine Learning*, August 2003.
11. M.F. Lopez, A. Gomez-Perez, J.P. Sierra, and A.P. Sierra. Building a chemical ontology using methontology and the ontology design environment. *"Intelligent Systems"*, 14:37–46, 1999.
12. Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *Proceedings of the ACM SIGMOD*, pages 427–438, 2000.
13. G. Rote. Computing the minimum hausdorff distance between two point sets on a line under translation. *Inf. Process. Lett.*, 38(3):123–127, 1991.
14. Gang Wang, Hsinchun Chen, and Homa Atabakhsh. Automatically detecting deceptive criminal identities. *Commun. ACM*, 47(3):70–76, 2004.
15. S. Wasserman and K. Faust. *Social network analysis.* Cambridge University Press, 1994.
16. Jennifer Xu and Hsinchun Chen. Untangling criminal networks: A case study. In *ISI*, pages 232–248, 2003.

Digging in the Details: A Case Study in Network Data Mining

John Galloway [1,2] and Simeon J. Simoff [3,4]

[1] Complex Systems Research Centre, University of Technology Sydney,
PO Box 123 Broadway NSW 2007 Australia
john.galloway@uts.edu.au
[2] Chief Scientist, NetMap Analytics Pty Ltd,
52 Atchison Street, St Leonards NSW 2065 Australia
[3] Faculty of Information Technology, University of Technology Sydney,
PO Box 123 Broadway NSW 2007 Australia
simeon@it.uts.edu.au
[4] Electronic Markets Group, Institute for Information and Communication Technologies,
University of Technology Sydney - PO Box 123 Broadway NSW 2007 Australia
http://research.it.uts.edu.au/emarkets

Abstract. Network Data Mining builds network linkages (network models) between myriads of individual data items and utilizes special algorithms that aid visualization of 'emergent' patterns and trends in the linkage. It complements conventional and statistically based data mining methods. Statistical approaches typically flag, alert or alarm instances or events that could represent anomalous behavior or irregularities because of a match with pre-defined patterns or rules. They serve as 'exception detection' methods where the rules or definitions of what might constitute an exception are able to be known and specified ahead of time. Many problems are suited to this approach. Many problems however, especially those of a more complex nature, are not well suited. The rules or definitions simply cannot be specified; there are no known suspicious transactions. This paper presents a human-centered network data mining methodology. A case study from the area of security illustrates the application of the methodology and corresponding data mining techniques. The paper argues that for many problems, a 'discovery' phase in the investigative process based on visualization and human cognition is a logical precedent to, and complement of, more automated 'exception detection' phases.

1 Introduction

The proliferation of data is both an opportunity and a challenge. It provides the details that governments need to find criminals and terrorists, that organizations need to fight fraud and that businesses need to solve problems and gain market advantage. At the same time, a large volume of data with different storage systems, multiple formats and all manner of internal complexity can often hide more than it reveals.

Data mining – "the process or secondary analysis of large databases aimed at finding unsuspected relationships that are of interest or value to the database owners" (Klösgen and Zytkow 2002) (p. 637), emerged as an "eclectic discipline" (Klösgen

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 14 – 26, 2005.

and Zytkow 2002) that addresses these large volumes of data. Although the data mining researchers have developed methods and techniques that support a variety of tasks, the main interest of analytics practitioners has been focused on predictive modeling. The dominant scenario in predictive modeling is the "black box" approach, where we have a collection of inputs and one or more outputs, and we try to build an algorithmic model that estimates the value of the outputs as a function of the values of the inputs. The key measure of the quality of the model is the accuracy of predictions, rather than the theory that may explain the phenomena.

In practice, the focus on predictive accuracy in the "black box" approach (inherited from regression analysis and automatic control systems) makes perfect sense. For example, the more accurate is a classifier of tumors based on the characteristics of mammogram data, the better the aid it provides to young practitioners (Antonie, Zaiane et al. 2003). In business areas like marketing, such an approach makes sense (for example, determining which few books Amazon should also offer to someone who searches a particular book). Numerous examples from different areas of human endeavor, including science engineering are presented in (Nong 2003). However, data mining applications in many areas, including businesses and security (Nong 2003) require deeper understanding of the phenomena. Such complex phenomena can be modeled with techniques from the area of complex systems analysis. Social network analysis (Wasserman and Faust 1994; Scott 2000) deals with this type of data analysis, looking at data sets that describe explicitly some relationships between the individual entities in them.

In addition to the explicitly coded relationships, there could be *implicit* relationships between the entities described by the data set. Any attribute can come into play for establishing such relations depending on the point of investigation. Such uncovering of implicit relationships between entities is the focus of network data mining methods.

These different perspectives infer different sets of models. The structure of the relationships between the individual entities is revealed by the *network models*. Such models, which include the topology of the network and the characteristics of its nodes, links, and clusters of nodes and links, attempt to explain observed phenomena through the interactions of individual nodes or node clusters. During recent years there has been an increasing interest in this type of model in a number of research communities (see (Albert and Barabási 2002; Newman 2003) for extensive surveys of the research work on network models in different areas). Historically, sociologists have been exploring the social networks among humans in different social settings. Typical social network research scenarios involve data collection through questionnaires or tables, where individuals describe their interactions with other individuals in the same social setting (for example, a club, school, organization, across organizations, etc.). Collected data is used then to infer a social network model which nodes represent individuals and edges represent the interactions between these individuals. Classical social network studies deal with relatively small data sets and look at "positions" of individuals in the network, measured by centrality (for example, which individuals are best connected to others, which ones are more influential, or which ones are in critical positions) and connectivity (paths between individuals or clusters of individuals through the network). The body of literature is well covered by the following complementary books (Wasserman and Faust 1994; Scott 2000).

Works looking at the *discovery of network models* beyond the "classical" social network analysis date back to the early 1990s (for example, the discovery of shared interests based on the history of email communications (Schwartz and Wood 1993)). Recent years have witnessed the substantial input of physicists (Albert and Barabási 2002), mathematicians (Newman 2003) and organizational scientists (Borgatti 2003) to the network research, with the focus shifting to large scale networks, their statistical properties and explanatory power. The areas include a wide spectrum of social, biological, information, technological and other heterogeneous networks (see Table II in (Newman 2003) and Table I in (Albert and Barabási 2002), the works by Valdis Krebs[1] and Vladimir Batagelj[2]). Recent works in the data mining community are looking at network models for predictive tasks (for example, predicting links in the model (Liben-Nowell and Kleinberg 2003), or the spread of influence through a network model (Domingos and Richardson 2001; Richardson and Domingos 2002; Kempe, Kleinberg et al. 2003). The interest towards link analysis and network models has increased during recent years, evidenced by the number of workshops devoted to the topic (for example, see the presentations at recent workshops on link analysis at ACM KDD conference series[3] and SIAM Data Mining Conference[4]).

Such interest towards network models and new approaches to derive them have been driven largely by the availability of computing power and communication networks that allow one to collect and analyze large amounts of data. Early social network research investigated networks of tens, at most hundreds of nodes. The networks that are investigated in different studies in present days may include millions (and more) nodes. This change of scale of the network models required change in the analytics approach (Newman 2003). Network data mining addresses this challenge.

We define *network data mining as the process of discovering network patterns and models in large and complex data sets*. The term denotes the methods and techniques and can be used in the following contexts:

- mining network models out of data sets
- mining network data (i.e. data generated by the interaction of the entities in a network, for example, in communications networks that can be the network traffic data).

We now briefly discuss the "loss of detail" problem and the "independency of attributes" assumption in knowledge discovery.

The "loss of detail" problem in data mining. Most data mining and analysis tools work by statistically summarizing and homogenizing data (Fayyad, Piatetsky-Shapiro et al. 1996; Han and Kamber 2001; Nong 2003), observing the trends and then looking for exceptions to normal behavior. In addition, as pointed in (Fayyad 2003) "data mining algorithms are "knowledge-free", meaning they are brittle and in real applications lack even the very basic "common sense reasoning" needed to recover even from simple situations. This process results in a *loss of detail* which, for

[1] http://www.orgnet.com
[2] http://vlado.fmf.uni-lj.si/
[3] http://www-2.cs.cmu.edu/~dunja/LinkKDD2004/
[4] http://www-users.cs.umn.edu/~aleks/sdm04w/

intelligence and detection work, can defeat the whole purpose as it is often in the detail where the most valuable information is hidden. More generally, the identifying of exceptions to the 'norm' requires a top down-approach in which a series of correct assumptions needs to be made about what is normal and abnormal. For many complex problems it can be difficult even to start this process since it is impossible to be specific about normal behavior and what might constitute an exception.

The "independency of attributes" assumption in data mining is accepted in some data mining techniques (for example, classifiers building algorithms like Naïve Bayes (Dunham 2002)). Under this assumption, the distributions of the values of the attributes are independent of each other. Unfortunately, real-world data rarely satisfies two-tier independency assumption: the independence of the 'input' attributes and the attribute value independence assumption. In fact some data mining techniques like association and correlation analysis (Han and Kamber 2001), techniques that look at causality and the discovery causal networks (for example, Bayesian network models (Ramoni and Sebastiani 2003)) make exactly the opposite assumption. Moreover, there are situations where the assumption sounds counterintuitive as well (e.g., it is natural for the salary value to be correlated with the values of the age in a sample).

The logic is clear: by missing detail or making the wrong assumptions or simply by being unable to define what is normal, an organization may fail to discover critical information buried in its data.

Further in the paper we present a human-centered knowledge discovery methodology that addresses these issues and present a case study that illustrates the solutions that the network data mining approach offers.

2 Network Data Mining – The Methodology

Where there are few leads and only an open-ended specification as to how to proceed with an analysis, *discovery* is all important. Network data mining is concerned with discovering relationships and patterns in linked data, i.e. the inter-dependencies between data items at the lowest elemental level. These patterns can be revealing in and of themselves, whereas statistically summarized data patterns are informative in different but complementary ways.

Similar to visual data mining (Wong 1999) network data mining integrates the exploration and pattern spotting abilities of the human mind with the processing power of computers to form a powerful knowledge discovery environment meant to capitalize on the best of both worlds. This human-centered approach creates aa extra dimension of value. However, to realizethe most complete solution the discovery phase needs to be repeated at regular intervals so that new irregularities that arise and variations on old patterns can be identified and fed into the *exception detection* phase. The overall network data mining process is illustrated in **Fig. 1**. In a network data mining scenario, the data miner is in a role similar to Donald Schön's "reflective practitioner" (Schön 1991), originally developed in the analysis of design processes. What is relevant to our claim is Schön's view that designers put things together (in our case, the miner replaces the designer, and the things s/he put together are the visual pieces of information) and create new things (in our case the miner creates new

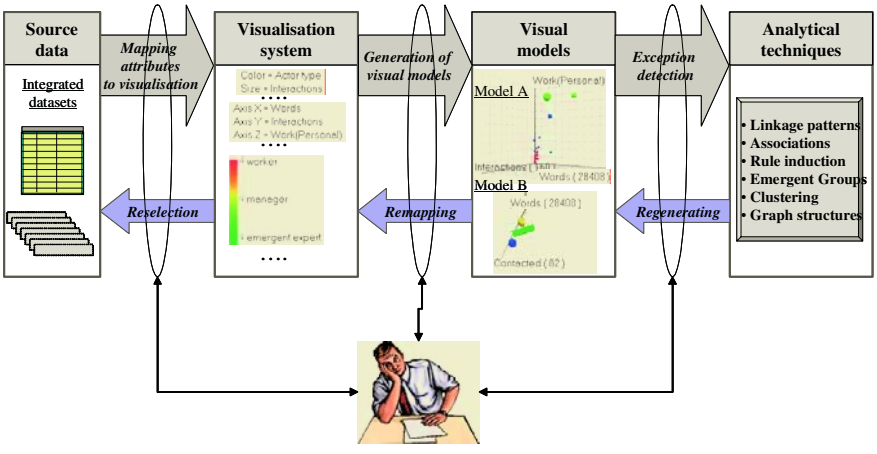


Fig. 1. Network data mining as a human-centered knowledge discovery process

chain of inquiries interacting with the views) in a process involving large numbers of variables (in our case these are the attributes and the linkages between them).

Almost anything a designer does involves consequences that far exceed those expected. In a network data mining approach, the inquiry techniques may lead to results that far exceed those expected and that in most cases may change the line of the analysis of the data into an emerging path. Design process is a process which has no unique concrete solution. In network data mining we operate with numerous network slices of the data set, assisting in revealing the different aspects of the phenomena. Schön also states that he sees a designer as someone who changes an indefinite situation into a definite one through a reflexive conversation with the material of the situation. By analogy, network data miner changes the complexity of the problem through the reflexive investigative iterations over the integrated data set. The reflective step is a revision of the reference framework taken in the previous step in terms of attributes selection, the set of visual models and corresponding guiding techniques, and the set of validation techniques.

The main methodological steps and accompanying assumptions of network data mining (NDM) approach include:

Sources of data and modeling: NDM provides an opportunity to integrate data so as to obtain a single address space, a common view, for disparately sourced data. Hence, a decision as to which sources are accessible and most relevant to a problem is an initial consideration. Having arranged and decided the sources, the next question relates to modeling. Which fields of data should serve as entities/ nodes (and attributes on the entities) and which should serve as links (and attributes on the links)[5]. Multiple data models can and often are created to address a particular problem.

[5] A tool that interfaces to relational databases and any original sources of data (e.g. XML files) is basic to NDM, and a capability provided in the NetMap software suite which was used in this case.

Visualization: The entities and a myriad of linkages between them must be presented to screen in meaningful and color coded ways so as to simplify and facilitate the discovery of underlying patterns of data items that are linked to other items and on-linked to still other items. This is especially so with large volumes of data, e.g. many hundreds of thousands and up to millions of links and entities which the user must be able to easily address and readily make sense of from an analytical and interpretative point of view.

'Train of thought' analysis: Linkage between data items means that the discovery of patterns can be a process whereby the analyst uses the reflective practitioner approach mentioned earlier. Explicit querying is less often the case; rather the analyst may let the intuition guide him or her. For example, "Why are all those red links going over there?", "What are they attached to, and in turn what are they attached to?" Such 'train of thought' processes invariably lead the analyst to discover patterns or trends that would not be possible via more explicit querying or exception based approaches – for the specification for the quires is not known.

Cognition and sense-making: An integral assumption in NDM is that the computer in the analyst's mind is more powerful by orders of magnitude that the one on the desktop. Hence, intuition and cognition are integral, and need to be harnessed in the analytical process especially at the discovery phase in those cases where there is only limited domain knowledge as a guide to analysis and understanding.

Discovery: An emergent process, not prescriptive one. It is not possible to prescribe ahead of time all the query rules and exception criteria that should apply to a problem, if domain knowledge is less than perfect. And of course in many if not most cases it is, otherwise the problem would already have been solved. By taking an emergent or bottom up approach to what the data are 'saying', patterns and linkages can be discovered in a way that is not too different from 'good old fashioned' policing – only now it is in the form of electronic data.

Finding patterns that can be re-discovered: Any linkage *pattern* observed on screen is simply that, an observation of potential interest. In the context of retail NDM for example, any sales assistant with a high ratio of refunds to sales (statistically flagged) might attract attention. In a case in point known to the authors, the perpetrators of a scam knew about such exception criteria. As longer term employees in the know, they could easily duck under them. They had taken it in turns to report levels of refunds always just under the limits no matter what the limits were varied to over an extensive period. NDM was able to show collusive and periodic reporting linkages to supervisors – patterns discovered through visualization and algorithms that facilitate the intuition. Such patterns are of particular interest, and in fact often an objective of the network mining approach. They are termed *scenarios* and characterized as definable and re-usable patterns. Their value is that they are patterns that have now become 'known'. Hence they can be defined, stored in a knowledge base, and applied at the front end of a process as, for example, in a predictive modeling environment. The important methodological step is that the definitions need to be discovered in the first place before they can be further applied.

Network data mining is particularly useful in the *discovery phase*, of finding things previously unknown. Once discovered, particular patterns and abnormal behaviors

and exceptions can be better defined. The section below illustrates the network data mining approach using a real-world case study

3 Real World Application of the Network Data Mining Approach

Many cases could be used to illustrate a network data mining approach and each case would be instructive of different features. Nonetheless, this particular case is not atypical of the methods involved in knowledge discovery in network data mining. The steps that the analyst actually took are described below.

The case involved analysis of approximately twelve months of motor vehicle insurance claims from one company. Five thousand records were available as a pilot project from one state of Australia. Note that all the information has been de-identified. The brief from the company was essentially open-ended and without specification. A concern was expressed that there *could* be unusual transactions or fraudulent activity but there were no known persons or transactions of interest. The case clearly necessitated a discovery process.

Using the NetMap software from NetMap Analytics, the analyst first built a set of linkages from the available fields. These were between persons, addresses, claim numbers, telephone numbers, and bank accounts into which claim monies had been paid. Initially the analyst only looked at the three fields of data and the linkages identified between them, as shown in Fig. 2 and Fig. 3.

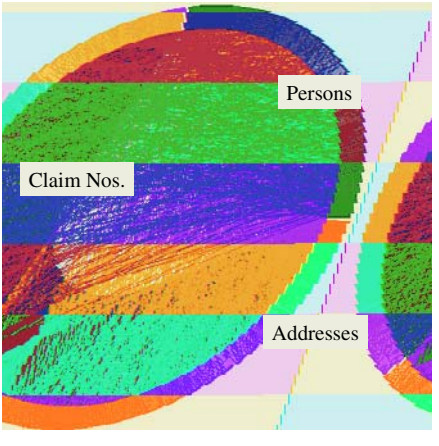


Fig. 2. Overview of links between persons, addresses and claim numbers

Fig. 3. Close up at approximately 3 o'clock of the linked data shown in Fig. 2

To start to make sense of the data overviewed in Fig. 2 and Fig. 3, the analyst then processed the data through an algorithm in NetMap to produce the display shown in Fig. 4 and Fig. 5. Seemingly regular patterns were observed. These appeared as small triangles of data items comprising a person, a claim number and

an address, all fully inter-linked. The explanation was simple: most people had just one claim and one address. However, the 'bumps' were observed as irregularities. They comprised persons linked to multiple claims and/or addresses. By taking the regular patterns out of the picture (the triangles), the analyst was quickly able to produce a short list of potentially suspicious transactions and inter-related behaviors. That is, from a larger amount of data she was able to quickly focus *where* to look to drill for more details.

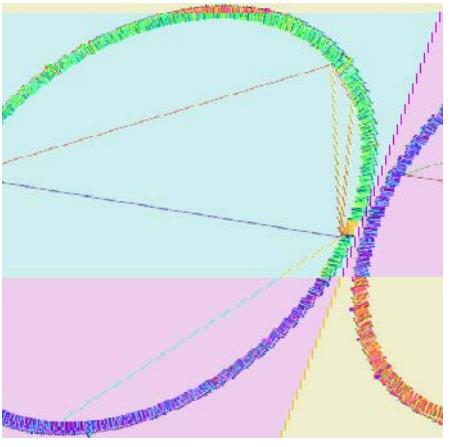


Fig. 4. Data from Fig. 2 and Fig. 3 processed to show certain patterns of irregularities (see detail in Fig. 5)

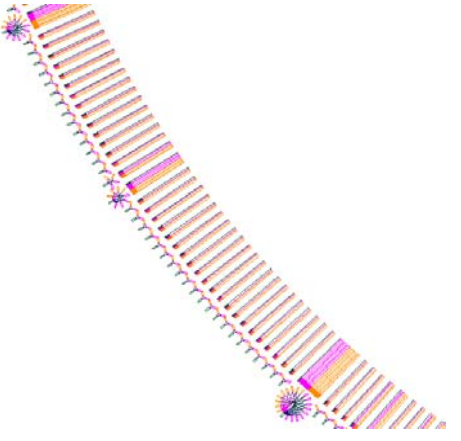


Fig. 5. Close up at 4 o'clock showing what appeared to be regular patterns (small triangles of linkage) and also irregularities (the larger 'bumps')

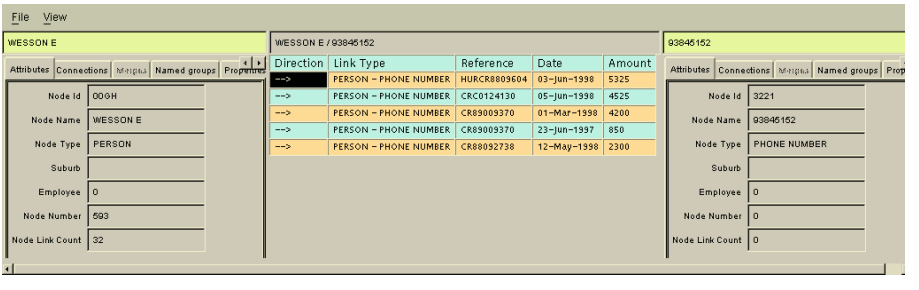


Fig. 6. More detailed information underlying any 'entity' or 'link' was able to accessed by clicking on that data element. The analyst could then quickly qualify observed patterns

Other and more interesting irregular patterns in this case proved to be the links seen in Fig. 4 and extending across the middle. The analyst could see clearly that most of the data items did not have links across the middle and that the linkage seemed to emanate from about 3 o'clock in Fig. 4. Accordingly, she zoomed into the 3 o'clock area and selected one of the inter-linked data items for 'step-link' purposes.

Step-linking from a selected data item follows the network links out through increasing numbers of degrees of separation. Here the analyst stepped out from a selected entity at 3 o'clock to 'infinity' degrees of separation, which effectively brought to screen all that party's indirect linkage through the data set. Obviously, if further sources of data could have been added (which is often done but extra sources were unavailable in this case) richer indirect linkages may have further assisted the inquiry.

The particular extract of linked data in Fig. 7 and Fig. 8 was then processed through an algorithm called *emergent groups*. The resulting pattern is shown in Fig. 9.

From the extract of data in Fig. 7 and Fig. 8, four emergent groups were identified as shown in Fig. 9a. An emergent group comprises closely inter-related data items; they have more links within the group than outside to any other group. Thus they are bottom-up defined 'out of' the data ('what are the data telling us?') rather than prescriptively.

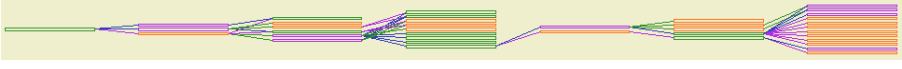


Fig. 7. Step-linking from a selected entity to infinity degrees of separation

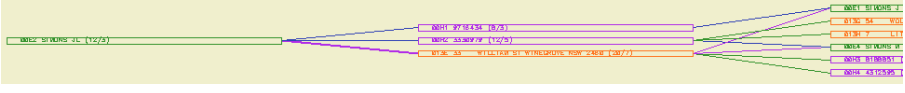
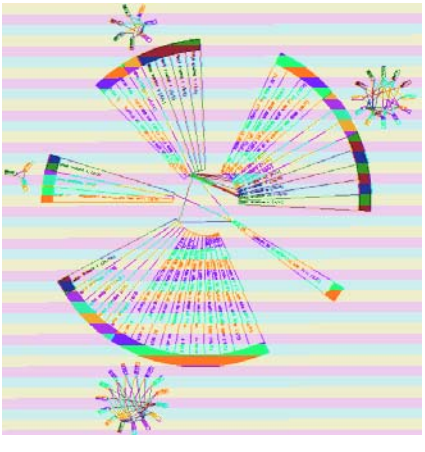
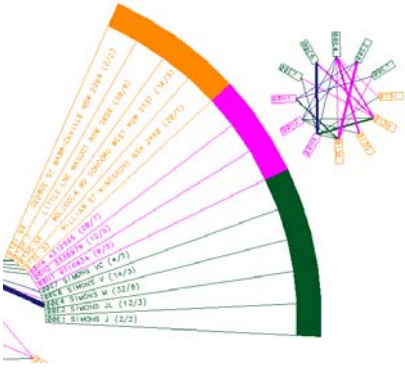


Fig. 8. Close up of portion of the data shown in Fig. 7



a. Four emergent or 'bottom up' groups were identified



b. Close up of one of the emergent groups on the left (in 'a')

Fig. 9. Emergent groups identified from the extract of data shown in Fig. 7 and Fig. 8

The emergent group on the right in Fig. 9b comprised five people called Simons, three claims and four addresses, all closely inter-related. They were linked across to another group at 11 o'clock comprised of more people called Simons and somebody

called Wesson. That Wesson (initial A) was linked down to the group at 5 o'clock to E Wesson via a common claim. That in turn took the analyst over to the address at 4 o'clock and then to a Mr Verman at 9 o'clock.

The above 'train of thought' analysis led the analyst to Verman, who the analyst's intuition indicated she wanted to look at more closely although she could not have been specific as to why. To cut a long story short however, when Verman was investigated and went to jail it was learned that he had had a 'regular' pattern (comprising one claim and one address – see Fig. 5). He reckoned this was a good way to make money. So, he recruited the Simons and the Wessons as the active parties in a scam of staged vehicle accidents while he tried to lie as low as he could. He was careless however, since recorded on a claim form somewhere he had left a link to another address (the one at 4 o'clock in Fig. 9a – note, he must have been a witness or a passenger). This link indirectly implicated him back into the activity surrounding the Simons and Wessons.

The analyst did not know this of course, but could sense that Verman was a few steps removed from the activity involving the Simons and thought she would like to quickly qualify him if possible. She firstly took Verman with all his indirect linkage (see Fig. 10).

She then enriched the linkage in Fig. 10 as a potential case that she would like to qualify on the spot if possible, by adding in extra linkage (see Fig. 11). In this case, she only had two extra fields available: bank account information and telephone numbers. Nonetheless, she quickly discovered one extra and crucial link that helped qualify the potential case, that one of Verman's two telephone numbers was also linked to A Wesson. That additional link provided the extra knowledge that essentially served as a tipping point leading to the recommendation that Verman be investigated. This led to his conviction on fraud charges. It subsequently transpired that Verman had been the mastermind behind a claims scam totaling approximately $150,000. He could not have been discovered by traditional data mining methods since no reasonable sets of exception rules or querying would apply. He would have rated under any such scenario definitions.

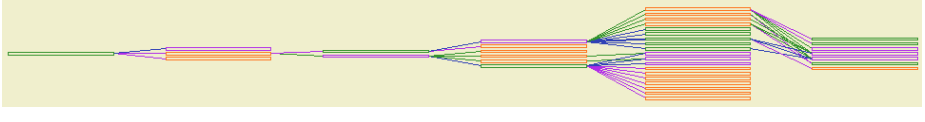


Fig. 10. A potential suspect on the left with all his indirect linkage to other data items

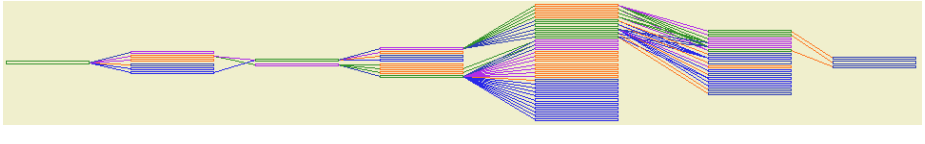


Fig. 11. Enrichment of the linkage in Fig. 10 by adding in extra fields of data

The analyst then enriched what she had discovered as a potential case (see Fig. 12).

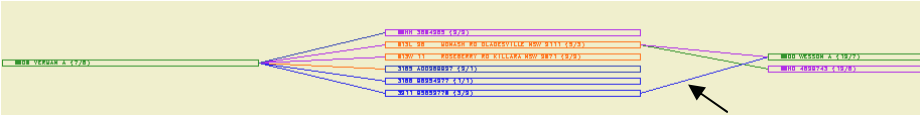


Fig. 12. Close up of portion of Fig. 11 showing the extra 'tell tale' link (arrowed: a telephone number in common) that led to the successful prosecution of Verman, the person shown on the left

That is, in order to qualify on the spot whether Verman should be flagged for the attention of investigators, by adding in extra linkage. In this case, only two extra fields were available: bank account information and telephone numbers. Nonetheless, that extra linkage showed that one of Verman's two telephone numbers was also linked to A Wesson. That insight served as a tipping point, in that discovery of the extra link gave her with the confidence to recommend that Verman be investigated. This led to his conviction on fraud charges. It transpired that Verman had been the mastermind behind a claims scam totaling approximately $150,000.

4 Conclusions

This paper presented the network data mining approach – a human-centred process to knowledge discovery. The approach has been illustrated with a real-world case, relevant to the problems of security. We can summarize the main steps in the knowledge discovery process (Fayyad, Piatetsky-Shapiro et al. 1996) as follows: (1) define scenarios in terms of query specifications and exception rules; (2) process the data, and; (3) interpret or initiate action. The discovery phase, which network data mining gives emphasis to, logically precedes and feeds into step 1. This prior step we refer to as step (0) *discover patterns and qualify them as scenarios.*

Note also that discovery (step 0) and exception detection (step 1) are inter-related. The discovery of new scenarios means that the rules underlying exceptions and querying need to be modified or updated. There is a continuing necessity for periodic updates of more automated and predictive processes.

We presented a case which illustrated the cornerstones of the network data mining approach. The party in the case who was eventually convicted (Verman) would have escaped detection if a traditional data mining approach had been used. He had only had one claim, no 'red flag' information was involved, and nothing particularly anomalous occurred with respect to him. By all accounts he would have slipped under any exception detection procedures. Successful discovery did occur through the use of the NDM methods outlined. Train of thought analysis in an NDM context enabled a discovery to be made in the data that would have been highly unlikely otherwise. Querying could have not have been successful since there is no way that a relevant query could have been framed. Exception rules could not have been specified since there was no information as to what could constitute an exception that would have discovered Verman. He was essentially below the radar.

Masterminds concealing their behavior, tend to be as unobtrusive as possible. They also often know the rules and the exceptions and know what they need to do if the rules and exceptions change so as to avoid being caught. Hence, any discovery tool must go beyond programmatic and prescriptive exception detection.

In our view, network data mining and traditional data mining are essentially complementary. A complementary approach could have been used in the above case as follows. Several of the underlings recruited by Verman had the same family name, Simons and the name Simons appeared more than would have been the case in terms of its occurrence, say, in the telephone directory. Hence, a rule could have been used to display names occurring more often than expected in the telephone population and so the name Simons would have been flagged. This flagging could have been used to aid in the discovery phase, since in an NDM linked data environment one is able to 'step out' numerous steps from, say, Simons and so discovery the mastermind, Verman.

This twin approach has been used to great effect and essentially leverages the advantages of both network and non-network data mining.

References

1. Albert, R. and A.-L. Barabási (2002): "Statistical mechanics of complex networks." *Reviews of Modern Physics* **74**(January 2002), 47-97.
2. Antonie, M.-L., O. R. Zaiane, et al. (2003): Associative classifiers for medical images. *Mining Multimedia and Complex Data*. O. R. Zaiane, S. J. Simoff and C. Djeraba. Heidelberg, Springer, 68-83.
3. Borgatti, S. P. (2003): "The network paradigm in organizational research:
4. A review and typology." *Journal of Management* **29**(6), 991-1013.
5. Domingos, P. and M. Richardson (2001): Mining the network value of customers. *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, ACM Press, 57-66.
6. Dunham, M. H. (2002): *Data Mining: Introductory and Advanced Topics*, Prentice Hall.
7. Fayyad, U. M. (2003): "Editorial." *ACM SIGKDD Explorations* **5**(2), 1-3.
8. Fayyad, U. M., G. Piatetsky-Shapiro, et al. (1996): From data mining to knowledge discovery: An overview. *Advances in Knowledge Discovery and Data Mining*. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy. Cambridge, Massachusetts, AAAI Press/The MIT Press, 1-34.
9. Han, J. and M. Kamber (2001): *Data Mining: Concepts and Techniques*. San Francisco, CA, Morgan Kaufmann Publishers.
10. Kempe, D., J. Kleinberg, et al. (2003): Maximizing the spread of influence through a social network. *Proceedings ACM KDD2003*, Washington, DC, ACM Press
11. Klösgen, W. and J. M. Zytkow, Eds. (2002). *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press.
12. Liben-Nowell, D. and J. Kleinberg (2003): The link prediction problem for social networks. *Proceedings CIKM'03*, November 3–8, 2003, New Orleans, Louisiana, USA., ACM Press
13. Newman, M. E. J. (2003): "The structure and function of complex networks." *SIAM Review* **45**, 167-256.
14. Nong, Y., Ed. (2003). *The Handbook of Data Mining*. Mahwah, New Jersey, Lawrence Erlbaum Associates.

15. Nong, Y. (2003): Mining computer and network security data. *The Handbook of Data Mining*. Y. Nong. Mahwah, New Jersey, Lawrence Erlbaum Associates, 617-636.
16. Ramoni, M. F. and P. Sebastiani (2003): Bayesian methods for intelligent data analysis. *Intelligent Data Analysis: An Introduction*. M. Berthold and D. J. Hand. New York, NY, Springer, 131-168.
17. Richardson, M. and P. Domingos (2002): Mining knowledge-sharing sites for viral marketing. *Proceedings of the Eighth International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, ACM Press, 61-70.
18. Schön, D. (1991): *Educating The Reflective Practitioner*. San Francisco, Jossey Bass.
19. Schwartz, M. E. and D. C. M. Wood (1993): "Discovering shared interests using graph analysis." *Communications of ACM* **36**(8), 78-89.
20. Scott, J. (2000): *Social Network Analysis: A Handbook*. London, Sage Publications.
21. Scott, J. (2000): *Social Network Analysis: A Handbook.* London, Sage Publications.
22. Wasserman, S. and K. Faust (1994): *Social Network Analysis: Methods and Applications*. Cambridge, Cambridge University Press.
23. Wong, P. C. (1999): "Visual Data Mining." *IEEE Computer Graphics and Applications* September/October, 1-3.

Efficient Identification of Overlapping Communities*

Jeffrey Baumes, Mark Goldberg, and Malik Magdon-Ismail

Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA
{baumej, goldberg, magdon}@cs.rpi.edu

Abstract. In this paper, we present an efficient algorithm for finding overlapping communities in social networks. Our algorithm does not rely on the contents of the messages and uses the communication graph only. The knowledge of the structure of the communities is important for the analysis of social behavior and evolution of the society as a whole, as well as its individual members. This knowledge can be helpful in discovering groups of actors that hide their communications, possibly for malicious reasons. Although the idea of using communication graphs for identifying clusters of actors is not new, most of the traditional approaches, with the exception of the work by Baumes et al, produce disjoint clusters of actors, de facto postulating that an actor is allowed to belong to at most one cluster. Our algorithm is significantly more efficient than the previous algorithm by Baumes et al; it also produces clusters of a comparable or better quality.

1 Introduction

Individuals (actors) in a social community tend to form groups and associations that reflect their interests. It is common for actors to belong to several such groups. The groups may or may not be well publicized and known to the general social community. Some groups are in fact "hidden", intentionally or not, in the communicational microcosm. A group that attempts to intentionally hide its communication behavior in the multitude of background communications may be planning some undesirable or possibly even malicious activity. It is important to discover such groups before they attempt their undesirable activity.

In this paper, we present a novel *efficient* algorithm for finding *overlapping communities* given only the data on who communicated with whom. Our primary motivation for finding social communities is to use this information in order to further filter out the hidden malicious groups based on other criteria, see [2]. However, the knowledge of all communities can be crucial in the analysis of social behavior and evolution of the society as a whole, as well as its individual members [9].

The need for an automated tool to discover communities using only the presence or absence of communications is highlighted by the sheer impracticality of

* This research was partially supported by NSF grants 0324947 and 0346341.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 27–36, 2005.

conducting such a discovery by looking at the volumes of the actual contents of the communications, or trying to interview actors regarding the social groups to which they belong (or think they belong). The idea of using communication graphs for identifying "clusters" of users is not new. It is the guiding principle for most classical clustering algorithms such as distance-based algorithms [7]; flow-based algorithms [6]; partitioning algorithms [3, 8]; and matrix-based algorithms that employ the SVD-technique [5]. The main drawback of these approaches is that they all produce *disjoint clusters* of actors, de facto postulating that an actor is allowed to belong to at most one cluster. This is a severe limitation for social communication networks. A serious need exists for efficient tools to produce *overlapping communities* or *clusters*.

The mathematical formulation of the problem of determining clusters that may possibly be overlapping was introduced in [1], which defines a cluster as a *locally optimal* subgraph with respect to a given *metric*. We briefly review this notion of a cluster in Section 2. Since locally optimal subgraphs may overlap, this formulation allows for overlapping clusters. The algorithm for finding such locally optimal subgraphs, presented in [1] consists of two parts: **initialization**, RaRe, which creates *seed clusters*; and **improvement**, IS, which repeatedly scans the vertices in order to improve the current clusters until one arrives at a locally optimal collection of clusters.

The focus of this paper is to provide a new *efficient* algorithm for initializing the seed clusters and performing the iterative improvements. Specifically, our main contributions are a procedure List Aggregate (LA) for initializing the clusters and a procedure IS^2 which iteratively improves any given set of clusters. The combined algorithm develops overlapping subgraphs in a general graph. Our algorithm is a significant improvement over the algorithms from [1]. In particular, our algorithm can be applied to large ($\sim 10^6$) node networks. The computational experiments, comparing the new algorithm with that from [1], show that the new algorithm is an order of magnitude faster, and simultaneously produces clusters of a comparable or better quality.

2 Clusters

In this paper we adopt the idea formulated in [2] that *a group C of actors in a social network* forms a community if its communication "density" function achieves a local maximum in the collection of groups that are "close" to C. We call two groups close if they become identical by changing the membership of just one actor. Several different notions of density functions were proposed in [2]; here we consider and experiment with one more notion; specifically, the density of a group is defined as the average density of the communication exchanges between the actors of the group.

Thus, a group is a community if adding any new member to, or removing any current member from, the group decreases the average of the communication exchanges. We call a *cluster* the corresponding subgraph of the graph representing

the communications in the social network. Thus, our definition reduces the task of identifying communities to that of graph clustering.

Clustering is an important technique in analyzing data with a variety of applications in such areas as data mining, bioinformatics, and social science. Traditionally, see for example [4], clustering is understood as a partitioning of the data into disjoint subsets. This limitation is too severe and unnecessary in the case of the communities that function in a social network. Our definition allows for the same actor to be a member of different clusters. Furthermore, our algorithm is designed to detect such overlapping communities.

3 Algorithms

3.1 The Link Aggregate Algorithm (LA)

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, [1], as opposed to random edges in the communication network. We discuss a different, efficient initialization algorithm here.

procedure LA($G = (V, E), W$)
$C \leftarrow \emptyset$;
Order the vertices $v_1, v_2, \ldots, v_{|V|}$;
for $i = 1$ to $|V|$ **do**
  $added \leftarrow$ false;
  for all $D_j \in C$ **do**
    if $W(D_j \cup v_i) > W(D_j)$ **then**
      $D_j \leftarrow D_j \cup v_i$; $added \leftarrow$ true;
  if $added =$ false **then**
    $C \leftarrow C \cup \{\{v_i\}\}$;
return C;

The Rank Removal algorithm (RaRe) [1] begins by ranking all nodes according to some criterion, such as Page Rank [10]. 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.

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 are 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 Rank Removal algorithm with this improvement, unless otherwise stated.

Since the 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 Link Aggregate LA focuses on efficiency, while still capturing good initial clusters. The pseudocode is given above. 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, 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 1. *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 $O(|C_i|deg(v_i))$, 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: either the other endpoint of the edge is 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

$$\sum_{deg(v_i)>0} |C_i|deg(v_i) + \sum_{deg(v_i)=0} 1 \leq \sum_{i=1}^{|V|} |C_i|deg(v_i) + \sum_{i=1}^{|V|} 1$$

$$\leq \sum_{i=1}^{|V|} |C|deg(v_i) + |V| = 2|C||E| + |V| = O(|C||E| + |V|).$$

3.2 Improved Iterative Scan Algorithm (IS^2)

The original algorithm IS explicitly constructs a cluster that is a local maximum w.r.t. a density metric by starting at a "seed" candidate cluster and updating it by adding or deleting one node at a time as long as the metric strictly improves. The algorithm stops when no further improvement can be obtained with a single change. The original process consists of iterating through the entire list of nodes over and over until the cluster density cannot be improved.

The new algorithm IS^2, based on IS, is given in pseudocode format to the right. In order 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 adjacent 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. Note that this algorithm is not strictly the same as the original IS algorithm, since potentially a node absent from N could become a neighbor of the cluster while the nodes are being examined. This node has a chance to join the cluster in the original IS algorithm,

```
procedure IS²(seed,G,W)
C ← seed; w ← W(C);
increased ← true;
while increased do
  N ← C;
  for all v ∈ C do
    N ← N ∪ adj(v);
  for all v ∈ N do
    if v ∈ C then
      C′ ← C \ {v};
    else
      C′ ← C ∪ {v};
    if W(C′) > W(C) then
      C ← C′;
  if W(C) = w then
    increased ← false;
  else
    w ← W(C);
return C;
```

while in IS^2 it is skipped. This is not an issue since the node will have a chance to join the cluster in the next iteration of IS^2.

This algorithm provides both a potential decrease and increase in runtime. The 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, the original algorithm IS is preferable.

4 Experiments

A series of experiments were run in order to compare both the runtime and performance of the new algorithm with its predecessor. In all cases, a seed algorithm was run to obtain initial clusters, 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 range from 1,000 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

$$W_{ad}(C) = \frac{2|E(C)|}{|C|},$$

where $E(C)$ is the set of edges with both endpoints in C.

The runtime for the algorithms is presented in Figure 1. The new algorithm remains quadratic, but both the seed algorithm and the refinement algorithm run-times are improved significantly for sparse graphs. In the upper left plot in Figure 1, the original version of RaRe is also plotted, which recalculates the node ranks a number of times, instead of precomputing 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 five 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 about 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 2 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

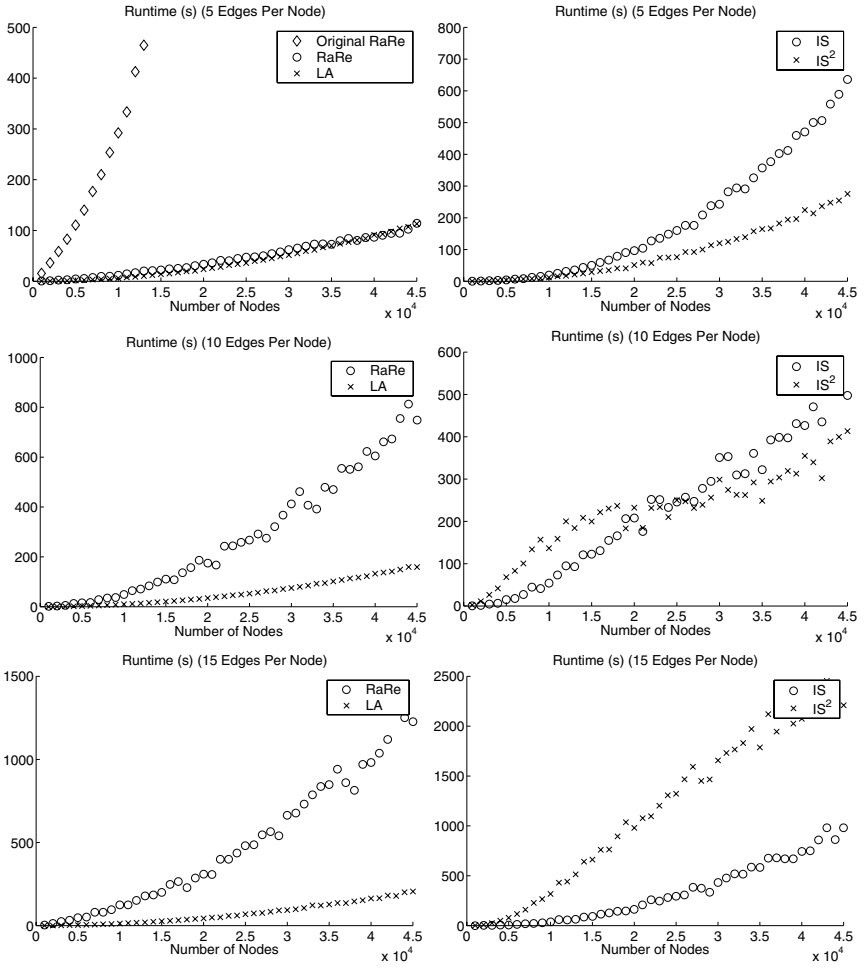


Fig. 1. Runtime of the previous algorithm procedures (RaRe and IS) compared to the current procedures (LA and IS^2) with increasing edge density. On the left is a comparison of the initialization procedures RaRe and LA, where LA improves as the edge density increases. On the right is a comparison of 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

runtime improvements. Figure 3 compares the average density of the clusters found for both the old and improved algorithms. A higher 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

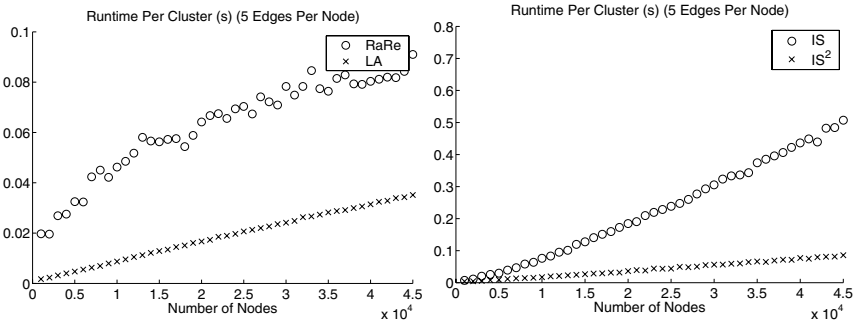


Fig. 2. Runtime per cluster of the previous algorithm (RaRe followed by IS) and the current algorithms (LA followed by IS^2). These plots show the algorithms are linear for each cluster found

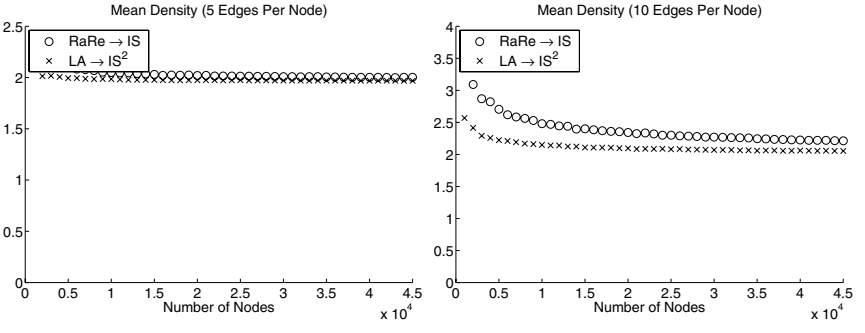


Fig. 3. Performance (average density) of the algorithm compared to the previous algorithm

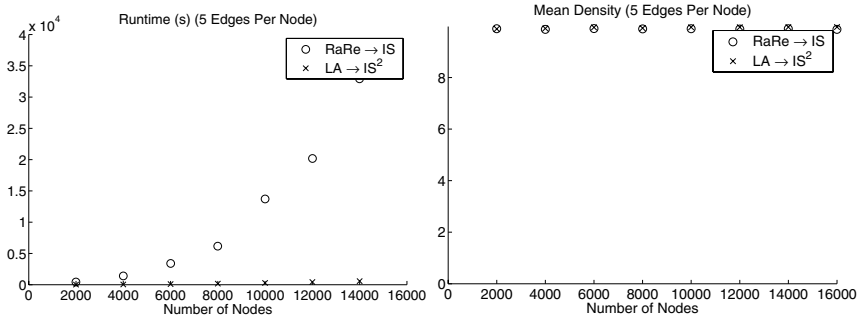


Fig. 4. Runtime and performance of the previous algorithm (RaRe followed by IS) and the current algorithm (LA followed by IS^2) for preferential attachment graphs

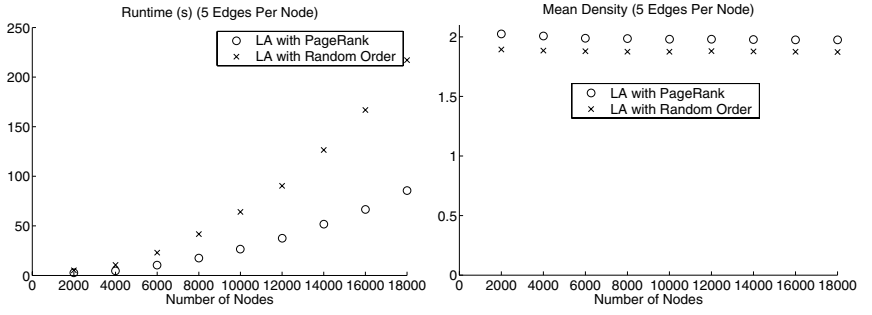


Fig. 5. Runtime and performance of LA with two different ordering types

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 five links per node, and the number of nodes ranging from 2,000 to 16,000. Figure 4 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 all 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 Figure 4 shows that the quality of the clusters are not compromised by using the significantly faster new algorithm $\mathsf{LA} \rightarrow \mathsf{IS}^2$.

Figure 5 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 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 1. For all cases other than the web graph, the new algorithm produced a clustering of higher quality.

Table 1. Algorithm performance on real-world graphs. 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 (4,526 nodes). The Fortune 500 graph is the network connecting companies to members of their board of directors (4,262 nodes)

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
Algorithm	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

5 Conclusions

We have described a new algorithm for the discovery of overlapping communities in a communication network. This algorithm, composed of two procedures LA and IS^2, was tested on both random graph models and real world graphs. The new algorithm is shown to run significantly faster than previous algorithms presented in [1], while keeping the cluster quality roughly the same and often better. In addition, we demonstrated that the LA procedure performs better when the nodes are ranked, as opposed to a random order. Surprisingly, though the ranking process initially takes more time, the procedure runs more quickly overall.

Directions for our ongoing work are to test different metrics, and to apply the algorithms to a variety of networks ranging from social communication networks to protein networks, ranging in sizes from hundreds of nodes to a million nodes. There are a variety of options for parameter settings available to the user, and it will be useful to provide the practitioner with an exhaustive database of test-cases, giving guidelines for how to set the parameters depending on the type of input graph.

References

1. J. Baumes, M. Goldberg, M. Krishnamoorty, M. Magdon-Ismail, and N. Preston. Finding communities by clustering a graph into overlapping subgraphs. *Proceedings of IADIS Applied Computing 2005*, pages 97–104, February 2005.
2. J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. Wallace. Discovering hidden groups in communication networks. *2nd NSF/NIJ Symposium on Intelligence and Security Informatics*, 2004.
3. J. Berry and M. Goldberg. Path optimization for graph partitioning problem. *Discrete Applied Mathematics*, 90:27–50, 1999.

4. U. Brandes, M. Gaertler, and D. Wagner. Experiments on graph clustering algorithms. *Lecture Notes in Comuter Science*, Di Battista and U. Zwick (Eds.):568–579, 2003.
5. P. Drineas, R. Kannan, A. Frieze, S. Vempala, and V. Vinay. Clustering in large graphs and matrices. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1999.
6. G.W. Flake, K. Tsioutsiouliklis, and R.E. Tarjan. Graph clustering techniques based on minimum cut trees. Technical report, NEC, Princeton, NJ, 2002.
7. A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, 1988.
8. B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49:291–307, 1970.
9. M. E. J. Newman. The structure and function of complex networks. *SIAM Reviews*, 45(2):167–256, June 2003.
10. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Stanford Digital Libraries Working Paper*, 1998.

Event-Driven Document Selection for Terrorism Information Extraction

Zhen Sun[1], Ee-Peng Lim[1], Kuiyu Chang[1], Teng-Kwee Ong[2], and Rohan Kumar Gunaratna[2]

[1] Centre for Advanced Information Systems, School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore
aseplim@ntu.edu.sg

[2] International Center for Political Violence and Terrorism Research, Institute of Defence and Strategic Studies, Nanyang Technological University, Singapore 639798, Singapore

Abstract. In this paper, we examine the task of extracting information about terrorism related events hidden in a large document collection. The task assumes that a terrorism related event can be described by a set of entity and relation instances. To reduce the amount of time and efforts in extracting these event related instances, one should ideally perform the task on the relevant documents only. We have therefore proposed some document selection strategies based on information extraction (IE) patterns. Each strategy attempts to select one document at a time such that the gain of event related instance information is maximized. Our IE-based document selection strategies assume that some IE patterns are given to extract event instances. We conducted some experiments for one terrorism related event. Experiments have shown that our proposed IE based document selection strategies work well in the extraction task for news collections of various size.

Keywords: Information extraction, document selection.

1 Introduction

1.1 Objectives

Information about a certain terrorism event frequently exists across several documents. These documents could originate from different portal and news websites around the world, with varying amount of information content. Clearly, it is not always necessary for a terrorism expert to find all documents related to the event since they may carry duplicate information. As far as the expert is concerned, it is important to read only a small subset that can give a complete and up-to-date picture so as to maximize his/her efficiency.

We therefore propose an extraction task, for which several pattern-based document selection strategies were studied. The extraction task aims to incrementally select a set of documents relevant to a terrorism event, using various

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 37–48, 2005.

document selection strategies designed to aid in the extraction task. A set of patterns for extracting event specific entity and relationship instances from a document is assumed to be given. We also assume that some seed entity instances are given to bootstrap the extraction process. The overall objective is to find all entity and relationship instances related to the given event from the smallest possible subset of documents.

In the following, we summarize our contribution to this research:

- We formally define the extraction task which incorporate a document selection strategy to find event related entity and relationship instances. This task has not been studied before and our work therefore establishes the foundation for this field.
- We propose a few document selection strategies to identify the smallest possible subset of documents for event related instances. Each document selection strategy aims to maximize the novelty of the set of entity and relationship instances that can be found in the next document to be extracted. In this way, one can hopefully reduce the number of documents that the terrorism experts have to review.
- We have created two datasets to evaluate our extraction task and document selection strategies. The experimental results show that our strategies performs well and appears to scale well to large document collection.

1.2 Paper Outline

In the remaining portion of this paper, we present related work in Section 2, followed by formal definitions of the extraction task in Section 3. Next, we describe our proposed document selection strategies in Section 4. Experiments and results are given in Sections 5 and 6 respectively. Section 7 gives the conclusion and future work.

2 Related Work

Finding entity and relation instances of a certain event is our research focus and this is related to named entity recognition. Named Entity [4] Recognition deals with extracting specific classes of information (called "entities") such as person names, locations, and organizations from plain text. Michael Chau et al. [14] addressed the problem on extracting entities from police narrative reports. In their work, they built a neural network-based extraction method.

Named Entity Recognition can be viewed as a kind of single-slot extraction. Single-slot extraction such as AutoSlog [6] and its extensions [7, 8, 9] have been developed and which have demonstrated high extraction accuracy. Multi-slot extraction [2, 5, 10] refers to extracting multiple slots with some relationships from a document. Our work utilizes both multi-slot extraction and named entity recognition in the extraction task.

New Event Detection (NED) is a document selection task to identify the first story of an event of interest from an ordered collection of news articles. Since the

first story does not necessarily contain all the entity and relation instances of an event, NED methods cannot be applied directly to the event-driven extraction task [11, 12]. These methods often do not involve information extraction except in [13], where Wei and Lee proposed an information extraction based event detection (NEED) technique that uses both information extraction and text categorization methods to conduct NED.

Finn and Kushmerick proposed various active learning selection strategies to incrementally select documents from a large collection for user labelling so as to derive good extraction patterns [1]. In contrast, our work focuses on finding documents containing both novel and related information with the help of extraction patterns.

3 Event-Driven Extraction with Document Selection

3.1 Event Representation Using Entity and Relation Instances

In our extraction task, we represent a terrorism event by a set of entity and relation instances. The entity instances describe the people, organisations, locations, dates/times and other information relevant to the event. The relation instances provide the links between entity instances so as to understand their inter-relations. Prior to the extraction task, we assume that a terrorism expert wishes to derive all entity and relation instances for a single event. To ensure that only relevant instances are extracted, a set of entity and relation classes are assumed to be known apriori.

Let E be a set of *entity classes*, i.e. $E = \{E_1, E_2, ..., E_n\}$, and R be a set of *relation classes*, $R = \{R_1, R_2, ..., R_m\}$. E and R together describe the information to be extracted for a target terrorism event. An entity class E_i denotes a set of entity instances of the same type, and each entity instance is usually a noun or noun phrase appearing in the document. Each relation class R_i represents a semantic relationship from an entity class $SourceEnt(R_i)$ to an entity class $TargetEnt(R_i)$ and is associated with an *action class* A_i. A_i refers to a set of verbs or verb phrases that relate source entity instances in $SourceEnt(R_i)$ to target entity instances in $TargetEnt(R_i)$. Each relation instance comprises a source entity instance from $SourceEnt(R_i)$, a target entity instance from $TargetEnt(R_i)$, and an action instance from A_i, i.e., $R_i \subseteq SourceEnt(R_i) \times A_i \times TargetEnt(R_i)$, where $SourceEnt(R_i), TargetEnt(R_i) \in E$.

3.2 Event-Driven Extraction Task

Suppose we are given a set of extraction patterns EP, a collection of documents D, and a set of seed entity instances W relevant to an event. Let E and R represent the entity classes and relation classes relevant to the event. We use $\mathcal{E}$ to denote the set of all entity instances contained in E, i.e. $\mathcal{E} = \cup_{i=1}^{n} E_i$, and $\mathcal{R}$ to denote the set of all relation instances in R i.e., $\mathcal{R} = \cup_{i=1}^{m} R_i$. W is a small subset of $\mathcal{E}$ useful for bootstrapping the extraction of other instances. To ensure that all instances will be extracted given the seed entity instances W, we require

Algorithm 1. Event-Driven Extraction Task

inputs: EP, D, W
for each document d_j in D **do**
  apply EP on d_j to obtain $\mathcal{E}'_j$, $\mathcal{R}'_j$
  score d_j using **InitialScore** function $score(d_j)$
end for
repeat
  Find the document d_s with highest $score(d_s)$, move d_s from D to S
  Extract (manually by an expert user) entity and relation instances from d_s
  Add newly extracted instances from d_s to $\mathcal{E}$ and $\mathcal{R}$
  for each document d_j in D **do**
    re-score d_j using a score function $score(d_j)$ based on $\mathcal{E}'_j$, $\mathcal{R}'_j$, $\mathcal{E}$, $\mathcal{R}$
  end for
until termination condition is satisfied
outputs: S, $\mathcal{E}$, $\mathcal{R}$

all event instances $\mathcal{E}$ to be directly or indirectly linked to W through the relation instances in $\mathcal{R}$.

In the **event-driven extraction task**, documents for extracting event related instances are selected one at a time. At the beginning, the seed entity instances set W is given to identify the relevant documents. Each time a document is selected, it is given to the expert user for manual extraction of entity and relation instances. Note that manual extraction is conducted to ensure that no instances are missed. This process repeats until all event related entity and relation instances are found.

The detailed description of the task is depicted in Algorithm 1. During the extraction task, the extraction patterns EP are used to find the existence of entity and relation instances that could be relevant to the event. The extraction patterns can be for single-slot, or multi-slot extraction. The former is appropriate for extracting entity instances while the latter can be used for extracting both entity and relation instances. The entity and relation instances extracted from a document d_j using EP are stored in $\mathcal{E}'_j$'s and $\mathcal{R}'_j$'s respectively.

Assuming that the expert user has in mind a set of entity and relation instances to be extracted for an event. We can then define a set of documents containing relevant instances as the *relevant set* denoted by L. The objective of the event-driven extraction task on the other hand is to select the smallest subset O of L that covers all relevant instances. We call O the **optimal set**. Let $\mathcal{E}_j$ and $\mathcal{R}_j$ denote the set of entity and relation instances in document d_j. Then O is an optimal set if and only if it satisfies the following two conditions:

1. $(\cup_{d_j \in O} \mathcal{R}_j = \mathcal{R})$ and $(\cup_{d_j \in O} \mathcal{E}_j = \mathcal{E})$
2. $\nexists O'$ *s.t.* $(\cup_{d_j \in O'} \mathcal{R}_j = \mathcal{R})$ and $(\cup_{d_j \in O'} \mathcal{E}_j = \mathcal{E})$ and $(|O'| < |O|)$

4 Pattern-Based Document Selection Strategies

We have developed several document selection strategies using different score functions in the proposed extraction task. Each document selection strategy adopts a different score function to rank documents. In general, documents containing significant novel and related information should have higher scores. Since these strategies rely on extraction patterns to identify potentially relevant entity and relation instances, we call them *pattern-based document selection strategies.*

4.1 InitialScore

This is the default strategy that selects documents based on the given seed entity instances W. This document selection strategy is therefore used in the first iteration only. The primary objective of scoring is to assign higher scores to documents that have more extraction patterns fired. The first term of the score formula below considers proportion of W that is extracted. This is to ensure a relevant document is selected initially.

$$score(d_j) = \frac{|\mathcal{E}'_j \cap W| + \frac{|W|}{\gamma}}{|W|} \cdot log_2(|EP_j|) \cdot \sum_{k=1}^{|EP_j|} f_k \tag{1}$$

where $\gamma \gg |W|$ is a smoothing factor that prevents the first term from becoming zero if $\mathcal{E}'_j \cap W$ is empty, EP_i is a subset of EP that fired on document d_j, and f_j is the number of relation instances extracted by extraction pattern $ep_{j,k}$. In our experiment, we used $\gamma = 100$ with $|W| = 4$.

4.2 DiffCompare

This strategy examines the amount of overlap between relation instances extracted from the current document d_j with the accumulated relation instance set $\mathcal{R}$. The smaller the overlap, the higher the score. In addition, the amount of intersection between the extracted entity instances $\mathcal{E}'_j$ and W is also considered. This is to assign higher score for documents having direct links to the seed set. Contribution from the two factors are linearly weighted by $\alpha \in [0, 1]$. Equation (2) shows the score function:

$$score(d_j) = \alpha \cdot \frac{|\mathcal{R}'_j - \mathcal{R}|}{\max_{d_l \in D} |\mathcal{R}'_l|} + (1 - \alpha) \cdot \frac{|\mathcal{E}'_j \cap W|}{|W|} \tag{2}$$

where N is the total number of documents in D.

4.3 CombineCompare

This strategy combines the amount of intersection and dissimilarity between relation instances extracted from d_i with instances in $\mathcal{R}$. A modifier $\beta \in [0, 1]$ is used to adjust the relative importance of overlapping relation instances compared with novel relation instances (i.e., relevant relation instances that have not been

extracted so far). When the former is more important, $\beta > 0.5$. When $\beta = 0.5$, both are treated equally important. Equation (3) gives the score function of this strategy. Note that when $\beta = 0$, this is equivalent to DiffCompare.

$$score(d_j) = \alpha \cdot \frac{\beta \cdot |\mathcal{R}'_j \cap \mathcal{R}| + (1-\beta) \cdot (|\mathcal{R}'_j - \mathcal{R}|)}{\max_{d_l \in D} |\mathcal{R}'_l|} + (1-\alpha) \cdot \frac{|\mathcal{E}'_j \cap W|}{|W|} \quad (3)$$

4.4 PartialMatch

In this document selection strategy, we want to select documents with relation instances linked to those entity instances that have already been found. This requires a partial match between the former and latter. Note that all entity instances in the event are connected with others using relation instances. This applies even in the midst of extraction task. Hence, we only need to conduct partial match between a relation instance extracted using EP and the relation instances found so far.

Given two relation instances $r_s = (e^s_s, a_s, e^t_s)$ and $r_t = (e^s_t, a_t, e^t_t)$, the partial match of r_s and r_t denoted by $PartialMatch(r_s, r_t)$ is defined by:

$$PartialMatch = \begin{cases} 0 \text{ if } e^s_s \neq e^s_t \wedge e^t_s \neq e^t_t \wedge (a_s \in A_p, a_t \in A_q, p \neq q) \\ 0 \text{ if } e^s_s = e^s_t \wedge e^t_s = e^t_t \wedge (a_s \in A_p, a_t \in A_q, p = q) \\ 0 \text{ if } e^s_s \neq e^s_t \wedge e^t_s \neq e^t_t \wedge (a_s \in A_p, a_t \in A_q, p = q) \\ 1 \text{ otherwise} \end{cases}$$

With $PartialMatch$ measuring the novelty of instances, we now define the score function for the partial match document selection strategy in equation (4):

$$score(d_j) = \alpha \cdot \frac{\sum_{k=1}^{M_j} \sum_{h=1}^{|\mathcal{R}|} PartialMatch(r'_{j,k}, r_h)}{|\mathcal{R}'_j| \cdot |\mathcal{R}| + 1} + (1-\alpha) \cdot \frac{|\mathcal{E}'_j \cap W|}{|W|} \quad (4)$$

where M_j is the number of relation instances extracted from d_j using EP; $r'_{j,k}$ is the kth relation instance from $\mathcal{R}'_j$; and r_h is the hth instance in $\mathcal{R}$.

5 Experimental Setup

5.1 Construction of Experiment Datasets

We used two datasets covering the terrorism event of Australian Embassy bombing (AEB) in Jakarta, September 2004. They are the AEB and AEB-1000 datasets. Both datasets were created by downloading documents from an online news website and converting the documents to plain text. The seed words used for the extraction task are "*Australian Embassy*", "*Australian Embassy Bombing*", "*Suicide Bombers*" and "*Elisabeth Musu*". Among them, Elisabeth Musu is a victim who was injured during the event. Based on the above seeds, other entity and relation instances about the event were determined by an expert familar with the event as shown in Figure 1. In the figure, relation instances are

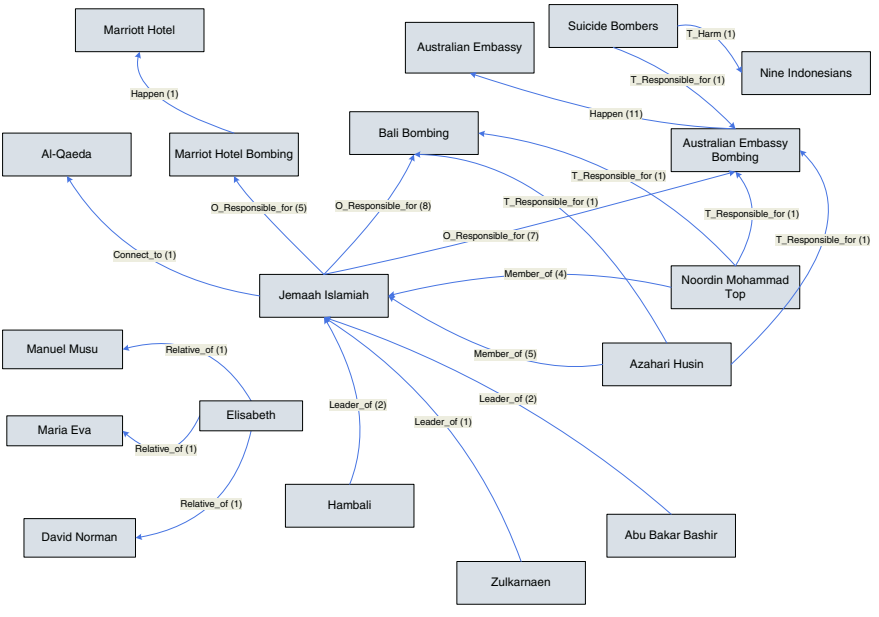


Fig. 1. Entity and Relation Instances of Australian Embassy Bombing Event

represented by directed edges. The numbers in brackets show the occurrences of the relation instance in all relevant documents.

AEB dataset has 100 documents consisting of 34 relevant documents and 66 irrelevant documents. The 34 relevant documents were selected to cover all instances of the bombing event. The 66 irrelevant documents were selected from more than 10,000 documents downloaded during the week after the event occurred. These documents were intentionally selected to describe other similar criminal events such as murdering and kidnapping. In other words, both the relevant and irrelevant documents describe some criminal events and they have a certain similarity content-wise. This also increases the level of difficulty in the document selection.

AEB-1000 dataset has 1000 documents. The relevant documents in AEB-1000 are identical to that of AEB. In addition to the 66 irrelevant documents, we randomly selected 900 irrelevant documents for AEB-1000. With AEB-1000, we can evaluate the performance of our document selection strategies for a larger dataset.

There are altogether 7 entity classes and 10 relation classes that we are interested. The 7 entity classes are: Victim, Terrorist, Terrorist Organization(Org), Event, Location, Employer and Relative. And the 10 relation classes are: T_{Harm}, O_{Harm}, Connect_to, $T_{Responsible\_for}$, $O_{Responsible\_for}$, Member_of, Leader_of, Happen, Work_for and Relative_of. Table 1 shows these 10 relation classes with their source and target entity classes. Table 2 shows more detailed information about the two datasets.

Table 1. 10 relation classes with their source and target entity classes

⟨Rel⟩	(⟨SrcEnt⟩,⟨TgtEnt⟩)	⟨Rel⟩	(⟨SrcEnt⟩,⟨TgtEnt⟩)
T_{Harm}	(Terrorist, Victim)	O_{Harm}	(Terrorist Org, Victim)
Connect_to	(Terrorist Org, Terrorist Org)	$T_{Responsible\_for}$	(Terrorist, Event)
$O_{Responsible\_for}$	(Terrorist Org, Event)	Member_of	(Terrorist, Terrorist Org)
Leader_of	(Terrorist, Terrorist Org)	Happen	(Event, Location)
Work_for	(Victim, Employer)	Relative_of	(Victim, Relative)

Table 2. Detailed Information of the two datasets

	$\|\mathcal{E}\|$	$\|\mathcal{R}\|$	# of Relevant docs	# of Optimal docs	Total docs
AEB	19	20	34	9	100
AEB-1000	19	20	34	9	1000

5.2 Construction of Extraction Patterns

The IE system chosen for the experiment is Crystal [2]. We have manually created a set of extraction patterns for extracting the entity and relation instances. These extraction patterns were created based on some common linguistic structures of the English language in order to be applied in a generic extraction task. For example: ⟨Subject⟩ ⟨Verb⟩ ⟨Object⟩, ⟨Subject⟩ ⟨Verb⟩ ⟨Prepositional Phrase⟩, ⟨Verb⟩ ⟨Object⟩ ⟨Prepositional Prase⟩ and ⟨Subject⟩ ⟨Prepositional Phrase⟩ are four common structures we used. By constraining one or more part of each structure by words, we have the extraction patterns. For example: ⟨Subject⟩ in one extraction pattern can be constrained by the Terrorist entity class, while the ⟨Object⟩ is constrained by the Victim entity class. An extraction pattern is not going to fire on a sentence unless some constraints have been met. In other words, we use instances in the action and entity classes to guard the invocation of extraction patterns.

As we are interested in terrorism events, we use WordNet [3] to obtain some words for initializing the entity and action classes in the extraction patterns. These are generic words that can be used to describe the action classes relevant to terrorism and the names of already known terrorists. In our experiments, 21 terrorists' names found on FBI website[1] and 54 terrorist organization's names found on ICT website[2] have been included into entity class *Terrorist* and *Terrorist Organization* instance sets to form the extraction patterns.

5.3 Evaluation Settings

In our experiment, we set $\alpha = 0.6$ to place a higher emphasis on the relation instances with respect to the seed entity instances. For CombineCompare, we

[1] http://www.fbi.gov/mostwant/terrorists/fugitives.htm
[2] http://www.ict.org.il

set $\beta = 0.8$ to give more weight to documents containing larger number of relation instances already found. The experiment was conducted by running the extraction task for 45 iterations on AEB, and 50 iterations on AEB-1000.

We also propose a set of performance metrics defined below, which were evaluated after every 5 documents have been selected. These performance metrics focus on how much relevant instances the selected documents contain and how well each document selection strategy perform.

1. Evaluation on Extracted Entity and Relation Instances
   Suppose we have all relevant entity instances in set $\mathcal{E}_r$ and all relevant relation instances in set $\mathcal{R}_r$. To evaluate the resultant sets obtained in extraction task i.e., $\mathcal{E}$ and $\mathcal{R}$, let $|\mathcal{E}_a|$ be the number of intersection between sets $\mathcal{E}_r$ and $\mathcal{E}$, i.e. $|\mathcal{E}_a| = |\mathcal{E}_r \bigcap \mathcal{E}|$, and $|\mathcal{R}_a|$ be the number of intersections between sets $\mathcal{E}_r$ and $\mathcal{R}$, i.e. $|\mathcal{R}_a| = |\mathcal{R}_r \bigcap \mathcal{R}|$. The recall measure is defined as follows:

   - $Recall_{average} = \frac{1}{2}(\ Recall_{entity} + Recall_{relation})$
     where $Recall_{entity} = \frac{|\mathcal{E}_a|}{|\mathcal{E}_r|}$ and $Recall_{relation} = \frac{|\mathcal{R}_a|}{|\mathcal{R}_r|}$

2. Evaluation on Document Selection
   Let L be the set of all relevant documents and S denote the set of selected documents. The precision and recall measures with respect to relevant documents are defined as follows:

   - $Precision_{rel\_doc} = \frac{|S \bigcap L|}{|S|}$
   - $Recall_{rel\_doc} = \frac{|S \bigcap L|}{|L|}$

   Suppose there are v different optimal sets among all relevant documents (since the optimal set is usually not unique). Let $\mathcal{O}$ denote the set of all optimal sets, i.e., $\mathcal{O} = \{O_1, O_2, ..., O_v\}$. We have $|O_1| = |O_2| = ... = |O_v|$. The recall and precision measures respect to optimal set are defined as follows:

   - $Recall_{opt\_doc} = \frac{\max_{O_i \in O} |O_i \bigcap S|}{|O_1|}$
   - $Precision_{opt\_doc} = \frac{\max_{O_i \in O} |O_i \bigcap S|}{|S|}$

5.4 Ideal Document Selection Strategies

We introduce an ideal document selection strategy here to compare our proposed document selection strategies. The ideal selection strategy selects a document that gives the largest increase in the proportion of performance measurement during each iteration and the selected document must be a relevant document. We assume all documents have been manually annotated with entity and relation instances. Therefore, the score formula for ideal selection strategy is defined as follows:

$$score(d_i) = M(d_i)$$

where M refers to the improvement of a chosen performance metric brought by selecting document d_i.

Note that the ideal document selection strategy is not achievable in practice as we cannot accurately determine the instances in the documents to be selected. We however would like to use it to examine how far worst are the other document selection strategies.

6 Experimental Results

Figure 2(a) shows the $Recall_{average}$ of AEB dataset. PartialMatch gives the best performance as it reaches almost perfect recall with the smallest number of iterations (documents selected). DiffCompare is the runner-up, followed by CombineCompare ($\beta = 0.8$). PartialMatch extracted 95% entity and relation instances with almost less than half the number of documents compared to other pattern-based strategies.

We conclude from Figures 2(b) and 2(c) that PartialMatch consistently performs well in selecting relevant documents for the AEB dataset. It maintained perfect $Precision_{rel\_doc}$ until more than 20 iterations. While $Precision_{rel\_doc}$ of other strategies oscillate below 1 indicating that they are not able to select the relevant documents all the time. Although not shown in the figure, PartialMatch performs better on selecting the optimal documents. It selected 90% optimal documents at the $21th$ iteration, which is much better than DiffCompare ($41th$ iteration) and CombineCompare ($44th$ iteration).

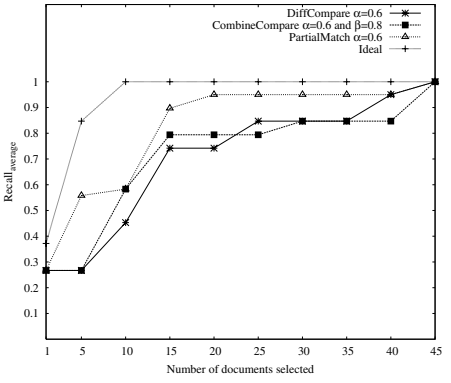
For the AEB-1000 dataset, PartialMatch reached 95% $Recall_{average}$ in the $21th$ iteration as shown in Figure 2(d). This is followed by CombineCompare ($\beta = 0.8$) and DiffCompare. The extraction task selected only 5% of the total number of documents and obtained almost all entity and relation instances with PartialMatch. In other words, even the worst pattern-based strategies can find more than 65% instances by selecting only 5% documents.

Although PartialMatch is the best among all strategies, it's performance is lower than that of Ideal selection. Figure 2(a) shows that the Ideal strategy reaches perfect $Recall_{average}$ in the $9th$ iteration, while PartialMatch requires 44 iterations. Therefore, there is still some room for improvement.

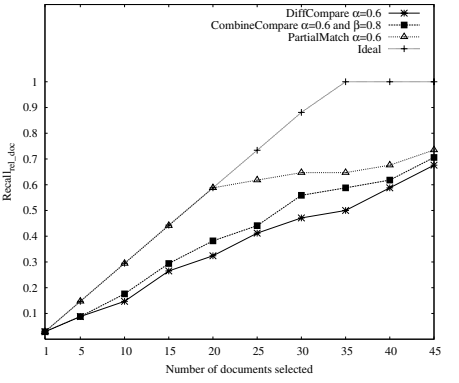
7 Conclusions

We have proposed a new event driven extraction task and four pattern-based document selection strategies in this paper. This task is applied to the terrorism event information extraction. Our objective is to select as few documents as possible to construct the event related entity and relation instances.

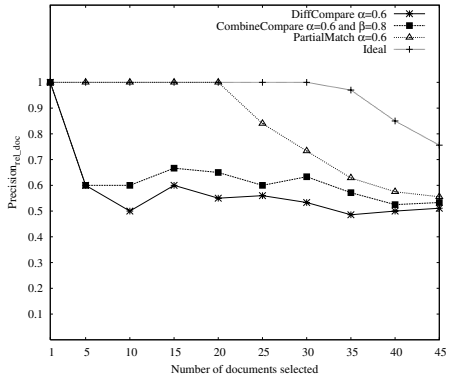
We have defined performance metrics to compare the proposed document selection strategies and conducted several experiments on 2 datasets. Experimental results conclude that our proposed strategies perform well on the extraction task.



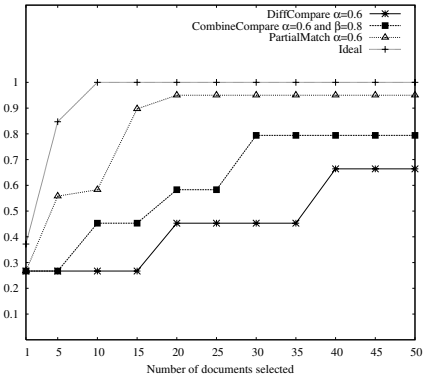
(a) Extracted Instances (AEB)



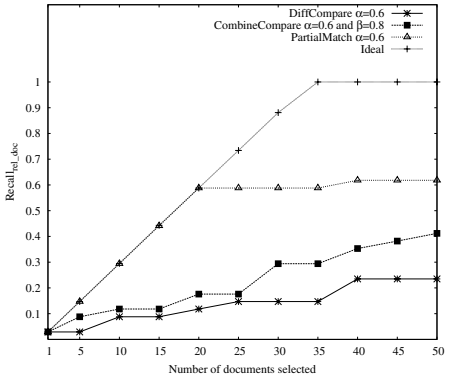
(b) Document selection recall (AEB)



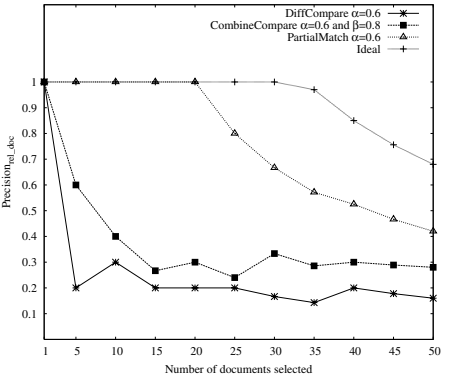
(c) Document selection precision (AEB)



(d) Extracted Instances (AEB-1000)



(e) Document selection recall(AEB-1000)



(f) Document selection precision(AEB-1000)

Fig. 2. Experimental Results

Among our proposed strategies, PartialMatch shows the best performance. Especially for a dataset containing 1000 documents, it managed to extract 95% of the required event related information by selecting only 5% of the documents.

References

1. Finn, A., Kushmerick, N.: Active learning selection strategies for information extraction. In: Proceedings of ATEM. (2003)
2. Soderland, S., Fisher, D., Aseltine, J., Lehnert, W.: Crystal: Inducing a conceptual dictionary. In: Proceedings of the 14th IJCAI. (1995)
3. Fellbaum, C.: Wordnet: An electronic lexical database. MIT Press (1998)
4. Maynard, D., Tablan, V., Ursu, C., Cunningham, H., Wilks, Y.: Named entity recognition from diverse text types. In: Proceedings of Natural Language Processing 2001 Conference. (2001)
5. Huffman, S.: Learning information extraction patterns from examples. In: Proceedings of IJCAI-95 Workshop on new approaches to learning for natural language processing. (1995)
6. Riloff, E.: Automatically constructing a dictionary form information extraction tasks. In: Proceedings of the 11th National Conference on Artificial Intenlligence. (1993)
7. Riloff, E.: Automatically generating extraction patterns from untagged text. In: Proceedings of the 13th National Conference on Artificial Intenlligence. (1996)
8. Riloff, E., Jones, R.: Learning dictionaries for information extraction by multi-level bootstrapping. In: Proceedings of the 16th National Conference on Artificial Intenlligence. (1999)
9. Thelen, M., Riloff, E.: A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing. (2002)
10. Agichtein, E., Gravano, L.: Snowball: Extracting relations from large plain-text collections. In: Proceedings of the Fifth ACM International Conference on Digital Libraries. (2000)
11. Allan, J., Papka, R., Lavrenko, V.: On-line new event detection and tracking. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. (1998)
12. Kumaran, G., Allan, J.: Text classification and named entities for new event detection. In: Proceedings of the 27th annual international conference on Research and development in information retrieval. (2004)
13. Wei, C.P., Lee, Y.H.: Event detection from online news documents for supporting environmental scanning. Decis. Support Syst. **36** (2004) 385–401
14. C, Michael., J, Xu., Chen, Hsinchun.: Extracting Meaningful Entities from Police Narrative Reports. In: Proceedings of the National Conference for Digital Government Research. (2002)

Link Analysis Tools for Intelligence and Counterterrorism

Antonio Badia and Mehmed Kantardzic

University of Louisville, Louisville KY 40292, USA
{abadia, mmkant}@louisville.edu

Abstract. Association rule mining is an important data analysis tool that can be applied with success to a variety of domains. However, most association rule mining algorithms seek to discover statistically significant patterns (i.e. those with considerable support). We argue that, in law-enforcement, intelligence and counterterrorism work, sometimes it is necessary to look for patterns which do not have large support but are otherwise significant. Here we present some ideas on how to detect potentially interesting links that do not have strong support in a dataset. While deciding what is of interest must ultimately be done by a human analyst, our approach allows filtering some events with interesting characteristics among the many events with low support that may appear in a dataset.

1 Introduction

In data mining, most algorithms seek to discover underlying patterns and trends hidden in large amounts of data. Following established statistical practice, in association rule mining one of the attributes of such trends is that they are present in a large portion of the data, i.e. that they have large support. This allows analysis to focus on patterns that are very unlikely to arise by chance. Thus, this point of view has served data analysis well for a long time in traditional domains. However, law-enforcement, intelligence and counterterrorism analysts may need to look for small nuggets of highly significant data. Because they work in an *adversarial* situation, where the opponents try very hard to hide traces of their activities, it is likely that low-support trends could reveal something of interest. Note that it is perfectly possible for high-support trends to be meaningful in this environment: for instance, ATT is said to have examined its huge data warehouse after 9/11 and discovered that the volume of calls from Afghanistan to the U.S. in the months preceding the tragedy tripled its normal volume. Thus, there is certainly strength in numbers. However, events with low support may also carry some significance, and are disregarded by most data mining tools (although there has been some research on the significance of such events; see section 4). Unfortunately, focusing only on events with very low support means having to separate random, meaningless events from potentially significant ones. In most domains, a quasi-normal distribution means that there

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 49–59, 2005.

will be a large number of low-support events. The challenge, then, is to explore such low-support events and be able to determine which ones are potentially interesting.

Recently, it has become quite fashionable to analyze data by building a graph (network) out of a set of observations and applying tools and techniques of graph theory. The recent emphasis in *social network theory* ([16]) is one example of this approach. Usually, the network is created in such a way that nodes correspond to observational units of interest (for instance, people), while links between nodes reflect the strength of some known connection (i.e. the higher the support for the connection, the higher the strength of the link). Thus, even though the graph allows the analyst to focus on localized interactions (between two individuals, or institutions, or places, at a time), it still carries a notion that high-support events are more important than low-support ones.

In this paper, we extend previous work in which we formalized the process of graph creation and pointed out several new ways of looking at the resulting graph (reference withhold). In particular, we introduce ways to analyze the strength of connections regardless of the support, to determine whether they happen by chance, and to assess their rarity in the dataset. We believe that such measures may enable the analyst to determine which connections are potentially interesting. We argue that the ideas introduced in this work can be fruitfully applied to intelligence and counterterrorism tasks.

In section 2 we briefly overview our process for network creation. In section 3 we introduce the idea of *interesting* and *significant* connections and show, through examples, how it can be used to filter relevant links from large amounts of data in a network format. In section 4 we overview some related research. We close with an analysis of the work and directions for future efforts.

2 Constructing Graphs From Data

To consider our analysis of low-support events in context, we study the creation of a graph or network from a set of raw data. To analyze this process, we try to decide how nodes are characterized, and how links are established. We assume that, regardless of how raw data is present, all the information can be put into a *table*. A table is taken here in the sense of relation, as in relational databases: a schema or set of attributes is given, and each entry in the table (row) contains one value per attribute. This step is needed to introduce some uniformity, as raw data may be given in a large number of different formats. We realize that the process of fitting raw data into the table can be quite complex, and may involve what is usually called cleaning, standardization, and other steps ([9, 12]). While this is a very important part of the mining process, we do not consider it here here, instead focusing on creating the graph from the information in the table.

We consider the process to create a graph out of a basic set of observations as guided by the analysis that a user has in mind; therefore, it is possible to build multiple graphs from the same set of basic facts. To choose which graph to build, the user must make two basic choices: to determine what constitutes an *object* or

Emails			
From	**To**	**Date**	**Content**
P_1	P_2	1	A
P_2	P_5	1	A
P_3	P_4	2	C
P_1	P_3	3	B
P_2	P_3	3	B
P_4	P_1	4	F
P_1	P_2	5	C
P_3	P_6	5	C
P_2	P_5	6	D
P_3	P_4	1	A
P_4	P_1	1	C
P_1	P_3	2	C
P_2	P_3	5	C
P_5	P_3	2	C
P_3	P_4	1	A
P_3	P_3	5	C
P_4	P_2	1	A

Fig. 1. Set of raw data

node; and to determine a measurement of *object connectivity*, from which links will be defined. Given a collection of objects C with attributes $\mathcal{A} = A_1, \ldots, A_n$ (i.e. a table with schema $\mathcal{A}$), the user choses a non-empty, strict subset of the attributes $\mathcal{O} \subset \mathcal{A}$, called an *object definition*. Intuitively, these are the attributes that determine what an object is. The set of objects under consideration is given by the different values for those attributes.

As an example that we use throughout the paper, let a table *Emails* have attributes `From, To, Date, Content`, as in Figure 1. This is meant to be a list of emails, each row being a particular email. The attribute `Content` is the text of the email; in order not to introduce complex text analysis issues ([1, 2]), in our example we simplify the content to a category. In this table, a user may choose `From` to be the object definition (in which case the set of objects of type `From` is the set of all people sending messages: $\{P_1, P_2, P_3, P_4, P_5\}$), which means the analysis is centered in who sends the messages. Or, the user could choose `Date` (in which case the set of objects of type `Date` is the set of all dates in which messages were sent: $\{1, 2, 3, 4, 5, 6\}$), which means the analysis is centered on time. Choosing `From, Date` as a pair (in which case the set of objects of type `From, Date` is the set of all pairs (s,d) where s is a person, d is a date and s sent a message in date d: $\{(P_1, 1),\ (P_2, 1),\ (P_3, 2), (P_1, 3), (P_2, 3), (P_4, 4), (P_1, 5), (P_3, 5), (P_2, 6), (P_3, 1), (P_4, 1), (P_1, 2), (P_2, 5), (P_5, 2)\}$) means that the analysis is centered on who sends the messages when.

Given an object o in a collection C, the *support set* of o, in symbols $S(o)$, is defined as the set of tuples in C that have o as the object definition. Continuing

with our previous example, the support set of P_1 is the set of all tuples that have the value P_1 in attribute `From`: $\{(P_1, P_2, 1, A), (P_1, P_3, 3, B), (P_1, P_2, 5, C), (P_1, P_3, 2, C)\}$.

Once an object definition chosen, next the user choses a set of attributes $\mathcal{L} \subseteq \mathcal{A} - \mathcal{O}$, called the set of *link dimensions*. For instance, in our previous example, if the user chose `From` as the object definition, any set with attributes `To, Date, Content` can be used. Note that $\mathcal{L} \cap \mathcal{O} = \emptyset$. The attributes in this set will be used to determine which links, if any, exist between pairs of objects. The *link support set* of an object o with dimensions $\mathcal{L}$, in symbols $S(\mathcal{L}, o)$, is defined as the set of values in $\mathcal{L}$ for all the tuples in the support set of o. Thus, following on our previous example, assume `From` is the object definition, and `Date, Content` is the set of link dimensions. Given the support set of P_1 as before, the link support set is $\{(1, A), (3, B), (5, C), (2, C)\}$.

Once the user has chosen link dimensions (and a support set), the question still remains of how to determine when a link exists between two objects o_1 and o_2. There are several well-known proposals, all of them based on the size of $S(\mathcal{L}, o_1) \cap S(\mathcal{L}, o_2)$ (i.e., what the two objects have in common) ([3]). We call this the *common set* of the two objects. These measures are usually normalized, meaning that they always take a value between 0 and 1, regardless of the size of the common set. This has the advantage that the measures of different pairs of objects may be meaningfully compared, but has the disadvantage that a threshold value must be chosen to consider two objects connected. Setting this threshold is problematic, as any number is bound to be somewhat arbitrary.

We propose two new, non-normalized measures that are based on the following intuition: two objects are connected if they have more in common than they have different. Thus, we define

$$S_1(o_1, o_2) = \begin{cases} \alpha & \text{if } S(\mathcal{L}, o_1) = S(\mathcal{L}, o_2) \\ \frac{|S(\mathcal{L},o_1) \cap S(\mathcal{L},o_2)|}{max(|S(\mathcal{L},o_1) - S(\mathcal{L},o_2)|, |S(\mathcal{L},o_2) - S(\mathcal{L},o_1)|)} & else \end{cases}$$

$$S_2(o_1, o_2) = \begin{cases} \alpha & \text{if } S(\mathcal{L}, o_1) = S(\mathcal{L}, o_2) \\ \frac{|S(\mathcal{L},o_1) \cap S(\mathcal{L},o_2)|}{|S(\mathcal{L},o_1) - S(\mathcal{L},o_2)| + |S(\mathcal{L},o_2) - S(\mathcal{L},o_1)|} & else \end{cases}$$

In the above, α denotes a symbol larger than any positive number. These measures are not normalized, in that they can return any positive number; we can make a virtue out of this by noting that now a natural threshold emerges: 1. In effect, if $S_2(o_1, o_2) > 1$, this means that o_1 and o_2 have more in common (with respect to the chosen link dimensions) than they have different; and if $S_1(o_1, o_2) > 1$, this means that o_1 and o_2 have more in common than any one of them has different from the other. Thus, we consider o_1, o_2 *disconnected* when $S_i(o_1, o_2) = 0$; *connected* when $S_i(o_1, o_2) > 1$; and neither connected nor disconnected in all other cases ($i = 1, 2$).

Note that most data mining tasks focus on large sets; our approach covers both large and small sets. Of particular note is the fact that when both $S(\mathcal{L}, o_1)$ and $S(\mathcal{L}, o_2)$ are small, or when both $S(\mathcal{L}, o_1)$ and $S(\mathcal{L}, o_2)$ are large, the con-

nection depends solely on how much they have in common. But when one of $S(\mathcal{L}, o_1)$ and $S(\mathcal{L}, o_2)$ is large and the other one is small, our definition will rule out a connection. We consider this situation analogous to the *negative correlation* problem for association rules, in which what seems like a statistically significant overlap tends to be caused by the overall distribution of the data. Thus, we wish to avoid such cases (and our approach effectively does this) as they are suspect and likely not to carry any meaning.

As usual, a *graph* is a collection of objects (from now on called *nodes*) and of links (from now on called *edges*). Edges can be classified according to their *directionality* (they may be one-way or directed, or two-way or undirected), and their *character* (they may be positive, denoting closeness of two objects, or negative, denoting differences o distance between two objects). The above measures S_1 and S_2 can be combined to determine not only whether a link between two objects exist, but also, if it exists, whether it is directional or not, and its character ([5, 14]).

Given a collection of facts C, a graph or network for C is a pair $(\mathcal{O}, \mathcal{L})$, where $\mathcal{O}$ is an object definition and $\mathcal{L}$ is a link dimension for $\mathcal{O}$. The nodes in this graph are the objects o determined by $\mathcal{O}$ in C, and the edges are the links created according to link measure S_1 and S_2 as follows: create a negative edge if $max(S_1(o_1, o_2), S_2(o_1, o_2)) < 1$; a positive edge (with weight i) if $min(S_1(o_1, o_2), S_2(o_1, o_2)) = i > 1$; and no edge otherwise. An edge can be given direction as follows: if $S_1(o_1, o_2) << S_2(o_1, o_2)$ the edge should be given direction from the object with smaller support set to the one with the larger; if $S_1(o_1, o_2) \approx S_2(o_1, o_2)$, the edge is bidirectional. Note that using more than one measure to determine direction is necessary since by definition measures are symmetric, while a directional edge is not symmetric.

Following with our example, assume a user chooses `From` as the object definition, and `{Date, Content}` as link dimensions, the graph means: person o_1 and person o_2 are sending messages at the same (or close) time with the same (or similar) content. We don't show the whole resulting graph, focusing on the link between nodes P_1 and P_2. The support sets are as follows: for P_1, {(1,A), (3,B), (5,C), (2,C)}; for P_2, {(1,A), (3,B), (6,D), (5,C)}. Then, S_1 is computed as the size of the intersection of such sets (which is 3, as the intersection is {(1,A), (3,B), (5,C)}) divided by the symmetric difference (which is 2, as there is one element in P_1 minus P_2, and one element in P_2 minus P_1), for a value of 1.5 (if we use S_2, we divide by the maximum of the two, which is 1, leading to a value of 3). Therefore, we conclude that this is a strong link.

There are many ways in which a law-enforcement or counterterrorism analyst may be interested in constructing a graph, possibly incorporating some type of constraints in the process. Assume that, for or example data set, we build a graph where nodes correspond to people, and we simply link two nodes a and b if a ever sent some email to b or vice-versa. This measure may be too crude: it may happen that many emails are sent *en masse* (with a large number of addresses in the `cc` field, for instance). Then most people are going to be connected. It may be more interesting to link two people only if they have exchanged a certain

number of emails, or if they have exchanged a certain number of emails per day for a minimum number of days, or if they both have sent out emails (not necessarily to each other) on the same topic at (about) the same time. We try to capture some intuitions about what makes a connection interesting. Unlike other work in data mining, we disregard the size of the support of a link, and try to define its *interestingness*.

3 Interesting and Significant Connections

Note that in our approach, the absolute size of the support sets is not used. Therefore, we will consider a connection based on its strength alone. However, there may be many such connections that are of no interest. In the next subsections, we try to isolate connections that are potentially interesting among all connections that have a strong link.

3.1 Significant Connections

We still have not determined when a connection (link) is interesting. We believe that it is impossible to formalize completely a notion of *interestingness* or *relevance*, since such a notion is bound to be context-dependent and somewhat subjective. We can attempt, however, to capture some aspects of the concept.

Our approach is based on the following intuition: the formal definitions introduced before captured what is usually known as the *confidence* of a link, i.e. the strength that we associate with it. In order to be interesting, a link should also be *statistically significant* (i.e. should not occur by chance), and should be *rare* (not frequent). These two ideas seem to be in contradiction, as traditionally statistical significance has been taken to mean high frequency. Here, however, we take it to be simply that the events in question are more related than they would be by chance. To implement that, we use a χ^2 test for categorical data. This allows us to consider low-support events, and simply disregard those that are likely *noise*.

We also require events to be infrequent but to contribute significantly to a link. The intuition here is that rare events should barely appear in any given connection; if they appear often, that is worth noting. In our example, if messages of a certain type are rare (they occur very infrequently), but most of such messages happen to be between two people, that should call our attention. Note that this does not *per se* guarantee that the event is interesting -only the analyst can ultimately make that determination- but it certainly warrants some attention.

We explain our approach with an example. We continue with the example of the link between P_1 and P_2, as derived from the data in Figure 1. We have already determined the link is strong. We then apply the χ^2 test as follows: the expectation matrix in Figure 2 is created. Note that the square corresponding to $\bar{P_1}, \bar{P_2}$ has the number of values *in the whole table* that are not in the support set of P_1 or in the support set of P_2 -in the example, this corresponds to {(4,F),

	P_2	$\bar{P}_2$	Total
P_1	3	1	4
$\bar{P}_1$	1	2	3
Total	4	3	7

Fig. 2. Expectation Matrix for P_1, P_2

(1,C)}. Then the traditional χ^2 test yields a value of 3.8, which means that the link is statistically significant ($p > 0.05$). Now we compute our third value, *rarity*. We use three measures for this: *Selective Rarity (SR)*, *local Rarity (LR)* and *Global Rarity (GR)*. We compute, for each element in the common set of the link, how many times the element appears in the whole dataset, and compare this number of the number of appearances in the common set under consideration. For instance, (1,A) occurs 5 times in the table, of which 2 are in the link between P_1 and P_2; there are two occurrences of (3,B) in the database, and both are in the link between P_1 and P_2; and (5,C) occurs 4 times in the whole database, 1 of those in the link between P_1 and P_2. We then compute, for each data point, the fraction of those two values as *number of occurrences in the link support set / total number of occurrences*. For (1,A), this value is $\frac{2}{5} = .4$; for (3,B), the value is $\frac{2}{2} = 1$ (the maximum possible value); for (5,C), the value is $\frac{1}{4} = .25$. SR chooses the maximum of all these values, which in our example is 1. This is as high as it can get, and therefore we deem the event interesting. LR takes the average instead of the maximum; so $avg(0.4, 1, 0.25) = .55$. This measure should be above 0.5, so we also deem the link interesting under this measure. Finally, our last check includes the fact that the element under consideration is indeed infrequent in the data. We notice that (1,A) occurs 5 times out of 17 observations, (3,B) occurs 2 times out of 17, while (5,C) occurs 4 times out of 17. Note that the minimum frequency is $\frac{1}{N}$, where N is the total number of facts under consideration (17, in this example). The GR of a value is defined as the ratio of this fraction to the SR of the value. For instance, (3,B) occurs infrequently (2/17) while its SR is as high as possible (1). These two facts combined indicate that we should consider this an interesting event: one that is rare and yet, happens frequently between P_1 and P_2.

We believe that this is of interest to law-enforcement, counter-terrorism and intelligence work because, as state above, in these lines of work it may be necessary to detect small but potentially significant events. For instance, in a very large database of calls, a few of them mentioning nitrates and other fertilizer material that can be also used to make explosives may go unnoticed (and, say, 10 calls out of one million may be simply noise). But if most of the calls happen to be between the same two people (say, 8 out of 10 calls were between two individuals), this is a fact that may be worth investigating. While there may be many several cases (small number of calls about a given subject), the χ^2 test should help weed out those that are likely happening by accident, and the tests introduced above should allow the analyst to focus on a few, potentially relevant leads.

3.2 Interesting Connections

Now that we have determined that the connection between P_1 and P_2 is interesting (according to our measures), a question arises as to how this can be expanded to other nodes in order to locate, for instance, interesting subgraphs.

Assume now that we have a link l_1 between objects o_1 and o_2, and a link l_2 between objects o_2 and o_3, and we have already determined that l_1 and l_2 are interesting. We would like to find out if objects o_1 and o_3 are also related in an interesting way. The connection can then be defined based on the underlying common sets for links l_1 and l_2; that is, the links may be composed only if their underlying common sets pass some test. Assume, for instance, that the link dimensions on the graph were {`Date,Body`} (so that common sets are made up of pairs $< d, b >$, where d is the date of an email and b is its body). Then one can define an *interesting connection* between o_1 and o_3 to exists iff $S((\mathcal{L}, o_1) \cap S(\mathcal{L}, o_2)) \cap (S(\mathcal{L}, o_2) \cup S(\mathcal{L}, o_3)) \neq \emptyset$. Note that the intersection here is computed on sets of pairs. This idea can be extended without difficulty to *paths* (or, to be technically accurate, to paths of length > 1). We can use this idea to derive a graph G' from a graph G as follows: for each set of nodes $o_1, o_2, o_3 \in N(G)$ such that there is a link between o_1 and o_2 and a link between o_2 and o_3, compute the interesting connection. If one exists, create a node named (o_1, o_2) and a node named (o_2, o_3) and a link between them. Note that different topologies in G given raise to nodes with different *labels* in G'. For instance, the example just given is one of a path; cycles give raise to other labeling, as do *stars* (i.e. graphs in which one central node o is connected to all the rest of the nodes $o_1, \ldots, o_n$, which are not connected among them). Three example topologies are given in Figure 3, assuming that all connections are found interesting (the original graph G on the left, the resulting G' to its right). Given two arbitrary objects o_1 and o_2 in graph G, the question *Is there a connection between* o_1 *and* o_2*, and if so, what is its nature?* can be answered as follows: if there is a path in the original graph G, compute G', G'', G''',... that is, compute the transitive closure of the graph under the interesting connection relationships. If a direct link is established between o_1 and o_2, then there is indeed a connection; the nature of the connection can be explained by looking at the successive common sets computed along the way.

Note that we can relax the idea of an interesting connection. If the original linking was based on {`To, Content`}, a and b were connected if they sent messages to the same person d about the same topic t, and b and c were connected if they sent messages to the same person d' about the same topic t'. We could be interested in connecting a and c if they sent messages out to the same person (i.e. $d = d'$), or if they sent messaged out about the same topic (i.e. $t = t'$). Of course, we could also be interested in them only if they sent out message to the same person about the same topic (i.e. if $d = d'$ *and* $t = t'$). Relaxing or tightening the constraints imposed on the connection, we can determine subgraphs of the original graph which are more or less tightly (and interestingly) connected. Continuing with our example in the previous subsection, once we have determined that persons a and b are talking about some suspicious substances a lot

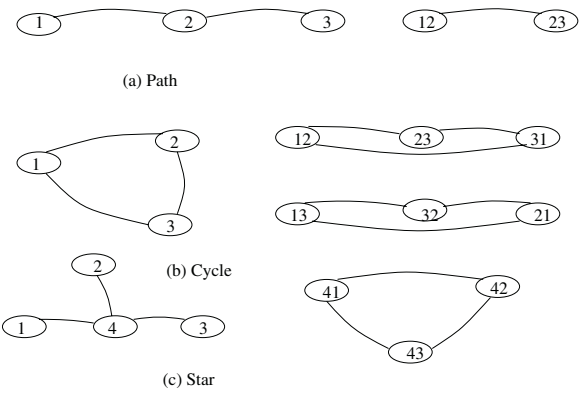


Fig. 3. Different Interesting Connections

more than could be expected by chance, we may want to see who a and b talked to about the same subject, or about the same time that they were talking to each other (or both). Again, we insist that it is ultimately the analyst who must decide, based upon further analysis and context knowledge, whether an event is of interest. The purpose of our process is to allow the analyst to focus on a few facts of potential interest, among a large volume of data.

4 Related Work

There is a considerable amount of work that can be regarded as somewhat relevant to this project; we focus here on research that is directly related to the work presented herein. There has been research on determining when a relationship between two variables is interesting, usually in the context of association rule mining ([15, 8, 6, 10, 11, 3]). However, some of this work still assumes that the typical support and confidence measures are used to filter association rules before any further analysis; since support excludes small sets, such work does not cover the same ground as ours. Most of the work derives from statistical concepts and therefore focuses on statistical relevance ([8]). One exception is [3], which proposes a measure that, like ours, is independent of support. This measure, usually called *Jacquard's coefficient*, was well-known before [3], but this work presents its use in the context of association rules, and defines a probabilistic algorithm to compute the measure very efficiently. Various notions of rarity are introduced to measure the interestingness of paths in bibliographical datasets in [13]. While experiments show promising results, better handling of noise, corruption and directionality of links is desirable. Messages with rare words are analyzed in [14] using *independent component analysis (ICA)*, when correlated words with "wrong" frequencies trigger detection of interesting messages/ Other work has looked at connections in the context of Web searches ([7]). Applying work on network analysis to law enforcement and related tasks is a flourishing area of research ([17]).

5 Conclusion and Further Work

We have argued that in law-enforcement, intelligence and counter-terrorism work, it is sometimes necessary to look for small-support patterns and determine those that are of interest. This is a very difficult task, as in many large collections of data, small-support sets may be frequent and patterns involving them may arise by chance. Determining which patterns are random and which ones are interesting is an intuitive, unformalized process that analysts in these fields must deal with. Here we have presented some measures that are formal and therefore can be supplied to a computer for efficient processing of large datasets. We stress that, even though we have kept our presentation intuitive and used examples to introduce the main ideas, the approach presented here can be completely formalized. In this paper, our goal was to stress the applicability of tools like the one we are developing to law enforcement, counter-terrorism and intelligence work.

Currently, we are testing our approach on some real-life datasets ([4]), refining our concepts and developing algorithms that allow for efficient computation of the desired results over very large datasets. Our approach allows for several graphs to be created from the same data; we are developing ways to combine those graphs in an overall view of the data. We also plan to organize all steps involved in going from the raw data to the final analysis in a common framework.

References

1. Belew, R. *Finding Out About*, Cambridge University Press, 2000.
2. Baeza-Yates, R. and Ribeiro-Neto, B. *Modern Information Retrieval*, Addison-Wesley, 1999.
3. Cohen, E., Datar, M., Fujiwara, S., Gionis, A., Indyk, P., Motwani, R., Ullman, J.D. and Yang, C. *Finding Interesting Associations without Support Pruning*, in Proceedings of the 16th International Conference on Data Engineering (ICDE) 2000, San Diego, California, USA, IEEE Computer Society.
4. Enron data set, available at `http://www-2.cs.cmu.edu/7Eenron/`.
5. Domingos P., *Multi-Relational Record Linkage*, Proceedings of the KDD-2004 Workshop on Multi-Relational Data Mining, Seattle, CA, ACM Press, 2004, pp. 31-48.
6. Hussain, F., Liu, H., Suzuki, E. and Lu, H. *Exception Rule Mining and with Relative Interestingness Measure*,in Proceedings of the Knowledge Discovery and Data Mining, Current Issues and New Applications, 4th Pacific-Asia Conference (PADKK 2000), Kyoto, Japan, 2000, Lecture Notes in Computer Science 1805, Springer.
7. Hue G., et al, *Implicit Link Analysis for Small Web Search*, SIGIR-03, Toronto, Canada, 2003.
8. Imberman, S.P., Domanski, B. and Thompson, H. W. *Using Dependency/Association Rules to Find Indications of Computarized Tomography in a Head Trauma Dataset*, Artificial Intelligence in Medicine, 26(1-2), September-October 2002, Elsevier.
9. Jarke, M., Lenzerini, M., Vassiliou, Y. and Vassiliadis, P. *Fundamentals of Data Warehousing*, Springer-Verlag, 2000.

10. Jaroszewicz, S. and Simovici, D.A. *A General Measure of Rule Interestingness*, in Proceedings of Principles of Data Mining and Knowledge Discovery, 5th European Conference (PKDD 2001) Freiburg, Germany, 2001, Lecture Notes in Computer Science 2168, Springer.
11. Jaroszewicz, S. and Simovici, D.A. *Interestingness of frequent itemsets using Bayesian networks as background knowledge*, in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 2004, ACM Press.
12. Kuramochi M., Karypis G., *An Efficient Algorithm for Discovering Frequent Subgraphs*, Technical report, Department of Computer Science, University of Minnesota, 2002. http://www.cs.umn.edu/kuram/papers/fsg-long.pdf. http://citeseer.ist.psu.edu/kuramochi02efficient.html
13. Lin S., Chalupsky H., *Unsupervised Link Discovery in Multi-Relational Data via Rarity Analysis*, Proceedings of the Third IEEE International Conference on Data Mining (ICDM '03), 2003.
14. Skillicorn D. B., *Detecting Related Message Traffic*, Workshop on Link Analysis, Counterterrorism, and Privacy, SIAM International Conference on Data Mining 2004, Seattle, WA, USA, August, 2004.
15. Tan, P., Kumar, V. and Srivastava, J. *Selecting the Right Interestingness Measure for Association Patterns*, in Proceedings of the Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (SIGKDD'02) 2002, Edmonton, Alberta, Canada, ACM Press.
16. Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.
17. Xu J. J., Chen H., *Using Shortest Path Algorithms to Identify Criminal Associations*, Decision Support Systems, Vol. 38, 2004, pp. 473-487.

Mining Candidate Viruses as Potential Bio-terrorism Weapons from Biomedical Literature

Xiaohua Hu[1], Illhoi Yoo[1], Peter Rumm[2], and Michael Atwood[1]

[1] College of Information Science and Technology, Drexel University, Philadelphia, PA 19104
{thu, michael.atwood}@cis.drexel.edu
{iy28, pdr26}@cis.drexel.edu
[2] School of Public Health, Drexel University, Philadelphia, PA 19104

Abstract. In this paper we present a semantic-based data mining approach to identify candidate viruses as potential bio-terrorism weapons from biomedical literature. We first identify all the possible properties of viruses as search key words based on Geissler's 13 criteria; the identified properties are then defined using MeSH terms. Then, we assign each property an importance weight based on domain experts' judgment. After generating all the possible valid combinations of the properties, we search the biomedical literature, retrieving all the relevant documents. Next our method extracts virus names from the downloaded documents for each search keyword and identifies the novel connection of the virus according to these 4 properties. If a virus is found in the different document sets obtained by several search keywords, the virus should be considered as suspicious and treated as candidate viruses for bio-terrorism. Our findings are intended as a guide to the virus literature to support further studies that might then lead to appropriate defense and public health measures.

1 Introduction

The threat of bio-terrorism is real. The anthrax mail attack in October, 2001 terrorism caused 23 cases of anthrax-related illness and 5 deaths. The threat of the use of biological weapons against public is more acute than any time in U.S. history due to the widespread availability of biological/chemical agents, widespread knowledge of production methodologies, and potential dissemination devices. Therefore, the discovery of additional viruses as bio-terrorism weapon and preparedness for this threat is seemingly vital to the public health and home land security.

Because it is very difficult for laypeople to diagnose and recognize most of the diseases caused by biological weapons, we need surveillance systems to keep an eye on potential uses of such biological weapons [2]. Before initiating such systems, we should identify what biological agents could be used as biological weapons. Geissler identified and summarized 13 criteria (shown in Table 1) to identify biological warfare agents as viruses [6]. Based on the criteria, he compiled 21 viruses. Figure 1 lists the 21 virus names in MeSH terms. The viruses in Figure 1 meet some of the criteria described in Table 1.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 60 – 71, 2005.

Table 1. Geissler's 13 Criteria for Viruses

1	The agent should consistently produce a given effect: death or disease.
2	The concentration of the agent needed to cause death or disease the infective dose should be low.
3	The agent should be highly contagious.
4	The agent should have a short and predictable incubation time from exposure to onset of the disease symptoms.
5	The target population should have little or no natural or acquired immunity or resistance to the agent.
6	Prophylaxis against the agent should not be available to the target population.
7	The agent should be difficult to identify in the target population, and little or no treatment for the disease caused by the agent should be available.
8	The aggressor should have means to protect his own forces and population against the agent clandestinely.
9	The agent should be amenable to economical mass production.
10	The agent should be reasonably robust and stable under production and storage conditions, in munitions and during transportation. Storage methods should be available that prevent gross decline of the agent's activity.
11	The agent should be capable of efficient dissemination. If it cannot be delivered via an aerosol, living vectors (e.g. fleas, mosquitoes or ticks) should be available for dispersal in some form of infected substrate.
12	The agent should be stable during dissemination. If it is to be delivered via an aerosol, it must survive and remain stable in air until it reaches the target population.
13	After delivery, the agent should have low persistence, surviving only for a short time, thereby allowing a prompt occupation of the attacked area by the aggressor's troops

- Hemorrhagic Fever Virus, Crimean-Congo
- Lymphocytic choriomeningitis virus
- Encephalitis Virus, Venezuelan Equine
- Encephalitis Virus, Western Equine
- Encephalitis Virus, Eastern Equine
- Encephalitis Virus, Japanese
- Encephalitis Viruses, Tick-Borne
- Encephalitis Virus, St. Louis
- Arenaviruses, New World
- Marburg-like Viruses
- Rift Valley fever virus
- Yellow fever virus
- Chikungunya virus
- Dengue Virus
- Ebola-like Viruses
- Hantaan virus
- Hepatitis A virus
- Orthomyxoviridae
- Junin virus
- Lassa virus
- Variola virus

Fig. 1. Geissler's 21 Viruses

Based on the criteria, government agencies such as CDC and the Department of Homeland Security compile and monitor viruses which are known to be dangerous in

bio-terrorism. One problem of this approach is that the list is compiled manually, requiring extensive specialized human resources and time. Because the viruses are evolving through mutations, biological or chemical change, some biological substances have the potential to turn into deadly virus through chemical/genetic/biological reaction, there should be an automatic approach to keep track of existing suspicious viruses and to discover new viruses as potential weapons. We expect that it would be very useful to identify those biological substances and take precaution actions or measurements.

2 Related Works

The problem of mining implicit knowledge/information from biomedical literature was exemplified by Dr. Swanson's pioneering work on Raynaud disease/fish-oil discovery in 1986 [11]. Back then, the Raynaud disease had no known cause or cure, and the goal of his literature-based discovery was to uncover novel suggestions for how Raynaud disease might be caused, and how it might be treated. He found from biomedical literature that Raynaud disease is a peripheral circulatory disorder aggravated by high platelet aggregation, high blood viscosity and vasoconstriction. In another separate set of literature on fish oils, he found out the ingestion of fish oil can reduce these phenomena. But no single article from both sets in the biomedical literature mentions Raynaud and fish oil together in 1986. Putting these two separate literatures together, Swanson hypothesized that fish oil may be beneficial to people suffering from Raynaud disease [11][12]. This novel hypothesis was later clinically confirmed by DiGiacomo in 1989 [4]. Later on [10] Dr. Swanson extended his methods to search literature for potential virus. But the biggest limitation of his methods is that, only 3 properties/criteria of a virus are used as search key word and the semantic information is ignored in the search procedure. In this paper, we present a novel biomedical literature mining algorithms based on this philosophy with significant extensions. Our objective is to extend the existing known virus list compiled by CDC to other viruses that might have similar characteristics. We hypothesize, therefore, that viruses that have been researched with respect to the characteristics possessed by existing viruses are leading candidates for extending the virus lists. Our findings are intended as a guide to the virus literature to support further studies that might then lead to appropriate defense and public health measures.

3 Method

We propose an automated, semantic-based data mining system to identify viruses that can be used as potential weapons in bio-terrorism. Following the criteria established by Geissler and the similar ideas used by Swanson [10], in the mining procedure, we consider many important properties of the virus such as *the genetic aspects of virulence*; *airbone transmission of viral disease*; and *stability of viruses in air or aerosol mixtures* etc.. Our objective is to identify which viruses have been investigated with respect to these properties. The main assumption of the proposed approach is that the more criteria are met by a virus, the more suspicious the virus is a potential candidate

for bio-terrorism. In other words, if a virus is commonly found in the different document sets searched by several search keywords, the virus should be considered as suspicious.

We introduce an automated semantic-based search system, called Combinational Search based Virus Seeker (CSbVS), to identify viruses that can be used as potential weapons in bio-terrorism. The method is based on Dr. Swanson's method with the following enhancements:

(1) Search keywords (SK) are more complete based on Dr. Geissler's 13 criteria.
(2) The importance of search key words are reflected by different weighs based on the properties of the virus.
(3) In [10], only 3 properties/criteria of a virus are used as search key word, we consider all the meaningful combinations of the properties/criteria of the virus. And different search keywords have different weight; if a virus is found to meet the criteria in many search keywords, the virus is more suspicious. Therefore, the result is more reliable. Each virus has its own score so that the viruses can be ranked while Swanson just listed the viruses without any ranking.

In order to find all the suspicious viruses in the biomedical literature, we first identify all the possible properties of viruses as search key words based on Geissler's 13 criteria; the identified terms are then defined in MeSH terms (a biomedical ontology developed by the National Library of Medicine, http://www.nlm.nih.gov/mesh/meshhome.html). These properties are shown in Figure 2. Then, we assign each

- "Virulence"[MeSH]
- "Disease Outbreaks"[MeSH]
- "Viral Nonstructural Proteins"[MeSH]
- "Cross Reactions"[MeSH]
- "Mutation"[MeSH] AND "Virus Replication"[MeSH]
- "Insect Vectors"[MeSH]
- "severe acute"
- cause OR causing
- mortality
- death AND disease
- encephalitis OR encephalomyelitis
- epidemics OR epidemiologically
- etiologic
- fatal
- febrile
- fever
- hemorrhagic
- infect OR infecting
- mosquito-borne
- transmission OR transmit
- survive
- viability OR viable
- airborne

Fig. 2. The Properties/Criteria of Suspicious Viruses

property an importance weight based on domain experts' judgment. For example, it is believed that "virulence" as MeSH term is much more important than "cause" in searching the potential virus. Therefore, for each search keyword, a weight is given based on the importance; this is the domain knowledge, which may lead to better results to identify suspicious virus. After generating all the possible valid combinations

of the properties, CSbVS performs the searches for each combination, retrieving all the relevant documents. Each combination has its importance, which is the sum of the weights of the key words used in the combination. Next CSbVS extracts virus names from the downloaded documents for each search keyword and identifies the novel connection of the virus with these properties. The viruses, as a result of each search combination, have the same importance as the search combination. Based on the importance, all the viruses are ranked. Figure 3 shows the data flow of CSbVS.

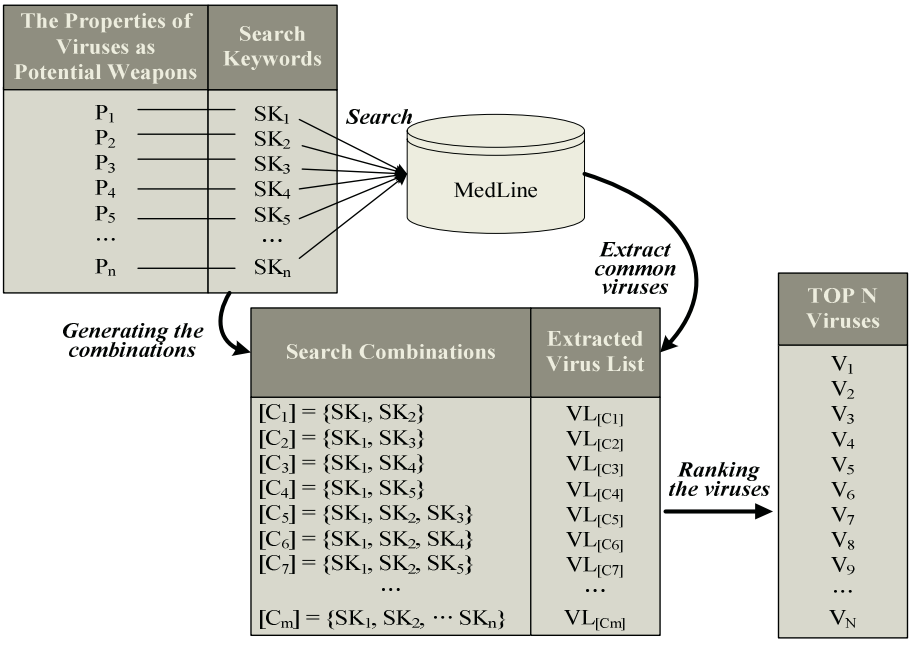


Fig. 3. The Data Flow of CSbVS

Figure 4 shows an example of the combinational search. Each circle (e.g., $Virus_{sk1}$) indicates a virus list from the documents by a search keyword (e.g., SK1). Each

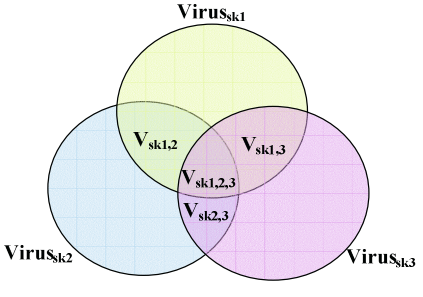


Fig. 4. An Example of Combinational Search

intersection (e.g., $V_{sk1,2}$) contains the viruses that are commonly found in the original document sets. Therefore, we can guess the viruses in $V_{sk1,2,3}$ are more suspicious than the viruses in $V_{sk1,2}$.

Input: Search keywords with their weights
All virus names in MeSH terms
Output: the top N viruses
Procedural
STEP 1: Generating all the possible valid combinations of the search keywords.
STEP 2: Searching every keyword and extracting virus names based on the virus category in the MeSH hierarchy in the downloaded documents for each search keyword
STEP 3: Finding common viruses in various search combination
STEP 4: Accumulate the scores of the common viruses with the sum of the weights of the search keywords involved
STEP 5: Sorting all viruses based on their accumulated scores in descending order.

Fig. 5. The Algorithm of CSbVS

STEP 1: Generating all the possible valid combinations of search keywords. For example, if there are 4 search keywords (e.g., A, B, C, D), all the possible valid combinations used in the approach are the following. {AB, ABC, ABCD, AC, AD, BCD, BD, CD, ABD}

Here we assume that a virus has to meet at least two criteria to be considered as a potential virus for bio-terrorism. The combinations that consist of only a single criterion are not considered.

STEP 2: Searching every keyword against Medline (PubMed) and download the documents relevant to each search keyword. For better recall and precision, we included "Viruses" and "Human" as MeSH terms into the combinational search. For example, for the "Virulence" search keyword, the complete search keyword against PubMed is the following

"Virulence"[MeSH] AND "Viruses"[MAJOR] AND "Human"[MeSH]

After downloading the relevant documents, the system extracts virus names from these documents for each search keyword; the targets of the extraction are *MajorTopic* virus names assigned to Medline articles. In order to identify virus names, we collected all the MeSHs in B04 category ("Viruses") of MeSH Categories that are shown Appendix 1; the total number of MeSH terms in the category is 487.

STEP 3: Finding common viruses in each combination. If a combination consists of A, B & C as search key, the virus list contains the viruses that are "commonly" found is in every document set by the search key as shown in Figure 6.

STEP 4: Accumulate the scores of the common viruses based on the weights of the search keywords involved.

STEP 5: Sorting all viruses based on their accumulated scores in descending order. Finally the top N viruses are the output.

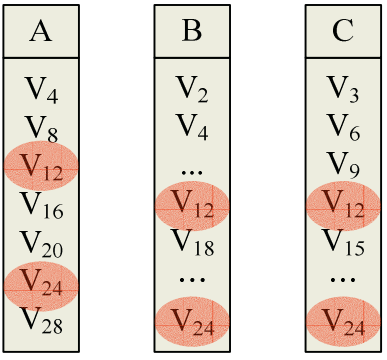


Fig. 6. Viruses commonly found in A, B, and C document sets

4 Experimental Results

In our experiment, first of all, a weight (1 to 5) is carefully assigned to each property based on domain expert's opinion. Then, CSbVS download the documents relevant to each keyword from Medline and extracted virus names for each combination of search keywords. Table 2 shows the number of documents for each search keyword. After searching every search keyword against Medline and generating all the possible valid combination of search keywords, common virus names are extracted for each combination search. Finally, based on the importance of the combinational search, all the viruses are ranked. Table 3 shows the top 143 suspicious viruses; which included all the 21 viruses identified by Geissler (marked in bold.)

Table 2. The Search Keywords and the number of Documents by them

Search Keywords	# of Doc.
Virulence[MeSH]	1455
Disease Outbreaks[MeSH]	3141
Viral Nonstructural Proteins[MeSH]	1262
Cross Reactions[MeSH]	1559
("Mutation"[MeSH] AND "Virus Replication"[MeSH])	1742
Insect Vectors[MeSH]	413
severe acute	487
(cause OR causing)	5907
mortality	3279
(death AND disease)	1052
(encephalitis OR encephalomyelitis)	4398
epidemics OR epidemiologically)	17825
etiologic	988
fatal	1372
febrile	746

Table 2. (*Continued…*)

Search Keywords	# of Doc.
fever	3629
hemorrhagic	1412
(infect OR infecting)	2883
mosquito-borne	98
(transmission OR transmit)	11166
survive	196
(viability OR viable)	1081
airborne	61
total	**66152**

Table 3. The top 143 suspicious viruses

Ranking	Virus Name in MeSH
1	Hepacivirus
2	West Nile virus
3	***Dengue Virus***
4	***Encephalitis Viruses, Tick-Borne***
5	Hantavirus
6	Bunyaviridae
7	Vaccinia virus
8	Herpesvirus 3, Human
9	Enterovirus
10	Respiratory Syncytial Viruses
11	Adenoviruses, Human
12	Cytomegalovirus
13	Adenoviridae
14	Influenza A Virus, Human
15	Herpesvirus 4, Human
16	Enterovirus B, Human
17	Influenza A virus
18	Herpesvirus 2, Human
19	Herpesviridae
20	Herpesvirus 1, Human
21	Simplexvirus
22	HIV-1
23	***Ebola-like Viruses***
24	Human T-lymphotropic virus 1
25	***Encephalitis Virus, Venezuelan Equine***
26	Herpesvirus 6, Human
27	Rotavirus
28	Sindbis Virus
29	Respirovirus
30	Flavivirus
31	***Yellow fever virus***
32	Arboviruses
33	Encephalitis Viruses
34	Herpesvirus 8, Human
35	***Orthomyxoviridae***
36	Polioviruses
37	Bunyamwera virus
38	Rabies virus
39	Influenza B virus
40	HIV
41	Respiratory Syncytial Virus, Human
42	Measles virus
43	Alphavirus
44	***Encephalitis Virus, Japanese***
45	Deltaretrovirus
46	Parvovirus B19, Human
47	Picornaviridae
48	RNA Viruses
49	Reassortant Viruses
50	SARS Virus

Table 3. (*Continued…*)

Ranking	Virus Name in MeSH	Ranking	Virus Name in MeSH
51	***Hantaan virus***	88	Bacteriophages
52	Influenza A Virus, Avian	89	Parvoviridae
53	Flaviviridae	90	Newcastle disease virus
54	Retroviridae	91	Endogenous Retroviruses
55	***Rift Valley fever virus***	92	Rubella virus
56	Norovirus	93	Papillomavirus
57	Vesicular stomatitis-Indiana virus	94	Coliphages
		95	Semliki forest virus
58	Hepatitis A Virus, Human	96	HIV-2
59	Hepatovirus	97	Hepatitis Viruses
60	Virion	98	***Lymphocytic choriomeningitis virus***
61	Mumps virus		
62	Influenza A Virus, Porcine	99	Proviruses
63	Poliovirus	100	Coronaviridae
64	Rhinovirus	101	Caliciviridae
65	Morbillivirus	102	***Encephalitis Virus, Eastern Equine***
66	Papillomavirus, Human		
67	SIV	103	Herpesvirus 7, Human
68	Paramyxovirinae	104	Salmonella Phages
69	***Hemorrhagic Fever Virus, Crimean-Congo***	105	Lentivirus
		106	Lyssavirus
70	Parainfluenza Virus 3, Human	107	Echovirus 9
71	Poxviridae	108	Parainfluenza Virus 1, Human
72	Hepatitis Delta Virus	109	Distemper Virus, Canine
73	***Arenaviruses, New World***	110	Encephalomyocarditis virus
74	***Encephalitis Virus, St. Louis***	111	Simian virus 40
75	Astrovirus	112	Metapneumovirus
76	Arenaviridae	113	Norwalk virus
77	Hepatitis E virus	114	***Chikungunya virus***
78	Oncogenic Viruses	115	Aphthovirus
79	JC Virus	116	Ross river virus
80	DNA Viruses	117	Viruses, Unclassified
81	***Lassa virus***	118	SSPE Virus
82	***Marburg-like Viruses***	119	Filoviridae
83	Rhabdoviridae	120	Monkeypox virus
84	Reoviridae	121	Herpesvirus 1, Cercopithecine
85	Coronavirus	122	Encephalitis Virus, Murray Valley
86	Defective Viruses		
87	Polyomavirus	123	Theilovirus

Table 3. (*Continued…*)

Ranking	Virus Name in MeSH	Ranking	Virus Name in MeSH
124	BK Virus	133	***Hepatitis A virus***
125	Orthopoxvirus	134	Papillomaviridae
126	***Variola virus***	135	Echovirus 6, Human
127	Paramyxoviridae	136	Leukemia Virus, Murine
128	Murine hepatitis virus	137	Phlebovirus
129	Borna disease virus	138	Muromegalovirus
130	Transfusion-Transmitted Virus	139	Baculoviridae
131	***Encephalitis Virus, Western Equine***	140	Parvovirus
132	Dependovirus	141	Coronavirus OC43, Human
		142	Herpesvirus 1, Suid

Although Geissler's 21 viruses, compiled in 1986, would not be the full list of the viruses used as potential weapons at present, we compare our Top 50, 100 & 143 Viruses with Geissler's 21 viruses as a golden standard in terms of recall and precision; all of the Geissler's 21 viruses are found within our top 143 viruses. As Figure 7 shows, the recalls are consistent for the three groups. In other words, Geissler's 21 viruses are equally distributed in our Top 143 virus list. It is very important to note that our system is able to find "West Nile virus" and "SARS Virus", and ranks them in 2nd and 50th respectively.

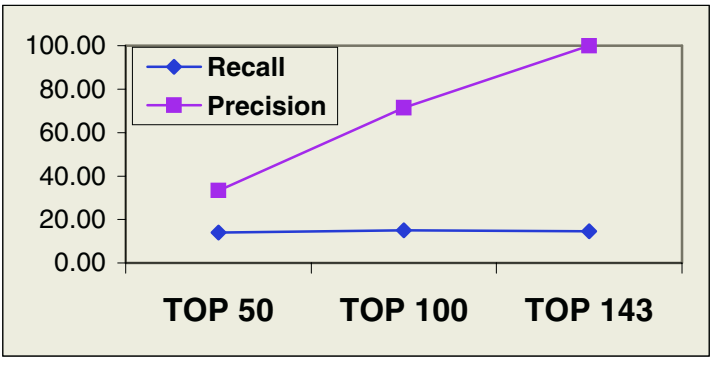


Fig. 7. Recalls and Precisions for Top 50, 100 and 143

5 Potential Significance for Public Health and Homeland Security

As such a list shows there are many potential viral threats that could affect the health of the public on a wide scale if disseminated effectively. This situation is worrisome to public health officials who are concerned that the public health system might not yet be prepared fully for such a crisis as a release of a viral agent in the U.S. population. Certainly, there have been steps made in laboratory preparedness and public

health preparedness to identify such threats but potential gaps remain [1][2][6][7][8]. These viruses vary in their biological capability to survive, replicate, and be infective. The U.S. Department of Health and Human Services Agency for Health Care Research and Quality working with University of Alabama has recently put out a list of what they think are the most probable public health biological threats with following caveats [13]:

The U.S. public health system and primary health-care providers must be prepared to address varied biological agents, including pathogens that are rarely seen in the United States. High-priority agents include organisms that pose a risk to national security because they

- can be easily disseminated or transmitted person-to-person
- cause high mortality, with potential for major public health impact
- might cause public panic and social disruption; and require special action for public health preparedness

The Category A viral agents that most fit this bill currently according to AHRQ, the CDC, and the University of Alabama (and other experts) are: Smallpox, viral hemorrhaghic agents (there are many of these – see below), SARS and Monkeypox.

Besides Smallpox, which is known to have stores in secured locations in the U.S. and Russia but may be in other sites [3], most of the literature currently focuses on the usage of hemorrhagic viruses as the most probably viral bio-terrorism threats.

Other viruses which our federal government conceives be used as terrorist weapons on a population scale, but not considered as great a threat as the filoviruses include:

Arenaviruses: Lassa Fever (Africa) and the New World Hemorrhagic Fevers - Bolivian Hemorrhagic Fever (BHF, Machupo virus), Argentine Hemorrhagic Fever (AHF, Junin virus), Venezuelan Hemorrhagic Fever (Guanarito virus), and Brazilian Hemorrhagic Fever (Sabia virus)
Bunyaviruses: Crimean-Congo Hemorrhagic Fever (CCHF), Rift Valley Fever (RVF)
Flaviviruses: Dengue, Yellow Fever, Omsk Hemorrhagic Fever, and Kyasanur Forest disease [9]

That being said, there are other potential viruses that this data search have identified such as rabies, which is a highly infective agent that if introduced into the food chain, although perhaps, not infective would certainly be likely to cause panic.

Others such as adenovirus which has caused huge outbreaks in susceptible military populations could conceivably be more of a disabling virus that also affected populations [5], while HantaVirus has been associated with recent outbreaks in the United States [7].

Therefore, it is hard to discount completely that in some form that most of the viruses on this list could at least create fear and panic in populations, simply by their introduction – we only need to look at the recent shortage of influenza vaccine to see that populations may not behave rationally in regards to risk when dealing with infectious diseases. Therefore, such a list may at least remind us that there are other viral agents that potentially cause disease and/or terror in populations as well as those commonly known groups.

References

1. The Association of State and Territorial Health Officials (ASTHO). Public Health Preparedness: A Progress Report – First Six Months (ATAIP Indicators Project) (2003)
2. Büchen-Osmond C. Taxonomy and Classification of Viruses. In: Manual of Clinical Microbiology, 8th ed, Vol 2, p. 1217-1226, ASM Press, Washington DC (2003)
3. Diaz, Rumm et. al. National Advisory Committee on Children and Terrorism – Report to the Secretary of DHHS (2003)
4. DiGiacome, R.A, Kremer, J.M. and Shah, D.M. Fish oil dietary supplementation is patients with Raynaud's phenomenon: A double-blind, controlled, prospective study, American Journal of Medicine, 8, 1989, 158-164.
5. Frist B. When Every Moment Counts – What You Need to Know About Bio-terrorism from the Senate's Only Doctor, Rowman and Littlefield (2002)
6. Geissler, E. (Ed.), Biological and toxin weapons today, Oxford, UK: SIPRI (1986)
7. Gray GC, Callahan JD, Hawksworth AK, Fisher CA, and Gaydos JC. Respiratory diseases among U.S. military personnel: countering emerging threats. Emerging Infectious Disease, Vol 5(3): 379-87 (1999)
8. Gursky EA, Drafted to Fight Terror, U.S. Public Health on the Front Lines of Biological Defense, ANSER (2004)
9. Lane SP, Beugelsdijk T, and Patel CK. FirePower in the Lab – Automation in the Fight Against Infectious Diseases and Bioterrorism, John Henry Press, DC (1999)
10. Swanson, DR, Smalheiser NR, & Bookstein A. Information discovery from complementary literatures: categorizing viruses as potential weapons. JASIST 52(10): 797-812 (2001)
11. Swanson, DR., Fish-oil, Raynaud's Syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine 30(1), 7-18 (1986)
12. Swanson, DR., Undiscovered public knowledge. Libr. Q. 56(2), pp. 103-118 (1986)
13. Web site updated regularly by the Agency for Health Care Research and Quality, US DHHS on bioterrorism and emerging infectious disease agents accessible at: http://www.bioterrorism.uab.edu/EIPBA/vhf/summary.html

Private Mining of Association Rules

Justin Zhan, Stan Matwin, and LiWu Chang

[1] School of Information Technology & Engineering,
University of Ottawa, Canada
[2] School of Information Technology & Engineering,
University of Ottawa, Canada
{zhizhan, stan}@site.uottawa.ca
[3] Center for High Assurance Computer Systems,
Naval Research Laboratory, USA
lchang@itd.nrl.navy.mil

Abstract. This paper introduces a new approach to a problem of data sharing among multiple parties, without disclosing the data between the parties. Our focus is data sharing among two parties involved in a data mining task. We study how to share private or confidential data in the following scenario: two parties, each having a private data set, want to collaboratively conduct association rule mining without disclosing their private data to each other or any other parties. To tackle this demanding problem, we develop a secure protocol for two parties to conduct the desired computation. The solution is distributed, i.e., there is no central, trusted party having access to all the data. Instead, we define a protocol using homomorphic encryption techniques to exchange the data while keeping it private. All the parties are treated symmetrically: they all participate in the encryption and in the computation involved in learning the association rules.

Keywords: Privacy, security, association rule mining.

1 Introduction

In this paper, we address the following problem: two parties are cooperating on a data-rich task. Each of the parties owns data pertinent to the aspect of the task addressed by this party. More specifically, the data consists of instances, and all parties have data about all the instances involved, but each party has its own view of the instances - each party works with its own attribute set. The overall performance, or even solvability, of this task depends on the ability of performing data mining using all the attributes of all the parties. The two parties, however, may be unwilling to release their attributes to other parties that are not involved in collaboration, due to privacy or confidentiality of the data. How can we structure information sharing between the parties so that the data will be shared for the purpose of data mining, while at the same time specific attribute values will be kept confidential by the parties to whom they

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 72–80, 2005.

belong? This is the task addressed in this paper. In the privacy-oriented data mining this task is known as data mining with vertically partitioned data (also known as heterogeneous collaboration [8].) Examples of such tasks abound in business, homeland security, coalition building, medical research, etc.

Without privacy concerns, all parties can send their data to a trusted central place to conduct the mining. However, in situations with privacy concerns, the parties may not trust anyone. We call this type of problem the *Privacy-preserving Collaborative Data Mining problem*. As stated above, in this paper we are interested in heterogeneous collaboration where each party has different sets of attributes [8].

Data mining includes a number of different tasks, such as association rule mining, classification, and clustering. This paper studies the association rule mining problem. The goal of association rule mining is to discover meaningful association rules among the attributes of a large quantity of data. For example, let us consider the database of a medical study, with each attribute representing a characteristic of a patient. A discovered association rule pattern could be "70% of patients who suffer from medical condition C have a gene G". This information can be useful for the development of a diagnostic test, for pharmaceutical research, etc. Based on the existing association rule mining technologies, we study the *Private Mining of Association Rules* problem defined as follows: two parties want to conduct association rule mining on a data set that consists all the parties' private data, but neither party is willing to disclose her raw data to each other or any other parties. In this paper, we develop a protocol, based on homomorphic cryptography, to tackle the problem.

The paper is organized as follows: The related work is discussed in Section 2. We describe the association rule mining procedure in Section 3. We then present our proposed secure protocols in Section 4. We give our conclusion in Section 5.

2 Related Work

2.1 Secure Multi-party Computation

A Secure Multi-party Computation (SMC) problem deals with computing any function on any input, in a distributed network where each participant holds one of the inputs, while ensuring that no more information is revealed to a participant in the computation than can be inferred from that participant's input and output. The SMC problem literature was introduced by Yao [13]. It has been proved that for any polynomial function, there is a secure multi-party computation solution [7]. The approach used is as follows: the function F to be computed is firstly represented as a combinatorial circuit, and then the parties run a short protocol for every gate in the circuit. Every participant gets corresponding shares of the input wires and the output wires for every gate. This approach, though appealing in its generality and simplicity, is highly impractical for large datasets.

2.2 Privacy-Preserving Data Mining

In early work on privacy-preserving data mining, Lindell and Pinkas [9] propose a solution to privacy-preserving classification problem using oblivious transfer protocol, a powerful tool developed by secure multi-party computation (SMC) research. The techniques based on SMC for efficiently dealing with large data sets have been addressed in [4, 8].

Random perturbation-based approaches were firstly proposed by Agrawal and Srikant in [3] to solve privacy-preserving data mining problem. In addition to perturbation, aggregation of data values [11] provides another alternative to mask the actual data values. In [1], authors studied the problem of computing the kth-ranked element. Dwork and Nissim [5] showed how to learn certain types of boolean functions from statistical databases in the context of probabilistic implication for the disclosure of statistics.

Homomorphic encryption [10], which transforms multiplication of encrypted plaintexts into the encryption of the sum of the plaintexts, has recently been used in secure multi-party computation. For instance, Freedmen, Nissim and Pinkas [6] applied it to set intersection. The work most related to ours is [12], where Wright and Yang applied homomorphic encryption to the Bayesian networks induction for the case of two parties. However, the core protocol which is called *Scalar Product Protocol* can be easily attacked. In their protocol, since Bob knows the encryption key e, when Alice sends her encrypted vector $(e(a_1), \cdots, e(a_n))$, where a_is are Alice's vector elements, Bob can easily figure out whether a_i is 1 or 0 through the following attack: Bob computes $e(1)$, and then compares it with $e(a_i)$. If $e(1) = e(a_i)$, then $a_i = 1$, otherwise $a_i = 0$. In this paper, we develop a secure two-party protocol based on homomorphic encryption. Our contribution not only overcomes the attacks which exist in [12], but applies our secure protocol to tackle collaborative association rule mining problems.

3 Mining Association Rules on Private Data

Since its introduction in 1993 [2], the association rule mining has received a great deal of attention. It is still one of most popular pattern-discovery methods in the field of knowledge discovery. Briefly, an association rule is an expression $X \Rightarrow Y$, where X and Y are sets of items. The meaning of such rules is as follows: Given a database D of transactions, $X \Rightarrow Y$ means that whenever a transaction R contains X then R also contains Y with certain confidence. The rule confidence is defined as the percentage of transactions containing both X and Y with regard to the overall number of transactions containing X. The fraction of transactions R supporting an item X with respect to database D is called the support of X.

3.1 Problem Definition

We consider the scenario where two parties, each having a private data set (denoted by D_1 and D_2 respectively), want to collaboratively conduct association rule mining on the concatenation of their data sets. Because they are concerned

about their data privacy, neither party is willing to disclose its raw data set to the other. Without loss of generality, we make the following assumptions about the data sets (the assumptions can be achieved by pre-processing the data sets D_1 and D_2, and such a pre-processing does not require one party to send her data set to other party): (1) D_1 and D_2 contain the same number of transactions. Let N denote the total number of transactions for each data set. (2) The identities of the ith (for $i \in [1, N]$) transaction in D_1 and D_2 are the same.

Private Mining of Association Rule Problem: Party 1 has a private data set D_1, party 2 has a private data set D_2. The data set $[D_1 \cup D_2]$ forms a database, which is actually the concatenation of D_1 and D_2 (by putting D_1 and D_2 together so that the concatenation of the ith row in D_1 and D_2 becomes the ith row in $[D_1 \cup D_2]$). The two parties want to conduct association rule mining on $[D_1 \cup D_2]$ and to find the association rules with support and confidence being greater than the given thresholds. We say an association rule (e.g., $x_i \Rightarrow y_j$) has confidence $c\%$ in the data set $[D_1 \cup D_2]$ if in $[D_1 \cup D_2]$ $c\%$ of the transactions which contain x_i also contain y_j (namely, $c\% = P(y_j \mid x_i)$). We say that the association rule has support $s\%$ in $[D_1 \cup D_2]$ if $s\%$ of the transactions in $[D_1 \cup D_2]$ contain both x_i and y_j (namely, $s\% = P(x_i \cap y_j)$). Consequently, in order to learn association rules, one must compute the candidate itemsets, and then prune those that do not meet the preset confidence and support thresholds. In order to compute confidence and support of a given candidate itemset, we must compute, for a given itemset C, the frequency of attributes (items) belonging to C in the entire database (i.e., we must count how many attributes in C are present in all transactions of the database, and divide the final count by the size of the database which is N.). Note that association rule mining works on binary data, representing presence or absence of items in transactions. However, the proposed approach is not limited to the assumption about the binary character of the data in the content of association rule mining.

3.2 Association Rule Mining Procedure

The following is the procedure for mining association rules on $[D_1 \cup D_2]$.

1. L_1 = large 1-itemsets
2. **for** (k = 2; $L_{k-1} \neq \phi$; k++) **do begin**
3. C_k = **apriori-gen**(L_{k-1})
4. **for** all candidates $c \in C_k$ **do begin**
5. **Compute *c.count*** [1]
6. **end**
7. $L_k = \{c \in C_k | c.count \geq min\text{-}sup\}$
8. **end**
9. Return L = $\cup_k L_k$

[1] *c.count* divided by the total number of transactions is the support of a given item set. We will show how to compute it in Section 3.3.

The procedure **apriori-gen** is described in the following (please also see [2] for details).

apriori-gen(L_{k-1}: large (k-1)-itemsets)

1. insert into C_k
2. select $p.item_1$, $p.item_2$, $\cdots$, $p.item_{k-1}$,$q.item_{k-1}$
3. from L_{k-1} p, L_{k-1} q
4. where $p.item_1 = q.item_1, \cdots, p.item_{k-2} = q.item_{k-2}, p.item_{k-1} < q.item_{k-1}$;

Next, in the *prune* step, we delete all itemsets $c \in C_k$ such that some (k-1)-subset of c is not in L_{k-1}:

1. for all itemsets $c \in C_k$ do
2. for all (k-1)-subsets s of c do
3. if($s \notin L_{k-1}$) then
4. delete c from C_k;

3.3 How to Compute *c.count*

In the procedure of association rule mining, the only steps accessing the actual data values are: (1) the initial step which computes large 1-itemsets, and (2) the computation of *c.count*. Other steps, particularly computing candidate itemsets, use merely attribute names. To compute large 1-itemsets, each party selects her own attributes that contribute to large 1-itemsets. As only a single attribute forms a large 1-itemset, there is no computation involving attributes of the other party. Therefore, no data disclosure across parties is necessary. However, to compute *c.count*, a computation accessing attributes belonging to different parties is necessary. How to conduct this computations across parties without compromising each party's data privacy is the challenge we address.

Alice
1
0
1
1
1

Bob
1
1
1
1
0

Fig. 1. Raw Data For Alice and Bob

If all the attributes belong to the same party, then *c.count*, which refers to the frequency counts for candidates, can be computed by this party. If the attributes belong to different parties, they then construct vectors for their own attributes

and apply our secure protocol, which will be discussed in Section 4, to obtain *c.count*. We use an example to illustrate how to compute *c.count*. Alice and Bob construct vectors C_{k1} and C_{k2} for their own attributes respectively. To obtain *c.count*, they need to compute $\sum_{i=1}^{N}(C_{k1}[i] \cdot C_{k2}[i])$ where N is the total number of values in each vector. For instance, if the vectors are as depicted in Fig.1, then $\sum_{i=1}^{N}(C_{k1}[i] \cdot C_{k2}[i]) = \sum_{i=1}^{5}(C_{k1}[i] \cdot C_{k2}[i]) = 3$. We provide a secure protocol in Section 4 for the two parties to compute this value without revealing their private data to each other.

4 Collaborative Association Rule Mining Protocol

How the collaborative parties jointly compute *c.count* without revealing their raw data to each other presents a great challenge. In this section, we develop a secure protocol to compute *c.count* between two parties.

4.1 Introducing Homomorphic Encryption

In our secure protocols, we use homomorphic encryption [10] keys to encrypt the parties' private data. In particular, we utilize the following characterizer of the homomorphic encryption functions: $e(a_1) \times e(a_2) = e(a_1 + a_2)$ where e is an encryption function; a_1 and a_2 are the data to be encrypted. Because of the property of associativity, $e(a_1 + a_2 + .. + a_n)$ can be computed as $e(a_1) \times e(a_2) \times \cdots \times e(a_n)$ where $e(a_i) \neq 0$. That is

$$e(a_1 + a_2 + \cdots + a_n) = e(a_1) \times e(a_2) \times \cdots \times e(a_n) \tag{1}$$

4.2 Secure Two-Party Protocol

Let's assume that Alice has a vector A_1 and Bob has a vector A_2. Both vectors have N elements. We use A_{1i} to denote the *ith* element in vector A_1, and A_{2i} to denote the *ith* element in vector A_2. In order to compute the *c.count* of an itemset containing A_1 and A_2, Alice and Bob need to compute the scalar product between A_1 and A_2.

Firstly, one of parties is randomly chosen as a key generator. For simplicity, let's assume Alice is selected as the key generator. Alice generates an encryption key (e) and a decryption key (d). She applies the encryption key to the addition of each value of A_1 and $R_i * X$ (e.g., $e(A_{1i} + R_i * X)$), where R_i is a random integer and X is an integer which is greater than N. She then sends $e(A_{1i} + R_i * X)$s to Bob. Bob computes the multiplication $\prod_{j=1}^{n}[e(A_{1j} + R_i * X) \times A_{2j}]$ when $A_{2j} = 1$ (since when $A_{2j} = 0$, the result of multiplication doesn't contribute to the *c.count*). He sends the multiplication results to Alice who computes $[d(e(A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X)]) mod X = (A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X) mod X$ and obtains the *c.count*. In more detail, Alice and Bob follow the following protocol:

Protocol 1 *(Secure Two-Party Protocol)*

1. Alice generates a cryptographic key pair (d, e) of a homomorphic encryption scheme. Let's use $e(.)$ denote encryption and $d(.)$ denote decryption. Let X be an integer number which is chosen by Alice and greater than N (i.e., the number of transactions).
2. Alice randomly generates an integer number R_1 and sends $e(A_{11} + R_1 * X)$ to Bob.
3. Bob computes $e(A_{11} + R_1 * X) * A_{21}$.
4. Repeat Step 2 - 3 until Bob gets $E_1 = e(A_{11} + R_1 * X) * A_{21}$, $E_2 = e(A_{12} + R_2 * X) * A_{22}$, $\cdots$ and $E_N = e(A_{1N} + R_N * X) * A_{2N}$. Since A_{2i} is either 1 or 0, $e(A_{1i} + R_i * X) * A_{2i}$ is either $e(A_{1i} + R_i * X)$ or 0. Note that R_1, R_2, $\cdots$, and R_N are unrelated random numbers.
5. Bob multiplies all the E_is for those A_{2i}s that are not equal to 0. In other words, Bob computes the multiplication of all non-zero E_is, e.g., $E = \prod E_i$ where $E_i \neq 0$. Without loss of generality, let's assume only the first j elements are not equal to 0s. Bob then computes $E = E_1 * E_2 * \cdots * E_j = [e(A_{11} + R_1 * X) \times A_{21}] \times [e(A_{12} + R_2 * X) \times A_{22}] \times \cdots \times [e(A_{1j} + R_j * X) \times A_{2j}] = [e(A_{11} + R_1 * X) \times 1] \times [e(A_{12} + R_2 * X) \times 1] \times \cdots \times [e(A_{1j} + R_j * X) \times 1] = e(A_{11} + R_1 * X) \times e(A_{12} + R_2 * X) \times \cdots \times e(A_{1j} + R_j * X) = e(A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X)$ according to Eq. 1.
6. Bob sends E to Alice.
7. Alice computes $d(E) mod X$ which is equal to *c.count*.

4.3 Analysis of Two-Party Protocol

Correctness Analysis. Let us assume that both parties follow the protocol. When Bob receives each encrypted element $e(A_{1i} + R_i * X)$, he computes $e(A_{1i} + R_i) * A_{2i}$. If $A_{2i} = 0$, then *c.count* does not change. Hence, Bob computes the product of those elements whose A_{2i}s are 1s and obtains $\prod e(A_{1j} + R_j) = e(A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X)$ (note that the first j terms are used for simplicity in explanation), then sends it to Alice. After Alice decrypts it, she obtains $[d(e(A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X))] mod X = (A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X) mod X$ which is equal to the desired *c.count*. The reasons are as follows: when $A_{2i} = 1$ and $A_{1i} = 0$, *c.count* does not change; only if both A_{1i} and A_{2i} are 1s, *c.count* changes. Since $(A_{11} + A_{12} + \cdots + A_{1j}) \leq N < X$, $(A_{11} + A_{12} + \cdots + A_{1j} + (R_1 + R_2 + \cdots + R_j) * X) mod X = (A_{11} + A_{12} + \cdots + A_{1j})$. In addition, when $A_{2i} = 1$, $(A_{11} + A_{12} + \cdots + A_{1j})$ gives the total number of times that both A_{1i} and A_{2i} are 1s. Therefore, *c.count* is computed correctly.

Complexity Analysis. The bit-wise communication cost of this protocol is $\alpha(N + 1)$ where α is the number of bits for each encrypted element. The cost is approximately α times of the *optimal* cost of a two-party scalar product. The optimal cost of a scalar product is defined as the cost of conducting the product of A_1 and A_2 without privacy constraints, namely one party simply sends its data in plaintext to the other party.

The computational cost is caused by the following: (1) the generation of a cryptographic key pair; (2) the total number of N encryptions, e.g., $e(A_{1i}+R_i*X)$ where $i \in [1, N]$; (3)at most 3N-1 multiplications; (4) one decryption; (5) one modulo operation; (6) N additions.

Privacy Analysis. All the information that Bob obtains from Alice is $e(A_{11} + R_1 * X)$, $e(A_{12} + R_2 * X)$, $\cdots$ and $e(A_{1N} + R_N * X)$. Bob does not know the encryption key e, R_is, and X. Assuming the homomorphic encryption is secure, he cannot know Alice's original element values. The information that Alice obtains from Bob is $\prod[e(A_{1i} + R_i * X) * A_{2i}]$ for those that $A_{2i} = 1$. After Alice computes $[d(\prod e(A_{1i} + R_i * X) * A_{2i})] mod X$ for those that $A_{2i} = 1$, she only obtains *c.count*, and can't exactly know Bob's original element values. From symmetric point of view, we could let Alice and Bob be the key generator in turn. When computing the first half of their vector product, Alice is selected as the key generator; when computing the second half vector product, Bob is selected as the key generator.

5 Conclusion

In this paper, we consider the problem of private mining of association rules. In particular, we study how two parties can collaboratively conduct association rule mining on their joint private data. We develop a secure collaborative association rule mining protocol based on homomorphic encryption scheme. In our protocol, the parties do not need to send all their data to a central, trusted party. Instead, we use the homomorphic encryption techniques to conduct the computations across the parties without compromising their data privacy. Privacy analysis is provided. Correctness of our protocols is shown and complexity of the protocols is addressed as well. As future work, we will develop a privacy measure to quantitatively measure the privacy level achieved by our proposed secure protocols. We will also apply our technique to other data mining computations, such as secure collaborative clustering.

References

1. G. Aggarwal, N. Mishra, and B. Pinkas. Secure computation of the k th-ranked element. In *EUROCRYPT pp 40-55*, 2004.
2. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 207–216, Washington D.C., May 1993.
3. R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 439–450. ACM Press, May 2000.
4. W. Du and Z. Zhan. Building decision tree classifier on private data. In *Workshop on Privacy, Security, and Data Mining at The 2002 IEEE International Conference on Data Mining (ICDM'02)*, Maebashi City, Japan, December 9 2002.

5. C. Dwork and K. Nissim. Privacy-preserving datamining on vertically partitioned databases.
6. M. Freedman, K. Nissim, and B. Pinkas. Effiicent private matching and set intersection. In *EUROCRYPT pp 1-19*, 2004.
7. O. Goldreich. Secure multi-party computation (working draft). http://www.wisdom.weizmann.ac.il /home/oded/public_html/foc.html, 1998.
8. J.Vaidya and C.W.Clifton. Privacy preserving association rule mining in vertically partitioned data. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*.
9. Y. Lindell and B. Pinkas. Privacy preserving data mining. In *Advances in Cryptology - Crypto2000, Lecture Notes in Computer Science*, volume 1880, 2000.
10. P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptography - EUROCRYPT '99, pp 223-238, Prague, Czech Republic*, May 1999.
11. L. Sweeney. k-anonymity: a model for protecting privacy. In *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5), pp 557–570*.
12. R. Wright and Z. Yang. Privacy-preserving bayesian network structure computation on distributed heterogeneous data. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2004.
13. A. C. Yao. Protocols for secure computations. In *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science*, 1982.

Design Principles of Coordinated Multi-incident Emergency Response Systems

Rui Chen[1], Raj Sharman[1], H. Raghav Rao[1, 2], and Shambhu J. Upadhyaya[2]

[1] Department of Management of Science and Systems,
State University of New York at Buffalo,
Buffalo, NY 14260
{ruichen, rsharman, mgmtrao}@buffalo.edu
[2] Department of Computer Science and Engineering,
State University of New York at Buffalo,
Buffalo, NY 14260
shambhu@cse.buffalo.edu

Abstract. Emergency response systems play an important role in homeland security nowadays. Despite this, research in the design of emergency response systems is lacking. An effective design of emergency response system involves multi-disciplinary design considerations. On the basis of emergency response system requirement analysis, in this paper, we develop a set of supporting design concepts and strategic principles for an architecture for a coordinated multi-incident emergency response system

1 Introduction

Due to the increasing threat of terrorist attacks, the need for effective emergency response systems has been recognized worldwide. Several researchers have studied Emergency Response systems recently [1], [2], [3] and [4]. However, these approaches did not deal with multi-incident situations, nor do they consider redeployment of assets as may be needed in a multi-incident emergency.

In this paper, we start by presenting a scenario analysis that includes multiple incidents that deal with the same type of emergency. An analysis of the resource related issues crucial for multi-incident emergencies is presented. Finally, we present propositions and methodology for future research and discuss our current stage of research.

2 Emergency Scenario Description

In this section we present a conceptual and a process view of the system design principles. In order to highlight some of the critical issues in dealing with a multi-incident situation we begin by presenting a scenario of a possible emergency incident. We have adopted this approach to help us understand the complexity as well as the pitfalls in the architecture currently described in literature.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 81 – 98, 2005.

2.1 Scenario: A Chemical Attack

The threat of a chemical attack against civilian targets has increased dramatically during the last decade [8], [9], and [26]. These attacks could potentially result in catastrophic damage to soft city infrastructures such as shopping centers, subway systems, schools and airports.

In our scenario, we assume that terrorists have launched three separate and independent chemical attacks within a short interval of time in a mid-size city as illustrated in Figure 1. For the purposes of this illustration, we assume that the attacks have focused on soft targets such as a shopping mall, a college campus, and an airport. We further, assume that this attack is targeted using a parcel of radioactive chemical materials. The terrorists ignite the parcel and cause chemical fire in the above three places. Each incident of attack generates an impact zone, within which the normal society activities are affected negatively.[1]

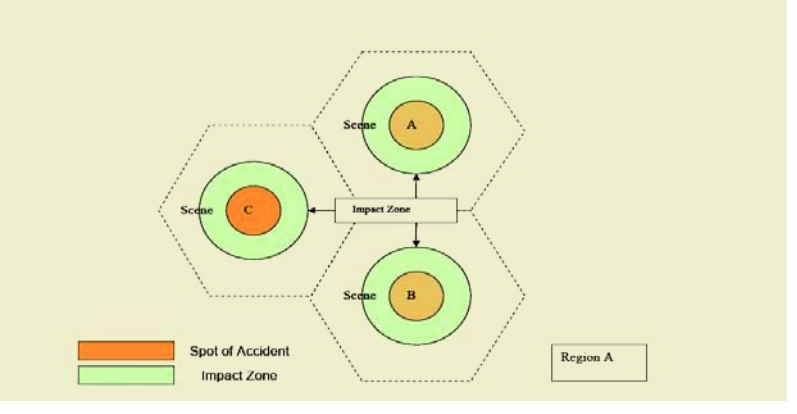


Fig. 1. Simultaneous Multiple Attacks

2.2 Response Services and Multiple Tiers of Emergency Responders

The nature of each emergency determines the emergency response services needed. Response services prescribe the mitigation actions to be taken, policies to be executed, rules to be followed, and personnel required. Example response services are firefighting, victim rescue, and security.

To provide such services, a number of emergency response personnel are called to the scene during a given incident of emergency [10]. Individual emergency responders may be assigned multiple services. For instance, firefighters perform firefighting and victim rescue at the same time in most cases of emergency response. Affected by the uncertainty of the nature of emergency response, the response services assigned to an individual emergency responder may be dynamically reconfigured, updated, and revoked.

[1] The impact zones are illustrated as circular for simplicity purpose only. The environmental effects such as wind or rain may change the shape and development pattern of an impact zone.

In an incident of chemical fire with radioactive contamination, emergency response services may include firefighting, victim rescue, first aid, decontamination, security, resident evacuation, traffic and etc. The following responders are usually involved[2]:

(1) FBI, Police, Firefighters, and Emergency Medical Service (EMS) workers
(2) Hazardous Material (HAZMAT) workers and medics and hospitals
(3) Security Inspector, Waste Disposal Technicians, and Government agencies

The reason we divide responders into three different groups is because they are activated in accordance to a time sequence, i.e., the first responders in the first group are called to react to the emergency before responders in the second and third groups are called. Further, from currently available data estimates we know it takes about an hour to activate all the responders in category 1. HAZMAT and decontamination units would be in place in about four hours. We refer to the responders in the first group as 1st tier responders, the responders in the second group as 2nd tier responders, the responders in the third group as 3rd tier responders.

Each response has to be assessed for its severity. The severity level allows the incident commander to invoke laws such as Article 2B (New York State) to involve local, state and federal assets. The severity assessment is usually based on incomplete, imprecise information. Such information is flowing presumably via a 911-call center call.

Most emergencies require multiple tiers of responders. Among them, the personnel in the 1st tier responders are easy to identify. They include the Police force and EMS workers in many cases. However, the need for other tiers of responders is unclear until a need for them is recognized. Such a need is recognized mainly by the existing responders on the scene, and they will request for follow-up tier responders subsequently.

For example, when firefighters are putting out a fire, they might notice the existence of toxic chemical fumes and radioactive contamination. Hence, they request for HAZMAT workers as the follow-up tier responders. We will describe in depth the role and involvement of different tiers of responders later in this paper. Meanwhile, the local Emergency Response Center must coordinate and mange the emergency situations for the duration of the emergency.

2.3 Sequence of Response and Role of First Responders

As soon as 9-1-1- receives a call, the emergency response process is activated in our architecture. The Emergency response Center coordinates the different groups of first responders while the latter carry out individual tasks [14], [15].

A typical response to an emergency is typically characterized by the following phases:

Phase I:

(1) Unified incident command center chooses the 1st tier responders and dispatches the emergency information to those responders

[2] The responders and activities under study are more general in nature. The response pattern may be different from natural incidences to terrorist attacks.

(2) 1st tier responders prepare themselves, drive to the scene, and carry out their tasks
(3) 1st tier responders request unified incident command center for follow-up responders

Phase II:

(1) Unified incident command center chooses the 2nd tier responders and dispatches the emergency information to those responders
(2) 2nd tier responders prepare themselves, drive to the scene, and carry out their tasks
(3) 2nd tier responders request unified incident command center for follow-up responders

More and more responders may be requested due to the need recognized by the existing responders. This chain of responders terminates when the emergency response finishes.

2.4 Time Delay in Emergency Response

We have assumed that there are time delays within the process of emergency response. This is exemplified in Figure 2. Within each phase, a number of time periods have been consumed before an effective response is performed. By "effective response", we mean the tasks performed to mitigate the incident.

Typical time delays in each phase are as follows:

I. Time spent by unified incident commander to choose the right response. Fast coordination and decision making in the face of fuzzy, incomplete and imprecise information should be made. A decision support system incorporating and adjusting the decision is vital. The system should be able to remove noisy information from the decision making information. However, the outlier information should not be discarded but analyzed for possible useful content.
II. Time spent by responders to get prepared before they carry out effective response such as suiting, etc. This preparation can be facilitated by information, right equipment, enough personnel, help from other groups, and task-related knowledge. Information should be receivable by first responders from the unified command and decision support system. First responders should also be able to publish information into the system providing valuable feedback though multiple channels. A secure "publish / subscribe" [28] system is an already existing methodology that can be leveraged in this regard. This issue is discussed in a later section as well as incorporated in the architecture. This usually leads to better situational awareness.
III. Time spent by responders on communication, information seeking, and decision making before an action is taken. Once again the "publish / subscribe" system can be leveraged to minimize the loss of time.

Time delays I, II and III reflect the efficiency of an Emergency Response Center. Several elements influence this time delay:

1. Availability (volume and completeness) of information on emergency incident
2. Correctness of information on emergency incident
3. Efficiency of decision making about possible incident type and required respond
4. Completeness of information on available first responders pool
5. Clear selection criteria of personnel appointment
6. Efficient communication tools for contacting first responders
7. Quality of dispatched emergency information to first responders

2.5 An Illustration of Emergency Response

One can understand the emergency sequence, roles of responders, time delays, and many other metrics of emergency response system design from the following diagram shown in Figure 2.

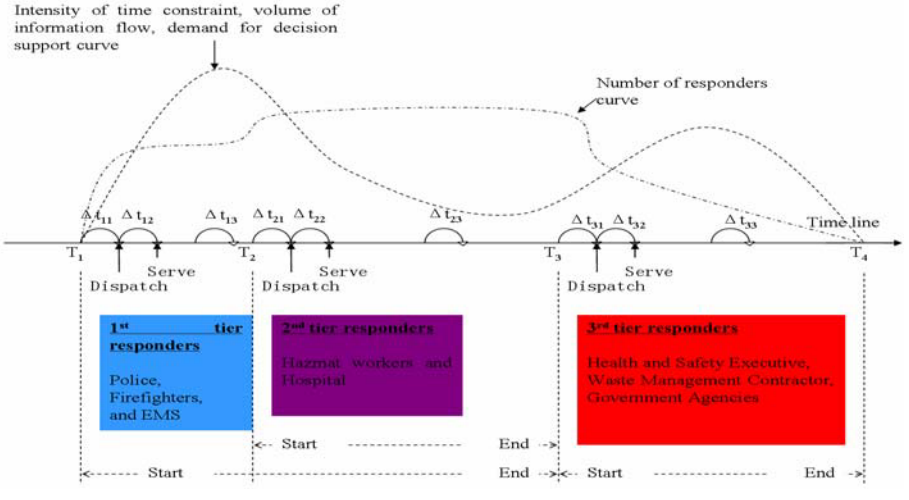


Fig. 2. Illustration of the Emergency Response

Horizontal coordinate (axis) represents the "Time Line". From left to right, the events occurred follow the time sequence.

We use "T_1", "T_2", "T_3", and "T_4" to represent the start of Phase I, Phase II, Phase III and the end of the emergency response respectively. We refer to the first kind of time delay in Phase I as Δt_{11}, the second kind time delay in Phase I as Δt_{12}, and the third kind of time delay in Phase I as Δt_{13}. A similar tagging rule is applied on the Phase II and Phase III.

In Figure 2, three kinds of time delays are shown in each individual phase of the emergency response. Unlike the first two kinds of time delays occurring at the beginning of each phase, the third kind of delay occurs some where during the process.

The intensity of "Time constraint, Volume of information flow, and Demand for decision support" curve represents the changes in those values at different time during the emergency response. Higher value implies that the response taken during that time period is relatively time critical, high volume of information flow, and high demand for decision support.

The "Number of Responders" curve represents the changes in the number of responders at different times. Higher value implies that the number of responders is relatively larger.

We observe that the peak of intensity of "Time constraint, Volume of information flow, and Demand for decision support curve" occurs during the T_1 and T_2 time periods. The curve declines as the emergency response moves from T_2 into T_3 time period. This is because more information is exchanged as to offset the disadvantage of incomplete information. Besides, unexpected events in the early phase of emergency response require more decisions rather than following routine procedures. Lastly, the level of emergency decreases as result of follow up reactions are taken, and thus, the events in the later phase are not as time critical as events in the earlier phase. However, this trend alters during T_3 and T_4 time periods. In addition to the tedious tasks of recovery, extensive activities on resource replenishment are carried out. Meanwhile, analysis on current attacks is done and Federal agencies design prevention strategies to avoid future attacks on potential targets.[3]

Emergency responses taken in the first two time periods are more complex than that in the third time period. Thus, the overall performance in the first two time periods greatly affects the overall emergency response performance. However, the performance in the last time period affects the speed of disaster recovery and the prevention for future attacks.

The parameters shown in Table 1 characterize the changes during the shifts among different phases of emergency response. These parameters suggest that an effective emergency response system must be able to work at the maximum levels of workload among all possible phases.

Table 1. Parameters of Response in Different Phases

	Completeness of Info	**Volume of Info Flow**	**Demand for Decision Making**	**Quality of Decision Making**	**Complexity**	**System Load**	**Time Constraint**
T1-T2	Low	High	High	Low	High	High	High
T2-T3	Medium	Medium	Medium	Medium	High	High	High
T3-T4	High	High	High	High	Medium	High	High

This essentially includes such things as high bandwidth for data transmission, good load balancing performance, quick integration of multimedia data from multiple sources, and fast data processing features [5], [16].

[3] Thanks to the suggestions from anonymous viewers.

2.6 Decision Structure and Information Flow

Emergency Response is a tem effort and team performance is critical. In team performance, the relationship between information structure and performance is mediated by variables related to the operating structure in which the information structure is embedded.

Team Performance = function (Information Structure, Decision Structure, Environment) (17)

We paraphrase Baligh and Burton's [17] definition of an organization structure in terms of the three elements (See Rao, Chaudhury and Chakka [18] for details.)

A. The Decision-Structure
   1. The decision-makers in the team
   2. The decision problems that the team is solving; the decisions to be made in each period by each decision-maker
   3. The decision rules and mechanism used by each decision-maker; given the information each members can produce his/her decision

B. Information Structure
   1. Identify the information supplied by decision maker and identify who supplies the information

C. The environment
   1. The reward-system under which the decision makers are operating
   2. The degree of turbulence in the environment where the information originates

Discussion of the environment in Emergency Response is beyond the scope of this research. We represent the decision structure and information flow that occurs in Phase I as shown in Figures 3 and 4. A similar decision structure and information flow occurs in Phase II and Phase III.

2.6.1 Decision Structure

The emergency response system is organized as a two-level hierarchy, with coordinators from the Emergency Response Center at one level and police, firefighters, and EMS squad on the other level. They work together as a team that constitutes the set of decision makers. The decision structure is illustrated as below:

Four types of decision making could occur as illustrated in Figure 3: Emergency Response Center makes decisions alone, First responders make decisions alone, Emergency Response Center delegates decision making privilege to first responders and First responders request decision making help from Emergency Response Center.

The coordinators' decision-problem is to dispatch the right responders to react to the emergency and distribute information to first responders. Their purpose is to increase the effectiveness of response, enhance efficiency of response, decrease response time, and improve coordination between groups. The individual first responder teams' decision-problem deals with the correct actions taken to finish their job and request for help when necessary. The decision making model is validated through the observation of "Table Top Simulation Exercise" of emergency response.

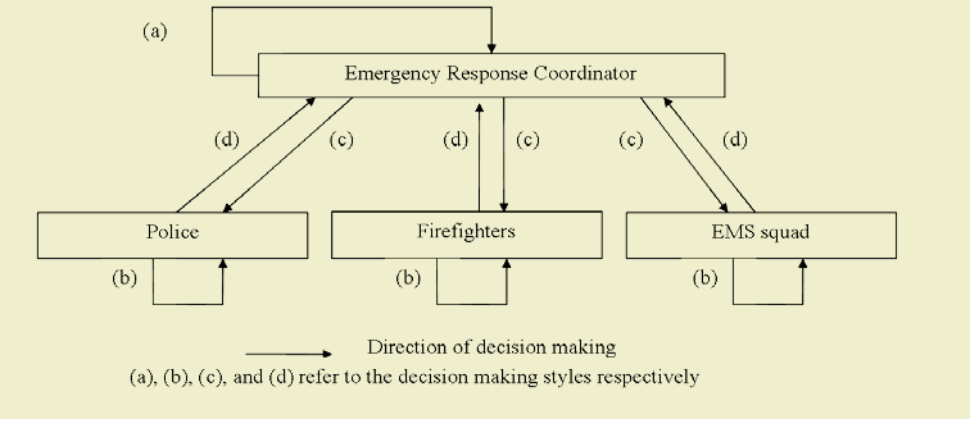


Fig. 3. Decision Making Model

2.6.2 Information Structure

The information obtained by Emergency Response Center includes real time information from first responders, real time information from other divisions, and reference data from collective knowledge base in their existing systems.

The information available to fist responders includes real time information obtained at the scene, real time information shared from other groups at the scene and information provided by Emergency Response Center as depicted in Figure 4.

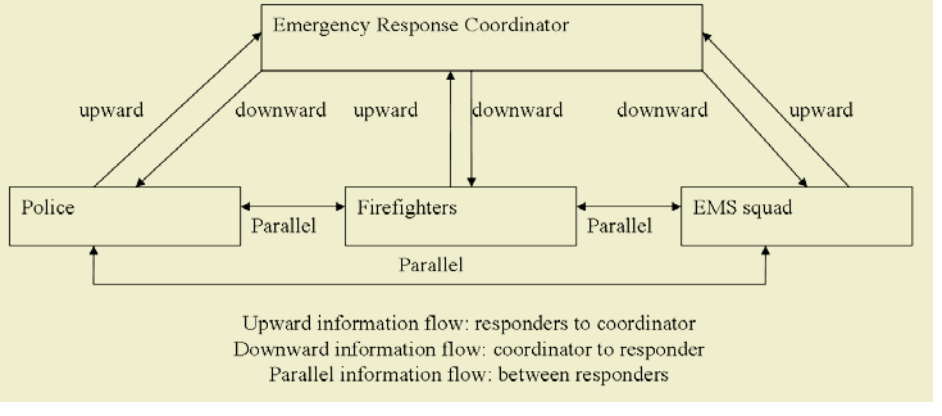


Fig. 4. Information Structure

Two-level hierarchy allows us to set up three separate information structures: upward information flow, downward information flow, and parallel information flow. The information structure is validated through the observation of "Table Top Exercise" of emergency response.

2.7 Fundamental Requirements of Emergency Response System Design

From the above analysis, we conclude the following requirements of an effective emergency response architecture [23], [24], [25]. Notice that this is not a conclusive list of requirements. We would introduce several more in the later sections.

Table 2. Fundamental Requirements of System Design

Requirement	Description
Effective Information Collect and Analysis	Emergency response data collection, compilation, analysis, and storage.
Directory of First Response Resource	Database of personnel, equipment, and tools with their availability, amount, and properties
Knowledgebase of task related information	Police, legal regulation, code, reference, and maps
Communication Support	Multiple communication channel, mobile, robust, and secure communication
Open Information Share	Information share, exchange, access
Decision Making Support	DSS, autonomic decision making, decision role delegation, and expert system
Response Tracking Support	Updating and tracking of personnel location, resource assumption, and task progress
Multimedia Support (including GIS)	Visualization tools for representing, decision making, and communication
Security Support	Secured information flow and access control
Fault Tolerance and Redundancy Support	Data backup, distributed data storage, load balance, and mirrored hot servers
Modularity and Scalability Support	Scalable architecture facilitating local, state, and federal agencies, easy integration and upgrade of modules

The performance metrics of the above requirements have not been developed in this paper and is a part future research.

3 Coordination Theory in Emergency Response System

Coordination has been a long-standing interest of researchers. Coordination Theory [18] has proved to be a very successful theory in resource and tasks coordination by research in multiple disciplines [5], [19], and [20]. The coordination theory defined coordination as “managing dependencies between activities”. Dependency refers to the dependent relationships between tasks and resources. Theoretical framework for analyzing coordination in complex processes was introduced and coordination mechanisms for common dependencies coordination were presented.

The main claim of coordination theory is that dependencies and the mechanisms for managing them are general. Commonly observed dependencies and related mechanisms are illustrated as follows.

Table 3. Sample Dependencies and Mechanisms

Dependency	Examples of coordination processes for managing dependency
Shared resources	"First come, first serve", priority order, budgets, managerial decision, market-like bidding
Task/Resource assignments	Same as for "Shared resources"
Producer / Consumer relationships	
→Prerequisite constraints	Notification, sequencing, tracking
→Transfer	Inventory management (e.g., "just in time", "economic order quality")
→Usability	Standardization, ask users, participatory design
→Design for manufacturability	Concurrent engineering
Simultaneity constraints	Scheduling, synchronization
Task / Subtask	Goal selection, task decomposition

During a response to emergencies, coordination plays an important role for managing response resource and tasks. Consequently, we integrate coordination functionality into the proposed architecture design.

Within the context of emergency response, respond resource may include personnel, equipment, and tools. Respond resource coordination plays an important role as a solution to existing resource allocation conflict, particularly in the event of multi-incident. Resource coordination offers optimal resource allocation schemes, coordinates resource supplies and redeployment, and prevents or reduces the potential of resource exhaust.

3.1 Coordination of Resource Allocation

In case of resource allocation conflict among multiple incidents, two allocation schemes may be employed. "Winner-take-all" scheme appoints resource to meet the need of one emergency response first. "Collaboration" scheme appoints resource among all requesting emergency response. Example collaboration scheme is such as "priority order".

Consider the following example. Emergencies A and B request non-sharable resource R_A. Assume that 5 units is of resource R_A is available, and the demand from the two emergencies A and B for the resource is for 4 and 6 units respectively. Using winner-take-all scheme, emergency A may be allocated 4 units of resource R_A when emergency A beats emergency B, leaving 1 unit for B. Using collaborative scheme, emergency A may be allocated 2 units while emergency B may be allocated 3 units if they are equal of equal priority. There are several variations to resource sharing that are possible which results in a more prudent and optimal allocation.

3.2 Coordination of Tasks

Coordination of emergency tasks involves the management of the sequence of emergency response. During emergency response, three major tasks control patterns exist. According to the Theory of Workflow Patten by Van der Alast [21], the three

patters are termed as “Sequence”, “Parallel Split”, and “Synchronization”. Sequence pattern states that tasks execute in sequence. Parallel split pattern states that tasks executed in parallel. Synchronization pattern states that tasks are synchronized in execution.

For tasks of sequence pattern, mechanisms such as notification, sequencing, tracking, and standardization are employed. For tasks of parallel split and synchronization, mechanisms such as scheduling, notification, and synchronization are employed.

Coordination effort itself is among the many resources require coordination support. In addition, tasks coordination incurs new resource coordination need. E.g., after a task is performed in respond to an incident of emergency/attack, not only the follow up tasks will be notified to start, but also the resource involved in the finished task might be dynamically allocated to help other incidents of emergency response, lessening the potential of resource exhaustion.

3.3 Supplementary Requirements of Emergency Response System Design

Based on the discussion about coordination of resource and tasks, the following support models provide supplement to the above fundamental requirements of system design.

Table 4. Fundamental Requirements of System Design

Requirement	Description
Resource Coordination Support	Respond resource allocation and logistics management
Task Coordination Support	Task assignment and task flow control

4 Emergency Response to Multi-incident

Emergency response becomes complicated when the number of emergency incidents increases. Most critical issues arise in such multi-incident event is closely related to emergency response resource.

4.1 Resource Demand, Allocation and Exhaustion

The resource available to first responders critically influences the outcome. Such resource includes personnel (professional responders and volunteers), equipment, transport tools, coordination effort and etc. Due to the high cost of personnel training and equipment, availability of response resource within a region may be limited [11]. Acquisition from neighboring regions comes at the expense of a precious resource called time.

Each emergency incident demands specific types and amount of resources [12], [22]. In the case of multiple incidents of emergency, specific types of resources may be requested by multiple responders. When resources are allocated among multiple incidents, two schemes are usually employed: “Shareable” and “Exclusive Allocation”. When sharing the same resource without conflict for multiple incidents,

the optimization algorithm has to arrive at a recommendation based on constraints such as accessed need, relocation time, etc. From a slightly different perspective another example of a sharable resource is the information database. Correlating information generated by one incident can be included in arriving at decisions for another incident. There are several resources that have to be allocated exclusively to one incident and therefore become inaccessible to others. An example of exclusively possessed resource is the personnel. Policy driven decision support systems are needed to assist the unified incident commander to help him arrive at a decision.

In the case of multiple incidents of emergency/attacks, exclusive allocation of resource may incur the exhaustion of resources. Two reasons exist for this potential problem. The first reason is insufficient inventory of resource. The second reason is sufficient inventory of resource but poor allocation schemes.

The illustration of resource exhaustion due to poor allocation of exclusively possessed resource is given in Table 2. In this example, emergency A, B, and C occur within a short time period. We assume that the overall resource inventory for six possible resources is ten units of R_A, eight units of R_B, ten units of R_C, eight units of R_D, six units of R_E, and eight units of R_F. Assume that they are all exclusively possessed meaning to say they cannot be shared.

At time T1, emergency A occurs. We assume that the on-scene authority for emergency A requests for resource A, B, C, and D. We further, assume that the optimal amount of resource necessary to mitigate emergency A is as follows: three units of resource R_A, four units of resource R_B, two units of resource R_C, and five units of resource R_D. With a poor allocation scheme, we assume that the coordinator allocates five units of resource R_A, four units of resource R_B, four units of resource R_C, and five units of resource R_D to the response of emergency A. Notice that more than enough resource of R_A and R_C are allocated.

At time T2, emergency B occurs right after emergency A. We further assume that optimal allocation of resources for emergency B is three units of resource R_A and R_C, four units of resource R_E and six units of R_F. With a poor allocation scheme, we assume that the coordinator allocates three units of resource R_A, four units of resource R_C, four units of resource R_E, and six units of resource R_F to the response of emergency B. Notice that more than enough resource of R_C is allocated.

Table 5. Illustration of Resource Exhaustion in Exclusive Allocation Scheme

Resource Type	**Inventory**	**Emergency A**		**Emergency B**		**Emergency C**	
		Optimal	Allocation	Optimal	Allocation	Optimal	Allocation
R_A	10	3	5 (surplus)	3	3	4	2 (inadequate)
R_B	8	4	4	0	0	4	4
R_C	10	2	4 (surplus)	3	4 (surplus)	5	2 (inadequate)
R_D	8	5	5	0	0	2	2
R_E	6	0	0	4	4	0	0
R_F	8	0	0	6	6	0	0

At time T3, emergency C occurs right after emergency B. We further assume that the optimal amount of resources needed to deal with emergency C is as follows: four

units of resource R_A four units of resource R_B, five units of resource R_C, and two units of resource R_D. Because of sub-optimal allocation of resources to emergency A and B, response to emergency C is starved and resource R_A and R_C are inadequate. Hence, the quality of emergency response to emergency C is greatly affected. Notice that the sum of the optimal amount of R_A and R_C in emergency A, B, and C will not exceed the overall inventory, given a proper allocation scheme.

4.2 Resource Redeployment

Due to inventory constraints of first response resources, resource redeployment is an appealing solution. We refer to "resource redeployment" as the reallocation of resources after it has been used in some other emergency incidents. An example to the point is the redeployment of a group of firefighters to emergency B after they have finished their duties in response of emergency A.

Resource redeployment helps the other ongoing emergency response activities. By redeploying first response resources, the aggregation of resource at a currently active emergency enhances efficiency. Notice that resource redeployment also helps with the resource exhaustion problem.

A better understanding in resource redeployment requires knowledge of "Resource Reusability". In the response to emergency/attack, we refer to resource reusability as the ability of a given resource to be reused without interruption in time frame. We have assumed this for simplicity here. However real life applications will require relaxation of this constraint and needs to be built into the model. Resources such as personnel, transport tools, communication tools are generally regarded as reusable. Resources such as disposable facilities are generally regarded as not reusable. Equipment such as decontamination equipment is also considered as not reusable.

Notice that resource redeployment may be invoked at any point of emergency response. It is not necessary that resource be redeployed only after their duties are done.

In the event of multi-incident, concurrent incidents might have unequal levels of priority. For example, a fire to an abandoned house is considered to be of less priority than an explosion in a subway station. In case that there is a lack of firefighters to respond to the explosion in subway station, the firefighters working on the abandoned house fire might be redeployed to the subway station. In this way, limited number of resource can provide more mitigation and better overall performance. As a matter of fact, terrorists frequently launch trivial attacks prior to a serious attack. The trivial attacks are employed as to attract response effort and resource. In this way, the response and resource to the follow up serious attack are likely to be inadequate and performed at low performance. Therefore, decision support model of redeployment is necessary to assist responders to such series attacks.

4.3 Supplementary Requirements of Emergency Response System Design

Based on the discussion about resource exhaust and resource redeployment, the following support models provide supplement to the above fundamental requirements of system design.

Table 6. Fundamental Requirements of System Design

Requirement	Description
Inventory Monitoring Support	Response resource management, alert auto generation, and logistics management
Optimal Resource Allocation Support	Optimal amount of resource allocation suggestion
Resource Redeployment Support	Situation-aware resource redeployment analysis, and logistics management

5 Research Propositions and Methodology

Based on the above discussion and analysis of emergency response process and its characteristics, we present several research propositions.

5.1 Research Issues in Resource Allocation Schemes

Frequently two major types of resource allocation schemes are utilized. The first type is referred to as "Full Resource Allocation" mechanism. In this scheme more than enough resource is allocated at a time to contain the incident and avoid possible lack of resource in the future. This scheme is premised on the fact that more resources allow for a quick mitigation of the emergency and thus reduce the damage.

The second style is referred to as "Conservative Resource Allocation" mechanism, in which less resource is allocated at first and progressively more resource is added as the need arises and as it becomes evident that there is no other competing needs. The initial amount of first responders assigned to a given incident may set to be at a fixed portion less than 100%. The speed and the amount of supplemental resources differ from situation to situation.

In dealing with terrorist attacks such as 9/11, conservative resource allocation mechanism may be preferred. In such complex cases, conservative resource allocation mechanisms serve as a means to prevent resource exhaustion. In dealing with routine emergencies such as a house fire, full resource allocation mechanisms may be preferred. In such cases, full resource allocation mechanism prevents possible lack of resources and allow us to bring the fire under control more quickly.

A possible and useful study in this area would include propositions such as:

P1: *Conservative resource allocation mechanism in the initial stage of an unknown emergency reduces potential for emergency resource exhaustion*

P2: *Full resource allocation mechanism outscores conservative allocation in routine emergency response*

The above propositions are to be validated through simulation approaches in the future research. Refer to section 5.3 for more information.

5.2 Research Issues in Time Delay

As discussed above, response time is a critical parameter of evaluating emergency response [14]. Based on the outcome, the response time can be divided into two

categories. The first type of response time is called as "Supportive Response Time" and the second type of response time is called "Effective Response Time".

By "Supportive Response Time", we refer to the time spent on seeking information, decision making, communicating and coordinating tasks. The amount of time spent helps first responders to make quality response. The length of supportive response time can be reduced on the tradeoff of less amount of information, lower decision quality, less communication, and poor coordination.

By "Effective Response Time", we refer to the time spent on performing response tasks as an immediate remedy to an emergency. Effective response includes "putting out the fire" and "giving medical treatment". The length of effective response time is determined by the size and the particular nature of the individual emergency. For example, the length of time consumed to put out fire depends on the size of the fire and the number of fire engines.

For performance of response, we rely heavily on the availability of information collected, the quality of decision making, the effectiveness of communication, and the level of coordination [16]. However, first responders can not obtain the above information without consuming a significant amount of time. For example, in case of emergencies, information is typically limited, incomplete, and inaccurate. As a result, first responders will have to sacrifice a portion of time to gather enough information for work. Hence, responders and coordinators face the problem of how to allocate the amount of supportive response time.

To deal with routine emergencies, existing procedures are followed and previous experience guides the responders through the emergency response process. In such a routine, the amount of supportive time spent may negatively correlate with performance.

In contrast, an unexpected complicated emergency/attack may require a more deliberated response scheme and improvising. In such a case, the amount of supportive time spent may positively correlate with the performance. Rather than passively responding to an attack, a well prepared and coordinated response scheme leads to a response that decreases the overall detrimental impact.

Hence, we propose the following hypothesis regarding the relationship between response time and response performance.

P3: In a routine emergency, supportive response time is negatively correlated with the emergency response performance.

P4: In a non-routine emergency/attack, the supportive response time is positively correlated with the emergency response performance.

The above propositions are to be validated through simulation approaches in the future research. Refer to section 5.3 for more information.

5.3 Methodology

Simulation approaches have been proposed and studied by researchers for emergency response systems since 1960th [34],[35],[36],[37],[38]. Simulation enables the study of a myriad of possible scenarios and strategies efficiently [18]. To further this study in the future, integrated agent- and activity-centric simulation [33] is to be employed for system performance evaluation and research proposition validation. Activity-

centric approach provides support for process modeling by capturing the mechanic components of the emergency response activities while agent approach provides specific aspects of the human components for decision making. Multi-incident scenarios are to be exercised, performance metrics designed, and performance on different decision strategies studied with different response strategies.

6 Conclusion

In this research, we have presented a set of system design requirements for an architecture of coordinated multi-incident emergency response system. The design requirements provide the fundamental functionalities for emergency response to multiple emergency incidents/attacks. It supports well coordinated emergency response and manages resource optimal allocation and redeployment. The implementation of such a system would provide strong support to the emergency management and benefit the community of emergency responders. The unified incident command system to deal with multiple incidents is very complex and daunting task. The discussion presented in this paper provides an analysis of several major issues. Uncovered issues due length considerations integrating decision making modules with existing systems, public relations, public information feed system, etc. Some of the areas for future research include: Simulation design on the basis of the proposed design requirements for multi-incident emergency response system, System performance metrics design and performance test, Propositions test with different strategies, etc.

References

1. Probasco, K., Mogle, R.: The 21st Century First Responder: The Vision. prepared for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830 (1998)
2. Dykstra, E.: Toward an International System Model in Emergency Management. Public Entity Risk Institute Symposium (2003)
3. Neal, D.: Transition from Response to Recovery: A Look at the Lancaster, Texas Tornado. Journal of Emergency Management, Vol. 2, No. 1 (2004)
4. Arens, Y., Rosenbloom, P.:Responding to the Unexpected. Communication of ACM, September 2003, vol, 46, no. 9 (2003)
5. Shen, S., Shaw, M.: Managing Coordination in Emergency Response Systems with Information Technologies. IT for Emergency Response System Coordination. Proceedings of the Tenth American Conference on Information Systems. New York, New York (2004)
6. Mehrotra, S.: Project Rescue: Challenges in Responding to the Unexpected. Proceedings of 16th Annual Symposium on Electronic Imaging Science and Technology, San Jose, CA (2004)
7. Gillis, B.: E-Safety: Saving Lives and Enhancing Security for Washington Citizens. WSU Center to Bridge the Digital Divide. White Paper 2003.06.01 (2003)
8. Macko, S.: The Threat of Chemical and Biological Attack. Emergency Net News (ENN) Daily Report, 08/27/96 (1996)

9. Shea, D.: Small-scale Terrorist Attacks Using Chemical and Biological Agents: An Assessment Framework and Preliminary Comparisons. Congressional Research service, May 20 (2004)
10. Jump, P., Bruce, J.: The Response Factor. Electric Perspectives, May/Jun (2003), 28, 3, ABI/INFORM Global, pg 22
11. Department of Homeland Security, http://www.dhs.org/
12. Green, L.: Improving Emergency Responsiveness with Management Science. Emergency Service Model, pg. 1 (2000)
13. Carafano, J.: Preparing Responders to Respond: The Challenges to Emergency Preparedness in the 21st Centry. Heritage Lectures, No. 812, November 20 (2003)
14. Wimberly, R.: How Quickly Can We Respond. Occupational Health & Safety, Apr (2004), 73, 4, ABI/INFORM Global pg. 46
15. Walks, I.: Emergency Response outside the Envelope. Security Management, 53 (2003)
16. Turoff, M.: The Design of A Dynamic Emergency Response Management Information System (DERMIS). Journal of Information Technology Theory and Application, (2004) 5, 4, ABI/INFORM Global pg. 1
17. Baligh, H., Burton, R.M., Obel, B.: Designing organization structures: An expert system method. in: J.L. Roos, ed.. Economics and Artificial Intelligence (Pergamon, Oxford, 1985) 177—181 (1995)
18. Rao, R., Chaudhury, A., Chakka, M.: Modeling Team Processes: Issues and a Specific Example. *Information Systems Research*, Vol 6, Number 3, pg 255-285 (1995)
19. Malone, T.: The Interdisciplinary Study of Coordination. ACM Computing Surveys, Vol. 26, No. 1 (1994)
20. Crowston, K.: Coordination Theory. Human-Computer Interaction in Management Information Systems, Vol. I.M.E. Sharpe (2001)
21. Crowston, K.: A Coordiantion Theory Approach to Process Description and Redesign. In T. W. Malone, K. Crowston & G. Herman (Eds.). Organizing Business Knowledge: The MIT Process Handbook. Cambridge, MA:MIT Press (2002)
22. Van Der Aalst, W.: Workflow Patterns. Distributed and Parallel Databases, Vol. 14, Issue 1, July (2003), Pg. 5-51, ISSN:0926-8782
23. Chaffee, M.: DVATEX: Navy Medicine's Pioneering Approach to Improving Hospital Emergency Preparedness. Journal of Emergency Management, Vol. 2, No. 1 (2004)
24. Business & Finance Bulletin IS-3 Electronic Information Security, November 12, (1998), University of California
25. Sawyer, S.: Mobility and the First Responder, Communications of the ACM, March (2004), vol. 47, no. 3
26. Anderson, P.: Information Technology: Where is It in the Coordination of Emergency Services. Asia Pacific Police Technology Conference (1991)
27. Jenvald, J.: Simulation-supported Live Training for Emergency Response in Hazardous Environments. Simulation & Gaming (2004)
28. Huang, Y. and Garcia-Molina, H.: Publish/Subscribe in a Mobile Environment. 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE) (2001)
29. NIMS Document Directory March 1, 2004 version: http://www.nimsonline.com/nims_3_04/index.htm
30. State and Local Guide (SLG) 101: Guide for All-Hazard Emergency Operations Planning: First eponders – FEMA: http://www.fema.gov/fema/first_res.shtm
31. Incident Command System, US Coast Guard Site: http://www.uscg.mil/hq/gm/mor/Articles/ICS.htm

32. Consequences Assessment Tool Set (CATS): http://www.saic.com/products/simulation/cats/cats.html
33. Raghu, T.S., Jayaraman, B., Rao, H.R.: Toward an Integration of Agent- and Activity-Centric Approaches in Organizational Process Modeling: Incorporating Incentive Mechanisms. Information Systems Research, Vol. 15 Issue 4 (2004)
34. Savas, E. S.: Simulation and Cost-effectiveness Analysis of New York's Emergency Ambulance Service. Management Science, Vol. 23, Issue 2, 146-158 (1976)
35. Fitzsimmons, J.A.: A Methodology for Emergency Ambulance Deployment. Management Science, Vol 19, Issue 6, 627-636 (1973)
36. Rider, K.L.: A Parametric Model for the Allocation of Fire Companies in New York City. Management Science, Vol. 23 Issue 2, 146-158 (1976)
37. Green, L.: A Multiple Dispatch Queuing Model of Police Patrol Operations. Management Science, Vol. 30, Issue 6, 653-664 (1984)
38. Jenvald, J. and Morin, M.: Simulation-supported Live Training for Emergency Response in Hazardous Environments. Simulation and Gaming, Vol. 35, No. 3, 363-377 (2004)

Multi-modal Biometrics with PKI Technologies for Border Control Applications

Taekyoung Kwon and Hyeonjoon Moon

Sejong University, Seoul 143-747, Korea
{tkwon, hmoon}@sejong.ac.kr

Abstract. It is widely recognized that multi-modal biometrics has the potential to strengthen border protection by reducing the risk of passport fraud. However, it may take high costs to issue smart-card enabled passports over the world and to process a huge amount of biometric information (on-line). A public key cryptography is another useful tool for verifying a person's identity in a stringent way, but a key management is one of critical problems arising from the use of cryptographic schemes. For example, a passport-holder should keep a private key in a smart-card-level device while an inspecting officer accesses a corresponding public key in an authentic manner. In this paper, we present a low-cost but highly-scalable method that uses multi-modal biometrics based on face and fingerprints, and public key infrastructures (PKIs) for border control applications. A digital signature in PKIs and multi-modal biometrics are carefully applied in our scheme, in order to reduce the possibility of undesirable factors significantly at nation's borders without requiring any hardware device in passports. We could print a (publicly readable) barcodes on the passport instead of requiring the smart-card-level devices.

1 Introduction

Due to the rapidly growing interconnectivity over the world, the significance of border security and safety has been discussed from the perspective of national defense. The border control applications usually require proper means of identifying travelers, for example, by what they have (passports, visas, and travel documents), and what they know (answers to several questions). However, the importance of adding new elements of who they are (biometrics) is now being observed and recognized by many countries, in the post September 11 era. For instance, the U.S. government initiates the USVISIT (United States Visitor and Immigrant Status Indicator Technology) program and has a concrete plan to install biometric equipment and software at all national ports of entry [28].

Challenges. Biometrics is actually the science of using digital technologies to identify or verify a human being based on the individual's unique biological (say physiological or behavioral) characteristic such as fingerprint, voice, iris, face, retina, handwriting, thermal image, or hand geometry [16, 19]. Among those various biometric features, fingerprint, iris pattern, facial image, and hand print

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 99–114, 2005.

are regarded as most suitable for border control applications for their relatively-accurate measuring. It is widely recognized that multi-modal biometrics (multiple biometrics) has the potential to strengthen border protection by reducing the risk of passport fraud and improving the accuracy and performance of measurement. However, depending on biometrics one could face different challenges in border control applications: 1) It may take high costs to process a huge amount of biometric information (on-line) for $1 : n$ identification and to issue smart-card-enabled passports over the world for $1 : 1$ verification; 2) biometrics is still remaining as a useful technology in small scale applications, excluding world-wide border control applications; and 3) a passport holder may feel reluctant to provide his or her biometrics to an inspecting officer because of inconvenience and privacy infringement.

This Paper. The main concern of this paper includes the first and second challenges above. We aim to present a low-cost but highly-scalable method that uses multi-modal biometrics based on face and fingerprints (or iris patterns in the future study) without requiring any hardware devices in passports. We exploit the existing tools in software and consider a public key cryptography for verifying a person's identity in a stringent way, in public key infrastructures (PKIs) for border control applications. We carefully define a formal model, explore such tools satisfying our model, and then present our scheme by exploiting them as presented in the following sections. A digital signature in PKIs and multi-modal biometrics are carefully applied in our scheme, in order to reduce the possibility of undesirable factors significantly at nation's borders without requiring any hardware device in passports. We could print a (publicly readable) barcode on the passport instead of requiring the smart-card-level devices.

The rest of this paper is organized as follows. Section 2 describes the overview of our scheme while Section 3 and 4 will describe more details of the proposed scheme. Section 5 will conclude this paper.

2 Overview of Our Scheme

2.1 Definitions

Security Parameters. Let κ and ℓ denote security parameters where κ is a general one (say 160 bits) and ℓ is a special one for public keys (say 1024 bits).

Face Space. Face space is a representation of face as a point (Figure 1). A face is represented by its projection onto a subset of eigenvectors in face space. In the face recognition literature, the eigenvectors can be referred to as *eigenfaces* [27].

Digital Signature Scheme. A digital signature is a term used to describe a data string which associates a digital message with an assigned person only. It has various applications in information security such as authentication, data integrity, and non-repudiation. Formally a digital signature scheme is denoted by $\Sigma = (\mathcal{G}_\Sigma(1^\ell), \mathcal{S}, \mathcal{V})$ where $\mathcal{G}_\Sigma$ is a probabilistic algorithm returning a public-private key pair from input 1^ℓ, and $\mathcal{S}$ and $\mathcal{V}$ are respectively signing and verifying algorithms, which run in polynomial time [15].

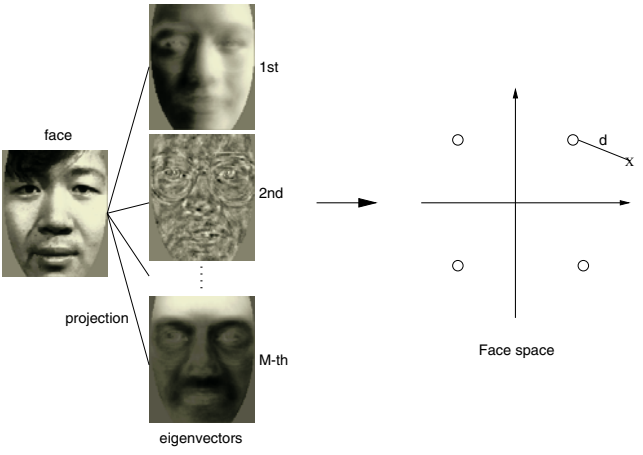


Fig. 1. Representation of face (face space)

Public Key Infrastructure. For an authorized assertion about a public key, we use digital certificates issued by a trusted entity called the certificate authority (CA) in the existing public key infrastructure (PKI) [6].

2.2 Basic Concept

Requirements. A passport holder should present to an inspecting officer his or her biometrics along with the passport for 1 : 1 verification. The first requirement is that the existing passport (issued by each national body over the world) must be refined in low-cost (say without mandatory requirement of embedding any hardware device on it) for accommodating biometrics. In other words, we should be able to print a barcode on the passport along with human readable text, instead of requiring a smart-card-level device for biometric information. The passport should still remain passive in that sense. The barcode is the dominant automatic identification technology that fits our purpose [22]. Especially 2D codes provide higher information capacity than conventional ones. For example, a 2D bar code symbol (such as PDF 417 and QR code) can hold up to about 4,300 alphanumeric characters or 3,000 bytes of binary data in a small area [13].

One drawback of storing biometric information in publicly readable form is its vulnerability to a potential biometric attack known as a *hill-climbing attack*. This attack could occur when an attacker has access to the biometric system and the user's template upon which (s)he wishes to mount a masquerade attack [25]. The attacker could exploit the compromised biometric template to produce a new image that exceeds the threshold of the biometric system and use that image again as input to the system to which the original template belongs. The second requirement is that a cryptographic scheme should be applied to the biometric information printed on the passport. A digital signature scheme is appropriate for providing integrity in a stringent way. We explore the most suitable schemes by presenting a formal model.

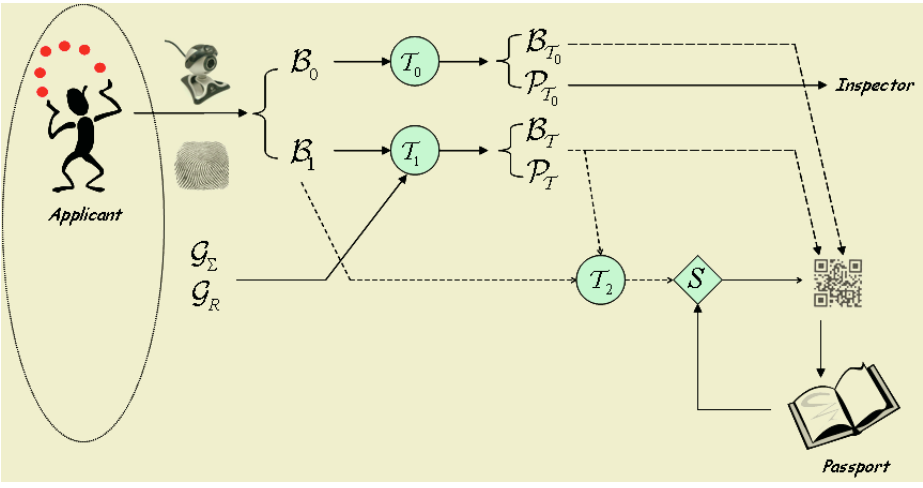


Fig. 2. Passport Issue

Formal Model. In order to authenticate a passport holder using biometrics without smart-card devices, we postulate that the passport holders can be screened with regard to their biometrics and demographic information in the passport that is not protected directly by hardware. So the passport holder is defined formally as $\mathcal{U} = \{\mathcal{B}, \mathcal{P}\}$ where $\mathcal{B}$ and $\mathcal{P}$ are defined as user's biometrics and possession (passport) respectively. $\mathcal{B}$ is regarded as a probabilistic algorithm returning user's biometrics while $\mathcal{P}$ is deterministic. As for multi-modal biometrics, $\mathcal{B}$ can be regarded as a set of biometrics, for example, $\mathcal{B} = \{\mathcal{B}_0, \mathcal{B}_1\}$.

Based on our biometrics scheme, we manipulate the user's biometrics with regard to feature representation. We define the following transformation:

- $\mathcal{T}_0 : \mathcal{B}_0 \rightarrow \langle \mathcal{B}_{\mathcal{T}_0}, \mathcal{P}_{\mathcal{T}_0} \rangle$

where $\mathcal{B}_{\mathcal{T}_0}$ and $\mathcal{P}_{\mathcal{T}_0}$ are feature representation and eigenvector respectively.

Given a digital signature scheme Σ, we have to manipulate the key returned by $\mathcal{G}_\Sigma$ to be linked with both the user's biometrics and possession. Therefore, we define the following transformation:

- $\mathcal{T}_1 : \langle \mathcal{G}_\Sigma(1^\ell), \mathcal{G}_R(1^\kappa), \mathcal{B}_1 \rangle \rightarrow \langle \mathcal{B}_\mathcal{T}, \mathcal{P}_\mathcal{T} \rangle$ and
- $\mathcal{T}_2 : \langle \mathcal{B}_1, \mathcal{B}_\mathcal{T}, \mathcal{P}_\mathcal{T} \rangle \rightarrow \mathcal{G}_\Sigma$,

where $\mathcal{G}_R$ is a probabilistic algorithm returning a random integer from input 1^κ, and $\mathcal{B}_\mathcal{T}$ and $\mathcal{P}_\mathcal{T}$ are respective transformed values. We define $\mathcal{P} = \{\mathcal{B}_{\mathcal{T}_0}, \mathcal{B}_\mathcal{T}, \mathcal{P}_\mathcal{T}\}$ while $\mathcal{P}_{\mathcal{T}_0}$ is manipulated as an eigenvector that might be known to an inspector. Note that the inverse transformation is possible for $\mathcal{T}_0$, while it is computationally infeasible for $\mathcal{T}_1$ and $\mathcal{T}_2$. It is impractical for the latter transformations to measure $\mathcal{B}$ by feature extraction which cannot guarantee enough entropy.

Passport Issue. For the issue of passport, an applicant must provide his or her biometrics (facial image and fingerprint) to the passport issuing department

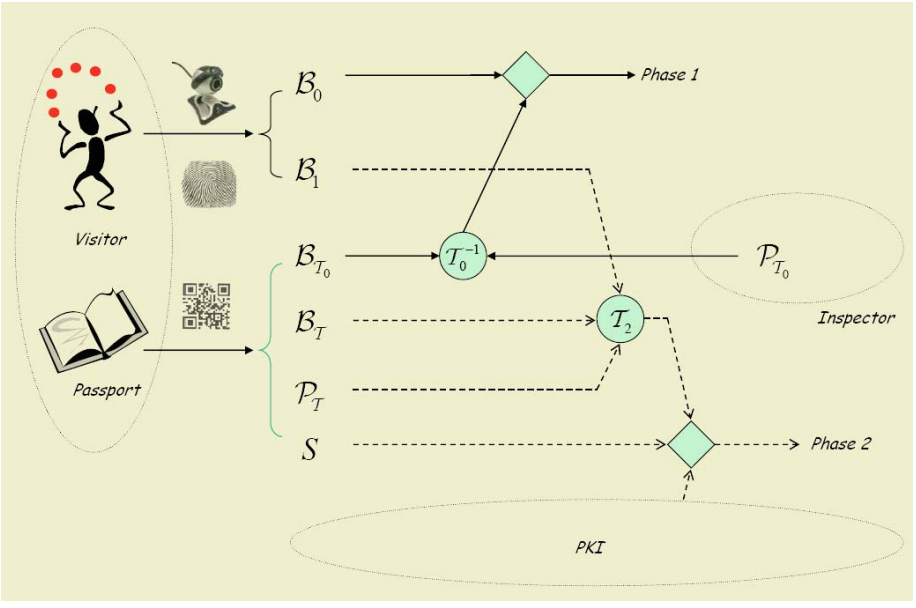


Fig. 3. Authentication Procedure

along with the legacy authorization materials such as proof of citizenship, photographs, and passport application form (e.g., DSP-11 in the States). Then, Figure 2 shows how the passport is composed of in our formal model.

Authentication Procedure. Figure 3 describes the proposed authentication procedure. A passport holder (or visitor) $\mathcal{U}$ presents $\mathcal{B}$ and $\mathcal{P}$ to an inspecting officer who has $\mathcal{P}_{\mathcal{T}_0}$, a set of eigenvectors. In *Phase 1* (solid line), the inspecting officer retrieves an inverse transformation of $\mathcal{T}_0$ from $\mathcal{B}_{\mathcal{T}_0}$ and $\mathcal{P}_{\mathcal{T}_0}$ [27], and verifies its validity by retrieved information. In *Phase 2* (dashed line), the inspecting officer computes transformation $\mathcal{T}_2$ for deriving a corresponding cryptographic key, and verifies a digital signature S on the information including $\mathcal{P}$. A PKI supports the guaranteed verification of the keying values. In this paper, we technically define $\mathcal{B}_0$ as a face and $\mathcal{B}_1$ as a fingerprint. However, various biometrics can be applied to our scheme in the further study, for example, iris codes for $\mathcal{B}_1$ [10, 11]. More details of Phase 1 and Phase 2 are described in Section 3 and 4 respectively.

2.3 Basic Tools

Principal Component Analysis. Principal component analysis (PCA) is a statistical dimensionality reduction method, which produces the optimal linear least squared decomposition of a training set [14, 17]. In a PCA-based face recognition algorithm, the input is a training set $\mathbf{t}_1, ..., \mathbf{t}_W$ of N images such that the ensemble mean is zero ($\sum_i \mathbf{t}_i = 0$). Each image is interpreted as a point in $\Re^{n \times m}$,

where the image is n by m pixels. PCA finds a representation in a $(W-1)$ dimensional space that preserves variance. PCA generates a set of $N-1$ eigenvectors $(e_1, ..., e_{N-1})$ and eigenvalues $(\lambda_1, ..., \lambda_{N-1})$. We normalize the eigenvectors so that they are orthonormal. The eigenvectors are ordered so that $\lambda_i > \lambda_{i+1}$. The λ_i's are equal to the variance of the projection of the training set onto the ith eigenvector. Thus, the low order eigenvectors encode the larger variations in the training set (low order refers to the index of the eigenvectors and eigenvalues). The face is represented by its projection onto a subset of M eigenvectors, which we will call *face space* (see Figure 1). Thus the normalized face is represented as a point in a M dimensional face space.

Biometric Encryption and Digital Signature. Since it is not easy to derive a cryptographic key from biometric information which has variations, much work have been done to use an independent, two-stage process to authenticate the user through biometrics and release the key from hardware storage [9]. Most recently, an innovative technique has been developed by C. Soutar et al [26]. It links the key with the biometric at a more fundamental level during enrollment and then retrieve it using the biometric during verification. Subsequently, a novel technique that generates a digital signature from biometrics has been developed [18].

3 Phase 1: Face Recognition

3.1 Proposed Face Recognition System

Currently, projection based (principal component analysis, independent component analysis, etc.), wavelet based (Gabor), and local feature analysis (LFA) based face recognition systems have been broadly used in face recognition community. We have designed a projection based modular face recognition system which includes preprocessing, feature extraction, and recognition. These are common and essential components for existing (and commercial) face recognition system however, the major difference is design of feature extraction method. In our face recognition system, the feature extraction module has been designed to train feature vector (face space) by precalculating the vectors to maximize discriminating power between intra-class information and minimize differences between inter-class information. We have carefully designed preprocessing and recognition module to optimize overall performance for our projection based face recognition system.

Our face recognition system consists of three modules and each module is composed of a sequence of steps (see Figure 4). The first module performs preprocessing of the input images, $\mathcal{B}_0$. The goal of the preprocessing is to transform the facial image into a standard format that removes variations that can affect recognition performance. This module consists of four steps. The first step filters or compresses the original image. The image is filtered to remove high frequency noise in the image. An image is compressed to save storage space and reduce transmission time. The second step places the face in a standard geometric posi-

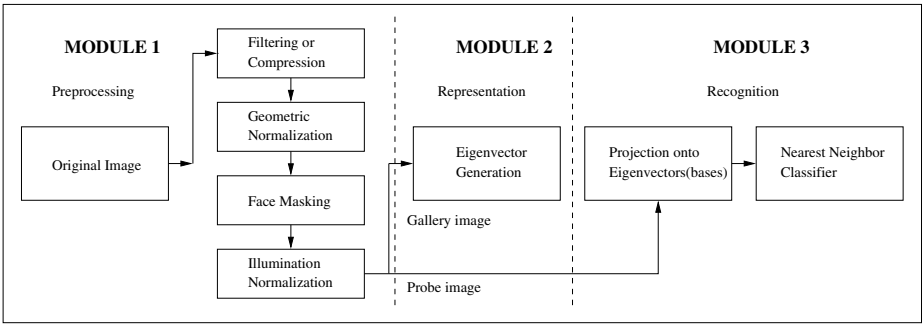


Fig. 4. Block Diagram of Phase 1 System

tion by rotating, scaling, and translating the center of eyes to standard locations. The goal of this step is to remove variations in size, orientation, and location of the face. The third step masks out background pixels, hair, and clothes to remove unnecessary variations which can interfere identification process. The fourth module removes some of the variations in illumination between images. Changes in illumination are critical factors in algorithm performance. The second module performs representation based on the PCA decomposition on the training set. This produces the eigenvectors (eigenfaces) $\mathcal{P}_{\mathcal{I}_0}$ and eigenvalues. The third module identifies the face from a normalized image, and consists of two steps. The first step projects the image onto the eigenface representation, $\mathcal{B}_{\mathcal{I}_0}$. The critical parameter in this step is the subset of eigenvectors that represent the face. The second step recognizes faces using a nearest neighbor classifier. The critical design decision in this step is the similarity measure in the classifier. We present performance results using L1 distance $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sum_{i=1}^{k} |x_i - y_i|$, L2 distance $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 = \sum_{i=1}^{k} (x_i - y_i)^2$, angle between feature vectors

$$d(\mathbf{x}, \mathbf{y}) = -\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = -\frac{\sum_{i=1}^{k} x_i y_i}{\sqrt{\sum_{i=1}^{k} (x_i)^2 \sum_{i=1}^{k} (y_i)^2}}$$

and Mahalanobis distance $d(\mathbf{x}, \mathbf{y}, \mathbf{z}) = -\sum_{i=1}^{k} x_i y_i z_i$ where

$$_i = \sqrt{\frac{\delta_i}{\delta_i + \alpha^2}} \simeq \frac{1}{\sqrt{\delta_i}}, \alpha = 0.25$$

and δ_i = eigenvalue of ith eigenvector. Additionally, Mahalanobis distance was combined with L1, L2, and angle between feature vectors mentioned above [21].

3.2 Test Design

The baseline algorithm has the following configuration: The images are not filtered or compressed. Geometric normalization consists of rotating, translating,

Table 1. Size of galleries and probe sets for different probe categories.

Probe category	duplicate I	duplicate II	**FB**	**fc**
Gallery size	1196	864	1196	1196
Probe set size	722	234	1195	194

and scaling the images so the center of the eyes are on standard pixels. This is followed by masking the hair and background from the images. In the illumination normalization step, the non-masked facial pixels were normalized by a histogram equalization algorithm. The non-masked facial pixels were transformed so that the mean is equal to 0.0 and standard deviation is equal to 1.0. The geometric normalization and masking steps are not varied in the experiments in this paper. The training set for the PCA consists of 501 images (one image per person), which produces 500 eigenfaces. Faces are represented by their projection onto the first 200 eigenvectors and the classifier uses the L_1 norm.

Testing was performed with the FERET protocol [23] and all images are from the FERET database. The target set contained 3323 images and the query set 3816 images. All the images in the target set were frontal images. The query set consisted of all the images in the target set plus non-frontal images and digitally modified images. We report results for four different probe categories. The size of the galleries and probe sets for the four probe categories are presented in table 1. The **FB**, **fc**, and duplicate I galleries are the same. The duplicate II gallery is a subset of the other galleries.

3.3 Experimental Results

Variations in the normalization module (Illumination normalization). We experimented with three variations to the illumination normalization step. For the baseline algorithm, the non-masked facial pixels were transformed so that the mean was equal to 0.0 and standard deviation was equal to 1.0 followed by a histogram equalization algorithm. First variation,the non-masked pixels were not normalized (original image). Second variation, the non-masked facial pixels were normalized with a histogram equalization algorithm. Third variation, the non-masked facial pixels were transformed so that the mean was equal to

Table 2. Performance results for illumination normalization methods. Performance score are the top rank match

	Probe category			
Illumination normalization method	duplicate I	duplicate II	**FB** probe	**fc** probe
Baseline	0.35	0.13	0.77	0.26
Original image	0.32	0.11	0.75	0.21
Histogram Eq. only	0.34	0.12	0.77	0.24
$\mu = 0.0$, $\sigma = 1.0$ only	0.33	0.14	0.76	0.25

0.0 and variance equal to 1.0. The performance results from the illumination normalization methods are presented in table 2.

Variation in Representation module (Number of low order eigenvectors). The higher order eigenvectors which are associated with smaller eigenvalues encode small variations and noise among the images in the training set. One would expect that the higher order eigenvectors would not contribute to recognition. We examined this hypothesis by computing performance as a function of the number of low order eigenvectors in the representation. Figure 5 shows the top rank score for **FB** and duplicate I probes as the function of the number of low order eigenvectors included in the representation in face space. The

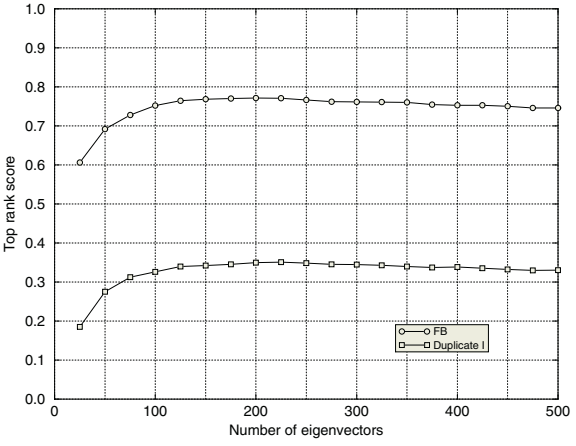


Fig. 5. Performance on duplicate I and FB probes based on number of low order eigenvectors used. (Number of images in gallery = 1196; Number of FB images in probe set = 1195, Number of duplicate I in probe set = 722)

Table 3. Performance scores based on different nearest neighbor classifier. Performance scores are the top rank match

	Probe category			
Nearest neighbor classifier	duplicate I	duplicate II	**FB** probe	**fc** probe
Baseline (L_1)	0.35	0.13	0.77	0.26
Euclidean (L_2)	0.33	0.14	0.72	0.04
Angle	0.34	0.12	0.70	0.07
Mahalanobis	0.42	0.17	0.74	0.23
L_1 + Mahalanobis	0.31	0.13	0.73	0.39
L_2 + Mahalanobis	0.35	0.13	0.77	0.31
Angle + Mahalanobis	0.45	0.21	0.77	0.24

representation consisted of $e_1, \ldots, e_n,\ n = 50, 100, \ldots, 500$, where e_is are the eigenvectors generated by the PCA decomposition.

Variation in Recognition module (Nearest neighbor classifier). We experimented with seven similarity measures for the classifier. They are presented in table 3, along with the results. Based on our experiment, performance score for **fc** probes shows most variation among different category of probes.

3.4 Discussion

We conducted experiments that systematically varied the steps in each module based on our PCA-based face recognition system. The goal belongs to understand the effects on performance scores from these variations. In the preprocessing module, we experimented with varying the illumination normalization step. The results show that performing an illumination normalization step improves performance, but which implementation that is selected is not critical. The results also show that compressing or filtering the images does not significantly effect performance. In the representation module, we varied the number of low order eigenvectors in the representation from 50 to 500 by steps of 50. Figure 5 shows that performance increases until approximately 200 eigenvectors are in the representation and then performance decreases slightly. Representing faces by the first 40% of the eigenvectors is consistent with results on other facial image sets that the authors have seen. In the recognition module, the similarity measure in the nearest neighbor classifier was changed. This variation showed the largest range of performance. For duplicate I probes, performance ranged from 0.31 to 0.45, and for **fc** probes the range was from 0.07 to 0.39. For duplicate I, duplicate II and **FB** probes, the angle+Mahalanobis distance performed the best. For the **fc** probes, the L_1+Mahalanobis distance performed the best. But, this distance was the worst for the duplicate I probe. Because of the range of performance, it is clear that selecting the similarity measure for the classifier is the critical decision in designing a PCA-based face recognition system. However, decision of selecting similarity measure is dependent on the type of images in the galleries and probe sets that the system will process.

Our experimental results has a number of implications for border control applications. First, face recognition system should include a range of images in terms of quality. For example, when measuring the concord between algorithm and human performance, the results should be based on experiments on multiple probe categories. Second, the fine details of algorithm implements can have significant impact on results and conclusion. Our face recognition system can be easily extended for border control applications since the majority of the algorithms in the literature are view-based and have the same basic architecture.

4 Phase 2: Digital Signature Manipulation

4.1 Assumption

Our digital signature manipulation system consists of three modules and each module is composed of a sequence of steps (see Figure 6). The first and second modules perform key generation and signature generation, respectively, and run as a preprocessing for issuing the passport. The third module runs for actual authentication of a passport holder to an inspecting officer. Thus this module may only run in real time.

In our scheme, we suppose to use a simple hash-and-sign RSA (Rivest, Shamir, Adleman) primitive in a probabilistic manner (with κ_Σ-bit random numbers). The original public-private keys are respectively $\langle e, N\rangle$ and $\langle d, N\rangle$ where N is the product of two distinct large primes p and q, and $ed \equiv 1 \bmod \phi(N)$ for the Euler totient function $\phi(N) = (p-1)(q-1)$ [24]. The public key is postulated to be certified by the CA but not in the original form (see Section 4.2). We assume $\mathcal{S}$ returns signature on a message m; $\langle s, r\rangle$ where $s \leftarrow H(m, r)^d \bmod N$ and $r \leftarrow_R \{0,1\}^{\kappa_\Sigma}$. The two-party RSA is the case that the private key is split into two shares such that $d \equiv d_1 d_2 \bmod \phi(N)$ [1, 5]. For our manipulation, a drawback of RSA is the huge size of key. Though we have chosen RSA for wide acceptance, it is considerable to use a different signature scheme in a GDH (Gap Diffie-Hellman) group over E/F_{3^l} [4] or an elliptic curve group for more spatial efficiency and easier manipulation on a short cryptographic key.

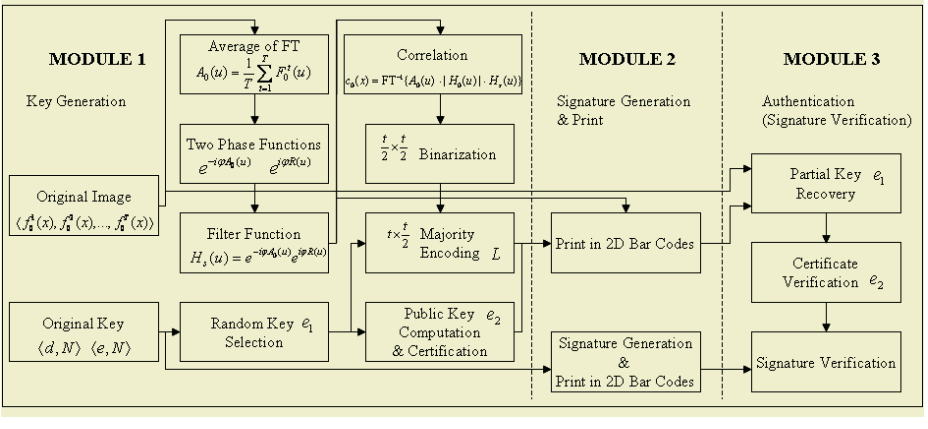


Fig. 6. Block Diagram of Phase 2 System

As for acquiring fingerprint images, the mechanism of correlation is the basis for it [26]. Let $f(x)$ denote a two-dimensional input image array and $F(u)$ its corresponding Fourier Transform (FT) mate, where x denotes the space domain and u the spatial frequency domain. Then correlation is normally used to provide a single scalar value which indicates the degree of similarity between one

image, $f_1(x)$, obtained during verification and another obtained during enrollment, $f_0(x)$, that is represented by the filter function, $H(u)$, derived from a set of $T(\geq 1)$ training images $\langle f_0^1(x), f_0^2(x), ..., f_0^T(x)\rangle$. The correlation function is formally defined as

$$c(x) = \int_{-\infty}^{\infty} f_1(v) f_0^*(x+v) dv$$

where $*$ implies the complex conjugate. In practice, it is obtained by computing the inverse Fourier Transform (FT^{-1}) such that $c(x) = \mathrm{FT}^{-1}\{F_1(u)F_0^*(u)\}$, where $F_0^*(u)$ is represented by $H(u)$ that must be the biometric template tolerant to distortion in correlation-based biometric systems [26]. Let $A_0(u)$ be an average of $F^t(x)$'s for $0 \leq t \leq T$. The stored filter function is defined as

$$H_s(u) = e^{-i\varphi_{A_0}(u)} e^{i\varphi_R(u)}$$

where the phase of the complex conjugate of the training set images, $e^{-i\varphi_{A_0}(u)}$, and the random phase-only function, $e^{i\varphi_R(u)}$, are only multiplied. The magnitude terms of the optimal filter can be calculated on-the-fly during either enrollment or verification and are denoted by $|\cdot|$.

4.2 Key Generation and Signature Generation

$\mathcal{G}_\Sigma(1^\ell)$ outputs an RSA public-private key pair, $\langle e, N\rangle$ and $\langle d, N\rangle$. We assume that the exponent e is chosen from a reasonably large space, say in κ-bit length, so that another d can be approximately in λ-bit length. As for the public exponent, a t-bit integer e_1 is chosen at random to be relatively prime to $\phi(N)$ and e_2 is computed for a large integer k as follows:

$$e_2 = ee_1^{-1} \bmod \phi(N) + k\phi(N)$$

Note that e_2 is to be huge, for example, about $\log k + \ell$ bits, while e_1 is small, for example, only 128 bits [18]. The main difference from the ordinary RSA signature is that the public exponent is the splits such as e_1 and e_2, not e. The value e is discarded and will never be known to others. Note that the CA may certify $\langle e_2, N\rangle$ (with e_1 being proved), instead of $\langle e, N\rangle$. Thus, e_1 can be regarded as a secret value.

A series of input images $\langle f_0^1(x), f_0^2(x), ..., f_0^T(x)\rangle$ are given as $\mathcal{B}_1$ and combined with a random phase array to create two output arrays, $H_s(u)$ and $c_0(x)$, where $H_s(u) = e^{-i\varphi_{A_0}(u)} e^{i\varphi_R(u)}$ and $c_0(x) = \mathrm{FT}^{-1}\{A_0(u) \cdot |H_0(u)| \cdot H_s(u)\}$ [26].

Given the partial key e_1, the central $\frac{t}{2} \times \frac{t}{2}$ portion of $c_0(x)$ must be extracted and binarized for marority-encoding e_1. A complex element $a + bi$ at position (x, y) of the $\frac{t}{2} \times \frac{t}{2}$ portion of $c_0(x)$ will be fragmented in the way that a will appear at (x, y) and b at $(x + \frac{t}{2}, y)$ in the $t \times \frac{t}{2}$ binarized template [26]. Now the binarized template, bt, contains $\frac{t^2}{2}$ real values that can be binarized with respect to 0.0, i.e., set as 1 if they are equal to or greater than 0.0, and otherwise 0. From bt, we can compose a lookup table, L, which may encode e_1 in the way that a number of locations whose element values are equal to each bit of e_1 are stored in each corresponding column.

Finally the user's possession $\mathcal{P}(=\langle \mathcal{B}_{\mathcal{T}_0}, \mathcal{B}_{\mathcal{T}}, \mathcal{P}_{\mathcal{T}} \rangle)$ is defined as $\mathcal{B}_{\mathcal{T}} = \{H_s(u), L\}$ and $\mathcal{P}_{\mathcal{T}} = \{e_2, N\}$, while $\mathcal{B}_{\mathcal{T}_0}$ was obtained in Phase 1. $\mathcal{P}$ is encoded and printed by an arbitrary 2D bar code on the user's passport. A digital signature on the passport holder's information including $\mathcal{P}$ is generated by using $\langle d, N \rangle$. Given the information M (say, identification contents in a passport), $\mathcal{S}$ raises it to the power of d for obtaining the corresponding signature $\sigma = M^d \bmod N$. The signature S is also printed on the passport in barcodes. Note that $\mathcal{P}_{\mathcal{T}}(= \{e_2, N\})$ can be certified by authority in PKIs, so that the inspecting officer could verify the key integrity without solely depending on paper tamper-proofing techniques in passports.

4.3 Signature Verification

A passport holder $\mathcal{U}$ provides a series of fingerprint images $\langle f_1^1(x), f_1^2(x), ..., f_1^T(x) \rangle$ as input $\mathcal{B}_1$ along with $\langle \mathcal{B}_{\mathcal{T}_0}, \mathcal{B}_{\mathcal{T}}, \mathcal{P}_{\mathcal{T}} \rangle$, say $\langle H_s(u), L, e_2, N \rangle$, in 2D bar codes. We assume $\langle e_2, N \rangle$ is provided with a certificate and verified in this step. We also postulate that $\mathcal{B}_{\mathcal{T}_0}$ is already verified by an inspecting officer under the comparison of photo id in the passport.

A series of input images are combined with $H_s(u)$ to create a new output array, $c_1(x)$ where $c_0(x) = \mathrm{FT}^{-1}\{A_1(u) \cdot |H_1(u)| \cdot H_s(u)\}$.

Given the lookup table L, the central $\frac{t}{2} \times \frac{t}{2}$ portion of $c_1(x)$ must be extracted and binarized for majority-decoding e_1. A method to obtain the new binarized template, bt', is exactly the same to that of key generation process. From bt' and L, we can compose a new table L' which may majority-decodes e_1 in the way that a majority bit in each column is derived to each location in e_1.

Given the signature S from the passport, $\mathcal{V}$ raises it to the power of e_1 and subsequently the result to the power of e_2 for verifying $M \equiv (S^{e_1})^{e_2} \pmod N$. This is obvious because $e \equiv e_1 e_2 \equiv e_1\{ee_1^{-1} \bmod \phi(N) + k\phi(N)\} \pmod{\phi(N)}$.

4.4 Discussion

Security. There have been many theoretical and practical attacks on RSA with regard to the length of public and private exponents, and the property of modulus N [2]. For example, M. Wiener first showed that instances of the RSA cryptosystem using low private exponents (that are not exceeding approximately one-quarter of the length of N) are insecure in 1990 [29], while Coppersmith et al. presented an attack on the instance using low public exponents [8]. Also, Boneh and Durfee improved the related attack and showed the higher boundary for the Wiener's attack [3]. As for those attacks against RSA, our scheme is secure because we used a large private exponent and a public exponent in reasonable length. The value e_1 is secure without any tamper-resistant device. It is only recovered by live biometrics and 2D bar codes. The main difference from the ordinary RSA signature is that the public exponent is the splits such as e_1 and e_2, not e. The value e is discarded and is never known to others. Note that our public exponents are manipulated carefully in linking biometric information.

Practicality. In our system, the size of e_2 was assumed about $k + \ell$ bits where k is less than ℓ. So the liveness check for $\mathcal{B}$ is additionally necessary while its minimized template can be stored in $\mathcal{P}$, say exactly $\mathcal{P}_{\mathcal{T}}$, under the easy consideration of the hill-climbing attack. Note that a passport holder possess $\mathcal{P}$. When we consider the number of the most expensive modular N multiplications [20], our RSA signature verification using the repeated square-and-multiply algorithm will take $t + k + \ell$ (approximately 2ℓ) modular squarings and expected $\frac{t+k+\ell}{2}$ (approximately ℓ) modular multiplications. This means only the double of the usual RSA signature generation time. Note that we could apply discrete logarithm based digital signature schemes using smaller private exponents in much easier ways. However, the difficulty of certificate revocation is one drawback of our scheme and must be resolved in the future study.

5 Conclusion

In this paper, we explored a low-cost but highly-scalable method that uses multi-modal biometrics based on face and fingerprints, and public key infrastructures (PKIs) for border control applications. It is widely recognized that multi-modal biometrics has the potential to strengthen border protection by reducing the risk of passport fraud. However, it may take high costs to issue smart-card enabled passports over the world and to process a huge amount of biometric information (on-line). A digital signature in PKIs and multi-modal biometrics are carefully applied in our scheme, in order to reduce the possibility of undesirable factors significantly at nation's borders without requiring any hardware device in passports.

In our scheme, a passport holder presents his or her biometrics (face and fingerprint) along with the passport to an inspecting officer who can access a set of eigenvectors. Our scheme is proceeded in two distinct phases. In Phase 1, the inspecting officer retrieves an inverse transformation of $\mathcal{T}_0$ from the biometric representation $\mathcal{B}_{\mathcal{T}_0}$ and the eigenvectors $\mathcal{P}_{\mathcal{T}_0}$ [27]. Subsequently, the officer verifies its validity by retrieved information (under a demographic picture in the passport). In Phase 2, the inspecting officer computes transformation $\mathcal{T}_2$ for deriving a corresponding public exponent (that is camouflaged in length 2λ), and verifies a digital signature on the demographic information including $\mathcal{P}$. Note that we can consider iris patterns for transformation $\mathcal{T}_1$ and $\mathcal{T}_2$, instead of fingerprints, in the future study.

Acknowledgement

This research was supported in part by University IT Research Center Project, and also was supported in part by Korea Research Foundation Grant (R08-2003-000-11029-0).

References

1. M. Bellare and R. Sandhu, "The security of practical two-party RSA signature schemes," Manuscript, 2001.
2. D. Boneh, "Twenty years of attacks on the RSA cryptosystem," Notices of the American Mathematical Society (AMS), vol. 46, no. 2, pp.203-213, 1999.
3. D. Boneh and G. Durfee, "Cryptanalysis of RSA with private key d less than $N^{0.292}$," Eurocrypt '99, LNCS (1592), Springer-Verlag, pp.1-11, 1999, and IEEE Trans. on Information Theory, vol. 46, no. 4, 2000.
4. D. Boneh, H. Shacham, and B. Lynn, "Short signatures from the weil pairing," Asiacrypt '01, LNCS vol. 2139, Springer-Verlag, pp.514-532, 2001.
5. C. Boyd, "Digital multisignatures," *Cryptography and Coding*, Oxford University Press, pp.241-246, 1989.
6. S. Brands, *Rethinking public key infrastructures and digital certificates*, The MIT Press, p.11 and pp.219-224, 2000.
7. H. E. Burke, "Handbook of bar Coding Systems," *Van Nostrand Reinhold*, New York, N.Y., 1984.
8. D. Coppersimth, M. Franklin, J. Patarin, and M. Reiter, "Low-exponent RSA with related messages," Eurocrypt 1996, LNCS, vol. 1070, pp.1-9, 1996.
9. Daon Inc., "Biometric Authentication & Digital Signatures for the Pharmaceutical Industry," White paper available at `http://www.daon.com/downloads/publications/esignature.pdf`
10. J. Daugman, "High confidence personal identifications by rapid video analysis of iris texture," IEEE International Carnahan Conference on Security Technologies, pp.50-60, 1992.
11. J. Daugman, "High confidence personal identifications by a test of statistical independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, no.11, pp.648-656, November 1993.
12. G. Davida, Y. Frankel, and B. Matt, "On enabling secure applications through offline biometric identification," IEEE Symposium on Security and Privacy, pp.148-159, 1998.
13. Denso Inc., "QRmaker: User's Manual," *Denso Corporation*, Aichi, Japan, 1998.
14. K. Fukunaga, "Introduction to statistical pattern recongition," Academic Press, Orlando, FL, 1972.
15. S. Goldwasser, S. Micali, and R. Rivest, "A digital signature scheme secure against adaptive chosen-message attacks," *SIAM Journal of Computing*, vol.17, no.2, pp.281-308, Apr. 1988.
16. A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM*, February 2000.
17. I. Jolliffe, "Principal Component Analysis," Springer-Verlag, 1986.
18. T. Kwon, "Practical digital signature generation using biometrics," Proceedings of ICCSA 2004, LNCS, Springer-Verlag, 2004.
19. V. Matyáš and Z. Říha, "Biometric authentication - security and usability", Manuscrypt available at `http://www.fi.muni.cz/usr/matyas/cms_matyas_riha_biometrics.pdf`
20. A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, pp.287-291, pp.312-315, 1997.
21. H. Moon, "Performance Evaluation Methodology for Face Recognition Algorithms," Ph.D. Thesis, Dept. of Computer Science and Engineering, SUNY Buffalo, 1999.

22. Roger. C. Palmer, "The Bar Code Book," *Helmers Publishing*, Peterborough, N.H., 3rd Ed., 1995.
23. P. Phillips and H. Moon and S. Rizvi and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Pattern Analysis and Machine Intelligence*, vol.22, pp.1090-1104, 2000.
24. R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol.21, pp.120-126, 1978.
25. C. Soutar, "Biometric system performance and security," Manuscrypt available at `http://www.bioscrypt.com/assets/bio_paper.pdf`, 2002.
26. C. Soutar, D. Roberge, A. Stoianov, R. Golroy, and B. Vijaya Kumar, "Biometric Encryption," ICSA Guide to Cryptography, McGraw-Hill, 1999, also available at `http://www.bioscrypt.com/assets/Biometric_Encryption.pdf`
27. M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, No. 1, pp. 71-86, 1991.
28. U.S. DoS, USVISIT, `http://fpc.state.gov/20738.htm`.
29. M. Wiener, "Cryptanalysis of short RSA secret exponents," *IEEE Transactions on Information Theory*, vol.36, no.3, May 1990.

Risk Management Using Behavior Based Bayesian Networks

Ram Dantu and Prakash Kolan

Department of Computer Science,
University of North Texas
{Rdantu, prk0002}@cs.unt.edu

Abstract. Security administration is an uphill task to implement in an enterprise network providing secured corporate services. With the slew of patches being released by Microsoft, HP and other vendors, system administrators require a barrage of tools for analyzing the risk due to these vulnerabilities. In addition to this, criticalities in patching some end hosts (eg., in hospitals) raises serious security issues about the network to which the end hosts are connected. In this context, it would be imperative to know the risk level of all critical resources (e.g., Oracle Server in HR department) keeping in view the everyday emerging new vulnerabilities. We hypothesize that sequence of network actions by an attacker depends on the social behavior (e.g., skill level, tenacity, financial ability). We extended this and formulated a mechanism to estimate the risk level of critical resources that may be compromised based on attacker behavior. This estimation is accomplished using behavior based attack graphs. These graphs represent all the possible attack paths to all the critical resources. Based on these graphs, we calculate the risk level of a critical resource using Bayesian methodology and periodically update the subjective beliefs about the occurrence of an attack. Such a calculated risk level would be a measure of the vulnerability of the resource and it forms an effective basis for a system administrator to perform suitable changes to network configuration. Thus suitable vulnerability analysis and risk management strategies can be formulated to efficiently curtail the risk from different types of attackers (script kiddies, hackers, criminals and insiders).

1 Introduction

The increase in the size of the enterprise network is an ever-growing process. With the increase in number of hosts connected to the network, there is always a mounting risk of protecting computers from the outside attacks. In addition to this, improper configuration of network hosts result in host vulnerabilities because of which the hosts are susceptible to outside attacks. Accurate vulnerability analysis require a deep understanding of failure and attack modes and their impact on each of the network components, and the knowledge of how these components interact with each other during normal as well attack modes of operation. For managing the security of a network, security engineers identify security holes by probing the network hosts, asses the risk associated with the vulnerabilities on the computer hosts and fix host vulnerabilities by using patches released by the vendors.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 115 – 126, 2005.

Patching up network hosts is a short-term solution for avoiding an attack, but this requires fixing the vulnerabilities in all of the network hosts and its components. We see frequent release of patches from product vendors (Microsoft, IBM, HP) etc. to reduce the effect of vulnerability once it is reported. The product vendors, for the process of vulnerability assessment, focus on active prevention methodologies of closing the vulnerabilities before they are exploited. But this process of patching end hosts requires a great deal of human intervention, time and money. It involves frequent monitoring of end systems using a set of monitoring tools by the admin staff to identify and prevent intrusion. The situation worsens when the already present state of the art monitoring tools are not effective in identifying new vulnerabilities.

Risk management refers to process of making decisions that would help in minimizing the effects of vulnerabilities on the network hosts. In context of high exploit probability, risk management is a nightmare to plan with. And also, it is very difficult to identify new exploits and vulnerabilities. For many years security engineers have been doing risk analysis using economic models for the design and operation of risk-prone, technological systems([1], [3], [4], [6]) using attack profiles. Considerable amount of research has been reported in developing profiles of the attacker based on the evidence he leaves behind during an attack. The evidence collected can be used in estimating the type of attacker. Based on the type of attacker identified, effective risk management policies can be formulated for the network.

Simultaneously, a great deal of psychological and criminological research has been devoted to the subject; but the security engineers do not use these studies. We believe that integrating this research could improve the process of risk analysis. Many articles explain how intruders break into systems ([14], [15]). Companies like *Psynapse, Amenaza*, and *Esecurity* have built products using the behavior of intruders. To our knowledge, no work has been reported on integrating behavior-based profiles with sequence of network actions for computing the vulnerability of resources. *The overall goal of this research is to estimate the risk of a critical resource based on attacker behavior and a set of vulnerabilities that can be exploited.* This implies a more fine-grained repertoire of risk mitigation strategies tailored to the threat rather than blanket blocking of network activity as the sole response.

2 Background

A considerable amount of work has been reported on attacker profiles and risk management on an individual basis. But none of them attempted in integrating risk analysis with attacker behavior. Jackson[4] introduces the notion of behavioral assessment to find out the intent behind the attack. The proposed Checkmate intrusion detection system distinguishes legitimate use from misuse through behavior intent. But this does not propose any detail on vulnerable device identification based on the assessed behavior. Rowley[17] views risk analysis to involve threat identification, risk assessment and steps to be taken for mitigating the risk. The issues that are identified to be of potential threats are identified and an estimate of damage the threat could pose is calculated. There is a need of integrating risk mitigation strategies with attack behavior to help reduce the possibility of impending attacks.

The psychological and criminological research on hacker community attempts to define different categories of hackers based on their intent, skill and attack proficiency. Categories of hackers like novices, crackers and criminals have been defined. Each of the hacker groups has their own knowledge and motivation for carrying on the attacks. Rogers[16] proposed different categorizations of a hacker community and advices derivation of hacker profiles using intruder behavior. Yuill[1] profiles detection of an on-going attack by developing a profile of the attacker using the information he reveals about himself during his attacks. Kleen[9] developed a framework by reviewing existing methods and advances in a way that hackers are classified and profiled, with the goal of better understanding their values, skills and approaches to hacking. There are several works in the literature on the hacker profiles ([6], [7], [8]) but none of them tie the profiles to any exploits in the network. All the theories proposed account for the hacker behavior, but don't attempt to relate the reasons behind hacker behavior to exploits and vulnerability utilization.

On the other hand, attack graphs are beginning to be used to formalize the risk for a given network topology and exploits. Sheyner[12]attempts to model a network by constructing attack graph for the model using symbolic model checking algorithms. Moore[11] documents attacks on enterprises in the form of attack trees, where each path from the root to the end node documents how an attacker could realize his desire of exploiting the host and ultimately the network. However, current research [10] [11] [12] does not combine the behavior with these graph transitions.

Loper[5][13] indicates that mapping network actions to social motives is sustained by the available data. It is relatively well established in social science that measurable attitudes and observable actions can predict specified behavior (within a known level of error). This paper marries profiling with chain of exploits, and detects highly vulnerable resources in the network. In addition, behavior profiles are used for calculating the trust of a given attack path. Our work uses the theory from criminology, statistical analysis, behavioral-based security, and attack graphs.

3 Methodology

Attack graphs or attack trees are been increasingly formalized to be a model for representing system security based on various attacks. An attack tree can be visualized to be a graph consisting of a group of nodes with links interconnecting them. Attack graphs can be drawn to represent the sequence of network actions for exploiting each network resource and ultimately the whole network. We use attack graphs for calculating the vulnerability level and risk of a critical resource in a given network for different attacker profiles. There are five steps in our procedure. The five steps are repeatedly executed until optimum security is achieved. *Our hypothesis is that there is a relation between network actions and social behaviour attributes.*

3.1 Step 1: Creation of an Attacker Profile

The profile an attacker gives the expendable resources associated with the attacker. These resources can be any of cost, computer and hacking skills, tenacity, perseverance, motives like revenge, reputation etc. that the attacker would expend to

exploit a vulnerability. Different attack profiles have different behavioral attribute values for attacker resources. For example, a corporate espionage has more money compared to a script kiddie who tries to hack for fun with little money. A corporate insider has more knowledge regarding the enterprise network topology compared to a hacker. One example for assigning relative attributes for a profile is for a hacker who has low level of funding (e.g., 0.2), medium level of skill (e.g., 0.6) and high level of tenacity (e.g., 0.8).

3.2 Step 2: Creation of Attack Graphs

An attack graph can be created using network topology, interconnection between hosts, and various vulnerabilities of each host ([10], [11], [12]). In this graph, each path identifies a series of exploits. Using this graph, we can learn how intruders culminate sequence of state transitions for achieving an attack. For example, an attack path in an attack graph [see Fig. 1] can be a sequence of events like overflow *sshd* buffer on host1(H1), overwrite *.rhosts* file on host2(H2) to establish *rsh* trust between H1 and H2, log-in using *rsh* from H1 to H2, and finally, overflow a local buffer on H2 to obtain root privileges. An attack graph can be shown as a causal graph with each node representing a cause and its child node representing an effect. Each node in the graph represents an event, and a path from root to leaf represents a successful attack.

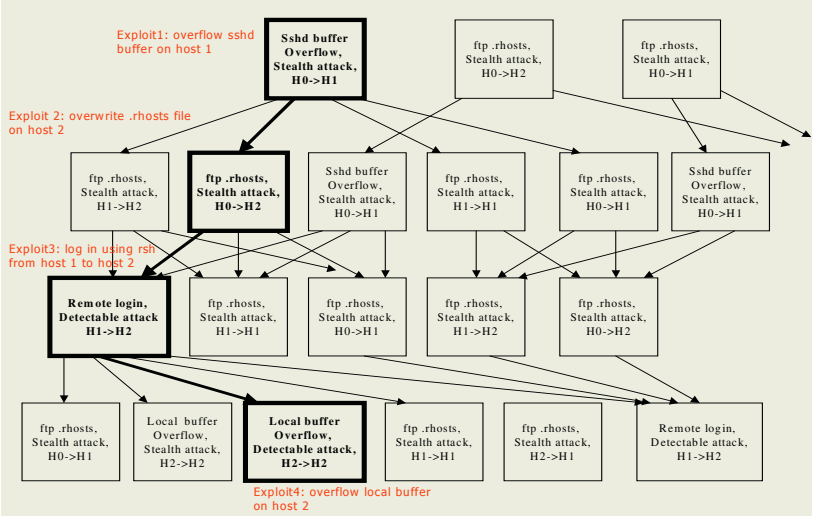


Fig. 1. An example attack graph with a chain of exploits

3.3 Step 3: Assigning Behavior Attributes to Attack Graph Nodes

For a given attacker profile, the nodes of the attack graph can be labelled using a set of behaviour attributes like: i) computer skills, ii) Hacking skills iii) tenacity iv) cost of attack v) techniques for avoiding detection etc. for carrying out the events

represented by them. We are in midst of conducting a survey which would help in profiling attack behavioural aspects such as how different people would behave in attack scenarios given expendable resources at their disposal[20]. Using this, for a given profile, the attack graphs based on that profile are constructed by documenting all the attack paths that could be possibly executed by that profile. For example, Fig. 2 represents attack graphs constructed for two example profiles A & B respectively for three example attributes cost, skill and tenacity. These profile based attack graphs give a source of analysis for inferring profile based attacks.

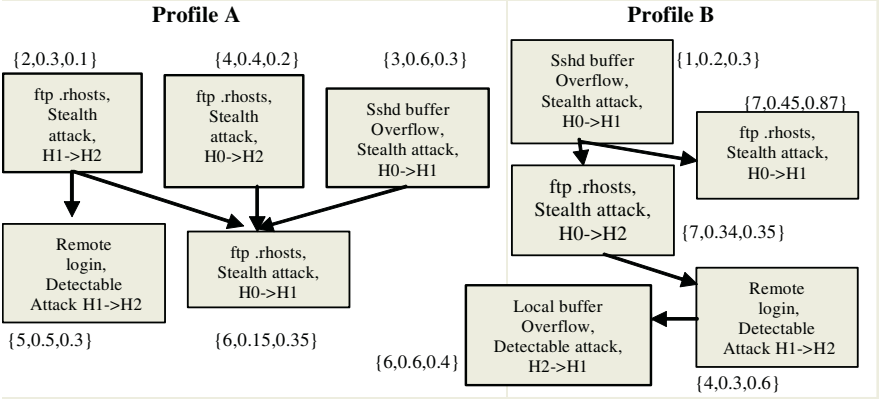


Fig. 2. Attack paths based on profiles from Fig. 1 (3 Tuple{Cost, Skill, Tenacity})

3.4 Step 4: Risk Computation

In this step, a risk level for all the critical resources is calculated based on the set of paths, attributes and attacker type (e.g., script kiddie, hacker, corporate insider etc.). Bayesian networks based estimation can be used for calculating the aggregated risk value of the resource. Next, a resource is marked as attack prone if this value is more than a threshold.

3.4.1 Deriving Risk of an Attack Path

Based on the type of the attacker, the attack paths are considerably different depending on the type of quantifying variable in consideration. The eventual path of the attacker would be his optimized use of the quantifying variables such as cost, skill, tenacity etc. Thus the final attack path "Θ" taken by the attacker would be a function of individual attack paths i.e. $\Theta = (f_1, f_2 \ldots fn)$ where each f_i is the attack path that an attacker would take for an identifier variable "i". Each f_i can be calculated by documenting individual attack paths of the attack graph. An attack path with nodes of "n" number of attributes in Fig. 3(a) can be represented as in Fig. 3(b).

Table 1 describes all the behavioral attributes for each attack path and exploits. Given an attacker profile, all the attack paths that the attacker can move are described in this table. In this way we can derive the relationship between sequence of network actions and the social motives behind the attacker to carry out the attack.

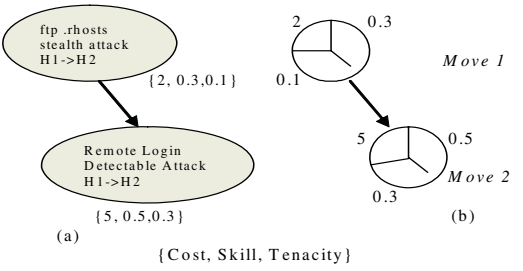


Fig. 3. An attack path of profile A in Fig. 2

Table 1. Probabilities for each move in Fig. 3

Path

Move	Skill	Tenacity	Cost	….
1	0.3	0.1	2	
2	0.5	0.3	5	
….				

3.4.2 Bayesian Networks for Risk Inference

A Bayesian network is a graphical model for showing probabilistic relationships among a set of variables (representing nodes) in a graph. Each node or variable is associated with a set of Probability Distribution Functions. Therefore the attack graphs can be modelled by reducing them to causal graphs and associate the nodes with probabilities. Using monitoring or intrusion detection systems, protocol state machines and traffic patterns observed between various states in the state machine, the initial subjective beliefs can be formulated. Any deviation from normal behaviour gives the evidence for calculating the posterior probabilities using Bayesian inference techniques. Bayesian statistics helps us to quantify the available prior probabilities or knowledge based on the evidence colleted at any node in the network. The evidence thus collected updates the subjective belief of all the other random variable probability distributions. The new posterior probability distributions designate the updated subjective beliefs or the possibilities of the intermediate network actions to achieve the overall goal of exploiting the vulnerabilities existing in the network and its components. These posterior probability calculations are done before and after the exploits are patched to estimate the new risk level of the critical resources.

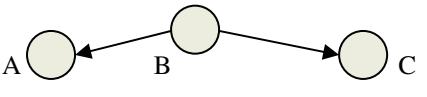


Fig. 4. Representing Conditional Probability

Fig. 4 is a simple Bayesian network with three nodes. The joint prob. distribution function for the figure can be shown to be P(A,B,C) i.e. P(A/B)*P(B)*P(C/B).

Therefore for set of variables in $X = X_1, X_2, \ldots\ldots X_N$, the prob. distribution would be

$$P(X_1,X_2,\ldots\ldots X_N) = \prod_{i=1}^{n} P(X_i / parent(X_i)) \quad (1)$$

3.4.3 Inference Based on Attacker Profiles

As shown in the previous sections the posterior probabilities of nodes converge depending upon the statistical dependencies existing due to various parent child relationships. Similarly, the attack graph for a given profile is initialized using expert knowledge and past observations[20]. Expert knowledge provides with profile information about all the probabilities of attack. The observations using expert knowledge can provide a basis for Bayesian probability estimation.

For example, consider one segment of attack graph shown in Fig. 1. Fig.5 represent the quantifying variables {cost, skill, tenacity} required by a profile to exploit the remote login and ftp .rhosts attacks. With the available initial knowledge, we compute the inferred probability for observed evidence at node E. Assume each of the nodes to be in two states "yes" or "no", and the probability values obtained from expert knowledge to the nodes are

P(A = yes) = 0.1, P(B = yes) = 0.35, P(C = yes) = 0.2
P(D=yes |A=yes)=0.3, P(D=yes |A=no)=0.4
P(E=yes |C=yes ,B=yes ,A= no) = 0.25 P(E=yes |C=yes ,B=yes ,A=yes) = 0.15

Then, if an attacker is using the *.rhosts* stealth attack at node E, then prob. that .rhosts attack at node A was carried out can be calculated by P(A/E, D,C,B).

$$P(A/E, D,C,B) = \frac{P(E, D, C, B, A)}{\sum P(E, D, C, B, A^1)} \quad (1)$$

$$= \frac{P(E / C, B, A) * P(D / A) * P(A)}{\sum P(E / C, B, A^1) * P(D / A^1) * P(A^1)} \quad (2)$$

$$= \frac{(0.15 * 0.7 * 0.1)}{(0.15 * 0.7 * 0.1) + (0.25 * 0.6 * 0.9)} = 0.0721 \quad (3)$$

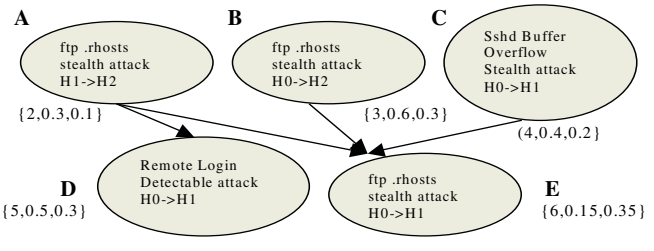


Fig. 5. A small Bayesian Causal graph

The probability before was 0.1, but the inferred probability is 0.0721 based on the values of other variables. For a given resource, we document all attack paths and calculate Bayesian probabilities of the root nodes of each attack path when the evidence regarding the leaf is available. Tab. 2 describes inferred probabilistic values for a given profile and attack path AE of Fig. 5. This procedure is carried out for all attack paths and profiles that are capable for carrying out an attack. Hence, for a given resource, all probable attack paths that can lead to the exploitation of it can be inferred.

Table 2. Bayesian prob. at the root node of attack path given evidence at the leaf

Path	Skill	Tenacity	Cost	…	…	…	…
1	0.072	0.33	1.82				
2							
….							

3.4.4 Relating Risk, Behavior and Penetration

As we described before we believe that sequence of network actions carried out by an attacker relate to social behaviour. We attempt to derive the relation between vulnerability of a given resource and the penetration an attacker can achieve in exploiting the network. This can be achieved by defining the probability of each event in the attack path and inferring the posterior probability given evidence at a node, usually the leaf node i.e. the node representing the final event for a successful attack. Fig. 6 is a part of an attack graph of Fig. 1 and the prob. of the nodes are represented using Conditional Probability Tables (CPT). The CPT tables give the probability of nodes given the value of its parents. Assume each node to be in two states “yes” or

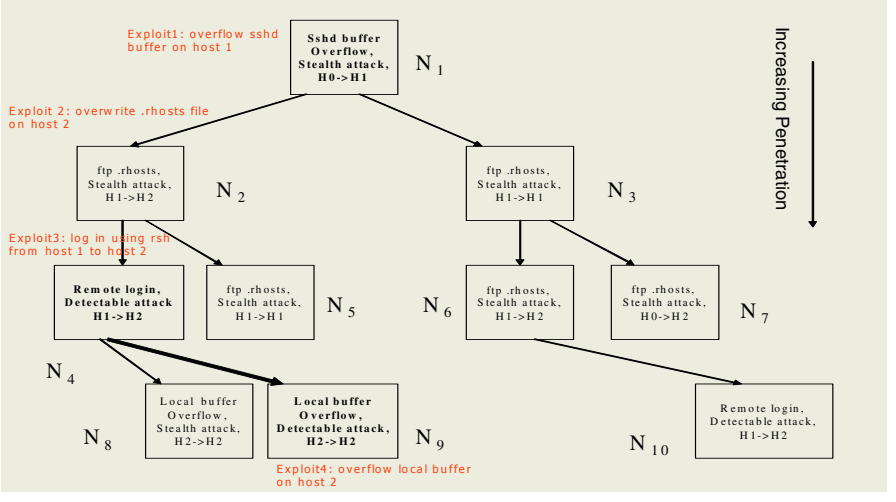


Fig. 6. Example Sub-Attack Graph from Fig. 1

"no". Representing Node 1 as N_1, Node 2 as N_2 etc. and probability of Node X as $P(N_X)$. For the four different profiles Corporate Insider, Corporate espionage, hacker and script kiddie, and five nodes in analysis N_1, N_2, N_4, N_8 and N_9, the CPT tables would be as in Tab 3a and Tab 3b. (*In reality, initialisation of CPT tables is carried out by analysing a statistical data from an interview or a survey*[20]).

Table 3a. Prob. of nodes N_1, N_2 ,N_4 of Fig. 6 given their parents (N1 does not have a parent)

Probability Profile	$P(N_1)$	$P(N_2)$ given N_1 = yes	$P(N_2)$ given N_1 = no	$P(N_4)$ given N_2 = yes	$P(N_4)$ given N_2 = no
Corp. Insider	0.8	0.75	0.82	0.85	0.70
Corp. Espionage	0.62	0.65	0.71	0.67	0.63
Hacker	0.6	0.7	0.31	0.51	0.46
Script Kiddie	0.4	0.52	0.36	0.48	0.32

Table 3b. Prob. of nodes N_8 and N_9 of [Fig. 8] given their parents

Probability Profile	$P(N_8)$ given N_4 = yes	$P(N_8)$ given N_4 = no	$P(N_9)$ given N_4 = yes	$P(N_9)$ given N_4 = no
Corp. Insider	0.71	0.83	0.69	0.77
Corp. Espionage	0.7	0.65	0.72	0.64
Hacker	0.3	0.57	0.52	0.63
Script Kiddie	0.62	0.41	0.44	0.62

The values given by the CPT tables describe the behaviour of each of the profiles. For example, a corporate insider who is in the enterprise has more profound knowledge of the corporate network topology and thus the risk posed by the corporate insider is more compared to a corporate espionage. The probabilities of the hacker are understandingly less because a hacker tries to compromise the network resources and the risk associated with this is much less compared to corporate espionage and an insider. Script Kiddie has the least skill, tenacity and knowledge of an enterprise network and hence with limited attributes tries to hack into the network by downloading some network scanning and monitoring tools.

Table 4. Bayesian Inference for directly affected nodes due to evidence at node N_9. N_1 represents minimum and N_4 the maximum penetration

Profile	$P(N_1)$	$P(N_2)$	$P(N_4)$	$P(N_8)$
Corp. Insider	0.8002	0.7609	0.7975	0.7343
Corp. Espionage	0.6199	0.6738	0.6829	0.6841
Hacker	0.5991	0.5416	0.4395	0.4513
Script Kiddie	0.3980	0.4112	0.3102	0.4751

In the figure, if evidence regarding the happening of the event represented by Node 9 is known to happen, nodes N_1, N_2, N_4, N_8 are directly inferred. For example, what is

the probability of a hacker penetrating through N_I-N_2-N_4-N_9 given an event that N_9 is attacked. The inferred probabilities of the nodes directly affected by this evidence using our analytical model (See Sec 3.4.3) are given in Tab. 4. Network Penetration is given by the extent to which the attacker would be able to penetrate i.e., the level of the graph on the attack path.

Fig. 7 represents the relationship between risk, behaviour and network penetration for all the profiles for a given attacker skill level. In reality, the CPT values will be a range instead of single value. The values for two profiles are given in Table 5a and Tab. 5b (α/β for each node represents the range of probabilities of given profile in which the prob. for all attributes of the profile fall into[5]). The range of probabilities of the nodes that are directly inferred by the event at node N_9 are given in Tab 6.

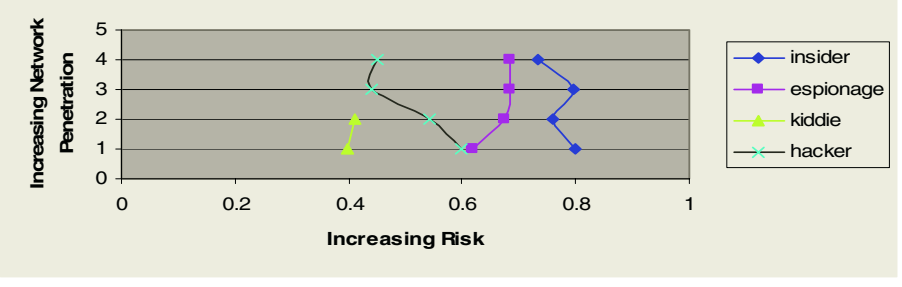


Fig. 7. Relating risk, behavior and penetration for an attribute of all profiles

Table 5a. For given two profiles, range of probabilities of nodes N_1, N_2 and N_4 of Fig. 8

Profile	$P(N_1)$	$P(N_2)$ given N_1 = yes	$P(N_2)$ given N_1 = no	$P(N_4)$ given N_2 = yes	$P(N_4)$ given N_2 = no
Corp. Insider	0.8/0.92	0.75/0.87	0.82/0.94	0.85/0.97	0.70/0.82
Script Kiddie	0.4/0.65	0.52/0.71	0.36/0.56	0.48/0.73	0.32/0.67

Table 5b. For given two profiles, range of prob. of nodes N_8 , N_9 of Fig. 8 given their parents

Profile	$P(N_8)$ given N_4 = yes	$P(N_8)$ given N_4 = no	$P(N_9)$ given N_4 = yes	$P(N_9)$ given N_4 = no
Corp. Insider	0.71/0.83	0.83/0.94	0.69/0.82	0.77/0.89
Script Kiddie	0.62/0.84	0.41/0.63	0.44/0.68	0.62/0.81

Table 6. Inferred prob. range of all the attributes for the given two profiles for directly affected nodes N1, N_2, N4, and N8. N_1 represents minimum and N_4 the maximum penetration

Profile	$P(N_1)$	$P(N_2)$	$P(N_4)$	$P(N_8)$
Corporate Insider	0.8/0.92	0.76/0.874	0.7975/0.974	0.734/0.835
Script Kiddie	0.398/0.649	0.411/0.655	0.310/0.672	0.475/0.771

From Tab. 6, we infer that for all attributes, the probability of node N_1 falls in the range [0.8, 0.92] for the corporate insider and in the range [0.398, 0.649] for the script kiddie. For the given two profiles and two attributes, the relation between the risk, behaviour and penetration looks as in Fig. 8(a). Fig. 8(a) can be extrapolated for all the four profiles and attributes, and can be shown as in Fig. 8(b). Fig. 8(b) shows the relation between behaviour, risk, depth in the graph (relates to sequence of moves) and critical resources. Certain behaviours overlap regardless of the type of threat (e.g. corporate insiders may share some behaviour with outside espionage).

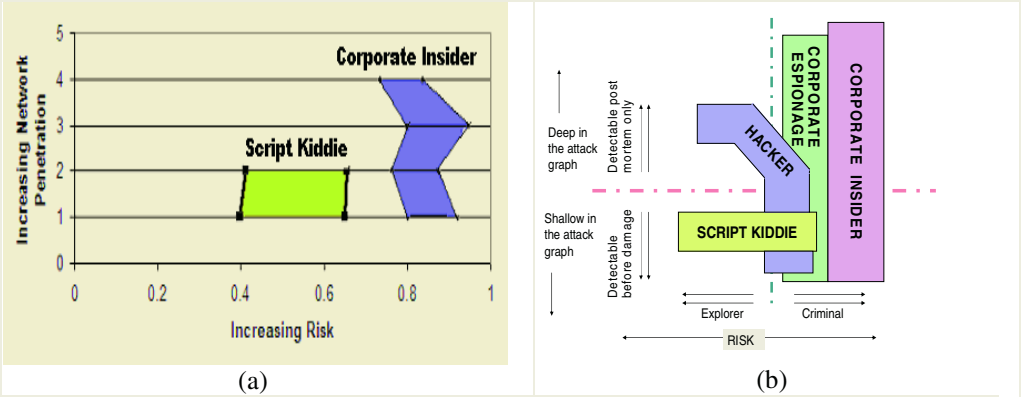


Fig. 8. Relation between risk, behavior and network penetration

3.5 Step 5: Optimizing the Risk Level

In a typical network, patching vulnerability may impact other network elements. For example, after patching some exploits and changing the network configuration (e.g., moving the firewall to another location in the topology or changing the firewall rules, deploying Intrusion detection systems etc.), the steps (Sec 3.1 to 3.4) outlined need to be performed repeatedly for an optimum risk value**.** This estimated risk value would help in processes like patch management and penetration testing etc.

4 Conclusion

Our *hypothesis* is that there is a relation between sequence of network actions and attacker behaviour and that the behaviour can be used for network risk analysis.. This analysis is based on sequence of actions carried out by an attacker and his social attributes. We used attack graphs for representing all possible attacks on a critical resource in the network. We have described a *five-step model* of vulnerable device detection and risk estimation of a network using attack graphs and attacker behaviour. The creation of attack graphs helps us in identifying the possible attacks on a network component. We formulated a mechanism through Bayesian estimation to quantitatively analyze the attack graphs and derive attack paths based on attacker attributes. Risk computation is carried out using Bayesian probability distributions of a set of identifiers at each node in an attack graph. This gives a more appropriate

prediction of risk and threat identification. Finally we suggest optimizing the network by patching the identified vulnerable devices or reconfiguration of network components till a comfortable security level is achieved. Using this methodology, a set of security policies can be formulated to reduce the vulnerability of a network and its hosts to external attacks. Future work includes applying our method to real-world network configurations and testing the methodology on data collected during past attacks.

References

1. Jim Yuill, J., Wu, S.F., Gong, F., Ming-Yuh H.: "Intrusion Detection for an on-going attack", RAID symposium.
2. Scheiner, B.: "Attack Trees: Modeling Security Threats", Dr. Dobb's Journal Dec 99.
3. Desmond, J.: "Checkmate IDS tries to anticipate Hackers Actions", www.esecurityplanet.com/prodser, 12th June, 2003.
4. Jackson, G.: "Checkmate Intrusion Protection System: Evolution or Revolution", Psynapse Technologies, 2003.
5. Loper, K.: "The Criminology of Computer Hackers: A qualitative and Quantitative Analysis", Ph.D. Thesis, Michigan State University, 2000.
6. Modern Intrusion Practicies, CORE security technologies,
7. Know Your Ennnemy: Motives, The Motives and Psychology of the Black-hat Community, 27th June, 2000.
8. Rogers, M.: "Running Head: Theories of Crime and Hacking", MS Thesis, University of Manitoba, 2003
9. Kleen, L.: "Malicious Hackers: A Framework for Analysis and Case Study", Ph.D. Thesis, Air Force Institute of Technology, Ohio, 2001.
10. Swiler, L.P., Phillips, C., Ellis, D., Chakerian, S.: "Computer-Attack Graph Generation Tool", IEEE Symposium on Security and Privacy 2001.
11. Moore, A.P., Ellison, R.J., Linger, R.C.: "Attack Modeling for Information Security and Survivalility", Technical Note,CMU/SE1-2001-TN-001, March 2001.
12. Sheyner, O., Joshua Haines, J., Jha, S., Lippmann, R., Wing, J.M.: "Automated Generation and Analysis of Attack Graphs", IEEE Symposium on Security and Privacy, 2002.
13. McQuade S., Loper, D.K.: "A Qualitative Examination of the Hacker Subculture Through Content Analysis of Hacker Communication", American Society of Criminology, November, 2002.
14. Chandler, A.: "Changing definition of hackers in popular discourse", International Journal of Sociology and Law, 24(2), 229-252, 1996.
15. Jasanoff, S.: "A sociology of Hackers", The Sociological Review, 46(4), 757-780, 1998.
16. Rogers, M.: " A New Hacker's Taxonomy" University of Manitoba
17. Rowley, I.: "Managing In An Uncertain World: Risk Analysis And The Bottom Line", Systems Engineering Contribution to Increased Profitability, IEE Colloquium on , 31 Oct 1989
18. WINBUGS - http://www.mrc-bsu.cam.ac.uk/bugs
19. HUGIN DEMO - http://www.HUGIN.com/
20. Dantu, R., Loper, K., Kolan, P.: Survey of Behavior Profiles, University of North Texas Internal Document 2004.

Sensitivity Analysis of an Attack Containment Model

Ram Dantu[1], João W. Cangussu[2], and Janos Turi[3]

[1] Department of Computer Science, University of North Texas
rdantu@unt.edu
[2] Department of Computer Science, University of Texas at Dallas
[3] Department of Mathematical Sciences, University of Texas at Dallas
{cangussu, turi}@utdallas.edu

Abstract. A feedback control model has been previously proposed to regulate the number of connections at different levels of a network. This regulation is applied in the presence of a worm attack resulting in a slow down of the spreading worm allowing time to human reaction to properly eliminate the worm in the infected hosts. The feedback model constitutes of two queues, one for safe connections and another for suspected connections. The behavior of the proposed model is based on three input parameters to the model. These parameters are: (i) the portion of new connection requests to be sent to the suspect queue, (ii) the number of requests to be transferred from the suspect to the safe queue, and (iii) the time out value of the requests waiting in the suspect queue. The more we understand the effects of these parameters on the model, the better we can calibrate the model. Based on this necessity, a sensitivity analysis of the model is presented here. The analysis allows for the computation of the effects of changing parameters in the output of the model. In addition, the use of a sensitivity matrix permits the computations of not only changes in one parameter but also combined changes of these parameters. From the sensitivity analysis we have verified our assumption that the changes in the input parameters have no effect on the overall system stability. However, there will be a short period of instability before reaching a stable state.

1 Introduction

In the past active worms have taken hours if not days to spread effectively. This gives sufficient time for humans to recognize the threat and limit the potential damage. This is not the case anymore. Modern viruses spread very quickly. Damage caused by modern computer viruses (example - Code red, sapphire and Nimda) [1, 2] is greatly enhanced by the rate at which they spread. Most of these viruses have an exponential spreading pattern [2]. Future worms will exploit vulnerabilities in software systems that are not known prior to the attack. Neither the worm nor the vulnerabilities they exploit will be known before the attack and thus we cannot prevent the spread of these viruses by software patches

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 127–138, 2005.

or antiviral signatures [3]. A feedback control model has been proposed to slow down the spreading rate of a worm in order to allow proper human reaction. Experiments have shown that the proposed model is an effective alternative to achieve such goals.

The feedback model constitutes of two queues, one for safe requests and one for suspected requests. A set of parameters determines the behavior of the model based on: (i) the portion of new requests to be sent to the suspect queue (also called delay queue), (ii) the number of requests to be transferred from the suspect to the safe queue, and (iii) the time out value of the requests waiting in the suspect queue. The accuracy and behavior of the model depends on the selection of these parameters. In order to use the model effectively, it is important to determine how a small change in one parameter impacts the overall results. For a given network topology, a set of hosts, firewall rules, and IDS signatures, we need to configure a set of input parameters. However, the selection of these parameters is critical for optimum operation of the state model. For example, the selection decides on how fast we can converge to a given set point. Sensitivity analysis aims to ascertain how a given model depends on its input parameters [4]. This analysis will help us in determining how confident are we in our results and how much will the results change if our selected values are slightly wrong. We will give a completely different outcome or change the outcome only slightly. Also, parameters can be changed at any point in time and the study conducted here allows the determination of the side effects of these changes at different stages of a system. That is, what are the consequences if changes are performed before the infection, at early stages of the infection, or later once almost all the nodes have already been affected? In summary, the goal of this paper is to improve the understanding of the behavior of the model to allow a better selection of parameter values to consequently improve its performance under distinct circumstances.

This paper is organized as follow. The state model used for the feedback control of an attack is described in Section 2. The technique used to compute the sensitivity matrix to analyze the model as well as a detailed description of the results on the sensitivity analysis are presented in Section 3. Conclusions and remarks are drawn in Section 4.

2 Feedback Model

The architecture of the proposed approach [5] is composed of (centralized or distributed) controllers that are responsible for collection and monitoring of all the events in the network. These controller are knowledgeable about the network topology, firewall configurations, security policies, intrusion detections and individual events in the network elements. They are logical functions and can be deployed anywhere in the network. The controllers are represented by state models and are described next.

2.1 State Model

It is assumed that, as the number of requests increases, a portion of them will be sent to a delay queue to be served later. The remaining ones are sent to a safe queue to be served immediately. The overall structure of this queuing system is depicted in Figure 1. Parameters related to the size of the delay queue and the number of dropped (time-out) connections are used to control the total number of connections resulting in a slow down of a spreading worm. Sapphire worm spreading is taken as an example to show the applicability of our approach. The number of requests generated by the worm increase according to an s-shape format [2]. The goal of the example is to slow down the spreading velocity of a worm by controlling the total number of connections ($C(t)$) detected by a firewall. A model capturing the behavior of the system, i.e., how the number of total connections is changing is needed to achieve this goal. We assume here that as the number of request increases, a portion of them will be sent to a delay queue to be served later. Parameters related to the size of the delayed queue and the number of dropped connections are used to control the total number of connections resulting in a slow down of a spreading worm [5]. The rate of change of the number of requests ($\frac{\partial C}{\partial t}$) is proportional to the number of served requests ($-C(t)$), plus the number of connections transferred from the delayed queue ($\beta \times D(t)$, where β is the transfer rate and $D(t)$ is the number of delayed requests on the queue) and a portion of the new requests sent directly to the safe queue ($\alpha \times u(t)$, where α is the percentage of not delayed requests). This results in Eq. 1 below.

$$\dot{C}(t) = -C(t) + \beta D(t) + \alpha u(t) \tag{1}$$

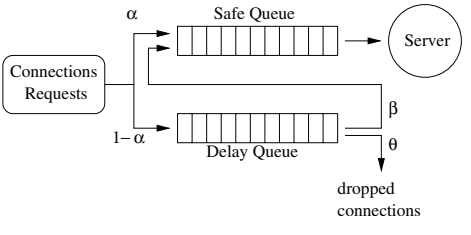


Fig. 1. Queue system which the state model is constructed upon

The rate of change in the size of the delayed queue ($\frac{\partial D}{\partial t}$) is proportional to a fraction of the new incoming requests sent to the queue ($(1-\alpha) \times u(t)$) minus the requests transferred to the safe queue ($-\beta \times D(t)$, where β is the transfer rate), minus the time-out requests($-\theta \times D(t)$, where θ is the time-out rate). This leads to Eq. 2.

$$\dot{D}(t) = -\beta D(t) - \theta D(t) + (1-\alpha)u(t) \tag{2}$$

It is assumed here that all requests in the safe queue, at one point in time, are served. Therefore, at each time a server is allocated to handle the requests at the safe queue, it serves all the new incoming requests plus all the ones transferred from the delayed queue at the previous time period. Combining Eqs. 1 and 2 in a state variable format leads to system in Eq. 3.

$$\begin{bmatrix} \dot{C}(t) \\ \dot{D}(t) \end{bmatrix} = \begin{bmatrix} -1 & \beta \\ 0 & -(\beta+\theta) \end{bmatrix} \begin{bmatrix} C(t) \\ D(t) \end{bmatrix} + \begin{bmatrix} \alpha \\ 1-\alpha \end{bmatrix} u(t) \tag{3}$$

The output part ($Y(t)$) of the system from Eq. 3 can be designed as needed, as long it is a function of the state vector $X(t) = [C(t)\ D(t)]^T$. Let us assume we are interested in the size of both queues $C(t)$ and $D(t)$ in addition to the rate of change in the safe queue ($\dot{C}(t)$). The size of the queues are the values of the state vector. The first derivative of $C(t)$ is represented by Eq. 1. The three desired outputs specified above and the respective equations lead to the output part ($Y(t)$) of a state model as presented by Eq. 4.

$$\begin{bmatrix} C(t) \\ \dot{C}(t) \\ D(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & \beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C(t) \\ D(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha \\ 0 \end{bmatrix} u(t) \tag{4}$$

Since Eqs. 3 and 4 represent our model, we refer to them, hereafter as the `FBAC Model` (Feedback Attack Containment). Now let us assume an input $u(t)$ as in Eq. 5 presenting an S-shape behavior.

$$\dot{u}(t) = Ku(t)(1 - \frac{u(t)}{s}) \tag{5}$$

where s represents the saturation point of the S-curve and K is the growth parameter. Eq. 6 represents the solution of Eq. 5 where $u(0)$ represents the initial condition. It has an S-shape format mimicking the behavior of a spreading worm.

$$u(t) = \frac{s}{1 + \left(\frac{s}{u(0)} - 1\right)\ e^{-Kt}} \tag{6}$$

Now consider a system with the following parameter values: $\alpha = 0.3$ (only 30% of the requests are served immediately while the remaining 70% go to the delay queue), $\beta = 0.15$ (15% of the requests in the delay queue are transferred to the safe queue) and $\theta = 0.2$ (20% of the requests on the delay queue are timed-out at each time stamp). Also, consider the input $u(t)$ as specified in Eq. 6 with parameters $K = 0.08$ and $s = 30$. The output of the model for the three specified values are shown in Figure 2. As can be seen in the figure, the size of both queues follow the S-shaped behavior of the input function. However, the safe queue saturates with eighteen requests while the delayed queue saturates with sixty requests per time unit. The velocity $\dot{C}(t)$ (rate of change of requests) increases initially until a inflection point is reached and it starts to decrease until the saturation level in the size of the save queue is achieved and the velocity goes to zero. The effects of the values of the parameters and consequences of changes in these values are analyzed next in Section 3.

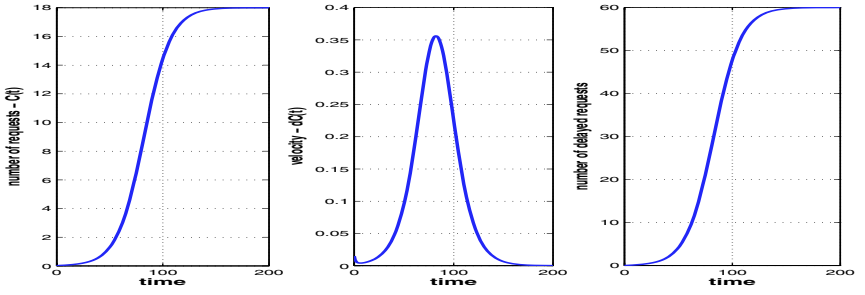


Fig. 2. Output behavior of the `FBAC Model` for $\alpha = 0.3$, $\beta = 0.15$, and $\theta = 0.2$

3 Sensitivity Analysis

The understanding of how the state variables change according to changes in the input parameters of the model can provide information to help on the determination of the values of these parameters. In other words, we are interested in analyzing how sensitive is the model to changes in its parameters. Three parameters are of interest in the state model from Eqn. 3: β, θ, and α. Consider now a parameter vector $G = [\beta\ \theta\ \alpha]^T$. The goal is to compute $\dfrac{\partial X(t)}{\partial G}$, where $X = [C\ D]^T$ is the state vector of interest. A sensitivity matrix can be computed to achieve this goal [4].

3.1 Computation of the Sensitivity matrix

One alternative for the computation of the sensitivity matrix would be the use of the Kronecker product [6] that allows the differentiation of vectors with respect to a matrix. However, the fact that the model in Eq. 3 presents a nice upper triangular format combined to the fact that we have a vector and not a parameter matrix allows an easier way to achieve the same goals. The differential equations for the elements of the state vector X are defined by Eqs. 7 and 8 below.

$$\dot{C}(t) = -C(t) \ + \ \beta D(t) \ + \ \alpha u(t) \tag{7}$$

$$\dot{D}(t) = -(\beta + \theta)D(t) \ + \ (1 - \alpha)u(t) \tag{8}$$

Solving Eqs. 7 and 8 result in Eqs. 9 and 10 below.

$$C(t) = e^{-t} c0 + \int_0^{\tau} e^{-t+\tau} \left(\beta\, D(\tau) + \alpha\, u(\tau)\right)\ d\tau \tag{9}$$

$$D(t) = e^{t(-\beta-\theta)} d0 + \int_0^{\tau} e^{(-\beta-\theta)(t-\tau)} \left(1 - \alpha\right) u(\tau)\ d\tau \tag{10}$$

A Jacobian matrix [7] for the functions $C(t)$ and $D(t)$ with respect to the parameters β, θ, and α can be compute as in Eqn. 11. This matrix represents the sensitivity matrix $SM(t)$ for the model in Eqn. 3.

$$SM(t) = \begin{bmatrix} \frac{\partial C(t)}{\partial \beta} & \frac{\partial C(t)}{\partial \theta} & \frac{\partial C(t)}{\partial \alpha} \\ \frac{\partial D(t)}{\partial \beta} & \frac{\partial D(t)}{\partial \theta} & \frac{\partial D(t)}{\partial \alpha} \end{bmatrix} \tag{11}$$

where the elements of the matrix above are computed according to Eqs. 12,..., 17.

$$\frac{\partial D(t)}{\partial \beta} = -te^{t(-\beta-\theta)}\, d0 - \int_0^{\tau} (t-\tau)\, e^{(-\beta-\theta)(t-\tau)}\, (1-\alpha)\, u\; d\tau \tag{12}$$

$$\frac{\partial D(t)}{\partial \theta} = -te^{t(-\beta-\theta)}\, d0 - \int_0^{\tau} (t-\tau)\, e^{(-\beta-\theta)(t-\tau)}\, (1-\alpha)\, u\; d\tau \tag{13}$$

$$\frac{\partial D(t)}{\partial \alpha} = -\int_0^{\tau} e^{(-\beta-\theta)(t-\tau)} u\; d\tau \tag{14}$$

$$\frac{\partial C(t)}{\partial \theta} = \int_0^{\tau} e^{-t+\tau}\beta \left(-\tau\, e^{\tau\,(-\beta-\theta)}\, d0 - \int_0^{z} (\tau - z)\, e^{(-\beta-\theta)(\tau-z)}\, (1-\alpha)\, u\; dz \right) d\tau \tag{15}$$

$$\frac{\partial C(t)}{\partial \beta} = \int_0^{\tau} e^{-t+\tau} \left(e^{\tau\,(-\beta-\theta)}\, d0 + \int_0^{z} e^{(-\beta-\theta)(\tau-z)}\, (1-\alpha)\, u\; dz + \beta \times \left(-\tau e^{\tau(-\beta-\theta)}\, d0 - \int_0^{z} (\tau - z)\, e^{(-\beta-\theta)(\tau-z)}\, (1-\alpha)\, u\; dz \right) \right) d\tau \tag{16}$$

$$\frac{\partial C(t)}{\partial \alpha} = \int_0^{\tau} e^{-t+\tau}\beta \int_0^{z} e^{-(\beta+\theta)*(\tau-z)} * u\; dz\; d\tau \tag{17}$$

Matrix $SM(t)$ can now be used to compute variations in the state variables in response to perturbations in specific parameters or combinations thereof. Eq. 18 is used to compute the variations.

$$\begin{bmatrix} \Delta C(t) \\ \Delta D(t) \end{bmatrix} \;=\; SM(t) \;\times\; \Delta_m \qquad where\ \Delta_m = \begin{bmatrix} \Delta\beta \\ \Delta\theta \\ \Delta\alpha \end{bmatrix} \tag{18}$$

To obtain the variations for individual or combined changes in the values of parameters, we set the Δ_m matrix. For example, to analyze the effects of a 5% change in β and a -10% change in θ, we set Δ_m to $[0.05\beta \;\; -0.1\theta \;\; 0]^T$, which when substituted in Eq. 18 gives us $[\Delta C(t)\;\; \Delta D(t)]^T$. $\Delta C(t') < 0$ at time $t = t'$ implies a decrease in the number of connections in the safe queue. Logically, $\Delta C(t') > 0$ implies an increase in the corresponding queue. A positive slope represents an increase in the queue size while a negative one represents a decrease. That is, though the overall size of the queue maybe smaller than the queue with no changes, a positive or negative slope represents how the modified system is changing according to time. $\Delta D(t')$ presents a similar behavior with respect to the size of the delayed queue.

3.2 Results of the Sensitivity Analysis

We now use the sensitivity matrix to analyze the sensitivity of the state variables of `FBAC Model` to variations in its parameters. Unless stated otherwise, we assume that `FBAC Model` represents a system with parameter values: $\beta = 0.2$, $\theta = 0.21$, and $\alpha = 0.3$. These values are arbitrary and do not affect the results of the analysis reported here. We also assume here that the input $u(t)$, representing the number of new requests follows the behavior of Eq. 6 fort $K = 0.08$ and $s = 20$.

Assuming the input above and the parameters values specified before, Figure 3 shows the behavior of the state variables C and D with respect to time for a 10% increase in the value of β ($\Delta_m = [0.1 \times \beta\ 0\ 0]^T$). As can be seen in the figure, the increase in the β parameter has opposite effects with respect to the safe and the delayed queue. β determines how many connections are removed from the delayed queue and send to the safe queue. Therefore, increasing β will naturally decrease the size of the last while increasing the size of the former. Also, $D(t)$ is more sensitive to this change than $C(t)$ as the overall absolute value of Δ_D is larger than Δ_C. The fact that α is less than 0.5 means that more requests are being sent to the suspect queue than to the safe queue. Consequently, the first one grows faster and larger then the second and more elements are transferred when β is increased. When no changes for the parameters are observed, the safe queue saturates around 13 connections per time unit and the delayed queue saturates at 34 connections per time unit. The size of the queues at saturation point shows that a 2.7% change in the safe queue is observed when α is increased by 10%. The delayed queue is more sensitive to the same change presenting a 5% increase.

Changes in α present a similar behavior as described for β as it increases the size of the safe queue and decreases the size of the delayed queue. However, the fact that α is larger than β magnifies the sensitivity. Consequently, a combined change of α and β also has a similar behavior. However, the sensitivity in this case is a result of the sum of the isolated changes for α and β. That is, $\Delta_{C_{\Delta_\alpha=10\%\Delta_\beta=10\%}} = \Delta_{C_{\Delta_\alpha=10\%}} + \Delta_{C_{\Delta_\beta=10\%}}$. The behavior of Δ_D being similar.

When the value of θ is increased both queues present a decay in their size. As θ represents the number of time out requests at the delayed queue, an increase in its value will naturally decrease the size of the queue as more requests are

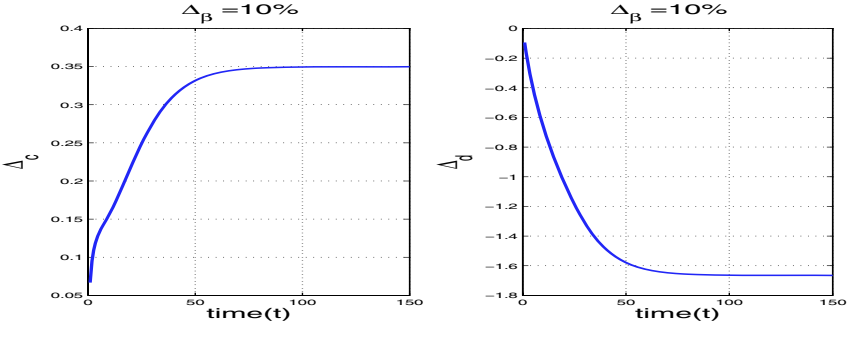


Fig. 3. Sensitivity of $C(t)$ and $D(t)$ for a 10% increase in the value of β

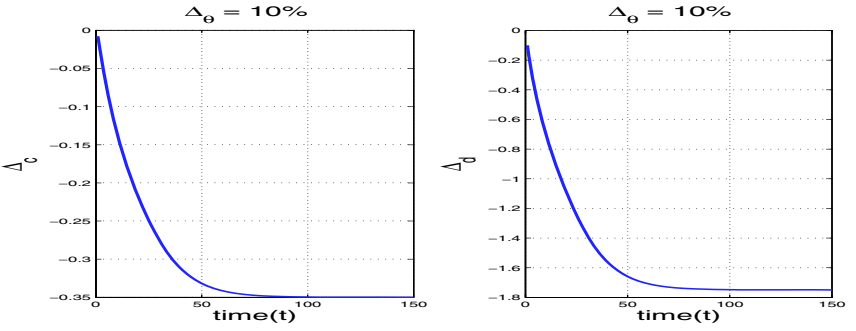


Fig. 4. Sensitivity of $C(t)$ and $D(t)$ for a 10% increase in the value of θ

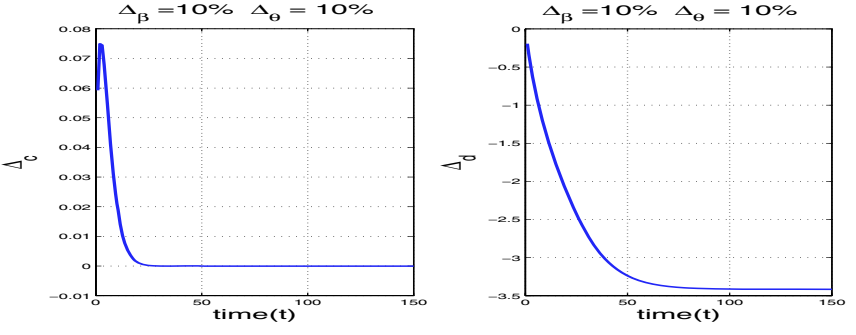


Fig. 5. Sensitivity of $C(t)$ and $D(t)$ for a 10% increase in the value of β and θ

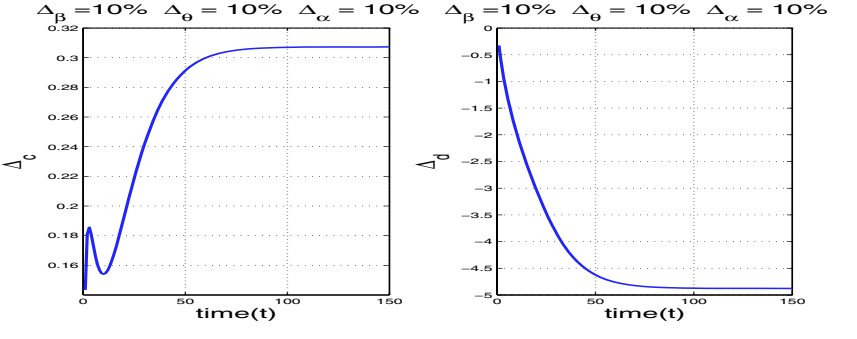


Fig. 6. Sensitivity of $C(t)$ and $D(t)$ for a 10% increase in the value of α, β, and θ

timing-out. Consequently, a smaller number of requests are transferred from the delayed to the safe queue which also experiences a decrease. Such behavior is shown in Figure 4. Combined changes for θ and α present a behavior similar to Figure 4. Though more requests are sent to the safe queue, more are dropped from the delayed queue and less are transferred from the last to the former.

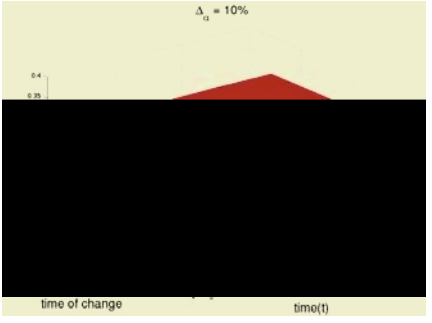


Fig. 7. Sensitivity of the safe queue with respect to a 10% change in the α parameter at different time stamps

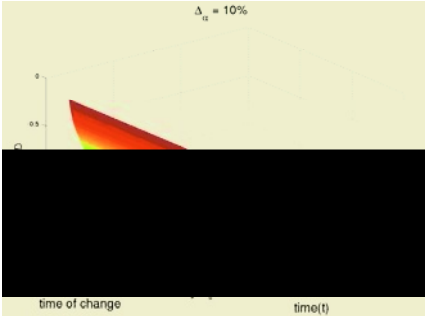


Fig. 8. Sensitivity of the delayed queue with respect to a 10% change in the α parameter at different time stamps

A combined change of θ and β produces initially small increase in the size off the safe queue. Though more requests are being transferred, initially, the dropped (due to time out represented by θ) requests are more than the transferred ones since θ is slightly larger than β. After some time, the changes average out and the sensitivity goes to zero as noticed from Figure 5.

A different behavior is observed when all the three parameters are subjected to 10% changes. In this case, the safe queue suffers an overall increase with some oscillation before saturating as can be seen in Figure 6. The oscillation is due to the fact that the overall number of requests (represented by $u(t)$) grows faster at the beginning and since α has also being increased, it compensates the changes in θ and β. After sometime, the changes in θ and β overcomes the growth of $u(t)$ and the change in α and the systems stabilizes.

All the results above are for changes of the parameters at time $t = 0$, i.e., at the beginning of the spreading of a worm. Next, we analyze not only the changes in the parameters but also the effect of delaying these changes ($t > 0$). The following plots represent the time, the time of the change of the parameter(s), and the sensitivity. As depicted in Figures 7 and 8, the time of change of α has no effect on the behavior of both queues. Though more requests are sent directly to the safe queue, less are being transferred from the delayed to the safe at any point in time. This explains the behavior of Figures 7 and 8.

Changes in β represent the behavior shown in Figures 9 and 10. The size of the delayed queue builds up with time. The later the change in β, the larger the size of the delayed queue at that time. Therefore, more requests are transferred from the delayed to the safe queue and an immediate increase is observed in the last. However, the size of the delayed queue starts to decrease and so does the size of the safe queue. This is observed until the increase in the number of requests (input $u(t)$) compensates this behavior and the system starts to present the same behavior as of changing β at time $t = 0$. This oscillation can be seen as one step toward the equilibrium point, i.e., the saturation level of the queues and the input. The behavior of the delayed queue for a 10% change in θ is similar

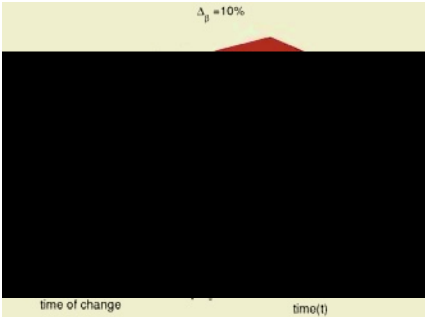


Fig. 9. Sensitivity of the safe queue with respect to a 10% change in the β parameter at different time stamps

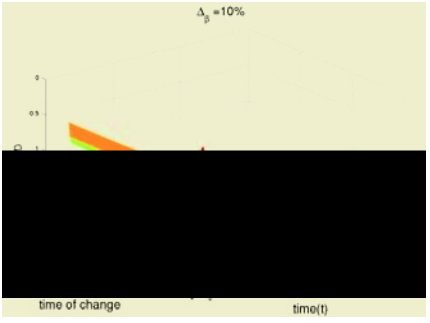


Fig. 10. Sensitivity of the delayed queue with respect to a 10% change in the β parameter at different time stamps

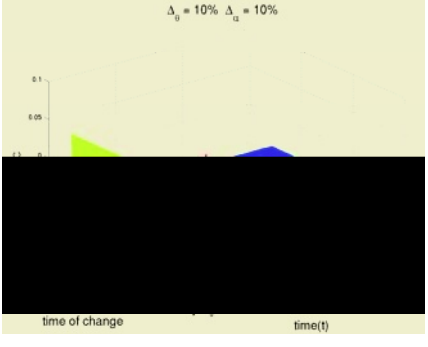


Fig. 11. Sensitivity of the safe queue with respect to a 10% change in α and θ at different time stamps

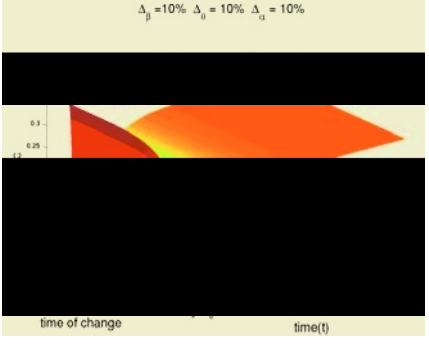


Fig. 12. Sensitivity of $C(t)$ and $D(t)$ for a 10% increase in the value of α, β and θ

to the one in Figure 10 (as it is any change that involves either β, θ, or both). However, the safe queue presents a continuous decrease up to the saturation point as less requests are now transferred. As with the changes in the parameters at time $t = 0$, the combined change of α and β presents a similar behavior as the change of β (Figures 9 and 10). However, as expected, the magnitude of the changes are amplified when both parameters suffer a simultaneous perturbation. The sensitivity of `FBAC Model` with respect to a combined change of 10% in parameters α and θ at increasing time stamps is depicted in Figure 11. In this case, at the beginning of the change when the input values are smaller, the change in α overcomes the change in θ. That is, initially the additional requests sent directly to the safe queue (a consequence of increasing α) are larger than the additional time-out requests (a consequence of increasing θ). However, as the delayed queue saturates, the time-out requests become larger than the extra

requests send to the safe queue which then presents an overall decrease as seen in Figure 11. The behavior for combined changes of β and θ has a similar behavior as described above. However, the magnitude of the sensitivity and the reasons for the behavior are different. In the first case the maximum increase of the safe queue is around 0.3% which is amplified by more than six times when the changes are for β and θ. Until the size of the delayed queue starts to decrease due to the change in θ, more requests are transferred to the safe queue which presents an initial increase in its size. After some time period, the changes compensate each other and the sensitivity goes to zero.

The plot in Figure 12 shows the sensitivity of the safe queue of `FBAC Model` when all the three parameters are increased by the same factor. A prompt increase in the size of the queue is observed at the beginning. The later the changes of the parameters, the larger the increase. After that, the queue size suffers a decrease, denoted by the initial negative slope in Figure 12, followed by an increase until a stable point is achieved. This oscillation is due to the same arguments presented earlier when discussing Figure 6.

4 Concluding Remarks

Automatic containment of the propagation of computer worms is a difficult task since we do not know the behavior of future worms. But we do know from past experience that worms propagate by attempting large number of connections to other hosts. If successful, the infected remote machine will try to infect the other machines and the process continues. We are researching a containment model based on the feedback control theory. In order to validate our model, we have carried out sensitivity analysis on the input parameters (α, β, and θ). We observed the change in size of these two queues with res pect to time. In addition we observed another dimension (time of change) with respect to the state of the worm propagation.

We have increased the values of α, β, and θ in increments of 10%. In all the three cases, the system became stable before 50 time units and this is well before the beginning of the attack period. Hence from our analysis we observe that there is no impact of changing the individual input parameters on the system accuracy or the stability in containing the attacks. But when we apply the change at a later time, say during 100 time units, a sudden increase in β will result in sudden increase in C and we consider this as a merging or clumping effect. This is due to sudden transfer of connections from the delay queue to safe queue. Next we have increased two out three parameters simultaneously in increments of 10%. In all the cases we observed a small oscillation where the size of safe queue increases/decreases and suddenly reverses the trend. For example, simultaneous increase of θ and α will result in a sudden increase of safe queue but trend reverses as θ (drop of connections in the delay queue) increases. This is due to the fact that initially large number of connections are enqueued to the safe queue in addition to transfer of connections from delay to the safe queue. However, as the connections are timed out and dropped at the delay queue, less

and less number of connections are transferred to the safe queue. Finally we have increased all three parameters simultaneously and we found similar oscillation effects in the beginning. This is expected behavior when compared with the two-parameter-change.

In all the experiments described above, the size of the oscillation depends on the instant of change with respect to the S-shaped input. The later the changes, the larger the oscillation but the system comes back to a stable state within short period of time and hence there is no impact of this change on the outcome of the model. Hence from the sensitivity analysis we have verified our assumption that the changes in the input parameters have no effect on the overall system stability. But there will be short period of instability in the beginning and after that the system reaches a stable state and stays there for rest of the time. We have analyzed the impact due to small changes in the input parameters. The overall system behavior may be different for large variations of the input parameters. The analysis of these changes is subject of future work.

References

1. S. Staniford, V. Paxson, and N. Weaver, "How to own internet in your spare time," in *Proceedings of the USENIX Security Symposium*, pp. 149–167, August 2002.
2. D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "The spread of the sapphire worm." http://cs.berkely.edu/ nweaver/sapphire.
3. M. M. Williamson, "Throttling viruses: Restricting propagation to defeat malicious mobile code," in *18th Annual Computer Security Applications Conference*, pp. 61–68, December 2002.
4. *Sensitivity Analysis*. John Wiley & Sons, 2000.
5. R. Dantu, J. W. Cangussu, and A. Yelimeli, "Dynamic control of worm propagation," in *Proceedings. ITCC 2004. International Conference on Information Technology*, vol. 1, pp. 419–423, April 5-7 2004.
6. J. W. Brewer, "Matrix calculus and the sensitivity analysis of linear dynamic systems," *IEEE Transactions on Automatic Control*, vol. 23, pp. 748–751, August 1978.
7. R. A. DeCarlo, *Linear systems : A state variable approach with numerical implementation*. Upper Saddle River, New Jersey: Prentice-Hall, 1989.

Toward a Target-Specific Method of Threat Assessment

Yael Shahar

Senior Researcher, International Policy Institute for Counter-Terrorism,
Interdisciplinary Center Herzliya
webmaster@ict.org.il

Abstract. The threat assessment model used here has been used by researchers at ICT to estimate the "attractiveness" of specific facilities to terrorist organizations. The model uses on-site evaluations of vulnerabilities to build a portfolio of possible attack scenarios. These scenarios are then analyzed using known or estimated sensitivities and target-assessment criteria for the different organizations. The result is a means of rating the different scenarios according to their attractiveness to different types of organization. This enables decision-makers to concentrate resources on most probably scenarios, rather than on worst-case scenarios. The model has provided credible results for actual venues.

1 Introduction

The venue-specific risk assessment model discussed here has been used by ICT researchers for the past three years in the framework of security consulting contracts. This model estimates the "attractiveness" of specific facilities to terrorists. While the model's methodology is quite general, the particular target characteristics generate very specific risk vs. impact numbers for that target. We have used this model for evaluating threat / impact risks for tourist and transportation facilities with credible results. The end goal is to turn this model into a computational apparatus that can be used by those responsible for installation security, who may not have extensive knowledge of the terrorist groups most likely to target their facility. This will require extensive data-mining capabilities to allow the model to "self-adapt" and "learn" from open source material.

In the absence of empirical data on threats, most planners and pundits fall back on "worst case" scenarios. However, this has often meant the over-allocation of resources to some of the least likely scenarios. Preparing for the "low risk / high consequence" attack is not the only (or even the best) response. Scarce resources can be better exploited if a way is found to classify threats according to likelihood, and not just according to the severity of consequences.

Naturally, our assessment of which scenarios are most attractive must be made from the point of view of the terrorist. We need to be on the same wavelength as our enemies; we need to know our enemies from the inside out, and this requires that any preparedness program be built on sound intelligence. The stages in preparing an operational counter-terrorism plan would be, first, to identify those groups or individuals who are both *motivated and capable* of attacking the facility under consideration; secondly, to concentrate intelligence gathering resources on these

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 139 – 152, 2005.

groups in order discern possible methods and targets; and thirdly, to conduct tactical operations against these groups. At the same time, our understanding of who the enemy is and what he is likely to do should inform our defensive preparedness as well, allowing us to increase security around likely targets.

2 Methodology

The probability of a terror attack on a particular target is dependant not only on the characteristics of the target—its symbolic or strategic value and its vulnerabilities—but also on the ambition, capabilities, and sensitivities of the relevant terrorist organizations.

I would like to present an overview of a statistical method for evaluating the threat presented by different types of terrorist groups to particular venues. The method builds on input regarding the known characteristics of the terrorist groups active in the region in question, combined with the characteristics of each potential *modus operandi*. The goal is not so much to give a precise prediction of who, how, and where; but rather to provide a better basis for deciding how best to allocate finite counter-terrorism resources.

The stages used in the proposed model are:

- *Organization-specific factors* – Develop a means of determining which organizations present the greatest threat.
- *Venue-specific factors* – Categorize scenarios according to the target selection strategies of these particular organizations.
- *Numerical synthesis* – Combine these two stages into a numerical analysis of what type of attack would appeal to which organization.

To evaluate the likelihood that any particular type of attack will be carried out by a particular terrorist organization, we identify factors relating to the difficulty of carrying out an attack, along with the desirability of the outcome of the attack from the point of view of the terrorist organization. For each terrorist organization type, we include factors to weigh the terrorists' sensitivity to these attack-specific factors. The result is a score that indicates the net "attractiveness" of a particular type of attack to a particular type of organization. For example, some organizations may be more deterred by physical difficulty or by attacks requiring years of planning, while others would be less so. Some may see achieving maximum casualties as part of their goals, while others may be unwilling to risk causing high casualties.

This method is useful for getting an idea of the types of attacks that each organization might be inclined to carry out on the selected venue, based on our understanding of the organization's target-selection methods. Thus, the method builds on considerable research into the characteristics of each organization.

The same care must be taken to examine the vulnerabilities of the venue in question, from the point of view of the potential terrorist—seeking weaknesses that might be exploited in the various scenarios under consideration.

Naturally, the resulting numbers are only as good as the information that goes into the model. In effect, the model should be viewed as merely a template for better

organizing our knowledge; without that knowledge, the template is empty of all content. However, when based on reliable information, the model is a very useful tool for data visualization and risk assessment.

3 Organization-Specific Indicators

For the purposes of illustration, we will base the following example on a tourist venue in the United States. An analysis of organizations in the ICT database and their characteristics leads us to believe that the greatest threat to tourist sites in the United States is likely to be posed by groups or cells having the following characteristics:

- *Motivated by religion or quasi-religious ideology.* This entails a preference for mass-casualty attacks, as well as the potential to carry out suicide attacks.
- *Supported by a state or sub-state entity.* The stronger such support, the greater the resources at the disposal of the group, and the more ambitious a potential attack is likely to be.
- *Has carried out attacks outside its local sphere of influence.* Often, this is a function of how much assistance the group receives from a state or sub-state sponsor. Those foreign-based organizations that have demonstrated capability to act far from their home ground are naturally to be considered a greater threat.
- *Has carried out attacks against similar targets.* For example, an organization that has singled out tourism targets for attack is one whose goals generally include inflicting economic damage on the target country. This is true of some domestic groups, such as ETA, as well as of international terrorist groups. However, organizations that single out tourist sites popular with international travelers generally have a more globe-spanning goal. International tourism-related targets are often chosen because they represent the antithesis of the terrorists' own worldview: they stand for openness, diversity, and globalization.
- *Has carried out attacks against targets or interests belonging to the country under consideration.* For example, a history of attacks on American targets indicates not only a pre-existing enmity toward the United States, but more importantly, a readiness to transform this enmity into action.
- *Loose network affiliated with other like-minded cells or groups around the world.* Such international connections are particularly useful in the planning stages of an attack, and facilitate the escape of those directly involved in supervising and perpetrating the attack. Groups posing the greatest danger are those with a broad-based support system around the world. Such a support system often consists of fundraising and political activists living outside the group's normal sphere of operations.
- *Is able to blend in with immigrant groups in the target country.* Because of the need for extensive planning and pre-operational intelligence, the members of a potential attack cell would be required to spend long periods under cover inside the target country. Connections with a particular ethnic group within that country could greatly facilitate the existence of "sleeper" cells.

- *Shows past history of hatred of the target country/countries.* Most terrorist groups are fairly vocal about their agenda, in order to recruit like-minded individuals to their camp. Although not all vocally anti-Western groups pose a threat to national security, it is probable that those that do pose a threat will make no secret of their intentions.
- *Has been, or expects to be, targeted by the international counter-terror campaign.* In general, potentially the most dangerous groups are likely to be those directly affected by Western policies. In the past, this meant that the groups to worry about would be those adversely affected in some way by the target country's foreign policy. However, in light of the U.S.-led campaigns in Afghanistan and in Iraq, we can expect the circle of those with an axe to grind against the U.S. and its allies to have grown considerably wider. Because of the inter-relatedness of the threat, the potential risk is spread out among dozens of semi-autonomous organizations in a number of countries. Many of these countries are not themselves in any way directly affected by the U.S. campaign.

These indicators of potential threat will be given a relative weight according to the proportional contribution to a given organization's potential to attack similar targets in the target country. Table 1 shows one example of how these indicators might be chosen in the case of a tourism-related venue.

Table 1. Organization-specific indications of potential threat

Indicator	**Relative Weight**
Structure and reach	
Supported by a state or sub-state entity.	*20*
Part of international network of groups or cells.	*15*
Ability to blend in with immigrant groups in the United States.	*10*
	Max: 45
Attack History	
Past history of attacks outside its own sphere of influence.	*15*
Past history of attacks against similar targets	*10*
Past history of attacks against American interests	*5*
	Max: 30
Motivation	
Motivated by religion or quasi-religious ideology.	*15*
Has threatened to attack American interests.	*5*
Has been, or expects to be, targeted by international counter-terror campaign.	*5*
	Max: 25

The values of the variable chosen here are not necessarily the values that would be used for an actual threat assessment. Since we have not specified a real venue, we are using these values as an example only. The exact values of the variables would be determined by the researchers according to the nature of the venue under consideration. The total of all the variables should equal to 100.

3.1 Example of Organization Input

Table 2. Gama'ah al-Islamiyya: Relative risk from analysis of indicators

Structure and Reach	
	Supported by a state or a sub-state entity
0	The Egyptian Government believes that Iran, Bin Ladin, and Afghan militant groups have all provided financial support to the Gama`ah al-Islamiyya. However, the group does not benefit from significant state-sponsorship.
	Part of an international network of like-minded groups and cells
5	The Gama`ah al-Islamiyya operates mainly in southern Egypt, and in urban centers. However, the group has been indirectly involved in international terror, through 'Arab-Afghans' (Egyptian citizens who trained in the Mujahideen camps after the end of the war) who operate as individuals within autonomous Islamic terrorist cells abroad. Such cells exist in Sudan, the United Kingdom, Afghanistan, Austria, and Yemen.
	Ability to blend in with immigrant groups in the United States
5	Other than a small group of radicals, the Islamic Group has only a limited presence in the United States. However, the group has been known to run fundraising operations in the U.S. and in Canada.
Attack History	
	International terrorist activity
	From 1993 until the cease-fire, al-Gama`ah carried out attacks on tourists in Egypt, most notably the attack at Luxor that killed 58 foreign tourists.
15	The Egyptian organizations, predominantly Al-Gama`ah al-Islamiyya, were indirectly involved in international terror, through 'Arab-Afghans' operating as individuals within autonomous Islamic terrorist cells abroad. Attacks carried out, or claimed by these cells, include a suicide bomb attack against the police station in Rijeka, Croatia (October 1995) and the attempted assassination of Hosni Mubarak during his visit to Ethiopia (June 1995).
	The Gama`ah has never specifically attacked U.S. citizens or facilities, and has not conducted an attack inside Egypt or abroad since August 1998.
10	Attacks against tourism-related interests
	Until the ceasefire, GIA was known for its attacks on tourism in Egypt. However, the organization has not attacked tourist targets outside of Egypt.
5	Attacks against American Targets
	American tourists were attacked in Egypt, along with nationals of other countries.
Motivation	
	Motivated by religious ideology
15	The Islamic Group was greatly influenced by the militant ideology of Sayyid Qutb. The blind Egyptian cleric, Sheikh Omar Abdel Rahman, the supreme spiritual authority of Al-Gama`ah al-Islamiyya and the Egyptian 'Jihad,' issued a *fatwa* sanctioning terrorist activity.
	Has expressed willingness to attack Western interests
5	The group has sided with Osama bin Ladin in the past, including signing his 1998 fatwa calling for attacks on U.S. civilians. While the Gama`ah's traditional goal was the overthrow of the Egyptian Government, Taha Musa's faction has expressed an interest in attacking U.S. and Israeli interests.
	Targeted by international anti-terrorism campaign
5	The U.S.-led campaign may have cut off some of the group's funding, and driven Musa Taha, the leader of the radical faction, out of Afghanistan. The US treasury has added the GI to the list of groups targeted by economic sanctions.
Overall relative risk: 65	

A similar analysis would be carried out for each organization deemed to present a potential threat. Those having the highest overall risk score would be those used in the analysis that follows. The organization above would most likely not be one of those included in the analysis, due to its relatively low risk score.

3.2 Venue-Specific Factors Affecting Target Selection

In order to evaluate the net "attractiveness" of a particular type of attack to a particular type of organization, each type of attack receives a combined Difficulty score, as well as three separate Impact scores, as shown below in Table 3:

Table 3. Attack-type parameters

Parameters	Relative Weight
Difficulty, based on:	
Planning time required	33
Resources required	33
Physical difficulty	33
Impact, based on	
Physical/economic damage	33
Casualties	33
Symbolic value (includes psychological impact)	33

3.3 Example Scenarios

The following scenarios illustrate how these factors are applied to particular scenarios. Since we are not dealing with a particular venue in these examples, these samples are very general in nature. An actual threat assessment would require that the specific vulnerabilities of the venue be factored into the scenario rating.

Remotely Detonated Bomb

This is the simplest type of attack and requires only moderate investment. However, in order for this type of attack to be successful, the organization must carry out at least minimal pre-operational intelligence to determine the best time and location for bomb placement. For example, if the intention is to inflict casualties, the bomb will need to be set in an area with heavy traffic and time to go off during working hours. On the other hand, if the aim is to gain media attention without causing casualties, then the bomb might be placed in the same area at midnight.

Bomb in public area

Leave bomb package in lobby or in garbage container. Zone: A / B	A bomb placed in a travel bag or plastic bag is left in a public area or in a garbage container. The bomb can then be activated by timer or remote control mechanism.	Planning time: 20 Resources: 10 Difficulty: 10

	Result: Possible large damage to property Disruption of operations for short while Possible moderate number of casualties	Physical/Economic Damage: 25 Casualties: 20 Symbolic: 25

Suicide Attacks

The human bomb is extremely motivated either by ideological and/or religious impulses. His/Her main aim is mass killing and destruction, without regard for his or her own life or the lives of others.

Spiritually, these people are ready to die, and as a result they are calm and determined. Until they get to the specific spot where they will operate the explosives, only a highly trained eye would notice anything suspicious.

Such a human bomb can approach any public facility—particularly a crowded place such as a bar or restaurant—and operate a charge of a few dozen kilograms. Such charges are usually stuffed with nails and shrapnel, in order to maximize casualties and loss of limbs.

The effects of this type of bombing are generally limited by the amount of explosives that the attacker can carry into the target zone. However, unlike a static charge, the attack can choose both the time and the place of detonation, thus maximizing damage and casualties.

Suicide bomb

Suicide bomber	A suicide bomber conceals the explosives on his body or in a bag and manually activates the device.	Planning time: 20 Resources: 15 Difficulty: 15
	Result: Possible large number of casualties; Closure of facility for several days / weeks	Physical/Economic Damage: 25 Casualties: 25 Symbolic: 25

Carbomb Attack on Main Entrance or External Walls

An additional and potentially even more devastating kind of attack would be to break into the lobby of a public building with a carbomb. We've seen the effects of such an attack at the American embassy in Beirut, in some Israeli official facilities in Lebanon and the bombings of the American embassies in Kenya and Tanzania, not to mention the more recent attacks on the Marriot Hotel in Jakarta and the Paradise Hotel in Mombassa.

Such a vehicle can carry a few hundred kilograms of explosives. Operating such a charge at the center of the building will cause many casualties and large-scale damage to the structure.

Terrorists are well aware of the potential impact of this type of attack; it is exactly the sort of outcome they aim for. The fall of the World Trade Center towers in NYC illustrates the possible result far better any other description.

Suicide Car bomb attack on lobby / external wall

Explosives-laden vehicle attack on public building. Zone: B	Drive a car or truck filled with explosive into the entrance or lobby. (This was the method used in the attack on the Paradise Hotel in Mombassa in November 2002.)	Planning time: 20 Resources: 20 Difficulty: 15
	Result: Possible large number of casualties; Closure of building for several weeks / months	Physical/Economic Damage: 30 Casualties: 28 Symbolic: 25

More "professional" terrorists could carry out a carbomb attack by crashing the vehicle into the lobby through the main entrance, leaping out of the vehicle (an action requiring no more than 30 seconds) and escaping through a secondary entrance while everyone in the vicinity is in shock. The explosives can be detonated a few minutes later by timer or remote control.

Carbomb in Underground Parking Lot

One of the most vulnerable and crucial weak points of many public building is the underground parking lot. In many cases, these parking lots are open to the public, and there is no control whatsoever over who parks there. Some of the cars may be parked in the lot on a permanent basis, either daily or periodically. It would be extremely easy for hostile elements to park a vehicle packed with a few hundred kilograms of explosives, which could be timed to detonate after the perpetrators have left the country.

Car bomb – underground car park

Place car or small truck loaded with high explosives or a gas container in car park – detonate by timer or remote control.	One or more cars, vans, or small trucks are bought, hired, or stolen. The vehicles are packed with explosives or gas tanks and parked in the underground parking. The vehicles are placed in such a way as to cause maximum structural damage. The vehicles are detonated using remote control devices or timers.	Planning time: 25 Resources: 25 Difficulty: 28
	Result: Possible large number of casualties; Possible major/total structural damage; Disruption/cessation of operations	Physical/Economic Damage: 33 Casualties: 33 Symbolic: 33

In quite a few cases, the layout of underground parking lots allows a vehicle to be parked near the main load-bearing pillars of the building. A very large charge

detonated in such a place could cause major damage, many casualties, and at the very least, disruption of operations for a long time. Structural damage could be serious enough to lead to the complete closing of the facility, if not to even worse consequences.

A regular sedan type car is capable of carrying a few hundreds kilo of explosives. Such a charge can be operated by timer or remote control mechanism. A blast of a few hundred Kilos will shock the entire building. There is no doubt that the blast wave will impact the entire building. Depending on the exact location of the blast, it is possible that a section of the building will collapse as well.

Indiscriminate Shooting Attacks

Indiscriminate shooting attacks aim to kill as many people as possible, regardless of who they are. Such a method can be executed in a number of ways, among them, entering the building and opening fire, shooting from a moving vehicle, etc.

One of the main problems facing security authorities in many countries is the fact that they are not allowed to carry weapons themselves. This means that in the case of a shooting attack, the chances of effective return fire are slim. The security staff can only attempt to minimize the damage by preventive measures and by damage control after the attack is over. In such a scenario, time lost equates to lives lost. Since the security staff cannot respond, the attackers will have a free hand to continue the attack; civilians are similarly unarmed and will thus not be able to stop the attack. Any emergency plan should include a response to such an attack, including evacuation of wounded and direction of people to a shelter.

Shooting attack in public building (eg. Shopping Mall)

Use automatic weapons and explosive in crowded public area.	Terrorists enter the building with weapons concealed in bags or under coats. The gunmen start shooting and throwing grenades. A hostage-taking scenario can evolve from here or the operation can be a suicide attack, ending with a firefight with the authorities and the eventual death of the attackers. If the building has entrances to other public areas, it would be easy to move from one location to the other.	Planning time: 5 Resources: 10 Difficulty: 15
	Result: Possible large number of casualties; Closure of the facility for several hours / days; Creation of an incident in which security personnel may shoot at each other.	Physical/Economic Damage: 15 Casualties: 20 Symbolic: 15

Obviously, the list of potential scenarios is huge. It is the job of the on-site investigative team to come up with as many potential attacks as possible based on the venue's vulnerabilities.

Evaluation of Organizational Sensitivities

Each category of terrorist organization receives four "sensitivity" scores, as shown in Table 4. These scores will be combined with each attack-type's Difficulty and Impact factors to determine the importance of these factors to the organization type. When the resulting "Weighted Difficulty" is subtracted from the resulting "Weighted Impact," the result is the "Attractiveness" of that attack to that organization type.

Table 4. Sensitivity parameters

(Negative interest indicates aversion to causing this type of damage.)

Parameters	Range
Sensitivity to Difficulty	0 – 100%
Interest in causing physical/economic damage	-100% – 100%
Interest in causing casualties	-100% – 100%
Interest in symbolic value of targets	-100% – 100%

As mentioned above, the organizations selected for analysis will be those having the highest overall threat score, estimated in Stage 1. Based on our familiarity with the characteristics of these organizations, we can estimate their sensitivities. Table 4 below shows a sample of such a categorization; since we have not specified a venue, we will use group types, rather than actual organizations, for the analysis. Note that this table is based only on a generic analysis; in order to be truly effective, it would require far more parameters than those shown—an analysis that is beyond the scope of this paper.

Table 5. Organization-type sensitivities

	Al-Qaida "Central"	Local Islamists	Ad hoc Groups
Sensitivity to Difficulty	25%	60%	90%
Interest in Physical / Economic Damage	100%	60%	95%
Interest in Casualties	100%	80%	-70%
Interest in Symbolic Value of Targets	100%	100%	90%

4 Synthesis

Putting all these factors together, we can derive the following numerical breakdown, indicating the attractiveness of specific types of attack to the groups discussed above.

Attack Type	Al-Qaida Central	Local Extremist Groups	Ad Hoc Groups
Open Space Bomb	60	32	8
Restaurant Bomb	58	25	-8
Suicide Bomb	63	30	-3
Suicide Car Bomb	69	32	-3
Static Car Bomb	54	16	-20
Car-Park Bomb	80	32	-13
Open Space Shooting Spree	43	22	-4
Mortar/Missile Attack	16	-9	-29
Food Poisoning	23	0	-15
Bio-Agent in Food	34	-1	-32
Explosive Bio-Device	57	16	-24
Pressurized Bio-Device	55	13	-28
Bio-Agent in A/C	53	7	-36
Bio-Agent in Fire Sprinklers	52	4	-40
Suicide Bio-Aerosol	54	15	-16
Leave Radioactive Substance	40	6	-25
Radioactive in A/C	33	-11	-47
Crash Private Plane	27	1	-11
Arson	20	7	1

By graphing this information, we can get a more visually satisfying view of the data:

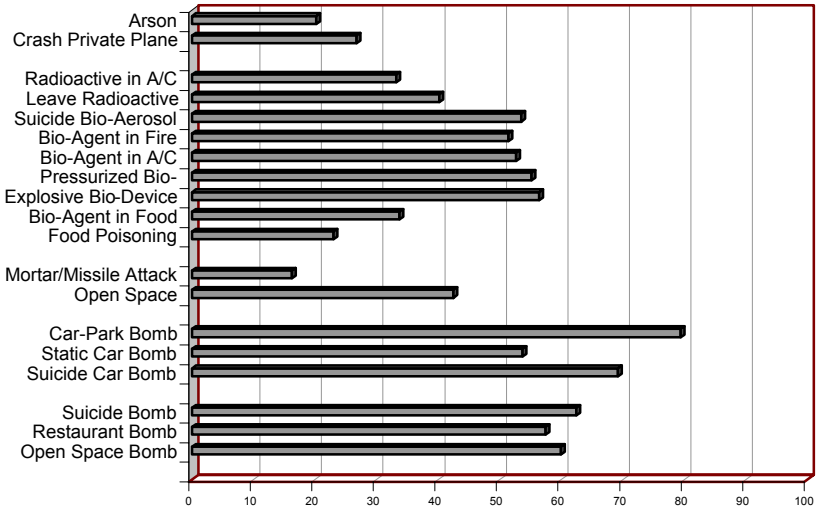


Fig. 1. Attacks attractive to al-Qaida "Central"

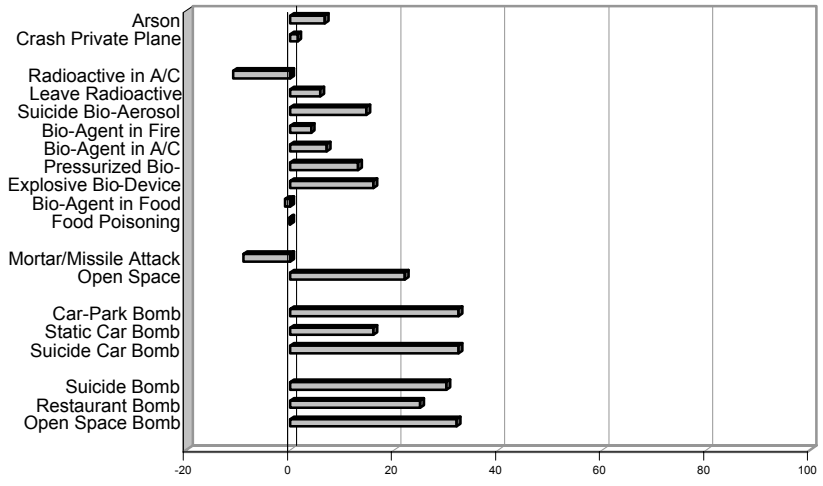


Fig. 2. Attacks attractive to local Islamists

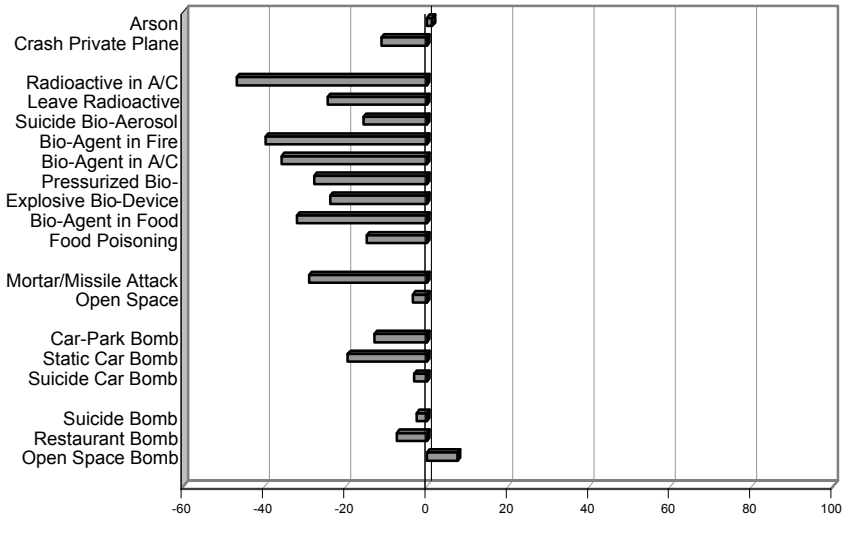


Fig. 3. Attacks attractive to Ad hoc Groups

4 Summary: Scenarios Most Likely to Be Carried Out by Relevant Terrorist Groups

As described above, we have attempted to establish criteria for judging which attack scenarios might prove "attractive" to the various categories of terrorist groups that might be interested in attacking the venue in question. As our judgments regarding the affinities and sensitivities of these groups are necessarily approximate, and our judgments as to the difficulties and impact of the various forms of attack are not

necessarily the same as the judgments made by the terrorists themselves, the specific numbers and rankings should be taken only as a general guide.

Nonetheless, we can draw some conclusions from the analysis:

- The most "attractive" types of attack for al-Qaida and its associates are "high quality" explosive attacks—particularly car-park bombings aimed at destroying or severely damaging the entire facility. Many other types of attack are reasonably "attractive" as well, but there is no pressing reason for these organizations to use non-conventional weapons where explosives will serve well enough.
- Local extremist groups are likely to select from a smaller menu of attack types; they are most likely to carry out bombings or shooting attacks, but may also attempt a low-level non-conventional attack.
- Ad hoc groups, in our view, should not find a tourist venue to be all that attractive a target. As most of the attack scenarios we envision involve significant casualties, and we believe that these groups generally prefer to avoid killing many people, it would seem that the venue should not be high on their list of targets. However, should one of these groups nevertheless choose our venue as a target, certain types of attack are likely to be more attractive than others; and they may also elect to perform bombings with warnings phoned in before the explosives detonate, in order to cause large-scale damage while minimizing casualties. As such an attack has in fact occurred at tourist venues in the past, we must assume that it can happen again.

5 Further Enhancements – Advanced Data-Mining Capabilities

This method can be further enhanced to include data-mining abilities using open source information and building on the ICT database of terrorist incidents. This raw data would provide the basis for a continual reevaluation of the organizational factors that enter into the equation, as well as the availability of resources that could be used in the perpetration of certain types of attack. The result will be a dynamic threat assessment apparatus which could be programmed to "flag" certain types of incidents and organization when the relevant threat indices reach a critical level.

This enhancement would build on current data-mining technologies and off-the shelf database tools. The end product would be applicable to governments, installation security, public transport officials, and a host of other sectors that could be targeted by terrorists.

The enhancement of the model to include advanced data-mining technologies will enable the model to become self-adjusting to changes in terrorist organization make-up, local circumstances, and political realities.

6 Conclusion

The value of the model discussed here lies not in its ability to reveal new information, but rather in its usefulness as a planning and pedagogical aid: a means to visualize data so that trends may be readily apparent where they might otherwise have been lost

in a sea of data. Such data-visualization tools are important in helping us better manage our resources.

However, such models, and the intelligence on which they are based, should be seen as only one component in the fight against terrorism—including non-conventional terrorism. We must never forget that terrorism is psychological warfare and its real targets are located not on the battlefield, but on the home front. Thus, a significant portion of our resources must go toward thwarting the effectiveness of terrorism here, at its true target. Terrorism works somewhat like a virus; by hijacking society's own means of communication—the news media—terrorism succeeds in disseminating its own messages. In a sense, the media plays a multi-level role in the chain of terrorism effectiveness—it carries scenes of death and destruction into the homes of every member of society, while at the same time carrying the fears and anxieties of society to the ears of the terrorists. By telling the terrorists what we are most afraid of, we play an active role in their selection of targets and modes of attack.

That terrorists use the media to further their own aims has become a cliché, and yet this knowledge is too often ignored in the heat of the moment. The reaction of the media toward the threat of non-conventional terrorism is a case in point. Education is the only way to break the chain of terrorism—by informing the public of the real extent of the threat. Here too, the media can play a role—this time a positive one. By helping citizens understand what terrorists can and cannot do, and what the individual can do to defend his or her society, we can decrease the effectiveness of terrorism as a means of policy-making. It is up to us, the counter-terrorism professionals, to see that this information reaches both the media and the citizenry.

Incident and Casualty Databases as a Tool for Understanding Low-Intensity Conflicts

Don Radlauer

Associate, International Policy Institute for Counter-Terrorism
DonRadlauer@yahoo.com

Abstract. Today's "low-intensity" conflicts do not involve victory or defeat in the conventional sense; instead, each side attempts to achieve a psycho-political victory by influencing people's thoughts and feelings about the issues in dispute. Casualty statistics are an important element in forming these thoughts and feelings; in turn, a robust incident and casualty database can be an important tool in coming to an accurate understanding of complex conflicts with multiple actors and incident types. For a casualty database to produce meaningful, informative, and accurate results, it must have a rich array of well-defined categories to which to assign incidents and casualty data. It must also be conceived, designed, and administered with a strict adherence to accuracy rather than advocacy as a primary goal.

1 The Need for Casualty Data in Low-Intensity Conflicts

In traditional warfare, victory or defeat is ultimately a matter of physical realities: at some point, one side is no longer able to continue fighting, and either negotiates an end to the conflict or surrenders unconditionally. In this type of conflict, accurate casualty figures are useful to generals and political leaders who need to assess the situation – how much damage each side has suffered and how much fighting capacity each side retains – but it is clear that these statistics (as opposed to the lives they represent) are not in themselves a vital objective of the combatants. They are important only in that they clarify the situation "on the ground".

In contrast, today's "low-intensity" conflicts do not involve victory or defeat in the conventional sense. While these conflicts often have a significant military component, they are more accurately described as "psycho-political" conflicts rather than military ones, in that each side's primary goal is to influence the course of events by changing people's thoughts and feelings regarding the issues and parties involved in the dispute. In such conflicts, casualty statistics have become an important lever for gaining political influence, quite apart from any strictly military significance they may have; in fact, actors in low-intensity conflicts often pursue policies that would at first glance appear antithetical to their own interests, simply because of the political value (positive or negative) of casualty statistics.

While adversaries in low-intensity conflicts might wish to achieve a conventional victory, the nature of low-intensity conflict itself precludes such an accomplishment.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 153 – 170, 2005.

Non-state actors in these conflicts lack the resources, infrastructure, and (in general) disciplined troops to achieve anything more than abject defeat in traditional, formation-against-formation warfare against a competent regular military; and while terror and guerilla attacks against a state actor can exact a heavy price, they do not constitute an existential threat. But at the same time – and paradoxically – the advantages of even the best-trained and equipped army are of little value in defeating non-state opponents. [1] Since non-state militants routinely shelter among non-combatant civilians, and since some of the most deadly weaponry they use (including suicide bombs and fertilizer-based car bombs, for example) can easily be constructed using simple "civilian" ingredients and techniques, only the most draconian measures taken by a state actor can entirely prevent terrorist and guerilla attacks. Such measures carry a very high political price, and in any case seem to have only local and temporary effect.

Neither side, then, can win (or can afford to win) a physical victory in a low-intensity conflict; instead each side must instead struggle to achieve psychological and political victory, or at least to avoid psycho-political defeat. Non-state actors seek to influence decision-making and alter the political climate of the state actor, to gain the attention and sympathy of non-participant states and groups, and gain support and recruits from their own social milieu. State actors seek to pursue their policies without internal or external interference. Victory can be achieved only in the hearts and minds of the antagonist societies themselves, as well as in the hearts and minds of influential non-participants.

In turn, people's thoughts and feelings are largely a product of the information they receive about the course of the conflict. Casualty statistics are a major component of this information: They have huge emotional impact, as they represent human lives lost and damaged; and yet, as numbers, they carry the intellectual *gravitas* of mathematical factuality. In their most over-simplified form, they are also short and memorable, making them an easily remembered "sound bite" in an age of instant – and often superficially reported – news. But despite their importance, these statistics have not in general been subjected to the level of analysis they deserve.[2]

Low-intensity conflicts tend to be highly complex. Each side consists of multiple elements – some civilian, some military or quasi-military, and others somewhere in between. The non-state side in a low-intensity conflict may not even be a "side" in any conventional sense – it is perhaps better to view non-state actors as collections of more-or-less allied "sides" acting semi-independently, rather than as unitary entities. Even state actors can behave in fragmentary ways at times, with civilian groups acting as terrorists or vigilantes.

---

[1] Maj. Hong Kian Wah. "Low-Intensity Conflict," Journal of the Singapore Armed Forces, Vol 26 n3. July-September 2000. Ministry of Defense, Singapore. This brief article gives a good summary of current thinking on the subject, emphasizing the limitations of conventional military approaches in fighting low-intensity conflicts.

[2] See, for example, Lara Sukhtian and Josef Federman, "Children under 17 caught in Mideast crossfire," Associated Press. Houston Chronicle, 5 March, 2005. This piece, while far better than most media coverage of the Israeli-Palestinian "numbers game", entirely neglects the issue of gender in analyzing "Intifada" mortality statistics.

Further, low-intensity conflicts consist of many different types of incident, with conventional battle being only one of the less-common possibilities. Fatalities can be the result of terror attacks, guerilla attacks, counter-terror operations, arrest attempts, internecine violence, riots, "work accidents", and so on. With all the different possible types of incident, all the different groups participating in these incidents, and of course innocent victims on both sides, the simplistic body-counts usually used to describe low-intensity conflicts completely fail to convey an accurate picture of what is really going on.

1.1 Competitive Victimization

As stated above, many actors in low-intensity conflicts realize the political importance of fatality statistics, and have begun to craft policy accordingly. The most obvious "value" of fatalities is to promote a sense of victimization, and thus to gain the sympathy and support of well-meaning people in non-participant societies and among the opposing side; this sense of victimization has also been used by some actors to maintain enthusiasm for continued fighting among their own population. This political use of fatality statistics is seen in several current low-intensity conflicts, notably the Israeli-Palestinian "al-Aqsa Intifada".

The asymmetry of low-intensity conflicts applies to the political use of fatality statistics as well as to the combatant forces themselves. It is common for advocates of the state actor in a low-intensity conflict to wish to emphasize the extent of their own side's victimization by the non-state side; but in fact this strategy, while it works well for the underdog (virtually always the non-state actor), can have a boomerang effect on a state actor. States benefit from an atmosphere of stability: They profit from trade, encourage capital investment, and benefit from tourism. By emphasizing the extent to which a state is victimized by terrorism and other violence associated with low-intensity conflict, the state's partisans are likely to do the state more harm than good.

2 Extracting Clarity from Chaos

It is clear, then, that while casualty statistics have become centrally important in low-intensity conflicts, the extreme complexity of the conflicts themselves, in addition to the intense emotional and political impact of casualty data, makes it quite difficult to arrive at an accurate and comprehensive picture of the deaths and injuries incurred. However, a robustly-constructed and carefully-administered database system can do a great deal to clarify the history of a low-intensity conflict; and perhaps, by improving the understanding of the nature of the casualties incurred and exposing the extent to which lives are being sacrificed to achieve political "points", such a database may help to reduce the human cost on both sides.

In order to achieve these ends, a database for low-intensity conflict must provide features not found in traditional terror-incident databases. In particular, it must include detailed data for each fatality, since the usual practice of including only aggregate casualty figures does not allow for demographic and other analyses of those killed.

2.1 Basic Structure and Potential Accomplishments of a Casualty Database

In order to improve our understanding of a low-intensity conflict, a database system must go far beyond the usual reporting of each side's total body count. Both incidents and victims must be classified according to as rich a set of criteria as is realistically possible. Some of these criteria and their utility will be discussed in this section; the next section will cover in greater detail some of the "real world" difficulties and dilemmas that must be confronted in applying such a classification scheme.

While this paper is intended to address low-intensity conflicts as a generality, specific reference will be made to the International Policy Institute's "al-Aqsa Intifada" Database to demonstrate the potential accomplishments of a low-intensity-conflict database application. Examples will be provided showing some of the significant findings of the "al-Aqsa Intifada" Database Project; but these are included only to demonstrate the methodology used. Discussion of these findings in relation to the issues involved in the Palestinian-Israeli conflict itself is outside the scope of this paper.

For reasons that will be discussed in the next section, it has proven impractical to deal directly with injury statistics in our work. Instead, we have chosen to focus almost exclusively on fatalities, on the assumption that deaths, as a subset of total casualties, can be used as a representative sample for purposes of analysis. Discussion of various aspects of our database will reflect the fact that while its design permits tracking of injuries as well as of fatalities, in practice we work only with the latter.

2.2 Combatant Levels

One of the most significant – and contentious – categories for understanding casualty statistics is the combatant status of those killed and injured. The most emotionally significant casualty in any conflict is the innocent victim: While most people take the death or injury of an armed soldier or militant more or less in stride, almost everyone regrets and condemns the killing of the defenseless and innocent civilian. It is important, then, to determine how many of those killed and injured on each side were combatants. In order to enable detailed and flexible reporting of the combatant status of those killed in the conflict, we have divided fatalities into nine categories:

- *Full Combatant.* A "full combatant" is a soldier on active duty, an active member of a terrorist group, a civilian independently choosing to perpetrate an armed attack against the opposing side, or someone using a "hot" weapon to defend against an attack. In general, rock-throwers are not considered to be combatants; an exception to this generalization would be, for example, someone dropping large rocks from a bridge onto fast-moving traffic. A rioter throwing "Molotov cocktails", grenades, or the like can be considered a full combatant.

  Mere possession of a weapon does not imply combatant status. A civilian driving with a weapon in his/her car, or a pedestrian with a holstered pistol, is

normally considered a non-combatant. However, a civilian who encounters a terror attack in progress and draws his/her weapon in an attempt to stop or prevent the attack *is* a combatant once the weapon is out of its holster.

- *Probable Combatant* A "probable combatant" is someone killed at a location and at a time during which an armed confrontation was going on, who appears most likely – but not certain – to have been an active participant in the fighting. For example, in many cases where an "Intifada" incident has resulted in a large number of Palestinian casualties, the only information available is that an individual was killed when Israeli soldiers returned fire in response to shots fired from a particular location. While it is possible that the person killed had not been active in the fighting and just happened to be in the vicinity of people who were shooting, it is reasonable to assume that the number of such coincidental deaths is not particularly high. Where the accounts of an incident appear to support such a coincidence, the individual casualty has been given the benefit of the doubt, and assigned a non-combatant status.
- *Violent Protester* A "violent protester" may not be a full-time militant, but has taken an active and violent part in rioting or vigilante activity – such as throwing incendiary devices. This category is a subset of "full combatants", created to enable more specific reporting of riot-related fatalities.
- Full Combatants, Probable Combatants, and Violent Protestors are normally aggregated into a total "Combatant" figure; all the remaining categories are considered "Non-combatants".
- *Non-Combatant* A non-combatant is an innocent bystander – a person whose death or injury has no justification other than nationality or ethnicity. This category is used as a "catch-all" for those non-combatants who do not fall into one of the more specific non-combatant categories.
- *Health Related* A "health related" fatality is someone who died from a cause only indirectly related to violence – for example, due to a heart attack following an incident, tear-gas inhalation, or a roadblock delay that prevented an ill person from receiving medical treatment in a timely manner.
- *Uniformed Non-Combatant* A "uniformed non-combatant" is a non-civilian, but is not actively involved in the conflict. This category can include civil police as well as soldiers in uniform but not on active duty.
- *Protestor Unknown* A "Protestor Unknown" is anyone who was killed during a protest for whom information as to violent behavior is unavailable.
- *Suspected Collaborator* This is a special category for people targeted by militants of their own nationality who suspect them of aiding the enemy.
- *Unknown* In some cases, the information at hand may be insufficient to decide the circumstances of death for a given casualty. In our experience, this was especially true in the early days of the "Intifada", when many Palestinians were reported as having been killed, but with minimal information as to the circumstances surrounding their death. Even when detailed reports are available, however, there are cases in which it is impossible to draw a conclusion as to the combatant status of those killed – especially when each side's report of a given incident is radically different from the other side's, with both versions appearing reasonably credible.

The fact that all "Unknowns" are classified as Non-combatants serves as a "safety factor" in our analysis: Since a large number of Palestinian fatalities are classified as "Unknown" compared to a tiny number of Israelis, and since it is reasonable to assume that a substantial proportion of these "Unknowns" were in fact combatants, we can state with high confidence that our figures for total combatants among Palestinians are not inflated.

As an example of the results of our combatant-level classification system, the following two pie charts show the combatant status of Israeli and Palestinian fatalities:

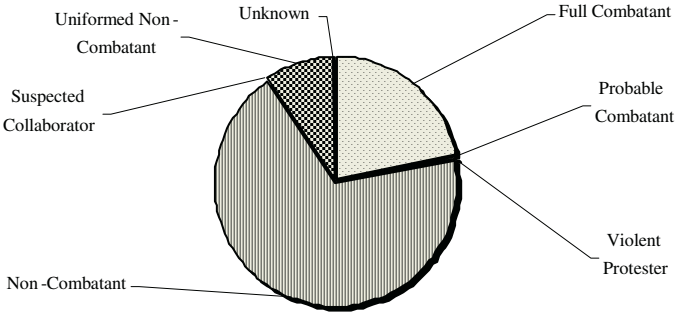


Fig. 1. Breakdown of Palestinian fatalities by Combatant Level

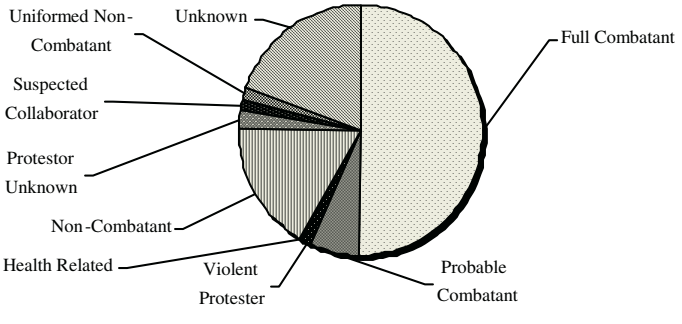


Fig. 2. Breakdown of Israeli fatalities by Combatant Level

2.3 Incident Types

As mentioned above, low-intensity conflicts are typically a mixture of many different types of incident. In order to make some sense of the confusion, we have categorized incidents at two levels: Incident Type and Attack Type. Incident Type refers to the general nature of what happened, and includes the following possibilities:

Terror Attack
Guerilla Attack
Attempted Infiltration
Crossfire
Violent Clash
Riot/Violent Demonstration
Roadblock Confrontation
Counter-Terror Operation/Interception
Initiated Military Operation
Targeted Killing
Work Accident
Internecine Violence
Unclaimed Killing
Unrelated Accident
Unknown

Attack Types are used to provide more detail as to exactly what form of violence took place, and are in turn divided into several "meta-attack-types":

Bombings (non-suicide)
- Bombing
- Car Bomb
- Letter Bomb
- Bomb Threat

Suicide Bombings
- Suicide Bomb
- Suicide Car Bomb

"Cold Weapon" Attacks
- Knife Attack
- Stoning
- Arson
- Vandalism

Armed Assaults
- Shooting
- Ground-to-Ground Missile
- Artillery
- Vehicle Attack
- Hand Grenade/RPG
- Infiltration
- Incendiary Device
- Mortar Attack
- Helicopter Missile

Kidnappings
- Kidnapping
- Hostage-taking
- Hijacking
- Lynching

Unknown/Other Attacks
- Chemical Attack
- Unknown

Incident Type and Attack Type classifications are highly useful in answering "nuts and bolts" questions about a low-intensity conflict, such as "How many people were killed in suicide bombings?" and the like. Combined with other criteria, such as demographic categories (see below), they can provide some highly interesting and unexpected results.

2.4 Demographics

The age and gender of each person killed have turned out to be two of the most useful pieces of information in our database. Not only are these data easily obtained and uncontroversial (unlike Combatant Level, for example); they have yielded some very interesting results.

We treat victims' ages in three ways: For most of our demographic analyses, we use five-year age categories – i.e. 0-4 Years, 5-9 Years, and so on. For our analysis of the deaths of young people, we need a more detailed view – so we look at age year-by-year. And to compare the impact of the conflict on specific population sectors, it has proven useful to work with specific age categories:

- Children: Ages 0 through 11
- Adolescents: Ages 12 through 17
- Young Adults: Ages 18 through 29
- Adults: Ages 30 through 44
- Mature Adults: Age 45 and over*
- Adolescents Plus Young Adults: Ages 12 through 29

Age and gender analysis has yielded some of the most surprising results for the "al-Aqsa Intifada". For example, age distributions comparing combatants with non-combatants are very different for the two sides:

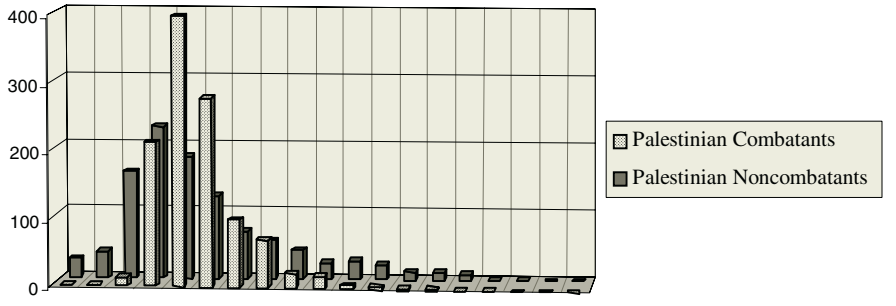


Fig. 3. Israeli combatant and noncombatant fatalities, broken down by five-year age categories. Note the very regular narrow distribution of ages of combatants, compared with a broader and "sloppier" spread for the ages of noncombatants

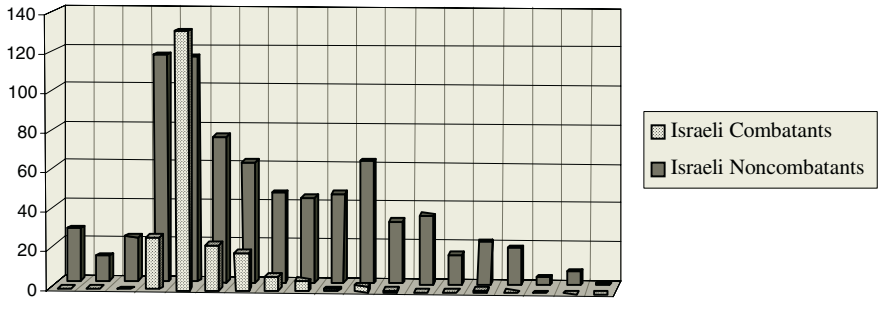


Fig. 4. Age distribution of Palestinian combatant and noncombatant fatalities. Note that while the spread of ages for Palestinian combatants is broader than for Israeli combatants, the age profile of Palestinian noncombatant fatalities is much narrower and more regular than is the case for Israeli noncombatants

The detailed breakdown of childhood and adolescent fatalities by age and gender again showed patterns that we had not expected beforehand:

* As the author would otherwise be entering this category during the course of the year 2005, it is possible that the minimum age for "Mature Adult" will be preemptively increased to forestall this eventuality.

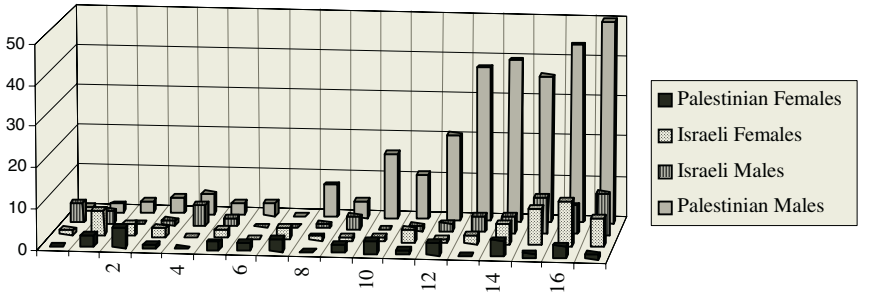


Fig. 5. Young victims of the conflict, arranged by age and gender. Note that fatalities are relatively uncommon among young children on both sides and of both genders. Israeli fatalities begin to increase at around age 13, with an essentially even balance between males and females. Palestinian fatalities show a tremendous increase between the ages of 10 and 13 among boys, but no increase with age among girls

One of our earliest findings was that Palestinian fatalities, including those classified as non-combatant, were overwhelmingly male, while Israel fatalities were much more balanced – about sixty per cent male:

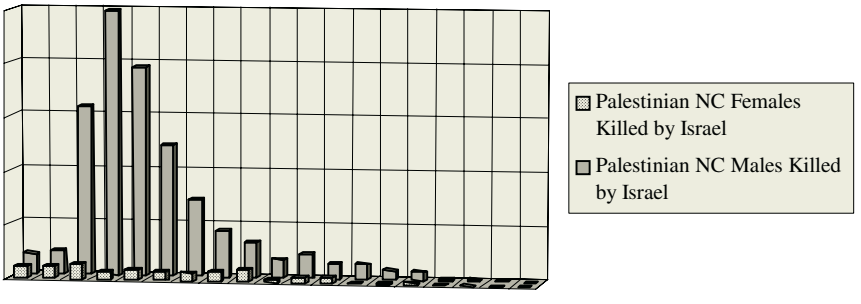


Fig. 6. Age distribution (in five-year brackets) of Palestinian noncombatants killed by Israel, with males and females separated. The preponderance of males is obvious; in addition, males show a very regular age distribution, while females do not

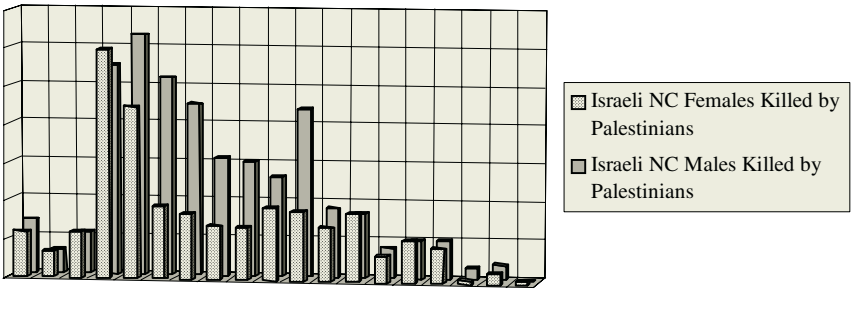


Fig. 7. Age distribution of Israeli noncombatants killed by Palestinians. The distribution is broad and "sloppy" for both genders; fatalities are essentially equally balanced between males and females, except for those between 20 and 54 years old

The combination of demographic information with Incident Type and Attack Type shows considerable promise as an analytical approach. While we have only begun to explore the possibilities of this type of analysis, our initial attempt yielded some surprising results regarding the demographics of Palestinians unintentionally killed in Israeli "targeted killings" compared to the demographics of Palestinian non-combatants killed in other types of incident:

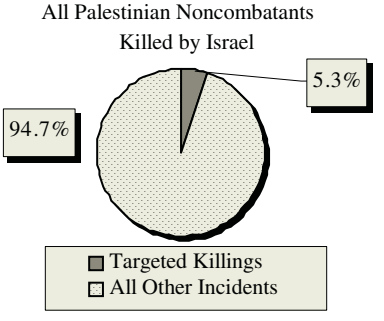
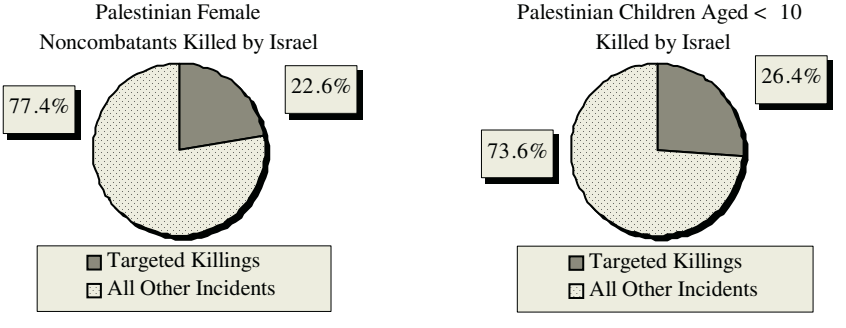


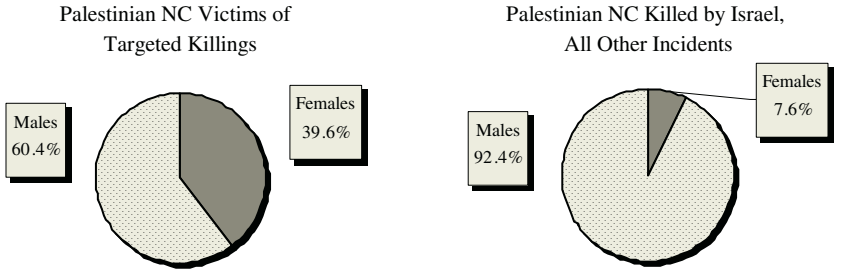
Fig. 8. The prevalence of "collateral victims" of Israeli "targeted killings" among Palestinian noncombatants killed by Israel



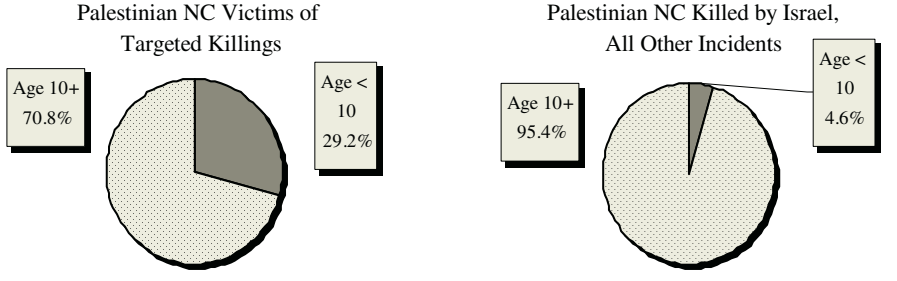
Figs. 9-10. The prevalence of "collaterals" among Palestinian females and young children killed by Israel. While **Figure 8** showed that "collaterals" amount to just over five percent of all Palestinian noncombatants killed by Israel, these figures show that they include a disproportionate share of female and young victims

2.5 Time

Another basic area of analysis is the way in which a low-intensity conflict can change over time. It is quite common to talk and write about these conflicts as if they were single, unchanging events; but in fact they show significant change over time, and it is quite typical for them to consist of a number of fairly distinct phases, with different rates of death and injury, and even differences in the demographic characteristics of the victims. These phases are signaled by various events: mediation efforts, truces,



Figs. 11-12. Another look at the comparative demographics of "collaterals" versus other noncombatant Palestinians killed by Israel. Among unintended victims of "targeted killings", some 40% were female – comparable to the percentage of females among Israeli noncombatants killed by Palestinians. But among Palestinian noncombatants killed in other types of incident, the number of females is much lower



Figs. 13-14. The same type of comparison gives a similar result for young children: they make up a large proportion of the unintended victims of "targeted killings", but a very small proportion of Palestinian noncombatants killed in other types of incident

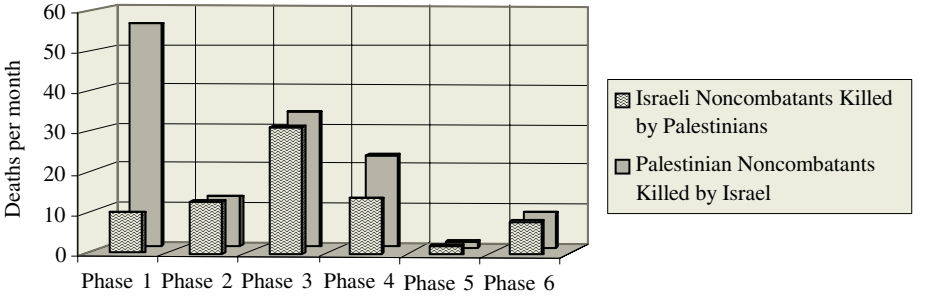


Fig. 15. Change in death rates of noncombatants on each side, separated by phase and normalized per standard 30-day month

major attacks or incursions, or even significant outside events such as the 9/11 attacks. (The latter attacks, for example, marked the onset of the third phase of the "al-Aqsa Intifada", the most chaotic nine months of the conflict to date.) Of course,

nobody in a position of authority officially announces that Phase X of a given conflict has ended and Phase Y has begun; only somewhat after the fact is it apparent that a particular event has actually changed the nature of the conflict.

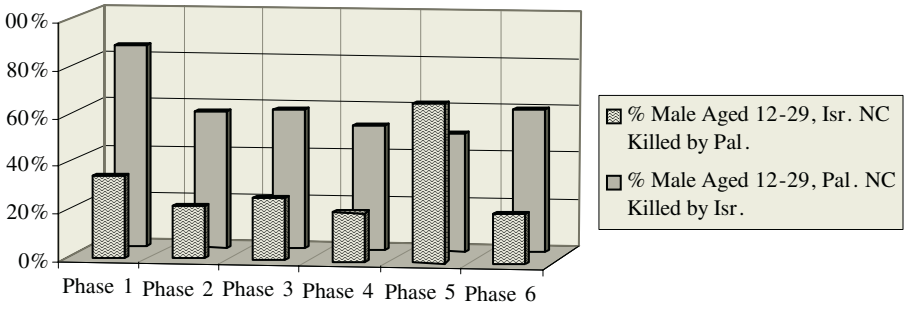


Fig. 16. Change in the percentage of young males among noncombatants killed on each side, separated by phase

3 Classification of Victims and Incidents – Issues, Dilemmas, and Solutions

3.1 Avoiding Bias

In order for the results of a low-intensity-conflict database to be meaningful, data must be gathered and classified in ways that minimize bias and inaccuracy. This is not always as easy and straightforward as it might seem: The information available on incidents and casualties is often unreliable, incomplete, and sometimes contradictory; and some of the classification criteria involve the exercise of judgment, which invites unconscious bias even if those performing the classification are attempting to be scrupulously fair. It would perhaps be best if these databases could be created and maintained by completely disinterested outsiders; but, sadly, the disinterested are generally uninterested in committing the requisite time and effort to investigating someone else's conflict. (In any case, the truly disinterested investigator is probably mythical.)

While the problem of bias is not completely solvable, there are a number of steps that can be taken to minimize the possibility of biased or unreliable results:

- From the outset, it must be made very clear to everyone who works on a low-intensity-conflict database that the goal of the project is an accurate picture of the conflict, rather than to prove one side or another "correct".
- Data must be gathered from a variety of sources, including "enemy" media, non-governmental organizations, and so on, as well as "friendly" media. Reports in "friendly" media that favor the investigators' side in the conflict must be treated with an extra degree of skepticism, since the human tendency is to grant extra credibility to favorable news.

- Data should be gathered, entered into the database, and categorized by one team, and a separate team should perform all analysis. By separating these functions, one avoids contaminating the judgment of the person entering data with thoughts of "How will this look on my graphs?"
- Categories should be defined clearly and specifically, to leave as little as possible to the discretion of the person assigning categories. We have found it helpful to test our criteria on difficult or strange hypothetical cases, to help establish the boundary conditions for category assignment.*
- Categorization should as much as possible be based on physical rather than ethical or political criteria. As discussed below, while we do track incident-level responsibility (i.e. who started any particular incident), virtually all our analysis is based on casualty-level responsibility – meaning that we track responsibility for fatalities based purely upon who killed whom, not upon why a particular person was killed. This approach avoids the infinite regress of "who started first" – where each side claims that its own actions are a just and proper response to the other side's previous actions.
- "Safety reserves" (such as our categorization of all Palestinians of "Unknown" combatant level as non-combatants) should be created wherever categories unavoidably involve the use of judgment.
- The more that can be done with the least controversial data – such as the age and gender of victims – the better. To the extent that we can learn interesting things from simple demographic data, we significantly reduce the possibility that bias will color our results.
- Raw data and categorizations of each incident and casualty should be made available to the general public. By making a database inquiry function available on a website, the investigator makes it plain that the general public does not have to take the published aggregate statistical results on faith. In addition, skeptical (and energetic) members of the public may find and report genuine problems with the categorization of specific incidents and casualties, improving the quality of the data. (Sadly, our experience has been that while many people accuse us of bias simply because we are Israeli, virtually none of them make the effort to check the accuracy of our work.)

* One rather gruesome hypothetical case was this: A pregnant woman attempts a suicide bombing, but is identified by security forces before she can set off her explosives. In order to prevent her from triggering her explosives, a policeman fires at her; the bullet sets off the explosive. Which side, then, is responsible for the death of her unborn child? After some deliberation, we decided that in this case the would-be suicide bomber is responsible, even though she did not in fact set off the explosives – since she placed the explosives in such a way that they would cause the unborn child's death upon detonation, and one way or another they were almost certainly going to detonate. The security forces, however, would be responsible for the woman's death, even though she was intending to blow herself up and thus is held responsible for the incident itself. Had the unborn child been killed by gunfire, rather than the explosion of the bomb, the security forces would be responsible for both deaths.

3.2 Civilians and Non-combatants

Media reports frequently discuss the casualties of low-intensity conflict in terms of the number of "civilian fatalities" on each side. We have deliberately avoided this usage. In any conflict between a country with conventionally-organized military and police forces and an opposing force mostly composed of non-uniformed "irregulars", the state actor's forces cannot avoid killing a disproportionate number of "civilians" – since even their most deadly opponents are usually not members of an official military or security force, and in many cases have perfectly respectable "day jobs". Further, people on either side of a low-intensity conflict may act in different capacities at different times – for example, many members of Palestinian security forces combine their official service with membership in one or more unofficial groups such as Hamas or the various arms of Fatah. When Palestinians in this situation have killed Israelis, they have generally done so in their "civilian" capacity.

At first glance, it should be easier to determine which state-actor fatalities are "civilians". However, even here the distinction between "civilians" and members of official security forces paints a somewhat distorted picture. A substantial number of Israeli fatalities, especially those killed inside "Israel proper", have been members of the civil police, or noncombatant members of the Israel Defense Forces such as office workers and mechanics. By most internationally accepted definitions, such individuals are considered to be non-combatants. (See, for example, the U.S. State Department's definition of "noncombatants" in their "Patterns of Global Terrorism" reports. It is worth noting, however, that we differ from the State Department in our definition of terrorism: They define terrorism as attacks against *non-combatant* targets, while we define it as attacks against *civilian* targets.)

As a result of all these factors, dividing a low-intensity conflict's fatalities into "civilians" and "non-civilians" over-emphasizes the "civilian" status of many of the non-state side's victims, and to a degree distorts the significance of the state actor's fatalities as well. At best, such categorization paints an inaccurate picture of the conflict; and in some instances, those who use these categories are clearly being disingenuous in claiming the deaths of active militants as "civilian casualties".

For this reason, we chose to classify those killed by their actual combatant status, according to the criteria laid out in the "Combatant Level" section above. While this method requires a degree of judgment in categorizing those killed, it offers some hope of making sense of an asymmetrical conflict; whereas the alternative system, while easier to apply, cannot provide meaningful results.

3.3 When a Terrorist's Not Engaged in His Employment

The purpose of using our "combatant/non-combatant" criteria is, in essence, to produce a fair picture of the extent to which each side is limiting its attacks to legitimate targets – that is, targets that are properly considered military or quasi-military. Thus, as mentioned above, an armed civilian is considered a non-combatant – until s/he draws a weapon; and an armed soldier on duty is a combatant, even if he was asleep at his post and never fired a shot. But terrorists (or guerilla fighters – in the "Intifada" both types of attack are carried out by the same organizations) present a problem: Are they combatants when they are not actively involved in carrying out an attack?

This question triggered an internal debate. If an army general were killed in his home, we would normally classify him as a Uniformed Non-Combatant. But if the head of a terrorist organization were killed in *his* home, we were classifying him as a Full Combatant. This struck some of us as fundamentally unfair. Finally, we decided that our initial approach was correct – or at least more correct than any other approach we could think of – because it came closest to meeting our fairness criteria. As with so many other problems in dealing with low-intensity conflict, the difficulty here stems from the conflict's asymmetry. IDF generals spend a large amount of their time at their job, just as other soldiers do. During these periods, they are considered combatants even if they are not actively fighting. In essence, the nature of being a soldier is that one spends a large percentage of one's time being a legitimate military target. Terrorists and guerilla fighters – and even more so the leaders of terrorist/guerilla groups – do not "report for duty" as soldiers do; they perform no "routine patrols", and in general do not allocate a regulated portion of their time to being official "military targets". Since so little of their time is spent in actual "military" activity (and none at all for the leaders of terrorist groups), we have to consider them as essentially "full-time combatants". (This is similar to the status of undercover espionage agents in enemy territory, who are legally considered to be full-time combatants even when they are doing nothing overt to harm their target country.)

This is not a perfect solution to the problem, of course. It illustrates the fact that creating a "level playing field" for analyzing a fundamentally asymmetrical conflict is a goal to which we aspire, but which we can never fully achieve.

3.4 Assigning Responsibility

As mentioned above, we assign responsibility at both an incident and casualty level. In both cases, we use "responsibility" in a strictly physical sense. Thus, an arrest attempt is the responsibility of the security forces making the attempt, without regard to the guilt or innocence of the person they are attempting to arrest.

Incident-level responsibility can present some interesting borderline cases, particularly in regard to attacks that are foiled before they can be carried out. Once a terrorist or guerilla reaches his target and begins his attack, the incident is clearly "his". On the other hand, if security forces receive advance warning of an attack and are able to intercept the attacker before he has reached his target, responsibility for the incident rests with the security forces. What happens, though, when an attacker reaches the area of his target, but is intercepted before he can succeed in carrying out his attack?

Our practice here is to judge based on whether security forces intercepted the attacker based upon advance warning of the attack, or whether they were alerted by the actual approach of the attacker (who may have set off alarms cutting through a fence, for example). If security forces were responding to an alarm the attacker set off, the incident-level responsibility rests with the attacker even if he never succeeded in firing a shot or triggering an explosive charge.

Our statistical analysis of the "al-Aqsa Intifada" has so far made use almost entirely of casualty-level responsibility: who killed whom, without regard for who was responsible at the incident level. This approach enables a less "politicized" approach, particularly in analyzing situations, such as violence at riots, in which it is impossible to determine who is responsible for escalating the incident to the point where lives were lost.

4 Technical Issues – Basics, Bells and Whistles

4.1 Platform Choice

We chose to build our database using standard, "off-the-shelf" tools. The database runs on Microsoft Access; but while data-entry screens have been built using Access-specific features, all the database tools for analysis have been built using SQL commands. This approach has the advantage that it enables future portability to other platforms with minimal inconvenience, since SQL – unlike Microsoft Access's application-building tools – is a standardized, multi-platform language. In future, we are likely to transfer the database to Microsoft SQL Server, with Access retained as a data-entry front end if possible.

A large suite of SQL queries (on the order of 50 queries) was created to make data available for graphing and analysis. A series of four Microsoft Excel spreadsheets extract data from the database using these queries, and then process the query results in preparation for graphing. Another set of Excel spreadsheets performs the actual graphing; this division avoids problems we experienced with Excel crashes due, apparently, to the complexity and size of the spreadsheets. Separate graphing spreadsheets also make it easier to produce differently formatted graphs and charts based on the same data and computations.

Our website is based on Cold Fusion; this allows the entire website, including almost all article text, to be database-driven. Cold Fusion allows us to make on-line database query functions available to the public, as well as a summary Breakdown of Fatalities screen which presents the most recent figures from our on-line database.

4.2 Database Structure

The "foundation" tables in our database are Incidents and Casualties. In order to facilitate the required range and complexity of queries we require for analysis, we have added a rather large number of "support" tables:

- *Age* (in five-year brackets)
- *AgeBrackets* (e.g. Adolescents, Adults)
- *AgeHR* ("high-resolution" – that is, one row in the table for each year of age)
- *AttackTypes* and *MetaAttackType* (described above, under "Incident Types")
- *CasualtyTypes* (various levels of injury, from "lightly injured" to "killed")
- *CombatantLevel* (described at length above)
- *ConfidenceLevels* (1 = extremely low, 3 = questionable, 5 = extremely high)
- *Countries*
- *Gender* (male / female / unknown)
- *IncidentTypes* (described above)
- *Months* (one row for each month of the conflict; used by time-series SQL queries)
- *Organizations* (shared with our other terrorism databases)
- *Side* (Israel / Probably Israel / Palestinian / Probably Palestinian / Unclear / None. This table is used in assigning responsibility to incidents and casualties. In practice, we treat the "probable" assignments of responsibility the same as the definite ones.

- *Targets* and *MetaTarget* (respectively, specific targets such as Bank or Marketplace, and general target categories such as Transportation, Civilian Personnel, and Military)

A few relatively unimportant tables have been left out of this list.

4.3 Summary Statistic Generation Infrastructure

To enable rapid, automated, and flexible generation of aggregate and computed statistics for multiple time periods, the database includes a powerful table-configured summary statistic generator. This generator consists of a series of SQL queries that make use of the following tables:

- *Intervals*: This table includes one row for each time interval for which summary statistics will be generated, such as the entire conflict, each calendar year, and the various phases that have been designated. Intervals may be flagged as "open", meaning that their ending date is automatically extended each time summary statistics are generated.
- *SummaryInstructions*: This table tells the SQL queries what calculations to perform. Up to 40 counter variables and 30 computed variables can be defined. For counter variables, we can specify the following criteria:
  - Variable Number (1-40) and Name
  - From Table (Casualties or Incidents)
  - Age Bracket (uses the brackets defined in the AgeBrackets table)
  - Gender (uses the genders defined in the Gender table)
  - Combatant (combatants, non-combatants, or all)
  - Side Responsible (casualty level responsibility, Israeli / Palestinian / Unknown)
  - Nationality (Israeli / Palestinian / Foreign)
  - Incident Type (from the IncidentType table)
  - Attack Type (from the AttackType table)
  - Organization (applied to the organization responsible for the incident when From Table is Incidents, or to the organization a casualty belonged to when From Table is Casualties)

For computed variables, we can select any two counter variables ("X" and "Y") that we have defined using the criteria listed above, and perform one of the following operations on them:

- Percentage computes X as a percentage of X plus Y.
- Ratio computes X as a percentage of Y.
- Daily computes X as a daily rate over the length of the time interval.
- Monthly computes X as a monthly rate over the length of the time interval, normalized to 30 days per month.
- Yearly computes X as an annual rate over the length of the time interval.

The daily, monthly, and annual rate computations allow conflict phases of different lengths to be compared on an equal basis.

When the summary statistic generator is triggered, all the variables defined in SummaryInstructions are generated for all the time periods defined in Intervals; the results are stored in the SummaryData table, with one row for each Interval.

5 Conclusion

A well-constructed database application is one of the best tools available for understanding the complexity of low-intensity conflicts. However, the most meaningful and accurate results can be achieved only if a number of problems are addressed, not all of them technical:

- The database must be sufficiently "rich" in detail that a large number of different criteria can be recorded and analyzed.
- A substantial numerical "back end" is needed to process values from the database in order to find meaningful trends and relations, and to provide clear graphical data displays.
- Categories must be carefully and precisely defined, especially for politically sensitive issues like the combatant level of casualties.
- Those designing the database, administering the project, and analyzing the data must maintain a strong and consistent commitment to accuracy – even when the results will make their own side in the conflict look less than perfect. If the results produced cannot withstand the most skeptical scrutiny, the entire exercise will be a waste of time and effort.

Integrating Private Databases for Data Analysis*

Ke Wang[1], Benjamin C. M. Fung[1], and Guozhu Dong[2]

[1] Simon Fraser University, BC, Canada
{wangk, bfung}@cs.sfu.ca
[2] Wright State University, OH, USA
gdong@cs.wright.edu

Abstract. In today's globally networked society, there is a dual demand on both information sharing and information protection. A typical scenario is that two parties wish to integrate their private databases to achieve a common goal beneficial to both, provided that their privacy requirements are satisfied. In this paper, we consider the goal of building a classifier over the integrated data while satisfying the k-anonymity privacy requirement. The k-anonymity requirement states that domain values are generalized so that each value of some specified attributes identifies at least k records. The generalization process must not leak more specific information other than the final integrated data. We present a practical and efficient solution to this problem.

1 Introduction

Nowadays, one-stop service has been a trend followed by many competitive business sectors, where a single location provides multiple related services. For example, financial institutions often provide all of daily banking, mortgage, investment, insurance in one location. Behind the scene, this usually involves information sharing among multiple companies. However, a company cannot indiscriminately open up the database to other companies because privacy policies [1] place a limit on information sharing. Consequently, there is a dual demand on information sharing and information protection, driven by trends such as one-stop service, end-to-end integration, outsourcing, simultaneous competition and cooperation, privacy and security.

Consider a concrete scenario. Suppose a bank A and a credit card company B observe different sets of attributes about the same set of individuals identified by the common key SSN, e.g., $T_A(SSN, Age, Balance)$ and $T_B(SSN, Job, Salary)$. These companies want to integrate their data to support better decision making such as loan or card limit approval. However, simply joining T_A and T_B would reveal the sensitive information to the other party. Even if T_A and T_B individually do not contain sensitive information, the integrated data can increase the

* Research was supported in part by a research grant from Emerging Opportunity Fund of IRIS, and a research grant from the Natural Science and Engineering Research Council of Canada.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 171–182, 2005.

Table 1. Compressed tables

Shared		**Party** A		**Party** B		
SSN	**Class**	**Sex**	...	**Job**	**Salary**	...
1-3	0Y3N	M		Janitor	30K	
4-7	0Y4N	M		Mover	32K	
8-12	2Y3N	M		Carpenter	35K	
13-16	3Y1N	F		Technician	37K	
17-22	4Y2N	F		Manager	42K	
23-25	3Y0N	F		Manager	44K	
26-28	3Y0N	M		Accountant	44K	
29-31	3Y0N	F		Accountant	44K	
32-33	2Y0N	M		Lawyer	44K	
34	1Y0N	F		Lawyer	44K	

possibility of inferring such information about individuals. The next example illustrates this point.

Example 1. Consider the data in Table 1 and taxonomy trees in Figure 1. Party A and Party B own

$$T_A(SSN, Sex, \ldots, Class) \text{ and } T_B(SSN, Job, Salary, \ldots, Class)$$

respectively. Each row represents one or more original records and $Class$ contains the distribution of class labels Y and N. After integrating the two tables (by matching the SSN field), the "female lawyer" on (Sex, Job) becomes unique, therefore, vulnerable to be linked to sensitive information such as $Salary$. To protect against such linking, we can generalize *Accountant* and *Lawyer* to *Professional* so that this individual becomes one of many female professionals. No information is lost as far as classification is concerned because $Class$ does not depend on the distinction of *Accountant* and *Lawyer*. ■

In this paper, we consider the following *secure data integration* problem. Given two private tables for the same set of records on different sets of attributes, we want to produce an integrated table on all attributes for release to both parties. The integrated table must satisfy the following two requirements:

Privacy Preservation. The k-anonymity requirement: Given a specified subset of attributes called a "quasi-identifier," each value of the quasi-identifier must identify at least k records. The larger the k, the more difficult it is to identify an individual using the quasi-identifier. This requirement can be satisfied by generalizing domain values into higher level concepts. In addition, at any time in this generalization, no party should learn more detailed information about the other party other than those in the final integrated table. For example, *Lawyer* is more detailed than $Professional$.

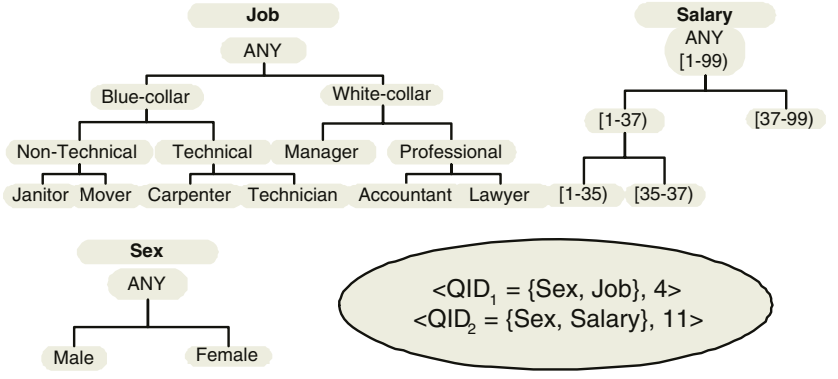


Fig. 1. Taxonomy trees and QIDs

Information Preservation. The generalized data is as useful as possible to classification analysis. Generally speaking, the privacy goal requires masking sensitive information that are *specific* enough to identify individuals, whereas the classification goal requires extracting trends and patterns that are *general* enough to predict new cases. If generalization is "carefully" performed, it is possible to mask identifying information while preserving patterns useful for classification.

There are two obvious approaches. The first one is "integrate-then-generalize": first integrate the two tables and then generalize the integrated table using some single table methods. Unfortunately, this approach does not preserve privacy because any party holding the integrated table will immediately know all private information of both parties. The second approach is "generalize-then-integrate": first generalize each table locally and then integrate the generalized tables. This approach does not work for a quasi-identifier that spans the two tables. In the above example, the k-anonymity on (Sex,Job) cannot be achieved by the k-anonymity on each of Sex and Job separately.

This paper makes two contributions. First, we define the secure data integration problem. The goal is to allow data sharing in the presence of privacy concern. In comparison, classic data integration assumes that all information in private databases can be freely shared, whereas secure multiparty computation allows "result sharing" (e.g., the classifier in our case) but completely prohibits data sharing. More discussions will be available in Section 2. In many applications, being able to access the actual data not only leads to superior results, but also is a necessity. For example, the medical doctor will not trust a given classifier without knowing certain details of patient records. Second, we present a solution to secure data integration where the two parties cooperate to generalize data by exchanging information not more specific than what they agree to share.

2 Related Work

Information integration has been an active area of database research. This literature typically assumes that all information in each database can be freely shared [2]. Secure multiparty computation (SMC), on the other hand, allows sharing of the computed result (i.e., the classifier in our case), but completely prohibits sharing of data [3]. Liang et al. [4] and Agrawal et al. [2] proposed the notion of minimal information sharing for computing queries spanning private databases. They considered computing intersection, intersection size, equijoin and equijoin size. Their model still prohibits the sharing of databases themselves.

The notion of k-anonymity was proposed in [5], and generalization was used to achieve k-anonymity in Datafly system [6] and μ-Argus system [7]. Preserving k-anonymity for classification was studied in [8][9][10]. All these works considered a single data source, therefore, data integration is not an issue. In the case of multiple private databases, joining all databases and applying a single table method would violate the privacy constraint private databases.

3 Problem Definition

We first define k-anonymity and generalization on a single table, then the problem of secure data integration.

3.1 The *k*-Anonymity

Consider a person-specific table $T(D_1, \ldots, D_m, Class)$. The $Class$ column contains class labels or distribution. Each D_i is either a categorical or a continuous attribute. Let $att(v)$ denote the attribute of a value v. The data provider likes to protect against linking an individual to sensitive information through some subset of attributes called a *quasi-identifier*, or QID. A sensitive linking occurs if some value of the QID is shared by only a "small" number of records in T. This requirement is defined below.

Definition 1. Consider p quasi-identifiers $QID_1, \ldots, QID_p$ on T. $a(qid_i)$ denotes the number of records in T that share the value qid_i on QID_i. The *anonymity* of QID_i, denoted $A(QID_i)$, is the smallest $a(qid_i)$ for any value qid_i on QID_i. A table T satisfies the *anonymity requirement* $\{<QID_1, k_1>, \ldots, <QID_p, k_p>\}$ if $A(QID_i) \geq k_i$ for $1 \leq i \leq p$, where k_i is the *anonymity threshold* on QID_i. ■

Note that if QID_j is a subset of QID_i, where $i \neq j$, and if $k_j \leq k_i$, then $<QID_j, k_j>$ is implied by $<QID_i, k_i>$, therefore, can be removed.

Example 2. $<QID_1 = \{Sex, Job\}, 4>$ states that every qid on QID_1 in T must be shared by at least 4 records in T. In Table 1, the following qids violate this requirement: <*M*, *Janitor*>, <*M*, *Accountant*>, <*F*, *Accountant*>, <*M*, *Lawyer*>, <*F*, *Lawyer*>. The example in Figure 1 specifies the k-anonymity requirement on two QIDs. ■

3.2 Generalization and Specialization

To generalize T, a *taxonomy tree* is specified for each categorical attribute in $\cup QID_i$. A leaf node represents a domain value and a parent node represents a less specific value. For a continuous attribute in $\cup QID_i$, a taxonomy tree can be grown at runtime, where each node represents an interval, and each non-leaf node has two child nodes representing some "optimal" binary split of the parent interval. Figure 1 shows a dynamically grown taxonomy tree for *Salary*. More details will be discussed in Section 4.

We generalize a table T by a sequence of specializations starting from the top most value for each attribute. A *specialization*, written $v \rightarrow child(v)$, where $child(v)$ denotes the set of child values of v, replaces the parent value v with the child value that generalizes the domain value in a record. A specialization is *valid* if the specialization results in a table satisfying the anonymity requirement after the specialization. A specialization is *beneficial* if more than one class are involved in the records containing v. A specialization needs to be performed only if it is both valid and beneficial. The specialization process pushes the "cut" of each taxonomy tree downwards. A *cut* of the taxonomy tree for an attribute D_j, denoted Cut_j, contains exactly one value on each root-to-leaf path. A *solution cut* is $\cup Cut_j$, where D_j is an attribute in $\cup QID_i$, such that the generalized T represented by $\cup Cut_j$ satisfies the anonymity requirement. Figure 2 shows a solution cut indicated by the dashed curve.

3.3 Secure Data Integration

We now consider two parties. Party A owns the table $T_A(ID, D_1, \ldots, D_t, Class)$ and B owns the table $T_B(ID, D_{t+1}, \ldots, D_m, Class)$, over the same set of records. These parties agree to release "minimal information" to form an integrated table T (by matching the ID) for conducting a joint classification analysis. The notion of minimal information is specified by the *joint anonymity requirement* $\{<QID_1, k_1>, \ldots, <QID_p, k_p>\}$ on the integrated table. QID_i is *local* if it contains only attributes from one party, and *global* otherwise.

Definition 2 (Secure Data Integration). Given two private tables T_A and T_B, a joint anonymity requirement $\{<QID_1, k_1>, \ldots, <QID_p, k_p>\}$, and a taxonomy tree for each categorical attribute in $\cup QID_i$, the *secure data integration*

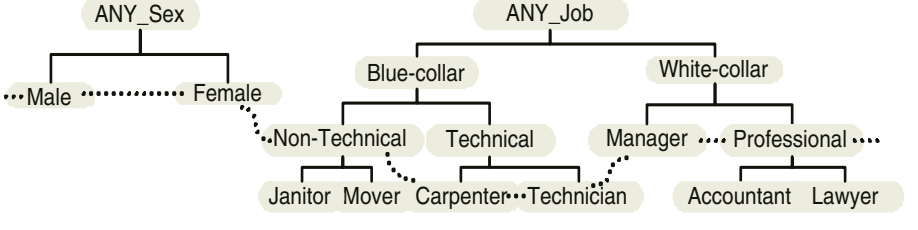


Fig. 2. A solution cut for $QID_1 = \{Sex, Job\}$

is to produce a generalized integrated table T such that (1) T satisfies the joint anonymity requirement, (2) T contains as much information as possible for classification, (3) each party learns nothing about the other party more specific than what is in the final generalized T. ■

For example, if a record in the final T has values F and $Professional$ on Sex and Job, and if Party A learns that $Professional$ in this record comes from $Lawyer$, condition (3) is violated. Our privacy model ensures the anonymity in the final integrated table as well as in any intermediate table.

4 An Unsecured Solution

One unsecured approach is first joining T_A and T_B into a single table T and then generalizing T. Though this approach does not satisfy the requirement (3) in Definition 2 (because the party that generalizes the joint table knows all the details of both T_A and T_B), the integrated table produced satisfies requirements (1) and (2) in Definition 2. Below, we briefly describe this unsecured approach. A secured approach that produces the same integrated table and satisfies the requirement (3) will be presented in Section 5.

A *top-down specialization (TDS)* approach has been proposed in [8] to generalize a single table T. It starts from the top most value for each attribute and iteratively specializes current values until the anonymity requirement is violated. Initially, Cut_i contains the top most value for each attribute D_i. At each iteration, it performs the best specialization w (i.e., of the highest $Score$) among the *candidates* that are valid, beneficial specializations in $\cup Cut_i$, and updates the $Score(x)$ and "valid/benefical" status of x in $\cup Cut_i$ that are affected. The algorithm terminates when there is no more candidate in $\cup Cut_i$.

The core of this approach is computing $Score$, which measures the goodness of a specialization with respect to privacy preservation and information preservation. The effect of a specialization $v \rightarrow child(v)$ can be summarized by "information gain," denoted $InfoGain(v)$, and "anonymity loss," denoted $AnonyLoss(v)$, due to the specialization. *Our selection criterion favors the specialization v that has the maximum information gain per unit of anonymity loss:*

$$Score(v) = \frac{InfoGain(v)}{AnonyLoss(v) + 1}. \quad (1)$$

We add 1 to $AnonyLoss(v)$ to avoid division by zero.

$InfoGain(v)$: Let $T[x]$ denote the set of records in T generalized to the value x. Let $freq(T[x], cls)$ denote the number of records in $T[x]$ having the class cls. Let $|x|$ be the number of elements in a set x. Note that $|T[v]| = \sum_c |T[c]|$, where $c \in child(v)$. We have

$$InfoGain(v) = I(T[v]) - \sum_c \frac{|T[c]|}{|T[v]|} I(T[c]), \quad (2)$$

where $I(T[x])$ is the *entropy* of $T[x]$ [11]:

$$I(T[x]) = -\sum_{cls} \frac{freq(T[x], cls)}{|T[x]|} \times log_2 \frac{freq(T[x], cls)}{|T[x]|}, \quad (3)$$

Intuitively, $I(T[x])$ measures the "mix" of classes for the records in $T[x]$, and $InfoGain(v)$ is the reduction of the mix by specializing v.

$AnonyLoss(v)$: This is the average loss of anonymity by specializing v over all QID_j that contain the attribute of v:

$$AnonyLoss(v) = avg\{A(QID_j) - A_v(QID_j)\}, \quad (4)$$

where $A(QID_j)$ and $A_v(QID_j)$ represents the anonymity before and after specializing v. Note that $AnonyLoss(v)$ not just depends on the attribute of v; it depends on all QID_j that contain the attribute of v.

5 Top-Down Specialization for 2 Parties

Now we consider that the table T is given by two tables T_A and T_B with a common key ID, where Party A holds T_A and Party B holds T_B. At first glance, it seems that the change from one party to two parties is trivial because the change of $Score$ due to specializing a single attribute depends only on that attribute and $Class$, and each party knows about $Class$ and the attributes they have. This observation is wrong because the change of $Score$ involves the change of $A(QID_j)$ that depends on the combination of the attributes in QID_j.

Suppose that, in the TDS approach, each party keeps a copy of the current $\cup Cut_i$ and generalized T, denoted T_g, in addition to the private T_A or T_B. The nature of the top-down approach implies that T_g is more general than the final answer, therefore, does not violate the requirement (3) in Definition 2. At each iteration, the two parties cooperate to perform the same specialization as identified in TDS by communicating certain information in a way that satisfies the requirement (3) in Definition 2. Algorithm 1 describes the procedure at Party A (same for Party B).

First, Party A finds the local best candidate and communicates with Party B to identify the overall winner candidate, say $w \rightarrow child(w)$. To protect the input score, Secure 2-party max [3] can be used. The winner candidate will be the same as identified in TDS because the same selection criterion is used. Suppose that w is local to Party A (otherwise, the discussion below applies to Party B). Party A performs w on its copy of $\cup Cut_i$ and T_g. This means specializing each record $t \in T_g$ containing w into those $t1', \cdots, tk'$ containing child values in $child(w)$, by examining the set of raw records generalized by t, denoted $T_A[t]$, and partitioning $T_A[t]$ among $T_A[t1'], \cdots, T_A[tk']$. Similarly, Party B updates its $\cup Cut_i$ and T_g, and partitions $T_B[t]$ into $T_B[t1'], \cdots, T_B[tk']$. Since Party B does not have the attribute for w, Party A needs to instruct Party B how to partition these records in terms of IDs.

Algorithm 1. TDS2P for Party A

1: initialize T_g to include one record containing top most values;
2: initialize $\cup Cut_i$ to include only top most values;
3: **while** there is some candidate in $\cup Cut_i$ **do**
4: find the local candidate x of highest $Score(x)$;
5: communicate $Score(x)$ with Party B to find the winner;
6: **if** the winner w is local **then**
7: specialize w on T_g;
8: instruct Party B to specialize w;
9: **else**
10: wait for the instruction from Party B;
11: specialize w on T_g using the instruction;
12: **end if**;
13: replace w with $child(w)$ in the local copy of $\cup Cut_i$;
14: update $Score(x)$, the beneficial/valid status for candidates x in $\cup Cut_i$;
15: **end while**;
16: output T_g and $\cup Cut_i$;

Example 3. Consider Table 1 and the joint anonymity requirement:

$\{<QID_1 = \{Sex, Job\}, 4>, \quad <QID_2 = \{Sex, Salary\}, 11>\}$.

Initially,

$T_g = \{<ANY\_Sex,\ ANY\_Job,\ [1\text{-}99)>\}$

and

$\cup Cut_i = \{ANY\_Sex,\ ANY\_Job,\ [1\text{-}99)\}$,

and all specializations in $\cup Cut_i$ are candidates. To find the candidate to specialize, Party A computes $Score(ANY\_Sex)$, and Party B computes $Score(ANY\_Job)$ and $Score([1\text{-}99))$. ■

Below, we describe the key steps: find the winner candidate (Line 4-5), perform the winning specialization (Line 7-11), and update the score and status of candidates (Line 14). For Party A, a *local attribute* refers to an attribute from T_A, and a *local specialization* refers to that of a local attribute.

5.1 Find the Winner Candidate

Party A first finds the local candidate x of highest $Score(x)$, by making use of computed $InfoGain(x)$, $A_x(QID_j)$ and $A(QID_j)$, and then communicates with Party B (using secure 2-party max as in [3]) to find the winner candidate. $InfoGain(x)$, $A_x(QID_j)$ and $A(QID_j)$ come from the update done in the previous iteration or the initialization prior to the first iteration. This step does not access data records. Updating $InfoGain(x)$, $A_x(QID_j)$ and $A(QID_j)$ is considered in Section 5.3.

5.2 Perform the Winner Candidate

Suppose that the winner candidate w is local at Party A (otherwise, replace Party A with Party B). For each record t in T_g containing w, Party A accesses the raw

records in $T_A[t]$ to tell how to specialize t. To facilitate this operation, we represent T_g by the data structure called *Taxonomy Indexed PartitionS (TIPS)*. The idea is to group the raw records in T_A according to their generalized records t in T_g.

Definition 3 (TIPS). TIPS is a tree structure. Each node represents a generalized record over $\cup QID_j$. Each child node represents a specialization of the parent node on exactly one attribute. A leaf node represents a generalized record t in T_g and the *leaf partition* containing the raw records generalized to t, i.e., $T_A[t]$. For a candidate x in $\cup Cut_i$, P_x denotes a leaf partition whose generalized record contains x, and $Link_x$ links up all P_x's. ■

With the TIPS, we can find all raw records generalized to x by following $Link_x$ for a candidate x in $\cup Cut_i$. To ensure that each party has only access to its own raw records, a leaf partition at Party A contains only raw records from T_A and a leaf partition at Party B contains only raw records from T_B. Initially, the TIPS has only the root node representing the most generalized record and all raw records. In each iteration, the two parties cooperate to perform the specialization w by refining the leaf partitions P_w on $Link_w$ in their own TIPS.

Example 4. Continue with Example 3. Initially, TIPS has the root representing the most generalized record <*ANY_Sex, ANY_Job, [1-99)*>, $T_A[root] = T_A$ and $T_B[root] = T_B$. The root is on $Link_{ANY\_Sex}$, $Link_{ANY\_Job}$, and $Link_{[1-99)}$. See the root in Figure 3. The shaded field contains the number of raw records generalized by a node. Suppose that the winning candidate w is *[1-99)* $\rightarrow$ {*[1-37), [37-99)*} on *Salary*.

Party B first creates two child nodes under the root and partitions $T_B[root]$ between them. The root is deleted from all $Link_x$, the child nodes are added to $Link_{[1-37)}$ and $Link_{[37-99)}$, respectively, and both are added to $Link_{ANY\_Job}$ and $Link_{ANY\_Sex}$. Party B then sends the following instruction to Party A:

IDs 1-12 go to the node for *[1-37)*.
IDs 13-34 go to the node for *[37-99)*.

On receiving this instruction, Party A creates the two child nodes under the root in its copy of TIPS and partitions $T_A[root]$ similarly. Suppose that the next winning candidate is $ANY\_Job \rightarrow \{Blue\text{-}collar, White\text{-}collar\}$.

The two parties cooperate to specialize each leaf node on $Link_{ANY\_Job}$ in a similar way, resulting in the TIPS in Figure 4. ■

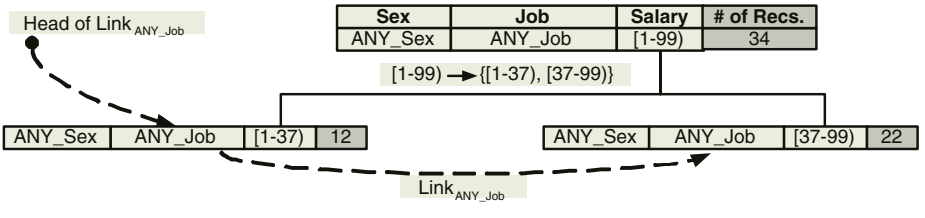


Fig. 3. The TIPS after the first specialization

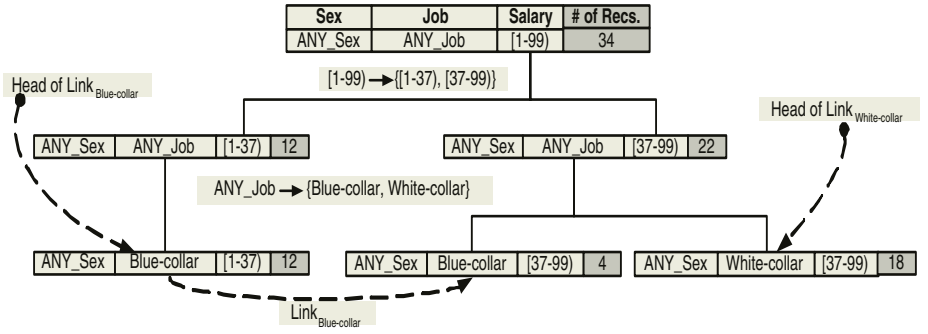


Fig. 4. The TIPS after two specializations

We summarize the operations at the two parties, assuming that the winner w is local at Party A.

Party A**.** Refine each leaf partition P_w on $Link_w$ into child partitions P_c. $Link_c$ is created to link up the new P_c's for the same c. Mark c as "beneficial" if the records on $Link_c$ has more than one class. Also, add P_c to every $Link_x$ other than $Link_w$ to which P_w was previously linked. While scanning the records in P_w, Party A also collects the following information.

- *Instruction for Party B.* If a record in P_w is specialized to a child value c, collect the pair (id,c), where id is the ID of the record. This information will be sent to Party B to refine the corresponding leaf partitions there.
- *Count statistics.* To update $Score$ without accessing raw records, some "count statistics" is maintained for each partition in the TIPS. This is done in the same scan as performing w described above. See the details in [8].

Party B**.** On receiving the instruction from Party A, Party B creates child partitions P_c in its own TIPS. At Party B, P_c's contain raw records from T_B. P_c's are obtained by splitting P_w among P_c's according to the (id, c) pairs received.

We emphasize that updating TIPS is the only operation that accesses raw records. Subsequently, updating $Score(x)$ (in Section 5.3) makes use of the count statistics without accessing raw records anymore.

5.3 Update the Score

$Score(x)$ depends on $InfoGain(x)$, $A_x(QID_j)$ and $A(QID_j)$. The updated $A(QID_j)$ is obtained from $A_w(QID_j)$, where w is the specialization just performed. Below, we consider updating $InfoGain(x)$ and $A_x(QID_j)$ separately.

Updating $InfoGain(x)$**.** $InfoGain(x)$ is affected in that we need to compute $InfoGain(c)$ for newly added c in $child(w)$. The owner party of w can compute $InfoGain(c)$ while collecting the count statistics for c in Section 5.2.

Updating $AnonyLoss(x)$**.** Recall that $A_x(QID_j)$ is the minimum $a(qid_j)$ after specializing x. Therefore, if $att(x)$ and $att(w)$ both occur in some QID_j, the spe-

cialization on w might affect $A_x(QID_j)$, and we need to find the new minimum $a(qid_j)$. The following $QIDTree_j$ data structure indexes $a(qid_j)$ by qid_j.

Definition 4 (QIDTrees). For each $QID_j = \{D_1, \ldots, D_q\}$, $QIDTree_j$ is a tree of q levels, where level $i > 0$ represents generalized values for D_i. A root-to-leaf path represents an existing qid_j on QID_j in the generalized data T_g, with $a(qid_j)$ stored at the leaf node. A branch is trimmed if its $a(qid_j) = 0$. $A(QID_j)$ is the minimum $a(qid_j)$ in $QIDTree_j$. ■

$QIDTree_j$ is kept at a party if the party owns some attributes in QID_j. On specializing the winner w, a party updates its $QIDTree_j$'s that contain the attribute $att(w)$: creates the nodes for the new qid_j's and computes $a(qid_j)$. We can obtain $a(qid_j)$ from the local TIPS: $a(qid_j) = \sum |P_c|$, where P_c is on $Link_c$ and qid_j is the generalized value on QID_j for P_c. $|P_c|$ is part of the count statistics for w collected in Section 5.2.

Example 5. Continue with Example 4. Figure 5 shows the initial $QIDTree_1$ and $QIDTree_2$ for QID_1 and QID_2 on the left. On performing *[1-99)* $\rightarrow$ {*[1-37)*, *[37-99)*}, <*ANY_Sex*, *[1-99)*> in $QIDTree_2$ is replaced with new qids <*ANY_Sex*, *[1-37)*> and <*ANY_Sex*, *[37-99)*>. $A(QID_2) = 12$.

Next, on performing *ANY_Job* $\rightarrow$ {*Blue-collar*, *White-collar*}, <*ANY_Sex, ANY_Job*> in $QIDTree_1$ is replaced with new qids <*ANY_Sex, Blue-collar*> and <*ANY_Sex, White-collar*>. To compute $a(vid)$ for these new qids, we add $|P_{Blue\text{-}collar}|$ on $Link_{Blue\text{-}collar}$ and $|P_{White\text{-}collar}|$ on $Link_{White\text{-}collar}$ (see Figure 4): $a(<ANY\_Sex, Blue\text{-}collar>) = 0 + 12 + 4 = 16$, and $a(<ANY\_Sex, White\text{-}collar>) = 0 + 18 = 18$. So $A_{ANY\_Job}(QID_1) = 16$. ■

Updating $A_x(QID_j)$. This is the same as in [8]. Essentially, it makes use of the count statistics in Section 5.2 to do the update. We omit the details here.

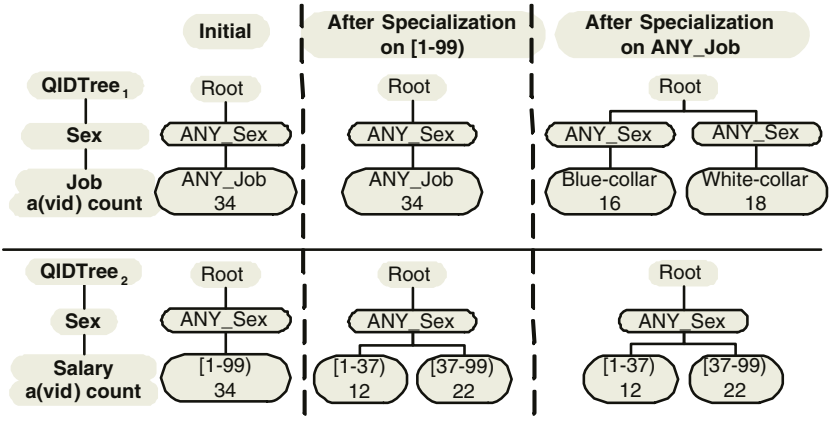


Fig. 5. The QIDTrees data structure

5.4 Analysis

Theorem 1. TDS2P produces exactly the same integrated table as the unsecured TDS on the joint table, and ensures that no party learns more detailed information about the other party other than what they agree to share. ■

This claim follows from the fact that TDS2P performs exactly the same sequence of specializations as in TDS in a distributed manner where T_A and T_B are kept locally at the sources. The only information revealed to each other is those in $\cup Cut_j$ and T_g at each iteration. However, such information is more general than the final integrated table that the two parties agree to share, thanks to the nature of the top-down approach.

We omit the empirical evaluation of the proposed method. Basically, this method produced exactly the same generalized data as in the centralized case where one party holds all attributes of the data (Theorem 1). The latter case has been studied in [8].

References

1. The House of Commons in Canada: The personal information protection and electronic documents act (2000) http://www.privcom.gc.ca/.
2. Agrawal, R., Evfimievski, A., Srikant, R.: Information sharing across private databases. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, San Diego, California (2003)
3. Yao, A.C.: Protocols for secure computations. In: Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science. (1982)
4. Liang, G., Chawathe, S.S.: Privacy-preserving inter-database operations. In: Proceedings of the 2nd Symposium on Intelligence and Security Informatics. (2004)
5. Dalenius, T.: Finding a needle in a haystack - or identifying anonymous census record. Journal of Official Statistics **2** (1986) 329–336
6. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems **10** (2002) 571–588
7. Hundepool, A., Willenborg, L.: μ- and τ-argus: Software for statistical disclosure control. In: Third International Seminar on Statistical Confidentiality, Bled (1996)
8. Fung, B.C.M., Wang, K., Yu, P.S.: Top-down specialization for information and privacy preservation. In: Proceedings of the 21st IEEE International Conference on Data Engineering, Tokyo, Japan (2005)
9. Wang, K., Yu, P., Chakraborty, S.: Bottom-up generalization: a data mining solution to privacy protection. In: Proceedings of the 4th IEEE International Conference on Data Mining. (2004)
10. Iyengar, V.S.: Transforming data to satisfy privacy constraints. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada (2002) 279–288
11. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993)

Applying Authorship Analysis to Arabic Web Content

Ahmed Abbasi and Hsinchun Chen

Department of Management Information Systems, The University of Arizona,
Tucson, AZ 85721, USA
aabbasi@email.arizona.edu, hchen@bpa.arizona.edu

Abstract. The advent and rapid proliferation of internet communication has allowed the realization of numerous security issues. The anonymous nature of online mediums such as email, web sites, and forums provides an attractive communication method for criminal activity. Increased globalization and the boundless nature of the internet have further amplified these concerns due to the addition of a multilingual dimension. The world's social and political climate has caused Arabic to draw a great deal of attention. In this study we apply authorship identification techniques to Arabic web forum messages. Our research uses lexical, syntactic, structural, and content-specific writing style features for authorship identification. We address some of the problematic characteristics of Arabic in route to the development of an Arabic language model that provides a respectable level of classification accuracy for authorship discrimination. We also run experiments to evaluate the effectiveness of different feature types and classification techniques on our dataset.

1 Introduction

Increased online communication has spawned the need for greater analysis of web content. The use of potentially anonymous internet sources such as email, websites, and forums provides an attractive communication medium for criminal activity. The geographically boundless nature of the internet has further amplified these concerns due to the addition of a multilingual dimension. Application of authorship identification techniques across multilingual web content is important due to increased internet communication and globalization, and the ensuing security issues that are created.

Arabic is one of the official languages of the United Nations and is spoken by hundreds of million people. The language is gaining interest due to its socio-political importance and differences from Indo-European languages [10]. The morphological challenges pertaining to Arabic pose several critical problems for authorship identification techniques. These problems could be partially responsible for the lack of previous authorship analysis studies relating to Arabic.

In this paper we apply an existing framework for authorship identification to Arabic web forum messages. Techniques and features are incorporated to address the specific characteristics of Arabic, resulting in the creation of an Arabic language model.

The remainder of this paper is organized as follows: Section 2 surveys relevant authorship analysis studies, highlighting important writing attributes, classification

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 183 – 197, 2005.

techniques, and experiment parameters used in previous studies. Section 3 emphasizes the vital Arabic linguistic characteristics and challenges. Section 4 raises important research questions and describes an experiment designed to address these questions. Section 5 summarizes the experiment results and important findings. We conclude with a summary of key research contributions and point to future directions in Section 6.

2 Related Studies

In this section we briefly discuss previous research relating to the different writing style features, classification techniques, and experimental parameters used in authorship identification.

2.1 Authorship Identification

Authorship analysis is the process of evaluating writing characteristics in order to make inferences about authorship. It is rooted in the linguistic area known as Stylometry, which is defined as the statistical analysis of literary style. There are several categories of authorship analysis; however we are concerned with the branch known as authorship identification.

Authorship identification matches unidentified writings to an author based on writing style similarities between the author's known works and the unidentified piece. Studies can be traced back to the nineteenth century where the frequency of long words was used to differentiate between the works of Shakespeare, Marlowe, and Bacon [23]. Perhaps the most foundational work in the field was conducted by Mosteller and Wallace [24]. They used authorship identification techniques to correctly attribute the twelve disputed Federalist Papers.

Although authorship identification has its roots in historical literature such as Shakespeare and the Federalist papers, it has recently been applied to online material. De Vel et al. conducted a series of experiments on authorship identification of emails [8, 9]. Their studies provided an important foundation for the application of authorship identification techniques to the internet medium. Zheng et al. expanded de Vel et al.'s efforts by adding the multilingual dimension in their study of English and Chinese web forum messages [35]. In this study we are primarily interested in applying authorship identification to Arabic online messages.

2.2 Writing Style Features for Authorship Identification

Writing style features are characteristics that can be extracted from the text in order to facilitate authorship attribution. There are four important categories of features that have been used extensively in authorship identification; lexical, syntactic, structural, and content-specific.

Lexical features are the most traditional set of features used for authorship identification. They have their origins dating back to the nineteenth century, when Mendenhall used the frequency of long words for authorship identification in 1887

[23]. This set of features includes sentence length, vocabulary richness, word length distributions, usage frequency of individual letters, etc. [33, 34, 16].

Syntax is the patterns used for the formation of writing. Word usage (function words) and punctuation are two popular categories of syntactic features. Baayen et al. confirmed the importance of punctuation as an effective discriminator for authorship identification [3]. Mosteller and Wallace were the first to successfully use function words, which are generic words with more universal application (e.g., "while," "upon") [24].

Structural features deal with the organization and layout of the text. This set of features has been shown to be particularly important for online messages. De Vel et al. and Zheng et al. measured the usage of greetings and signatures in email messages as an important discriminator [8, 9, 35].

Content-specific features are words that are important within a specific topic domain. An example of content-specific words for a discussion on computer monitors might be "resolution" and "display." Martindale and McKenzie successfully applied content-specific words for identification of the disputed Federalist Papers [20].

A review of previous authorship analysis literature reveals a couple of important points. Firstly, lexical and syntactic features are the most frequently used categories of features due to their high discriminatory potential. Secondly, there is still a lack of consensus as to the best set of features for authorship identification. In a study done in 1998, Rudman determined that nearly 1,000 features have been used in authorship analysis [28].

2.3 Techniques for Authorship Identification

The two most commonly used analytical techniques for authorship identification are statistical and machine learning approaches. Several multivariate statistical approaches have been successfully applied in recent years. Burrows was the first to incorporate principle component analysis in 1987, which became popular due to its high discriminatory power [5]. Other successful multivariate methods include cluster analysis and discriminant analysis [15, 19].

Many potent machine learning techniques have been realized in recent years due to drastic increases in computational power. Tweedie et al. and Lowe and Mathews used neural networks whereas Diedrich et al. and de Vel et al. successfully applied SVM for authorship identification [30, 20, 8, 9]. Zheng et al. conducted a thorough study involving machine learning techniques in which they used decision trees, neural networks, and SVM [35].

Typically machine learning methods have achieved superior results as compared to statistical techniques. Machine learning approaches also benefit from less stringent requirements pertaining to models and assumptions. Most importantly, machine learning techniques are more tolerant to noise and can deal with a larger number of features [22].

2.4 Multilingual Authorship Identification

Applying authorship identification across different languages is becoming increasingly important due to the proliferation of the internet. Nevertheless, there has been a

lack of studies focusing across different languages. Most previous studies have only focused on English, Greek, and Chinese. Stamamatos et al. applied authorship identification to a corpus of Greek newspaper articles [51]. Peng et al. conducted experiments on English documents, Chinese novels, and Greek newspapers [47]. Zheng et al. performed authorship identification on English and Chinese web forum messages [63]. In all previous studies, English results were better than other languages.

Applying authorship identification features across different languages is not without its difficulties. Since most writing style characteristics were designed for English, they may not always be applicable or relevant for other languages. Morphological and other linguistic differences can create feature extraction implementation difficulties. For example, Peng et al. noted that the lack of word segmentation in Chinese makes word-based lexical features (such as the number of words in a sentence) too difficult to extract [47]. They also found that the larger volume of words in Chinese makes vocabulary richness measures less effective.

2.5 Authorship Identification of Online Messages

Online messages present several problems for authorship identification as compared to conventional forms of writing (literary works, published articles). Perhaps the biggest concern is the shorter length of online messages. Ledger and Merriam claimed that authorship characteristics were less apparent below 500 words, while Forsyth and Holmes placed that number at 250 words [19, 13]. This problem is further amplified by the fact that online messages typically have a larger pool of potential authors to distinguish between. Most previous authorship identification studies performed on conventional writing involved 2-3 authors, with almost no studies exceeding 10 authors.

The less formal style of online messages can cause problems since there is a greater likelihood of misspelled words, use of abbreviations and acronyms (e.g., "j/k"), and unorthodox use of punctuations (e.g., ":)"). These differences can lead to inaccurate feature extraction. Such problems are more prevalent in online messages because authors are less likely to follow formal writing rules.

Despite all the challenges, the unique style of online messages may also provide helpful discriminators that are useful for identification. Structural features such as greetings, signatures, quotes, links, and use of phone numbers and email addresses as contact information can provide significant insight into an author's writing characteristics.

3 Arabic Characteristics

Arabic is a Semitic language, meaning that it belongs to the group of Afro-Asian languages which also includes Hebrew. It is written from right to left with letters being joined together, similar to English cursive writing. Semitic languages have several characteristics that can cause problems for authorship analysis. These challenges include properties such as inflection, diacritics, word length, and elongation.

3.1 Inflection

Inflection is the derivation of stem words from a root. Although the root has a meaning, it is not a word but rather a class that contains stem instances (words). Stems are created by adding affixes (prefixes, infixes, and suffixes) to the root using specific patterns. Words with common roots are semantically related. Arabic roots are 3-5 letter consonant combinations with the majority being 3-letters. Al-Fedaghi and Al-Anzi believe that as many as 85% of Arabic words are derived from a tri-lateral root, suggesting that Arabic is highly inflectional [2]. Beesley estimated that there are approximately 5,000 roots in Arabic [4].

Figure 1 shows an inflection example. For the root and stems, the top row shows the word written using English alphabet characters and the second row shows the word written in Arabic. The words KTAB ("book") and MKTB ("office/desk") are derived from the root KTB. KT**A**B is created with the addition of the infix "A" whereas **M**KTB is derived with the addition of the prefix "M".

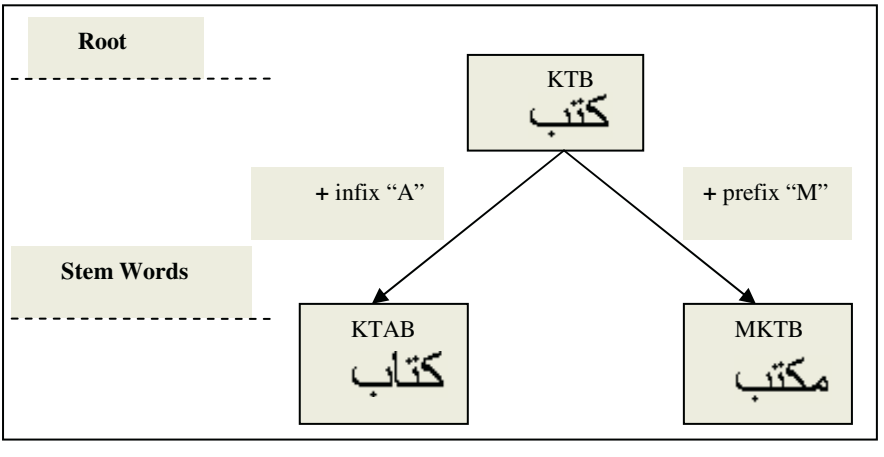


Fig. 1. Inflection Example

Larkey et al. stated that the orthographical and morphological properties of Arabic result in a great deal of lexical variation [18]. Inflection can cause feature extraction problems for lexical features because high levels of inflection increase the number of possible words, since a word can take on numerous forms. A larger word pool results in problems similar to those observed by Zheng et al. regarding Chinese [35]. In these circumstances vocabulary richness based measures such as Hapax, Lapax, Yule's, Honore's etc. will not be as effective.

3.2 Diacritics

Diacritics are markings above or below letters, used to indicate special phonetic values. An example of diacritics in English would be the little markings found on top of the letter "e" in the word résumé. These markings alter the pronunciation and

meaning of the word. Arabic uses diacritics in every word to represent short vowels, consonant lengths, and relationships between words.

Although diacritics are an integral part of Arabic, they are rarely used in writing. Without diacritics, readers can use the sentence semantics to interpret the correct meaning of the word. For example, the words "resume" and "résumé" would look identical without diacritics, however a reader can figure out the appropriate word based on the context. Unfortunately this is not possible for computer feature extraction programs. This, coupled with the fact that Arabic uses diacritics in every word, poses major problems as illustrated by Figure 2.

English	Definitions	With Diacritics	Without Diacritics
MIN	who, whoever	مِنْ	منْ
MUN	from, of, for, than	مَنْ	منْ

Fig. 2. Diacritics Example

The Arabic function words MIN and MUN are written using identical letters, but with differing diacritics. MIN uses a short vowel represented by a marking underneath and "MUN" has one above. Without diacritics, the two words look identical and are indistinguishable for machines. This reduces the effectiveness of word-based syntactic features, specifically function words. From a feature extraction perspective, it is impossible to differentiate between the function word "who" (MIN) and "from" (MUN).

3.3 Word Length

Arabic words tend to be shorter than English words. The shorter length of Arabic words reduces the effectiveness of many lexical features. The short-word count feature; used to track words of length 3-letters or smaller, may have little discriminatory potential when applied to Arabic. Additionally, the word-length distribution may also be less effective since Arabic word length distributions have a smaller range.

3.4 Elongation

Arabic words are sometimes stretched out or elongated. This is done for purely stylistic reasons using a special Arabic character that resembles a dash ("-"). Elongation is possible because Arabic characters are joined during writing. Figure 3 shows an example of elongation. The word MZKR ("remind") is elongated with the addition of four dashes between the "M" and the "Z" (denoted by a faint oval).

Although elongation provides an important authorship identification feature it can also create problems. Elongation can impact the accuracy of word length features. The example in Figure 3 causes the length of the word MZKR to double when elongated by fours dashes. How to handle elongation in terms of feature extraction is an important issue that must be resolved.

Elongated	English	Arabic	Word Length
No	MZKR	مذكر	4
Yes	M----ZKR	مـــذكر	8

Fig. 3. Elongation Example

4 Experiment

In this section we raise important research questions and discuss an experiment designed to address these questions. Specifically, we talk about the test bed, techniques, parameters, feature set, and experimental design used.

4.1 Research Questions

1. Will authorship analysis techniques be applicable in identifying authors in Arabic?
2. What are the effects of using different types of features in identifying authors?
3. Which classification techniques are appropriate for Arabic authorship analysis?

4.2 Experiment Test Bed

Our test bed consists of an Arabic dataset extracted from Yahoo groups. This dataset is composed of 20 authors and 20 messages per author. These authors discuss political ideologies and social issues in the Arab world.

4.3 Experiment Techniques

Based on previous studies, there are numerous classification techniques that can provide adequate performance. In this research, we adopted two machine learning classifiers; ID3 decision trees and Support Vector Machine.

ID3 is a decision tree building algorithm developed by Quinlan that uses a divide-and-conquer strategy based on an entropy measure for classification [27]. ID3 has been tested extensively and shown to rival other machine learning techniques in predictive power [6, 12]. Whereas the original algorithm was designed to deal with discrete values, the C4.5 algorithm extended ID3 to handle continuous values. Support Vector Machine (SVM) was developed by Vapnik on the premise of the Structural Risk Minimization principle derived from computational learning theory [32]. It has been used extensively in previous authorship identification studies [11, 35].

Both these techniques have been previously applied to authorship identification, with SVM typically outperforming ID3 [11, 35]. In this study we incorporated SVM for its classification power and robustness. SVM is able to handle hundreds and thousands of input values with great ease due to its ability to deal well with noisy data. ID3 was used for its efficiency. It is able to build classification models in a fraction of the time required by SVM.

4.4 Arabic Feature Set Issues

Before creating a feature set for Arabic we must address the challenges created by the characteristics of the language. In order to overcome the diacritics problem, we would need to embed a semantic-based engine into our feature extraction program. Since no such programs currently exist due to the arduous nature of the task, overcoming the lack of diacritics in the data set is not feasible. Thus, we will focus on the other challenges, specifically inflection, word length, and elongation.

4.4.1 Inflection

We decided to supplement our feature set by tracking usage frequencies of a select set of word roots. In addition to the inflection problem which impacts vocabulary richness measures, this will also help compensate for the loss in effectiveness of function words due to a lack of diacritics. Word roots have been shown to provide superior performance than normal Arabic words in information retrieval studies. Hmeidi et al. found that root indexing outperformed word indexing on both precision and recall in Arabic information retrieval [14]. Tracking root frequencies requires matching words to their appropriate roots. This can be accomplished using a similarity score based clustering algorithm.

4.4.1.1 Clustering Algorithm

Most clustering algorithms are intended to group words based on their similarities, rather than compare words to roots. These algorithms consist of several steps and some sort of equation used to evaluate similarity. The additional steps are necessary since word clustering is more challenging than word-root comparisons. Since we are comparing words against a list of roots, our primary concern is the use of a similarity-score based equation, and not necessarily all other parts of the algorithm. Two popular equations are Dice's equation and Jaccard's formula [1, 31]. Both these equations use n-grams to calculate the similarity score with the difference being that one places greater emphasis on shared n-grams that the other. The formulas are shown below:

SC(Dice) = 2 * (shared unique n-grams)/(sum of unique n-grams)

SC(Jac) = shared unique n-grams/(sum of unique n-grams – shared unique n-grams)

Although Dice's and Jaccard's equations are effective for English, they need to be supported by algorithms designed according to the characteristics of Arabic. De Roeck and Fares created a clustering algorithm based on Jaccard's formula, specifically designed for Arabic [7]. Their equation uses bi-grams, since they determined that bi-grams outperform other n-grams. The algorithm consists of five steps; however two steps require manual inspection and are not necessary for our purposes. Thus we omitted these parts and focused on Cross, Blank insertion, and applying Jaccard's formula.

Cross gives consonant letters greater weight by creating an additional bi-gram of the letter preceding and following a vowel. For example, in the word KTAB the letters before and after the vowel "A" form an additional bi-gram "TB", as shown in

Table 1. Word consonants are emphasized since roots are mostly composed of consonants, and giving consonants additional weight improves accuracy.

Table 1. Cross Example

Word	Without Cross	With Cross
KTAB	KT TA AB	KT TA TB AB

Blank insertion refers to the insertion of a blank character in the beginning and end of a word in order to give equal weight to border letters. For example, in the word KTAB without blank insertion the first and last letters are only used in a single bi-gram whereas the inner letters appear in 2 bi-grams each. Blank insertion allows all letters to be represented in an equal number of bi-grams, improving similarity score accuracy.

Table 2. Blank Insertion Example

Word	Without Blanks	With Blanks
KTAB	KT TA AB	*K KT TA AB B*

Table 3 shows a full example of our adaptation of De Roeck and Fares' clustering algorithm, using cross and blank insertion and applying Jaccard's formula. In this example, we compare the word KTAB against the roots KTB and KSB. The word is derived from the root KTB, and thus, it should receive higher similarity scores in comparisons with KTB as compared to KSB. In the comparison KTB scores 0.667 whereas KSB only gets a score of 0.25.

Table 3. Word-Root Comparison Example

Comparison	Total Unique bi-grams	Shared Unique bi-grams	Formula	SC
KSB KTAB	*K KS SB B* *K KA AT KT TB B*	*K B*	2/(10-2)	**0.25**
KTB KTAB	*K KT TB B* *K KA AT KT TB B*	*K KT B* KT	4/(10-4)	**0.67**

4.4.1.2 Applying the Word Root Feature

Root frequencies were extracted by calculating similarity scores for each word against a dictionary containing over 4,500 roots. The word was assigned to the root with the highest similarity score and the usage frequency of this root was incremented. Roots were sorted based on variance across authors. This was done based on the rationale that the roots with greater variance provide higher discriminatory potential. In order to determine the number of roots to include in the feature set, classification accuracy was used as the criteria. Between 0 and 50 roots were added to the complete Arabic feature set in order to determine the ideal quantity, which was tested using SVM as

the classifier. Such a trial-and-error approach had to be used due to the lack of previous studies relating to Arabic authorship identification. Stamamatos et al. used a similar method to determine the ideal number of functions words to include in their study relating to Greek [29]. Figure 4 shows that the optimal number of roots using SVM was found to be 30.

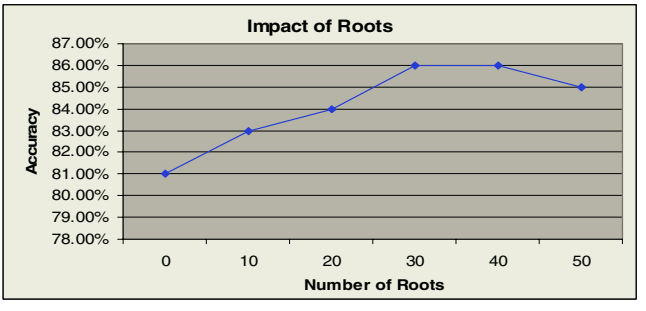


Fig. 4. Impact of Roots on Classification Accuracy

4.4.2 Word Length

Arabic words tend to be shorter than English words. Very long words are almost non-existent in Arabic. In order to test this hypothesis, we extracted the average length of our dataset and the number of words of length greater than 10. These values were then compared against the English dataset used by Zheng et al. which contained an equal number of web forum messages relating to computer software sales [35]. Table 4 shows that English words are approximately half a letter longer on average. More importantly, the number of English words longer than 10 letters is far greater. This disparity must be accounted for by tracking a smaller range for the Arabic word length distribution feature.

Table 4. English and Arabic Word Length Statistics

Data Set	Average Length	% Length > 10
English	5.17	6.125
Arabic	4.61	0.358

4.4.3 Elongation

The number of elongated words and the frequency of usage of elongation dashes should both be tracked since they represent important stylistic characteristics. Additionally, in order to keep word length features accurate, elongation dashes should not be included in word length measurements.

4.5 Arabic Feature Set

The Arabic feature set was modeled after the one used by Zheng et al. [35]. The feature set is composed of 410 features, including 78 lexical features, 292 syntactic

features, 14 structural features and 11 content specific features, as shown in Table 5. In order to compensate for the lack of diacritics and inflection, a larger number of function words and 30 word roots were used. A smaller word length distribution and short word count threshold were also included. A larger set of content-specific words was incorporated due to the more general nature of the Arabic topics of discussion.

4.6 Experiment Procedure

Four feature sets were created. The first feature set contained only lexical features. The second set contained lexical and syntactic features. Structural features were added to the lexical and syntactic features in the third set. The fourth set consisted of all features (lexical, syntactic, structural, and content-specific). Such a step-wise addition of features was used for intuitive reasons. Past research has shown that lexical and syntactic features are the most important and hence, form the foundation for structural and content-specific features. In each experiment five authors were randomly selected and all 20 messages per author were tested using 30-fold cross validation with C4.5 and SVM.

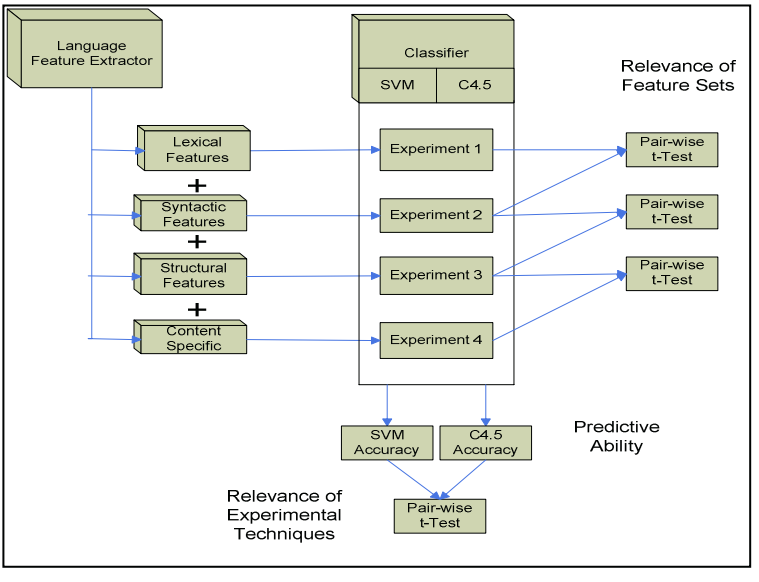


Fig. 5. Experiment Procedure

5 Results and Discussion

The results for the comparison of the different feature types and techniques are summarized in Table 6 and Figure 6. The accuracy kept increasing with the addition of more feature types. The maximum accuracy was achieved with the use of SVM and all feature types.

Table 6. Accuracy for Different Feature Sets across Techniques

	C4.5	SVM
F1	68.07%	74.20%
F1+F2	73.77%	77.53%
F1+F2+F3	76.23%	84.87%
F1+F2+F3+F4	81.03%	**85.43%**

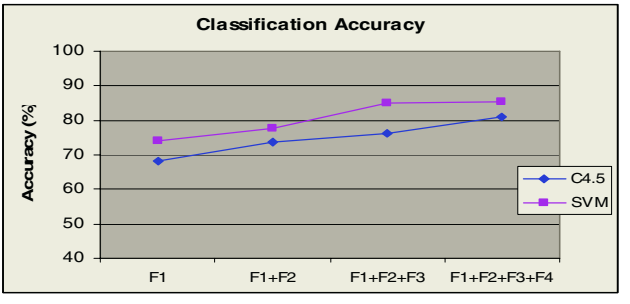


Fig. 6. Authorship Identification Accuracies for Different Feature Types

5.1 Comparison of Feature Types

Pairwise t-tests were conducted to show the significance of the additional feature types added. The results are shown in Table 7 below. An analysis of the t-tests and accuracies shows that all feature types significantly improved accuracy for the Arabic messages. Thus the impact of the different features types for Arabic was consistent with previous results found by Zheng et al. concerning online English and Chinese messages. The least effective set of features for Arabic was the content-specific words, however even this feature type had a significant impact with alpha set at 0.05. The lesser effectiveness of content-specific words could be attributed to the broad topic scope of the Arabic dataset.

Table 7. P-values of pairwise t-tests on accuracy using different feature types

t-Test Results with N=30		
Features/Techniques	C4.5	SVM
F1 vs F1+F2	**0.000*** **	**0.000*** **
F1+F2 vs F1+F2+F3	**0.000*** **	**0.000*** **
F1+F2+F3 vs F1+F2+F3+F4	**0.000*** **	**0.0134** **

**: significant with alpha = 0.05
***: significant with alpha = 0.01

5.2 Comparison of Classification Techniques

A comparison of C4.5 and SVM revealed that SVM significantly outperformed the decision tree classifier in all cases. This is consistent with previous studies that also

showed SVM to be superior [11, 35]. The difference between the two classifiers was consistent with the addition of feature types, with SVM outperforming C4.5 by 4%-8%.

Table 8. P-values of pairwise t-tests on accuracy using different classifier techniques

t-Test Results with N=30				
Technique/Features	F1	F1+F2	F1+F2+F3	F1+F2+F3+F4
C4.5 vs SVM	**0.000*** **	**0.000*** **	**0.000*** **	**0.000*** **

***: significant with alpha = 0.01

6 Conclusion and Future Directions

In this research we applied authorship identification techniques for the classification of Arabic web forum messages. In order to accomplish this we used techniques and features to overcome the challenges created by the morphological characteristics of Arabic. All feature types (lexical, syntactic, structural, and content-specific) provided significant discriminating power for Arabic, resulting in respectable classification accuracy. SVM outperformed C4.5 and the overall accuracy for Arabic was lower than previous English performance, both results being consistent with previous studies.

In the future we would like to analyze the differences between the English and Arabic language models using by evaluating the key features, as determined by decision trees. Emphasizing the linguistic differences between Arabic and English could provide further insight into possible methods for improving the performance of authorship identification methodologies in an online, multilingual setting.

Acknowledgements

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.

References

1. Adamson, George W. and J. Boreham (1974) The use of an association measure based on character structure to identify semantically related pairs of words and document titles. Information Storage and Retrieval,. Vol 10, pp 253-260
2. Al-Fedaghi Sabah S. and Fawaz Al-Anzi (1989) A new algorithm to generate Arabic root-pattern forms. Proceedings of the 11th National Computer Conference, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia., pp04-07.
3. Baayen, H., Halteren, H. v., Neijt, A., & Tweedie, F. (2002). An experiment in authorship attribution. Paper presented at the In Proceedings of the 6th International Conference on the Statistical Analysis of Textual Data (JADT 2002).

4. Beesley, K.B. (1996) Arabic Finite-State Morphological Analysis and Generation. Proceedings of COLING-96, pp 89-94.
5. Burrows, J. F. (1987). Word patterns and story shapes: the statistical analysis of narrative style. Literary and Linguistic Computing, 2, 61 -67.
6. Chen, H., Shankaranarayanan, G., Iyer, A., & She, L. (1998). A machine learning approach to inductive query by examples: an experiment using relevance feedback, ID3, Genetic Algorithms, and Simulated Annealing. Journal of the American Society for Information Science, 49(8), 693-705.
7. De Roeck, A. N. and Al-Fares, W. (2000) A morphologically sensitive clustering algorithm for identifying Arabic roots. In Proceedings ACL-2000. Hong Kong, 2000.
8. De Vel, O. (2000). Mining E-mail authorship. Paper presented at the Workshop on Text Mining, ACM International Conference on Knowledge Discovery and Data Mining (KDD'2000).
9. De Vel, O., Anderson, A., Corney, M., & Mohay, G. (2001). Mining E-mail content for author identification forensics. SIGMOD Record, 30(4), 55-64.
10. Diab, Mona**,** Kadri Hacioglu and Daniel Jurafsky. Automatic Tagging of Arabic Text: From raw text to Base Phrase Chunks. Proceedings of HLT-NAACL 2004
11. Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2000). Authorship attribution with Support Vector Machines. Applied Intelligence.
12. Dietterich, T.G., Hild, H., & Bakiri, G., (1990), A comparative study of ID3 and Backpropagation for English Text-to-Speech mapping, Machine Learning, 24-31.
13. Forsyth, R. S., & Holmes, D. I. (1996). Feature finding for text classification. Literary and Linguistic Computing, 11(4).
14. Hmeidi, I., Kanaan, G. and M. Evens (1997) Design and Implementation of Automatic Indexing for Information Retrieval with Arabic Documents. Journal of the American Society for Information Science, 48/10, pp 867-881.
15. Holmes, D. I. (1992). A stylometric analysis of Mormon Scripture and related texts. Journal of Royal Statistical Society, 155, 91-120.
16. Holmes, D. I. (1998). The evolution of stylometry in humanities. Literary and Linguistic Computing, 13(3), 111-117.
17. Hoorn, J. F., Frank, S. L., Kowalczyk, W., & Ham, F. V. D. (1999). Neural network identification of poets using letter sequences. Literary and Linguistic Computing, 14(3), 311-338.
18. Larkey, L. S. and Connell, M. E. Arabic information retrieval at UMass in TREC-10. In *TREC 2001*. Gaithersburg: NIST, 2001.
19. Ledger, G. R., & Merriam, T. V. N. (1994). Shakespeare, Fletcher, and the two Noble Kinsmen. Literary and Linguistic Computing, 9, 235-248.
20. Lowe, D., & Matthews, R. (1995). Shakespeare vs. Fletcher: a stylometric analysis by radial basis functions. Computers and the Humanities, 29, 449-461.
21. Martindale, C., & McKenzie, D. (1995). On the utility of content analysis in author attribution: The Federalist. Computer and the Humanities, 29(259-270).
22. Mealand, D. L. (1995). Correspondence analysis of Luke. Literary and Linguistic Computing, 10(171-182).
23. Mendenhall, T. C. (1887). The characteristic curves of composition. Science, 11(11), 237-249.
24. Mosteller, F., Frederick, & Wallace, D. L. (1964). Applied Bayesian and classical inference: the case of the Federalist papers (2 ed.): Springer-Verlag.
25. Mosteller, F., & Wallace, D. L. (1964). Inference and disputed authorship: the Federalist: Addison-Wesley.

26. Peng, F., Schuurmans, D., Keselj, V., & Wang, S. (2003). Automated authorship attribution with character level language models. Paper presented at the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2003).
27. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81-106.
28. Rudman, J. (1998). The state of authorship attribution studies: some problems and solutions. Computers and the Humanities, 31, 351-365.
29. Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2001). Computer-based authorship attribution without lexical measures. Computers and the Humanities, 35(2), 193-214.
30. Tweedie, F. J., Singh, S., & Holmes, D. I. (1996). Neural Network applications in stylometry: the Federalist papers. Computers and the Humanities, 30(1), 1-10.
31. Van Rijsbergen, C. J. Information retrieval. London: Butterworths, 1979.
32. Vapnik, V. (1995). The nature of statistical learning theory. New York: Springer Verlag.
33. Yule, G. U. (1938). On sentence length as a statistical characteristic of style in prose. Biometrika, 30.
34. Yule, G. U. (1944). The statistical study of literary vocabulary. Cambridge University Press.
35. Zheng, R., Qin, Y., Huang, Z., & Chen, H. (2003). Authorship Analysis in Cybercrime Investigation. Paper presented at the In Proceedings of the first NSF/NIJ Symposium, ISI2003, Tucson, AZ, USA.

Automatic Extraction of Deceptive Behavioral Cues from Video[1]

Thomas O. Meservy, Matthew L. Jensen, John Kruse,
Judee K. Burgoon, and Jay F. Nunamaker

Center for the Management of Information
University of Arizona, McClelland Hall Rm #114
1130 East Helen Street Tucson, AZ 85721-0108
{tmeservy, mjensen, jkruse, jburgoon,
jnunamaker}@cmi.arizona.edu
http://www.cmi.arizona.edu

Abstract. This research initiative is an initial investigation into a novel approach for deriving indicators of deception from video-taped interaction. The team utilized two-dimensional spatial inputs extracted from video to construct a set of discrete and inter-relational features. The features for thirty-eight video interactions were then analyzed using discriminant analysis. Additionally, features were used to build a multivariate regression model. Through this exploratory study, the team established the validity of the approach, and identified a number of promising features, opening the door for further investigation.

1 Introduction

Deception and its detection have been a source of fascination for centuries, with literally thousands of publications on the topic testifying to its importance in the conduct of human affairs. Researchers and practitioners have pursued a host of detection techniques, from divining trustworthiness from the shape of one's head and ears, to the use of physiologically-based instruments such as the polygraph and vocal stress analyzer, to reliance on behavioral cues as potentially telltale cues to deception. However, no single cue has proven to be an accurate indicator.

All these efforts notwithstanding, deception detection accuracy has typically hovered around 50-50, or chance, even among trained professionals. The most promising avenues for distinguishing deceit from truth lie in tools and techniques that utilize constellations of cues. Moreover, because deception indicators are subtle, dynamic, and transitory, they often elude humans' conscious awareness. If computer-assisted detection tools can be developed to augment human detection capabilities by discerning micro-momentary features of behavior and by tracking these highly elusive and

[1] Portions of this research were supported by funding from the U. S. Department of Homeland Security (Cooperative Agreement N66001-01-X-6042). The views, opinions, and/or findings in this report are those of the authors and should not be construed as official Department of Homeland Security positions, policies, or decisions.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 198 – 208, 2005.

fleeting cues over a course of time, accuracy in discerning both truthful and deceptive information and communications should be vastly improved.

This paper outlines initial attempts at identifying and validating a set of behavioral indicators extracted from video.

2 Literature Review

Although the problem of deception detection is extremely difficult, many attempts have been made to address this issue. Some of these attempts are described below. For ease of comparison, these methods have been classified as physiological methods, verbal methods, and nonverbal methods.

2.1 Physiological Methods

Perhaps out of all the approaches to deception detection, the most well-known is the polygraph or "lie-detector." The polygraph relies on heart, electro-dermal, and respiratory measures to infer deception. It is believed that such physiological measures are directly linked to conditions brought on by deception [1]. There are two main methods of deception detection which use the polygraph: the Control Question Test (CQT) and the Guilty Knowledge Test (GKT). The CQT uses a series of irrelevant control questions for comparison to crime-specific questions to ferret out possible deception; however, it has often been criticized as subjective, non-scientific, and unreliable [2]. The GKT determines whether an interviewee has knowledge about a crime that would only be known by the perpetrator. A series of crime-related objects or pictures may be shown to the interviewee and the interviewee's reaction is recorded. The GKT enjoys a more objective, scientific footing [2], however specific and confidential details about a crime must be obtained for its use.

Another rising method of deception detection is the analysis of brain activity. Improvements in functional magnetic resonance imaging (fMRI) have allowed the monitoring of brain activity during an interview. Some researchers have noticed differences between the brain activity of truth-tellers and deceivers [3, 4]. Currently the reliability and accuracy of deception detection based on brain activity is being debated.

Physiological methods of deception detection require the use of invasive sensors attached to the body. Thus, cooperation from the interviewee is required.

2.2 Verbal Methods

Two methods of verbal analysis 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' [1]. CBCA takes place during a structured interview where the interviewer scores responses according to predefined criteria. RM also uses a scoring mechanism to judge potential deception, however, it is based on the concept that truthful responses will contain more perceptual, contextual, and affective information than deceptive responses.

CBCA and RM require trained interviewers to conduct interviews. Although these verbal analysis methods offer more flexibility than the physiological methods, they require carefully trained interviewers and they do not provide immediate feedback.

2.3 Nonverbal Methods

Computerized voice stress analysis (CVSA) has been proposed as a method to automatically detect deception. Signals that accompany psychological stress are identified from multiple cues in the voice. Accuracy of CVSA is comparable to the polygraph [5], however, in some cases CVSA may require a controlled environment with little background noise.

Observation of behavioral cues is also used as a method of deception detection. Numerous studies have shown that deceivers act differently than truth-tellers [6-8]. Differences include lack of head movement [9] and lack of illustrating gestures which accompany speech [10]. Many people suspect 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 [1]. Additionally, the cognitive load of simultaneously tracking behavior and maintaining an interview can be too heavy for a single interviewer to manage.

2.3 Theory for Behavioral Analysis

To avoid the problems associated with the current methods for identifying deceit, we propose a new method of deception detection which builds upon current nonverbal methods. This method involves automated extraction and identification of behavioral cues which are associated with deception.

Interpersonal Deception Theory (IDT) [11] provides the foundation for our method of deception detection. IDT models deception as a strategic interaction between participants. Both the deceiver and the receiver approach an interaction with a number of preconceived factors that will influence the interaction. These factors might include expectations, goals, familiarity, suspicion, etc. During the course of the interaction, the deceiver and receiver may alter strategies as their effectiveness is observed. During the interaction both parties, will likely unintentionally reveal behavioral cues which signal their suspicion and deception. The intention of our method is to automatically identify these indicators of deception. Our method focuses on deriving cues from the head and hands since these areas are a proven source of reliable indicators for deception.

3 Research Approach

The research approach that we use to categorize nonverbal behavior into deceptive/truthful classes is commonly used in pattern classification— (1) raw data is split into discrete units, (2) general metrics are extracted from the discrete units, (3) features are extracted and inferred from these metrics, and (4) selected features are used to classify each unit. The input to this process is nonverbal behavior captured as video frames and the output is a classification or level of deception or truth. This

exploratory approach utilizes a number of novel and innovative methods during each step.

3.1 Identification of Head and Hands in Video

First, video streams are segmented into discrete units. In an interview setting these units would typically be responses by the interviewee to a specific question or topic.

General metrics are extracted from the video using a method called "blob analysis." We utilize a refined method of blob analysis developed by the Computational Biomedicine Imaging and Modeling Center (CBIM) at Rutgers University [12]. This method uses color analysis, eigenspace-based shape segmentation and Kalman filters to track head and hand positions throughout the video segment. Lu et al [13] explain this process in detail.

Metrics—including position, size, and angles—are produced by the software and are utilized when generating meaningful features. Fig. 1 shows a video frame that has been subjected to blob analysis.

Fig. 1. Video frame after blob analysis

3.2 Feature Extraction

A number of features can be extracted or inferred from the relatively simple metrics produced in the previous phase. Many of these features attempt to capture some of the key elements that interviewers would look for when trying to detect deception (e.g. behavioral over control). However, a number of additional features that typically aren't used by humans (e.g. distance of movement per frame) are extracted because they are potentially useful in an automated environment.

The kinesics-based deceptive features that are extracted from video can be subdivided into single frame features and multi-frame features. Single frame features, as the name implies, are calculated using information from a single video frame. These features can be further categorized as descriptive, relational, or multi-relational features. Multi-frame features require information from two or more frames. Multi-frame features can be classified as descriptive, angular movement, or directional features. Fig. 2 illustrates a taxonomy for these kinesics-based features.

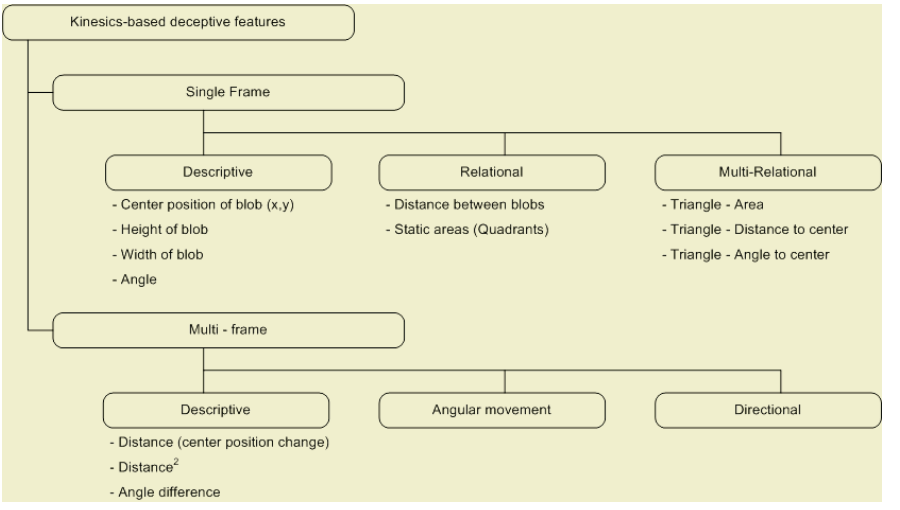


Fig. 2. Taxonomy for kinesics-based deceptive features in video data

Single Frame Features

The features in the descriptive category of single frame features are actually provided as the output of blob analysis. These features include the center position of each blob (x, y), the height (h) and width (w) of each blob, and the angle of the major axis (θ). Fig. 3 illustrates these features as they relate to a single blob.

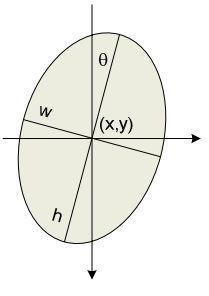


Fig. 3. Single frame descriptive features

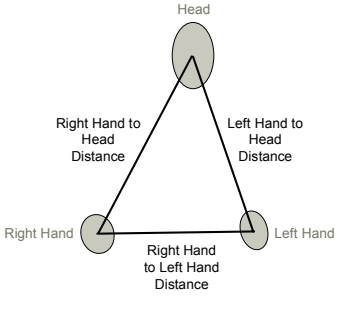


Fig. 4. Blob distance features

The relational category of single frame features contains features that represent relationships between two objects—either a blob to another blob or a blob to a reference point or area. In the former case, the distance between two blobs of interest is calculated using a simple Euclidean distance formula. Fig. 4 demonstrates some single frame distance features. The distance between the head, the right hand, and the left hand allows us to know when they are touching (distance is zero) or how far apart they are. This may hint at gestures that may indicate nervousness (such as preening, scratching, rubbing, etc), but only between the objects identified.

Features that track a blob in comparison to other regions are also part of the relational category. Quadrants derived from the head position and head width have been used in other experiments [14]. We extract quadrant information in a similar manner, creating one region for the area above the shoulders (quadrant 1) and three areas below (quadrants 2-4). Quadrant 3 is derived based on the width of the head. Quadrants 2 and 4 occupy any additional area to the left and right, respectively. Fig. 5 maps the quadrants onto the sample video frame. This feature allows us to understand how often each blob is in each quadrant. It was hypothesized that the hands of deceivers, who are more closed in their posture, would spend more time in quadrant 3 than truth tellers [14].

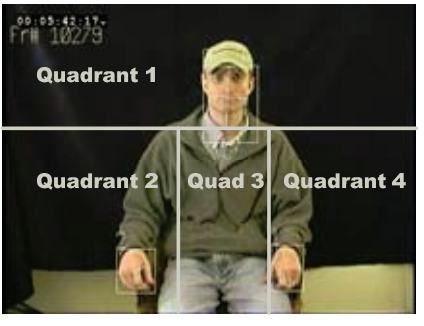


Fig. 5. Quadrant features

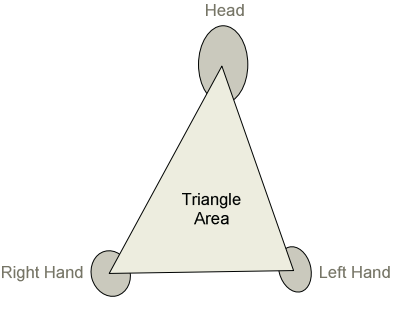


Fig. 6. Triangle Area Feature

The multi-relational category contains features that involve information from 3 or more blobs or objects. The features of interest in our approach include calculating the area of the triangle between the 3 blobs, the distance from each blob to the center point of the triangle, and the angle of each blob in relation to the center of the triangle.

The area of the triangle is calculated using the center points from the head and hands using Formula 1.

$$\frac{\left|\left(\left(x_{head} * y_{left} - x_{left} * y_{head}\right) + \left(x_{right} * y_{head} - x_{head} * y_{right}\right) + \left(x_{left} * y_{right} - x_{right} * y_{left}\right)\right)\right|}{2} \quad (1)$$

Fig. 6 depicts the triangle area feature. This feature to some degree shows the openness of the individual's posture. Additionally, we can get a feel for when the hands touch the head or each other (the triangle area is zero), however we have to rely on other features to determine which blobs are involved.

The distance from each blob to the center of triangle from uses a simple Euclidean distance formula. The formula uses the center point coordinates from one of the blobs and the center point of the triangle. Fig. 7 illustrates this feature. This feature captures the relationship of one blob to the two other blobs in a single metric and may provide insight into posture of an individual.

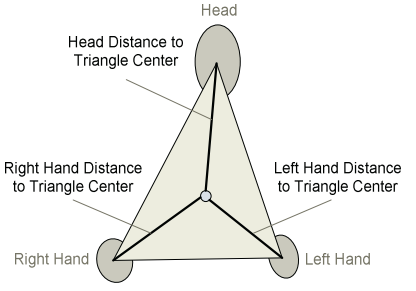


Fig. 7. Distance to triangle center

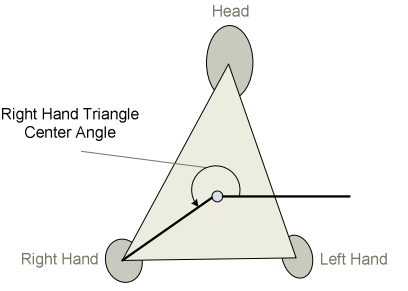


Fig. 8. Triangle center angle features

A blob's angle relative to the center of the triangle is also classified into the Multi-relational category of single frame features. The center point coordinate of a blob and the triangle center are used when calculating the angle. This angle, initially in radians, is converted to degrees. Fig. 8 illustrates the triangle center angle feature for a right hand blob. This feature can discriminate when one hand is up but not the other or when the body is asymmetric (both hands to one side of the head).

Multi-frame Features

The features in the descriptive category of multi-frame features are calculated using the descriptive single frame features. The distance feature tracks how much a single blob has moved from one frame to another. Fig. 9 illustrates this concept for a right-hand blob between frame one (dashed border) and frame two (solid border). Once again this distance is calculated using a simple Euclidean distance formula. Since we calculate this metric for each frame, which represents approximately $1/30^{th}$ of a second, we can also view this metric as a representation of speed of movement. This gives us insight into how much an individual moves her or his head and hands over time.

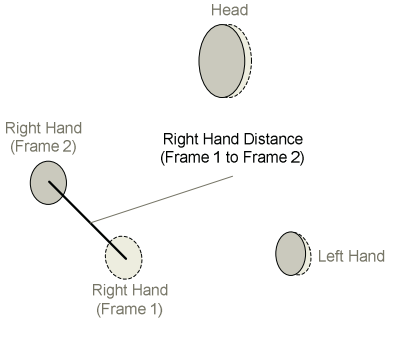


Fig. 9. Multi-frame Distance Feature

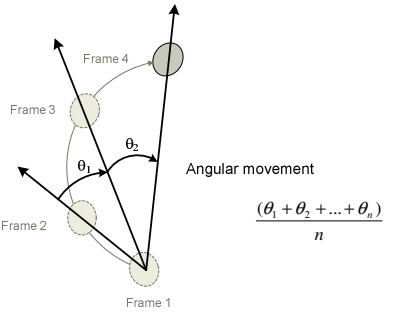


Fig. 10. Angular Movement Feature

Another feature in the multi-frame descriptive category is angle difference. To calculate this feature we simply subtract the last frame's angle for a specific blob from the current frame's angle for that same blob. Whether this feature is positive or negative captures not only the rotation of the major axis, but also the direction of rotation. It was hypothesized that the angle differences for the head may be useful, but not for the hands. By naïve observation it was noticed that the shape of the hand blobs and the major axis frequently change when the hands are pivoted at the wrist.

The angular movement category of multi-frame features contains features that attempt to capture arced movement of a blob over a period of time. Fig. 10 illustrates how this is calculated for a series of four frames. The first frame serves as a reference point and vectors are calculated using the first frame as the origination point and other frames as destination points. Polar coordinate angles are calculated for each vector. The angular movement feature consists of the average change in vector angles. Our calculations utilize 5 frames when calculating this feature. This feature was designed to capture patterns that are representative of illustrator gestures.

The multi-frame directional category of features contains information about the direction that a blob has moved. A vector is created using the center points of the same blob in two different frames and a polar coordinate angle is extracted and converted to degrees. This feature can further be split into binary-value classifications (e.g. up, up-left, down-left, right, etc). It is hypothesized that these values will be particularly useful when using time-series statistics and they will represent information contained in illustrator and adapter gestures.

3.3 Analysis of Summarized Features

In order for these extracted features to be analyzed by specific statistical methods, such as discriminant analysis and regression, they need to be summarized over the length of the video clip. Simple descriptive statistics (means and variances) were calculated for all of the features extracted. It was anticipated that the variance of some features would more telling than the feature itself. For example, by calculating the variance of the triangle area, we can quickly see how much overall change in hand and head position occurred during the length of the clip.

4 Experiments and Results

The method described above was used to identify deception in a Mock Theft experiment. The experiment and the preliminary results are reported below.

4.1 Description of Mock Theft Experiment

The purpose of the Mock Theft experiment [15, 16] was to reveal cues that can be used in deception detection. In this experiment, a wallet was left unattended in a classroom and some participants played the role of a thief who took the wallet. Other participants either observed the theft or were unaware of the theft. All the participants were interviewed by untrained and trained interviewers via chat, audio-only, and face-to-face channels and these interactions where recorded. The participants in the experiment were undergraduate students in a large mid-western university.

The face-to-face interactions from the Mock Theft experiment where used in this study. There were a total of 42 possible face-to-face interactions that could be included in the study, four were not used because of improper video work or because the participant did not follow instructions. Each interaction was composed of a number of question-answer exchanges. In this study only the theft narrative was included in the analysis. Of the 38 responses, 16 were truthful and 22 were deceptive.

4.2 Discriminant Analysis

Our initial approach in evaluating the features that were extracted from the video segments was through discriminant analysis. Each feature was reviewed for discriminating power between the deceptive and truthful conditions.

The final model included average and variance of the head position and angle, the average and variance of the positions of the head and hands, the average distance between the hands, the average and variance in the distance between the center of the triangle and the hands and head, the average triangle area, the variance of the center position of the triangle, and the average number of frames the hands are located in each quadrant

The results of the preliminary discriminant analysis are shown in Table 2. The deceptive and truthful participants were classified with an accuracy rate of 89.5 percent. However, when one participant is withheld from the analysis and then used for testing, the accuracy falls to 60.5 percent. This preliminary result is promising in its discriminatory ability however, the model is not significant ($p = .24$).

Table 2. Discriminant analysis accuracy results

		Predicted	
	Actual	Truthful	Deceptive
Original	Truthful	87.5	12.5
	Deceptive	9.1	90.9
Cross-validated	Truthful	56.3	43.8
	Deceptive	36.4	63.6

4.3 Regression Model

As part of the mock theft experiment, participants were asked to rate the level of honesty (a number from 1 to 10) that they displayed during the narrative portion of the experiment. We built a multivariate regression model with honesty level as the dependent variable.

We selected several promising features as the initial independent variables and then used backward elimination to remove variables from the model. The resulting model had an R^2 of .856, an adjusted R^2 of .644, and a standard error of estimate of 2.55.

Predictors for this model include variance and average of the triangle area, and distances from each blob to the triangle center; the variance of the center point of the triangle (x and y), the head and right hand x, and the left hand x and y; the average amount of time that the left hand is in quadrant 1 and 2, and the right hand in quad-

rants 1 and 4; the average head distance; the average head angle difference, and the variance of the left hand distance.

Features from the multi-relational category, both variances and averages, seemed to be most influential in the model. Only averages for features in the relational category were selected to be part of the model. However, only variances for features in the single frame descriptive category were included in the model. The multi-frame descriptive category had 2 average features and 1 variance feature. Variables from the angular movement and directional categories were not used in the model because the features in these categories were primarily designed for time series analysis.

Each of the predictors in this model were individually significant ($p < .05$) with the exception of the average time that the left hand was in quadrant 1 ($p=.094$) and the variance of the head x ($p=.052$). Many of the predictors are individually significant at levels where $p<=.001$. As a whole the model has a significance of $p=.004$. The ANOVA table for the overall model is shown in Table 3.

Table 3. Regression ANOVA

	Sum of Squares	df	Mean Square	F	Significance
Regression	575.802	22	26.173	4.037	.004
Residual	97.250	15	6.483		
Total	673.053	37			

While the regression model shows promise for the task of classifying the level of honesty of an individual under similar circumstances, the main purpose of applying the statistical methods was to help understand the contributions of the extracted features and validate our hypothesis about their usefulness. Additional testing and refinement of the models need to occur to validate the predictive ability. However, we argue that the novel features and the method of deception detection described in this paper are supported by this preliminary study.

5 Conclusion

The automated extraction of behavioral cues associated with deception overcomes many of the weaknesses which hinder other methods of deception detection. It is not invasive and can be done without cooperation from the interviewee. This method allows flexibility, could provide prompt feedback, and does not require specially trained interviewers. This method does not require the same controlled setting as CVSA and is firmly grounded in deception theory and past empirical studies.

5.1 Future Steps

Currently, only summary data from a video segment is being analyzed for deception. Future steps will include time series analysis so that a near real-time system will be possible. More data collection efforts are planned to diversify our existing data from the Mock Theft experiment. These new data will be larger in number and will come from actual criminal interviews where deception regularly takes place and the motivation to deceive is not induced by an experimenter. Finally, probabilistic models are

currently being considered in the place of linear regression and discriminant analysis for predicting deception. These models may more closely represent human behavior.

References

1. A. Vrij, Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice. West Sussex: John Wily & Sons Ltd, 2000.
2. G. Ben-Shakhar and E. Elaad, "The Validity of Psychophysiological Detection of Information with the Guilty Knowledge Test: A Meta-Analytic Review," Journal of Applied Psychology, vol. 88, pp. 131-151, 2003.
3. G. Ganis, S. M. Kosslyn, S. Stose, W. L. Thompson, and D. A. Yurgelun-Todd, "Neural Correlates of Different Types of Deception: An fMRI Investigation," Cerebral Cortex, vol. 13, pp. 830-836, 2003.
4. R. Johnson, J. Barnhardt, and J. Zhu, "The contribution of executive processes to deceptive responding," Neuropsychologia, vol. 42, pp. 878-901, 2004.
5. R. G. Tippett, "A Comparison Between Decision Accuracy Rates Obtained Using the Polygraph Instrument and the Computer Voice Stress Analyzer in the Absence of Jeopardy," vol. 2003: Florida Department of Law Enforcement, 1994.
6. B. DePaulo, J. Lindsay, B. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," Psychological Bulletin, vol. 129, pp. 74-118, 2003.
7. P. Ekman and W. V. Friesen, "Nonverbal Leakage and Clues to Deception," Psychiatry, vol. 32, pp. 88-105, 1969.
8. M. Zuckerman and R. E. Driver, "Telling Lies: Verbal and Nonverbal Correlates of Deception," in Nonverbal Communication: An Integrated Perspective, A. W. Siegman and S. Feldstein, Eds. Hillsdale, NJ: Erlbaum, 1985, pp. 129-147.
9. D. Buller, J. Burgoon, C. White, and A. Ebesu, "Interpersonal Deception: VII. Behavioral Profiles of Falsification, Equivocation and Concealment," Journal of Language and Social Psychology, vol. 13, pp. 366-395, 1994.
10. A. Vrij, K. Edward, K. P. Roberts, and R. Bull, "Detecting Deceit via Analysis of Verbal and Nonverbal Behavior," Journal of Nonverbal Behavior, vol. 24, pp. 239 - 263, 2000.
11. D. B. Buller and J. K. Burgoon, "Interpersonal Deception Theory," Communication Theory, vol. 6, pp. 203-242, 1996.
12. S. Lu, D. Metaxas, D. Samaras, and J. Oliensis, "Using Multiple Cues for Hand Tracking and Model Refinement," presented at IEEE CVPR 2003, Madison, Wisconsin, 2003.
13. S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, "Blob Analysis of the Head and Hands: A Method for Deception Detection," presented at Hawaii International Conference on System Science (HICSS'05), Hawaii, 2005.
14. J. K. Burgoon, M. Adkins, J. Kruse, M. L. Jensen, T. Meservy, D. P. Twitchell, A. Deokar, J. F. Nunamaker, S. Lu, G. Tsechpenakis, D. N. Metaxas and R. E. Younger, "An Approach for Intent Identification by Building on Deception Detection," presented at Hawaii International Conference on System Science (HICSS'05), Hawaii, 2005.
15. J. K. Burgoon, J. P. Blair, T. Qin, and J. F. Nunamaker, "Detecting Deception Through Linguistic Analysis," presented at NSF/NIJ Symposium on Intelligence and Security Informatics, 2003.
16. J. K. Burgoon, J. P. Blair, and E. Moyer, "Effects of Communication Modality on Arousal, Cognitive Complexity, Behavioral Control and Deception Detection During Deceptive Episodes," presented at Annual Meeting of the National Communication Association, Miami Beach, Florida, 2003.

Automatically Determining an Anonymous Author's Native Language

Moshe Koppel, Jonathan Schler, and Kfir Zigdon

Department of Computer Science, Bar-Ilan University,
51900 Ramat-Gan, Israel
{koppel, schlerj, zigdonk}@cs.biu.ac.il

Abstract. Text authored by an unidentified assailant can offer valuable clues to the assailant's identity. In this paper, we show that stylistic text features can be exploited to determine an anonymous author's native language with high accuracy.

1 Introduction

One of the first requirements for countering a security threat is identifying the source of the threat. In the case of a written threat, the text itself can be a valuable source of information regarding the identity of an assailant. If the text is long enough, stylistic analysis of the text might offer hints towards a psychological or demographic profiling of the text's author. For example, it has already been shown that automated text analysis methods can be used to identify an anonymous author's gender with accuracy above 80% [1].

In this paper, we will show that stylistic idiosyncrasies can be used to identify the native language of the author of a given English language text. Writers' spelling, grammar and usage in a second language are often influenced by patterns in their native language [2] [3]. Thus, it is plausible that certain writing patterns – function word selection, syntax and errors – might be particularly prevalent for native speakers of a given language.

Some work [4] has been done on categorizing transcripts of English speech utterances as by native or non-native English speakers. In our experiments, we know that the writer is not a native English speaker but we wish to determine which language is native to the author. We consider written text, which offers the benefit of grammar and spelling cues, but loses the benefit of mispronunciation cues. To the best of our knowledge, this is the first published work on the automated determination of author native language from written text.

2 Stylistic Features

Identifying an author's native language is a type of authorship attribution problem. Instead of identifying a particular author from among a closed list of suspects, we wish to identify an author class, namely, those authors who share a particular native language.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 209 – 217, 2005.

Researchers in authorship attribution typically seek the kinds of features use of which is roughly invariant for a given author (or author class) across topics but which might vary from one author (or author class) to another. Generally, researchers use feature sets that are relatively common. Thus, for example, the seminal authorship attribution work of Mosteller and Wallace [5] on the Federalist Papers used a set of several hundred function words, that is, words that are context-independent and hence unlikely to be biased towards specific topics. Other features used in even earlier work [6] are complexity-based: average sentence length, average word length, type/token ratio and so forth. Recent technical advances in automated parsing and part-of-speech (POS) tagging have facilitated the use of syntactic and quasi-syntactic features such as POS n-grams [7] [8] [9] [10]. Other recent work [11] considers language modeling using letter n-grams.

However, human experts working on real-life authorship attribution problems do not work this way. They typically seek idiosyncratic usage by a given author that serves as a unique fingerprint of that author. For example, Foster [12] describes his techniques for identifying a variety of notorious anonymous authors including the author of the novel, *Primary Colors*, and the *Unabomber*. These techniques include repeated use of particular types of neologisms or unusual word usage. Significantly, Foster identifies these linguistic idiosyncrasies manually. In the case of unedited texts, spelling and grammatical errors, which are typically eliminated in the editing process, can be exploited as well.

In this paper, we will use a variety of stylistic feature types that might be helpful for determining an author's native language. Very crudely, we can break these feature types into three broad categories.

1. Function words – As noted above, function words are useful for authorship attribution [5]. It stands to reason that such words might also be useful for native language identification since certain function words are liable to be used more or less frequently by native speakers of a given language, depending on the presence or absence of analogues for those words in the given language. A good example is the word *the*, which is typically used less frequently by native speakers of languages, such as Russian, that do not use a definite article.
2. Letter n-grams – As noted, letter n-grams have also been shown to be useful for authorship attribution [11]. It is likely that this is simply an artifact of variable usage of particular words, which in turn might be the result of different thematic preferences. In the case of native language attribution, however, letter n-grams might reflect the orthographic conventions of an author's native language (at least for those cases in which the native language uses the Latin alphabet).
3. Errors and Idiosyncrasies – As noted, errors and idiosyncrasies are the features commonly used by attribution experts analyzing data manually. Their utility for native language attribution is obvious: writers might be expected to transport orthographic or syntactic conventions from their native languages over to English in ways that result in non-conventional English. One of the main contributions of this paper will be to fully automate the process of idiosyncrasy detection. The next two sections of the paper will be devoted to describing this process.

3 Error Types

Flagging of various types of writing errors has been used in a number of applications including teaching English as a foreign language (EFL) [13] [14] student essay grading [15] and, of course, word processing. The approaches used are either partially manual or do not flag the types of errors relevant to our task. Consequently, we develop our own automated methods for error-tagging.

Our first challenge is to identify the error types we are interested in tagging. After that we will show how to automate the tagging process. The error types we consider fall into the following four categories:

1. Orthography – We consider here a range of spelling errors as follows:

    - Repeated letter (e.g. *remmit* instead of *remit*)
    - Double letter appears only once (e.g. *comit* instead of *commit*)
    - Letter α instead of β (e.g. *firsd* instead of *first*)
    - Letter inversion (e.g. *fisrt* instead of *first*)
    - Inserted letter (e.g. *friegnd* instead of *friend*)
    - Missing letter (e.g. *frend* instead of *friend*)
    - Conflated words (e.g *stucktogether*)
    - Abbreviations

The first six of these represents multiple error types since the specific letter(s) are specified as part of the error. For example, a missing *i* is a different error than a missing *n*. Thus, in principle, "Letter α instead of β" represents 26*25/2 = 325 separate error types. We will see below though that most of these occur so infrequently that we can consider only a small subset of them.

It should be emphasized that we use the term "error" or "idiosyncrasy" to refer to non-standard usage or orthography in U.S. English, even though often such usage or orthography simply reflects different cultural traditions or deliberate author choice.

2. Syntax – We consider non-standard usage as follows:

    - Sentence Fragment
    - Run-on Sentence
    - Repeated Word
    - Missing Word
    - Mismatched Singular/Plural
    - Mismatched Tense
    - *that/which* confusion

Our system supports these error types for use in a variety of applications not considered in this paper. These errors are not appropriate for the native language problem we consider here, so we do not use them in the experiments reported below.

3. Neologisms – In order to leverage an observation of Foster [12] that certain writers tend to create neologistic adjectives (like *fantabulous*) while others create neologistic verbs, nouns, etc., we note for each POS (other than proper nouns), entirely novel exemplars (i.e. those for which there is no near match) of that POS.

4. Parts-of-speech bigrams – We consider 250 rare POS bigrams in the Brown corpus [16]. Such pairs might be presumed to be in error, or at least non-standard. Chodorow and Leacock [15] flag errors for essay grading by checking those POS pairs which appear less frequently in the corpus than would be expected based on the frequency of the pair's constituent individual POS. For our purposes, any rare POS bigram that shows up in a text is worth noting.

4 Automated Error Tagging

Of course, we can conjure many sophisticated error types, based upon deeper linguistic analysis of the text, besides those that were presented in the previous section. However, we restrict ourselves here to those considered above because they can be identified with relative ease, as we now show.

In order to tag errors in the above list, we exploit existing tools. Thus, for most of the error types in categories 1 and 2 above, we use the following procedure:

We run a text through the MS-Word application and its embedded spelling and grammar checker. Each error found in the text by the spell checker is recorded along with the best suggestion (to correct the error) suggested by the checker. Each pair <error, suggestion> is then processed by another program, which assigns it an "error type" from among those in the list we constructed.

Obviously, automated spelling and grammar checkers are far from perfect: certainly, suggested corrections may not reflect an author's intention. Nevertheless, since we are not interested in any individual error but rather to gather statistics on error-type frequencies, such automated checkers are adequate for our purpose. Still, for certain classes of errors we found MSWord's spell and grammar checker to be especially inadequate, so we prepared scripts ourselves for capturing them. In particular, we found that MSWord's spell checker was very weak at handling non-standard words with grammatical suffixes (*-ism, -ist, -ble, -ive, -logy, -tion,* etc.)

For categories 3 and 4, we run a text through the Brill [17] tagger. For category 4, we juxtapose results from MSWord's spelling checker (and our own routines for words with identifiable grammatical suffixes) with results of the Brill tagger.

When we ran our entire corpus of flawed texts through this process, we found that many error types on our list are so infrequent as to not be worth considering. Consequently, we reduced our list of error types to only those 185 types that occurred at least three times in a large corpus of chat group posts used for gathering error statistics (in addition to the 250 rare part-of-speech bigrams).

5 Experimental Setup

We use the International Corpus of Learner English [18], which was assembled for the precise purpose of studying the English writing of non-native English speakers from a variety of countries. All the writers included in the corpus are university students (mostly in their third or fourth year) studying English as a second language. All are roughly the same age (in their twenties) and are assigned to the same proficiency level in English. We consider sub-corpora contributed from Russia, Czech Republic,

Bulgaria, France and Spain. The Czech sub-corpus, consisting of essays by 258 authors, is the smallest of these, so we take exactly 258 authors from each sub-corpus (randomly discarding the surplus). Each of the texts in the collection is of length between 579-846 words.

Each document in the corpus is represented as a numerical vector of length 1035, where each vector entry represents the frequency (relative to document length) of a given feature in the document. The features are:

- 400 standard function words
- 200 letter n-grams
- 185 error types
- 250 rare POS bigrams

We use multi-class linear support vector machines (SVM) [19] to learn models for distinguishing vectors belonging to each of the five classes. The efficacy of linear SVMs for text categorization is already well attested [20].

In order to test the effectiveness of models learned by SVMs to properly categorize unseen documents, we ran ten-fold cross-validation experiments: the corpus was divided randomly into ten sets of (approximately) equal size, nine of which were used for training and the tenth of which was used for testing the trained model. This was repeated ten times with each set being held out for testing exactly once.

6 Results

In Figure 1, we show accuracy results of ten-fold cross-validation experiments for various combinations of feature classes. As can be seen, when all feature types are used in tandem we obtain accuracy of 80.2%. The confusion matrix for the experiment with all features is shown in Table 1. It should be noted that a document is only regarded as being correctly classed if it is assigned to its correct class and to no other class. Thus, since we have five possible classes of roughly equal size, 20% accuracy is a reasonable baseline for this experiment.

The success of the system depends on the interaction of hundreds of features. Nevertheless, it is instructive to consider some of the features that proved particularly helpful in the learned models. Table 2 shows a variety of features along with the number of documents in each category in which the feature appears. We find that a number of distinctive patterns were exploitable for identifying native speakers of particular languages. For example:

- Some features that appear in more documents in the Bulgarian corpus than in the other corpora are the POS pair *most*–ADVERB and the somewhat formal function words *cannot* and *however*.
- A relatively large number of authors in the Spanish corpus had difficulty with doubling consonants, either doubling unnecessarily (*dissappear, fullfill, opening) or* omitting one of a double (*efect, inteligent*). Some errors that are almost exclusive to authors in the Spanish corpus derive directly from orthographic or pronunciation conventions of

Spanish: confusion of *m* and *n* (*confortable*) or *q* and *c* (*cuantity, cuality*).

- A relatively high number of authors of documents in the Czech corpus also doubled letters in a non-standard way.
- Documents in the French corpus are characterized by relatively frequent use of the word *indeed* as well as *Mr* (without the period) and, as in the Spanish corpus, incorrect use of the vowel o (*outhor, psychodelic*). The POS pair*s number—modal_verb* (*one must, one could*) and *there–to* (*there* and *to* are each assigned their own POS in the Brill tag set) also appears more frequently in the French corpus.
- Authors in the Russian corpus are more prone to use the word *over* as well as the POS pair NUMBER—*more*.

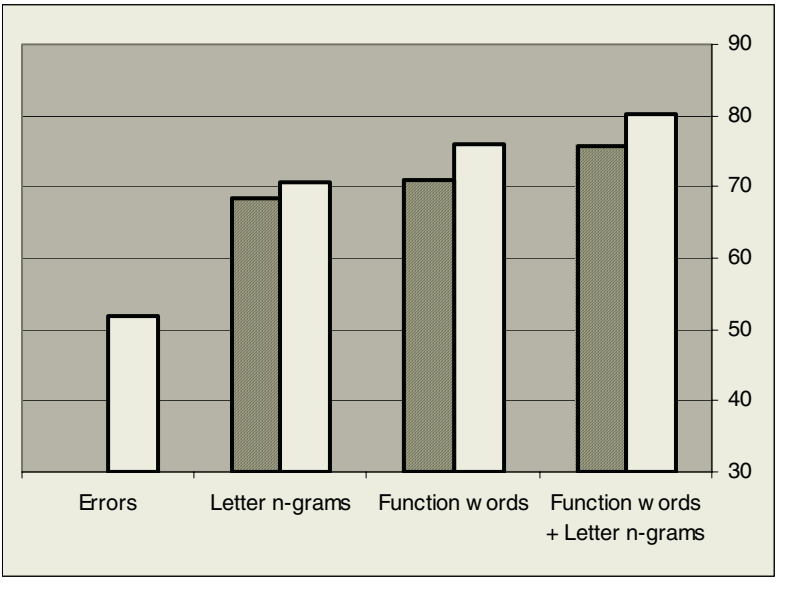


Fig. 1. Accuracy (*y-axis*) on ten-fold cross-validation using various feature sets (*x-axis*) without (diagonal lines) and with (white) errors, and classifying with multi-class linear SVM. Note that "errors" refers to all error types, including rare POS bigrams

Table 1. Confusion matrix for the author's native language identification using SVM

		Classified As				
		Czech	**French**	**Bulgarian**	**Russian**	**Spanish**
Actual	**Czech**	**209**	1	18	20	10
	French	9	**219**	13	12	5
	Bulgarian	14	8	**211**	18	7
	Russian	24	8	24	**194**	8
	Spanish	16	10	10	7	**215**

The above examples are all features that distinguish one language corpus from all the rest. Of course there are many features that distinguish some subset of languages from the others. For example, the frequency of the word *the* is significantly less frequent in the documents by Czech (47.0 per 1000 words), Russian (50.1) and Bulgarian (52.3) authors than in those by French (63.9) and Spanish (61.4) authors. (Russian and Czech both do not use a definite article and Bulgarian uses it only as a suffix.)

Unsurprisingly, as can be seen in Table 1, most mistakes were among the three Slavic languages (Russian, Czech, Bulgarian).

Table 2. A selection of features and the number of documents in each sub-corpus in which they appear

Feature	Bulgarian	Czech	French	Russian	Spanish
cannot	131	100	79	65	57
however	127	65	92	45	81
indeed	25	3	86	15	9
over	66	72	61	116	52
***most*_ADVERB**	20	3	3	8	1
NUMBER_*more*	5	6	2	17	5
there_to	2	2	14	2	0
NUMBER_MODAL	16	8	54	23	5
DOUBLED CONSONANT	31	106	91	46	108
MR (no period)	1	14	39	1	16
VOWELo	6	20	37	22	46
UNDOUBLED CONSONANT	43	113	59	45	167
CONSnm	1	8	2	2	47
CONScq	0	0	0	0	16

7 Conclusions

We have implemented a fully automated method for determining the native language of an anonymous author. In experiments on a corpus including authors from five different countries, our method achieved accuracy of above 80% in categorizing unseen documents.

The authors of these documents were generally reasonably proficient in English (and may have even used automated spell-checkers), which made the task particularly difficult. It may be, however, that we were able to take unfair advantage of differences in overall proficiency among the different sub-corpora. For example, the Bulgarian authors were on average considerably less prone to errors than the Spanish authors. One way to ensure robustness against such artifacts of the available data would be to run similar experiments in which error frequency is normalized not against document length but rather against overall error frequency.

The applicability of these methods in a law enforcement framework depends on a number of factors. Is the method precise enough to handle tens if not hundreds of different candidate native languages? How short can the documents be and still permit accurate categorization? Each of these questions requires further investigation.

References

1. Koppel, M., S. Argamon, A. Shimony. *Automatically categorizing written texts by author gender*. (2002) Literary and Linguistic Computing 17(4).
2. Lado, R. *Linguistics Across Cultures*, Ann Arbor: (1961) University of Michigan Press.
3. Corder, S. P. *Error Analysis and Interlanguage*. (1981) Oxford: Oxford University Press.
4. Tomokiyo, L.M. and R. Jones. *"You're Not From 'Round Here, Are You? Naive Bayes Detection of Non-native Utterance Text" (2001)* NAACL 2001.
5. Mosteller, F. and Wallace, D. L. *Inference and Disputed Authorship: The Federalist.* Reading, Mass. : Addison Wesley, (1964).
6. Yule, G.U. 1938. *On sentence length as a statistical characteristic of style in prose with application to two cases of disputed authorship*. Biometrika, 30, (1938) 363-390.
7. Baayen, H., H. van Halteren, F. Tweedie, *Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution.*In "Literary and Linguistic Computing, 11, (1996).
8. Argamon-Engelson, S., M. Koppel, G. Avneri. *Style-based text categorization: What newspaper am I reading?*. in Proc. of AAAI Workshop on Learning for Text Categorization, (1998), pp. 1-4
9. Stamatatos, E., N. Fakotakis & G. Kokkinakis, *Computer-based authorship attribution without lexical measures.* Computers and the Humanities 35, (2001) pp. 193—214.
10. Koppel, M., J. Schler. *Exploiting Stylistic Idiosyncrasies for Authorship Attribution*. in Proceedings of "IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis", (2003) Acapulco, Mexico
11. Peng, F., D. Schuurmans, S. Wang. *Augmenting Naive Bayes Classifiers with Statistical Language Models*. Inf. Retr. (2004) 7(3-4): 317-345
12. Foster, D. *Author Unknown: On the Trail of Anonymous*, (2000) New York: Henry Holt.
13. Dagneaux, E., Denness, S. & Granger, S. *Computer-aided Error Analysis. System*. An International Journal of Educational Technology and Applied Linguistics. Vol 26 (2), (1998) 163-174.
14. Tono, Y., Kaneko, T., Isahara, H., Saiga, T. and Izumi, E. *The Standard Speaking Test (SST) Corpus: A 1 million-word spoken corpus of Japanese learners of English and its implications for L2 lexicography*. Second Asialex International Congress, Korea, (2001) pp. 257-262
15. Chodorow, M. and C. Leacock. *An unsupervised method for detecting grammatical errors*, Proceedings of 1st Meeting of N. American Chapter of Assoc. for Computational Linguistics, (2000) 140-147
16. Francis , W. and H. Kucera. *Frequency Analysis of English Usage: Lexicon and Grammar*, (1982) Boston: Houghton Mifflin Company.
17. Brill, E. 1992. *A simple rule-based part-of-speech tagger*. Proceedings of 3rd Conference on Applied Natural Language Processing, (1992) pp. 152—155
18. Granger S., Dagneaux E. and Meunier F. *The International Corpus of Learner English. Handbook and CD-ROM*. (2002) Louvain-la-Neuve: Presses Universitaires de Louvain

19. Crammer, K. and Y. Singer *On the algorithmic implementation of multiclass kernel-based vector machines*, Journal of Machine Learning Research 2 (2001) pp.265-292
20. Joachims, T. *Text categorization with support vector machines: learning with many relevant features*. Proceedings of 10th European Conference on Machine Learning, (1998) pp.137--142

A Cognitive Model for Alert Correlation in a Distributed Environment[1]

Ambareen Siraj and Rayford B. Vaughn

Center for Computer Security Research,
Department of Computer Science and Engineering
{ambareen, vaughn}@cse.msstate.edu

Abstract. The area of alert fusion for strengthening information assurance in systems is a promising research area that has recently begun to attract attention. Increased demands for "more trustworthy" systems and the fact that a single sensor cannot detect all types of misuse/anomalies have prompted most modern information systems deployed in distributed environments to employ multiple, diverse sensors. Therefore, the outputs of the sensors must be fused in an effective and intelligent manner in order to provide an overall view of the status of such systems. A unified architecture for intelligent alert fusion will essentially combine alert prioritization, alert clustering and alert correlation. In this paper, we address the alert correlation aspect of sensor data fusion in distributed environments. A causal knowledge based inference technique with fuzzy cognitive modeling is used to correlate alerts by discovering causal relationships in alert data.

Keywords: Network security, intelligent alert fusion, alert correlation, fuzzy cognitive modeling.

1 Introduction

Research in IDS improvement has taken on new challenges in the last few years. One such contemporary and promising approach in this area is alert fusion in a multi-sensor environment. Increased demands for "more trustworthy" systems and the fact that a single sensor cannot detect all types of misuse/anomalies have prompted most modern information systems deployed in distributed environments to employ multiple, diverse sensors. Therefore, the outputs of the sensors must be fused in an effective and intelligent manner to provide an overall view of the status of the system.

Alert fusion, alert aggregation, alert clustering, alert correlation - all serve the same primary purpose – i.e., to provide some form of high level analysis and reasoning capabilities beyond low level sensor abilities. We refer to *alert fusion* as the process of interpretation, combination and analysis of alerts to determine and provide a

[1] This work is supported by NSF Cyber Trust Program Grant No: SCI-0430354, NSA IASP Grant No: H98230-04-1-0205, Office of Naval Research Grant number N00014-01-1-0678. and the Department of Computer Science and Engineering Center for Computer Security Research at Mississippi State University (http://www.cs.msstate.edu/~security).

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 218 – 230, 2005.

quantitative value for the system such that the value is representative of the degree of concern in the system. In a distributed environment, characterized by physically (and maybe geographically too) dispersed but networked systems, sensors are used to monitor security violations in the protected network. In such environment, fusion of sensor reported alerts is necessary for:

- *Alert Clustering*: To find structural relationships in data by grouping/aggregating alerts with common features. Alert clustering can aid in alert reduction and discovery of general attack patterns.
- *Alert Correlation*: To find causal relationships in data by associating alerts, which are parts of linked chains of events. Alert correlation can help to identify multi-staged attacks and to reduce alert volume.

In this paper, we address the alert correlation aspect of sensor data fusion for distributed environments. Here we illustrate the use of a causal knowledge-based inference technique with *Fuzzy Cognitive Modeling* to discover causal relationships in sensor data. The following sections will outline the research, provide necessary background information, describe the technique we use in alert correlation, report on experimental results on a benchmark dataset and lastly conclude.

2 Related Work

Research in the area of alert fusion/alert aggregation/alert clustering/alert correlation has emerged in last few years and primarily concerns information modeling and high level reasoning. Among them, the ones that are relevant to our work are the following.

Julisch introduces attribute generalization in alarm (i.e., alert) clustering as a method to support *root cause discovery* [3]. This work outlines a semi-automatic approach for reducing false positives in alarms by identifying the root causes with clustering of alerts by abstraction and then eliminating the root causes to reduce alarm overload. Ning et al. proposes an alert correlation model based on prerequisites and consequences of intrusion [8]. With knowledge of prerequisites and consequences, the correlation model can correlate related alerts by matching the consequences of previous alerts with prerequisites of later ones and then *hyper alert correlation graphs* are used to represent the alerts. In the prerequisite-consequence model, the authors conduct reasoning with predicate logic where predicates are used as basic constructs to represent the prerequisites and consequences of attacks. This approach requires extensive modeling of attacks in terms of specifying prerequisite and consequence of each alert type in the sensor report. Yu and Frincke propose a model for *Alert Correlation and Understanding (ACU)* based on *Hidden Colored Petri-Nets (HPCN)* [13]. HPCNs model agents, resources, actions, and functions of a system with transition and observation probabilities. To perform correlation, a model based on prerequisites and consequences is generated using domain knowledge. Training of model is required to best fit the model parameters. Qin and Lee generate high level aggregated alerts from low level sensor data and then conduct causal analysis based on a statistical technique, known as the *Granger Causality Test*, to discover new patterns of attack relationships [10]. Although this approach does not require apriori

knowledge of attacks behavior, it still requires some human intervention in background alert identification used in the statistical technique employed.

Fuzzy Cognitive Maps (FCMs) originated from the combination and synergism of fuzzy logic and neural networks. Researchers have used FCMs for many tasks in several different domains. Use of FCMs was first reported in [11] for fusing alert information in a multi-sensor intrusion detection environment to assess network health. Fuzzy Intrusion Recognition Engine, a network based IDS, also use FCMs in detecting attacks from features extracted from network traffic [12]. The work that we present in this paper differs from our previous work [11] primarily in the focus of the research, which is discovery of causal relationships between alerts rather than structural relationships and in the use of abstract FCM models to address issues of scalability and uncertainly.

3 A Cognitive Model for Alert Correlation with FCMs

The principal objective of our intelligent alert fusion model is to provide an overall condensed view of the distributed system by assessing the health of the primary resources in the network, which are essentially the computing nodes/hosts in the network. Therefore, our fusion model is *resource-centric*, i.e., analysis is centered upon the resources in the system in terms of the communication involving them. A resource-centric view inherently reduces alert volume by:

- presenting an overall picture of the compromised resources to the security administrator instead of alerts;
- not having to take account of information out of the scope of resource perimeter.

For any environment where performance heavily depends on system resources, such a view only seems natural.

A cognitive model is a "generalization over repeated experience" where knowledge - acquired through perception and experience – is organized into mental structures. One such cognitive modeling and inferencing technique uses *Fuzzy cognitive map (FCM)* that allows us to represent our perception of the real world in a structured way. We are using *fuzzy cognitive modeling* to correlate alerts in sensor data because it offers a straightforward structural representation of causal knowledge and allows what-if kinds of reasoning for causal analysis of data. Proposed by Kosko, FCMs model the world as concepts and causal relations between concepts in a structured collection [4,5,6]. *Concepts* (nodes) in an FCM (fig. 1) are events that originate in the system and whose values change over time. The causality links between concepts are represented by directed *edges* that denote how much one concept impacts the other(s). The concepts in the FCMs can be crisp or fuzzy. Concepts typically take values in the interval [0,1]. In the simplest case, a concept is either on (1) or off (0). A concept can also be represented by a fuzzy set and can fire to some degree. The edges typically have values between 0 and 1 or –1 and 1. Edges can also be fuzzy, and in those cases we can use linguistic words such as "a little," "highly," "somewhat," to represent the edges. When the edges between concepts are fuzzy values, fuzzy set operators like T-norms and T-conorms can be applied to the particular chain of concepts to infer the total effects of concepts in the chain.

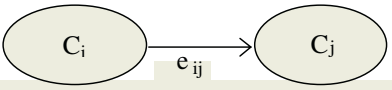


Fig. 1. Two FCM Concepts with a Causal Link

With the premise that every cause is bound to have an effect (whether high or low), our cognitive model views the alerts issued by the sensors as causes that have the potential to generate specific effects in systems. Different alerts in sensor report pertain to different actions of attackers with different objectives. The effects generated can potentially be coupled together in a causal chain to reveal the possible correlations between the alerts initiating them.

In this respect, our fusion model recognizes two kinds of events in systems:

- *Cause Events (CEvent*): This type of event is generated as a result of alerts seen in the sensor reports and corresponds to possible actions taken by the attacker to achieve some goal.
- *Effect Events (EEvent):* This type of event is generated as a result of activation of cause events and corresponds to specific security incidents in the system. Effect events can in turn act as cause events to further produce some other effect events (i.e., incidents).

To illustrate the use of such cognitive modeling for alert correlation, a common attack such as Distributed Denial of Service (DDoS) is examined. Suppose, a DDoS attack is to be launched using a known vulnerability of the *sadmind* service in *Solaris* systems. In this case, the following steps are usually carried out by an intruder [7]:

- Conduct *IPSweep* from a remote site to find out existence of hosts;
- Probe the hosts looking for *sadmind* daemon running on *Solaris* hosts;
- Break into host(s) using the *sadmind* vulnerability;
- Install *Trojan mstream* DDoS software on some host(s); and
- Launch the DDoS.

Fig. 2 is an FCM that models the scenario described above for the DDoS attack using cause and effect types of events (the EEvents shown here are similar to the consequences of hyper alert types as in [8]). The FCM in this figure denotes that an *IPSweep* alert in the sensor report will generate an *IPSweep* CEvent, from which *HostExits* EEvent can be inferred. Later, when the *SadmindPing* CEvent is generated from the alert report, the fusion model can associate this with a previously generated *HostExits* EEvent and the combination of both will generate a new EEvent *VulnerableToSadmind*. All alerts contributing to CEvents of a particular FCM model can be correlated as part of the attack scenario depicted by the FCM model.

But, problems with such specific/exact knowledge modeling are that:

- it does not scale well; and
- it does not work well when there are deviations in data (due to heterogeneous sources) or incomplete data (due to incomplete/imperfect source coverage).

Therefore, in order to address these issues and make FCM models more applicable to real world situations, our fusion model employs abstract or more generalized cognitive models such that:

- a particular model can accommodate variations of similar knowledge/evidence; and
- inference can still take place with incomplete/varied information.

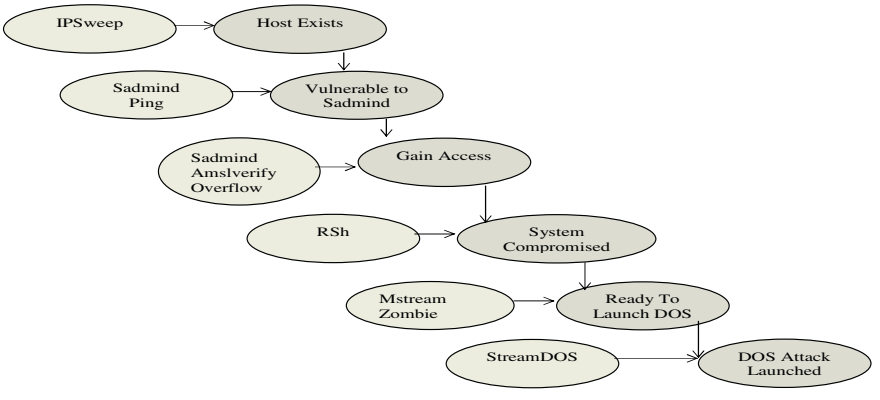


Fig. 2. An FCM Model for Detecting DDoS Attack Exploiting Sadmind Service Vulnerability

In this regard, our fusion model uses FCMs that employ more abstract or generalized events than specific ones (fig. 3.).

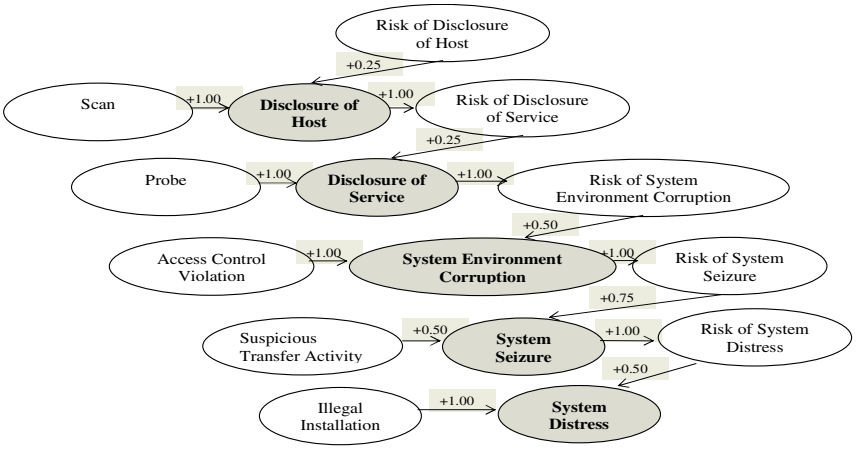


Fig. 3. An Abstract FCM Model for Detecting DDoS Attacks in General

For example, instead of generating specific CEvents like *Sadmind AmslverifyOverflow* (fig. 2), it abstracts alerts by using a generalization hierarchy such as in fig. 4, to activate more generalized CEvents like *AccessControlViolation*

(fig. 3). Note that the same CEvent will also be generated for similar types of alerts such as *StatdOverflow* or *SolarisLPDOverflow*. Also, instead of generating specific EEvents like *VulnerableToSadmind*, the fusion model activates more generalized EEvents like *DisclosureOfService*. Note that the same EEvent can also replace other specific EEvents like *VulnerableToStatd*, or *VulnerableToSolarisLPD*. Thus, a single such cognitive model of an abstract attack scenario can replace multiple explicit attack models and help with scalability issues in alert correlation modeling.

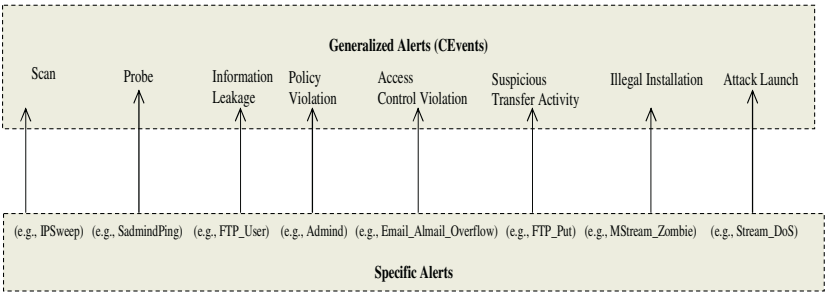


Fig. 4. Generalization of Low Level Sensor Alerts to Abstract CEvents

Different actions of an attacker targeted at a particular host activate different incidents for the host. The degree to which an incidence occurs depends not only on the report of the corresponding action taken by the attacker, but also on the existing risk of such an incidence taking place (Fig. 3). For example, the EEvent *SystemEnvironmentCorruption* primarily depends on the sensor reporting of CEvent *AccessControlViolation* (alert impact[1] designated by FCM edge of +1.00). But this type of action is not always successful and therefore, sensor notification of this alert does not guarantee that such an incidence actually took place in reality. Our model deals with this uncertainty by taking additional information into account - which is the existing risk of such incidence happening for the particular resource in question. Fig. 3 shows the risk impact designated by FCM edge of +0.50 for the incident *SystemEnvironmentCorruption*. Note the difference in between alert and risk impact. This is because usually, we pay more attention on the report of the alert itself than on the existing risks. But, sometimes when alerts such as - *rsh, Telnet XDisplay, ftp_put* - are issued by sensors, which may or may not result from actual malicious activities, the existing risk (or possibility) of such incident occurring should impact the incident

[1] The edge values used in the correlation model come from security experts' common sense judgment and experience. Note that edges represent how much a certain concept impact the other, on a scale of 0 to 1 or 0 to -1. Although these impact values are determined from expert knowledge and experience, once the values are initially set, their performance can be observed over time and their values can be tuned for optimal performance by the security administrator based on the empirical performance of the alerts generated. We have found that FCMs offer a highly flexible structure in this regard. A variety of both manual and automated techniques can potentially be used to fine-tune these parameters.

more than the alert itself. Hence impacts of such CEvents, like *Suspicious TransferActivity,* are less than the impact of associated risk.

The degree to which an incidence occurs for a particular resource designates its *incidence strength*. In order to compute incidence strengths and also to fuse the overall impact of the different incidents activated for each host, the resemblance between FCMs and neural networks is utilized [1]. In the neural network approach, the concepts of the FCMs are represented by neurons and the edges are represented by the weights of the connecting neurons. The incidents, treated as neurons, trigger activation of alert levels with different weights depicting causal relations between them. An adjacency matrix is used to list these cause and effect relationships between the FCM concepts. In an FCM, the runtime operation is observed by determining the value of the effect concept from the cause concepts and the connecting edge values.

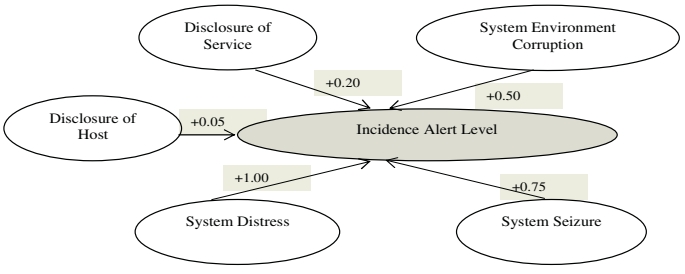


Fig. 5. FCM model for Fusing Evidences of Incidents

As incidents are activated for a resource, our model correlates the alerts that contribute in activating the incidents. Along with identifying the correlated alerts, our model also fuses the different incidence strengths of a particular resource to measure the degree of the *incidence alert level (IAL)* of the resource in the correlated alerts. Fig. 5 shows how the evidence of different incidents activated for a resource contribute to the IAL of the resource with different impacts. The degree of impact depends on the nature of the incidence and security policy. At any time, IAL of any particular resource collectively represent the effects of all the incidents activated for the resource at the time. Therefore in accordance with FCM inference [6], the IAL of a resource R_i at t_{n+1} time for each contributing incidents I_k with impact e_{ki}, can be represented as the following:

$$IAL\ (R_i)(t_{n+1}) = \left[\frac{\sum_{k=1}^{n} (I_k)(t_n) * e_{ki}(t_n)}{\sum_{k=1}^{n} e_{ki}(t_n)} \right]$$

IAL can be considered a confidence score given by the fusion model to represent the degree of concern for a particular resource in its involvement in correlated incidents resulting from multi-staged attacks. It should be pointed out that with FCM modeling of system events, the presence of all predecessor events is not mandatory in a correlation scenario for inferring subsequent events. Alert correlation with abstract fuzzy cognitive modeling allows inference to progress with missing/incomplete alerts

in sensor reports. Due to space constraints in this paper, discussion on dealing with missing alerts is left for future avenues.

4 Experimental Results

To evaluate the effectiveness of our alert correlation technique based on fuzzy cognitive modeling, we started out with experiments in a traditional distributed environment where our objective was to evaluate the alert fusion model's ability to correlate low level sensor alerts that are part of coordinated attacks.

We chose to use MIT Lincoln's Lab's DARPA 2000 Intrusion Detection Evaluation (IDEVAL) Scenario Specific Data Sets [7] for the experiments because it is a renowned benchmark dataset that contains simulated multi-staged attack scenarios in a protected environment. Use of this dataset also allows us to compare our experimental results to work by other researchers in this area. The dataset includes a series of attacks carried out over multiple network and audit sessions by an attacker who probes hosts, successfully breaks in some of them to prepare for and finally launch DDoS attacks against an off-site government website. In the DARPA 2000 dataset, there are three segments of a simulation network: a network inside an Air Force base, an internet outside an Air Force base and the demilitarized zone (DMZ) that connects the outside to inside [7]. Also, there are two attack scenarios: one that includes DDoS attacks carried out by a novice attacker (DDoS 1.0) who compromises three hosts individually and one that includes DDoS attacks carried out by a more sophisticated attacker (DDoS 2.0.2) who compromises one host and then fans out from it. The DARPA website provides a list of all the hosts in the three segments of the evaluation network [7].

Since we are interested in the fusion of sensor data, we needed to work on sensor alert report generated on the Lincoln Lab dataset. Such a sensor alert report by RealSecure network sensor (Version 6.0) [2], executed with Maximum Coverage Policy on the Lincoln Lab's datasets, has been made available by researchers at North Carolina State University as a part of the TIAA (A Toolkit for Intrusion Alert Analysis) project [9]. We used this sensor alert report in our experiments to evaluate the usefulness of our approach for alert correlation.

For our experiments, we used a similar abstraction hierarchy as shown in fig. 4 to generalize the alert types in the sensor alert report. The low level alerts reported by RealSecure were generalized to abstract categories with the help of attack signatures descriptions provided by ISS, Inc.'s X-Force database, a very comprehensive threats and vulnerabilities database (http://xforce.iss.net/). In addition, security experts were consulted for their valuable comments/suggestions on the generalization scheme.

As we ran our cognitive model on the DDoS 1.0 inside zone sensor alert report, we were able to correctly identify the three victim hosts (*mill*: 172.016.115.020, *pascal*: 172.016.112.050 and *locke*: 172.016.112.010), which the attacker compromised individually and then used to launch the DDoS attack. The graph in fig. 6 shows the DDoS attack scenario represented from the alerts correlated by our experiment. The graph shows four specific incidents identified by the model that represents four distinct phases of the attacks, which are as follows:

- Probing activities were conducted to discover services running on in the hosts, which resulted in the incidence *Discloser of Service* (DSV). (Here the attacker probed the hosts with *sadmind ping* to detect which ones had the *sadmind* service running.)
- Exploitation attacks were executed that resulted in the incidence *System Environment Corruption* (SEC). (Here the attacker used the vulnerability associated with the *sadmind* service to gain root access into the victim hosts.)
- Remote-to-root activities were carried out that resulted in the incidence *System Seizure* (SSZ). (Here the attacker uploaded necessary files for installing mstream software on the compromised hosts via *telnet* and *rsh*.)
- Attack tools were installed which resulted in the incidence *System Distress (SDT)*. (Here the attacker installed *Trojan mstream* DDoS software to carry out DDoS from the victim hosts.)

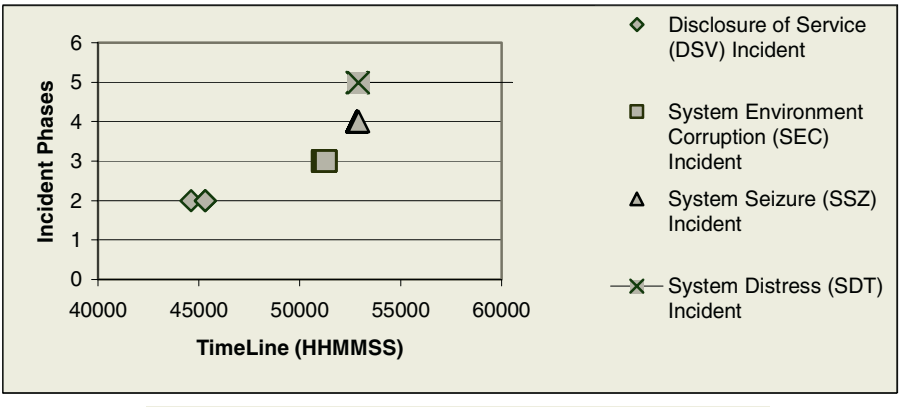


Fig. 6. Correlated Alerts depicting DDoS Attack Scenario

According to DARPA documentation [7], there are two additional phases in the coordinated attack and these were not reported by the FCM model. In DDoS 1.0, the attacker first conducted an *IPSweep* of the network from a remote site. Since the sensor RealSecure missed this alert, consequently our model was unable to analyze the corresponding alert data. Ning et al. report this same problem during their experiments with the DARPA data [8]. This highlights the fact that effectiveness of any high level analysis of sensor data is largely dependent on the quality of the sensor data itself. The final phase of the DDoS attack concerns launching of the *Stream_DoS* attack itself, which was also missed by the model. The reason is that our model is resource-centric and therefore, concentrates on communications to/from legitimate hosts in the network. Since this final DDoS attack was launched from a spoofed IP address and targeted a host outside of the network, our model did not consider the corresponding communication for analysis. However, it does not severely affect the alert situation awareness for the network because the critically significant *System Distress* incidence inherently places the concerned system under severe alert.

To show the alert correlation capability of our model, we use correlation performance metrics suggested by Qin and Lee [10]. Table 1. denotes the alert correlation results. We refer to the DARPA documentation to determine the number of causal relationships in data. It should be noted that the information in the table concerns only the correlated alerts in the sensor report. The abbreviations used in the table are the following:
SA: Sensor reported alerts, CR: Causal Relationships, CA: Correlated Alerts, CCA: Correctly Correlated Alerts, ICA: Incorrectly Correlated Alerts, MA: Missed Alerts, ARR: Alert Reduction Rate, TCR: True Causality Rate, FCR: False Causality Rate. The values in the Dataset column designate the following: 1: DDoS 1.0 Inside Zone, 2: DDoS 1.0 DMZ Zone, 3: DDoS 2.0.2 Inside Zone, and 4: DDoS 2.0.2 DMZ Zone.

Table 1. Correlation Performance of the FCM Model

Dataset	SA	CR	CA= CCA+ ICA	CCA	ICA	MA = CR-CA	ARR= (1-CA/ SA)	TCR= CCA/CR	FCR= ICA/CA
1	922	44	43	41	2	3	95.33%	93.18%	4.65%
2	891	57	58	55	3	1	93.49%	96.49%	5.17%
3	489	16	11	11	0	5	97.75%	68.75%	0%
4	425	6	4	4	0	2	99.06%	66.67%	0%

Ning et al. conducted alert correlation and reported experimental results on the same datasets using hyper-alert correlation graphs [8]. The work we present here is primarily different from Ning et al.'s work in the technique used. In [8], the authors refer to correctly correlated alerts as true alerts and report the following true alerts for the datasets: (1: 41, 2: 54, 3: 10, and 4: 3). Here we refrain from comparing our results to theirs because of the differences in the way we consider and count truth. For example, in counting causal relations in data we include alerts for telnet sessions that were initiated by an attacker in preparation for installing DDoS tools in the compromised hosts [7]. Every telnet session generated three alerts by RealSecure: *Telnet Terminal Type* (*TTT*), *Telnet Env_All* (*TEA*) *and Telnet XDisplay* (*TXD*). In the FCM model, *TTT, TEA* alerts are generalized to *InformationLeakage* CEvent, which does not generate any incidence that is part of a coordinated attack such as in fig. 3. Since we generalize *TXD* to *SuspiciousTransferActivity* (because this alert denotes actual initiation of a remote session), our model is able to correlate this alert to be part of the attack scenario. Therefore, we count the number of causal relations as 44 (the same 41s reported by Ning et al. in [8] plus three more for one telnet session). We also include two of the telnet alerts (*TTT* and *TEA*) as missed alerts along with the one for *Stream_DoS*. For dataset 3, the number of missed alerts is five (four for two telnet sessions plus one more for *Stream_DoS*). As we consider the missed alerts as false negatives, Table 1. shows that the TCRs we report are better for scenario one than for

scenario two. Also, our model incorrectly correlates two alerts in dataset 1 which are the *FTP_Syst* and *Email_Almail_Overflow* alerts generated for host 172.016.113.148. In dataset 2, in addition to the same two false positives as in dataset 1, there is an additional one for *UDP_Port_Scan* alert for host 172.016.112.050. While correlation of these alerts is justifiable but since this is not mentioned in the DARPA documentation [7], we count them as false positives.

A significant advantage of using alert correlation is reduction of alert volume such that the security administrator is not overwhelmed with a large volume of alert data. Table 1. also shows the effectiveness of our approach in reducing alert volume in terms of correlated data (i.e., data that are part of multi-staged attacks).

The ultimate goal of the FCM model is to provide the security administrator with a condensed view of the system in terms of the resources that are affected in alert situations and also their involvement in such situations by reporting their incidents strengths and incidence alert levels. Therefore, our model also reports overall situation of the hosts that are under alert and not just in terms of the alerts themselves. Fig. 7 is a snapshot of the alert situation discovered for the compromised host *mill* (172.016.115.020) in the coordinated multi-staged attacks as captured in all four datasets. Here the x-axis shows the incident strengths of the incidents that were activated for this host and y-axis shows the underlying dataset the analysis is based upon.

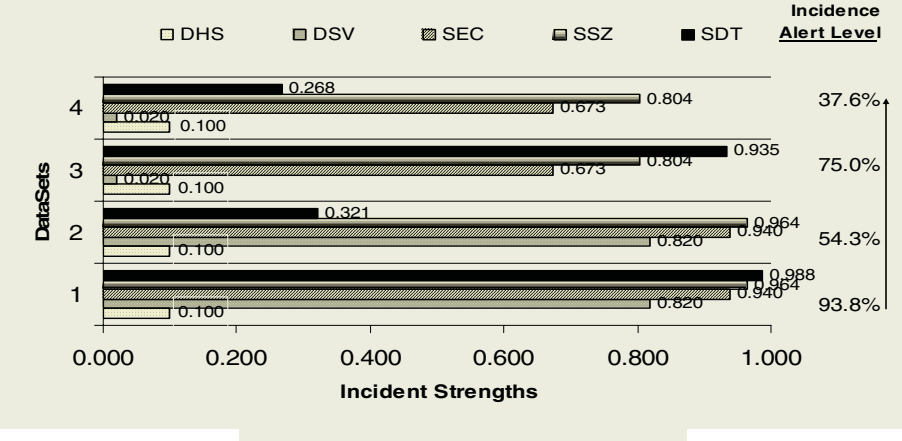


Fig. 7. Alert Situation for Host 'mill'

One can observe from fig. 7 how the existence of the intruder's action (that causes the incidence) and the existing risks (that give rise to possibility of incidence) jointly affect the strength of the incidence itself. For example, the *SystemEnvironment Corruption* (SEC) incidence for *mill* has a high strength of 0.94 for scenario one (dataset 1 & 2) and a moderate strength (0.673) for scenario two (dataset 3 & 4). This is because for scenario two, the attacker used more sophisticated probing technique (*DNS_HInfo*), which was not reported by the network sensors and therefore, no preceding EEvents were generated for the host *mill* that had the potential to place *mill* under risk of SEC. Consequently, as *mill* was under no initial risk of SEC, the incidence SEC activated with lesser strength. Fig. 7 also shows the total incidence

alert levels for *mill* for each of the datasets. It is highest in dataset 1 (because of activation of almost all incidents in correlation chain) and lowest in dataset 4 (because of absence of most detrimental incidents, such as *SystemDistress*). Table 2. denotes a list of the compromised hosts for all the datasets along with their incidence alert levels. The shaded rows show four hosts that we report as compromised but were not listed as compromised according to the DARPA documentation [7]. The attacker tried to compromise these hosts by exploiting the system vulnerabilities but the attempts were unsuccessful. Since the sensor cannot report on the success of these alerts, we justifiably correlate them. However absence of any further activity for these hosts result in low incidence alert level (25.4%) for them. With effective threshold schemes, it is possible to avoid these false positives (e.g., acceptable IAL level with threshold 33%).

In our experiments with the DARPA scenario specific dataset we were able to discover the multi-staged attack scenarios successfully, reduced the alert volume significantly and also reported extent of involvement of the compromised resources in the coordinated attacks.

Table 2. Incidence Assessment for Compromised Hosts

Dataset	# of Compromised Hosts Reported in Correlated Data	Hosts	Incidence Alert Level (IAL)
1	4	172.016.112.010	93.8%
		172.016.112.050	93.8%
		172.016.115.020	93.8%
		172.016.113.148	25.4%
2	7	172.016.112.010	54.3%
		172.016.112.010	54.3%
		172.016.112.050	54.3%
		172.016.113.148	25.4%
		172.016.114.010	25.4%
		172.016.114.020	25.4%
		172.016.114.030	25.4%
3	2	172.016.112.050	75.0%
		172.016.115.020	75.0%
4	1	172.016.115.020	37.6%

5 Conclusion and Future Work

In this paper, we described the use of cognitive modeling of cause and effect relationships to correlate alerts in a distributed environment. We used scenario specific DARPA 2000 dataset for our experimentations as an initial effort to evaluate the effectiveness of our alert correlation model. The results show potential for this simple but effective approach in alert correlation and in incident assessment. The limitations of this approach include inability to deal with unknown alerts and mapping requirement of sensor alert types into generalization hierarchy. A misuse type sensor would miss unknown/unfamiliar attacks and since our knowledge base depends on the sensors' knowledge, so would we. Although our approach requires

knowledge of attack behavior in terms of its impact, the use and encoding of this knowledge is straightforward. We have found FCMs to be particularly suitable in dynamic environment as they are flexible enough to capture adaptive nature of human knowledge. Currently, we are in the process of simultaneously generating multi-sensor data from the DARPA 2000 dataset to demonstrate usefulness of our approach in improving alert correlation by corroborating evidences in multi-sensor environment. Our ongoing research concentrates on developing a unified alert fusion model which will combine alert prioritization, alert clustering and alert correlation in a single framework and can be used to provide a security administrator with a better overall understanding of the health of the system resources. Also, we are developing a model that will be suitable for a high performance computing (HPC) cluster environment where assurance issues must not severely affect performance, which is the essence of HPC environment. In near future, we will experiment in the cluster environment with simulated cluster specific attacks.

References

1. D. Brubaker, "Fuzzy Cognitive Maps," *EDN Access*, Apr. 1996.
2. Internet Security Systems, "RealSecure Network 10/100," http://www.iss.net/products_services/enterprise_protection/rsnetwork/sensor.php (current January 30, 2005).
3. K. Julisch, "Mining Alarm Clusters to Improve Alarm Handling Efficiency," *Proceedings:* 17th Annual Computer Security Applications Conference (ACSAC'01), New Orleans, LA, December 10 - 14, 2001.
4. B. Kosko, "Fuzzy Cognitive Maps," *International Journal of Man-Machine Studies,* vol. 24, 1986, pp. 65-75.
5. B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence,* Prentice Hall, Englewood Cliffs, NJ, 1992.
6. B. Kosko, *Fuzzy Engineering,* Prentice Hall, Upper Saddle River , NJ, 1997.
7. M.I.T Lincoln Laboratory, "2000 DARPA Intrusion Detection Scenario Specific Data Sets," http://www.ll.mit.edu/IST/ideval/data/2000/2000_data_index.html (current January 30, 2005)
8. P. Ning, Y. Cui, and D.S. Reeves, "Constructing Attack Scenarios through Correlation of Intrusion Alerts," *Proceedings: ACM Conference on Computer & Communications Security*, Washington D.C., WA, Nov. 2002.
9. P. Ning, "TIAA: A Toolkit for Intrusion Alert Analysis," http://discovery.csc.ncsu.edu/software/correlator/ current January 30, 2005).
10. X. Qin, and W. Lee, "Statistical Causality Analysis of INFOSEC Alert Data," *Proceedings: Recent Advances in Intrusion Detection,* Pittsburgh, PA, Sep. 2003.
11. A Siraj, S.M. Bridges, and R.B. Vaughn, "Fuzzy Cognitive Maps for Decision Support in an Intelligent Intrusion Detection System," *Proceedings: International Fuzzy Systems Association/ North American Fuzzy Information Processing Society (IFSA/NAFIPS) Conference on Soft Computing*, Vancouver, Canada, Jul. 2001.
12. J.Q. Xin, J.E. Dickerson, and J.A. Dickerson, "Fuzzy Feature Extraction and Visualization for Intrusion Detection," Proceedings: FUZZ-IEEE, St. Louis, MO, 2003.
13. D. Yu and D. Frincke, "A Novel Framework for Alert Correlation and Understanding," Proceedings: International Conference on Applied Cryptography and Network Security (ACNS), Yellow Mountain, China, 2004.

Beyond Keyword Filtering for Message and Conversation Detection

D.B. Skillicorn

School of Computing, Queen's University
skill@cs.queensu.ca

Abstract. Keyword filtering is a commonly used way to select, from a set of intercepted messages, those that need further scrutiny. An obvious countermeasure is to replace words that might be on a keyword list by others. We show that this strategy itself creates a signature in the altered messages that makes them readily detectable using several forms of matrix decomposition. Not only can unusual messages be detected, but sets of related messages can be detected as conversations, even when their endpoints have been obscured (by using transient email addresses, stolen cell phones and so on).

1 Introduction

Groups with evil intent must communicate with one another, and most forms of communication, particularly in real time, can be intercepted. For example, the U.S.A., U.K., Canada, Australia and New Zealand use the Echelon [3] system to intercept radio, landline, and internet communication globally; several other countries have admitted the existence of similar, if smaller scale, systems. Encryption is the most obvious way to conceal the content of messages but, in many settings, the fact of encryption itself draws attention to the existence of the message. Security by obscurity is often a better choice – an extremely large number of messages are transmitted each day, so it is difficult to find the few needles of interest in the haystack of ordinary traffic. This is all the more so if the endpoints of a message (email addresses, phone numbers) are not connected to the physical structure of the terrorist organization, or are changed frequently.

We consider the problem of detecting messages between members of such groups, and of connecting these messages together into conversations when the usual endpoint markers are not present. Identifying such messages enables them to be read; putting them together enables higher-level structure within the group (e.g. command and control) to be determined, and also enables traffic analysis to be applied.

The fact that distinguishes messages among members of such groups from ordinary messages is that they must respond to the existence of keyword filtering. An awareness of the existence of keyword filtering is not the same as knowing which keywords are on the list. Although many likely words could probably be guessed, it is hard to know whether less obvious words are on the list (e.g.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 231–243, 2005.

'fertilizer'). Hence messages must avoid the use of words that *might* be on the list; since these messages must be about something, substitutions of other words must be made instead. We hypothesize that these substitutions themselves become a signature for messages that deserve further analysis, and this turns out to be the case.

There are two choices for the replacement strategy for words that should not be used. The first is to replace them with other words or locutions chosen at random. The frequency with which such words appear in messages, and especially in conversations, will differ from the their natural frequency and this discrepancy is sufficient to make such messages detectable. Fortunately, the presence of correlated messages using words, even rare words, with their natural frequency does not produce false positives; and nor does the presence of individual messages with unusual word frequency, so individuals with an idiosyncratic writing style do not cause false positives either.

The second choice for the replacement strategy is to replace words with other words of similar frequency. This strategy has a number of problems. First, it is extremely difficult in real time (phone calls), or near real time (email). People are not able to judge reliably how common a particular word is; the word 'nuclear' is the 1266th most common word in English (`www.fabrica.it/wordcount/main.php`), and words similar to it in frequency are 'pupils', 'meaning', 'increasing', and 'reach', several of which might plausibly have been guessed to be more frequent. Second, an alternative strategy such as using a codebook (a list of words of the right frequency to be used as substitutes) is hard to use in real time, particularly under pressure, and requires distribution and security for the codebook. Third, the use of an online resource, such as the web site above, and a standard offset (use the word 5 places down the list from the one you want to avoid) also creates patterns that may be detectable.

In this paper, we show how two matrix decompositions, singular value decomposition, and independent component analysis can be applied to message-word and message-rank matrices to detect messages with anomalous word usage. These techniques therefore complement keyword filtering, since the more messages are altered to avoid detection by keyword filtering the more likely they are to be detected by these other techniques.

2 Related Work

Messages for which senders and receivers are known are naturally regarded as a directed graph. Several different technologies have been used to understand such graphs. Social network analysis (or link analysis) examines properties of such graphs that reflect the position or power of the nodes (representing individuals), for example [2,4,9]. In an early paper, Baker and Faulkner [1] showed that the positions of individuals in a network of price fixing in the electrical industry were predictive of, e.g., sentence in criminal proceedings. Others have used patterns in email to determine group structure, e.g. [12], or to determine topics of mutual interest [11].

3 Preliminaries

The distribution of words in English (and most other languages) is highly non-uniform. The commonest words are extremely common, but frequency drops very quickly, and most words are quite uncommon. The actual distribution is a Zipf distribution. There are a number of explanations for this phenomenon, ranging from those that invoke deep language and cognitive properties, to those based solely on the increasing number of possible words with length [10]. The Zipf frequency distribution applies not only to all words, but also to individual parts of speech such as nouns.

We consider the word frequency distributions of messages as a kind of signature that can be used to compare messages. The high-frequency words are not useful discriminators – almost all messages will contain the word "the", the most common word in English. On the other hand, the low-frequency words are not helpful discriminators either, because almost all messages contain none of them. We will concentrate our attention on words of medium frequency, where we might expect differences of style and content to be most obvious.

It would probably also be useful to consider the digram word frequency structure – how often pairs of words appear adjacently in the message. This is quite difficult to exploit. The difference in likelihood between "I will deliver the bomb" and "I will deliver the unicorn" is quite obvious, but both sentences are grammatical, and it requires a reasonably deep semantic model to decide that the second is quite unlikely in almost any context.

A deeper difficulty is that the adjacency graph of English sentences has a small world property [7]: if we arrange words in concentric layers of frequency, and connect them by edges when they are used adjacently, there are no long paths away from the center. Rather sentences that use unusual words are formed by a sequence of short paths that go out far from the center and return almost immediately. There is probably some mileage to be obtained from consideration of longer units, but it seems difficult. Hence we restrict our attention to single-word usage patterns.

Linguists differentiate written and spoken utterance because they tend to have different properties. The context, and the ability to edit written text allows a more formal, more grammatical style to be used in written messages. By contrast, spoken messages are generated in real-time, cannot be edited, and hence allow a more informal style that is frequently ungrammatical, and contains many speech artifacts (ums, ers). Email tends to fall somewhere between these two extremes; although the opportunity exists to edit emails, anecdotal evidence suggests that this is rarely done.

4 Matrix Decompositions

Matrix decompositions express a matrix, representing a dataset, in a form that reveals aspects of its internal structure. Different matrix decompositions impose different requirements on the structure of the decomposition and so reveal differ-

ent structures. A typical matrix decomposition allows a matrix A to be expressed as a product

$$A = C\ F$$

where, if A is $n \times m$, the matrix C is $n \times r$ and F is $r \times m$. Sometimes a third, diagonal $r \times r$ matrix is also part of the decomposition. A natural interpretation of such a decomposition is *geometric*, and interprets the rows of F as axes in a transformed space, and the rows of C as coordinates in this space.

We will use two matrix decompositions: Singular value decomposition (SVD) [5] for which

$$A = U\ S\ V'$$

where U and V are orthogonal, and S is diagonal with non-increasing entries; and Independent Component Analysis (ICA) [6] for which

$$A = W\ H$$

where the rows of H are statistically independent.

SVD has the property that the first new axis is aligned along the direction of maximal variation in the data, the second axis along the direction of remaining maximal variation, and so on. Each axis is orthogonal to the others, so the 'factors' corresponding to each axis are linearly independent. The truncated representation for any $k \leq m$ is the most faithful possible in that number of dimensions. A useful property of SVD is that it transforms correlation in the original data into proximity in the transformed space. Fast algorithms for computing the SVD of a sparse matrix (with complexity proportional to r times the number of non-zero entries in A) are known.

A particularly useful property of SVD is that distance of a point from the origin in the transformed space (even when the number of dimensions is reduced by truncation) represents how interesting the point is in the sense of how strongly it is correlated with all of the other points. Hence points far from the origin are most anomalous, while those close to the origin are least anomalous. Both pieces of information can be useful.

ICA is similar to SVD but selects factors (rows of H) that are statistically independent. Typically, these factors do not have a natural ordering on them, as those of SVD do.

5 Datasets

We use artificial, but plausible, datasets for our experiments. We assume that messages have been processed to generate a frequency histogram giving the number of occurrences of each word of some (potentially large) dictionary in each message. We use message-word matrices, in which each row represents a message, each column a word (with the columns arranged in decreasing order of natural word frequency), and the entries are the frequencies of each word in each message.

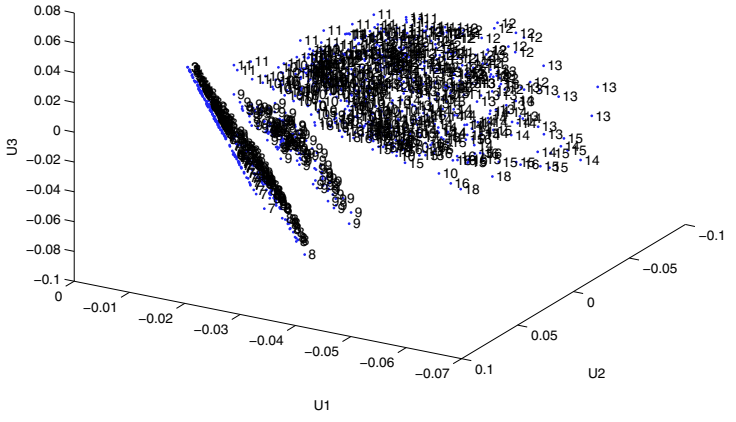


Fig. 1. SVD of example matrix with messages labelled by their length

The ijth entry of such a dataset is generated by sampling from a Poisson distribution with mean $f * 1/j + 1$, where f is a parameter that allows the overall frequencies to be altered, and the $1/j$ term decreases the probability of the occurrence of a word depending which column represents it (so that inherently infrequent words, supposed to be represented by the later columns, are unlikely to appear in any given document). This approximates the Zipf distribution.

A dataset with 1000 documents and 400 words, and $f = 3$ has about 16000 non-zero entries (4% sparse) and each document contains about 20 distinct words. This dataset is a reasonable representation of, say, the nouns in a collection of 1000 messages. We discard the first 200 columns of the dataset in most experiments, since the commonalities in the usage of these common words tends to obscure more interesting connections, and since it mimics the structure of real emails more closely.

We also generate the corresponding message-rank matrix whose rows are messages, and whose entries are ranked lists of words appearing in each message. So if a message contains the ith, jth and kth most frequent words in English, then the row contains the entries i, j, and k. Since messages contain different numbers of words, the length of the rows of this matrix are different, so we pad the right hand end of rows with 0s. The typical width of this matrix is 25–26, the number of words in a typical message in the message-word matrix. Notice that frequency information is lost in this representation.

Message-word matrices are common in information retrieval and their behavior with respect to SVD is well understood. In contrast, message-rank matrices have not received much attention, so we briefly describe the effect of SVD on such datasets. First, a dataset in its original high-dimensional space is extremely curved since each new rank is in a new dimension. This curved structure is, of course, preserved by SVD even in low dimension. Second, the entries in each

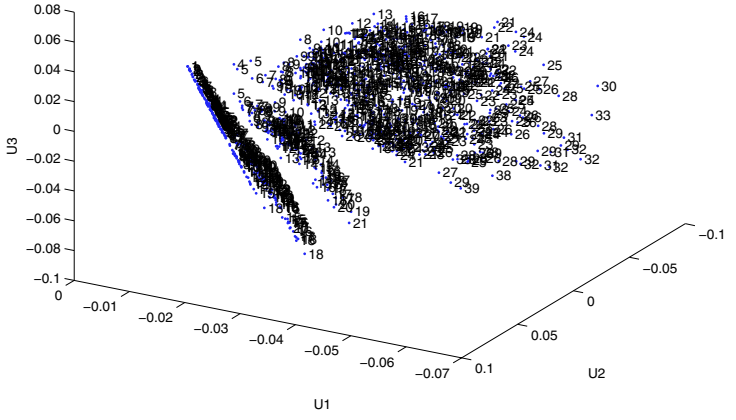


Fig. 2. SVD of example matrix with messages labelled by the average natural rank of the words they contain

row are strictly increasing (until there are no further words). A short message has only a few words, and hence only a few steps, and is unlikely to contain any words of high rank (magnitude). A longer message contains more steps and is more likely to reach words of higher rank.

Figure 1 shows a plot of the SVD of the example matrix above, with the messages labelled by their length (the number of distinct words they contain). Figure 2 is the same plot with the messages labelled by the average rank of the words they contain. The shape that arises from a typical Zipf-like distribution is a fan shape, where position around the fan corresponds to message length, and distance from the center to average rank, which in turn is a measure of how unusual overall the words in the message are. There is some correlation between these two factors because short messages have fewer chances to use unusual words.

6 Experiments

We begin with a dataset constructed as above, with a further 10 similar rows added to it (representing messages that we wish to detect). None of the techniques we use rely on the ordering of the rows so, without loss of generality, we can make these the last ten rows. In all of these matrix decompositions, we discard the first 200 columns of the dataset since the (mostly spurious) correlations among common words obscure the more interesting structure of the less-frequent words.

6.1 Correlated USE of Words with Unnatural Frequency Is Detectable

The first dataset illustrates a set of messages with correlated use of unusual words. A block of size 10 rows by 6 columns with uniformly random 1's and 0's

is generated and added to the dataset at rows 1001 to 1010 and columns 301 to 306. Each message therefore has an overlap of 2–3 words with every other message in the group. The natural frequency of occurrences of these words is around 1% so the use in these messages is well above the background usage. This dataset represents a typical scenario in which several objects are being discussed in messages, but different, less frequent words, are being used in place of the object names.

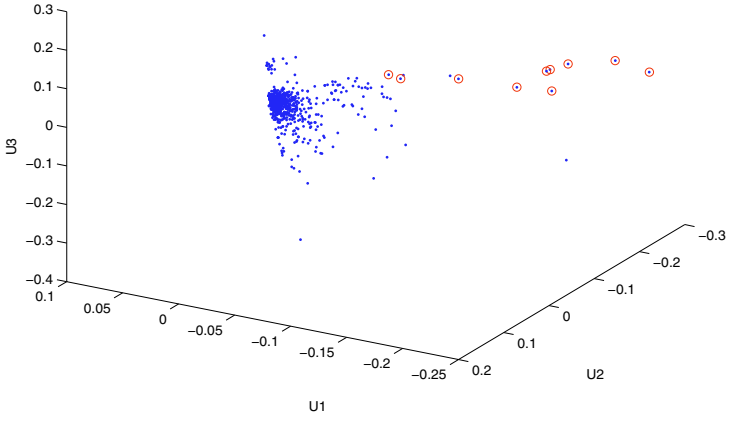


Fig. 3. The 3-dimensional plot of messages for a dataset containing correlated unusual word usage

Figure 3 shows such a plot using the first 3 columns of the U matrix from the decomposition. The messages in the correlated group are marked with circles. It is easy to see how they are separated from the other messages in the first and second dimensions. Figure 4 shows how an unusual group of messages shows up as a set of points in unusual positions using the first 3 columns of the W matrix.

6.2 Correlation with Typical Frequencies Is not Enough to Be Detectable

We now show that both correlation and unusual frequency are required in order to form detectable groups of related messages. We first add to the base dataset a block of correlated messages whose frequencies are natural. To do this, we generate a block of 5 rows by 6 columns and place non-zero entries in it with frequencies appropriate to columns 301–306 of the base dataset. We then insert this block twice, at rows 1001–1005 and 1006–1010. Because there is only approximately a 1% chance of a word of this frequency being used in a message, such blocks may contain very few 1's. However, even when $f = 30$ is used, so

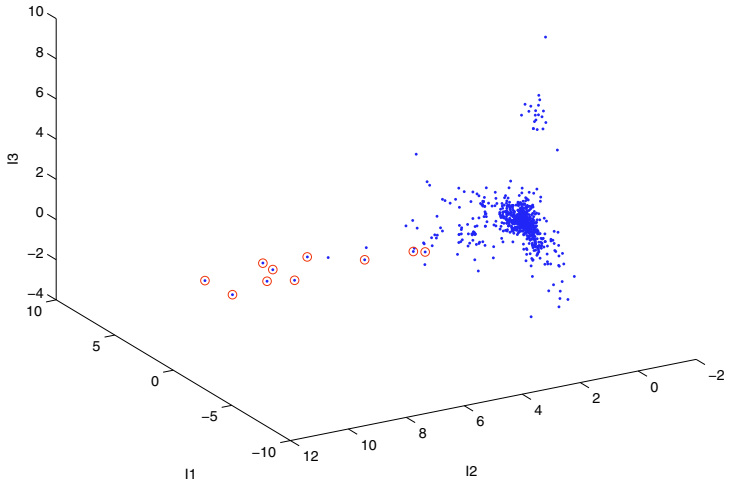


Fig. 4. The 3-dimensional plot from ICA of a dataset containing correlated unusual word usage

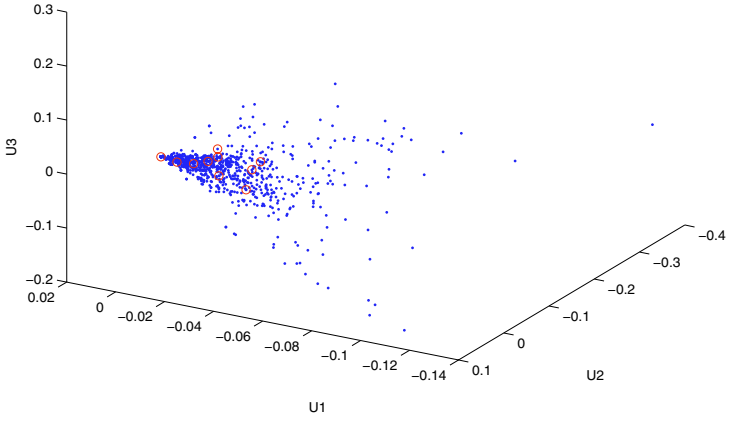


Fig. 5. 3-dimensional plot from SVD for a dataset with correlation but typical frequencies

that there are a significant number of 1's in the repeated block, no structure is seen.

Figure 5 shows that the points corresponding to rows 1001–1010 are not separated from the main mass of points. Figure 6 shows that ICA does not see any structure related to this group in the first 3 dimensions.

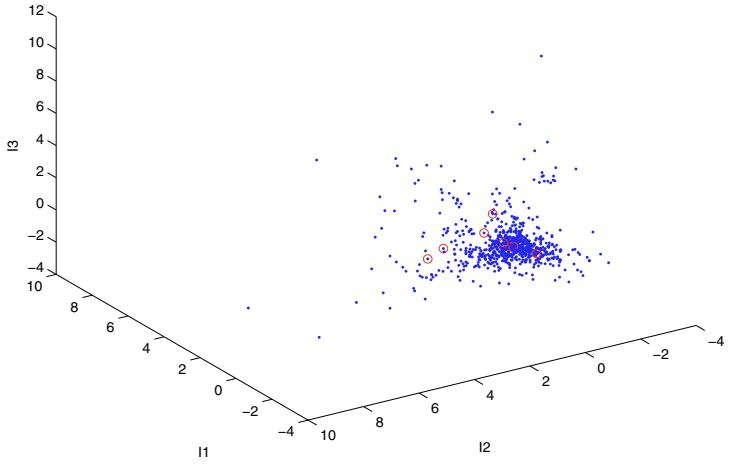


Fig. 6. 3-dimensional plot from ICA for a dataset with correlation but typical frequencies

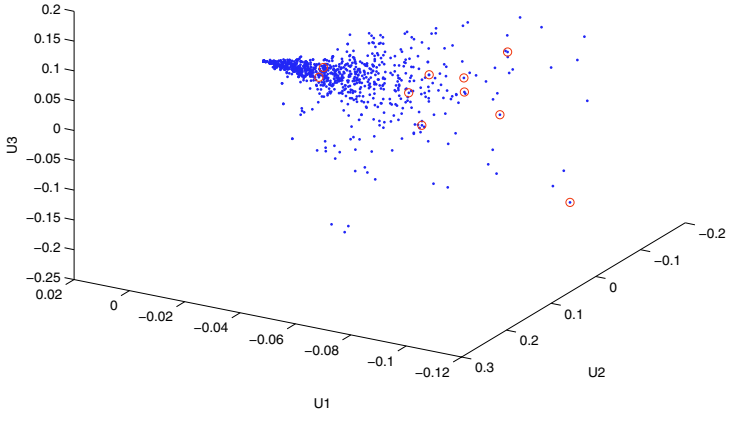


Fig. 7. 3-dimensional plot from SVD for a dataset with unusual word frequencies but not correlation

6.3 Unusual Frequency Is not Enough to Be Detectable

We now consider a dataset where unusual words uses are present but they are not correlated. We generate 10 independent vectors of size 1 by 6 with a uniform distribution of 0's and 1's (as in the first dataset) and then place each of these vectors in non-overlapping columns starting from column 280. The resulting dataset therefore has 10 final rows in which rare words are used with much greater than their natural frequency.

As expected Figure 7 shows the points corresponding to these messages scattered all over the plot. Notice, though, that several of these points are far along

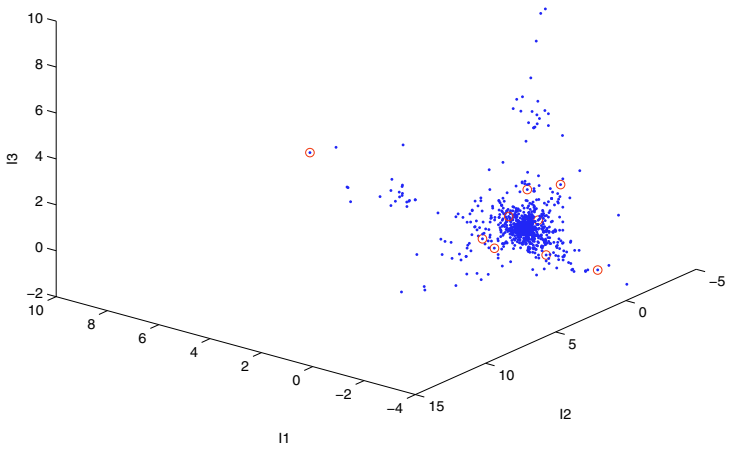


Fig. 8. 3-dimensional plot from ICA for a dataset with unusual word frequencies but not correlation

the $U1$ axis as a result of the unusual word usage they contain. Similar results for ICA can be seen in Figure 8.

6.4 Choosing Words with Similar Frequencies Is Detectable

We now consider what happens when an attempt is made to replace keywords by words of similar frequency. We generate a dataset as before, but add ten extra rows, each of which is a copy of an arbitrary row (row 1) shifted by 1,2, and so on. Hence these extra rows represent an attempt to send the message described by row 1 using other words of similar frequency.

In an SVD plot of this data, the messages are not particularly distinctive, since almost all of the time one message will have a zero value in a column in which another message has a positive value. Figure 9 shows the SVD plot of the ranked version of the matrix. Now the relationship between these messages (and their relationship to message 1) is much clearer. In the message-rank matrix, the rows corresponding to these messages have a similar profile, so that they are visibly correlated. As messages are shifted to the right, they contain less frequent words, so that their average word rank increases, moving them farther from the origin.

There are a number of potential ways to exploit the ability of SVD applied to the message-rank matrix to see structure. First, it shows the futility of a scheme whereby different participants in a conversation might use different shifts to avoid keywords – the commonality of structure reveals the conversation anyway. Second, the role of message 1 could be played by inserting an artificial message using a selection of keywords. Messages that had similar content but were attempting to hide by using shifting would then show up as neighbors in the SVD plot. Third, message 1 could be a previous message from a member of the

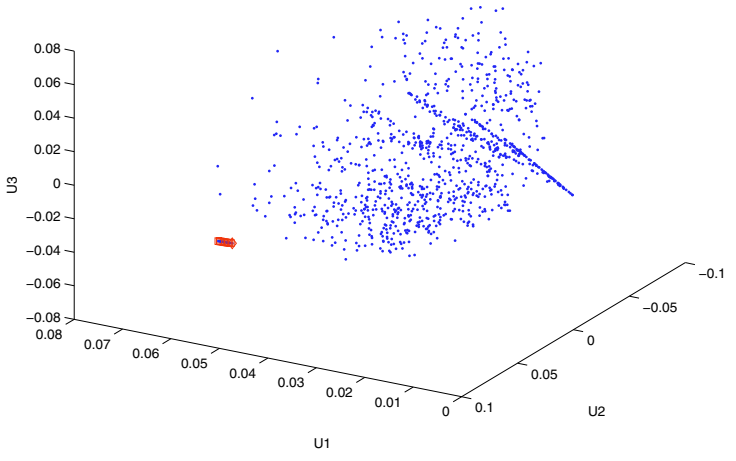


Fig. 9. SVD plot of message-rank matrix with words of similar frequency substituted (copies shifted by increasing amounts marked by boxes; original message marked by a diamond)

group. Messages with similar word ranking structure would once again show up as neighbors in the SVD plot.

6.5 Artificial Word Ranks

In the word-rank matrix we have assumed that the natural word frequency was used. SVD plots messages whose average word rank is large at positions far from the origin. Hence the distance from the origin is a surrogate for the unusualness of each message as an example of English text.

However, we assume that there is a list of keywords that are of more interest than other words. In a message-word matrix, we could add extra weight to keywords, and this would have the effect of moving messages that contained these keywords farther from the origin (and hence of making them more visible in a plot). This strategy does not pay off, however, because we are assuming that such words will not be used by those who want to conceal their conversations.

The use of the message-rank matrix allows a generalization of the idea of a keyword list, simply by reordering the ranking so that it reflects the usefulness of keywords for detection rather than their natural rank in English. We want instead to generate a ranking of words so that the most uninteresting words are at the beginning, and the words become more interesting farther down the list. Now there is no boundary between keywords and non-keywords, only a matter of degree.

Messages that contain words considered useful will now appear far from the origin in the SVD plot; and indeed distance from the origin can be used as an ordering on the interestingness of messages. Furthermore, the use of an SVD on such a message-rank matrix also extracts correlations among messages based on

similar word use. Hence this approach has both the benefits of keyword filtering and of correlation-based analysis.

Although we have shown results only for a particular base dataset and particular modifications to it, the results shown here are typical of similar datasets. Although as a matter of practicality, the datasets are small, they are not unreasonable as examples of email or phone conversations collected over a short period of time.

Notice that only a few dimensions of the decomposition are needed to give good results. Hence the complexity of these matrix decompositions is quite practical since the data matrices are sparse – the complexity is effectively linear in the number of messages considered.

7 Conclusion

We have shown that matrix decomposition techniques, applied to message-word and message-rank matrices, are a complement to standard techniques such as keyword filtering that are used to select an interesting subset from the flood of global messages. Attempts to defeat keyword filtering by different kinds of substitution strategies create signatures in individual messages, and especially in related messages, that can be detected. In particular, substitution of words by other words of different natural frequency in a set of messages creates an easily detectable pattern. Substitution of words of similar natural frequency makes detection more difficult, but the use of message-rank matrices can help. The Enron email dataset is a large set of real emails made available by the U.S. Department of Justice. We have begun to investigate the word use properties of this dataset to see how much actual word use agrees with the theoretical behaviors suggested here [8].

References

1. W.E. Baker and R.B. Faulkner. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review*, 58:837–860, December 1993.
2. T. Coffman, S. Greenblatt, and S. Marcus. Graph-based technologies for intelligence analysis. *CACM*, 47(3):45–47, March 2004.
3. European Parliament Temporary Committee on the ECHELON Interception System. Final report on the existence of a global system for the interception of private and commercial communications (echelon interception system), 2001.
4. L. Garton, C. Haythornthwaite, and B. Wellman. Studying online social networks. *Journal of Computer-Mediated Communication*, 3(1), 1997.
5. G.H. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.
6. A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.
7. R. Ferrer i Cancho and R.V. Solé. The small world of human language. *Proceedings of the Royal Society of London Series B – Biological Sciences*, pages 2261–2265, 2001.

8. P.S. Keila and D.B. Skillicorn. Structure in the Enron email dataset. In *Third Workshop on Link Analysis, Counterterrorism and Security, SIAM International Data Mining Conference*, pages 55–64, 2005.
9. V.E. Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.
10. W. Li. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEETIT: IEEE Transactions on Information Theory*, 38(6):1842–1845, 1992.
11. R. McArthur and P. Bruza. Discovery of implicit and explicit connections between people using email utterance. In *Proceedings of the Eighth European Conference of Computer-supported Cooperative Work, Helsinki*, pages 21–40, 2003.
12. J.R. Tyler, D.M. Wilkinson, and B.A. Huberman. Email as spectroscopy: Automated discovery of community structure within organizations. HP Labs, 1501 Page Mill Road, Palo Alto CA, 94304, 2003.

Content-Based Detection of Terrorists Browsing the Web Using an Advanced Terror Detection System (ATDS)

Yuval Elovici[1], Bracha Shapira[1], Mark Last[1], Omer Zaafrany[1], Menahem Friedman[2], Moti Schneider[3], and Abraham Kandel[4]

[1] Department of Information Systems Engineering, Ben-Gurion Univ. of the Negev, Beer-Sheva 84105, Israel
elovici@inter.net.il, {bshapira, mlast, zaafrany}@bgu.ac.il
[2] Department of Physics Nuclear Research Center – Negev Beer-Sheva, POB 9001, Israel
mlfrid@netvision.net.il
[3] School of Computer Science Netanya Academic College Netanya, Israel
motis@netanya.ac.il
[4] Department of Computer Sc. and Engineering, Univ. of South Florida Tampa, FL, USA
kandel@csee.usf.edu

Abstract. The Terrorist Detection System (TDS) is aimed at tracking down suspected terrorists by analyzing the content of information they access. TDS operates in two modes: a training mode and a detection mode. During the training mode TDS is provided with Web pages accessed by a normal group of users and computes their typical interests. During the detection mode TDS performs real-time monitoring of the traffic emanating from the monitored group of users, analyzes the content of the Web pages accessed, and issues an alarm if the access information is not within the typical interests of the group. In this paper we present an advanced version of TDS (ATDS), where the detection algorithm was enhanced to improve the performance of the basic TDS system. ATDS was implemented and evaluated in a network environment of 38 users comparing it to the performance of the basic TDS. Behavior of suspected terrorists was simulated by accessing terror related sites. The evaluation included also sensitivity analysis aimed at calibrating the settings of ATDS parameters to maximize its performance. Results are encouraging. ATDS outperformed TDS significantly and was able to reach very high detection rates when optimally tuned.

1 Introduction

Due to the availability and publishing ease of information on the Web, terrorists increasingly exploit the Internet as a communication, intelligence, and propaganda tool where they can safely communicate with their affiliates, coordinate action plans, raise funds, and introduce new supporters into their networks [1, ,9, 10]. Governments and intelligence agencies are trying to identify terrorist activities on the Web in order to prevent future acts of terror [11]. Thus, there is a need for new methods and technologies to assist in this cyber intelligence effort.

By means of content monitoring and analysis of Web pages accessed by a group of Web surfers, it is possible to infer surfers' areas of interest [2]. Using this approach, a

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 244 – 255, 2005.

real time Web traffic monitoring could be performed to identify terrorists and their supporters as they access terrorist-related information on the Internet [2, 5].

In this paper an Advanced Terrorist Detection System (ATDS) is presented aiming at tracking down suspected terrorists by analyzing the content of information they access. The system operates in two modes: the training mode activated off-line, and the detection mode operating in real-time. In the training mode ATDS is provided with Web pages of normal users from which it derives their normal behavior profile by applying data mining (clustering) algorithms to the training data. In the detection mode ATDS performs real-time monitoring of the traffic emanating from the monitored group of users, analyzes the content of the Web pages they access, and generates an alarm if a user accesses to information that is not expected from a normal user i.e., the content of the information accessed is "very" dissimilar to usual content in the monitored environment. The design goals behind the development of ATDS were:

1. Detecting terrorist activities based on their Content – ATDS should be able to detect terrorist related activities by monitoring the network traffic content. ATDS should focus only on network traffic content containing textual HTML pages. In order to achieve this goal ATDS has to be able to disregard irrelevant network traffic content.
2. On-line detection – ATDS should be able to detect on-line suspected terrorists accessing terrorist related content. Such on-line detection may enable law enforcement agencies to arrest suspected terrorists accessing the Web through public infrastructure such as public computer labs in a university campus or Internet cafés. The detection result should include the suspected terrorist IP address. The connection between IP and the real user identity is beyond the scope of this design.
3. Detection should be based on passive eavesdropping on the network – ATDS should monitor the network traffic without being noticed by the users or the monitored infrastructure provider. Passive eavesdropping can be achieved by a network snifter or by installing an agent in the users' computers acting as a proxy.
4. Profile accuracy – ATDS should monitor and collect several HTML pages for each monitored IP. The number of HTML pages that will be included in the detection process per IP should be a system parameter.
5. Detection threshold- ATDS should include two threshold parameters: The first parameter should indicate the required level of similarity between a user HTML page and one of the information interests of the normal users. The second parameters should control the number of collected pages per IP that should be similar to the normal users profile in order to consider the user as normal.

In this paper we present the advanced version of the basic TDS [2], ATDS. The main difference between TDS and ATDS is the replacement of the naive detection algorithm with an advanced one that can be tuned to maximize performance.

The reminder of the paper is organized as follows. In Section 2, a brief review on the content-based methodology for anomaly detection on the Web is presented which is the basis of ATDS. In Section 3, ATDS architecture is described in detail. The experiments conducted to evaluate the feasibility and performance of ATDS is discussed in Section 4. Section 5 concludes the paper with presenting future research issues.

2 Content-Based Methodology for Anomaly Detection: A Review

Anomaly detection relies on models of the intended behavior of users and applications and interprets deviations from this `normal' behavior as evidence of malicious activity [17,16,18]. The underlying intuitive assumption of this model is that content of users browsing reflects their interests. This assumption is the basis of many personalization models, algorithms and systems [15] that generate user "profiles" from the content of pages they browse. A stronger assumption is that users that have similar interests can be identified by the content of their browsing activities and represented as a group of users or stereotypes. Individual user profiles can then be compared to those stereotypes to identify whether they relate to any of these groups.

The new behavior-based anomaly detection model (implemented first in TDS [2, 5] and then in ATDS) uses the content of web pages browsed by a specific group of users as an input for detecting abnormal activities. In this study, we refer to the *textual* content of web pages only, excluding images, music, and other complex data types.

The basic idea is to maintain information about the interests of "normal" users in a certain environment (such as a campus), and detect users that are dissimilar to the normal group under a defined threshold of similarity.

The model has two phases, the learning phase and the detection phase:

1. The learning phase – during which the Web traffic of a group of users is recorded and transformed to an efficient representation for further analysis. We use the vector representation [14], i.e. each page is transformed to a vector of weighted terms. The learning phase is applied in the same environment where the detection would later be applied in order to learn the "normal" content of users in the environment. The collected data is used to derive and represent the group's areas of interest by applying a data-mining technique (cluster-analysis).
2. The detection phase – The detection phase is aimed at detecting abnormal users. Users dissimilar to the normal users are detected. The detection is performed by transforming the content of each page accessed by a user to a vector representation that can be compared against the representation of the groups for dissimilarity. The minimum number of suspicious accesses required in order to issue an alarm about a user are defined by the detection algorithm. The detection is performed on-line and should therefore be efficient and scalable. In the following sub-sections, we briefly describe the learning and the detection phases. The reader is referred to [2,5,6] for further details

2.1 Learning Phase

The learning phase, graphically described in Figure 1, generates a DB including representations of the interests of the monitored group of users. During the detection phase users accessed pages are compared to this DB.

The learning phase consists of the *Filter*, *Vectors-Generator* and the *Clusters-Generator* components. Following is a brief description of the functionality of these components: Each page of the training data is sent to the *Filter* for exclusion of non-appropriate pages, i.e. non-textual pages, and for omitting of images and tags related to the content format from the appropriate pages. The filtered pages are sent to the

Vectors- Generator component that transforms each page to a weighted terms vector. The vector entries represent terms and their values represent the importance of the term to the page defined by the relative frequency of the term to the page, and by other factors such as the position of the term in the page. The vectors are recorded for the clustering process that follows.

The *Clusters-Generator* module (Fig. 1) receives the vectors from the *Vectors-Generator* and performs cluster-analysis on them. A Cluster-Analysis process receives as input a set of objects with attributes and generates clusters (groups) of similar objects, so that objects within a cluster have high similarity and objects between groups are dissimilar. The objects in this model are the pages, where the attributes are the terms. The *n* clusters generated by the *Cluster-Generator* represent the *n* areas of interest of the defined group of users. The optimal *n* is defined by the clustering algorithm. For each cluster, the *Group-Representor* component computes a central vector (centroid), denoted by Cv_i for cluster *i*. In our model each centroid represents one area of interest of the group of users. The learning phase is a batch process and should be activated regularly to update the representation of the group.

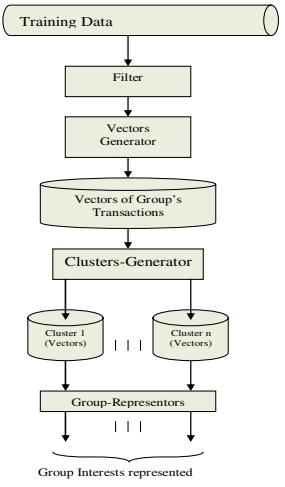


Fig. 1. The Learning phase of the model

2.2 The Detection Phase

During the detection phase, a group of computers in a certain environment is monitored to detect abnormal users. The content of transactions of the group members is collected on-line and compared to the set of pre-defined centroids (generated during the learning phase). The detection phase consists of the *Sniffer*, *Filter*, the *Vector – Generator* and the *Detector* which is the main component of this phase. The *Sniffer* captures the pages that users, (identified by their IPs), access at the network layer, and sends each page to the *Filter* for further processing. The *Filter* and the *Vector-Generator* have the same functionality as in the learning phase. The Detector receives the vectors (and their respective IPs) and decides whether to issue an alarm. The *De-*

tector measures the distance between each vector and each of the centroids representing an area of interest to the user. While in TDS the detection algorithm considers only one vector of the last user access, in ATDS, detection is based on the user's access history. The detection algorithm can be calibrated with certain parameters to fine-tune the detection. Some of the parameters are: 1) The dissimilarity threshold, 2) the minimum number of abnormal accesses by the same IP that would issue an alarm and, 3) the time frame of suspicious accesses by the same IP that would issue an alarm.

The detection algorithm issues an alarm when all the parameters are satisfied. The alarm consists of the suspicious IPs along with the data that caused the issuing of the alarm. The similarity between the vectors accessed by users and the centroids is measured by the Cosine of the angle between them. An access is considered abnormal if the similarity between the access vector and the nearest centroid is lower than the threshold denoted by *tr*. The following is the Cosine equation used:

$$Min\left(\frac{\sum_{i=1}^{m}\left(tCv_{i1}\cdot tAv_i\right)}{\sqrt{\sum_{i=1}^{m}tCv_{i1}^2\cdot\sum_{i=1}^{m}tAv_i^2}},\ldots,\frac{\sum_{i=1}^{m}\left(tCv_{in}\cdot tAv_i\right)}{\sqrt{\sum_{i=1}^{m}tCv_{in}^2\cdot\sum_{i=1}^{m}tAv_i^2}}\right)<tr \quad (1)$$

where Cv_i is the *i*-th centroid vector, Av - the access vector, tCv_{i1} - the *i*-th term in the vector Cv_i, tAv_i - the *i*-th term in the vector Av, and *m* – the number of unique terms in each vector.

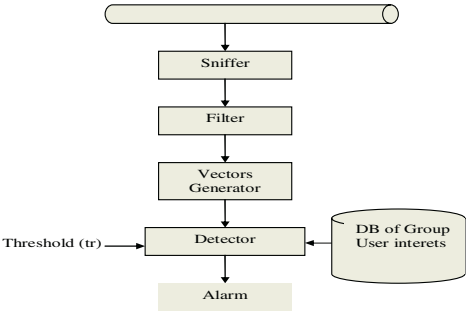


Fig. 2. The Detection Phase

3 ATDS Architecture

ATDS architecture is based on the content-based model for anomaly detection described in section 2. It includes four components (see Figure 2): 1) Online-HTML Tracer (OHT) – including the Sniffer and the Filter, 2) the Vector Generator, 3) the Cluster Generator, and 4) the Detector.

On-line-HTML Tracer. The On-Line HTML Tracer consists of the Sniffer and the Filter modules. The Sniffer is implemented using the WinPickup software tool [7]. The packets intercepted by the Sniffer are sent to the Filter module that scans the content of each intercepted packet and filters out packets not containing textual HTML pages. The remaining packets are reconstructed into HTML pages and sent to the Vector Generator.

Vector Generator. The Vector Generator component is implemented using the Extractor software tool [3] that receives an HTML page and generates a vector of up to 30 weighted key phrases scored by their importance to the page.

Cluster Generator. The Cluster Generator is based on the Vcluster program from the Cluto Clustering Tool [13] using the 'k-way' clustering algorithm (a version of k-means). The similarity between the objects is computed using the cosine measure [14]. The Cluster Generator applies clustering to all the vectors in the DB and generates a centroid vector for each of the clusters. The output is a set of centroids representing the areas of interests of the 'normal' users, or the "normal user behavior".

Detector. The Detector receives as input a new captured page represented as a vector along with its IP address. The Detector applies a similarity check (performed by the Similarity Checker) measuring Cosine similarity between the new incoming vector and the set of clusters representing the normal user's profile. If the similarity is above the predefined threshold, the new incoming vector is classified as normal. Only when none of the clusters is similar to the incoming vector, it is tagged as abnormal. The vector and its classification (normal or abnormal) are recorded into the queue of the specific IP Address in the Vector Sub-Queues data structure. The next step is performed by the Anomaly Finder which scans the sub queue of a certain IP Address and counts the vectors classified as abnormal in the queue, if the number is above the alarm threshold, the Detector generates an alarm implying that the user at the specific IP is accessing abnormal information. In this paper we describe experiments that examine the effect of tuning the queue size and the alarm threshold on detection quality.

4 Evaluation

To evaluate the feasibility and the performance of the model, we implemented it by developing ATDS. Following are some details on ATDS implementation:

ATDS was implemented using Microsoft Visual C++ and designed in a modular architecture. The computational resources and the complexity of the methodology required a careful design to enable real-time on-line sniffing.

The system was deployed on an Intel Pentium 4 2.4 GHz server with 512 MB RAM. The server was installed in the computation center of Ben-Gurion University (BGU) and was configured to monitor 38 stations in teaching labs at the Department of Information System Engineering (ISE) at BGU. The stations consisted of Pentium 4, 2.4GHz with 512MB of RAM, 1Gbps fast Ethernet Intel adapter and Windows XP professional operating system. The access to those stations is allowed only to students of the one department (Information Systems Engineering).

4.1 Data Preparation for the Simulations

We prepared data (Web pages) for the learning phase of the model and other data for the detection phase which included accesses to normal and abnormal (terror-related) content to represent "normal" and "abnormal" users.

For the learning phase we collected 170,000 students' accesses to the Web in the teaching labs during one month. The students were aware and agreed to have their accesses collected anonymously. After exclusion of non-English pages (ATDS currently does not handle them) the "normal" collection included 13,300 pages. We then ran the Filtering and Vector Generator processes on the 13,300 pages and received a 13,300 * 38,776 matrix (i.e., 38,776 distinct terms representing the 13,300 pages).

In the detection phase, we collected 582 terror-related pages for the simulation of abnormal accesses. We applied Filtering and Vector Generation to these pages resulting with a set of vectors representing terror related sites. We also randomly selected 582 vectors from the normal matrix for the simulation of normal users accessing only normal pages and abnormal users (terrorists) accessing a mix of normal and terror-related pages.

4.2 Evaluation Objectives and Measures

We evaluated ATDS performance by ROC (Receiver Operator Characteristic) curves using the following measures (based on [12]) :

True Positive (TP) (also known as Detection Rate or Completeness or hit rate) is the percentage of alarms in case of abnormal (positive) activity.

$$Tp = \frac{postive\_corrected\_classified}{total\_number\_of\_positives} \tag{2}$$

False Positive (FP) is the percentage of false alarms when normal (negative) activity is taking place.

$$Fp = \frac{negative\_incorrecctly\_classified}{total\_number\_of\_negatives} \tag{3}$$

ROC graphs graphically represent the tradeoff between the TP and FP rates for every possible cutoff. Equivalently, a ROC curve represents the tradeoff between sensitivity and specificity. In a ROC graph the x axis represents the FP rate while the y axis represents the TP alarm rate. A point on the graph represents the FP/TP for a specific similarity threshold. The experiments aimed at examining the following issues:

- Feasibility of the model, i.e., the system's ability to perform on-line sniffing and accurate Web pages reconstruction.
- The effect of the following system's parameters on its performance, (i.e., on the values of TP and FP):
- The size of the sub queue for each IP, i.e., the number of accessed pages kept for each IP used by the detection algorithm as the IP history.
- Alarm thresholds values – the threshold of the ratio between the number of accesses detected as abnormal and the total number of accesses in the user queue for which an alarm is issued.
- Number of clusters representing the normal profile of the monitored users.

4.3 The Experiment

During our experiment we ran three simulations:

The **first one** was run in order to evaluate the monitoring capabilities of OHT, which is crucial to the feasibility of ATDS. During the **second and the third** simulations we tested the feasibility of ATDS as well as the effect of system's parameters on the FP and TP rates running simulations with a range of values of the system's parameters, as detailed below.

First Simulation. In order to examine the ability of OHT to monitor and intercept all HTML pages accessed by users in the experimental environment, we ran one simulation measuring the percent of lost HTML pages during the capturing and reconstruction processes. We simulated heavy traffic to and from the Web, monitored the traffic, and compared the number of captured pages with the number of pages actually received by the user. All 38 computers were connected to a network switch configured to send all the packets to the network communication port of the system. A special program developed to emulate users access to the Web was installed on each of the stations. We did not experiment with real users since their traffic would be too slow to check the limits of OHT.

The simulations consisted of 13 iterations of access to a given list of 100 URLs including textual HTML files. All iterations were performed by all 38 stations in the lab. Thus, if ideally performed, every iteration would result in 3,800 reconstructed HTML pages. The emulators were started simultaneously from all stations. We controlled the time between succeeding accesses to the Web in each iteration to test the frequency of accesses the system was able to handle. A maximal time gap for the iterations was set to a fixed value, while the actual time gap between accesses within iterations varied randomly in a range between zero and the maximal time gap. The first iteration was set to a maximal time gap of 60 seconds and the time gaps decreased in steps of five seconds on the following iterations. Results are shown in Table 1, presenting the number of HTML that OHT captured for each maximal time gap, and the percent of the captured pages referred to as "Success Rate".

Table 1. Captuerd HTML success rate

	Captured HTML pages	Success Rate
60	3800	100%
55	3800	100%
50	3800	100%
45	3800	100%
40	3800	100%
35	3800	100%
30	3800	100%
25	3800	100%
20	3800	100%
15	3800	100%
10	3798	99.9%
5	3800	100%
0	3796	99.8%

OHT managed to capture almost all HTML pages accessed from the 38 stations with a very high success rate (100% for most time gaps). We were able to conclude that OHT, the most critical real-time component of the system, was feasible.

Simulation No. 2. In order to examine the feasibility of the suggested methodology (not only OHT) and the effect of the number of clusters on TP and FP rates, we simulated the learning phase by applying k-mean clustering on the vectors representing the normal user's behavior. In this simulation we used the basic TDS detection algorithm. We used its results to calibrate the number of clusters for ATDS. Also, this simulation became the baseline for comparing the performance of ATDS (operated in simulation no. 3) to the performance of the basic TDS. We then simulated "normal" and "abnormal" users and applied detection in order to detect the "abnormal" users. We simulated normal users accessing only "normal" web pages and abnormal users accessing terror related sites (abnormal content).We applied the simple TDS detection algorithm to detect the abnormal (potential terrorist) users. In this simulation we wanted to test the effect of number of clusters generated in the learning phase on the detection performance measured by TP and FP presented as ROC graphs. We ran the simulation three times with different number of clusters: 50, 100, and 200.

Upon receiving a vector of an accessed HTML page from the simulation of the normal or abnormal users, the system computes the similarity between the vector and the group profiles to locate the closest cluster. If the similarity is bellow the defined similarity threshold, the systems issues an alarm.

Figure 3 describes some results of the simulation showing the TP and FP rates for an abnormal user accessing abnormal pages and a normal user accessing only normal pages. Each curve presents a different number of clusters generated during the learning phase. As can be seen from the graph 50 clusters and 200 clusters resulted in very similar performance while the 100 clusters classifier obtained better performance for most ranges of TP and FP, specifically for $TP < 0.84$ and $FP < 0.4$ we will prefer the 100 clusters classifier. We believe that it is impossible to generalize a conclusion about an ideal number of clusters. It depends on the training data and the number of vectors included in the process. However, this simulation showed that the number of clusters might affect the system's performance; therefore a sensitivity analysis (such as this simulation) is required to find the optimal number of clusters for a given system. In addition, these simulations confirmed the system's feasibility, since even when a simple detection algorithm is applied that is not calibrated to optimize performance; we were able to obtain reasonable results (FP and TP).

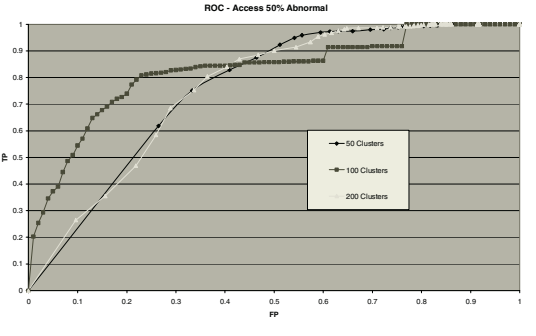


Fig. 3. TP and FP as a function of the number of clusters

Simulation No. 3. Our enhanced detection algorithm aims at applying continuous detection and issue an alarm based on several accesses to abnormal pages. Also, an alarm is issued only if the rate of suspected vectors in a user's queue exceeds a threshold (the alarm threshold). Simulation no. 3 examined the effect of the queue size and the alarm threshold on the detection quality measured by TP and FP; we therefore, ran several simulations applying the detection algorithm with different values to these parameters. We examined the queue size with values: 2, 8, 16 and 32, and the alarm thresholds with values of 50% and 100%. Since, these simulations might be affected by the order of incoming vectors that might alter the number of suspected pages in a user's access queue and the rate of the abnormal pages in the queue (alarm threshold) we repeated the detection ten times to cancel the effect. We present the results averaged for the ten repeated simulations. As in simulation no. 2, an abnormal user was simulated by accesses to abnormal pages. The graphs on Figures 4 and 5 show the effect of the alarm threshold, and those on Figure 6 and 7 show the effect of the queue size.

It can be seen from figures 4-7 that the detection performance increases with the increase of the queue size for both values of the alarm threshold. Also, the graphs show that for this data 100% alarm threshold is better than 50%. Actually the system reached an almost ideal detection for queue size =32 and alarm threshold 100% (see Figure 7). This result of superiority of the 100% alarm threshold over 50% cannot be generalized since it might depend on the data. However, the results suggest that with sensitivity tuning of the advanced detection algorithm's parameters it is possible to optimize system's performance.

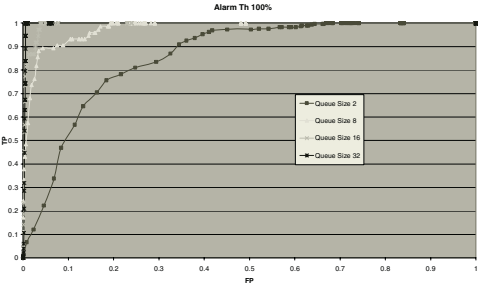


Fig. 4. TP and FP for 100% alarm threshold

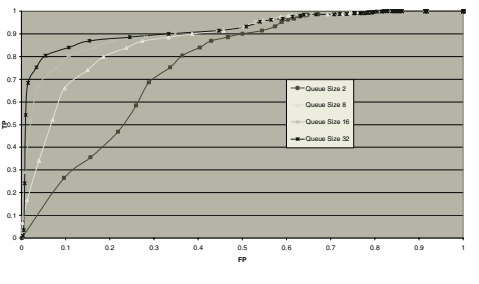


Fig. 5. TP and FP for 50% alarm threshold.

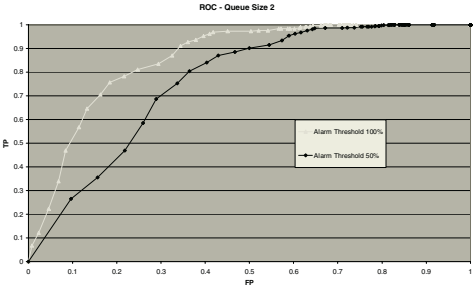


Fig. 6. TP and FP positive for queue size 2

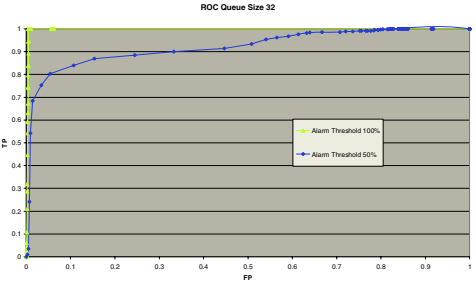


Fig. 7. TP and FP as for queue size 32

5 Summary and Future Research Issues

ATDS is aimed at tracking down suspected terrorists by analyzing the content of information that they access. ATDS is based on the Content-Based Methodology for Anomaly Detection on the Web suggested by [2,5]. In this paper we presented ATDS evaluation that confirmed its feasibility and the contribution of the advanced detection algorithm to the detection performance.

An important contribution of ATDS lies in the unique environment of its application. The detection component is planned to run in a real-time wide-area network environment and should be capable of online monitoring of many users. Therefore, a crucial design requirement was high-performance which called for enhancement of the algorithms involved especially the mining algorithm to high-performance and scalability. ATDS is an example of a successful application of data mining and machine learning techniques to the international cyber-war effort against terror.

As for future research issues, we plan to evaluate the systems simulating a more realistic model of the terrorist user, i.e., mixing accesses to normal and abnormal pages. In addition, we are developing a cross-lingual version of the system as most terror-related sites use languages other than English (e.g., Arabic). We are also planning to analyze non-textual content on pages, such as logos, pictures, colors and any other non-textual features that may identify normal or abnormal content. We also developed (but not yet implemented) an opposite (profile-based) model that collects and

learns the typical content of terror-related web pages and detects users viewing similar content rather than detecting users by content, which is dissimilar to the normal content. Integration of the behavior-based and the profile-based models is expected to reduce the false positive rate of the detection system.

References

1. Birnhack M., D. and Elkin-Koren, N. Fighting Terror On-Line: The Legal Ramifications of September 11, Internal Report, The Law and Technology Center, Haifa University. (http://law.haifa.ac.il/faculty/lec_papers/terror_info.pdf) (2003)
2. Elovici,Y., Shapira, B., Last, M., Kandell, A., and Zaafrany, O: Using Data Mining Techniques for Detecting Terror-Related Activities on the Web, J of Information Warfare, 3 (1) (2004) 17-28.
3. Extractor DBI Technologies (2003) http://www.dbi-tech.com
4. Fielding, R. Gettys, J. and Mogul, J. (1999) "RFC2616: Hypertext Transfer Protocol – HTTP/1.1" Network Working Group.
5. Last, M. Elovici, Y. Shapira, B. Zaafrany, O, and Kandel, A.: Using Data Mining for Detecting Terror-Related Activities on the Web, ECIW Proceedings (2003) 271-280.
6. Last, M. Elovici, Y. Shapira, B. Zaafrany, O., and Kandel, A.: Content-Based Methodology for Anomaly Detection on the Web, Advances in Web Intelligence, E. Menasalvas et al. (Eds), Springer-Verlag, Lecture Notes in Artificial Intelligence, Vol. 2663, 113–123 (2003).
7. Winpcap version 3.0 (2004) http://winpcap.polito.it/
8. Wooster, R., Williams, S. and Brooks, P.: HTTPDUMP: a network HTTP packet snooper. Working paper available at http://cs.vt.edu/~chitra/work.html (1996)
9. Kelley, J.: Terror Groups behind Web encryption", USA Today, http://www.apfn.org/apfn/WTC_why.htm (2002)
10. Lemos, R.: What are the real risks of cyberterrorism?, ZDNet, http://zdnet.com.com/2100-1105-955293.html (2002)
11. Ingram, M.: Internet privacy threatened following terrorist attacks on US, http://www.wsws.org/articles/2001/sep2001/isps24.shtml (2001)
12. Sequeira, K. and Zaki, M.: ADMIT: Anomaly-based Data Mining for Intrusions, Proceedings of SIGKDD 02, (2002) 386-395
13. Karypis, G.: CLUTO - A Clustering Toolkit, Release 2.0, University of Minnesota, http://www.users.cs.umn.edu/~karypis/cluto/download.html (2002)
14. Salton, G., and Buckley, C.: Term-Weighting Approaches in Automatic Text Retrieval, Information Processing and Management, 24(5), (1988) 513-523
15. Mobasher, M., Cooley, R., and Srivastava, J. :Automatic personalization based on Web usage mining *Communications of the ACM,* 43 (8) (2002) 142-151
16. Ghosh, A.K., Wanken, J., and Charron, F.: Detecting Anomalous and Unknown Intrusions Against Programs. In Proceedings of ACSAC'98, December (1998)
17. Tan, K., and Maxion, R.: "Why 6?" Defning the Operational Limits of Stide, an Anomaly-Based Intrusion Detector. Proceedings of the IEEE Symposium on Security and Privacy (2002) 188 -202
18. Lane,V, and Brodley, C.E.: Temporal sequence learning and data reduction for anomaly detection. In Proceedings of the 5th ACM conference on Computer and Communications Security, (1998) 150-158.

Modeling and Multiway Analysis of Chatroom Tensors

Evrim Acar, Seyit A. Çamtepe, Mukkai S. Krishnamoorthy, and Bülent Yener*

Department of Computer Science, Rensselaer Polytechnic Institute,
110 8^{th} Street, Troy, NY 12180
{acare, camtes, moorthy, yener}@cs.rpi.edu

Abstract. This work identifies the limitations of n-way data analysis techniques in multidimensional stream data, such as Internet chatroom communications data, and establishes a link between data collection and performance of these techniques. Its contributions are twofold. First, it extends data analysis to multiple dimensions by constructing n-way data arrays known as *high order tensors*. Chatroom tensors are generated by a simulator which collects and models actual communication data. The accuracy of the model is determined by the Kolmogorov-Smirnov goodness-of-fit test which compares the simulation data with the observed (real) data. Second, a detailed computational comparison is performed to test several data analysis techniques including SVD [1], and multiway techniques including TUCKER1, TUCKER3 [2], and PARAFAC [3].

1 Introduction and Background

Internet Relay Chat (IRC) is a multi-user, multi-channel and multi-server communication medium that provides text-based, real-time conversation capability [4]. Chatroom communication data offer valuable information for understanding how social groups are established and evolve in cyberspace. Recently, there has been intense research focus on discovering hidden groups and communication patterns in social networks (see [5, 6, 7, 8, 9] and references therein).

Chatrooms are attractive sources of information for studying social networks for several reasons. First, chatroom data are public, and anyone can join into any chatroom to collect chat messages. Second, real identities of *chatters* are decoupled from the *virtual identities* (i.e., nick names) that they use in a chatroom. For example, a 50-year old male chatter can participate in a teenager chatroom with multiple virtual identities one of which could be associated with a female persona. Thus, there is no privacy in chatroom communications. Indeed, it is partially this total lack of privacy that makes chatrooms vulnerable to malicious intent and abuse, including terrorist activities. Third, chatroom data are obtained from streaming real-time communications, and contain multidimensional and noisy information. Extracting structure information without understanding

* This research is supported by NSF ACT Award # 0442154.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 256–268, 2005.

the contents of chat messages (i.e., without semantic information) to determine how many topics are discussed, or which chatters belong to the same conversation topics, is quite challenging.

There are several efforts to extract information from chatroom communications [10, 11, 12, 8, 9]. However, current techniques have limited success since chatroom data may have high noise and multidimensionality. Thus, data analysis techniques, such as Singular Value Decomposition (SVD) [1] that rely on linear relationships in two-dimensional representation of data, may fail to capture the structure. In this work, we extend chatroom data analysis to multiple dimensions by constructing multiway data arrays known as *high order tensors*. In particular, we consider three-way arrays (i.e., cubes) with dimensions: (1) users (who), (2) keywords (what), (3) time (when). Accordingly, we consider generalization of SVD to higher dimensions to capture multiple facets of chatroom communications. Multidimensional SVD has been a focus of intensive research [13, 14, 15, 16, 17, 18, 19]. It is well understood that there is no "best" way to generalize SVD to higher dimensions. Computational methods favor greedy algorithms based on iterative computations such as *alternating least squares* (ALS) [16, 15, 13]. For example, most popular multiway data analysis techniques TUCKER3 [2] and PARAFAC [3] use ALS. While special cases (such as tensors with orthogonal decompositions [16]) are possible, in general enforcement of constraints in ALS remains as a challenge.

1.1 Our Contributions

The main goal of this work is to identify the limitations of n-way data analysis techniques, and establish a link between data collection (i.e., tensor construction) and performance of these techniques. More precisely, this paper has several contributions:

i. We present a model and its statistical verification using actual chatroom communications data. The model is used to implement a simulator for generating three dimensional **chatroom tensors** with *user* × *keyword* × *time*.
ii. We examine how two-way data analysis techniques such as SVD would perform on chatroom tensors to extract the structure information. We show that SVD may fail on chatroom tensors even with quite simple structure while three-way data analysis techniques such as TUCKER1 and TUCKER3 are successful.
iii. We investigate how the construction of chatroom tensors would impact the performance of both SVD and three-way data analysis techniques. In particular, we investigate the importance of noise filtering and dimensions of chatroom tensors, and show how sensitive the analysis techniques are.
iv. Finally we compare three-way analysis techniques with each other as a function of several metrics, such as number of components, explained variation, number of parameters and interpretability of the models. We show that high model complexity (w.r.t. the parameters), which is an indication of more modeling power and more explained variation, does not necessarily capture the right structure when data are noisy.

Organization of the paper: This paper is organized as follows. Section 2 describes the data collection procedure from chatroom channels and verifies the simulation model. In Section 3, we discuss the impact of data construction on two-way and multiway analysis techniques using both special cases and simulation data. This section also compares the performance of different data analysis techniques in extracting the internal structure of data.

2 Modeling and Simulating Chatroom Data

Data Collection: We have implemented an IRC bot similar to one used in [8]. The bot connects to an IRC server, and joins to given channel. It logs public messages and control sequences (nick, quit, kick, leave, etc.) flowing in the channel. We have collected 24 hours, 20 days (November 2004) of logs from *philosophy* channel in *dallas.tx.us.undernet.org undernet* server [1]. The log file is 25 MB in size, and includes 129,579 messages.

The logs are processed for 4-hour period between 16:00-20:00 for 20 days. There are average of 10 to 20 active users during this period. We keep track of *join* and *change nick* messages to determine a chain of nicks (e.g., change nick from A to B, B to C, etc.), and associate the chain with a single *user*.

Table 1 shows interarrival statistics obtained over 20-days of data. Rows of the table show the statistics when the interarrival time is bounded by the given value. Note that, 99.82% of the interarrival times are less then 300 seconds. Thus, we assume that no conversation could survive 300 seconds of silence.

Table 1. Interarrival time statistics over 20-day of log

Interarrival Time	Mean	Median	STD	Skewness	Kurtosis	Number of Messages	% of Messages
≤ 60	11.97	8	11.6	1.57	2.36	103,997	97.36
≤ 180	13.73	5	16.75	3.37	17.6	106,449	99.66
≤ 300	14.08	9	18.9	4.86	40.65	106,624	99.82

Table 2 presents the statistics for *message interarrival time*, *message size* in terms of word counts and *number of messages per user*. Message size and interarrival time fit to exponential distribution with parameters ($\mu = 0.0801$) and ($\lambda = 0.0677$), respectively. Number of messages per user obeys to a power law distribution with exponent ($\alpha = -1.0528$).

Identifying Keywords: Users in a chatroom talk about several topics which may overlap in time domain. We define a *conversation* as a sequence of posts made by at least two topic members.

[1] There was no specific reason for choosing the philosophy channel. It is one of the many channels with less junk information.

Table 2. Results of analysis on 4-hour x 20-day of log for message size, interarrival time and number of messages per user

	Mean	Std	Skewness	Kurtosis	# Samples	Distribution
Message Size	12.47	11.17	1.86	4.83	18,483	$f(x) = 0.0801\ e^{-0.0801x}$
Interarrival Time	14.76	19.29	4.68	37.58	18,430	$f(x) = 0.0677e^{-0.0677x}$
# Mess. per User	21.24	33.78	2.88	10.00	870	$f(x) = 0.2032x^{-1.0528}$

It is possible to find a set of specific keywords for each topic which are frequently used by the topic members. However, care must be taken to handle irregular verbs or verbs with $-ed$, $-ing$, $-s$ to treat them as the same word. We consult to the online *webster* (www.webster.com) dictionary to find the simple forms of these words. We consider common words among several topics as *noise* if they are not *specific keywords* of any topic. Noise also includes typos or other unresolved words by *webster*.

Model: We developed a model for chatroom communications based on the statistical observations on the real data. Model accepts five parameters: (i) distribution for interarrival time, (ii) distribution for message size in terms of word count, (iii) distribution for number of messages per user, (iv) noise ratio (NR), and (v) time period.

Given a topic-tuple $T_1, \cdots, T_n$ of n topics, the model computes the number of messages $m_{j,k}$ posted by user j on topic k. This number is assigned according to a power law distribution which is obtained from the statistics collected over real data. Once $m_{j,k}$ is determined, *message posting probability* for a user h is calculated as $m_{h,k} / \sum_{\forall j} m_{j,k}$. For the philosophy channel, interarrival time obeys exponential distribution which is generated by a Poisson arrival process with arrival rate of λ. Thus, conversation duration for a topic-tuple becomes: $\sum_{\forall j} \sum_{\forall k} m_{j,k} * 1/\lambda$. We model a chatroom log as a queue with multiple Poisson arrival processes. Suppose there are $T_1, ..., T_n$ of n topics each with $M_1, ..., M_n$ messages respectively. Then, the arrival rate for each topic will be $\lambda_1, ..., \lambda_n$ respectively where:

$$\lambda_i = \frac{M_i * \lambda}{\sum_{\forall j} M_j},\ 1 \leq i \leq n$$

Noise Modeling: In this work we use Gaussian noise to introduce a model parameter *noise ratio* (NR) as: *NR = (Topic Specific Words + Noise Words) / Noise Words*. Once message size is decided for a user, number of specific topic words and number of noise words are decided based on this ratio. Specific words are selected uniformly at random from the keyword set of the topic. Noise words are randomly selected according to Gaussian distribution. Gaussian distribution selects some of the words very frequently and some others very rarely. Frequently selected words represent the type of noise words which are used frequently by everybody in the chatroom. Rarely selected words represent typo like noise words which are used rarely in the chatroom. When all selected users in a topic post,

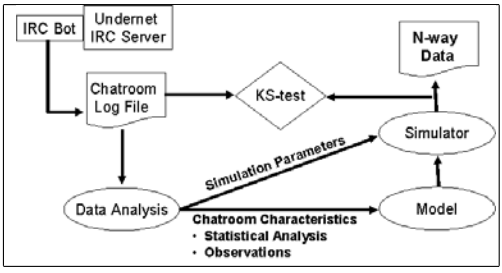


Fig. 1. System architecture for data collection, modeling and simulator

Table 3. Kolmogorov-Smirnov goodness-of-fit test (KS-test) results. *Interarrival time*, *message size* and *number of messages per user* data from model is compared to real chatroom data for the listed significance levels. KS-test does not reject null hypothesis which states that both *synthetic* and *chatroom data* come from the same distribution

	Synthetic Data # Samples	Chatroom Data # Samples	Asymptotic P-value	Significance Level
Interarrival Time	962	18,430	0.1800	10%
Message Size	1,034	18,483	0.2022	10%
# Mess. per User	57	870	0.0521	5%

number of posts for these selected users is decremented, and posting probabilities are recalculated.

Verification: Based on the model, we implement a simulator in Perl. Simulator receives its parameters from a configuration file, and generates chatroom-like communication logs according to the model to be used for verification. Figure 1 presents overall data flow. We perform goodness-of-fit test over synthetic and real data. Table 3 represents Kolmogorov-Smirnov goodness-of-fit test (KS-test) results for the listed significance levels. For all the cases, KS-test does not reject null hypothesis which states that both synthetic and chatroom data come from the same distribution.

3 Computational Comparison of 2-way and 3-way Data Analysis Techniques

We use specific datasets and simulation data to assess two-mode and three-mode analysis techniques. We demonstrate that two-way methods are not as powerful as three-way techniques in capturing the structure of data broken into a number of user groups. We define "user group" as the set of users who share a maximal keyword set in a given time period. Our analyses are conducted in Matlab using Tensor Class [20] for tensor operations and N-way ToolBox[21] for implementations of Tucker and PARAFAC models.

3.1 Impact of Tensor Data Construction

There are two types of data used in this section: (i) manually created data, and (ii) simulation data. For three-mode analysis, we rearrange the data into a tensor, $T \in R^{u \times k \times t}$, defined by *user x keyword x time* modes, where T_{ijk} shows the number of keyword j sent by user i during time slot k. For two-mode analysis based on SVD, we prepare two matrices: $UK \in R^{u \times k}$, where u and k are the number of users and keywords, respectively. Each entry UK_{ij} shows the number of keyword j sent by user i. Second matrix is matrix $UT \in R^{u \times t}$, where t is the number of time slots and UT_{ij} indicates the number of total keywords sent by user i in time slot j. Our objectives are twofold (i) to construct examples where SVD fails to discover the structure while three-way methods Tucker1 and Tucker3 succeed, and (ii) to generate tensors with the same properties as the actual chatroom-like communication data using the simulator and examine the impact of noise and time window size on the performance of analysis techniques.

Noise-Free Tensors with Disjoint Groups and Keywords: First dataset has a structure as shown in Table 4(a). Group 1 and 2 talk about the same topic using a common keyword set while Group 3 and 4 make use of a completely different keyword set. Group 1 and 3 always speak at odd time slots whereas Group 2 and 4 occupy even time slots. We note that data are *noise-free* thus there are no words that are not keywords and there are no users that do not belong to a group.

In such a setting, SVD on matrix UK tends to cluster the users using the same keywords. Similarly, SVD of matrix UT forms clusters containing the users that speak during the same time slots. Therefore, both methods fail to discover the internal structure of data, which actually contains 4 separate groups.

There are, in total, 10 users and 2 keyword sets each containing 2 keywords. Simulation time is 42 time slots. We can represent the sample data as a tensor A of size 10 x 4 x 42 or an unfolded matrix M with dimensions 10 x 168. Best ranks of matrices, UT, UK and M as well as best rank of each mode of tensor A are determined for rank reduction. The users are mapped on the spaces spanned by singular vectors chosen via rank reduction to identify the

Table 4. (a) First specific dataset where group membership and keywords are disjoint, (b) Second specific dataset where groups have common members but keywords are disjoint, and (c) Simulation dataset where groups have common members but keywords are disjoint

Groups	Members		
	(a)	(b)	(c)
1	User 1,2	User 1,2,3	User 1,2,3,4
2	User 3,4,5	User 2,3,4	User 3,4,5,6
3	User 6,7	User 5,6	User 7,8
4	User 8,9,10	User 7,8	User 9,10

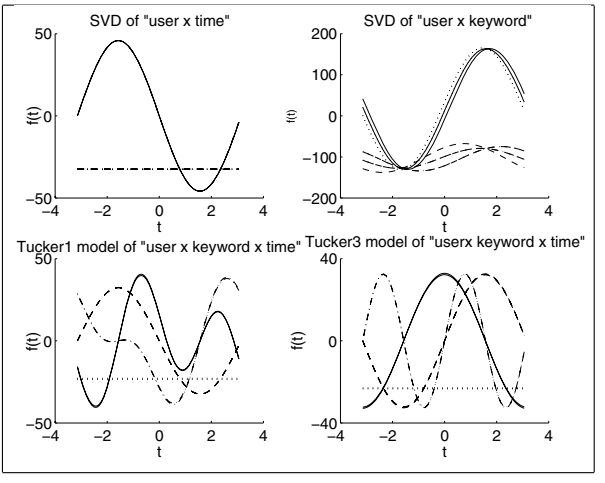


Fig. 2. SVD on *user x time*, UT and *user x keyword*, UK matrices are not powerful enough to find all four user groups while Tucker1 and Tucker3 are capable of extracting all groups from the data

clusters and the structure in the data. It is possible to have only two or three significant singular vectors but higher than three depending on the structure of data should also be anticipated. Let k be the number of most significant singular values identified by rank reduction. We multiply k significant singular values with their corresponding left singular vectors. If $U \in R^{nxm}$ and $S \in R^{mxm}$ represent left singular vectors and singular values of the original matrix, respectively, we compute matrix $F = U(:, 1:k) * S(1:k, 1:k)$ and regard $F \in R^{nxk}$ as a multidimensional dataset where each row represents a user and each column is one of the k properties of a user.

We represent the results of two-mode and three-mode methods using Andrew's curves [22], which transform multidimensional data into a curve, enable us to visualize graphically the structure of data stored in matrix F (i.e., how users are spread on the space spanned by more than 2 or 3 components) [1].

Noise Free Tensors with Overlapping Groups and Disjoint Keywords: The second dataset shown in Table 4(b) is similar to the first one except that overlapping user groups are allowed in order to inquire the performance of analysis methods in the presence of common users.

[1] To visualize the behavior of user i, i^{th} row of matrix F, F_i, is converted into a curve represented by the following function:

$$f_i(t) = \begin{cases} \frac{X_{i,1}}{\sqrt{2}} + X_{i,2}\sin(t) + X_{i,3}\cos(t) + ... + X_{i,p}\cos(\frac{p-1}{2}t) & \text{for } p \text{ odd} \\ \frac{X_{i,1}}{\sqrt{2}} + X_{i,2}\sin(t) + X_{i,3}\cos(t) + ... + X_{i,p}\sin(\frac{p}{2}t) & \text{for } p \text{ even} \end{cases}$$

where $t \in [-\pi, \pi]$

SVD of matrices UK and UT both perform poorly for this case: it can distinguish common users; but overlapping users are treated as a separate cluster. We consider this result as a failure because such analysis is capable of finding subsets of groups while missing the whole group structure. We note that in this case, Andrew's curves are not sufficient to differentiate common users. Therefore, we make use of an unsupervised clustering algorithm called fuzzy c-means [23], which returns membership degrees of each user for each group. Among many clustering algorithms, fuzzy c-means is an appropriate choice for chatroom data because it allows a data point to be in more than one cluster. C-means algorithm returns different membership values for each run since it is a nondeterministic algorithm. The results presented in Table 5, are the cases that represent the majority of 100 runs. SVD of UT gives the result in the table in 70% of the runs and SVD of UK returns the recorded result in 85% of the runs. The results for Tucker1 and Tucker3 are explained in detail below.

Rows named as "Groups", show which group each user is assigned to according to the results of membership values. SVD on UT, groups Users 1, 5 and 6

Table 5. Membership values for users given by fuzzy c-means clustering algorithm. Each user belongs to a group with certain probability represented by membership values. The highest probability determines the group each user belongs to

	User1	User2	User3	User4	User5	User6	User7	User8
SVD of UT								
Pr(Usr ∈ Grp1)	0.5000	0.0000	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000
Pr(Usr ∈ Grp2)	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp3)	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	1.0000
Pr(Usr ∈ Grp4)	0.5000	0.0000	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000
Groups	1 or 4	2	2	3	1 or 4	1 or 4	3	3
SVD of UK								
Pr(Usr ∈ Grp1)	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp2)	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp3)	0.0000	0.0000	0.0000	0.0000	0.0016	0.0124	0.9984	0.9876
Pr(Usr ∈ Grp4)	0.0000	0.0000	0.0000	0.0000	0.9984	0.9876	0.0016	0.0124
Groups	1	2	2	1	4	4	3	3
Tucker1								
Pr(Usr ∈ Grp1)	1.0000	0.0042	0.0055	0.0530	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp2)	0.0000	0.0005	0.0007	0.0499	1.0000	1.0000	0.0000	0.0000
Pr(Usr ∈ Grp3)	0.0000	0.9948	0.9931	0.8429	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp4)	0.0000	0.0005	0.0007	0.0541	0.0000	0.0000	1.0000	1.0000
Groups	1	3	3	3	2	2	4	4
Tucker3								
Pr(Usr ∈ Grp1)	0.0596	0.0005	0.0007	0.0000	1.0000	1.0000	0.0000	0.0000
Pr(Usr ∈ Grp2)	0.8105	0.9927	0.9905	0.0000	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp3)	0.0653	0.0062	0.0080	1.0000	0.0000	0.0000	0.0000	0.0000
Pr(Usr ∈ Grp4)	0.0646	0.0006	0.0008	0.0000	0.0000	0.0000	1.0000	1.0000
Groups	2	2	2	3	1	1	4	4

together although they just share the time space and do not form a group. We observe the same behavior for Users 4, 7 and 8. This is an expected outcome because SVD on matrix UT maps the users with similar chatting pattern in time closer to each other in the space spanned by left singular vectors. Another important point is that common users are clustered as a completely separate group from other users whom they are talking to. SVD on matrix UK, also shows the same behavior. For overlapping groups model, SVD of matrix UK performs poorly compared to SVD of matrix UT because it cannot capture that Users 1 and 4 are in different groups. However, it outperforms the results of SVD on time mode by correctly extracting Groups 3 and 4 from data. All in all, we observe that two -way analysis results do not reflect the real structure of data. Similarly, at first sight, neither Tucker1 nor Tucker3 seems to capture the exact structure but the cases shown for Tucker1 and Tucker3 occur approximately 50% of the time. If User2 and User3 are clustered with User4 in half of the runs, then they are clustered with User1 in the other half. Therefore, probability of User2 and User3's being in the same cluster with User1 and User4 are both close to 0.5. Thus, we can make a hypothesis for group membership based on these probabilities.

Noisy Tensors and Impact of Noise Ratio: Using the simulator, we implement a model where we demonstrate the effect of noise in extracting the structure of data for two-way and three-way methods. Noise is introduced in keywords mode by the use of a number of words shared by all user groups.

Experimental model consists of 4 groups of users as shown in Table 4(c). As in the scenario of special cases, Group 1 and 2, and similarly Group 3 and 4 use the same keyword sets. Keyword sets are distinct and contain 10 keywords each. Groups making use of distinct keyword sets can talk during the same time period with equal probability.

We create separate chat logs for a simulation time of 4 hours for different noise levels indicated by NR. We set the minimum and maximum number of messages that can be sent by each user as 30 and 70, respectively. We assess the relative performance of SVD, Tucker1 and Tucker3 models on different noise levels and demonstrate our results in Table 6. After the selection of significant components for user mode, we run fuzzy c-means clustering algorithm 100 times to see how often the pattern discovered by clustering method coincides with the internal structure of data. SVD on matrix UK fails to find the right data pattern because it tends to cluster users talking about the same topic regardless of their conversation slots. SVD on matrix UT can capture the structure with moderate success ratios while Tucker1 and Tucker3 have the best results. As noise level increases, we observe that all algorithms start to suffer at some threshold values. Analysis results suggest that Tucker3 performs better than Tucker1 on noisy tensor data.

There are different concepts of noise but our approach in introducing noise tends to form a single keyword cluster. When we observe the behavior of singular values under the effect of noise, we clearly see that noisy data approach to rank-1 in keyword mode while becoming full rank in user mode.

Table 6. Impact of NR (noise ratio) on success ratios of two-way and three-way analysis methods when time window is 240 seconds and tensor is of form 10x100x60

NR	SVD userXkeyword	SVD userXtime	TUCKER1	TUCKER3
No Noise	0%	56%	100%	100%
5	0%	55%	100%	100%
3	0%	47%	73%	100%
2	0%	0%	3%	16%

Impact of Sampling Time Window: On the same experimental set up, we show how different time window sizes affect the performance of analysis techniques. We work on a noise-free dataset to be able to observe solely the effect of window size on success ratios. In this setting, minimum and maximum number of messages sent by a user are 1 and 200, respectively. We create a single chat log of 4 hours and generate data for different time window sizes. Analysis results in Table 7 display that there exists a threshold window size below which none of the analysis methods can discover the structure of data. When time window is roughly 200-300 seconds, three way models have the best performance while SVD on UT can capture the structure only to some extent. Under 180 seconds, none of the methods succeeds. We also observe a performance degradation in all methods as time window size increases considerably. This is an indication of an upper bound on time window size over which users talking at different time slots are considered in the same time period. This hides the communication pattern completely. It is important to observe the relative performance of algorithms in Table 7 rather than exact success ratios or exact time window threshold values.

Table 7. Impact of sampling time window on success ratios of two-way and three-way analysis methods for noise-free data. Total simulation time is 14400 seconds. Tensor is constructed as*user x keyword x time*. As sampling time window changes, dimension of the tensor in time mode is adjusted accordingly

Tensor	Time Window (seconds)	SVD userXkeyword	SVD userXtime	TUCKER1	TUCKER3
10x20x1	14400	0%	0%	0%	0%
10x20x2	7200	0%	0%	13%	17%
10x20x8	1800	0%	0%	17%	22%
10x20x12	1200	0%	25%	24%	27%
10x20x24	600	0%	28%	26%	26%
10x20x48	300	0%	20%	100%	100%
10x20x72	200	0%	14%	100%	100%
10x20x80	180	0%	0%	0%	0%
10x20x160	90	0%	0%	0%	0%

Similar to the effect of noise case, when we inspect the behavior of singular values in user mode, we observe that as time window size gets smaller, we observe a structure close to full rank.

3.2 Performance Comparison of Multiway Techniques

We assess the performance of Tucker1, Tucker3 and PARAFAC with respect to several metrics such as number of components, explained variation, number of parameters and interpretability of models. Performance analysis results suggest that Tucker3 model provides the best interpretation. In case of noise-free data, there is no difference in using Tucker1 or Tucker3 decomposition in terms of data interpretation. However, for Tucker1, number of parameters, which is the total number of entries in matrices/ core tensors produced in tensor decomposition, is much larger than the number of parameters in Tucker3 and more parameters introduce complication in data interpretation. PARAFAC is not appropriate for modeling our data because of its strict modeling approach. It does not allow extraction of different number of components in different modes. Besides, while Tucker3 model enables us to decompose a tensor into orthogonal component matrices X, Y, and Z and estimate orthonormal bases, in PARAFAC, we can only do that if tensor is diagonalizable. In Table 8 we present the results of

Table 8. Performance comparison of N-way analysis techniques for time window 240 seconds and tensor 10x100x60. Tucker1, Tucker3 and PARAFAC are compared based on explained variation, number of parameters used in each model and success ratio of capturing the structure. Comparison of the models is presented for two different noise levels, NR=0 and NR=3

	NR	Number of Components	Explained Variation	Number of Parameters	Structure
Tucker1	0	5	84.6493	30050	100%
Tucker3	0	5 5 5	76.9814	975	100%
Tucker3	0	5 2 5	76.5944	600	100%
Parafac	0	5	48.753	850	0%
Tucker1	0	8	95.1406	48080	75%
Tucker3	0	8 8 8	84.8294	1872	100%
Tucker3	0	8 2 8	83.7563	888	100%
Parafac	0	8	49.4294	1360	0%
Tucker1	3	5	77.4148	30050	5%
Tucker3	3	5 5 5	62.7061	975	7%
Tucker3	3	5 2 5	62.2053	600	11%
Parafac	3	5	40.7648	850	0%
Tucker1	3	4	69.5659	24040	69%
Tucker3	3	4 4 4	58.3426	744	64%
Tucker3	3	4 2 4	58.2482	512	100%
Parafac	3	4	40.3238	680	0%

performance comparison for multiway techniques. Note that even if we extract the same number of components in each mode, Tucker3 model is more robust.

When data are noisy, we observe performance degradation in terms of interpretability in Tucker1 while Tucker3 can still capture the structure successfully if right number of components is determined for each mode. Table 8 demonstrates the importance of right component numbers in success ratio of data interpretation. Similarly, it gives an example of a case where selection of component numbers just taking into fit of the model into account does not necessarily imply better interpretation of data.

4 Conclusions

In this work we show how to generate three-way chatroom tensors and examine the performance of data analysis algorithms. We show that three-dimensional chatroom tensors contain multilinear structure that cannot be detected by SVD. The performance gap between SVD and multiway analysis techniques Tucker1 and Tucker3 grows as a function of increasing noise in the data. We also show that construction of the chatroom tensor with respect to sampling window size has significant impact on the performance of analysis techniques. We examine the performance of Tucker1, Tucker3 and PARAFAC with respect to several metrics such as number of components, explained variation, number of parameters and interpretability of the models. Our results suggest that there is no difference in using Tucker1 or Tucker3 decomposition if the data are noise-free. In general, Tucker3 model provides the best interpretation and has the advantage of less number of parameters compared to Tucker1. We note that one of the challenges left for further research is to determine the optimal number of components to obtain the most accurate structure information. It is evident from our study that how data are collected and represented have significant impact over discovering the structure hidden in them.

References

1. Golub, G., Loan, C.: Matrix Computations. 3 edn. The Johns Hopkins University Press, Baltimore, MD (1996)
2. Tucker, L.: Some mathematical notes on three mode factor analysis. Psychometrika **31** (1966) 279–311
3. Harshman, R.: Foundations of the parafac procedure: Model and conditions for an explanatory multi-mode factor analysis. UCLA WPP **16** (1970) 1–84
4. Kalt, C.: Internet Relay Chat. RFC 2810, 2811, 2812, 2813 (2000)
5. Krebs, V.: An introduction to social network analysis. http:// www.orgnet.com/ sna.html (accessed February 2004) (2004)
6. Magdon-Ismail, M., Goldberg, M., Siebecker, D., Wallace, W.: Locating hidden groups in communication networks using hidden markov models. In: Intelligence and Security Informatics (ISI'03). (2003)

7. Goldberg, M., Horn, P., Magdon-Ismail, M., Riposo, J., Siebecker, D., Wallace, W., Yener, B.: Statistical modeling of social groups on communication networks. In: First conference of the North American Association for Computational Social and Organizational Science (NAACSOS'03). (2003)
8. Camtepe, S., Krishnamoorthy, M., Yener, B.: A tool for internet chatroom surveillance. In: Intelligence and Security Informatics (ISI'04). (2004)
9. Camtepe, S., Goldberg, M., Magdon-Ismail, M., Krishnamoorty, M.: Detecting conversing groups of chatters: A model, algorithms, and tests. In: IADIS International Conference on Applied Computing. (2005)
10. Mutton, P., Golbeck, J.: Visualization of semantic metadata and ontologies. In: Seventh International Conference on Information Visualization (IV03), IEEE (2003) 300–305
11. Mutton, P.: Piespy social network bot. http://www.jibble.org/piespy (accessed January 2005) (2001)
12. Viegas, F., Donath, J.: Chat circles. In: ACM SIGCHI (1999), ACM (1999) 9–16
13. Kroonenberg, P.: Three-mode Principal Component Analysis: Theory and Applications. DSWO press, Leiden (1983)
14. Leibovici, D., Sabatier, R.: A singular value decomposition of a k-ways array for a principal component analysis of multi-way data, the pta-k. Linear Algebra and its Applications **269** (1998) 307–329
15. Lathauwer, L., Moor, B., Vandewalle, J.: On the best rank-1 and rank-(r1,r2,...,rn) approximation of higher-order tensors. SIAM J. Matrix Analysis and Applications **21** (2000) 1324–1342
16. Zhang, T., Golub, G.: Rank-one approximation to higher order tensors. SIAM J. Matrix Analysis and Applications **23** (2001) 534–550
17. Kolda, T.: Orthogonal tensor decompositions. SIAM J. Matrix Analysis and Applications **23** (2001) 243–255
18. Kofidis, E., Regalia, P.: On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J. Matrix Analysis and Applications **22** (2002) 863–884
19. Kolda, T.: A counter example to the possibility of an extension of the eckart-young low-rank approximation theorem for the orthogonal rank tensor decomposition. SIAM J. Matrix Analysis and Applications **24** (2003) 762–767
20. Kolda, T., Bader, B.: Matlab tensor classes for fast algorithm prototyping. Technical Report SAND2004-5187, Sandia National Laboratories (2004)
21. Andersson, C., Bro, R.: The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems. (2000)
22. Andrews, D.: Plots of high-dimensional data. Biometrics **28** (1972) 125–136
23. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York (1981)

Selective Fusion for Speaker Verification in Surveillance

Yosef A. Solewicz[1,2] and Moshe Koppel[1]

[1] Dept. of Computer Science, Bar-Ilan University, Ramat-Gan, Israel
[2] Division of Identification and Forensic Science, Israel National Police, Jerusalem, Israel
solewicz@013.net.il
koppel@netvision.net.il

Abstract. This paper presents an improved speaker verification technique that is especially appropriate for surveillance scenarios. The main idea is a meta-learning scheme aimed at improving fusion of low- and high-level speech information. While some existing systems fuse several classifier outputs, the proposed method uses a selective fusion scheme that takes into account conveying channel, speaking style and speaker stress as estimated on the test utterance. Moreover, we show that simultaneously employing multi-resolution versions of regular classifiers boosts fusion performance. The proposed selective fusion method aided by multi-resolution classifiers decreases error rate by 30% over ordinary fusion.

1 Introduction

Monitoring conversations in a communication network is a key tool in the counter-terrorism scenario. The goal may be either tracking known terrorists or even tracking suspicious conversations. In its basic form, trained personnel would listen to a sample of tapped conversations, sometimes guided by automatic word spot systems trained to detect suspicious vocabulary. Unfortunately, this solution is not feasible in practice for several reasons. First, the huge amount of simultaneous conversations precludes effective human surveillance. Moreover, criminals are often aware of tapped lines and communicate in codes, preventing the employment of word-spotting systems.

Automatic speaker recognition (ASR) and 'stress' detection systems could offer a practical solution at least towards reduction of the searching space. In theory, automatic systems could constantly sweep the network searching for enrolled criminals and suspiciously 'stressed' conversations. Relevant conversations could then be directed to human listeners. Current technology claims relative success both in the automatic speaker recognition field [1, 2] and stress detection [3]. ASR technology has been assessed by NIST's benchmarks [2] while detection of 'stressed' speech is still controversial, mostly due to the lack of proper databases for evaluations.

In this paper we present an improved speaker recognition system which takes into account the particular advantages and disadvantages of the surveillance scenario. The surveillance scenario is often characterized by long conversations that can be used as training data for learning speaker models. On the other hand, unlike commercial applications in which the user will normally use his personal line number and handset,

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 269 – 279, 2005.

in the surveillance scenario a variety of different lines could be used, thus adding noise to the recognition process. Our method first automatically characterizes a conversation according to varieties of detectable noise, including speaker stress. It then exploits abundantly available training data to learn speaker models for speaker verification, where the particular model used is a function of the type of noise detected.

In particular, we refine recent work in speaker verification that exploits fusion of low and high speech levels classifiers [4]. These classifiers are based on a variety of feature types, including acoustic, phonetic, prosodic and even lexical ones. The method proposed by Campbell et al. [4] uses a linear combination of classifiers, employing a meta-learner to obtain optimal weights for the respective component learners.

In this work, we propose that the constituent learner weights not be assigned uniformly. Rather, the type and degree of distortion found in the speech sample to be classified is taken into account as part of the classification task. We show that by considering pre-defined data attributes, it is possible to fine-tune the fusion method to improve results. Thus, for example, although acoustic features are generally far superior to all other feature types, there are circumstances under which more weight should be given to lexical features. In a previous paper [5], we showed that a similar approach, in which the selective fusion is controlled by means of a decision tree, improves verification in simulated noisy conditions by more than 20%. In this paper, we exploit various types of "noise", including channel characteristics and speakers' emotional and stress patterns detectable in test conversations. Moreover, we show that including multi-resolution representations of some classifiers enhances fusion capabilities in handling noisy patterns, thus increasing accuracy.

This method both provides significantly improved speaker recognition accuracy as well as pinpointing 'stressed' conversations as a by-product. It is thus ideally suited for surveillance.

The organization of this paper is as follows. In section 2, speech production levels involved in the experiments and their implementation are presented. Experimental settings are presented in section 3. Sections 4 and 5 are dedicated to the proposed meta-learning scheme and results are presented in Section 6. In Section 7, multi-resolution classifiers are considered. Finally, conclusions and future research are discussed in Section 8.

2 Speech Levels

Humans can activate different levels of speech perception according to specific circumstances, by having certain processing layers compensate for others affected by noise. Utterance length, background noise, channel, speaker emotional state are some of the parameters that might dictate the form by which one will perform the recognition process. The present experiments seek to mimic this process. For this purpose, four classifiers were implemented targeting different abstract speech levels:

- The *acoustic* level, covered by a standard CEPSTRUM-Gaussian Mixture Model (GMM) classifier. The term "acoustic" refers to the fact that the GMM spans the continuous acoustic space as defined by the CEPSTRUM features.

- The *phonetic* level, covered by a support vector machine (SVM) classifier using a feature set consisting of cluster indices provided by the GMM. We call this a "phonetic" classifier since it's based on counts of discrete acoustic units, namely, the GMM clusters. (To be sure, the term "phonetic" is not strictly appropriate, since we are not representing traditional phones, but rather abstract acoustic units resulting from clustering the CEPSTRUM space.) An alternative method would be to model cluster sequences [10].
- The *prosodic* level, covered by an SVM classifier using a feature set consisting of histogram bins of pitch and energy raw values and corresponding transitional tokens.
- The *idiolectal* level [6], covered by an SVM classifier using as a feature set frequencies of common words.

Generally speaking, the acoustic and phonetic levels are categorized as low-level as opposed to the higher prosodic and dialectal speech layers. Lower communication layers are normally constrained by the speakers' vocal-tract anatomy, while higher levels are more affected by behavioral markers.

Actually, the *acoustic* and *phonetic* levels are also represented in lower resolutions as analyzed in detail later in Section 7. In this case, less prototypical 'sounds' form the acoustic and phonetic space and each resolution is treated as a distinct (more abstract) level classifier. Let us now consider each of these in somewhat more detail.

2.1 GMM Classifier

Our GMM implementation comprises a Universal Background Model (UBM) from which client models are derived through cluster mean adaptation and is very similar to that described in [1]. Only voiced frames are used. This decision was originally taken mainly in order to attain compatibility with the prosodic vectors stream. In this way, the vectors for all classifiers are obtained in parallel over the same time frames. The GMM consists of 512 gaussians, jointly trained for male and female speakers, taken from NIST'03 evaluation and no score normalizations (such as T- or Z-norm) [7] are performed. Note that NIST'03 evaluation consists basically of cellular recordings, which are not ideal for modeling landline recording as in the present experiments. Moreover, unlike related work performed on this database [8], no echo-canceling procedures were adopted in order to pre-clean this database. Although the acoustic classifier represents a relatively poor baseline, it is particularly appropriate in the context of this work, since our objective is precisely to ascertain how non-acoustical sources can be used to compensate for the deficiencies of the GMM-acoustic approach.

2.2 SVM Classifiers

Three separate SVM classifiers, one for each of the feature types – phonetic, prosodic and idiolectal – are implemented using the *SVMlight* package [9]. After some preliminary calibration, Radial Basis Function (RBF) was the chosen kernel for all SVMs with a radius of 10 for the phonetic and prosodic feature sets and a radius of 100 for the idiolectal feature set.

The phone vector is formed by accumulating the occurrences of the closest 5 (out of 512) GMM centroids for all utterance frames. Intuitively, this represents the speaker specific 'sounds set' frequency.

The prosody vector is formed by an agglutination of the following component counts:

- 50 histogram bins of the logarithmic pitch distribution;
- 50 histogram bins of the logarithmic energy distribution;
- 16 bi-grams of pitch-energy positive/negative time differentiates;
- 64 tri-grams of pitch-energy positive/negative time differentiates;

(There are four possible combinations for positive or negative pitch and energy slopes. Therefore, respectively, 4x4 (16) and 4x4x4 (64) possible bi/tri-gram tokens)

The idiolectal vector is formed by the entries of the 500 most frequency words found in the conversation transcripts.

Fusion of the four speech levels presented is implemented through extra linear SVM learners.

3 Experimental Settings

In this work, experiments are performed following the NIST'01 'extended data' evaluation protocol [11], based on the entire SWITCHBOARD-I [12] corpus. Only the 8-conversation training conditions were used. These comprise 272 unique speakers, 3813 target test conversations and 6564 impostor test conversations. Conversation lengths are 2-2.5 minutes. The evaluation protocol dictates a series of model/test matches to be performed. The matches are organized in 6 disjoint splits, including matched and mismatched handset conditions and a small proportion of cross-gender trials. In all experiments, we use splits 1, 2 and 3 for training (fusion parameters or threshold settings for individual classifiers) and the others for testing and therefore speakers used for training do not appear in the test set. Errors are expressed in percentage of misclassified examples (and not in terms of equal error rates).

Besides speech files, automatic or manually generated transcripts are also available. In this work, we use BBN transcripts (available from NIST's site), which possess a word error rate of close to 50% (!) (Note that ordinary automatic transcripts can be easily obtained in surveillance applications.)

4 Data Attributes

A signal quality measure is needed as a means of controlling the fusion parameters, as a function of the degradation found in an utterance. We wish to use measurable attributes of the conversations to estimate the respective levels of three types of noise: communication channel, speaking style and speaker stress. Following is a brief description of the proposed attributes.

4.1 Channel

It is widely known that speaker recognition accuracy normally declines when the speech is conveyed through some communication channel. Real-world channels are band limited in frequency and often add some noise to the signal. Roughly speaking, an additive bias in acoustical features is introduced by different transmission lines. On the other hand, variance bias appears on the features due to additive (background) noise. Thus, means and standards deviations of the 20 filter bank outputs (byproduct of the MEL-CEPSTRUM extraction process) are retained as a representation of long-term channel behavior. In order to compress this information, which is highly redundant, DCT is applied to the means and the transformed first six components are kept. In addition, the mean of the individual filter bank standard deviations is calculated, as an estimation of additive noise level. One additional component will be added to the channel vector: the average GMM likelihood between the utterance frames and the background gaussian distribution (UBM). A low average likelihood will thus indicate an outlier feature distribution (possibly due to an unexpected channel).

4.2 Conversation Style

We approximate the emotional quality of an utterance through its pitch and energy averages and ranges, in addition to the estimated average articulation rate. In order to neutralize gender effects, pitch distributions are normalized per gender, so that male and female sets will posses the same mean pitch. Pitch and energy ranges are empirically measured as the number of histogram bins around the maximum value until a decay of, respectively, 1/4 or 2/3. Range asymmetry around the maximum is also included as the difference between the number of right and left bins. Articulation (or speech) rate is approximated as the average number of inflection points found in the first CEPSTRUM coefficient stream. A large average indicates a high rate of fluctuation of this parameter due to an accelerated speech rate.

4.3 Speaker Stress

Besides conversation style, we consider effects resulting from speaker stress. (Of course, the distinction between conversation style and speaker stress is somewhat arbitrary.) Features intended to identify stressed speech are based on the “Teager Energy Operator” (TEO) [3]. Our stress vector is composed of the mean and standard deviation of the TEO streams across 6 frequency bands.

5 Selective Fusion

In this section, we describe the proposed selective fusion method, which is depicted in Fig. 1.

In the training phase, k-means clustering is used to cluster the conversations according to respective attribute characteristics, namely channel, style and stress. (Note that only attributes and not explicit speaker verification features are employed in the selection phase.) Distinct fusion schemes are then learned for each cluster using linear

support vector machines. In testing mode, each conversation is first assigned to the appropriate cluster according to its attribute profile and then the corresponding learned fusion scheme is applied.

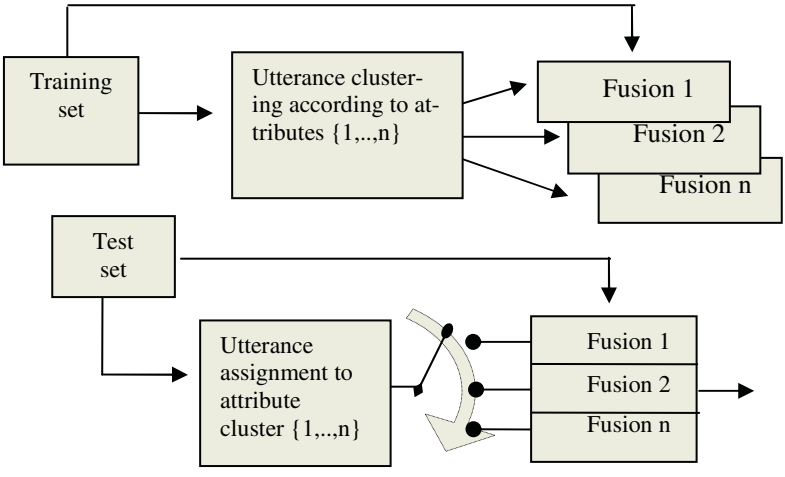


Fig. 1. Selective fusion method

Optimal 'k' (the number of clusters) and attribute vector composition (i.e. which of the attribute classes and components are the most efficient for data characterization) were selected through a nearly greedy search aiming at overall classification error reduction. Initially, full search was performed within the full three attribute classes separately. In a second step, the best candidates of each attribute class were concatenated in a composed vector and a new search was performed in order to determine optimal attribute vector composition. Clustering was performed on the basis of Euclidean distance after the vectors components were normalized to zero mean and unit standard deviation.

6 Results

In this section, we address the four different speech level classifiers, fusion results and the advantage obtained through the proposed meta-learning scheme. Results are analyzed also in light of the subjective ratings established by the SWITCHBOARD transcribers [12].

Individual classifier performance (in terms of percentage of misclassified conversation/speaker pairs, treating false positive and false negatives identically) and their relative weighting in ordinary fusion are depicted in Table 1. Fusion weights were obtained as the output of the fusion SVM, successively setting to 0 the scores corresponding to all individual classifiers except one. Bias terms were removed and the weights were then normalized to unity.

Table 1. Individual classifiers errors and weighting

	Acoust	Phon	Pros	Word
Error (%)	4.7	4.8	10.7	13.9
Weight	0.50	0.18	0.12	0.20

Table 2. summarizes fusion performance for the various attribute classes. Error rates shown are an average of 10 meta-learning runs (k-means clustering is non-deterministic).

Table 2. Fusion results

Attribute class	Error (%)
None (ordinary fusion)	2.84
Channel	2.63
Style	2.70
Stress	2.63
Channel + Style + Stress	2.39

As has been previously established [4], ordinary fusion offers better results than that obtained using any of the constituent speech levels individually. More significantly for our purposes, clustering and then fusing separately for each cluster offers improvement over ordinary fusion regardless of which attribute class is used. Maximal improvement is achieved, though, when all three attribute classes are considered. Let's focus on this case and consider a cluster scheme that proved to be optimal for our purposes. In Table 3 we show a stylized representation (in five quantization levels: {--, -, 0, +, ++}) for the attribute centroids derived from such clustering scheme. (Another successful cluster configuration included the 5th DCT channel component and the 2nd TEO band, instead of speech rate and energy asymmetry.) Recall that the optimal number of clusters ('k' in k-means) and attribute vector composition were found through greedy search.

Table 3. Fusion clusters

	speech rate	pitch range	energy asymm	3rd TEO	4th TEO
Cluster 1	+	+ +	+	+	+
Cluster 2	-	-	0	- -	- -
Cluster 3	+	-	+ +	+ +	+ +
Cluster 4	+	0	- -	+	+ +

Table 4 shows individual classifiers weighting and verification error for each cluster.

Table 4. Fusion weighting and error

	Acoust	Phon	Pros	Word	Error (%)
Cluster 1	0.65	0.11	0.11	0.12	2.80
Cluster 2	0.38	0.28	0.19	0.16	0.87
Cluster 3	0.50	0.20	0.07	0.23	3.07
Cluster 4	0.48	0.24	0.03	0.25	3.53

Let us briefly analyze this fusion scheme configuration. Cluster 2 is the most accurate fusion set. According to subjective rating, it contains conversations with the smallest amount of echo, static and background noise. We have observed that conversations containing echo seem to be around 20% correlated with stress (TEO) values (see Table 3), as in this case. Moreover, absence of echo is associated with a decrease in the acoustic classifier share in the fusion process along with an increase in the phonetic classifier weighting. Possibly, the phonetic classifier, operating in a winner-takes-all fashion is quite sensitive to noise effects, since small perturbations may lead to erroneous 'phone' identification. On the other hand, likelihood values estimated by the acoustic classifier are more smoothly distorted. Actually, we will show in the next section that including acoustic and phonetic classifiers with a smaller number of phonetic units will improve fusion strategies in noisy environments.

In terms of speaking style, we observe a correlation between subjective ratings and attribute values. Cluster 1 is rated as the most natural sounding and bearing high topicality. In fact, the style components (see Table 3) possess high values for this cluster indicating vivid conversations. On the other hand, Cluster 2 is rated as relatively unnatural and bearing low topicality. Indeed, the low-valued style components for this cluster centroid indicate the presence of a formal speaking mode.

Similarly, one can concentrate only on stylistic or stress attributes in order to efficiently detect suspicious conversations in surveillance applications, although more profound analysis of the functions of these attributes remains to be done using stress/deception oriented databases. In fact, an auditory analysis of some stressed labeled utterances revealed that prominent low-stressed conversations (male only) sound extremely bass and as "newscast" style. On the other hand, the impression caused by very high-'stressed' conversations (female only) seems more difficult to typify. It seems that the high-'stress' does not reflect high pitch only. In particular, some pitch normalization should be considered for TEO coefficients in future experiments.

7 Multi-resolution

In this section, we show that simultaneous classifiers covering multi-resolution partitions of the low-level feature space highly boosts fusion accuracy. The motivation behind multi-resolution classification is to make available (a combination of) coarse and refined feature space clusterizations, which can be freely selected according to the nature of incoming test. We expect that noisy data would be more safely classified

within a coarse segmented space, while clean data could explore the sharpness offered by a high-resolution mapping of the space.

For this purpose, we replicate the acoustic and phonetic SVM classifiers in 256, 128, 64 and 32-cluster resolutions, besides the original 512-cluster resolution. A greedy search was performed in order to find optimal ordinary fusion configurations. The following two configurations attained the lowest (**2.12%**) error rate:

- Acoust 512/256/64 + Phone 512/256/128 + Pros + Word
- Acoust 512/256/128/64 + Phone 512/256 + Pros + Word

Further error reduction can be achieved by applying selective instead of ordinary fusion. Optimal error reduction to **1.98%** was obtained for the former configuration. In this case, selective fusion is guided by two distinct attribute settings containing the 2nd and 6th TEO (stress) parameters and optionally one of the following: energy mean value or asymmetry. The following tables describe one of such settings. Table 5 schematically shows the attribute centroids for both clusters and the corresponding error rate for each fusion configuration. The unbalanced error rates are once more explained by the lesser amount of echo effects present in Cluster 1. This phenomenon is confirmed by the ratings assigned to the respective conversations and is in line with the low 'stress' assigned to Cluster 1.

Table 5. Stylized centroids

	Energy asymm	**2nd TEO**	**6th TEO**	**Error (%)**
Cluster 1	**0**	- -	- -	0.73
Cluster 2	**0**	**+**	+	2.72

Table 6 presents the weights assigned to each of the classifiers involved in the fusion process, for each cluster. It can be observed that for Cluster 2, the significance of low-resolution classifiers (Acoust 64 and Phone 128) is especially strong as compared to the corresponding higher-resolution versions. This is a reflection of the higher amount of noise in this cluster, requiring a decrease in feature-space resolution.

Table 6. Weights for individual classifiers

Classifier	**Cluster 1**	**Cluster 2**	**Classifier**	**Cluster 1**	**Cluster 2**
Acoust	0.20	0.64	**Phone 256**	0.10	-0.04
Acoust	0.20	0.31	**Phone 128**	-0.06	-0.16
Acoust 64	-0.02	-0.44	**Pros**	0.20	0.11
Phone 512	0.20	0.41	**Word**	0.18	0.19

8 Conclusions and Future Work

We presented in this paper a meta-learning scheme for fusion of several speech production levels. As opposed to standard classifier fusion, we introduce an utterance

quality measure, which adjusts the fusion scheme according to test signal idiosyncrasies. In addition, we show that multi-resolution low-level classifiers enhance fusion accuracy. Table 7 summarizes error reduction achieved with selective and multi-resolution fusion over the state-of-the-art GMM acoustic classifier and ordinary fusion of classifiers. Almost 60% error reduction could be achieved over the best individual classifier and 30% over ordinary fusion, with little calibration.

Table 7. Summary of fusion results

Classifiers configuration	Error (%)
Acoust (best individual classifier)	4.74
Ordinary fusion on: baseline (Acoust + Phone + Pros + Word)	2.84
Selective fusion on: baseline	2.39
Ordinary fusion on: baseline + Multi-resolution	2.12
Selective fusion on: baseline + Multi-resolution	1.98

The proposed scheme is well suited to surveillance applications. In this scenario, the presented sources of information can be easily extracted and are normally long enough to match the requirements for efficient fusion. In addition, the important function of detecting stressed conversation is already embedded in this scheme. Moreover, explicit stress detection can be achieved, pre-defining 'stressed' and 'non-stressed' conversation clusters within the selective fusion scheme, instead of unsupervised clustering through k-means. In this case, although optimum performance is not guaranteed anymore, suspiciously stressed conversations may be easily detected. Future work will focus on optimization of attribute characterization and selection and splitting of current speech features such as dynamic and static prosody and distinct dimensions of phonetic and acoustic representations. A deeper evaluation of stressed voiced detection is still pending the collection of appropriate databases.

References

1. Reynolds D., Quatieri T., Dunn R.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing, Vol. 10, no. 1 (2000) 19–41
2. NIST - Speaker Recognition Evaluations: http://www.nist.gov/speech/tests/spk/index.htm
3. Zhou G., Hansen J.H.L., Kaiser J.F.: Nonlinear Feature Based Classification of Speech under Stress. IEEE Transactions on Speech & Audio Processing, Vol. 9, no. 2 (2001) 201-216
4. Campbell J., Reynolds D., Dunn R.: Fusing High- and Low-Level Features for Speaker Recognition. Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech), Geneva, Switzerland (2003) 2665-2668
5. Solewicz Y. A., Koppel M.: Enhanced Fusion Methods for Speaker Verification. 9th International Conference "Speech and Computer" (SPECOM'04), St. Petersburg, Russia (2004) 388-392

6. Doddington G.: Speaker Recognition based on Idiolectal Differences between Speakers. Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), Aalborg, Denmark (2001) 2517-2520
7. Auckenthaler R., Carey M., Lloyd-Thomas H.: Score Normalization for Text-Independent Speaker Verification Systems. Digital Signal Processing, Vol. 10 (2000) 42–54
8. Andrews W. D., Kohler M. A., Campbell J. P., Godfrey J. J., Hernández-Cordero J.: Gender-Dependent Phonetic Refraction for Speaker Recognition. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP), Orlando, Florida (2002) 149-152
9. Joachims T.: Making large-Scale SVM Learning Practical. In: Schölkopf B. and Burges C., Smola A. (eds.): Advances in Kernel Methods - Support Vector Learning. MIT-Press (1999)
10. Ramaswamy G., Navratil J., Chaudhari U., Zilca R., Pelecanos J.: The IBM Systems for the NIST 2003 Speaker Recognition Evaluation. NIST-2003 Speaker Recognition Workshop, College Park, Maryland (2003)
11. Przybocki M., Martin A.: The NIST Year 2001 Speaker Recognition Evaluation Plan. http://www.nist.gov/speech/tests/spk/2001/doc/ (2001)
12. SWITCHBOARD: A User's Manual. Linguistic Data Consortium, http://www.ldc.upenn.edu/readme_files/ switchboard.readme.html

A New Conceptual Framework to Resolve Terrorism's Root Causes

Joshua Sinai

ANSER (Analytic Services)
joshua.sinai@verizon.net

To effectively resolve the violent challenges presented by terrorist groups to the security and well-being of their state adversaries, it is crucial to develop an appropriate understanding of all the root causes underlying such conflicts because terrorist insurgencies do not emerge in a political, socio-economic, religious or even psychological vacuum. It could be argued, in fact, that the root causes underlying an insurgency are the initial components driving the terrorist life cycle (TLC) and the terrorist attack cycle (TAC). The TLC refers to why and how terrorist groups are formed, led and organized, the nature of their grievances, motivations, strategies and demands vis-a-vis their adversaries, and the linkages that terrorist groups form with their supporting constituency. These components of the TLC, in turn, affect the TAC—a group's modus operandi, how they conduct the spectrum of operations, ranging from non-violent to violent activities, and their choice of weaponry and targeting.

To understand the context in which root causes relate to the TLC and TAC, it is necessary to conduct a comprehensive study of the magnitude of the warfare threat posed by a terrorist group against its adversary. The manifestations of the threat would then be 'drilled down' into their component warfare elements, such as conventional low impact (CLI) (e.g., warfare in which a few persons will be killed in a single attack involving conventional weapons warfare, such as explosives or shootings), conventional high impact (CHI) (e.g,, warfare in which conventional means are used to cause hundreds or thousands of fatalities), or warfare employing chemical, biological, radiological or nuclear (CBRN) (e.g., utilizing 'unconventional' means to inflict catastrophic damages). It is here, for example, where the latest advances in social science conceptual approaches, such as social network theory, would be applied to model how terrorist groups organize themselves, plan attacks, conduct recruitment, develop operational capabilities, link up with counterparts, etc. Other components of the TLC and TAC also would need to be addressed, such as why certain groups choose to embark on 'martyr'-driven suicide terrorism, as opposed to other forms of warfare where operatives seek to stay alive and escape from the scene of the incident.

Once the magnitude of the terrorist threat is identified and outlined (i.e., whether conventional low impact, conventional high impact, CBRN or a combination of the three), then one could begin the process of trying to understand the underlying conditions, or root causes, for why such warfare is being waged against a specific adversary (or adversaries). Thus, to attain the capability to anticipate and, in the most ideal cases preemptively contain or defeat on-going or emerging terrorist insurgencies, understanding the root causes underlying such conflicts must constitute the first line of analysis in a government's combating terrorism campaign's strategies and programs.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 280 – 286, 2005.

Therefore, to resolve terrorist insurgencies it is essential to research and systematically map the spectrum of root causes underlying a rebellion's origins, grievances and demands. In ideal cases, it is hoped that such mapping of root causes will then produce the knowledge and insight on the part of governments to formulate appropriate responses that would be most effective in terminating a terrorist insurgency, whether peacefully, militarily, by law enforcement, or through a combination of these measures. By incorporating such an understanding of a conflict's underlying root causes into a government's combating terrorism campaign, such response strategies and tactics could be effectively calibrated to address their specific challenges and threats. It is this paper's objective to provide an analytic framework to enable the combating terrorism community, whether in government or the academic sectors, to develop the conceptual capability and tools to resolve terrorist insurgencies using the most appropriate mix of coercive and conciliatory measures that address the general and specific root causes and other underlying factors that give rise to such insurgencies. Without understanding how to utilize such a root causes-based conceptual capability and tools, combating terrorism campaigns are likely to be ineffectual and terrorist insurgencies will become, due to lack of effective resolution, increasingly protracted and lethal in their warfare.

Why Root Causes Are Significant?

Terrorists, whether operating as small or large groups, are generally driven to commit acts of terrorism due to a variety of factors, whether rational or irrational, in which extreme forms of violence are utilized to express and redress specific grievances and demands. Root causes are the factors and circumstances underlying insurgencies that radicalize and drive terrorists – whether they are consciously or unconsciously aware of these root causes – into carrying out their violent actions. Root causes consist of multiple combinations of factors and circumstances, ranging from general to specific, global, regional or local, governmental-regime, societal or individual levels, structural or psychological, dynamic or static, facilitating or triggering, or other possible variations, some of which may be more important and fundamental than others.

Addressing a conflict's underlying root causes may not necessarily automatically lead to conflict termination. First, there may not be a direct correlation in every case between a specific root cause and a terrorist rebellion because of the myriad of alternative forms of action, ranging from non-violent to violent, that may be available to a group to express the underlying grievances and demands driving their group. In fact, a terrorist rebellion is likely to occur only when certain significant propitious circumstances in the form of political, economic, social, military, and other underlying trends coincide and coalesce, but even these trends may not be sufficient to launch such rebellions unless they are buttressed by the availability of effective leaders, organizational formations, including a willing cadre, access to particular types of weaponry and the logistical and other covert capabilities to carry out an operation against its adversary. Second, root causes should not be viewed as necessarily static, with some of the root causes that might play a significant role in the initial phase of a conflict becoming peripheral later on, while other root causes may emerge as paramount at a later phase in a conflict.

Thus, it is important to understand and map the spectrum of root causes underlying all phases in a terrorist rebellion because of their impact on future directions, includ-

ing influencing a group's choice of targeting and degree of lethality in its warfare. In fact, the intensity of how a group perceives its adversary and the strategies that it believes are required to redress the grievances against it, will affect the types of tactics and weaponry it will employ in the warfare against its stronger adversary. This is particularly the case in determining whether a group's warfare proclivity will be characterized by conventional low impact, conventional high impact, or CBRN warfare, with the latter form of warfare exponentially escalating the lethality threshold of casualties.

Another important consideration in understanding a rebellion's underlying root causes is to identify them from the varying perspectives of the insurgents, the threatened governments, and independent academic experts (who are likely to disagree among themselves) because these three general perspectives are likely to differ and, in some cases, even clash. Thus, for example, what the insurgents consider to be the underlying causes to their rebellion may be perceived entirely differently by the challenged government, which may deny the existence of such underlying factors. For example, while both Palestinians and Israelis agree that the central root cause underlying their conflict is the contention by two peoples over the same territory, there is disagreement over other possible root causes. To the Palestinian insurgents, the continued presence and expansion of Israeli settlements in the heart of the West Bank is claimed to constitute one of the primary root causes of their rebellion, whereas certain factions in the Israeli government may claim that such settlements should remain and are not an obstacle to reaching a peace accord. Independent academic experts may agree that such settlements may in fact represent an important root cause driving the conflict, because of the refusal by a minority of Israelis, in the form of the Jewish settlers, to give up their idea of a "Greater Israel" and live within the pre-June 1967 War confines of the Jewish state. At the same time, academic experts may find that the Palestinian insurgents engage in subterfuge on this issue because even if the settlements were evacuated many Palestinians would still refuse to ever recognize the legitimacy of Israeli rights to a homeland in a re-partitioned historical Palestine. Moreover, the Israeli government and academic experts, but not the Palestinian insurgents, may argue that an important root cause is the unresolved generational conflict among the Palestinians, with the younger generation, which is highly frustrated, much more militant and extremist than their elders, desiring to impose an Islamic theocracy over Palestinian society and reject a negotiated compromise with Israel.

To bridge the different interpretations between a government and its insurgent adversary, it is necessary for academic experts, who, as pointed out earlier, may even disagree among themselves, to provide as independent, impartial and objective as possible assessments of a conflict's root causes in order to assist the two adversaries to better understand the underlying problems that require resolution of their conflict.

In another important step, identifying and categorizing a conflict's underlying root causes will make it possible to hypothesize whether or not it may be possible to influence or resolve them so that long-term insurgency termination may take hold.

How to Resolve Root Causes?

Once the spectrum of a conflict's underlying root causes are mapped and identified—initially, as in most cases, at the academic level, and then at the governmental level, then it is up to governments and their security and military organizations to formulate the appropriate combating terrorism response measures to resolve these

underlying problems. For the underlying factors to be resolved, however, it is also up to the insurgents to incorporate into their demands grievances and other objectives that are amenable to the 'give and take' of compromise and negotiations because otherwise even addressing a conflict's root causes may not succeed in terminating the insurgency.

In this analytic approach, a government's combating terrorism campaign against an insurgent movement that utilizes terrorist tactics in order to overthrow that government, punish it for alleged transgressions, or seek independence against foreign rule, must be comprehensive and holistic in scope. This is because resolving terrorist insurgencies requires a much more thoroughgoing response than the narrower military or law enforcement orientations of most counter-guerrilla or counter-terrorist operations, which generally do not include crucial political, diplomatic, and socio-economic dimensions that are required to resolve a conflict's underling root causes.

The objective of the government's combating terrorism campaign therefore is to employ a mix of coercive (e.g., military or law enforcement) and conciliatory (e.g., political, diplomatic, or socio-economic) measures that either will militarily defeat the insurgents on the battlefield or peacefully terminate the insurgency by resolving the root causes and conditions that may prolong the conflict.

A successful combating terrorism campaign that seeks to address a conflict's underlying root causes must be based on the following three measures:

First, governments need to map, identify and prioritize what they consider to be the most significant underlying root causes driving the terrorist insurgency threatening them. To conduct such an assessment combating terrorism planners need to take into account their own perspectives, those of the insurgents, and academic experts. Once such a prioritized assessment is finalized, then the most appropriate measures need to be formulated on how these discrete root causes can be influenced and resolved. In fact, in determining the root causes associated with a terrorist insurgency, it is crucial to map all possible root causes, not just a select few that may be perceived as most likely. Such a comprehensive mapping effort will then generate the basis from which one could select those root causes whose resolution might yield the greatest benefit to eventual conflict termination. In this process, all perceived underlying root causes in a conflict would be itemized and categorized (e.g., poverty, lack of education, political inequality, foreign subjugation, religious extremism, psychological distress, nihilism, etc.) and codified (e.g., 1st order root cause, 2nd order root cause, 3rd order root cause, etc.).

Second, governments then need to formulate a clear definition in their directives and policies about the combating terrorism campaign's short-, medium- and long-term strategic objectives, including, as the final component, formulating a methodology to measure the effectiveness of their responses to the underlying root causes driving the terrorist insurgency. This involves formulating a mission area assessment that provides a roadmap for how strategic objectives can be implemented tactically on the ground for insurgency resolution to take place.

Third, the combating terrorism campaign must be coordinated and integrated at all levels of government, especially among the political, diplomatic, law enforcement, intelligence, and military establishments, resulting in a 'unity of effort.'

In ideal cases, when such a three-pronged combating terrorism campaign is implemented, in situations where an insurgent conflict is caused by political or socio-

economic deprivations or disparities that are exploited by the insurgents, a government's conciliatory policies that address and resolve that conflict's root causes is likely to succeed in peacefully winning the affected populations 'hearts and minds.' Also in ideal cases where a foreign power controls a territory that is inhabited by a hostile population, a combating campaign's conciliatory components is likely succeed in terminating the insurgency by providing autonomy or independence to that territory, following a consensual peace accord between the government and the insurgents.

Thus, in ideal cases, a conflict resolution-based combating terrorism strategy is likely to be the most effective way to resolve a protracted terrorist-based insurgency where the insurgents represent 'genuine' grievances that succeed in mobilizing the local population to support their cause. This does not imply that coercive measures are not necessary as an initial governmental response to nip the insurgency in the bud. In fact, during the initial phase, coercive measures are required to counteract the insurgency's violent threats to the maintenance of law and order. These coercive measures will likely take the form of military, police, and intelligence operations against the insurgent forces; governments will insist that no concessions be made to insurgent demands, which they perceive as illegitimate because violent means are used to express them; insurgent movements will be declared illegal; a state of emergency accompanied by prevention of terrorism laws will be imposed, particularly in insurgent areas; and diplomatic pressure will be exerted on the external patrons or supporters of the insurgency to cease such support.

While these coercive measures may be necessary in the initial stages of an insurgency, there are limits to the degree of coercion that democratic governments will employ in their combating terrorism campaign. Thus, for example, democratic governments, such as Israel, will refrain from employing crushing military force to wipe out civilian populations that provide the insurgents with support because of the damage that such devastation would inflict on their own democratic constitutional nature. This is the situation currently confronting Israel in its response to the al-Aqsa Intifada, where even the deployment of massive Israeli military force in Spring 2002 against Palestinian cities and towns in response to devastating Palestinian suicide terrorism against Israelis was not intended to massacre Palestinian civilians, but to ensure that terrorists, their operational handlers, and infrastructures were uprooted and destroyed so that a political settlement might be possible when conditions were considered ripe.

Moreover, even during the initial coercive phase of their response, democratic governments are likely to include certain limited conciliatory measures. These conciliatory measures will be restricted in scope, and will likely consist of limited degrees of political, legal, and socio-economic reforms, including permitting human rights groups to monitor the impact of the combating terrorism campaign on the affected population.

Authoritarian governments, on the other hand, are less inclined to act with such restraint against civilian supporters of an insurgency, as demonstrated by the crushing by Syrian forces of the Muslim Brotherhood insurgents in Hama in 1982, the Iraqi use of chemical weapons against the Kurdish villagers in early 1988, the 1998 bombardment by Serbian forces against the rebellious ethnic Albanian villagers in Kosovo, and Russia's military campaign against the Chechnyan terrorist separatists.

However, when an insurgency, even when it employs terrorism to achieve its objectives, succeeds in gaining the support of a significant segment of the population to its cause and in protracting the insurgency, and the government's coercive measures, accompanied by limited conciliation, are unable either to decisively defeat the insurgents on the battlefield or to resolve the insurgency peacefully, then a new combating terrorism strategy is required to resolve the conflict. Based on my research, I believe that in a situation of a protracted 'hurting stalemate' that is damaging to both sides, in which there is no military solution to end the insurgency, long-term resolution can only come about when governments begin to address the conflict's underlying root causes – but only when the insurgents' grievances are considered legitimate and grounded in some aspects of international law.

This recommendation does not imply that resolving a conflict's root causes will automatically terminate the insurgency peacefully. Some insurgent movements are inherently extremist and not interested in compromising their demands, such as militant religious fundamentalists who are intent on establishing highly authoritarian theocratic states (e.g., in Algeria, Egypt, Jordan and Lebanon), or are filled with unrelenting rage against a superpower (e.g., Usama bin Laden's al-Qaida's group and its network of affiliates), while other insurgents may be use narcotrafficking means to fund their political activities (e.g., the FARC in Colombia). Thus, in such cases no peaceful accommodation may be possible between governments and insurgents even when governments are willing to resolve a conflict's 'root causes,' such as socio-economic and political inequalities.

One way to determine whether it is possible for governments and insurgents to arrive at a negotiated compromise is by distinguishing between insurgents' legitimate and illegitimate grievances. Legitimate grievances may be defined as those that are anchored in international law, particularly in the areas of constitutionalism and human rights, and are politically, legally, economically, and geographically equitable to all relevant parties affected by the conflict. Illegitimate grievances, on the other hand, generally are based on anti-democratic, theocratic, religiously exclusionary, or criminal principles and objectives, as well as desiring the destruction or annihilation of the adversary.

Because of the different responses that are necessary to address legitimate and illegitimate demands being espoused by terrorist groups, employing conciliation to resolve a terrorist rebellion can be applied to certain types of insurgencies, but not others. In the case of the insurgency mounted by al-Qaida, for example, there may be no alternative but to pursue a full-scale military campaign, backed by intelligence and law enforcement measures, to round up as many of their insurgents as possible, because of their operatives' single-minded pursuit of causing as much catastrophic damage to their adversaries as possible, regardless of the consequences to their own societies. In fact, even under these circumstances, it is still possible to address the underlying conditions that facilitate recruitment and support for al-Qaida (such as the prevalence of Arab regimes that stifle opportunities for educated youths to attain socio-economic and political advancement) without giving in to al-Qaida's demands or long-term goals.

Similarly, for Israel, while it may be difficult to negotiate with insurgents such as Hamas, the Palestinian Islamic Jihad and the al Aqsa Martyrs Brigade, because of their determination to sabotage all efforts at a peace process by launching wave upon

wave of suicide bombers to achieve their goal of a theocratic Palestinian state in all of historical Palestine, the underlying conditions that perpetuate that conflict still need to be addressed. Thus, in spite of extremist demands by its terrorist adversaries, Israeli counterterrorism planners must map that conflict's root causes in order to generate responses that will effectively terminate or mitigate that insurgency. For example, if the presence of Jewish settlers in the heart of Palestinian territories in the West Bank and Gaza Strip is considered to constitute one of the underlying root causes for continued Palestinian hostility, then evacuating and resettling those settlers in Israel 'proper' may prove to be a solution to addressing those Palestinian demands that may be judged to be 'legitimate. In fact, there is a substantial segment of the Israeli leadership that supports the notion of 'unilateral disengagement' from such territories, even without a negotiating process with a counterpart Palestinian peace partner. Fortunately, such a conciliatory approach began to take shape when Prime Minister Ariel Sharon in early 2005 set the stage for the uprooting of Jewish settlements from the Gaza Strip as part of the Israeli disengagement from that territory, which coincided with the election of Mahmoud Abbas as President of the Palestinian Authority. This ushered in a more conciliatory era between the Palestinians and Israelis, which many hoped would result in finally resolving their long standing conflict.

However imperfect such an approach to conflict resolution exhibited by the Israelis, at least it recognizes that certain underlying problem areas can be resolved without appeasing the insurgents' extremist demands. Here, as in other cases, intransigence by insurgents should not preclude the need for the threatened governments facing protracted insurgencies to strive to resolve their conflicts' underlying problems by using as many creative and 'out of the box' measures as possible, because the alternative is continued suffering for all contending sides.

Analyzing Terrorist Networks: A Case Study of the Global Salafi Jihad Network

Jialun Qin[1], Jennifer J. Xu[1], Daning Hu[1], Marc Sageman[2], and Hsinchun Chen[1]

[1] Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721, USA
{qin, jxu, daning, hchen}@eller.arizona.edu
[2] The Solomon Asch Center For Study of Ethnopolitical Conflict, University of Pennsylvania, St. Leonard's Court, Suite 305, 3819-33 Chestmut Street, Philadelphia, PA 19104, USA
sageman@sas.upenn.edu

Abstract. It is very important for us to understand the functions and structures of terrorist networks to win the battle against terror. However, previous studies of terrorist network structure have generated little actionable results. This is mainly due to the difficulty in collecting and accessing reliable data and the lack of advanced network analysis methodologies in the field. To address these problems, we employed several advance network analysis techniques ranging from social network analysis to Web structural mining on a Global Salafi Jihad network dataset collected through a large scale empirical study. Our study demonstrated the effectiveness and usefulness of advanced network techniques in terrorist network analysis domain. We also introduced the Web structural mining technique into the terrorist network analysis field which, to the best our knowledge, has never been used in this domain. More importantly, the results from our analysis provide not only insights for terrorism research community but also empirical implications that may help law-reinforcement, intelligence, and security communities to make our nation safer.

1 Introduction

Terrorism threats span personal, organizational, and societal levels and have far-reaching economic, psychological, political, and social consequences. Only with a thorough understanding of terrorism and terrorist organizations can we defend against the threats. Because terrorist organizations often operate in a network form in which individual terrorists cooperate and collaborate with each other to carry out attacks [24], we could gain valuable knowledge about the terrorist organizations by studying various structural properties of terrorist networks. Such knowledge may help authorities develop efficient and effective disruptive strategies and measures.

However, even though the terrorism-related research domain has experienced tremendous growth since September 11th, studies of terrorist network structure have generated little actionable results. This is due to the difficulty in collecting and accessing reliable data and the lack of advanced network analysis methodologies in the field. Do terrorist networks share the same topological properties with other types of

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 287 – 304, 2005.

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 paper a case study of the analysis 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 and Web structural mining. We consider our case study 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 reliable data collected in a large-scale in-depth empirical study on the GSJ network [33]. Second, our study introduced multiple advanced network analysis methodologies into the study of terrorist networks including the Web structural mining techniques which, to the best our knowledge, has never been used in this domain. Third, our results provide not only insights for terrorism research community but also empirical implications that may help law-reinforcement, intelligence, and security communities to make our nation safer.

The remainder of the paper is organized as follows. Section 2 reviews various network analysis studies in different domains in relation to terrorist network analysis. In section 3, we provide some background information on the GSJ network and briefly describe how the GSJ network dataset was collected through the empirical study. In section 4, we present our methodologies and report our findings from the analysis. Section 5 concludes this paper with implications and future directions.

2 Literature Review

In this section, we review a few network analysis methodologies widely employed in other domains: social network analysis, statistical analysis of network topology, and Web link structure analysis. These techniques can be used to analyze terrorist networks. Different techniques reveal different perspectives of terrorist networks.

2.1 Social Network Analysis

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 [35, 36, 40]. A number of quantitative SNA methods have been employed to study organizational behavior, inter-organizational relations, citation analysis, computer mediated communication, and many other domains [18, 19, 22]. SNA has recently been recognized as a promising technology for studying criminal organizations and enterprises [27, 38]. Studies involving evidence mapping in fraud and conspiracy cases have recently been added to this list [6, 34].

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 pattern between sub-groups in terrorist networks. Several centrality measures can be used to identify key members who play important roles in a network. Freeman [16] 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, $c(i, a)$ is 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 is called "farness" [6, 16, 17]:

$$C_C(a) = \sum_{i=1}^{n} l(i,a)$$

where $l(i,a)$ is the length of the shortest path connecting nodes *i* and *a*.

Blokmodeling is another technique used in SNA to model interaction between clusters of network members. Blockmodeling chould reduce a complex network to a simpler structure by summarizing individual interaction details into relationship patterns between positions [42]. As a result, the overall structure of the network becomes more evident. In blockmodeling, a network is first partitioned into positions based on structural equivalence measure [26]. Two nodes are structurally equivalent if they have identical links to and from other nodes. Since perfectly equivalent members rarely exist in reality, this measure is relaxed to indicate the extent to which two nodes are substitutable in structure [40]. A position thus is a collection of network members who are structurally substitutable, or in other words, similar in social activities, status, and connections with other members. Position is different from the concept of subgroup in relational analysis because two network members who are in the same position need not be directly connected [26, 35].

After network partition, blockmodel analysis compares the density of links between two positions with the overall density of a network [5, 9, 42]. Link density between two positions is the actual number of links between all pairs of nodes drawn from each position divided by the possible number of links between the two positions. In a network with undirected links, for example, the between-position link density can be calculated by $d_{ij} = \frac{m_{ij}}{n_i n_j}$, where d_{ij} is the link density between positions *i* and *j*; m_{ij} is the actual number of links between positions *i* and *j*; n_i and n_j represent the number of nodes within positions *i* and *j*, respectively. The overall link density of a network is defined as the total number of links divided by the possible number of links in the

whole network, i.e., $d = \frac{m}{n(n-1)/2}$ where m is the total number of links; n is the total number of nodes in the network. Notice that for an undirected network the possible number of links is always $n(n\text{-}1)/2$.

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 $d_{ij} \geq d$, and absent otherwise.

In this study, however, we used blockmodeling to extract interaction pattern 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 [8, 14], small-world model [41], and scale-free model [7]. 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 [8], 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 [1]. The small-world model and scale-free model are significantly deviant from the random graph model [1, 29]. A small-world network has a significantly larger clustering coefficient [41] than its random model counterpart while maintaining a relatively small average path length [41]. The large clustering coefficient indicates that there is a high tendency for nodes to form communities and groups [41]. Scale-free networks [7], on the other hand, are characterized by the power-law degree distribution, meaning that while a big fraction of nodes in the network have just a few links, a small fraction of the nodes have a large number of links. It is believed that scale-free networks evolve following the self-organizing principle, where growth and preferential attachment play a key role in the emergence of the power-law degree distribution [7].

The analysis on the topology of complex systems has important implications to our understanding of nature and society. Research has shown that the function of a complex system may be to a great extent affected by its network topology [1, 29]. For instance, the small average path length of the World Wide Web makes cyberspace a very convenient, strongly navigable system, in which any two web pages are on average only 19 clicks away from each other [2]. It has also been shown that the higher tendency for clustering in metabolic networks is correspondent to the organization of functional modules in cells, which contributes to the behaviour and survival of organisms [30, 31]. In addition, networks with scale-free properties (e.g., protein-protein interaction networks) are highly robust against random failures and errors (e.g., mutations) but quite vulnerable under targeted attacks [2, 21, 37].

2.3 Web Structural Analysis

The Web is one of the largest and most complicated networks in the world. The Web, as a network of Web pages 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 [20]. 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, Web pages 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. Web pages are co-cited when they are linked to by the same set of parent Web pages and heavily co-cited pages are often relevant to the same topic. Co-citation is particularly helpful in finding relevant pages in some domains where pages with similar contents avoid linking to each other (e.g., commercial domains where providers of similar online contents are competitors). Researchers have developed many algorithms to judge the importance and quality of Web pages 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 [13]. The quality of these referring pages is also determined by PageRank. Thus, the PageRank of a page *p* is calculated recursively as follows:

$$\text{PageRank(p)} = 1 - d + d \times \sum_{\text{all q link to p}} \frac{\text{PageRank(q)}}{\text{c(q)}}$$

where d is a damping factor between 0 and 1 and c(q) is the number of out-going links in q.

PageRank is originally designed to calculate the importance of Web pages based on the Web link structure and is used in the commercial search engine Google [10] 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. In a co-authorship network, a link between authors implies the mutual endorsement relationship between them and the PageRank algorithm can be used to rank the authors based their importance in this co-authorship network. In the co-authorship analysis study conducted by Liu et al. [25], PageRank was used as one of the author ranking criteria along with other traditional SNA centrality measures. Similarly, we believe that PageRank can also used to rank the importance of terrorists within a properly constructed terrorist network.

3 Global Salafi Jihad Network

The Global Salafi Jihad (GSJ) is part of a violent worldwide Muslim revivalist movement. It is a new form of terrorism which threatens the worlds in different and horrifying ways from previous forms of this scourge. While all forms of terrorism

result in tragedies, the GSJ is driven by a fanatical determination to inflict maximum civilian and economic damages. It mainly targets the West, but its reckless operations and indiscriminately slaughter masses of humanity of all races and religions. With Al Qaeda as its vanguard, the GSJ includes many terrorist groups with members from different countries and forms a large global terrorist network. Through this network, the GSJ have successfully planed and launched many large-scale attacks against civilians across different countries. Examples include the 9/11 tragedy in 2001, the bombing in Bali in 2002, and the bombing in Morocco in 2003.

Collecting data on the GSJ terrorist presents many challenges, mostly because of a general lack of information. The GSJ data we used in this study was collected through a long-term empirical study on the GSJ members. The sources of information we used to collect data from were all in the public domain. The information was often inconsistent. We considered the source of information in selection facts to include in the dataset. In decreasing degrees of reliability, the information sources we favored 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. Data collected from these multiple sources were cross-validated to ensure maximum accuracy.

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. The central member clump mainly consists of the key Al Qaeda members. They 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 collected 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 (e.g., mental illness, personality, pathological narcissism, etc) features that could be the explanations of why these people became terrorists. More importantly, the data also captures all know 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. The dataset is presented in a form a spreadsheet with each raw containing the basic features of a certain GSJ member as well as all the other members that are related to this member through the various relationships or interactions mentioned above. We then calculated the "distance" between each pair of terrorists in the network based on the number of relationships between them and visualized the network using multi-dimensional scaling (MDS) technique. Our visualization provides an intuitive and clear view of the overall GSJ network (See Figure 1).

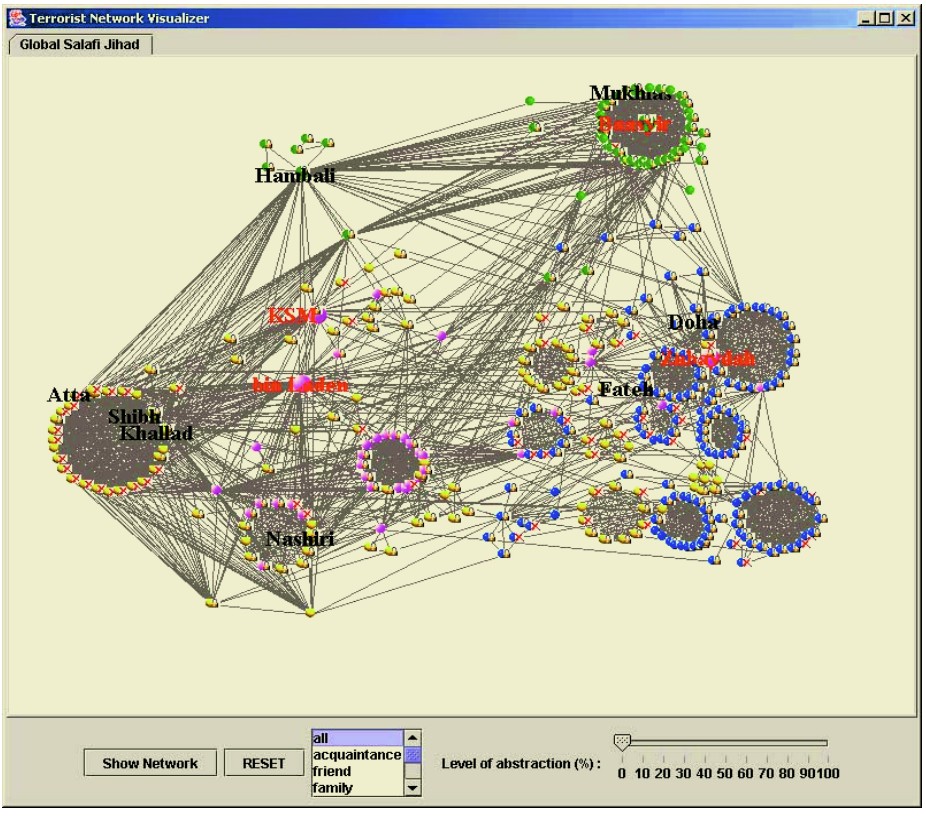


Fig. 1. Visualization of the Full GSJ Network

Figure 1 is the visualization of the GSJ network with all types of relations. Each node represents a terrorist. 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.

4 GSJ Network Analysis

To better understand how the GSJ network works, we employed the aforementioned network analysis techniques on the GSJ network dataset which are SNA, statistical analysis, and Web structural mining. In this section, we describe our analysis procedures and report our findings.

4.1 Social Network Analysis

This first analysis we conducted on the GSJ dataset was a social network analysis in which we used the centrality measures and block-modeling 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 person 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 1 summarizes the top 10 terrorists ranked by the 3 centrality measures in each of the 4 clumps.

Table 1. Terrorists with Top Centrality Ranks within Each Clump

Ranking	Leader	Gatekeeper	Outlier
Central Member			
1	Zawahiri	bin Laden	Khalifah
2	Makkawi	Zawahiri	SbinLaden
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

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 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.

To get a better understanding of how terrorists plan and coordinate attacks, we conducted more in-depth SNA on selected terrorist attack cases. Two large-scale terrorist attacks were selected: the Strasbourg cathedral bombing plot and the September 11th. For each case, we extract from the GSJ dataset the 1-hop network formed by the GSJ members who have participated in the attacks as well as those who were directly involved but had direct relationships with any of the involved members. Centrality analysis and blockmodeling analysis were then employed on these four 1-hop networks to identify key members who may have led or coordinated these attacks as well as to identify as well as the connections between different attacks.

The first case we investigated was the Strasbourg cathedral bombing plot in which was a plan to blow up a cathedral in Strasbourg, France by Al-Qaida in December 2000. We visualized the 1-hop network of the Strasbourg plot and showed it in Figure 2. In Figure 2, red nodes represent members who were directly involved in the plot

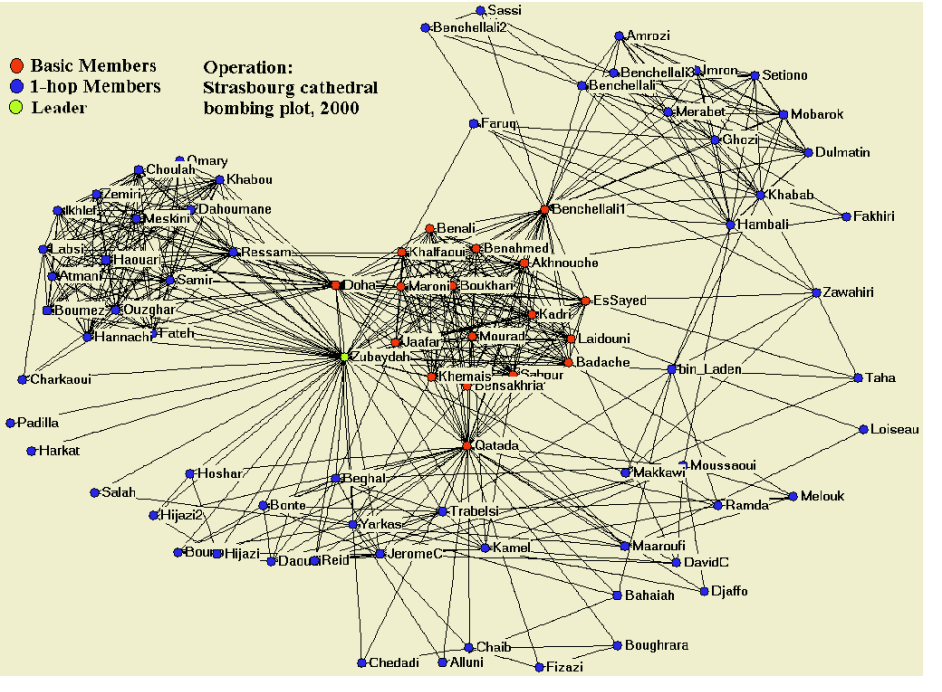


Fig. 2. The 1-hop Network of the Strasbourg Cathedral Bombing Plot

and blue nodes represent members who were 1-hop away from the directly involved members. Yellow node represents the leader of the attack identified by the SNA based on centrality measures.

In the Strasbourg plot network, Zubaydah Zain al-Abidin was identified by SNA as the leader (highest degree) and the gatekeeper (highest betweenness) of the Strasbourg plot. This means Zubaydah led the Strasbourg plot and he also acted as the major connector between the Strasbourg group and other GSJ terrorist. He also has the lowest closeness which means that he was in an excellent position to monitor the information flow in whole Strasbourg plot network. Intelligence collected by authorities during the investigation of the Strasbourg plot agrees with the SNA results pretty well. In fact, Zubaydah was a high-ranking member of al-Qaida and close associate of Osama bin Laden. And he is the highest-ranking al-Qaida leader in U.S. custody. Zubaydah was involved in all five Maghreb Arab plots (including the LAX airport bombing plot) as the central coordinator for Al Qaeda. These facts explain his high degree and high betweenness in the network. Zubaydah's role in Al Qaeda was to welcome new recruits in Peshawar. He was also the first one to brief them about security arrangements and screened the new arrivals. As such, he probably met with all the new recruits in Al Qaida. This may explain why he had low closeness in the network.

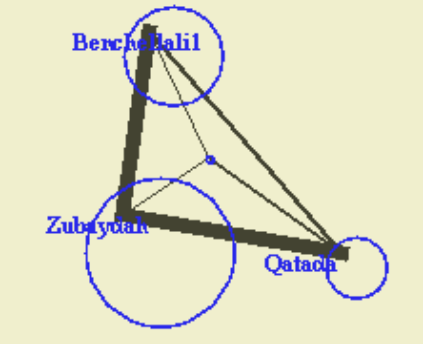


Fig. 3. Blockmodeling of the Strasbourg Cathedral Bombing Plot Network

Blockmodeling analysis was conducted on the Strasbourg plot network to identify the interaction patterns between the sub-groups within this network. In the blockmodeling result (See Figure 3), a circle represents a sub-group of terrorsts. In the Strasbourg plot network, three sub-groups were detected. They were the group that conducted the Strasbourg plot, the group that conducted the LAX airport bombing plot in 2001, and the group that conducted the Russian embassy bombing plot. The blockmodeling result also confirmed that Zubaydah was the leader of the Strasbourg plot and he also connected the Strasbourg group to the other two groups in the network.

The second case we investigated was the September 11th attack. 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 1-hop network of the September 11th attack which contained 161 members and covered nearly half of the whole GSJ network (See Figure 4).

SNA identified Osama bin Laden (the yellow node) as the leader and gatekeeper of the September 11th attack 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 directly participated in the attacks to all the four clumps of the GSJ network, which indicates a world-wide cooperation in the planning of the attacks.

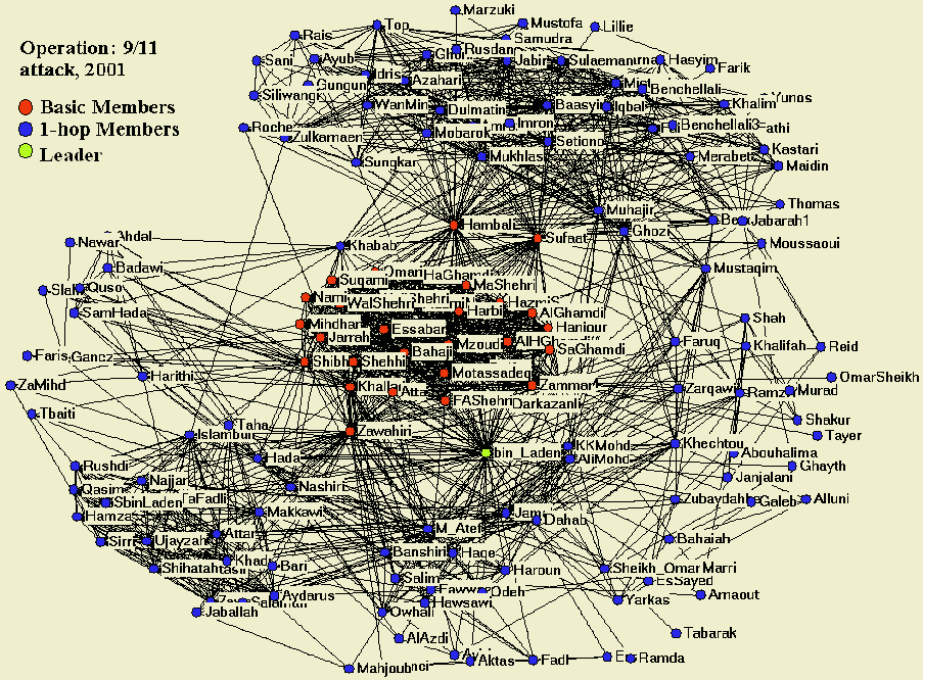


Fig. 4. The 1-hop Network of the September 11th Attack

In the blockmodeling result of the September 11th network (See Figure 5), 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 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 to be the link between KSM and the field as well as the general coordinator of the operation.

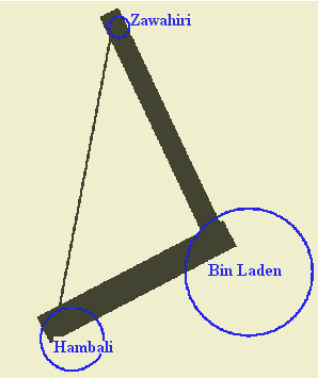


Fig. 5. Blockmodeling of the September 11th Attack Network

4.2 Statistical Analysis

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 2 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 2). The average path length and diameter [40] 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 [40]. 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.

Table 2. 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}

The other small-world topology, high clustering coefficient, is also present in the GSJ network (see Table 2). 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 networks [12, 33, 43, 44]. 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 [27].

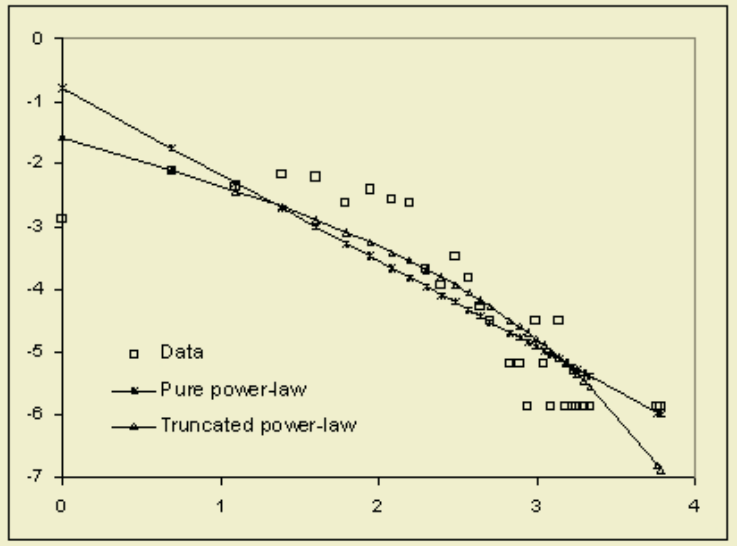


Fig. 7. Degree Distribution of the GSJ Network

Moreover, the GSJ network is also a scale-free system. The network follows an exponentially truncated power-law degree distribution [4, 28], $P(k) \sim k^{-\gamma} e^{-\frac{k}{\kappa}}$, with exponent γ=0.67 and cutoff κ=15.35. Different from other types of networks [2, 28, 41] 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. 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 [4]: (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 GSJ network where the terrorists preferred to attach to those who were their relatives, friends, or religious partners [33].

4.3 Web Structural Mining

In a social network, not all the members play equal roles. Instead, some members may have stronger social influences or higher social status than the others. In a terrorist network context, a terrorist may act as 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 to study the terrorist organization behavior. However, neither of our previous analysis allowed us to study the communication patterns in the GSJ network with such unequalized social relationships. To address this issue, we borrowed the 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, PageRank algorithm determines the importance of a specific member based on the structure of the whole network; while degree measure make the some judgment only based 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 important scores 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 report to. We then draw a directional link from each of the terrorists to their local leaders to visualize the terrorist social hierarchy and this graph is called a Authority Derivation Graph (ADG) [39]. Figure 6 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 as the role of "leader" and *A* acts as 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 type of relationship between the members. A thick link means there are personal relationships (kinship, family, friends, acquaintance, etc) 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. 2. The central component, consist-

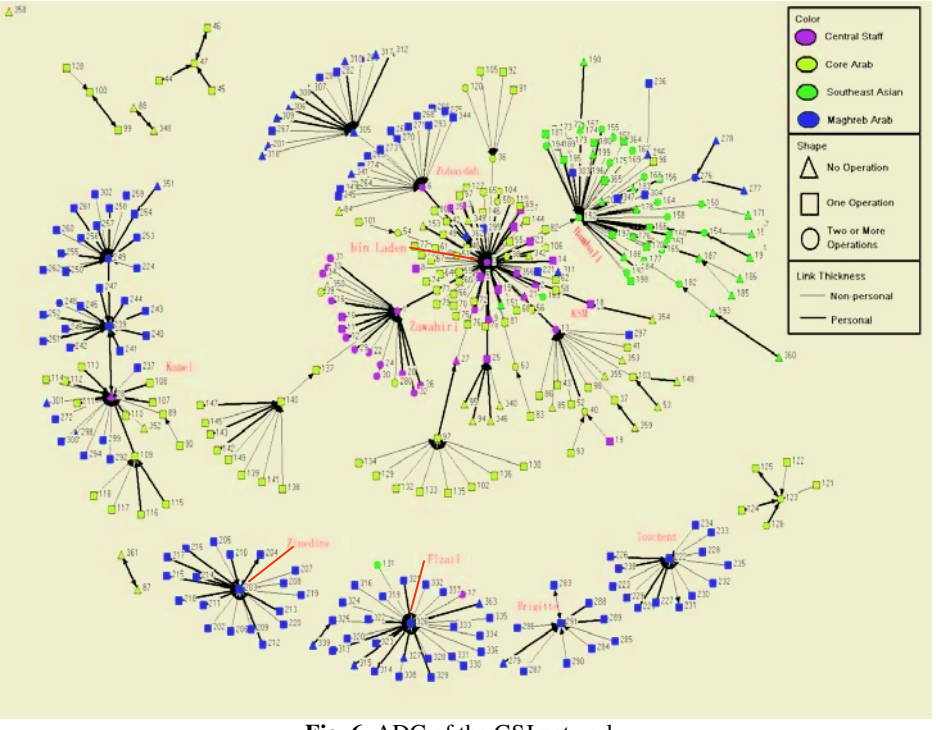


Fig. 6. ADG of the GSJ network

ing 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 3 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 functions 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 in-coming 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 trust-worthy "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.

5 Conclusions and Future Directions

It is very important for us to understand the functions and structures of terrorist networks to win the battle against terror. In this study, we employed several advanced network analysis techniques on a GSJ network dataset collected through a large scale empirical study. Our analysis results showed that centralities measures from 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. Such topological features help experts better understand how the GSJ network functions and also help them make more effective disruptive strategies. We also applied Web structural mining methodologies in the GSJ network analysis. This approach, to the best of our knowledge, has never been used in this domain before and it helps us study the terrorist organization structure under a social hierarchy assumption. This may provide insights into better understanding of terrorist organization behavior.

We have several future research directions to pursue. First, we are working with terrorism experts to fine tuning our algorithms to generate more accurate 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.

Acknowledgements

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. Joshua Sinai from the Department of Homeland Security, Dr. Rex A. Hudson from the Library of Congress, and Dr. Chip Ellis from the MIPT organization for their insightful comments and suggestions on our project. 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 Homa Atabakhsh, Cathy Larson, Chun-Ju Tseng, Ying Liu, Theodore Elhourani, Wei Xi, Charles Zhi-Kai Chen, Guanpi Lai, and Shing Ka Wu.

References

1. Albert, R., Barabasi A.-L.: Statistical Mechanics of Complex Networks. Reviews of Modern Physics **7**4(1) (2002) 47-97
2. Albert, R., Jeong, H., Barabasi, A.-L.: Diameter of the World-Wide Web. Nature 401 (1999) 130-131
3. Albert, R., Jeong, H., Barabasi A.-L.: Error and attack tolerance of complex networks. Nature 406 (2000) 378-382
4. Amaral, L. A. N., Scala, A., Stanley, H. E.: Classes of small-world networks. PNAS 97 (2000) 11149-11152
5. Arabie, P., Boorman, S. A., Levitt, P. R.f: Constructing blockmodels: How and why. Journal of Mathematical Psychology 17 (1978) 21-63
6. Baker, W. E., Faulkner, R. R.: The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry. American Sociological Review 58(12) (1993) 837-860
7. Barabasi, A.-L., Alert, R.: Emergence of Scaling in Random Networks. Science 286(5439) (1999) 509-512
8. Bollobas, B. Random Graphs. London, Academic, (1985)
9. Breiger, R. L., Boorman, S. A. Arabie, P.: An Algorithm for Clustering Relational Data, with Applications to Social Network Analysis and Comparison with Multidimensional Scaling. Journal of Mathematical Psychology 12 (1975) 328-383
10. Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems 30 (1998) 1–7
11. Burt, R. S.: Positions in networks. Social Forces 55 (1976) 93-122
12. Chen, H., Qin, J., Reid, E., Chung, W., Zhou, Y. Xi, W., Lai, G., Bonillas, A. A., Sageman, M.: The Dark Web Portal: Collecting and Analyzing the Presence of Domestic and International Terrorist Groups on the Web. Proceedings of The 7th Annual IEEE Conference on Intelligent Transportation Systems (ITSC 2004), Washington, D. C. (2004)
13. Cho, J., Garcia-Molina, H., Page, L.: Efficient Crawling through URL Ordering. Proceedings of the 7th International WWW Conference, Brisbane, Australia (1998)
14. Erdos, P., Renyi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5** (1960) 17-61
15. Evan, W. M.: An organization-set model of interorganizational relations. Interorganizational Decision-making. M. Tuite, R. Chisholm and M. Radnor. Chicago, Aldine (1972) 181-200
16. Freeman, L. C.: Centrality in Social Networks: Conceptual Clarification. Social Networks 1 (1979) 215-240
17. Freeman, L. C.: Visualizing social networks. Journal of Social Structure 1(1) (2000)
18. Galaskiewicz, J., Krohn, K.: Positions, roles, and dependencies in a community interorganization system. Sociological Quarterly 25 (1984) 527-550
19. Garton, L., Haythornthwaite, C., Wellman, B.: Studying online social networks. Doing Internet Research. S. Jones, Sage (1999)
20. Gibson, D., Kleinberg, J., Raghavan, P.: Inferring Web Communities from Link Topology. Proceedings of the 9th ACM Conference on Hypertext and Hypermedia, Pittsburgh, Pennsylvania, USA (1998)
21. Jeong, H., Mason, S. P., Barabasi A.-L., Oltvai, Z. N.: Lethality and centrality in protein networks. Nature 411(6833) (2001) 41
22. Kleinberg, J.: Authoritative Sources in a Hyperlinked Environment. Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA (1998)

23. Klerks, P.: The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands. Connections 24(3) (2001) 53-65
24. Krebs, V. E.: Mapping networks of terrorist cells. Connections 24(3) (2001) 43-52
25. Liu, X., Bollen, J., Nelson, M. L., Van de Sompel, H.: All in the Family? A Co-Authorship Analysis of JCDL Conferences (1994-2003). Proceedings of the IEEE/ACM Joint Conference on Digital Libraries 2004, Tucson, AZ (2004)
26. Lorrain, F. P., White, H. C.: Structural Equivalence of Individuals in Social Networks. Journal of Mathematical Sociology 1 (1971) 49-80
27. McAndrew, D.: The Structural Analysis of Criminal Networks. The Social Psychology of Crime: Groups, Teams, and Networks, Offender Profiling Series, III. D. Canter and L. Alison. Dartmouth, Aldershot (1999)
28. Newman, M. E. J. The structure of scientific collaboration networks. PNAS 98 (2001) 404-409
29. Newman, M. E. J. The structure and function of complex networks. SIAM Review 45(2) (2003) 167-256
30. Ravasz, E., A. L. Somera, Mongru, D. A., Oltvai, Z. N., Barabasi, A.-L.: Hierarchical Organization of Modularity in Metabolic Networks. Science 297 (2002) 1551-1555
31. Rives, A. W., Galitski, T.: Modular organization of cellular networks. PNAS 100(3) (2003) 1128–1133
32. Ronfeldt, D., Arquilla, J.: What next for networks and netwars? Networks and Netwars: The Future of Terror, Crime, and Militancy. J. Arquilla and D. Ronfeldt, Rand Press (2001)
33. Sageman, M.: Understanding Terror Networks. Philadelphia, PA, University of Pennsylvania Press (2004)
34. Saether, M., Canter, D. V.: A structural analysis of fraud and armed robbery networks in Norway. Proceedings of the 6th International Investigative Psychology Conference, Liverpool, England (2001)
35. Scott, J.: Social Network Analysis. London, Sage (1991)
36. Scott, M.: War's new front. CIO Insight (2001) 82-83
37. Solé, R. V., Montoya, J. M.: Complexity and fragility in ecological networks. Proc. R. Soc. B 268 (2001) 2039-2045
38. Sparrow, M. K.: The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects. Social Networks 13 (1991) 251-274
39. Toyoda, M., Kitsuregawa, M.: Creating a Web Community Chart for Navigating Related Communities. Proceedings of ACM Conference on Hypertext and Hypermedia, Århus, Denmark (2001)
40. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge, Cambridge University Press (1994)
41. Watts, D. J., Strogatz, S. H.: Collective Dynamics of 'Small-World' Networks. Nature 393 (1998) 440-442
42. White, H. C., Boorman, S. A.: Social Structure from Multiple Networks: I. Blockmodels of Roles and Positions. American Journal of Sociology 81 (1976) 730-780
43. Xu, J., Chen, H.: Untangling Criminal Networks: A Case Study. Proceedings of the 1st NSF/NIJ Symposium on Intelligence and Security Informatics (ISI'03), Tucson, AZ (2003)
44. Xu, J., Chen, H.: Criminal Network Analysis and Visualization: A Data Mining Perspective. Communications of the ACM (Forthcoming)

A Conceptual Model of Counterterrorist Operations

David Davis[1], Allison Frendak-Blume[1], Jennifer Wheeler[1],
Alexander E.R. Woodcock[2], and Clarence Worrell III[1]

[1] Peace Operations Policy Program,
George Mason University
{ddavis, afrendak, cworrel2, aerw}@gmu.edu
tiawheeler@earthlink.net
[2] Societal Dynamics Research Center,
George Mason University

Abstract. This paper describes the development of the Conceptual Model of Counter Terrorist Operations or the CMCTO. The CMCTO is a top down decomposition of the functions that are performed in the Counter Terrorist Domain. The models first decomposes the domain into Functions directed toward terrorists; Functions directed toward victims; and, Functions of support. Each of these functions is further decomposed to varying levels. The paper also includes a comprehensive review of the literature and of the process used.

1 Overview

Terrorism and the activities designed to counter terrorism are not new. However, in today's world it is important to ensure we understand these activities and their inter-relationships in a more profound and subtle way. In 2003, we were retained to develop a "Conceptual Model of Counterterrorist Operations" (CMCTO). This activity was a key component of a wider project titled *Implementation of a Conceptual Model of Counterterrorist Operations (CMCTO) as a Systems Dynamics Model*.

Our group has much experience with conceptual modeling, having performed similar work in the domain of peace operations over the last eight years. Recognizing that most experts in the field tend to focus on the concept of terrorism, in particular—How does one define terrorism? How does an individual become a terrorist?—the sponsor wished to see if our group could bring its conceptual modeling experience to bear in detailing operations seeking to counter terrorism.

2 Conceptual Modeling

In 1995, our group developed a Conceptual Model of Peace Operations (CMPO) to better study and understand the domain of peace operations. At the time, peace operations were relatively unknown in the military analysis community and what little investigation had been conducted on them had been performed in a flawed manner—

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 305 – 315, 2005.

shoe-horning the new query into old tools of analysis. At this time, the CMPO continues to serve as a framework for enhanced appreciation of the peace operations environment and as an organizing mechanism within which targeted analysis may be accomplished. Not only has it been used in an academic setting, but also in part, or in whole, in efforts and studies conducted by NATO Consultation, Command and Control Agency, the US Pacific Command, and World Vision International.

The general and integrated analysis of counterterrorism operations appears to be in much the same state today as peace operations was a decade ago. Similarly, many of the complex issues that our group earlier faced in portraying the peace operations domain also are present in the domain of counterterrorism operations: Multiple organizations with overlapping mandates; Definitional and vocabulary confusion; Lack of measures and metrics of success and progress; Functional relationships within and between stove-piped responses; and , Multiple planning viewpoints (i.e., threat-based or response/capability-based).

Due to these systematic similarities, our group envisioned responding to the research query through a three-part effort to develop the following informational models: Functional decomposition of counterterrorism operations to create a function tree; Task collection and analysis tied to the function tree; and, Organization collection and analysis tied to the task set. This paper, and the research conducted to date, is focused on the first activity—functional decomposition of counterterrorism activities.

Our group considers a conceptual model a functional decomposition of a domain where each level is created by asking "What are the constituent activities that define the next higher level?" For example, the current version of the CMCTO has at its highest point the process of *Countering Terrorists*. As will be illustrated in more detail below, when one asks "What are the activities which make up Countering Terrorists?" our research yielded three sub-functional responses: *Functions directed toward terrorists*, *Functions directed toward victims*, and *Functions of support*. At times during the modeling process there is a tendency to consider activities in light of the operators who perform them. This is actually not helpful in this type of exercise—for a conceptual model to be generically useful, it cannot be limited by *who* is responsible for a function, but merely must state the *type* of function. The aim of this form of conceptual modeling is to decompose functions to the extent that all processes within the domain are represented. This is a "top down" process. Thereafter, the "bottom up" process of delineating tasks associated with each function may be performed, as well as listing those organizations active in the domain and mapping the organizations to tasks performed.

Several tests were applied to determine whether the CMCTO adequately reflected the counterterrorism operations domain. The first is concerned with *completeness* of the functional description—ensuring that all functions, activities, and processes associated with the domain are represented. Two questions were asked: Is every function listed in the model contained in the function of its next higher level? Do functions exist that are part of a higher-level function and not contained in its decomposition? If the answer to the first question is "Yes" and the second "No," then one can assume completeness at each level.

The second test involved *balance*—the functions used to decompose or describe a higher-level function should be consistent in their apparent level. This can be very

much a subjective judgment. For example, the CMCTO presently decomposes *Functions directed toward terrorists* into *Prevent* and *Deter*. A review of counterterrorism literature and feedback provided by participants to several workshops on this subject revealed these sub-functions to be roughly balanced at this level. If, however, *Prevent* had been paired with *Use of Alarms*, a function three levels below *Deter*, it would have been recognized as out of balance. In a loop fashion, balance may be improved through constant comparative review of the decomposition for obvious clustering of functions and subsequent review of the clusters against the decomposition.

Finally, *acceptance* is the degree to which the model has undergone validation during its development. To justify the claim that the conceptual model incorporates a complete and balanced functional description of the domain, subject matter experts must accept the model. In the case of the CMCTO, this is an iterative process. The CMCTO will only be finalized after more experts in the field have had an opportunity to review the model and make comments. Acceptance is not a final element, but is one that continues to grow over time as the model is updated to reflect the thoughts and experiences of more collaborators in its construction.

3 Counterterrorism Literature

3.1 Definitions

As a first step in developing the CMCTO, the researchers surveyed scholarly literature to determine how counterterrorism and counterterrorism operations were defined, and ascertain whether models already existed to depict this domain. Whereas much literature may be found with respect to terrorists and terrorism, the same is not true of the subject of inquiry. To illustrate this difference, a search for the number of peer-reviewed journal articles with "terrorism" contained in either the title, abstract, or citation revealed over 2,500 items. The same type of search using the terms "antiterrorism," "anti-terrorism," "counterterrorism," "counter terrorism," and "counter-terrorism" yielded less than 250 articles.

The literature review revealed terrorism has been extensively studied since the late 1960s, however a definitional conundrum exists with respect to the term. In their 1988 survey of the field, Alex P. Schmid, Albert J. Jongman, and Michael Stohl identified 109 different definitions of the term (Brannan, Esler, and Strindberg, 2001, 11 quoting Schmid, Jongman, and Stohl, 1988). This phenomenon is not limited to academia. A sampling of the more than 40 government agencies, bureaus, and offices tasked with some counterterrorism functions yielded similar discord. For purposes of the model, our group settled on the following extensive definition:

> Terrorism is defined as the illegitimate use or threat of violence to further political objectives. It is illegitimate in that it targets civilians and/or non-combatants and it is perpetrated by clandestine agents of state and non-state actors in contravention of those laws of war and criminal statutes. It is symbolic and premeditated violence whose purpose is to communicate a message to a wider population

> than the immediate victims of violence. It is designed to affect this audience by creating psychological sates of fear in order to influence decision-makers to change policies, practices, or systems that are related to the perpetrators' political objectives. These objectives can be either systematic or sub-systematic and may be motivated by complex social forces including, but not limited to, ideology, ethno-nationalism, or religious extremism. (Cunningham 2002, 23).

Not surprisingly, there is no standard definition for counterterrorism operations found in academia or materials prepared by the entities performing this type of work. Presidential Decision Directive 39, *US Policy on Counterterrorism* considers a counterterrorism operation "an operation whose policy is to deter, defeat, and respond vigorously to all terrorist attacks on our territory and against our citizens, or facilities, whether they occur domestically, in international waters or airspace, or on foreign territory" (Clinton 1995, par. 1). The *US Government Interagency Domestic Terrorism Concept of Operations Plan* cites this form of operation as "any full range activity directed against terrorism, including preventive, deterrent, response, and crisis management efforts" (US Government 2001, B-1). In some cases, the definitions seem divided along defensive and offensive lines. For instance, the US Army uses "antiterrorism" operations to identify "those passive defensive measures taken to minimize vulnerability to terrorism" (National Interagency Civil-Military Institute n.d., 1-1). "Counterterrorism" operations, on the other hand, are the "full range of offensive measures to prevent, deter, and respond to terrorism" (Ibid.) In terms of the CMCTO model, counterterrorism operations are regarded not as favoring counterterrorism over anti-terrorism in definitional terms, but rather in the plain English meaning of *operations designed and conducted to counter terrorism*.

3.2 Models

Several models of counterterrorism operations are present in the literature. Peter Chalk notes characterizations of counterterrorism typically fall into one of two types. The *war model* "views terrorism as an act of revolutionary/guerilla warfare…where the onus of response is placed on the military and the use of such things as special forces, retaliatory strikes, campaigns of retribution, and troop deployment" (1996, 97). Counterterrorism operations are viewed in a range from non-force strategies—concessions, negotiations, diplomacy, and co-optation—to the undertaking of unilateral military action.

The second type, the *criminal justice model*, "views terrorism as a crime…where the onus of response is placed squarely within the bounds of the state's criminal legal system" (Chalk 1996, 97). Wardlaw (1989,66) cites D.B. Bobrow's four basic models of terrorist crisis resolution as possible alternatives available to governments to address this phenomenon.

Richard W. Leeman (1991, 10) asserts "by definition counterterrorism stands in opposition to terrorism." On a national level, when terrorists target a liberal democratic state they are attempting to force the abandonment of democratic principles. On an international level, terrorism may disrupt the consensual politics of

nations and peoples. Thus for Leeman, the "primary objective of international counterterrorism should therefore be to ensure international cooperation, especially for the purpose of preventing the escalation of conflict. A secondary objective of counterterrorism should be to prevent or lessen acts of terrorism" (1991, 11-12). This author acknowledges most counterterrorism studies focus on legal, diplomatic, or military responses to terrorism. Leeman believes one outstanding feature of the terrorism versus counterterrorism interplay is its dialogic nature.

Bruce Hoffman (2002, 313-314) likens terrorism to a shark in water, constantly moving forward to survive and succeed. He offers five recommendations on how the phenomenon should be viewed and addressed. Hoffman first acknowledges that terrorism has always been planned, purposeful, and premeditated. To respond, those conducting counterterrorism operations need to gain a better appreciation of what motivates terrorists and how these considerations impact their choice of targets and tactics. Secondly, terrorism is equated with psychological warfare, designed to elicit feelings of fear and intimidation in a target audience and undermine confidence in government and leadership. Hoffman recommends countermeasures that incorporate a full range of means—psychological, physical, diplomatic, military, economic, and persuasion. Third, the US and all democratic countries that value personal freedom and fundamental civil liberties will remain vulnerable to terrorism. Expectations about what can and cannot be achieved in countering terrorism must be made realistic. Fourth, many in the world have a negative view of the US and diplomatic efforts are needed to address this situation. Finally, terrorism owes its persistence to its ability to adapt to change and identify and exploit perceived vulnerabilities. Those countering terrorism must be as tireless, innovative, and dynamic.

Finally, Erik van de Linde, Kevin O'Brien, Gustav Lindstrom, Stephan de Spiegeleire, Mikko Vayrynen, and Han de Vries (2002) created an analytic framework for comparative analysis in the area of counterterrorism after ascertaining no others existed. The framework consists of four dimensions. The first three—Challenges, Measures, and Actors—are said to form a three-dimensional space representing the area of counterterrorism. The space is then positioned relative to the fourth dimension, the stage of an actual threat or attack—Pre-Attack Stage, Trans-Attack Stage, or Post-Attack Stage.

Turning to government documents, published shortly after 9/11 the US General Accounting Office (GAO) report *Combating Terrorism: Selected Challenges and Related Recommendations* summarized US policy as it had evolved the past 30 years:

The *National Strategy for Homeland Security* (2002) organizes strategy by critical mission areas: intelligence and warning, border and transportation security, domestic counterterrorism, protecting critical infrastructures and key assets, defending against catastrophic threats, and emergency preparedness and response. The recently released *National Strategy for Combating Terrorism* (2003) sets forth a four-part strategy for combating terrorism. Placing the goals and objectives into an outline form, it is compromised of the following elements: Defeat Terrorists and their Organizations; Deny Sponsorship, Support, and Sacctuary to Terrorists; Diminish the Underlying Conditions that Terrorists Seek to Exploit; and , Defend US Citizens and Interests at Home and Abroad.

As is demonstrated by this review, despite the more limited number of sources when compared with terrorism literature, there still exists a considerable body of thought concerning definitions and models for countering terrorism.

4 Developmental Process

The project's next phase involved extraction of functions from the literature and clustering concepts in a hierarchical (top-down) fashion. In this manner, the group believed it could obtain better coverage—satisfy the completeness criteria—than the actor- or response-based models dominating the literature. Functions present in both approaches would be collected, such as "retaliate" or "gather intelligence." The clustering of like items would eliminate repetition and aid in the process of balancing functions hierarchically. Looking at Wardlaw (1989), "introduce special anti-terrorist legislation" would therefore be placed at a higher level than "place legal limits on media" as the latter may be considered a more specific part of the first.

Two straw models were developed. One was guided more by its orientation to time—Was the action performed *Prior to an event* or *After an event*? The second was more concerned with the type of action taking place—*Information gathering*, *Actions directed toward perpetrators of violence*, and *Actions directed toward victims of violence*. Both were partial and included only what were believed to be top-level processes.

Several workshops were held for revision and validation of the decomposition. Participants included subject matter experts and practitioners. Somewhat mirroring the literature experience, the researchers discovered time and again there was a tendency in the proceedings to revert to a discussion of terrorists and terrorism. Those individuals who participated in the first workshop did not embrace either of the straw models and the dialogue refocused on the highest-level sub-functions to countering terrorism. The group determined certain functions had to be represented in the model—prevent, deter, defend, crisis management, consequence management, and respond—but there was disagreement about where they should be situated in the decomposition. The CMCTO was subsequently configured on a *Prevent-Respond* alignment, with the first related to functions conducted prior to an event and the second following an event. Merriam-Webster (2002) definitions were equated with each, thus *Prevent* meant "to keep from happening" and *Respond* meant "to act in response." A third function, *Support*, was added to capture those items that cut across both the *Prevent* and *Respond* functions.

During another workshop, it was determined that *Deter*, which had previously been a sub-function of *Prevent*, was improperly situated in the decomposition. While both assume an event may be stopped, the functions are distinct—*Prevention* occurs by directly targeting the threat and taking away its capability, while *Deterrence* occurs when the threat believes its capability is insufficient to be successful. *Mitigation*, here considered the lessening of the effects of an event, was also revealed as a missing function. Furthermore at this time, the researchers became aware that they were conflating timing and assumptions with the actors to which the functions were directed—whether terrorist or victim. This was resolved when the functions were organized along the actor. Thus, the highest-order functions for *Countering*

Terrorism became *Functions directed toward terrorists* and *Functions directed toward victims*, plus the previously designated *Functions of support*, with sub-functions as follows:

5 Narrative View of the CMCTO

5.1 Functions Directed Toward Terrorists

This function hosts those actions taken directly, or indirectly, against a terrorist threat to halt, or reduce, the effects of an attack. The two major sub-functions are ***Prevent*** (1.1) and ***Deter*** (1.2). Operations under the ***Prevent*** (1.1) sub-function are taken against terrorists to neutralize or thwart a threat by limiting the terrorist's capacity to act. This may be done by *Reducing the threat's capabilities to act* (1.1.1) or, by *Reducing the threat's physical and economic assets* (1.1.2).

Operations to ***Deter*** (1.2) seek to increase protection and presence to make terrorists perceive that their capabilities are insufficient to achieve their objectives. This may be achieved by *Developing physical security measures* (1.2.1) or *Increasing the presence of enforcement measures* (1.2.2).

5.2 Functions Directed Toward Victims

This function assumes that one cannot prevent all terrorist attacks from taking place and there are actions that may be taken to assist the victims of an attack, both prior to and following an incident. ***Mitigate*** (2.1) and ***Respond/consequence management*** (2.2) are the two major sub-functions. Operations under the ***Mitigate*** (2.1) sub-function seek to lessen the severity of an attack by implementing measures to assist victims to be better prepared for an attack, or be able to respond more effectively and feel less terrorized at the time, or following, an attack. This is done by *Preparing programs to allow rapid response to any action* (2.1.1) and *Providing public information on responding to an event* (2.1.2).

With respect to operations in the ***Respond/consequence management*** (2.2) sub-function, actions are taken to address the impact of a terrorist attack. Five functional categories have been distinguished. The first is to *Determine the situation* (2.2.1). The second is to *Plan for response* (2.2.2). The third is to *Decide on options for response* (2.2.3) and the fourth is to *Acquire resources* (2.2.4). The final category is to *Deploy resources* (2.2.5).

5.3 Support Functions

In developing the CMCTO, our group noted that certain functions were likely to be performed whether actions addressed terrorists or victims. Three major sub-functions have been identified: ***Supervision and synchronization*** (3.1), ***Information operations*** (3.2), and ***Logistics*** (3.3). ***Supervision and synchronization*** (3.1) is hands-on coordination, direction, and inspection conducted to direct the accomplishment of the various aspects of an assigned task or mission, particularly those involving numerous groups working together. *Consensus building* (3.1.1), *Coordinate and cooperate with*

others (3.1.2.), *Liaison with other actors* (3.1.3), and *Command center* (3.1.4) are its next lower functional decompositions.

Information operations (3.2) communicate knowledge, intelligence, news, or advice. Three functional categories have been distinguished. The first involves *Incoming information/intelligence* (3.2.1). The second is *Information management* (3.2.2) and the third *Information support* (3.2.3). Finally, ***Logistics*** (3.3) involve planning and/or carrying out the movement and maintenance of personnel and equipment, divided in the model by "Personnel" (3.3.1), "Supplies/services" (3.3.2), and/or "Transportation" (3.3.3).

6 Conclusion

Development of the CMCTO has resulted in a new level of understanding of the domain in question. Identification of *Functions directed toward terrorists*, *Functions directed toward victims*, and *Functions of support* provides a structure for implementation of the CMCTO as a "Systems Dynamics Model of Counterterrorist Operations." The latter furnishes a facility for studying the dynamical consequences of particular terrorist and counterterrorist adoption of actions and policies, and insight into the development of a systemic approach to identifying and responding to threats. Validation and verification of model processes, and parameter and coefficient values by appropriate subject matter experts are clearly necessary before the model could be used to study actual terrorist and counterterrorist actions. It is hoped that extensions and enhancement of both the CMCTO and the "Systems Dynamics Model of Counterterrorist Operations" will be undertaken in the near future.

References

1. Brannan, David W., Philip F. Esler, and N.T. Anders Strindberg. 2001. Talking to "terrorists": Towards an independent analytical framework for the study of violent substate activism. Studies in Conflict & Terrorism 24 : 3-24.
2. Clawson, Patrick. 1990. US options for combating terrorism. In The politics of counter-terrorism: The ordeal of democratic states, ed. Barry Rubin, 3-29. Washington, D.C.: The Johns Hopkins Foreign Policy Institute.
3. Clinton, William J. 1995. Presidential decision directive 39, US policy on counterterrorism. Universal resource link located at: www.fas.org/irp/offdocs/pdd39.htm; accessed 2 February 2003.
4. Crenshaw, Martha. 2001. Counterterrorism policy and the political process. Studies in Conflict & Terrorism 24 : 329-337.
5. De B. Taillon, J. Paul. 2001. The evolution of special forces in counter-terrorism: The British and American experiences. Westport and London: Praeger.
6. Falkenrath, Richard. 2001. Analytic models and policy prescription: Understanding recent innovation in U.S. counterterrorism. Studies in Conflict & Terrorism 24 : 159-181.
7. Hills, Alice. 2002. Responding to catastrophic terrorism. Studies in Conflict & Terrorism 25 : 245-261.
8. Hoffman, Bruce. 2002. Rethinking terrorism and counterterrorism since 9/11. Studies in Conflict & Terrorism 25 : 303-316.

9. Leeman, Richard W. 1991. The rhetoric of terrorism and counterterrorism. New York: Greenwood Press.
10. Merriam-Webster. 2002. The Merriam-Webster dictionary. Springfield: Merriam-Webster.National Interagency Civil-Military Institute, n.d. Preparing for and managing the consequences of terrorism: Resource guide. Universal resource link located at: www.nici.org/publications/publications/01%20%20PMC%20Resource%20Guide.pdf; accessed 13 February 2003.
11. Office of Homeland Security, 2002. National strategy for homeland security. Universal resource link located at: www.whitehouse.gov/homeland/book/nat_strat_hls.pdf; accessed 2 February 2003.
12. US General Accounting Office. 2001. Combating terrorism: Selected challenges and related recommendations. Universal resource link located at: www.ciaonet.org/cbr/cbr00/video/cbr_ctd/cbr_ctd_19a.pdf; accessed 12 February 2003.
13. US Government. 2003. National strategy for combating terrorism. Universal resource link located at: www.whitehouse.gov/news/releases/2003/02/counter_terrorism/counter_terrorism_ strategy.pdf; accessed 12 April 2003.
14. 14.2001. US Government Interagency domestic terrorism concept of operations plan. Universal resource link located at: www.fas.org/irp/threat/conplan.pdf; accessed 2 February 2003.
15. Van de Linde, Erik, Kevin O'Brien, Gustav Lindstrom, Stephan de Spiegeleire, Mikko Vayrynen, and Han de Vries. 2002. Quick scan of post 9/11 national counter-terrorism policymaking and implementation in selected European countries: Research project for the Netherlands Ministry of Justice. Santa Monica: RAND.
16. Veness, David. 2001. Terrorism and counterterrorism: An international perspective. Studies in Conflict & Terrorism 24 : 407-416.
17. Wardlaw, Grant. 1989. Political terrorism: Theory, tactics, and counter-measures, 2d ed. Cambridge: Cambridge University Press.

Appendix: The Conceptual Model of Counter Terrorist Operations (CMCTO)

1	Functions directed toward terrorists	1.1.2.2.	Halt fiscal activity	1.2.2.1.1.	Control borders
1.1.	Prevent	1.1.2.2.1.	Expose assets	1.2.2.1.2.	Control transportation hubs
1.1.1.	Reduce the threat's capabilities to act	1.1.2.2.2.	Deny access to assets	1.2.2.2.	Deter through presence
1.1.1.1.	Identify organizations and members	1.1.2.2.3.	Impair access to assets	1.2.2.2.1.	Deter by physical presence
1.1.1.1.1.	Identify location of organization	1.2.	Deter	1.2.2.2.2.	Deter by expected or virtual presence
1.1.1.1.2.	Identify location of members	1.2.1.	Develop physical security measures	2	Functions directed toward victims
1.1.1.1.3.	Identify roles in organization	1.2.1.1.	Protect perimeters	2.1.	Mitigate
1.1.1.2.	Directly neutralize or remove membrs	1.2.1.1.1.	Use detectors	2.1.1.	Prepare programs to allow rapid response to any action
1.1.1.2.1.	Pursue and capture members	1.2.1.1.2.	Use alarms	2.1.1.1.	Prepare public sector plans
1.1.1.2.2.	Punish members	1.2.1.1.3.	Use barriers	2.1.1.2.	Prepare private sector plans
1.1.1.3.	Indirectly neutralize or remove members	1.2.1.2.	Reinforce construction	2.1.1.3.	Prepare health sector plans
1.1.1.4.	Inhibit new membership	1.2.1.2.1.	Reinforce buildings	2.1.2.	Provide public information on responding to an event
1.1.1.4.1.	Destroy or limit driving spirit of orgs	1.2.1.2.2.	Reinforce transportation	2.1.2.1.	Provide information through educational institutions
1.1.1.4.2.	Reduce informational assets	1.2.1.3.	Increase surveillance	2.1.2.2.	Provide information to training programs
1.1.2.	Reduce the threat's physical and economic assets	1.2.1.3.1.	Increase visible surveillance	2.1.2.2.1.	Provide training for government
1.1.2.1.	Deny use of infrastructure	1.2.1.3.2.	Increase expected surveillance	2.1.2.2.2.	Provide training for individuals
1.1.2.1.1.	Deny use of facilities	1.2.1.4.	Implement checkpoints	2.2.	Respond/consequenc e management
1.1.2.1.2.	Deny transportation/mobility	1.2.2.	Increase presence of enforcement measures	2.2.1.	Determine situation
1.1.2.1.3.	Deny use of communications	1.2.2.1.	Control entry/exit	2.2.1.1.	Collect information
2.2.	Respond/consequenc e management	2.2.5.3.1.	Undertake long-term medical assistance	3.2.1.2.	Clandestine sources
2.2.1.	Determine situation	2.2.5.3.2.	Undertake economic asst	3.2.1.2.1.	Technical
2.2.1.1.	Collect information	2.2.5.3.3.	Undertake reconstruction	3.2.1.2.2.	Human intelligence
2.2.1.2.	Analyze information	3	Functions of support	3.2.1.3.	Other sources

2.2.1.2.1.	Review past performance	3.1.	Supervision and synchronization	3.2.2.	Information management
2.2.1.2.2.	Determine constraints	3.1.1.	Consensus building	3.2.2.1.	Communication model
2.2.1.2.3.	Place collected information in template/framework	3.1.2.	Coordinate and cooperate with others	3.2.2.1.1.	Source
2.2.1.3.	Report information	3.1.3.	Liaison with other actors	3.2.2.1.2.	Message
2.2.2.	Plan for response	3.1.3.1.	Local level	3.2.2.1.3.	Channel
2.2.2.1.	Determine immediate needs	3.1.3.2.	County level	3.2.2.1.4.	Receiver
2.2.2.2.	Determine sources of response	3.1.3.3.	State level	3.2.2.1.5.	Feedback
2.2.2.3.	Identify options for response	3.1.3.4.	Federal level	3.2.2.1.6.	Noise
2.2.3.	Decide on option for response	3.1.3.5.	International level	3.2.2.1.6.1.	Encoding/decoding
2.2.4.	Acquire resources	3.1.4.	Command center	3.2.3.	Information support
2.2.5.	Deploy resources	3.1.4.1.	Communication channels	3.2.3.1.	Information/intelligence sharing
2.2.5.1.	Deploy emergency response	3.1.4.2.	Procedures	3.2.3.2.	Language support
2.2.5.1.1.	Deploy control (1st response)	3.1.4.3.	Common operations picture	3.3.	Logistics
2.2.5.1.2.	Deploy medical resources	3.2.	Information operations	3.3.1.	Personnel
2.2.5.1.3.	Deploy evacuation resources	3.2.1.	Incoming Info/Intel	3.3.2.	Supplies/services
2.2.5.1.4.	Deploy search and rescue (urban)	3.2.1.1.	Open sources	3.3.3.	Transportation
2.2.5.1.5.	Deploy hostage rescue	3.2.1.1.1.	Mass media		
2.2.5.2.	Deploy immediate resp (2nd Resp [])	3.2.1.1.2.	Legally available documents		
2.2.5.3.	Undertake recovery actions	3.2.1.1.3.	Observation of political, military and economic activity		

Measuring Success in Countering Terrorism: Problems and Pitfalls

Peter S. Probst

Institute for the Study of Terrorism and Political Violence
Conterpro@aol.com

1 Introduction

One of the major problems in Intelligence analysis and counter-terrorism research is the use or, more precisely, misuse of metrics as a means to measure success. Such quantification may be admirable and necessary when dealing with rocket motors or physical phenomena but can be self-defeating and unrealistic when dealing with people and human events which, after all, are the ultimate underpinnings of terrorism, insurgency and political instability. Human behavior is notoriously hard to predict and outcomes without historical perspective difficult to assess. Measures of success that are touted as useful and accurate so often in the real world prove to be little more than intellectual snake oil. Hard quantifiable data that is meaningful is hard to come by, and so we often willingly settle for data that are easily accessible and quantifiable, hoping that our extrapolations are sufficiently accurate to guide or assess a course of action or the conduct of a conflict.

2 The Law of Unintended Consequences

Throughout my more than 30 years as an intelligence officer, I have seen many attempts to measure success in countering one's adversary, whether in terms of programs or personnel, or the development of systems that purport to objectively rank-order success against one's adversaries. I have seen some very sophisticated policy makers gulled into making some very foolish mistakes when trying to measure such elusive concepts as success, particularly when one fails to define the term or consider the secondary or tertiary consequences of operations against one's adversary.

By way of example there is a story, perhaps apocryphal, about efforts, years ago, by the U.S. Government to thwart Cuban influence in a particular Latin American country. To counter the Cubans and as part of a non-Peace Corps nation building effort, the United States covertly helped promote a cutting edge literacy campaign. Every month the progress reports reflected impressive gains with literacy rates among the target populations rising remarkably. In other words—short term success.

The Cubans, however, bided their time and then, after the Americans had taught the people to read, they flooded the area with Marxist propaganda that the people for the first time could now read and digest. Meanwhile in Washington, the budgetary winds shifted and for what ever reason the U.S. Government, regrettably, failed to engage in an equally ambitious counter-campaign. The Marxists ended up having the field pretty

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 316 – 321, 2005.

much to themselves. In other words short term success paved the way for long term failure and a significant setback in the war of hearts and minds. The point is defining success can sometimes be tricky depending on one's ultimate objectives and timeframe.

3 Measures of Success

Countless attempts to implement quantitative systems to evaluate success too often have backfired with unintended and serious consequences. In the early 1960s a process called Management by Objective was regarded by government as a cutting edge management tool. Its use by the Intelligence Community proved unfortunate and, too often, counterproductive. The principle was to define a series of professional objectives for intelligence officers to determine how well they measured up. The aim, of course, was to institute accountability and provide an objective tool to assess the relative success of field personnel and their operations.

Tremendous weight was given to the number of intelligence reports submitted by the Case Officer in the field. This was regarded as an objective gauge of effectiveness and worth. Case Officers also might be tasked to make a specific number of new agent recruitments each fiscal year. Of course the process was somewhat more complex, but this was the general idea. Performance depended on productivity and productivity was defined by numbers. Those serving overseas soon discovered that they had entered the numbers game big time.

Members of CIA's Clandestine Service have always had a reputation for being savvy with a healthy regard for self-preservation, and this extended to the bureaucratic arena as well. Many realized this was a whole new ballgame with a new set of rules. The result was adaptation. They began to take a solid, detailed report that the agent and case officer at no small risk had spent considerable time developing and, realizing that numbers were critical, would divide that report into two or three highly rated shorter reports; thereby, increasing their production numbers for that month. Nevertheless, such adjustments were rarely sufficient to overcome the weight of a large quantity of useful but not particularly valuable reporting that began to flood Headquarters as a result of the pressure to best the previous month's total or, at the very least, to maintain the numbers. As in academia, a "publish or perish" mentality became increasingly pervasive.

Of course, ways of weighing the value of the reports were ultimately introduced as a way to level the playing field but numbers too often trumped quality, and a system that had been introduced to measure success ended up measuring the wrong criterion, and introducing pressures that tended to compromise the integrity of the intelligence process and those participating in it.

A similar situation developed with regard to agent recruitment, an officer being tasked to recruit a minimum of new agents each fiscal year. The reality was that a Case Officer could literally work years to recruit a high-level communist party penetration; whereas, a colleague might spend only a couple of months to recruit two or three agents that were useful but of considerably lesser value. It was like comparing apples and oranges with the numerical comparisons creating a false sense of equivalence, objectivity and fairness. As a consequence there was a tremendous

temptation to make easy recruitments that might not have been of great value but had the virtue of keeping the Station's numbers up and Headquarters at bay.

4 Statistics: A Poor Reflection of Reality

The way we use statistics has a way of distorting reality and, frankly when it comes to Intelligence, one needs to regard statistics and those who tout them with no small measure of suspicion. There is a recent example of how a simple statistical error caused the U.S. government great embarrassment. It involved the publication of State Department's annual report, *Patterns of Global Terrorism, 2003*, which is an annual report mandated by Congress and used as something of a barometer to assess the success of efforts to counter international terrorism. At televised press conferences high-level State Department officials announced that terrorist attacks had dropped precipitously, and asserted that the decline in numbers proved that our terrorist adversaries were on the run and America was winning the War on Terrorism.

Much to State Department's chagrin it was subsequently disclosed that a major statistical undercount had seriously skewed the figures and, on redoing the numbers, it turned out terrorist attacks had almost doubled. If we were to have used State Department's logic that the number of attacks directly correlates with the success or failure of the War on Terrorism, the logical inference could have been drawn that we were losing this war and the terrorist had us in disarray. Everyone was terribly embarrassed.

In my view, the embarrassment was not so much that a computational error had been made but, rather, the manner in which the numbers had been interpreted.

The State Department had claimed that fewer terrorist attacks proved that our countermeasures were working. When you stop to think about it and, particularly, when you examine the history of terrorism you quickly realize that no such conclusion was warranted. There can be many reasons why terrorist attacks abate. The terrorists may be gathering for a massive onslaught and seeking to husband their resources. Or they may be seeking to lull a government into a false sense of security and, thereby, foster complacency and diminished vigilance that significantly increases government vulnerability. Or the terrorists may believe they are dealing from a position of strength and are ready to explore the political route to power. The terrorists, therefore, curb the violence to burnish their political image, while keeping their paramilitary capabilities under wraps. They may use the stand down for further training, recruitment and arms procurement in the event that success via the political route eludes them.

Conversely, a terrorist group that is on the ropes may launch a campaign of bold and bloody operations to prove to the government and other audiences that it is still a potent force. The numbers of attacks rise, but they are born of desperation and may, in reality, signal the death throes of the organization. In other circumstances, a spate of operations may be mounted simply to maintain the internal morale and cohesion of the terrorist group. Few things are more demoralizing to a terrorist group than to stand down and do nothing, while they feel the noose tightening. It foments restlessness and internal dissension with questions being raised as to the purpose and direction of the movement and the competence of the leadership.

The principle is very simple. Without reliable collateral reporting that provides a context for the numbers, the true meaning of an increase or decrease in terrorist attacks is often impossible to deduce.

5 Quality Versus Quantity

There are other weaknesses in how terrorist statistics are used. In the annual State Department report, the relative impact of the enumerated operations is not weighted. For example, an operation of the magnitude of the 9/11 attack on the World Trade Center would be tallied as a single event as would the bombing of a newsstand in Moscow. Obviously this makes little sense because there is no equivalence between the two attacks which is what the statistics would imply. It is true that the impact of a particular attack may be addressed in the narrative portion of the report, but this is rarely what the media picks up on. They report that terrorist attacks are "up 47 percent" and the public quails and Congress holds hearings. In other words, if the statistics seem at odds with the written analysis something is wrong somewhere and must be clarified. In this instance it would be a failure of methodology and a methodology we continue to employ.

What should be obvious by now is that more important than the number of attacks carried out against U.S. interests is the impact of those attacks, yet such impact is not reflected in the government's statistical tally. Some form of Bayesian analysis could help lend clarity and perspective, and make the statistics truly useful for the policy maker and the operator. The bottom line is that the wrong metrics are used to measure increase or decrease, success or failure. Impact of attacks rather than the number of attacks is the critical element and is what needs to be measured.

If we feel compelled to produce numbers, the enumerated attacks could be assigned values in some sort of a Terrorism Richter Scale with a Category One attack having minimal impact and a Category Ten attack being absolutely catastrophic.

The total number of attacks in each category could then be averaged to provide a numerical value for the year that would reflect the impact of terrorist operations over that 12 month period. There are, of course, issues inherent in such a system, such as how to measure and define impact, but at least such an approach would have us going down a more productive analytic path.

Another problem area is the failure to factor into our equation the number of failed or aborted attacks. I have never seen this addressed as part of the statistical analysis. Inclusion of such numbers, coupled with a measuring of attack impact would further refine our statistical analysis and help provide a more accurate gauge of the effectiveness of our counterterrorism efforts. For one thing, inclusion of the number of failed and aborted operations could provide a measure of how successful governments are in their efforts to disrupt terrorist operations. Such data, of course, would likely be fragmentary, but it could help broaden our understanding of the state-of-play and help us more accurately gauge our overall effectiveness.

6 The Stiletto over the Broadsword

Fighting an entrenched insurgency is much different than fighting a conventional war. There are problems inherent in relying principally on the traditional military in a non-

traditional conflict. Killing a large number of enemy is what a conventional military force seeks to do as a matter of course. Killing large numbers of terrorists or insurgents makes good news copy and carries the strong inference of success, but can be misleading. Killing a large number of people even if they are insurgents can be counterproductive. Killing an insurgent who poses a direct and immediate threat is, of course, necessary but far better than killing insurgents is to undermine their resolve and the ideology that motivated them to fight in the first place.

Demoralization is contagious and can destroy organizational cohesion just as surely as plague. Dead insurgents, however, do not exist in a social vacuum. They have parents, brothers, sisters, cousins, wives and children. Often the extended family, clan and circle of friends are impacted, and the local code of honor may demand that the death be avenged. The insurgent's death can act as a catalyst for commitment, propelling close friends and members of his familial circle to join or, at least, actively support the insurgency. What we should value; therefore, are operations that delegitimize the insurgent cause in the eyes of the insurgents, those who are close to them and the insurgents' various constituencies. Secondly, we should seek, when feasible, to minimize insurgent deaths with our emphasis being on quality, high value kills that target leaders, planners, financiers and political operatives.

Conventional military operations in an insurgency too often prove counterproductive. A high enemy body count may translate into high enemy recruitment. Rather than breaking his will and eroding his capabilities such casualties, particularly if indiscriminate, may serve only to strengthen resolve and swell their ranks.

Of course there always is the broadsword, whereby a government meets brutality with even greater brutality. And this can work if a country and its citizens have the stomach for it. However, most democracies don't. There is a tipping point when dealing with terrorism or an insurgency such as in Iraq. Western democracies with Western values are unlikely to go that route, because the public simply will not accept it. It violates our traditions and our Judeo-Christian values. Despots such as Iraq's Saddam Hussein and Syria's Hafez Assad do not share our queasiness. Saddam gassed and bombed the Kurds into virtual submission, and Assad leveled the Islamist stronghold of Hamma in a military assault of biblical brutality. Because of our understandable aversion to such near genocidal policies, we need to look elsewhere.

Targeted killings are one answer. Removal of key cadre is the type of operation that is extremely productive, depriving the insurgent force of direction and leadership. However to the uninitiated who link success to numbers, the statistical count in such operations is generally unimpressive—a mere handful at best. And yet, this handful can be critical to breaking the back of an insurgency.

The Military understands this, and aggressively pursues insurgent cadre. However, those who are enamored with numbers look at the four or five dead and shrug their shoulders, being much more impressed by conventional military assaults that result in scores of insurgent dead or wounded. And the bottom line remains that impressing the political movers and shakers as well as the general public directly translates into political support and burgeoning budget allocations. To correct such misperceptions is one reason why education and public information efforts are critical.

The other part of the equation is to break the morale of the insurgents. The cleanest and most cost effective way to do this is through a well thought out and professionally executed campaign of psychological warfare and covert action of which targeted

killings may comprise an integral part. But results of such operations are also difficult to quantify, and usually lack the drama of large enemy casualty counts.

It is important to realize that the mindset and value systems of the Jihadists are considerably different from ours. In general, they are not encumbered by our Western mindset. They may have their own issues, but a preoccupation with statistical counts is not one of them. Their operations are not tied to the fiscal year or an annual budget cycle. They are not saddled with annual project renewals. They take the long view.

The American mindset, in contrast, is tied to the fiscal year mentality and other artificial short-term and, too often, self-defeating constraints and pressures that include the political timetable of Presidential and Congressional elections. As one terrorism expert stated, "This is not a war that can be won by an impatient people." Yet we are impatient, and we demand immediate results. And this is one of our greatest strengths but also one of our greatest failings. Although such traits may be admirable in other circumstances, they are counterproductive when dealing with a Protracted Conflict, such as the one in which we are currently engaged against a determined terrorist foe such as al Qaeda and other global Jihadists.

7 Conclusion

Over-reliance on and misuse of statistical measurement not only has served to distort the intelligence product, but too often has corrupted the intelligence process as intelligence officers find themselves chasing the numbers with less time available to chase the hard but elusive information needed to advance the country's security interests. The mindset that produced the body count syndrome of the Vietnam War unfortunately is alive and well. It is part and parcel of our cultural baggage. As a consequence, it has caused us to fail to identify or misread critical trends and engage in practices that are transparently counterproductive. In seeking to measure success we look to measure things that are easy to quantify, but too often are off the mark, providing us only the illusion of accuracy and precision rather than a valid and accurate measure of meaningful progress.

Statistical analysis as used by the government to assess terrorism and counter-terrorism efforts remains primitive and, too often, dangerously misleading. We measure what can easily be quantified rather than what is truly meaningful. We strive to capture extremely complex phenomena in a simple sound bite, reinforced by seemingly compelling but simplistic statistical comparisons and then wonder why our instant analysis has failed to comport with reality, leaving us embarrassed and scratching our heads. Numbers, as we use them, provide a false sense of objectivity, accuracy and precision, too often leaving the decision makers frustrated and angry. And, too often, leaving the public with the feeling that somehow they have been conned.

In order to use our limited resources to best effect, we need to introduce concepts and analytic strategies that more accurately reflect the reality on the ground, enable us to better predict trends and more accurately assess the effectiveness of our efforts, our programs and our people.

Mapping the Contemporary Terrorism Research Domain: Researchers, Publications, and Institutions Analysis

Edna Reid and Hsinchun Chen

Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721, USA
{ednareid, hchen}@eller.arizona.edu

Abstract. The ability to map the contemporary terrorism research domain involves mining, analyzing, charting, and visualizing a research area according to experts, institutions, topics, publications, and social networks. As the increasing flood of new, diverse, and disorganized digital terrorism studies continues, the application of domain visualization techniques are increasingly critical for understanding the growth of scientific research, tracking the dynamics of the field, discovering potential new areas of research, and creating a big picture of the field's intellectual structure as well as challenges. In this paper, we present an overview of contemporary terrorism research by applying domain visualization techniques to the literature and author citation data from the years 1965 to 2003. The data were gathered from ten databases such as the ISI Web of Science then analyzed using an integrated knowledge mapping framework that includes selected techniques such as self-organizing map (SOM), content map analysis, and co-citation analysis. The analysis revealed (1) 42 key terrorism researchers and their institutional affiliations; (2) their influential publications; (3) a shift from focusing on terrorism as a low-intensity conflict to an emphasis on it as a strategic threat to world powers with increased focus on Osama Bin Laden; and (4) clusters of terrorism researchers who work in similar research areas as identified by co-citation and block-modeling maps.

1 Introduction

Contemporary terrorism is a form of political violence that evolved in the 1960s and characterized by an increase in terrorist attacks across international boundaries [33]. The recent escalation of contemporary terrorism has attracted many new and non-traditional research communities such as information science and human factors, whose scholars have a desire to do research in this area. This raises questions for new terrorism researchers as they try to adapt to the challenges in this domain "Who are the leading researchers in terrorism?" "What are their relevant publications?" "What are the dominant topics because I want to know if my ideas have already been explored?" "What types of data are used?" "Who should I work with?"

The task of responding to these questions is difficult because of the explosive growth in the volume of terrorism publications, the interdisciplinary and international

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 322 – 339, 2005.

nature of the field, and the lack of a professional association to nurture the terrorism research area and provide a platform for organizing and providing systematic access to terrorism studies [14;25]. For example, terrorism information is spread across many electronic databases, government and research center's websites, and a large number of journals that deal with various specialized aspects of the phenomenon [15].

With the interest in terrorism increasing, the findings of this study will be immensely useful in understanding the contributions of key terrorism authors in guiding terrorism-related research. This paper presents a brief review of analytical techniques and framework for knowledge mapping. Subsequent sections will describe the research design and results of our contemporary terrorism literature mapping with three types of analysis: basic analysis, content map analysis, and co-citation network analysis. The final section will provide conclusion.

2 Related Work

There is extensive literature on knowledge mapping of scholarly literature and patents to analyze the structure, the dynamics, social networks, and development of a field such as medical informatics and information science [5;13;16;31]. Mapping refers to an evolving interdisciplinary area of science aimed at the process of charting, mining, analyzing, sorting, enabling navigation of, and displaying knowledge [30]. Although it is useful to the subject expert for validation of perceptions and means to investigate trends, it provides an entry point into the domain and answers to domain-specific research questions for the non-expert [4].

Citation Data

Maps and snapshots of a field's intellectual space have been generated as a result of the pioneering work of Garfield and Small who stimulated widespread interest in using aggregated citation data to chart the evolution of scientific specialties [9]. By aggregating citation data, it is possible to identify the relative impact of individual authors, publications, institutions, and highlight emerging specialties, new technologies and the structure of a field [12].

The advent of citation databases such as the Institute for Scientific Information (ISI) Social Sciences Citation Index (SSCI) and Science Citation Index (SCI), which track how frequently papers are cited in a publication, and by whom, have created tools for indicating the impact of research papers, institutions, and authors [12]. The web-version of SSCI, SCI, and the Arts and Humanities Citation Index is the Web of Science (WoS). Web-based tools such as Google and ResearchIndex (formerly CiteSeer) have been created to harness the similarities between citation linking and hyperlinking [9;28]. Searching the digital citation indexes have resulted in enormous amounts of citation data that are difficult to analyze, extract meaningful results, and display using traditional techniques.

This was illustrated in earlier citation network studies of terrorism researchers in which the researcher used authors, institutions, and documents as units of analysis and the ISI databases to identify the invisible colleges (informal social networks) of terrorism researchers, key research institutions, and their knowledge discovery

patterns. This manual process was labor-intensive and relied on citation data [25;26;27]. While there are limitations in using the ISI citation data such as they are 'lagging indicators' of research that has already been completed and passed through the peer review cycle [12], they are widely used in visualization studies and are the basis for identifying key terrorism researchers, influential publications, and subgroups of terrorism researchers in this study.

Visualization Techniques

Recent developments in the field of domain visualization attempt to alleviate this "citation information overload problem" by applying information visualization techniques to interact with large-scale citation data [10]. Several techniques have been applied to citation visualization such as Pathfinder network scaling [5], social network analysis, and author co-citation analysis [5;31] which is particularly suited to investigation of intellectual structure because they provide the capability to interact with data and display it from different perspectives. Author co-citation map identifies interrelation among authors by analyzing the counts of the number of articles that cite pairs of authors jointly [32].

Content, or 'semantic', analysis is an important branch of domain analysis which relies on natural language processing techniques to analyze large corpora of literature [10]. The content map analysis technique produces content maps of large-scale text collections. The technique uses simple lexical statistics, key phrase co-occurrence analysis, and semantic and linguistic relation parsing. For example, Huang, et.al. [17] uses self-organizing map (SOM) algorithm to generate content maps for visualizing the major technical concepts appearing in the nanotechnology patents and their evolution over time.

Another visualization technique is block-modeling which seeks to cluster units that have substantially similar patterns of relationships with others [11]. It has been applied in criminal network analysis to identify interaction patterns between subgroups of gang members [6]. The application of visualization techniques to citation, content analysis, and author co-citation data provides a foundation for knowledge mapping. The techniques support the users' visual exploration of a domain to identify emerging topics, key researchers, communities, and other implicit knowledge that is presently known only to domain experts [30]. For example, the Namebase [24], mines names and organizations from terrorism books and periodicals included in its database and links names in a social network. Figure 1 provides an example of a terrorism social network for Brian M. Jenkins (name listed in the center in red), founder of terrorism research at Rand Corporation. It is based on the number of times a name is listed on the same page with Jenkins.

Although the Namebase visualization does not indicate whether there is a relationship between Jenkins and the other names listed on the page or the context of their relationships, it is the only web-based tool readily available for visualizing social networks of terrorism researchers. Additionally, no systematic study has been conducted that uses citation network, content map analysis, and author co-citation analysis for automatically mapping the terrorism research domain.

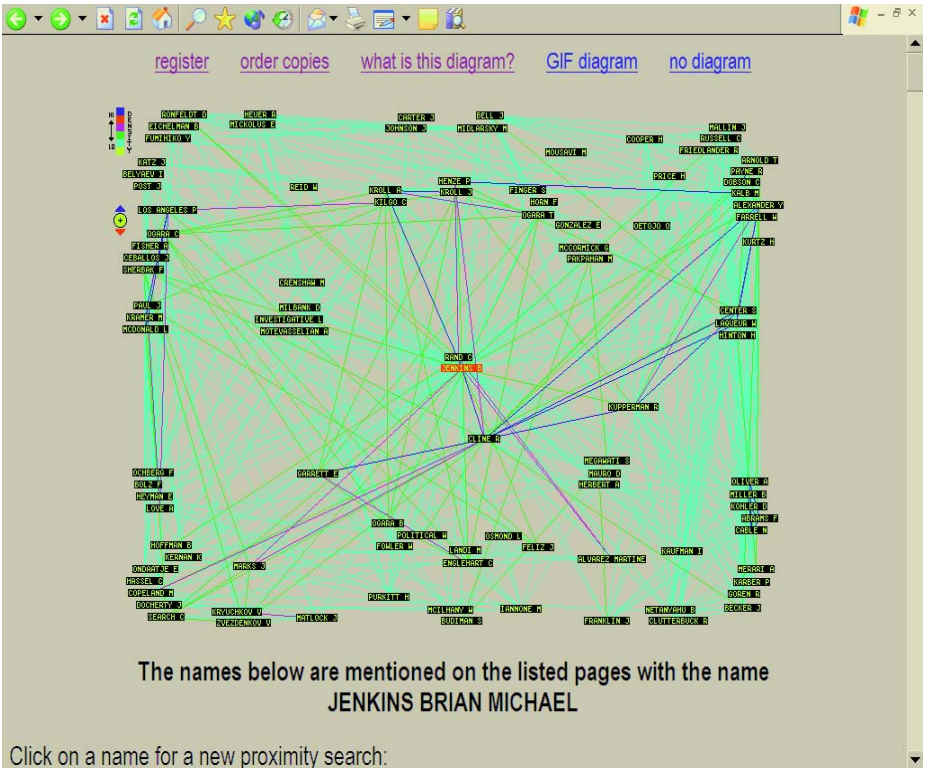


Fig. 1. Brian Jenkins's Social Network (http://www.namebase.org)

3 Research Design

This study purports to provide empirically based answers to the research questions (RQs) listed in Table 1. It adopts the integrated knowledge mapping framework proposed by Huang, et. al. [17] for patent analysis and used in Eggers study of medical informatics [10]. The framework includes three types of analysis: basic analysis, content map analysis, and citation network analysis to provide a multifaceted analysis of a research domain.

For the basic analysis, we analyze scientific output measures such as productivity (number of publications produced by a terrorism researcher) and impact (citation counts which allows one to find out how often a publication is cited). By analyzing documents and citation information, we identify key researchers, their influential terrorism publications, and research communities. The content map analysis visualizes the major subtopics and emerging concepts appearing in the publications while the co-citation map measures linkages and similarities among pairs of terrorism researchers as identified by citers. The co-citation data were also used in block-modeling to identify interaction patterns between subgroups of researchers within the terrorism scientific paradigms.

Table 1. Knowledge Mapping Framework and Research Questions

Type of Analysis	Unit of Analysis	Measure	Research Questions (RQs)
Basic analysis	Authors Publications Publication's citations	Productivity Impact	Who are key terrorism researchers? What institutions are they affiliated with? What are their influential terrorism publications? What are their collaboration patterns?
Content analysis	Documents Words	Coverage	What are the dominant terrorism topics? What are the new areas of research?
Co-citation analysis	Author's co-citations	Linkage	What groups of authors have papers with related content? What are the communities of researchers?

Basic Analysis

For the basic analysis, the initial step is to identify a set of key terrorism authors. We compiled a list of authors from several sources: terrorism publications [26;29], active terrorism experts identified by the KnowNet virtual community (organized by the Sandia National Laboratories), and terrorism research center portals identified on the Internet. A total of 131 unique names were identified. Names are for researchers primarily from think tanks, academic institutions, and governments located in 13 countries including UK (18), Israel (7), and France (5). Sixty-four percent are from the United States.

The second step in the basic analysis is to identify the researchers' terrorism publications. A bibliography of English-language terrorism publications was compiled for each researcher using commercial databases. The publications include journal articles, books, book chapters, reviews, notes, newspaper articles, conferences papers, and reports. Table 2 lists the ten commercial databases that were searched using author's name and terrorism-related keywords such as terrorism, hijacking, bombing, political violence, or bombing. The commercial databases were selected because of subject coverage and availability through our university library.

Bibliographical data and abstracts were downloaded, parsed and imported into a database for additional processing. After purging duplicate records, 2,148 bibliographic records were manually reviewed to identify other records that may be duplicates (non-obvious) or non-terrorism publications. Database searches for 22 researchers failed to retrieve any terrorism-related publications while no English publications were retrieved for 21 other recommended researchers. As a result, terrorism publications (bibliographic data and abstracts) were retrieved for only 88 researchers.

Table 2. Databases Used to Compile Bibliographies

Database	Discipline	Records Exported
ABI/Inform	Business, management, information sciences	164
Academic Search Premier (ASP)	Multi-disciplinary	496
Expanded Academic ASAP (EA)	Multi-disciplinary	439
International Biblographie der Zeitschriften Literature (IBZ)	International, European	161
ISI Web of Science	Social sciences, science, arts & humanities	360
PAIS International	Public affairs, business, social studies, international relations, economics	588
Political Science Abstracts (PSA)	Political science, international, politics	539
Science Direct	Science, technology, medicine	9
Sociological Abstracts	Sociology, family studies	279
WorldCat (materials cataloged by libraries around the world)	Multi-disciplinary	1,154
Total		4,129

The third step is to identify key terrorism researchers from the group of 88 researchers. The publications of the 88 terrorism researchers were analyzed using basic citation analysis to identify how frequently these are cited in the literature. Basic citation counts for each terrorism-related publication for each terrorism researcher were collected from the ISI Web of Science. Citations to each publication from any other article in the ISI dataset are counted, and each indexed author is credited with the full tally of citations to that article [20]. If an author's total number of citations for a publication in our collection is four or more then he is considered a key terrorism researcher. After an author is identified as a key researcher, his terrorism-related publication with the highest citation count is considered as his influential publication.

In addition, a coauthorship network was created to identify the collaboration patterns among the authors. The network covered the years 1965-2003. A hierarchical clustering algorithm was used to partition the core researchers who are connected if they coauthored a paper. This allows for visualization of collaboration, research teams, and institutions.

Content Map Analysis

The influential terrorism researchers' bibliographic data and abstracts were used in a content map analysis to identify the dominating themes and terrorism topics in 1965-2003. Since we want to examine more than simple frequency counts, we applied our

previous research in large-scale text analysis and visualization for content map technology to identify and visualize major research topics. The key algorithm of our content mapping program was the self-organizing map (SOM) algorithm [17]. It takes the terrorism titles and abstracts as inputs and provides the hierarchical grouping of the publications, labels of the groups, and regions of the terrorism document groups in the content map. Conceptual closeness was derived from the co-occurrence patterns of the terrorism topics. The sizes of the topic regions also generally corresponded to the number of documents assigned to the topics [22].

Co-citation Analysis

Author co-citation analysis was used to visualize the similarities among the researchers and their intellectual influence on other authors. It uses authors as the units of analysis and the co-citations of pairs of author (the number of times they are cited together by a third party) as the variable that indicates their distances from each other [1]. It was conducted based on co-citation frequencies for the key terrorism researchers, for the period 1965-2003. The co-citation map was created using a GIS algorithm developed in our lab.

We conducted terrorism keyword searches in the Web of Science to retrieve records related to the topic of terrorism. The records were used to create a terrorism citation collection and included bibliographic records for 7,590 terrorism-related articles that were downloaded. Results were parsed and loaded into a database which was used for the co-citation analysis. Table 3 summarizes the data sets used for this study.

Table 3. Data Sets Summary

Data	Web of Science (terrorism keyword searches)	10 Bibliographic Databases (author & keyword searches)
Publications	7,590	4,129
Authors	6,090	1,168
Cited References	67,453	Not retrieved
Cited Authors	32,037	Not retrieved

Program was developed to search the citation field of each bibliographic record and count the number of times two authors (or author pairs) were cited together. The result was the basis of the co-citation analysis portion of this study and offered a mapping of the field of terrorism research and the intellectual influence of the core researchers. Visualization of the relationships among researchers was displayed in a two-dimensional map that identifies their similarities, communities (clusters), and influence on emerging authors.

The co-citation data were also used in block-modeling to identify researchers' roles and positions in the terrorism research network. We used co-occurrence weight to measure the relational strength between two authors by computing how frequently they were identified in the same citing article [4]. We also calculated centrality measures to detect key members in each subgroup, such as the leaders [6]. The block-modeling algorithm is part of the social network analysis program reported in a crime data mining project.

4 Results

Basic Analysis

The basic analysis provides responses to the initial set of questions identified in Table 1. Forty-two authors were identified as key terrorism researchers. A total of 284 researchers (including coauthors) and their 882 publications made up the sample for this study.

Table 4 lists the 42 key researchers, the number of terrorism publications in our dataset, and the number of times the researchers' publications were cited in the ISI

Table 4. Forty-two Key Terrorism Researchers (based on citation score in ISI)

Author Name	# of Pubs*	# Times Cited	Author Name	# of Pubs*	# Times Cited
1. Wilkinson, Paul	87	**229**	22. Lesser, Ian O.	5	**23**
2. Gurr, T.R.	51	**214**	23. Bassiouni, M.C.	8	**22**
3. Laqueur, Walter	37	**191**	24. Carlton, David	1	**21**
4. Alexander, Yonah	88	**169**	25. Chalk, Peter	17	**20**
5. Bell, J.B.	47	**138**	26. Freedman, Lawrence	14	**20**
6.. Stohl, M.	30	**136**	27. Merari, Ariel	25	**19**
7.. Hoffman, Bruce	121	**100**	28. Post, Jerrold	12	**18**
8. Jenkins, Brian M.	38	**96**	29. Evans, Ernest H.	3	**17**
9.. Ronfeldt, David	20	**95**	30. Bergen, Peter	10	**16**
10. Crenshaw, Martha	40	**90**	31. Gunaratna, Rohan	14	**16**
11. Arquilla, John	20	**75**	32. Cline, R.S.	8	**15**
12. Mickolus, Edward F.	25	**73**	33. Friedlander, R.A.	4	**14**
13. Crelinsten, Ronald	19	**62**	34. Paust, Jordon J.	11	**13**
14. Schmid, Alex P.	6	**59**	35. Ranstorp, Magnus	8	**13**
15. Wardlaw, G.	25	**49**	36. Flynn, Stephen E	4	**12**
16.. Hacker, F.J.	3	**38**	37. Cooper, H.H.A	10	**11**
17. Rapoport, David	26	**37**	38. Wolf, J.B	7	**11**
18. Sloan, Stephen R	31	**30**	39. Horgan, John	13	**10**
19. Dobson, C.	6	**25**	40. Sterling, C.	5	**10**
20. Kepel, Gilles	6	**25**	41. McCauley, Clark	4	**8**
21. Stern, Jessica E	21	**25**	42. Merkl, Peter	6	**6**

* number of publications in our dataset

databases. They are mainly affiliated with academic institutions (23), think tanks (15), media organizations (3), and the government (1). Their bases of operation are located in nine countries including the US (29), UK (4), and Ireland (1).

The Appendix lists the most influential publication for each researcher which is based on the number of times cited in the ISI Web of Science. Table 5 lists the 12 most influential publications because they were cited more than twenty-five times in ISI databases.

Table 5. Most Influential Terrorism Publications

1. Publication	1. # Times cited	1. Topic	1. Author	1. Organization
2. 1.*Why men rebel*, 1970	**2. 145**	2. political violence	2. Gurr, Ted	2. Univ Maryland
3. 2. *Terrorism*, 1977	**3. 75**	3. terrorism historical aspects	3. Laqueur, Walter	3. Center for Strategic & Intl Studies (CSIS)
4. 3. *Terrorism & liberal state*, 1977	**4. 66**	4. terrorism prevention	4. Wilkinson, Paul	4. Univ Aberdeen (formerly), CSTPV
5. 4. *Inside terrorism*, 1998	**5. 47**	5. terrorism religious aspects	5. Hoffman, Bruce	5. Rand Corporation
6. 5. *Trans. Terrorism, a chronology*, 1980	**6. 41**	6. terrorism incidents	6. Mickolus, E.	6. CIA (formerly)
7. 6. *Crusaders, criminals*, 1976	**7. 34**	7. terrorism case study	7. Hacker, F.J. (deceased)	7. USC Medical & Law Schools
8. 7. *Time of terror*, 1978	**8. 33**	8. terrorism responses	8. Bell, J.B. (deceased)	8. Columbia Univ
9. 8. *State as terrorist*, 1984	**9. 32**	9. state sponsored terrorism	9. Stohl, M.	9. Purdue Univ
10.9. *Political terrorism theory, tactics*, 1982	**10.31**	10.terrorism prevention	10.Wardlaw, G	10.Australian Institute of Criminology
11.10. *Intl. terrorism national regional*, 1976	**11.30**	11.terrorism anthology	11.Alexander, Y.	11.CSIS; SUNY
12.11. *Political terrorism a new guide*, 1988	**12.29**	12.terrorism directory	12.Schmid, Alex P.	12.Royal Netherlands Academy of Arts & Science
13.12. *Intl. Terrorism a new mode*, 1975	**13.27**	13.terrorism	13.Jenkins, Brian M.	13.Rand Corporation

An investigation of the coauthorship patterns provides an understanding of the researchers' social network patterns. Figure 2 exhibits the coauthorship network of key researchers in scientific collaboration networks. The nodes represent researchers who coauthored papers.

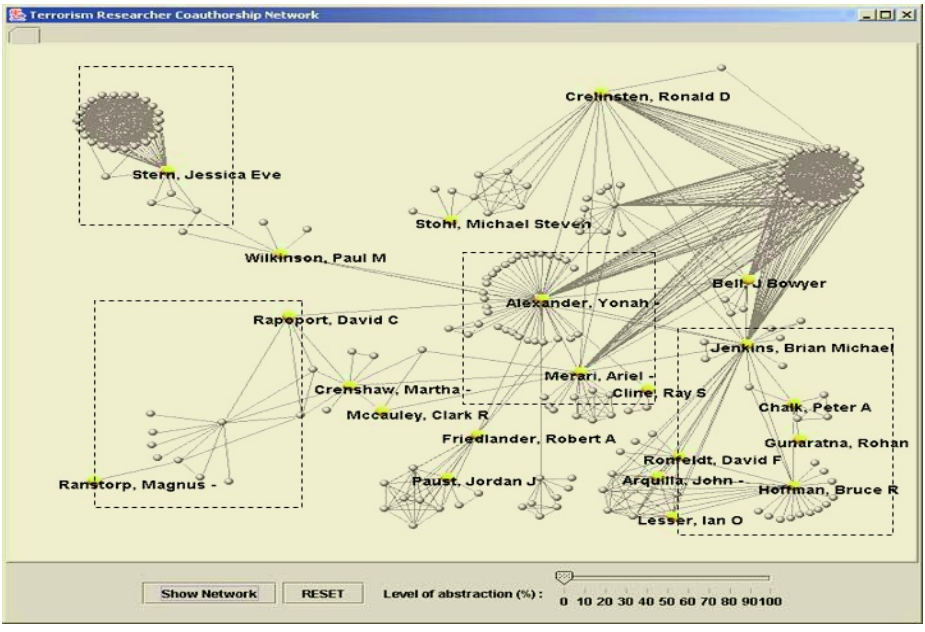


Fig. 2. Key Terrorism Researchers' Coauthorship Network

In the lower right corner of Figure 2, the Rand research teams led by Jenkins and Hoffman is one of the most active clusters. Except for Gunaratna, all of the researchers in the cluster are Rand's employees. Gunaratna coauthored publications with Chalk and Hoffman, his PhD advisor at St. Andrews University, Scotland, and founded the terrorism research center at the Institute of Defence and Strategic Studies, Singapore. Hoffman founded St. Andrews' Centre for the Study of Terrorism and Political Violence (CSTPV) and created the Rand-St. Andrews terrorism incident database which provides data for their studies [18].

For the cluster in the lower left corner that includes Ranstorp from CSTPV, it is sparse and shares few coauthorships. As chairman of the Advisory Board for CSTPV, Wilkinson has a few collaborations with Alexander but none with researchers at CSTPV who are in this sample. Another cluster includes researchers such as Alexander and Cline at the Center for Strategic and International Studies (CSIS). Since Alexander has 82 coauthors, this cluster displays a pattern of one to many coauthors. We found that coauthorships do not seem to be sustainable because many authors produce only a single publication with Alexander and did not publish with other terrorism researchers in this sample.

Content Map Analysis

Regarding the next set of questions identified in Table 1, several dominating terrorism topics have been identified for 1965-2003. Figure 3 displays the contemporary terrorism content map that was generated based on the title and abstracts of the 882 terrorism-related publications in our dataset. The topic map interface contains two components, a folder tree display on the left-hand side and a hierarchical content map

on the right-hand side [17]. The terrorism publications are organized under topics that are represented as nodes in the folder tree and colored regions in the content map. These topics were labeled by representative noun phrases identified by our programs. The number of terrorism publications that were assigned to the first-level topics is displayed in parenthesis after the topic labels.

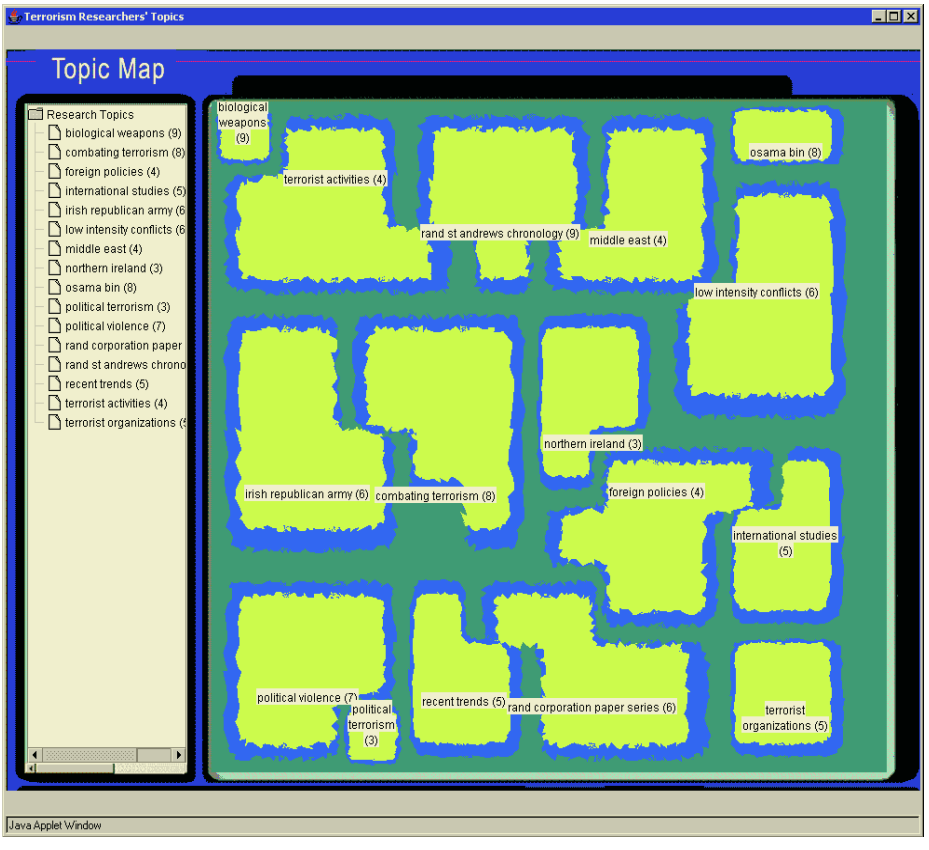


Fig. 3. Contemporary Terrorism Content Map: 1965-2003

Major terrorism topics (large regions with depth in the content map) include "low intensity conflicts," "rand corporation paper series," "osama bin,' "political violence," "rand st andrews chronology," and "irish republican army". The topics "rand corporation paper series" and "rand st andrews chronology" highlight the major roles that Brian Jenkins, one of the pioneers of modern terrorism studies [33], and Paul Wilkinson, Chairman of the St. Andrews' Centre for the Study of Terrorism and Political Violence (CSTPV), Scotland, played. They established terrorism research centers, created databases of terrorism incidents, secured funding for terrorism research projects, produced terrorism studies, and supervised student's research on terrorism [25].

Several interesting shifts in the cognitive structure of contemporary terrorism research are identified. A traditional terrorism topic, "low intensity conflicts," first appeared in 1991 and appeared seven other times in the 1990s but only one time in 2000s. Prior to 11th September, the conventional wisdom was that the use of terrorism was endemic in low intensity conflict but that it rarely, if ever, posed a strategic threat to the security of major international powers [33]. After 1997, there was an increasing appearance of the topic "osama bin" which first emerged in our dataset in 1998 as the subject of an article by Peter Bergen [2]. "Osama bin' referring to Osama Bin Laden is a new topic of interest.

Co-citation Analysis

For the final set of questions identified in Table 1, the author co-citation analysis is used to visualize the closeness of research interests among the key terrorism researchers and their intellectual influences on others. The raw co-citation data derived from keyword searches of the ISI Web of Science were used for the analysis conducted in this part of the study. We created author co-citation networks to identify which key researchers in terrorism are often cited together.

Figure 4 shows a sample of pairs of authors (researchers) linked by co-citation counts of 1-3. Authorship nodes are represented either by a square or circle followed by the last name of the first author, publication source, and year. The square node identifies a publication that cites the key terrorism researchers (circular nodes). The width of the arrows connecting authorship nodes have been made proportionate to their co-citation counts in size. The narrow arrow width reflects a count of one co-citation link while a thick one reflects a count of at least two co-citation links.

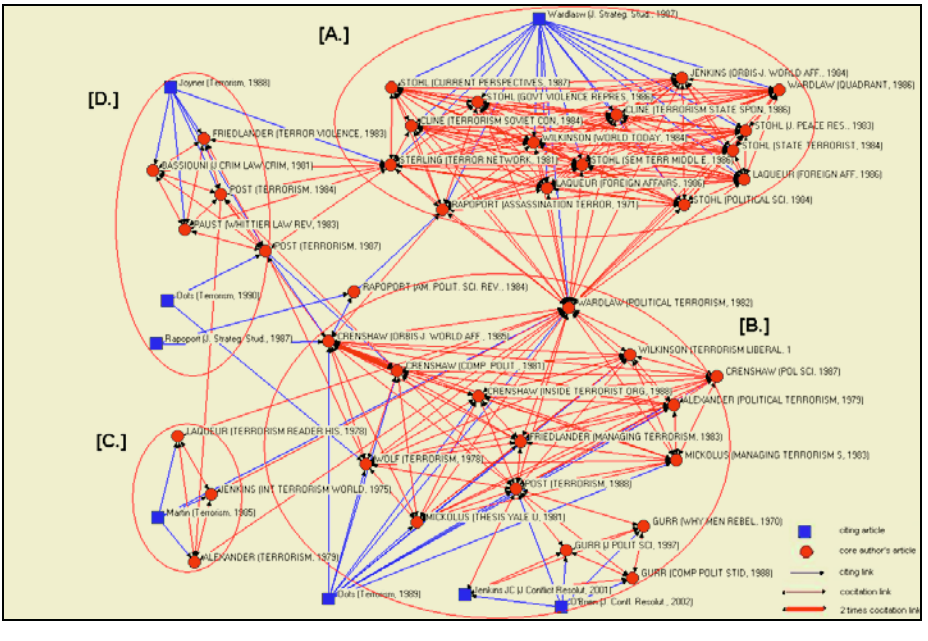


Fig. 4. Key Terrorism Researchers' Co-citation Network

To illustrate the findings represented through the author co-citation map, boundaries were drawn around clusters of researchers. Figure 4 illustrates four groupings of author co-citation patterns.

The groupings provide a way of clustering pairs of researchers who share areas of interests. For example, publications cited in Group A focuses on terrorism and foreign policy (based on terms from the titles and abstracts of their publications). In Group A, Wardlaw's article on terror as an instrument of foreign-policy is citing several of the most frequent co-cited pairs. The most frequently appearing author co-cited pairs are Laqueur and Wardlaw (13 times), Stohl and Wardlaw (12 times), and Cline and Stohl (12 times). Cline and Stohl specialized in state sponsored terrorism.

Group B emphasizes the organizational perspectives of terrorism. It includes Oots' publication entitled "Organizational Perspectives on the Formation and Disintegration of Terrorist Groups." Oots cites seven of the key researchers and identifies almost fifty author co-citation pairs. Group C's subject deals with historical aspects; while that of Group D is legal aspects of terrorism.

Another way of viewing subgroups and key members in contemporary terrorism research is to analyze their interaction patterns to identify the roles and positions that they play. It was found that, as Figure 5 shows, 18 terrorism researchers from the resulting network were co-cited in ISI.

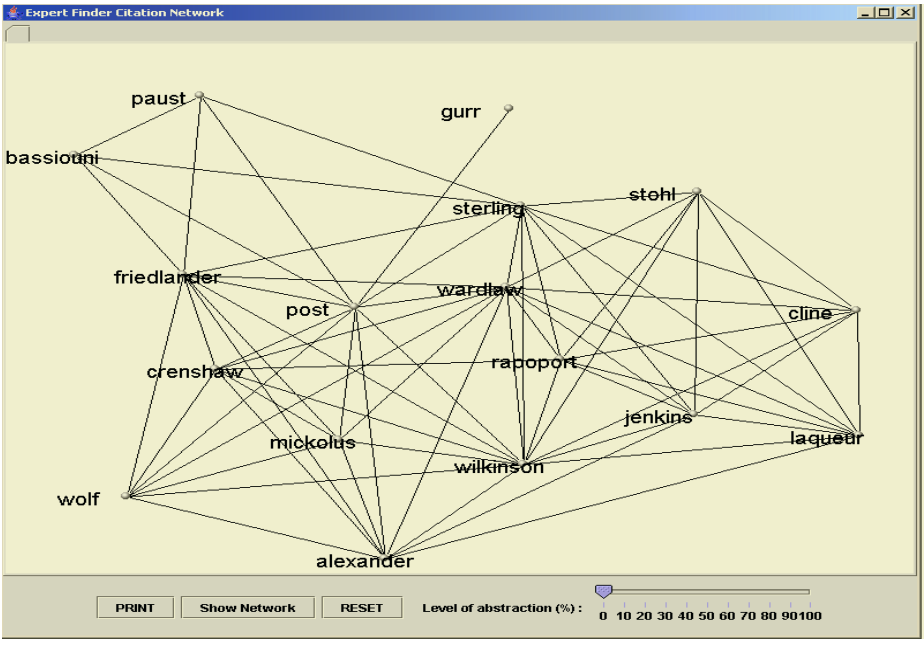


Fig. 5. The 18 Key Terrorism Researchers Who Were Co-cited in ISI

Figure 6 shows the subgroups identified by the system. They have the labels of their leaders' names (Crenshaw, Post, and Stohl). The thickness of the straight lines indicates the strength of relationships between subgroups.

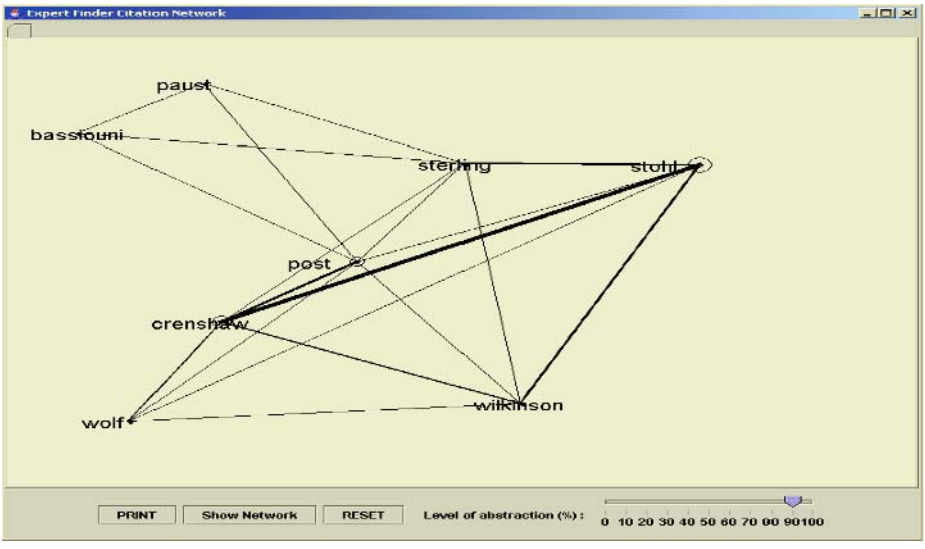


Fig. 6. Subgroups of Co-cited Authors and Tagged with Leaders' Names

For example, Crenshaw's group consists of Mickolous (cited with Crenshaw eight times), Post (cited with Crenshaw six times), Wolf (cited with Crenshaw six times), etc. Those familiar with terrorism research would not be surprised with the close co-cited relationship between Crenshaw and Post because they focus on the psychological aspects of terrorism with Crenshaw positing that there is no profile of the typical terrorist[8].

5 Conclusion

The mapping of contemporary terrorism research provides a perspective that heretofore has not been afforded. As such, the tools such as content map analysis and co-citation analysis can help individuals visualize scholarly development within the field. For instance, while those familiar with terrorism will already know that, say, Stohl and Cline worked in similar areas and are often cited together, those who are not well oriented with the field, particularly new researchers could find such information relevant.

Although there are benefits of using visualization techniques, visualization is not a substitute for extensive reading and detailed content analysis for understanding the development of a field. For new researchers, it provides an alternative approach for understanding quickly the structure and development of a field. Thus, the knowledge mapping framework and tools provided here, could allow the expanding group of non-traditional terrorism researchers to conduct systematic exploitation of the terrorism field and identify trends and research gaps in a short period of time. This approach helps identify influential researchers in a field, the amount they are cited, the topics that are being investigated, and the frequency of co-citation with other terrorism authors who perhaps work in similar subject areas. With the current

challenges in the interdisciplinary and international field of terrorism, new researchers must understand the intellectual structure of the field and how they can better frame their research questions.

We intend to supplement this work with other studies that will use time-series topic maps to present the development trends in terrorism across various periods to further examine the recent evolution and topic changes in the field. We will also include author content map analysis to group individual researchers based on their common research interests. In addition, we will use the results to develop a terrorism expert finder application that supports domain visualization and field test it with new and experienced terrorism researchers.

Acknowledgement

This research was funded by the National Science Foundation (NSF) and the ITR grant. Authors wish to thank the KnowNet Community, Dr. Jerold Post, and Dr. Marc Sageman for their support and assistance.

References

1. Andrews, J.E., Author Co-citation Analysis of Medical Informatics. Journal of the Medical Library Association, 2003. 91(1): p. 47-56.
2. Bergen, P., Holy War, Inc.: Inside the Secret World of Osama Bin Laden. 2001, New York: Free Press.
3. Borgman, C.L. and J. Furner, Scholarly Communication and Bibliometrics, In Annual Review of Information Science and Technology (ARIST). 2002, ASIST.
4. Boyack, K.W., Mapping Knowledge Domains: Characterizing PNAS. Arthur M. Sackler Colloquium of the National Academy of Sciences, N.A.o.S. (NAS), Editor. 2003, NAS: Irvine.
5. Chen, C. and R.J. Paul, Visualizing a Knowledge Domain's Intellectual Structure. IEEE Computer Society, 2001. 34(3): p. 65-71.
6. Chen, H., et al., Crime Data Mining: a General Framework and Some Examples. IEEE Computer Society, 2004: p. 50-56.
7. Chua, C., et al., Measuring Researcher-Production in Information Systems. Journal of the Association for Information Systems, 2002. 2: p. 146-215.
8. Crenshaw, M., Psychology of Terrorism: An Agenda for the 21st Century. Political Psychology, 2000. 21(1): p. 405-420.
9. Cronin, B., High-Fidelity Mapping of Intellectual Space: Early and Recent Insights from Information Science, in Spaces, Spatiality and Technology Workshop. 2002, Napier University, Edinburgh Scotland: Edinburgh.
10. Eggers, S., et al., Mapping Medical Informatics Research, in Medical Informatics: Knowledge Management and Data Mining in Biomedicine. Forthcoming, Springer Science.
11. Ferligoj, A., P. Doreian, and V. Batagelj, Optimizational Approach to Blockmodeling. Journal of Computing and Information Technology, 1996. 4: p. 63-90.
12. Garfield, E. and Welljams-Dorof, Citation Data: their Use as Quantitative Indicators for Science and Technology Evaluation and Policy-making. Science & Public Policy, 1992. 19(5): p. 321-327.

13. Garfield, E., A.I. Pudovkin, and V.S. Istomin, Algorithmic Citation-Linked Historiography: Mapping the Literature of Science, In ASIST 2002 Contributed Paper. 2002.
14. Gordon, A., Terrorism Dissertations and the Evolution of a Specialty: an Analysis of Meta-Information. Terrorism and Political Violence, 1999. 11(2): p. 141-150.
15. Gordon, A., Effect of Database and Website Inconstancy on the Terrorism Field's Delineation. Studies in Conflict & Terrorism, 2004. 27: p. 79-88.
16. Huang, Z., et al., Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field. Journal of Nanoparticle Research, 2003. 5: p. 333-363.
17. Huang, Z., et al., International Nanotechnology Development in 2003: Country, Institution, and Technology Field Analysis Based on USPTO Patent Database. Journal of Nanoparticle Research, 2004. 6: p. 325-354.
18. Hughes, G., Analyze This, in The Age. 2003.
19. Incites, Citation Thresholds. 2003, Institute for Scientific Information (ISI): Philadelphia.
20. ISI, How Does ISI identify Highly Cited Researchers? 2003, Institute for Scientific Information (ISI): Philadephia.
21. Kennedy, L.W. and C.M. Lum, Developing a Foundation for Policy Relevant Terrorism Research in Criminology. 2003, Rutgers University: New Brunswick.
22. Lin, X., H.D. White, and J. Buzydlowski. AuthorLink: Instant Author Co-citation Mapping for Online Searching. In National Online Proceedings 2001. New York City: Information Today.
23. McCain, K.W., Mapping Authors in Intellectual Space: a Technical Overview. Journal of the American Society of Information Science, 1990. 41(6).
24. NameBase, Public Information Research, Inc.: 2004. San Antonio.
25. Reid, E.O.F., Analysis of Terrorism Literature: a Bibliometric and Content Analysis Study. 1983, University of Southern California: Los Angeles.
26. Reid, E.O.F., Evolution of a Body of Knowledge: an Analysis of Terrorism Research. Information Processing & Management, 1997. 33(1): p. 91-106.
27. Reid, E., et al. Terrorism Knowledge Discovery Project: a Knowledge Discovery Approach to Addressing the Threats of Terrorism. In Second Symposium on Intelligence and Security Informatics, ISI 2004, June 2004 Proceedings. 2004. Tucson, Arizona: Springer-Verlag.
28. Reid, E.O.F. Identifying a Company's Non-Customer Online Communities. in Proceedings of the 36th International Conference on Systems Sciences (HICSS). 2004. Hawaii: HICSS.
29. Schmid, A. and A. Jongman, Political Terrorism: A New Guide to Actors, Authors, Concepts, Data Bases, Theories and Literature. 1988, Oxford: North Holland.
30. Shiffrin, R.M. and K. Borner. Mapping Knowledge Domains. In Arthur M. Sackler Colloquium of the National Academy of Sciences. Held May 9-11, 2003, at the Arnold & Mabel Beckman Center of the National Academies of Sciences & Engineering. 2004. Irvine, CA: NAS.
31. White, H.D. and K.W. McCain, Visualizing a Discipline: an Author Co-citation Analysis of Information Science 1972-1995. Journal of the American Society of Information Science, 1998. 49(4): p. 327-355.
32. White, H.D., X. Lin, and J. Buzydlowski. Co-cited Author Maps as Real-time Interfaces for Web-based Document Retrieval in the Humanities. In Joint International Conference o the Association for Computers and the Humanities and the Association for Literary and Linguistics Computing (ALLC). 2001. New York City: ACH/ALLC.
33. Wilkinson, P. Terrorism: Implications for World Peace. In Westermorland General Meeting Preparing for Peace Initiative. 2003. United Kingdom: Westermorland.

Appendix: List of 42 Influential Terrorism Researchers (as of Dec. 2003)

Author Name	No. of Pub.	Active Years	# times cited for pubs in collection	Most Frequently Cited Terrorism Publication	Date	# times cited
1. Alexander, Yonah	88	32	**169**	Intl. terrorism national regional	1976	30
2. Arquilla, John	20	30	75	Cyberwar is coming	1993	18
3. Bassiouni, M.C.	8	17	22	Intl. terrorism & political …	1975	16
4. Bell, J.B.	47	35	138	Time of terror	1978	33
5. Bergen, Peter	10	7	16	Holy war inc	2001	15
6. Carlton, David	1	2	21	Terrorism theory & practice	1979	21
7. Chalk, Peter	17	26	20	West European terrorism	1996	7
8. Cline, R.S.	8	14	15	Terrorism the Soviet	1984	14
9. Cooper, H.H.A	10	25	11	Chapter in Terrorism Interdiscip.	1977	7
10. Crelinsten, Ronald	19	28	62	Political terrorism a research guide	1993	22
11. Crenshaw, Martha	40	35	**90**	Why violence spreads	1980	23
12. Dobson, C.	6	14	25	Black September	1974	8
13. Evans, Ernest H.	3	4	17	Calling a truce	1979	17
14. Flynn, Stephen E	4	4	12	Beyond border	2000	8
15. Freedman, Lawrence Z.	14	21	20	Terrorism & Intl Order	1986	7
16. Friedlander, R.A.	4	10	14	Terror violence Terrorism documents	1983 1979	7 7
17. Gunaratna, Rohan	14	8	16	Inside al qaeda	2002	14
18. Gurr, T.R.	51	41	**214**	Why men rebel	1970	145
19. Hacker, F.J.	3	5	38	Crusaders, criminals	1976	34
20. Hoffman, Bruce	121	27	**100**	Inside terrorism	1998	45
21. Horgan, John	13	18	10	Technology vs terrorism	1986	5
22. Jenkins,	38	30	**96**	Intl. terrorism	1975	27

Author Name	No. of Pub.	Active Years	# times cited for pubs in collection	Most Frequently Cited Terrorism Publication	Date	# times cited
Brian M.				new mode		
23. Kepel, Gilles	6	4	25	Jihad expansion	2000	16
24. Laqueur, Walter	37	28	**191**	Terrorism	1977	75
25. Lesser, Ian O.	5	30	23	Intl. terrorism a chronology	1975	13
26. McCauley, Clark	4	12	8	Terrorism research & public	1991	8
27. Merari, Ariel	25	26	19	Readiness to kill & die	1990	8
28. Merkl, Peter	6	18	6	Political violence & terror	1986	6
29. Mickolus, Edward F.	25	28	73	Trans. terrorism, a chronology	1980	41
30. Paust, Jordon J.	11	30	13	Federal jurisdiction over …	1983	11
31. Post, Jerrold	12	19	18	Terrorist psycho logic	1990	12
32. Ranstorp, Magnus	8	13	13	Hizb'allah in …	1997	7
33. Rapoport, David	26	33	37	Assassination & terrorism	1971	20
34. Ronfeldt, David	20	30	**95**	Cyberway is coming Networks & netwars	1993 2001	18 18
35. Schmid, Alex P.	6	7	59	Political terrorism a new guide	1988	29
36. Sloan, Stephen R	31	34	30	Simulating terrorism	1981	10
37. Sterling, C.	5	7	10	Terror network	1981	10
38. Stern, Jessica E	21	13	25	Prospects of domestic bioterrorism	1999	12
39. Stohl, M.	30	28	136	State as terrorist	1984	32
40. Wardlaw, G.	25	23	49	Political terrorism theory, tactics	1982	31
41. Wilkinson, Paul	87	32	229	Terrorism & liberal state	1977	66
42. Wolf, J.B	7	16	11	Fear of fear	1981	5

Bold indicates most influential publications

Testing a Rational Choice Model of Airline Hijackings

Laura Dugan[1], Gary LaFree[1], and Alex R. Piquero[2]

[1] Department of Criminology and Criminal Justice, University of Maryland,
2220 LeFrak Hall, College Park, MD 20742
{ldugan, glafree}@crim.umd.edu
[2] Department of Criminology, Law and Society, University of Florida,
PO Box 115950, Gainesville, FL 32611
apiquero@crim.ufl.edu

Abstract. Using data that combines information from the Federal Aviation Administration, the RAND Corporation, and a newly developed database on global terrorist activity, we are able to examine trends in 1,101 attempted aerial hijackings that occurred around the world from 1931 to 2003. We have especially complete information for 828 hijackings that occurred before 1986. Using a rational choice theoretical framework, we employ econometric time-series methods to estimate the impact of several major counter hijacking interventions on the likelihood of differently motivated hijacking events and to model the predictors of successful hijackings. Some of the interventions examined use certainty-based strategies of target hardening to reduce the perceived likelihood of success while others focus on raising the perceived costs of hijacking by increasing the severity of punishment. We also assess which specific intervention strategies were most effective for deterring hijackers whose major purpose was terrorism related. We found support for the conclusion that new hijacking attempts were less likely to be undertaken when the certainty of apprehension was increased through metal detectors and law enforcement at passenger checkpoints. We also found that fewer hijackers attempted to divert airliners to Cuba once that country made it a crime to hijack flights. Our results support the contagion view that hijacking rates significantly increase after a series of hijackings closely-clustered in time. Finally, we found that policy interventions only significantly decrease the likelihood of non-terrorist-related hijackings.

1 Introduction

Over the past several decades, the rational choice perspective has been applied to a wide variety of criminal behavior, including drunk driving [31], burglary [50], robbery [51], shoplifting [41], income tax evasion [23], drug selling [19], and white collar crime [36], [46]. In this paper we use a rational choice perspective to develop a series of hypotheses about the success, benefits and costs of aerial hijacking. Rational choice theory would seem to be an especially appropriate theoretical perspective for understanding hijackings, given that many are carefully planned and appear to include at least some consideration for risks and rewards. We develop a series of hypotheses about hijackings and test them with a data base obtained from the Federal Aviation

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 340 – 361, 2005.

Administration with additional data from the RAND Corporation, and a newly developed data base on global terrorism [24]. Based on hazard modeling, our results support the conclusion that some certainty of apprehension measures (metal detectors and law enforcement at passenger check points) did significantly reduce the rate of new hijacking attempts. Also, a severity of punishment measure that made hijacking a crime in Cuba was significantly related to a drop in the hazard that a hijacked flight would be diverted to Cuba. We additionally found support for a contagion view that the rate of hijackings significantly increases following a series of hijackings closely-clustered in time. Finally, policy interventions only significantly impact the likelihood of non-terrorist-related hijackings.

Before we present the results of our analysis, we first provide an overview of rational choice theory and the prior research on rational choice theory and aerial hijacking.

2 Rational Choice Theory

The belief that credible threats of apprehension and punishment deter crime is as old as criminal law itself and has broad appeal to both policymakers and the public. As elaborated by social reformers like Bentham and Beccaria, or jurists like Blackstone, Romilly, or Feuerbach, rational actor perspectives assume that crime can be deterred by increasing the costs of crime or increasing the rewards of non-crime [16], [45], [34]. In particular, the principle of utility advanced by Bentham proposed that individuals act in view of their own self-interest and that the effective use of punishment serves to deter individuals from specific actions (including crime) that serve their self-interest.

Many contemporary rational choice models of crime [3], [6] express utilitarian philosophy in mathematical terms, with individuals maximizing satisfaction by choosing one of a finite set of alternatives, each with its particular costs and benefits [10], [9:5]. At their core, these rational choice models suggest that crime can be deterred through appropriate public policy. In general, the choice of crime is more appealing when legal options are less rewarding, when crime is less punishing, and/or when crime is more rewarding. Research on the rational choice perspective has increased our understanding of the costs and benefits associated with both crime and non-crime alternatives [38], [8], and recent evidence suggests that the criminal justice system can exert a deterrent effect on crime (for a review, see [29]).

Mathematically, a rational choice explanation of crime suggests that if p(success)*benefits > [1-p(success)]*costs, then crime is more likely to occur, and conversely, if p(success)*benefits < [1-p(success)]*costs, then crime is less likely to occur. The probability of *success*, p(success), is a function of the offender's perception. The rational choice perspective assumes that offenders calculate their probability of success when evaluating criminal opportunities. In general, a major goal of policy makers who design formal systems of punishment is to control or alter this calculation through policies aimed at reducing the certainty of success. In the case of policies on aerial hijacking for the past half century, this goal has been pursued primarily through target hardening including metal detectors, posting security personnel at airport gates, and baggage-screening.

According to the rational choice perspective, *benefits* can be both internal (e.g., monetary gain) and external (e.g., achieving political recognition) to offenders. Further, as prospective perpetrators witness others' hijacking successes, they may be more likely to use hijacking as a means to achieve their own goals. Piquero and Pogarsky [40] and others [48], [35], [39] have found that this vicarious experience with punishment avoidance is an important determinant of both the perception of sanctions and criminal behavior. Examples of such benefits in the case of aerial hijacking include the rapid growth of hijackings to Cuba in the late 1960s and early 1970s (before Cuba defined hijacking as a crime) and the rash of hijackings for the extortion of money after the widely publicized success of D.B. Cooper in November 1971.[1] The role of benefits in rational choice theory is closely related to the concept of contagion, which we discuss below.

The rational choice perspective also posits that offenders interpret and weigh the *costs* associated with their offending decisions. Such costs include the probability of punishment, as well as the severity of punishment experiences. Accordingly, policymakers try to raise the perceived costs of aerial hijacking by increasing the certainty of detection and by strengthening the severity of punishment. For example, several laws passed in the United States during the 1960s and 1970s were aimed at increasing punishment severity for airplane hijacking including a Cuban law in October 1970 that for the first time made hijacking a crime in Cuba. At the same time, policies such as posting security personnel at airport gates and placing sky marshals on aircraft were efforts aimed at increasing the certainty of punishment.

To summarize, the rational choice perspective predicts that the frequency of aerial hijackings will decrease if the probability of success is decreased, the perceived benefits are reduced, and the perceived costs are increased. In addition to testing specific hypotheses developed from rational choice theory, our analysis permits us to explore whether these general expectations hold equally well depending on the location of the incident and the likely motivation of hijackers. In particular, we distinguish in the analysis between hijacking incidents that originated in the U.S., those that originated elsewhere, offenders whose major purpose appears to be transportation to Cuba, and offenders who we classify as having a terrorist purpose.

3 Prior Research

We were able to identify three early studies that explicitly examined the rational choice perspective within the context of aerial hijacking [7], [25], [28]. All three of these studies focus only on the cost component of the rational choice framework. Chauncey [7] examined five deterrence-based policy efforts (two representing changes in the probability of success or certainty, two representing changes in severity, and one combining the two) related to hijacking incidents and found that only the two certainty events produced reductions in the rate of attempts, with the

[1] A hijacker using the name D.B. Cooper seized control of a Northwest Orient airliner and threatened to blow it up during a flight from Portland to Seattle. After he extorted $200,000 he parachuted from the flight and has never been found. This event gained national attention and the fact that Cooper successfully avoided detection gave him folk legend status with admirers [12].

largest reduction being a function of the metal detector screening/carry-on baggage inspection policy implemented in the first quarter of 1973 in U.S. airports. Minor [28] applied deterrence/prevention concepts to understand skyjacking in the U.S. and worldwide, and concluded that there was no major deterrent effect of skyjacking control programs before 1973, but that there was a prevention effect in 1973 and 1974 due to the implementation of baggage screening and metal detectors. Unfortunately, neither Chauncey nor Minor offer systematic statistical tests of their hypotheses about deterrence and prevention.

Following Becker [3] and Ehrlich [13], Landes [25] developed and tested an economic model of hijacking, conducting a quarterly analysis of mainly U.S. aircraft hijacking between 1961 and 1976. His results show that an increase in the probability of apprehension, the conditional probability of incarceration, and the length of sentence for those convicted of hijacking were all associated with significant reductions in hijacking during the 1961 to 1976 period. Additionally, using regression estimates from the sample period ending in 1972, Landes developed forecasts of the number of hijackings that would have taken place between 1973 and 1976 if (1) mandatory screening had not been instituted and (2) the probability of apprehension (once the hijacking was attempted) had remained constant and equal to its 1972 value. He concluded that without these interventions there would have been between 41 and 67 additional hijackings during the 1973 to 1976 period compared to the 11 that actually occurred.

While they do not specifically adopt a rational choice perspective, Hamblin, Jacobsen and Miller [17] and others [44], [37] rely on contagion or diffusion explanations of hijacking attempts to make predictions that are closely related to the reward component of the rational choice perspective. Thus, researchers supporting a contagion model assume that when potential aerial hijackers perceive that previous hijacking attempts have been rewarded (e.g., successful outcomes, avoidance of punishment) and that they can avoid punishment in the commission of a hijacking, they will be more likely to offend. For example, Holden [18] argues that successful airline hijackings will foster more airline hijackings while unsuccessful episodes will lead to fewer new skyjacking attempts. Related arguments include Rich's [44] claim that a "skyjack virus" may be transmitted through the media; Phillips' [37] argument that imitation explains the frequency of hijackings; and Hamblin, Jacobsen, and Miller's [17] argument that hijackings spread by diffusion and modification of a basic invention, as new hijackers attempt to outdo previous ones by inventing more effective hijacking strategies.

In the most detailed empirical study of the contagion hypothesis to date, Holden [18] develops a mathematical model of contagion and applies it to aircraft hijackings in the U.S. between 1968 and 1972. Defining contagion as an increase in the rate of new hijacking attempts, Holden [18:886] tests five hypotheses. First, the rate of aircraft hijacking attempts in the U.S. will increase following other hijacking attempts. Second, the rate of aircraft hijacking attempts in the U.S. will increase following publicized hijacking attempts, but not following unpublicized attempts. Third, compared to unsuccessful attempts, successful (i.e., rewarded) hijacking attempts will have a greater stimulating effect on additional hijackings. Fourth, because the motivation for transportation and extortion hijacking attempts may be very different and because history shows that the peak periods for transportation (1969-1970) and extortion (1972) hijackings were separated by three years,

transportation hijackings should be stimulated only by prior transportation hijackings, and extortion hijackings only by prior extortion hijackings. And finally, the stimulating effect on the U.S. hijacking rate will be far greater for hijackings on U.S. carriers than non-U.S. carriers.

Holden's research shows that successful hijackings generate additional hijacking attempts of the same type (transportation or extortion), but finds no contagion effects of unsuccessful hijacking attempts in the U.S. or successful or unsuccessful hijacking attempts outside the U.S. In particular, each successful transportation hijacking in the U.S. generated an average of .75 additional attempts, with a median delay of 60 days. This effect accounted for 53% of the total rate of U.S. transportation hijacking attempts in Holden's analysis. Each successful extortion hijacking in the U.S. generated an average of two additional hijacking attempts, with a median delay of 44 days, accounting for 85% of the total rate of U.S. extortion hijacking attempts. Holden's results also show (pp. 898-899) that while U.S. hijackers were not influenced by incidents outside the U.S., the likelihood of foreign extortion-based hijackings (including parachute hijackers) were increased by hijackings in the U.S.[2]

Although instructive, prior research on aerial hijacking from the rational choice perspective is limited in several ways. First, while there is some descriptive information available on overall trends in hijacking events [27], [21], much less is known about the effect of hijackers' motives on the frequency and success of the crime in the U.S. and elsewhere. Second, much of the prior research does not use formal statistical tests to determine if deterrent/preventive policies significantly reduce hijacking. Third, most studies [7], [28] have focused on the costs component of the rational choice framework and the only major study to examine the benefits component [18], did so through a contagion approach using data from a limited time span (1968 to 1972). And finally, past efforts have not examined the specific variables that are associated with hijacking success. For example, Holden's research distinguished successful from unsuccessful hijackings, but he includes no analysis of the variables that predict successful hijackings. Our study specifically addresses these limitations.

4 Current Focus and Hypotheses

We employ hazard modeling [11] to identify how a set of theoretically relevant variables (e.g., success and purpose of attack) affect the time between hijacking incidents.[3] This approach allows us to determine the variables that reduce the

[2] Holden's [18:879] extortion category "includes incidents involving both extortion (i.e., demands other than for transportation) and diversion to a particular destination because the primary motive in these cases is presumed to be other than transportation."

[3] Although we do not empirically distinguish between deterrent and preventive effects, it is useful to briefly explain the two. Prevention, according to Andenaes [2] and Jeffery [20] refers to the elimination of the opportunity for crime through modification of the environment in which crime occurs. Zimring and Hawkins [52:351] suggest that: "..if the probability that a particular type of offender will be apprehended is greatly increased, then the increased apprehension rate may achieve a substantial *preventive* effect which is quite independent of the *deterrent* effect of the escalation in enforcement…Nevertheless…it is crime prevention rather than deterrence which is the ultimate object of crime control measures."

temporal frequency of hijacking incidents. We then use logistic regression analysis to identify the qualities of hijacking attempts that are most likely to contribute to their success.

We develop five hypotheses derived from success, benefits and cost-related assumptions of the rational choice perspective.[4] For the purposes of this paper, we adopted the FAA's [15] definition of a successful hijacking as one *in which hijackers gain control of the plane and reach their destination, whether by landing or by a parachute escape, and are not immediately arrested or killed on landing; unsuccessful hijackings are those in which hijackers attempt but fail to take control of an aircraft*.[5] Our success-related hypothesis:

H1: A new hijacking attempt will be less likely when the certainty of apprehension is increased.

Hypothesis 1 is based on the fundamental rational choice prediction that the chances of additional prohibited behavior will decline when perpetrators can be expected to believe that the likelihood of success has lessened. We discuss below how we will use the timing of two certainty-based security policies to test this hypothesis. We also conduct an exploratory analysis to determine which flight characteristics and policies actually do increase the chances that hijackers will be apprehended.

The three benefits-related hypotheses are based on the premise that offenders will be more likely to attempt aerial hijackings when the expected benefits of hijacking increase:

H2a: New hijacking attempts will be more likely following two hijacking attempts that are closely spaced in time.

H2b. New hijacking attempts will be more likely following a series of successful hijackings.

H2c: Compared to those who hijack for other reasons, hijacking attempts by terrorists will be less affected by counter hijacking measures that raise the severity or certainty of punishment.

Consistent with Holden's [18] arguments about contagion, in Hypothesis 2a we predict that the incentives to hijack may manifest externally when prospective hijackers witness the hijacking attempts of others. Such attempts are likely to generate much media attention. Also consistent with Holden's arguments, in Hypothesis 2b, we examine whether successful hijacking attempts affect the hazards of additional attacks. We distinguish these two hypotheses to determine the necessity of the hijackings' success for contagion. Finally, perceived benefits of hijacking could depend on the type of motivation; thus, compared to hijackers who are trying to

[4] Because we have no direct data on actors' perceptions, our research is similar to other macro-level tests of deterrence/rational choice theory (e.g.,[5], [30], [26]) in assuming that potential hijackers' decisions were based at least in part on their knowledge of the probability of success and the costs of failure.

[5] While the FAA definition of success seems justifiable for the time period examined in our hazard models (1931-1985), we note in passing that according to the FAA definition, the hijackings of September 11, 2001 would have been classified as unsuccessful—even though the immediate goals of the hijackers in this case were apparently realized.

extort money for personal gain, hijackers wishing to make ideological statements through terrorist actions might be less affected by the risks of apprehension and punishment. Although we cannot directly measure the differential motivation for terrorists to hijack an aircraft, in H2c we hypothesize that compared to those who hijack for monetary gain or for transportation to another country (most often Cuba), terrorist hijackers will be less affected by traditional counter terrorist measures that increase the certainty or severity of punishment.

Our final hypothesis is derived from the cost-related portion of rational choice theory:

H3: A new hijacking attempt will be less likely after harsher punishments are announced.

This hypothesis is based on the deterrence/rational choice expectation that sanction severity will reduce criminal activity.

5 Developing an Aerial Hijacking Database

To examine long-term trends in global hijacking we obtained data on 1,101 aerial hijackings (285 originated from U.S. airports and 816 originated from non-U.S. airports) from 1931 to 2003. Much of the data from 1931 to 1985 are from the FAA and include 268 hijackings that originated from U.S. airports and 560 hijackings that originated elsewhere. We updated the original FAA data base with published FAA reports through 1999[6] and collected hijacking event data from 2000 to 2003 from the aviation safety network (http://aviation-safety.net/index.shtml). We then supplemented the resulting FAA data base with 39 additional hijacking cases identified from publicly available data from RAND (http://www.db.mipt. org/index.cfm) and from our own newly created data base on terrorist events [24], 2004). Data for 828 cases from 1931 to 1985 are especially complete, including whether the event was successful, as well as information on city/country of origin/destination, number of passengers, and weapons used.

In order to distinguish terrorist hijackings from other hijackings, we relied on the RAND data and our own terrorism data base. For the purposes of this study, we defined terrorist hijackings as those that involve *the threatened or actual use of illegal force and violence to attain a political, economic, religious or social goal through fear, coercion or intimidation* [24]. For example, an incident identified in our data base as a terrorist hijacking happened on January 31, 1980 when three Shi'ite Moslems hijacked an Air France airliner with pistols and a grenade over Beirut, Lebanon to draw attention to the disappearance of spiritual leader Iman Musa Sadr in

[6] Until the mid-1980s FAA criminal acts data were publicly and freely available in hard copy format. However, after the publication of a 1986 report that contained an impressive amount of detailed information (much of which is used in this study), the FAA criminal acts reports covering the period between 1986 and 2000 contained far less detailed information and are currently available for a fee from the National Technical Information Service (NTIS). Since the last published report (2003), which listed the cutoff date for criminal acts against civil aviation as December 31, 2000, we were unable to identify any publicly available reports from the NTIS or FAA regarding criminal acts against aviation.

Libya [24].[7] The resulting composite data base includes information on all known aerial hijackings from 1931 to 2003 and more detailed information on hijackers, their affiliations, and their main purpose for hijacking an aircraft from 1931 to 1985 (828 cases). Because our analysis includes an independent variable that incorporates information on two previous incidents (described below) we drop the first two airline hijackings (1931 and 1947) leaving us with 826 cases for the quantitative analysis.

6 Aerial Hijacking and Counter Hijacking Measures, 1947 to 2003

Figure 1 shows trends in total hijackings of flights originating in the U.S. and outside of the U.S. Because our data include no hijackings between 1931 and 1946, we limit Figure 1 to hijackings between 1947 and 2003. According to Figure 1, the total number of U.S. and non-U.S. skyjackings never rose above ten per year until the mid-1960s. In fact, our data show no non-U.S. hijackings for the years 1954, 1955 and 1957 and following the first U.S. hijacking in 1961. There were no reported U.S. hijackings in the years 1963 and 1966. But the total number of U.S. and non-U.S. skyjackings rose dramatically after the mid-1960s. Both U.S. and non-U.S. annual hijackings first exceeded ten in 1968 (20 U.S.; 15 non-U.S.). Figure 1 shows an especially sharp rise in both U.S. and non U.S. hijackings from 1968 to 1973. The highest number of hijackings of flights originating in the U.S. was in 1969 (39) and for flights originating in other countries it was 1970 (64).

Following 1973, there was a sizeable decline in hijackings, especially for flights originating in the United States. In fact, from a high point of 39 hijackings in 1969, U.S. hijacking counts declined to only two cases in 1973. Declines in non-U.S. hijackings were less dramatic, but still substantial. From a record high of 64 hijackings in 1970, non-U.S. hijackings dropped to a total of 14 in 1975. Following the early 1970s, non-U.S. hijackings experienced several smaller increases, with high points in 1990 (39), 1985 and 1993 (34 and 31 respectively), 1977 (28), and 2000 (21). Compared to the non-U.S. hijacking trends, the U.S. experienced lower post-1973 total hijackings, with high points in 1983 (21) and 1980 (20). However, there were no recorded hijackings of U.S.-origin flights from 1992 until an unsuccessful hijacking attempt by a lone offender in 2000.[8] After this event, the next incident was the deadly attack of September 11, 2001 involving four hijacked aircraft.

Not surprisingly, as aerial hijackings increased in the 1960s and 1970s, policy makers in the U.S. and elsewhere responded with a growing number of counter-hijacking strategies. Although there were also many less important policy efforts, we identified nine major policy changes aimed at reducing aerial hijackings from 1947 to 1986 which happened to have fallen within a four year period: (1) in February 1969, the FAA authorized physical searches of passengers to be conducted at the airlines'

[7] We had separate research assistants identify the terrorism cases independently. The correlation in selection of terrorism cases across assistants was 0.91. We reexamined disagreements and resolved discrepancies.

[8] The lone U.S. hijacking in 2000 occurred on July 27th and involved an individual who boarded a plane at Kennedy Airport in New York City with the intent of hijacking it, but was thwarted before the plane left the ground.

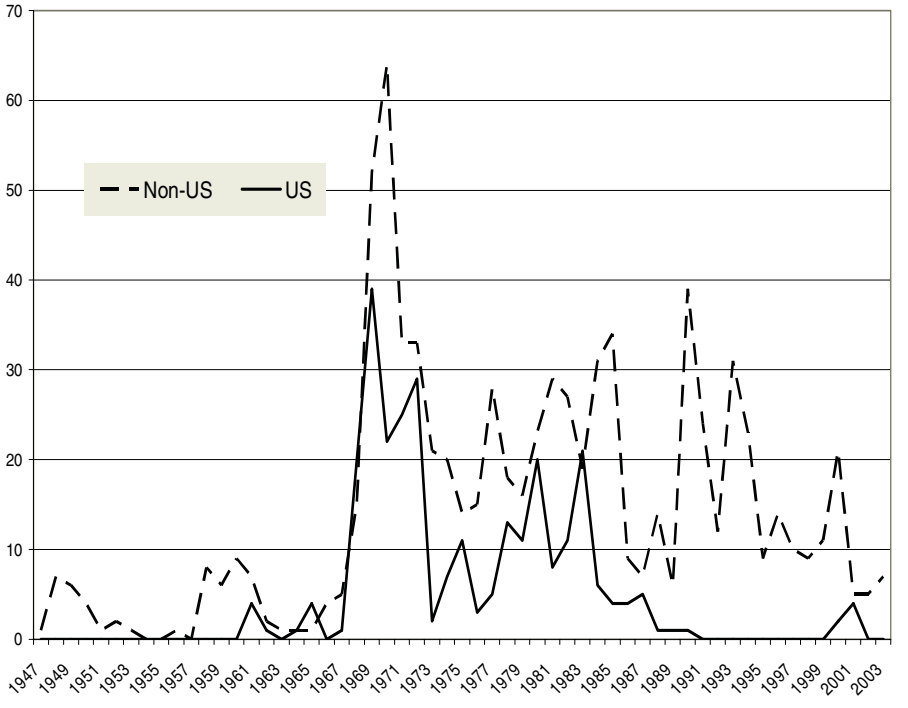


Fig. 1. U.S. and Non-U.S. Hijacking Attempts, 1947-2003

discretion; (2) beginning in October 1969, three major U.S. airlines implemented an FAA system that used weapons detection devices for passengers that fit a behavioral profile of past hijackers; (3) on November 1, 1969, Cuba (for the first time) returned six American hijackers to the U.S. to face criminal charges; (4) in October 1970, the Cuban government made skyjacking a crime; (5) in January 1972, the FAA issued rules ordering tighter screening of all air passengers and baggage using one or more suggested methods: "behavioral profile, magnetometer, identification check, physical search" [32: 6]; (6) in August 1972, the FAA mandated that airlines refuse to board any passengers who fit a hijacking behavioral profile before they were physically or electronically searched; (7) on January 5, 1973, metal detectors were installed in U.S. airports and although the dates and times differ substantially, similar devices were gradually introduced to major airports around the world; (8) on February 3, 1973, the U.S. and Cuba signed a Swedish-brokered agreement that defined hijacking as a criminal act in both nations and promised to either return hijackers or put them on trial; and (9) on February 5, 1973 the FAA required that local law enforcement officers be stationed at all passenger check points during boarding periods.

7 Estimating the Hazards of Aerial Hijacking

To test our hypotheses, we use Cox proportional hazard models to estimate the impact of flight context, hijacking motives, and policy intervention on the hazard of hijacking

attempts.[9] Previous applications of the Cox model estimate the hazard of a single event using many observations. Often this event can only occur once, such as death. However, more recently, analysts have extended the model to account for repeated events such as childhood infectious diseases [22] or purchases [4]. Here, we further extend the approach, by examining multiple repeated events (hijacking attempts) for the entire world. Our approach also differs from a conventional time-series analysis because instead of imposing fixed time intervals, we let the time between hijacking attempts determine each interval between events. This method allows us to use the exact date of implementation to estimate policy effects. Also, we can exploit the temporal spacing between events to account for changes in patterns that may have otherwise been collapsed into aggregate units for time-series analysis.

The main methodological concern with repeated events in hazard modeling is that the timing to the current event might depend upon the timing of other events experienced by the same observation [1]. Because we condition our models on the entire sample, this is not a problem here. However, the timing to the next hijacking attempt might depend on the timing to the current attempt, thus biasing our standard errors downward. To determine whether this is a problem, Allison suggests testing for dependency by including the length of the previous "spell" (in this case, the length of time between hijackings) as a covariate in the model for the current spell. As noted below, we already include in our models the time since the previous hijacking to test for contagion (H2a).[10] To assure that after controlling for the previous spell length our observations are independent, we tested the significance of the spell lagged twice and found its effect to be null ($p > 0.10$). Thus, we are confident that the standard errors produced by our method are unbiased.

To test the hypotheses outlined above, we estimate models separately for six subsets of hijacking attempts: (1) total, (2) those originating in the U.S., (3) those originating outside of the U.S., (4) those diverted to Cuba, (5) terrorist-related, and (6) non-terrorist-related. We use the following specification for the hazard models in the analysis:

$$h(\text{Next Attempt}) = \lambda_0(\text{Next Attempt}) \exp(\beta_1 \text{Policies} + \beta_2 \text{MajorPurpose} + \beta_3 \text{Context}) \quad (1)$$

We estimate the hazard of a new hijacking attempt (measured by the number of days until the next event) as a function of an unspecified baseline hazard function and other risk or protective variables represented by the vectors *Policies*, *Major Purpose*, and *Context*, which reflect our hypotheses and a set of control variables.

In Figure 2 we show the temporal ordering of the anti-hijacking policies described above. The most striking feature of Figure 2 is that while we are interested in policy changes since 1947, all nine major policy interventions happened over only a four year period: February 1969 through February 1973. This, of course, makes it more challenging to evaluate the individual impact of specific policies.

---

[9] We use the exact method to resolve ties in survival time [1]. This method assumes that the underlying distribution of events is continuous rather than discrete. This is the most appropriate strategy because airline hijacking can occur at any time.

[10] Because the previous spell is already included in the model, even if it is significantly related to the current spell, by the Markov property, the model conditions on all sources of dependence, thus any two contiguous observations are independent [49].

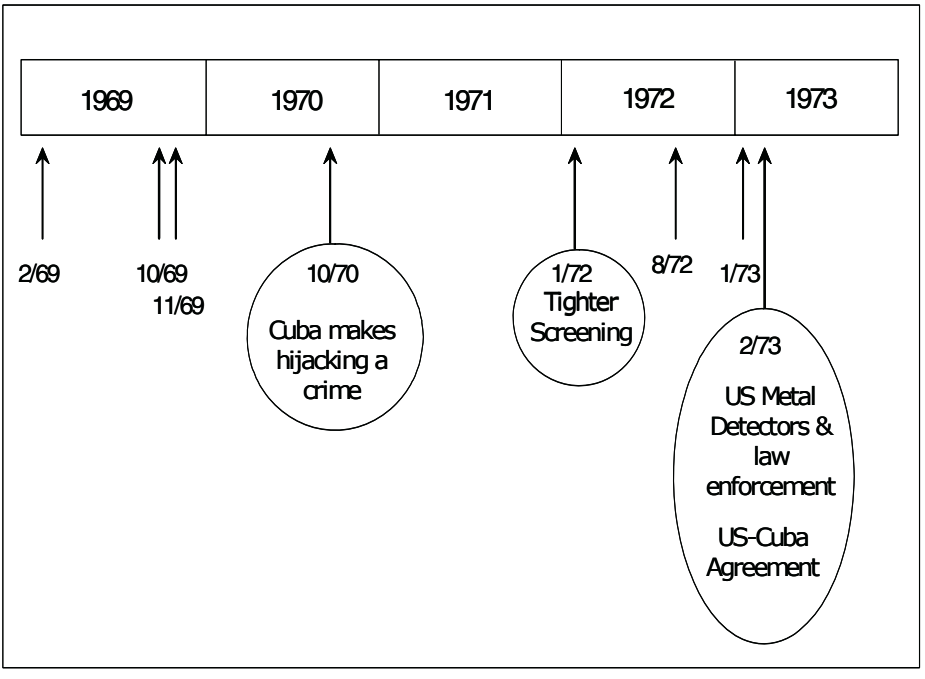


Fig. 2. Anti-Hijacking Policies

Based on the temporal ordering of the anti-hijacking policies, we identified three strategic policy dates.[11] First, we chose October 31, 1970, the date that Cuba made hijacking a crime (*Cuba Crime*), because the policy goal was specific to Cuban hijacking and therefore gives us a direct way to examine its effects: if there is truly a policy impact as a result of this law it should only have a significant effect in the model that uses data from hijackings diverted to Cuba.[12] Second, we chose the FAA policy of ordering *Tighter Screening* of all aircraft passengers and baggage, enacted on January 31, 1972. This policy intervention is strategic for two reasons. First, because it was imposed by the FAA only for flights from U.S. airports, any policy effect should be limited to the United States. And second, although several policy interventions are clustered closely during this period, tighter screening was implemented more than a year after the prior policy intervention, thus reducing the chance of simultaneous effects of the two interventions.

[11] Five cases in the data base were missing information on specific dates. For three of these cases, month of the hijacking was available and we estimated the dates by using the last day of the month (February 1931, August 1966, and November 1978). This assures that policy interventions occurred prior to the event. For the remaining two cases we knew only that the case occurred in the "Fall" and we therefore set the dates equal to October 31 of the appropriate year—the middle of the Fall season.

[12] To be sure, this measure could also be interpreted as increasing the degree of certainty [7], but we chose to retain its focus as one of severity because of its reliance on the administration and degree of punishment.

Finally, we selected three major policies that were all implemented in February 1973 (labeled *Metal Detectors*). While these policies were implemented at the same time, we might expect them to have somewhat different effects on the sub-samples being analyzed. Metal detectors should have an especially strong impact on flights taking off from U.S. airports—because these policies were first implemented in the United States [14]. But at the same time, these policies spread fairly quickly to other highly industrialized nations and were gradually adopted by most other nations of the world. By contrast, the agreement between Cuba and the U.S. should only affect Cuba-U.S. flights.

We distinguish between three major hijacking purposes in our models: *Terrorism*, *Extortion*, and *Transportation to Cuba*. We classified hijackings as terrorist when the hijackers made political, economic, religious or social demands. We classified as extortion all cases in which the hijackers demanded money. Finally, if the hijackers used the flight to get to Cuba, we classified the case as transportation. Altogether, we classified 51.8 percent of the cases as having at least one of these three purposes. In 35 cases (4.2%) we classified a single event in two of these categories and in two cases (0.2%) we classified a single event in three of these categories. One of the cases included in all three categories happened on November 10, 1972 when three members of the Black Panther Party hijacked (terrorist) a Southern Airways jet to Havana, Cuba (transportation to Cuba) and demanded $2 million in ransom (extortion; [43]).

We include five variables to measure the context of each incident: *Last Hijack*, *Success Density*, *Private Flight*, *U.S. Origin*, and *Year*. Last hijack measures the number of days since the previous hijacking attempt. Using the current and two previous flights, we calculated a success density measure as the ratio of the proportion of those flights that were successful over the number of months spanning the three events.[13] Thus, a large success density indicates that most events were successful over a relatively short time period. We also include indicators of whether planes were privately owned, whether flights originated from U.S. airports, and the year of each incident.

8 Predicting the Hazard of Hijacking Attempts

Table 1 shows the coefficient estimates and standard errors for the hazard models for total incidents, U.S. originated incidents, non-U.S. incidents, Cuba diverted incidents, terrorist-related incidents, and non-terrorist-related incidents. In each model, the dependent variable is the number of days until the next event. A positive coefficient suggests that the variable increases the hazard of another hijacking attempt in a shorter time while a negative value decreases the hazard of another hijacking attempt.

[13] We operationalize the success density with the following function, $\frac{P(\text{successful of last three flights})}{(\text{event date}_t - \text{event date}_{t-2})/365}$., thus incorporating the distribution of success among the three incidents with the temporal range. We initially calculated this measure using 3, 5, 7, 10, 15, 20, 30 and 40 incidents. We decided to report results for three incidents here because this strategy retained the most observations while providing the strongest results. Also, there is no clear theoretical reason to expect airline hijackings before the prior two to highly influence future decisions.

Table 1. Coefficients and Standard Errors for Cox Proportional Hazard Models

	All	US Origin	Non-US Origin	Cuba Diverted	Terrorist	Non-Terrorist
	n=826	n=265	n=556	n=272	n=123	n=700
Policies:						
Cuban Crime	-0.095	-0.500*	0.233	-0.421*	0.782	-0.145
	0.147	0.232	0.199	0.219	0.569	0.155
Tighter Screening	-0.084	0.686	-0.505*	0.070	-1.020	-0.016
	0.184	0.311	0.246	0.381	0.637	0.198
Metal Detectors	-0.949**	-1.598**	-0.653**	-0.967*	-0.644	-0.996**
	0.166	0.371	0.204	0.434	0.410	0.184
Major Purpose:						
Terrorism	0.146	0.359	0.163	0.311		
	0.104	0.470	0.109	0.246		
Extortion	0.147	0.142	0.052	0.178	-0.239	0.218
	0.139	0.260	0.176	0.412	0.331	0.154
Cuba	0.171*	0.086	0.287**		0.439	0.141
	0.092	0.148	0.119		0.275	0.099
Context:						
Last Hijack	-0.004**	-0.003**	-0.003**	-0.002*	0.001	-0.004**
	0.001	0.001	0.001	0.001	0.001	0.001
Success Density	0.002**	0.002	0.002*	0.001	0.000	0.002*
	0.001	0.001	0.001	0.001	0.001	0.001
Private Flight	-0.098*	-0.037	0.009	-0.130	0.517	-0.107
	0.119	0.193	0.161	0.238	1.152	0.120
US Origin	0.050			0.029	0.533	0.052
	0.087			0.137	0.532	0.089
Year	0.078**	0.074**	0.075**	0.091**	0.041	0.081**
	0.010	0.028	0.011	0.031	0.031	0.010

- * = $p \leq 0.05$ and ** = $p \leq 0.01$, all one tailed tests.

Hypothesis 1 predicts that the hazard of hijacking attempts will decrease following the adoption of measures that increase the certainty of apprehension. We examined the effect of two certainty-based measures: tighter U.S. security screening adopted in January 1972 and the adoption of metal detectors and enhanced U.S. airport security adopted in February 1973. The results show partial support for the certainty of apprehension hypothesis. Consistent with hypothesis 1, the hazard of hijacking in the U.S. origin model significantly dropped following the adoption of metal detectors and other target hardening policies in 1973. In fact, the 1973 policies were the only interventions that significantly reduced hijacking hazards in all models, except those

limited to terrorism.[14] In contrast, increasing certainty of apprehension through tighter U.S. screening protocols introduced in January 1972 reduced the hazard of non-U.S. origin flights but failed to do so for U.S. flights. In fact, there was a short-term *increase* in the hazard of U.S. origin hijacking attempts following the implementation of the 1972 screening policy.

Our next set of hypotheses examines the impact of perceived benefits of hijacking on the hazard of new hijacking attempts. Hypothesis 2a is a test of the contagion hypothesis that new hijacking attempts will be more likely after two hijackings that have occurred in a relatively short time period. In support, Table 1 shows that that the hazard of another hijacking increases significantly if the current and previous hijackings were attempted in close temporal proximity to one another. Similarly, in Hypothesis 2b we examine whether a series of successful hijackings increases the likelihood of additional hijackings. In support, Table 1 shows that if the three most recent events were primarily successful and close together, the hazard of a new hijacking attempt increased for the full sample as well as for non-U.S. and non-terrorist hijackings. Thus, both frequently attempted and successful hijackings increase the likelihood of additional hijackings.

Our other benefits-related hypothesis (H2c) predicts that compared to those who hijack for other reasons, those with terrorist-related motives will be affected less by the counter terrorist measures being examined here. The results are shown in the last two columns of Table 1. The null associations of the coefficients for tighter screening and the Cuban crime policy neither support nor reject the hypothesis because neither policy significantly impacted terrorist or non-terrorist related hijackings. By contrast, the 1973 policies (Metal Detectors) are significantly related to non-terrorist hijackings while null for terrorist events thus supporting the hypothesis. However, we should note that the differences in magnitude between the coefficient in the terrorism model (-0.644) and the non-terrorism model (-0.996) suggest only weak support for the hypothesis (z=0.78).

Hypothesis 3 predicts that as the severity of punishment increases, the hazard of a new hijacking will decline. We test this hypothesis by including a variable that indicates when it became a crime in Cuba to hijack a plane. Indeed, the hazard of hijacking decreased substantially after this policy was enacted for both Cuban flights and for U.S. origin flights. The latter finding makes sense because 57.5 percent of flights diverted to Cuba originated in the United States. Note also the null impact of this policy on other types of hijackings not closely related to Cuban flights.

9 Variables Associated with Hijacking Success

The significant effect of our success density measure strongly suggests that a successful hijacking attempt (as defined by the FAA) will likely lead to more attempts. Yet, little is known about the characteristics of successful hijackings. How closely do prospective hijackers' perceptions of the likelihood of success correspond to their actual likelihood of success? In the next part of the analysis, we examine the

[14] To be sure that this result is specific to the date, we re-estimated the model replacing February 5, 1973 with later dates. None of these re-estimates were significant.

determinants of successful hijackings. Our detailed hijacking data allows us to track trends in successful and non successful U.S. and non U.S. hijackings from 1947 to 1985.[15] Figure 3 shows that while the total numbers of successful hijackings originating in U.S. and non-U.S. airports are highly correlated until the 1970s, they diverge somewhat thereafter, with successful hijackings of U.S. origin flights declining more rapidly than successful hijackings of non-U.S. flights for most years after 1973 (the exceptions are 1975, 1980 and 1983). And as we have seen above, there are no hijackings originating in the U.S. from 1991 through 1999. In short, both the total number of hijackings and the total number of successful hijackings falls off more sharply for the U.S. than for other countries following 1972.

In Table 2 we summarize the effects on hijack success of variables measuring *Policies*, *Major Purpose*, and *Context*. All variables are constructed in the same way as those in Table 1, except that instead of using the success density measure, we

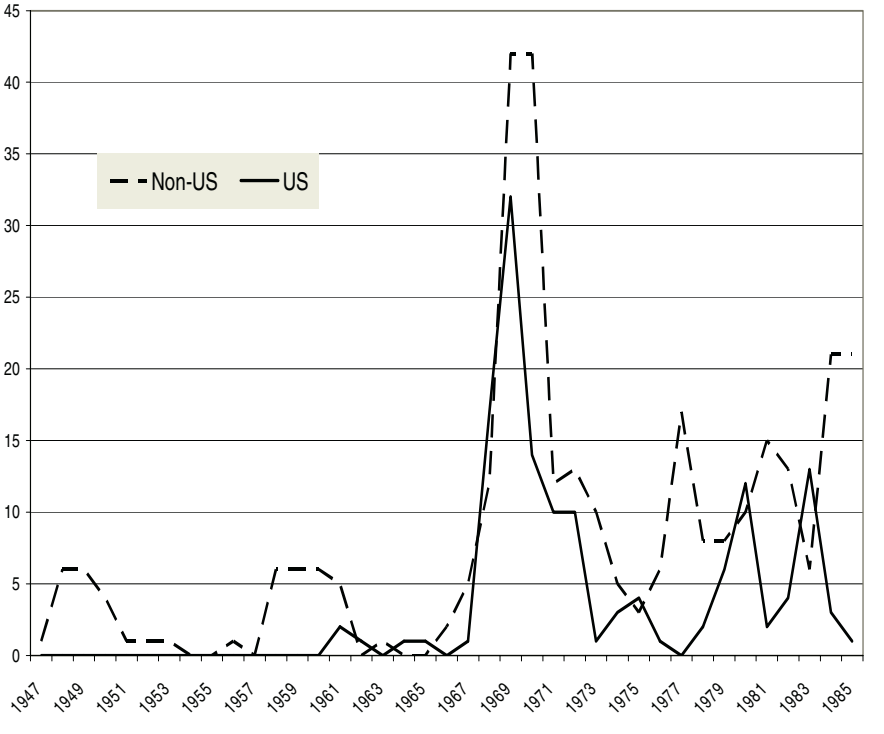


Fig. 3. U.S. and Non-U.S. Successful Hijackings, 1946-1985. A success is defined when hijackers gain control of the plane and reach their destination, whether by landing or by parachute escape, and are not immediately arrested or killed on landing. Unsuccessful hijackings are those in which hijackers attempt but fail to take control of an aircraft [15]

[15] The first incident in 1931 was excluded because two of the independent variables measure the previous incident.

Table 2. Odds Ratios and Standard Errors for Logistic Models Predicting Success

	All	US Origin	Non-US Origin	Cuba Diverted	Terrorist	Non-Terrorist
	n=827	n=267	n=559	n=273	n=119	n=702
Policies:						
Cuban Crime	0.286**	0.238*	0.254**	0.157**	1.112	0.251**
	0.091	0.132	0.105	0.077	1.406	0.085
Tighter Screening	1.528	3.782	1.143	3.598	0.554	1.563
	0.643	2.987	0.607	3.638	0.763	0.753
Metal Detectors	1.021	0.144*	1.506	0.081*	0.691	1.021
	0.379	0.130	0.659	0.088	0.619	0.447
Major Purpose:						
Terrorism			3.369**	6.157*		
	3.604**					
	0.852		0.820	4.830		
Extortion		0.728	0.378*	0.171	2.871	0.223**
	0.418**					
	0.140	0.488	0.152	0.192	2.444	0.101
Cuba		14.044**	1.843*		2.661	3.648**
	3.623**					
	0.755	5.919	0.482		1.862	0.810
Context:						
Last Hijack	1.004	1.005	1.001	0.999	1.000	1.004
	0.003	0.003	0.001	0.001	0.001	0.003
Last Success	1.226	0.973	1.064	0.463*	0.961	1.061
	0.198	0.321	0.205	0.168	0.443	0.191
Private Flight		7.344**	2.520**	2.522		2.961**
	2.813**					
	0.758	4.117	0.902	1.684		0.814
US Origin	0.660*			1.642		0.650*
	0.129			0.538		0.132
Year	0.992	1.093	0.981	1.149*	1.048	0.994
	0.020	0.075	0.021	0.076	0.069	0.021

* = $p \leq 0.05$ and ** = $p \leq 0.01$, all one tailed tests.

include an indicator of whether the previous flight was successful (*Last Success*). Because Table 2 reports odd ratios, all coefficients less than one indicate a negative effect and all coefficients greater than one indicate a positive effect.

Turning first to the policy results, perhaps most striking is that all hijackings except terrorist-motivated events were less likely to succeed following the passage of a Cuban law making hijacking a crime. The magnitudes of these results are quite large. For example, the ratio for Cuban flights suggests that the odds that an attempted hijacking to Cuba was successful dropped by 84.3 percent (100-15.7) after the policy was implemented. Table 2 also shows that following the implementation of metal

detectors and other interventions in 1973 there was a significant decline in the likelihood of success for both hijackings originating in the U.S. and those diverted to Cuba. Again, the magnitudes of these reductions are quite large. Finally, the results show that the tighter screening policy had no effect on hijacking success.

The next series of findings relate to the major purpose of the hijackers. Because there were only five cases of terrorism-related hijacking that originated in the U.S. and four of these were successful, we dropped the U.S. origin model from this part of the analysis.[16] Table 2 shows that compared to other flights, flights hijacked by terrorists are much more likely to be successful for total, non-U.S., and Cuban diverted incidents. Conversely, flights motivated by extortion were much less likely to be successful for total flights, non-U.S. origin flights and non-terrorism related flights. Flights diverted to Cuba were more likely than other flights to be successful in the analysis of total incidents, U.S. origin incidents, non-U.S. origin incidents, and non terrorist incidents. In fact, the odds of a successful hijacking originating in the U.S. are more than 14 times higher if the purpose of the hijacking was transportation to Cuba. This last finding likely reflects the long-standing policy of not offering physical resistance to hijackers who had forced aircraft to fly to Cuba on the assumption that this response was least likely to result in casualties [18:881], [37].

Finally, turning to the findings related to the context of the flight we see that a previous success only produces significant reductions in the success of Cuban flights. The odds of another successful Cuban hijacking after a successful Cuban hijacking are less than half of those that follow unsuccessful attempts. This finding might be due to the fact that a successful hijacking produces greater vigilance on the part of authorities, making subsequent successful attempts less likely—especially immediately after the successful hijacking. However, if this is the case, it is unclear why this effect is limited to the Cuban flights.

Table 2 also shows that the likelihood of success is unrelated to the time that has passed since the last attempted hijacking. While our analysis of the probability of new hijackings (Table 1) showed that private planes were no more likely to be hijacked than commercial aircraft, the results in Table 2 show that when private planes are hijacked, the hijacking is more likely to be successful—for all flights except Cuban.[17] Finally, flights originating from U.S. airports faced a lower probability of success both for the full sample and for the non terrorist cases.

10 Discussion and Conclusions

Based on a rational choice perspective we developed a set of five hypotheses about the likelihood of hijacking attempts and used data from the FAA, RAND and a newly developed terrorist events data base to determine whether aerial hijacking attempts respond to situations and policies expected to affect the probability of hijacking success and its perceived benefits and costs. Our results support three main conclusions. First, and most policy relevant, we found considerable support for the

[16] Without more cases and variation we are unable to accurately detect an effect for US origin.

[17] Private flight is omitted from the terrorism model because it predicts failure completely: there was only one terrorist hijacking of a private flight and it failed.

conclusion that new hijacking attempts are less likely to be undertaken when the certainty of apprehension or severity of punishment increases. But in this regard one of the certainty measures we examined (metal detectors and increased enforcement) had significant effects while another certainty measure (tighter baggage and customer screening) did not. Perhaps the implementation of metal detectors and increased law enforcement at passenger check points was simply a more tangible and identifiable intervention than the tighter screening policies introduced 18 months earlier. However, the fact that these policies were implemented closely in time also raises the possibility that it was the accumulation of policies as opposed to one specific policy that made the difference. The drop in the hazard of hijacking attempts after the Cuban crime policy was implemented strongly suggests that the threat of sanctions was useful here.

Second, we did find considerable support for a contagion view of hijacking: the rate of hijackings significantly increased after two relatively close hijacking attempts and following a series of successful hijackings.

Finally, we found that the counter-hijacking policies examined had no impact on the hazard of hijacking attempts whose main purpose was terrorism. By contrast, we found that the adoption of metal detectors and increased police surveillance significantly reduced the hazard of non-terrorist related hijackings. Moreover, tighter screening significantly reduced the hijacking hazard of non-U.S. flights and a policy making hijacking a crime significantly reduced hijackings to Cuba. Similarly, the policies examined had no significant impact on the success of terrorist-related hijackings. But in contrast, metal detectors and increased police surveillance significantly reduced the likelihood that U.S. origin and Cuba diverted flights would be successful and a policy criminalizing hijacking in Cuba significantly reduced the likelihood of success of all non-terrorist related flights.

While we have assembled the most comprehensive data base on international hijackings of which we are aware, our study has several limitations. Like many earlier macro-level tests of the deterrence/rational choice perspective, we had no perceptual data that would have allowed us to examine the individual motivations of hijackers. Although data on individual motivations from hijackers or would-be hijackers appear especially difficult to collect, such information would allow researchers to better understand how hijackers actually interpret policies and sanctions. Second, because most of the major anti-hijacking interventions happened very close in time, it was difficult to separate out independent effects. Thus, our analysis of the three policies passed in January and February of 1973 had to be combined. Third, although our data base includes many of the variables shown by prior research to be associated with aerial hijackings, it is certainly plausible that other variables not available to us (and likely unavailable elsewhere) would be useful to have. This is especially the case regarding our measure of benefits. For example, a successful hijacking could heighten a group's popularity and increase its membership or could increase the chances that political prisoners would be released.

With these limitations in mind, it would be useful to briefly consider this study's implications for theory, future research, and policy. With regard to theory, the results of the current study provide mixed evidence regarding the effectiveness of deterrence-/rational choice-based policies. The certainty-based 1973 metal detector and police surveillance policies appear more effective than the 1972 tighter screening policy.

There was evidence that the Cuba crime policy was effective in reducing Cuba-related hijackings. Taken together, these findings support Nagin's [30] conclusion that some deterrence efforts do work. At the same time, they also suggest that there is considerable variation in the effectiveness of the hijacking counter measures that were implemented.

Our results also suggest that policy interventions had less impact on the success of terrorist hijackings than on the success of other hijacking types. In fact, none of the three policies examined were significantly related to the attempts or success of terrorism-related hijackings. Perhaps the rational choice perspective is not the most appropriate theoretical framework for understanding terrorist-motivated hijackings. Alternatively, it could be that the traditional rational choice measures of costs/benefits (such as those used in this study) are less meaningful for terrorist-related hijackings. These results may suggest that we need different measures of costs and benefits in the study of terrorist-motivated hijackings.

While this study is an initial attempt at applying the deterrence/rational choice framework to aerial hijacking using data that has heretofore been unexamined, much remains to be documented and understood. We envision at least three additional projects. First, because aerial hijacking occurs over space and time, it is important to examine the specific sources of this variation. Perhaps certain countries or airlines are more hijack-prone than others at various times. Second, we need to better understand the motivations of terrorists. Are they really unrelated to the underlying assumptions of deterrence/rational choice so that punishment-based policies simply will not resonate with them? Given Pogarsky's [42] finding that deterrence is inoperable among a subgroup of 'incorrigible' offenders, it will be interesting to more thoroughly document the motivations across different types of hijackings and hijackers. Finally, because our data were confined to the pre-1986 period, we cannot comment on the many recent efforts (e.g., sky marshals, reinforced cockpit doors) currently employed by the U.S. and other governments at thwarting aerial skyjacking. Research on these policies will be important in order to determine their effectiveness weighed against their costs. Additionally, it is likely that such policies will be effective only to the extent that would-be offenders recognize these efforts and consider them in their decision-making. Because knowledge about the relationship of sanction risk perceptions to policy is virtually non-existent, such information is important for designing effective deterrence policies [29:1, 36-37].

From a policy perspective, our analysis indicates that some certainty- and severity-based interventions were effective at reducing some types of hijacking attempts and lowering the probability of some types of successful hijackings. That some policies are more effective at certain times and places and for certain kinds of acts than others is consistent with the policy implications emanating from situational crime prevention [8], [47], an approach based largely on the assumptions underlying the deterrence/rational choice framework that the motivations associated with certain kinds of crimes are not necessarily the same as the motivations associated with other kinds of crimes. Policy makers need to carefully study and understand the effectiveness of their policies, continue implementing the ones that work, modifying the ones that may work, and abandoning the ones that do not work.

In short, while we find substantial evidence that the deterrence/rational choice framework provides important insight into understanding different types of hijackings

and the probability of success among such hijackings, our understanding of aerial hijacking could clearly benefit from other perspectives [24]. Clearly, much work needs to be done to better understand whether common criminological theories are useful for improving our understanding of terrorist behavior.

References

1. Allison, Paul D. Survival Analysis Using the SAS System A Practical Guide. Cary, NC: SAS Institute Inc. (1995)
2. Andenaes, J. Punishment and Deterrence. Ann Arbor, MI: University of Michigan Press (1974)
3. Becker, G.S. Crime and punishment: An economic approach. Journal of Political Economy (1968) 76:169-217
4. Bijwaard, Govert E., Phillip Hans Franses, and Richard Paap Modeling Purchases as Repeat Events. Econometric Institute Report EI (2003) 2003-45
5. Blumstein, Alfred, Jacqueline Cohen, and Daniel S. Nagin Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates. Washington, DC: National Academy of Sciences (1978)
6. Carroll, John S. A Psychological Approach to Deterrence: A Psychodynamic Study of Deviations in Various Expressions of Sexual Behavior. New York: Citadel Press (1978)
7. Chauncey, R. Deterrence: Certainty, severity, and skyjacking. Criminology (1975) 12:447-473
8. Clarke, R.V. and D.B. Cornish Modeling offenders' decisions: A framework for research and policy. In M. Tonry and N. Morris (Eds.), Crime and Justice: An Annual Review of Research, Volume 6. Chicago: University of Chicago Press (1985)
9. Clarke, R.V. and M. Felson Introduction: Criminology, routine activity, and rational choice. In R.V. Clarke and M. Felson (Eds.), Routine Activity and Rational Choice, Advances in Criminological Theory, Volume 5. New Brunswick, NJ: Transaction (1993)
10. Cornish, Derek B. and Ronald V. Clarke, eds. The Reasoning Criminal. New York: Springer-Verlag (1986)
11. Cox, D. Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series B (Methodological) (1972) 34:187-220
12. Dornin, R. D.B. Cooper is gone but his legend lives on. CNN Interactive, Cable News Network, Inc. November 23, 1996 (1996)
13. Ehrlich, Isaac. Participation in illegitimate activities: A theoretical and empirical investigation. Journal of Political Economy (1973) 81: 521-565
14. Enders, W. and T. Sandler The Effectiveness of Anti Terrorism policies: A Vector–Autoregression Intervention Analysis. American Political Science Review (1993) 87: 829-44
15. Federal Aviation Administration (FAA) Aircraft Hijackings and Other Criminal Acts Against Civil Aviation Statistical and Narrative Reports. Washington, DC: FAA, Office of Civil Aviation Security (1983)
16. Gibbs, J. Crime, Punishment, and Deterrence. Amsterdam, The Netherlands: Elsevier Scientific Publishing Co. (1975)
17. Hamblin, R.J., R.B. Jacobsen, and J.L.L. Miller A Mathematical Theory of Social Change. New York: Wiley-Interscience (1973)
18. Holden, R.T. The contagiousness of aircraft hijacking. American Journal of Sociology (1986) 91:874-904

19. Jacobs, Bruce Crack Dealers and Restrictive Deterrence: Identifying Narcs. Criminology (1996) 34:409-431
20. Jeffery, C. R. Crime prevention through environmental design. Beverly Hills, CA: Sage (1971)
21. Karber, Phillip A. Re-constructing global aviation in an era of the civil aircraft as a weapon of destruction. Harvard Journal of Law and Public Policy (2002) 25:781-814
22. Kelly, Patrick J. and Lynette L-Y. Lim Survival analysis for recurrent event data: An application to childhood infectious diseases. Statistics in Medicine (2000) 19:13-33
23. Klepper, Steve and Daniel S. Nagin Tax Compliance and Perceptions of the Risks of Detection and Criminal Prosecution. Law & Society Review (1989) 23:209-240
24. LaFree, Gary, and Laura Dugan *Global Terrorism Database*. University of Maryland (2004)
25. Landes, W.M. An Economic Study of U.S. Aircraft Hijackings, 1961-1976. Journal of Law and Economics (1978) 21:1-31
26. Levitt, Steven D. Testing the Economic Model of Crime: The National Hockey League's Two Referee Experiment. Contributions to Economic Analysis and Policy 1:1-19 (2002)
27. Merari, Ariel Attacks on civil aviation: Trends and lessons. In P. Wilkinson and B.M. Jenkins (eds.), Aviation Terrorism and Security. London: Frank Cass (1999)
28. Minor, W.W. Skyjacking crime control models. Journal of Criminal Law and Criminology (1975) 66:94-105
29. Nagin, Daniel S. Criminal deterrence research at the outset of the twenty-first Century. In M. Tonry (ed.), Crime and Justice: A Review of Research, vol. 23. Chicago: University of Chicago Press (1998)
30. Nagin, Daniel S. General Deterrence: A Review of Empirical Evidence. In A. Blumstein, J. Cohen, and D. Nagin (eds.), Deterrence and Incapacitation: Estimates and Effects of Criminal Sanctions on Crime Rates. Washington, D.C.: National Academy of Sciences (1978)
31. Nagin, D.S., and R. Paternoster Enduring individual differences and rational choice theories of crime. Law and Society Review (1993) 27:467-496
32. National Materials Advisory Board Airline Passenger Security Screening: New Technologies and Implementation Issues. Washington, D.C.: National Academies Press (1996)
33. O'Brien, R.M. Measuring the convergence/divergence of "serious crime" arrest rates for males and females: 1960-1995. Journal of Quantitative Criminology (1999) 15:97-114
34. Paternoster, R. The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues. Justice Quarterly (1987) 4:173-217
35. Paternoster, R. and A. Piquero Reconceptualizing deterrence: An empirical test of personal and vicarious experiences. Journal of Research in Crime and Delinquency (1995) 32: 251-286
36. Paternoster, Raymond and Sally S. Simpson Sanction threats and appeals to morality: testing a rational choice model of corporate crime. Law and Society Review (1996) 30(3):549-583
37. Phillips, D. Skyjack: The Story of Air Piracy. London: Harrap (1973)
38. Piliavin, I., R. Gartner, C. Thorton, and R. Matsueda Crime, deterrence, and rational choice. American Sociological Review (1986) 51:101-119
39. Piquero, A. and R. Paternoster. An application of Stafford and Warr's reconceptualization of deterrence to drinking and driving. Journal of Research in Crime and Delinquency (1998) 35:3-39

40. Piquero, A.R. and G. Pogarsky Beyond Stafford and Warr's reconceptualization of deterrence: Personal and vicarious experiences, impulsivity, and offending behavior. Journal of Research in Crime and Delinquency (2002) 39:153-186
41. Piquero, Alex R. and Stephen G. Tibbetts offenders' decision making: toward a more complete model of rational offending. Justice Quarterly (1976) 13(3):481-510
42. Pogarsky, Greg Identifying "deterrable" offenders: Implications for research on deterrence. Justice Quarterly (2002) 19:431-452
43. RAND Black Panthers attack airlines and airports target. MIPT Terrorism Knowledge Base, MIPT (National Memorial Institute for the Prevention of Terrorism) Chronology Data (2001) 1968-1997
44. Rich, E. Flying Scared. New York: Stein & Day (1972)
45. Ross, Laurence and Gary LaFree Deterrence in Criminology and Social Policy. In Behavioral and Social Science: Fifty Years of Discovery, edited by N.J. Smelser and D.R. Gerstein. Washington DC: National Academy Press (1986)
46. Simpson, Sally S., Nicole L. Piquero, and Raymond Paternoster Exploring the Micro-Macro Link in Corporate Crime Research. In Peter Bamberger William J. Sonnenstuhl (Eds.), Research in the Sociology of Organizations. JAI Press: Greenwich, CT (1998)
47. Smith, Martha J. and Derek B. Cornish Theory for Practice in Situational Crime Prevention. Crime Prevention Studies, Volume 16. New York: Criminal Justice Press (2004)
48. Stafford, M. and M. Warr A reconceptualization of general and specific deterrence. Journal of Research in Crime and Delinquency (1993) 30:123-135
49. Whittaker, J. Graphical Models in Applied Multivariate Statistics. New York: Wiley (1990)
50. Wright, Richard T. and Scott H. Decker Burglars on the Job: Streetlife and Residential Breakins. Boston, MA:Northeastern University Press (1994)
51. Wright, Richard T. and Scott H. Decker Armed Robbers in Action:Stickups and Street Culture. Boston, MA:Northeastern University Press (1997)
52. Zimring, F. and G. Hawkins Deterrence. Chicago: University of Chicago Press (1973)

Analysis of Three Intrusion Detection System Benchmark Datasets Using Machine Learning Algorithms

H. Güneş Kayacik and Nur Zincir-Heywood

Dalhousie University, Faculty of Computer Science,
6050 University Avenue, Halifax, Nova Scotia. B3H 1W5, Canada
{kayacik, zincir}@cs.dal.ca

Abstract. In this paper, we employed two machine learning algorithms – namely, a clustering and a neural network algorithm – to analyze the network traffic recorded from three sources. Of the three sources, two of the traffic sources were synthetic, which means the traffic was generated in a controlled environment for intrusion detection benchmarking. The main objective of the analysis is to determine the differences between synthetic and real-world traffic, however the analysis methodology detailed in this paper can be employed for general network analysis purposes. Moreover the framework, which we employed to generate one of the two synthetic traffic sources, is briefly discussed.

1 Introduction

Along with benefits, the Internet also created numerous ways to compromise the stability and security of the systems connected to it. Although static defense mechanisms such as firewalls and software updates can provide a reasonable level of security, more dynamic mechanisms should also be utilized. Examples of such dynamic mechanisms are intrusion detection systems and network analyzers. The main difference between intrusion detection and network analysis is the former aims to achieve the specific goal of detecting attacks whereas the latter aims to determine the changing trends in computer networks and connected systems. Therefore network analysis is a generic tool, which helps system administrators to discover what is happening on their networks.

In this paper, we employed two machine learning algorithms for network analysis. The first method is a clustering algorithm, which aims to find the natural groupings in the dataset. The second method is based on a neural network algorithm called Self-Organizing Maps (SOMs) where topological models are built for the given dataset. Both methods act like network analyzers to discover similarities between the datasets. The objective of the analysis is to determine the robustness of our synthetic dataset in terms of its similarity to real-world datasets. However, the analysis methods could be employed for wide variety of purposes such as anomaly detection, network trend analysis and fault detection.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 362 – 367, 2005.

2 Framework for Generating Synthetic Traffic

Because of the privacy concerns and the extensive storage requirements, employing the data captured from a live network is not practical. Moreover, the captured data itself reveals very little information without further analysis. Having recognized the need for synthesizing data for training and testing intrusion detection systems, we developed a framework [1], which can: (1) Develop models of normal behavior from its observations. (2) Generate synthetic activity based on the developed analytic models. Our synthetic traffic generation framework focuses on modeling hypertext transfer protocol (HTTP). HTTP is selected because considerable amount of the Internet traffic is made up of HTTP traffic [2]. Our framework is implemented for HTTP however it could easily be tailored for many other protocols, which involve file transfers with variable user think times.

The framework [1] is divided into four tasks. (1) HTTP is not a session-based protocol; that is to say web server logs only contain the records of requested documents without any information about where a session ends. Session concept is imperative because it acts as a placeholder to group the activities of a user. The first task is to construct sessions from web server logs. (2) The second task is to develop session models by employing first order discrete Markov models [3]. Each web page is considered as a state and Markov models are developed by maintaining the frequencies of web page transitions. Models are developed for page transitions as well as transition delays (user think times). (3) Developed models are employed to generate synthetic sessions. A synthetic session starts from a special purpose start state and transitions (i.e. web page requests) are appended to the session stochastically until the special purpose end state is reached. This way, the model can simulate sessions with arbitrary lengths. (4) Synthetic sessions are processed and web page requests are passed to the web browser installed on the machine to generate the network traffic.

3 Analysis and Results

In order to determine the robustness of the generated synthetic dataset, we employed a SOM based intrusion detection system [4] and a clustering algorithm [5] as network traffic analyzers to compare the characteristics of synthetic and real-world datasets. The objective of the analysis is to determine whether our synthetic dataset shows improvements over the benchmark dataset, however similar analysis can be performed to discover the changing traffic trends in a network.

Data Description

In case of all three data sources, network traffic is recorded in *tcpdump* format. We compared our synthetic data with the standard intrusion detection system benchmark dataset, namely KDD 99 dataset, which is based on Lincoln Lab. DARPA 98 dataset. DARPA datasets, to the best of authors' knowledge, are the most comprehensive undertaking in intrusion detection system benchmarking. Similar to the preprocessing stage of KDD 99 competition, in order to summarize packet level data into high-level events such as connections, we employed Bro network analyzer. For KDD 99 compe-

tition, Bro was customized to derive 41 features per connection. This customized version is not publicly available furthermore the publicly available version can only derive 6 features per connection. Therefore, in our analysis, we employed 6 basic features from KDD datasets and employed publicly available Bro to summarize *tcpdump* data into connection records with these 6 features. The features are duration of the connection; protocol; service; connection status; total bytes sent to destination host; total bytes sent to source host. Among three datasets provided in KDD 99 competition, 10% KDD dataset, which was the original training set, was employed. Since our synthetic dataset only contains HTTP connections, other datasets were filtered to contain only HTTP connections. Outliers in each dataset were determined and removed by box plot with fences method [6].

In addition to the synthetic traffic that we generated from Ege University Vocational School web server logs and synthetic dataset from KDD 99 competition, we employed a real-world traffic, which is the captured traffic from Dalhousie University Faculty of Computer Science server Locutus, which has hundreds of users. The traffic was recorded in one day on December 2003. The approximate size of the recorded traffic is 2 gigabytes.

Analysis with Clustering

The objective of the k-means algorithm [5] is to partition the dataset into k groups or clusters where the number of desired clusters k is fixed *a priori*. Each cluster is assigned a cluster center (centroid), which defines the geometric center of the cluster. K-means clustering utilizes an iterative algorithm to minimize the distances between the dataset instances and the centroids. Training is carried out until the position of the centroids stop changing. Resulting clusters depend on the initial placement of the centroids.

By utilizing k-means clustering, centroids are calculated for each dataset. Since the initial placement and the number of centroids influence the resulting clusters, k-means clustering was employed with 6, 18 and 36 centroids. Different number of centroids produced similar results therefore results on 36 centroids are shown in Table 1. Each value in Table 1 is a coefficient of similarity and expressed as *C (D1, D2),* where *D1* is the dataset specified in the column header and *D2* is the dataset specified in the row header. The coefficient determines the similarity between the training dataset *D1* and test dataset *D2*. Coefficient of similarity is calculated by utilizing the centroids of the training dataset. For each instance in the test dataset, Euclidean distance to the closest training centroid is calculated. Calculated distances are averaged over all test dataset instances to form the coefficient of similarity. Given there are 6 normalized features, the coefficient ranges between 0 and 2.45, where 0 indicates high similarity.

Table 1. Coefficient of similarity matrix

	KDD	Ege	Locutus
KDD	0.001	0.355	0.341
Ege	0.826	0.007	0.257
Locutus	0.789	0.127	0.227

The coefficient *C (D, D)* shows the degree of dispersion in dataset *D*. *C (KDD, KDD)* and *C (Ege, Ege)* indicate that the synthetic dataset instances are close to the centroids, which results in low dispersion, whereas *C (Locutus, Locutus)* indicates higher dispersion in real-world datasets. Additionally, results detailed in the KDD column, i.e. *C (KDD, Ege)* and *C (KDD, Locutus)*, demonstrate that KDD is dissimilar to other two datasets. Compared with KDD dataset, i.e. *C (Locutus, KDD)*, our synthetic data shows more similarities to real-world data, *C (Locutus, Ege)*.

Analysis with SOM Based Intrusion Detection System

In Kayacik et al. [4], a two level SOM hierarchy was developed for intrusion detection. The system utilizes the 6 basic features of a connection, which can be derived from packet headers without inspecting the packet payload. SOM [5] is a neural network algorithm, which employs unsupervised learning for training. At the first level, six SOMs are trained, one for each feature where the general objective is to encode temporal relationships within the features. The second level combines the information from the first level SOMs. Other than the representation of temporal features and labeling of second level neurons, no further *a priori* information is employed.

Since real-world datasets are unlabeled, hit histograms are employed to analyze datasets. Hit histograms summarize how often the neurons are excited (i.e. selected as the best matching neuron). As a neuron is excited for more patterns in a dataset, the area of the hexagon being colored becomes larger. If the training and the test datasets have similar statistical properties such as similar range, probability distribution and dispersion, then the test dataset would populate a considerable portion of the SOM.

In Figure 1, the hierarchy trained on KDD demonstrates that, out of 36 neurons, mainly one or two are excited for any given dataset. This suggests that the organization of the SOM trained on KDD is unable to distinguish the characteristics of any dataset other than its training set (to a minimal degree). The system trained on our synthetic dataset (Figure 2) showed improvements compared to the system trained on KDD. In Ege SOM hierarchy, patterns in real-world dataset Locutus start to excite multiple neurons, specifically the lower right region. This indicates that the organization of the SOM hierarchy can comparatively distinguish different patterns in a real-world dataset.

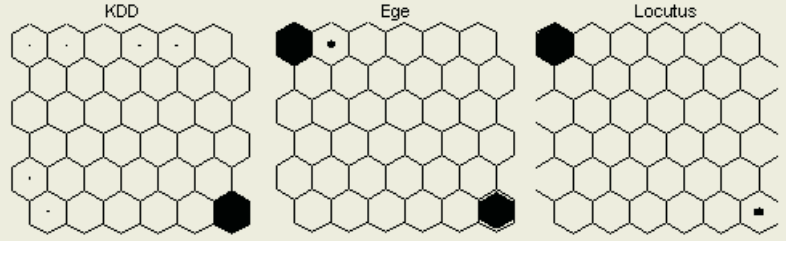


Fig. 1. Hit histograms from SOM hierarchy, which is trained on KDD dataset

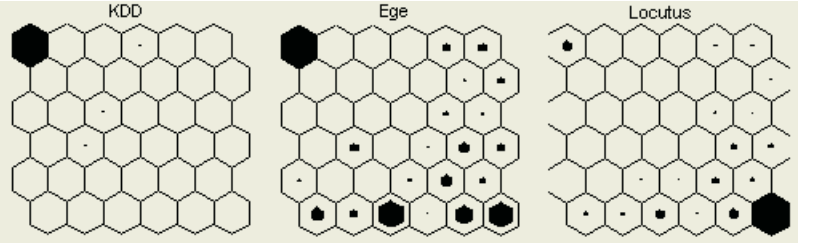


Fig. 2. Hit histograms from SOM hierarchy, which is trained on our synthetic dataset Ege

4 Conclusions and Future Work

In this paper, a framework for generating synthetic network traffic is discussed along with two techniques to analyze network traffic. The framework generates synthetic network traffic based on the models developed from web server logs. The set of components that we developed within the framework comprise a comprehensive toolkit, which can develop usage models and generate traffic. Previous benchmarking approaches (particularly, KDD dataset) were criticized because: (1) Simulated normal usage is said to be similar to what one would observe in a real-world environment. However this claim was not sufficiently validated. This raises many questions about how the normal usage is modeled [7]. (2) Internet traffic is not well behaved. Moreover, real-world data is more diverse than the synthetic traffic. Mahoney et al. [8] established that the KDD dataset is clean and known to have idiosyncrasies that can lead to detection.

In order to develop anomaly based detectors that can withstand in real-world environments; it is crucial to synthesize training data that has the similar characteristics to real-world data. Although our framework currently supports one protocol, by providing analytic modeling and facilitating repeatable simulations, it fills a gap in intrusion detection system benchmarking by generating synthetic traffic similar to real-world traffic.

In addition to generating synthetic network traffic, we employed two machine learning techniques to analyze the datasets. In k-means clustering, the objective is to find the centers of mass in the dataset. If two datasets are similar, their centers of mass will be close therefore the coefficient of similarity will be small. SOM intrusion detection system aims to achieve topological arrangement with respect to the training set, if the training and the test datasets have similar characteristics, the test dataset would excite a substantial portion of the SOM. Analysis results showed that the synthetic dataset generated by the framework shows improvements over KDD dataset in terms of being more similar to the real-world dataset.

The aforementioned techniques can also be employed for forensic analysis on network traffic. For instance given the traffic recorded at different periods from the same source, analysis can reveal the changing trends in the network traffic. Visualization methods of SOM can be employed to see the topological relationships of the datasets. Moreover analysis of the outliers can provide information about the anomalous

activities. The future work will investigate the use of other datasets from commercial and governmental organizations to develop models of usage and will employ other machine learning techniques to analyze the datasets.

Acknowledgements

This work was supported in part by NSERC Discovery and CFI New Opportunity Grants from the Canadian Government. Further information regarding the work described in this publication is available through the project homepage of the NIMS research group http://www.cs.dal.ca/projectx/.

References

1. Kayacik, G. H., Zincir-Heywood, A. N., "Generating Representative Traffic for Intrusion Detection System Benchmarking", Proceedings of the IEEE CNSR 2005 Halifax, Canada, May 2005.
2. Odlyzko A., "Internet traffic growth: Sources and implications", 2003, http://www.dtc.umn.edu/~odlyzko/doc/itcom.internet.growth.pdf Last accessed Nov. 2004.
3. Norris J.R., "Markov Chains", Cambridge University Press, 1997, ISBN 0-521-48181-3
4. Kayacik, G. H., Zincir-Heywood, A. N., Heywood, M. I., "On the capability of SOM based intrusion detection systems," Proceedings of the 2003 IEEE IJCNN, Portland, USA, July 2003.
5. MacQueen, J.B., "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297, 1967.
6. Chambers J., Cleveland W., Kleiner B., and Tukey P., 1983, "Graphical Methods for Data Analysis", Wadsworth.
7. McHugh J., "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory", ACM Transactions on Information and System Security, Vol. 3 No.4 November 2000.
8. Mahoney M.V., Chan P.K. "An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection", In RAID 2003 Symposium, Pittsburgh, USA, September 8-10, 2003, Springer 2003, ISBN 3-540-40878-9

Discovering Identity Problems: A Case Study

Alan G. Wang[1], Homa Atabakhsh[1], Tim Petersen[2], and Hsinchun Chen[1]

[1] Department of Management Information Systems, University of Arizona, Tucson, AZ 85721
{gang, homa, hchen}@eller.arizona.edu
[2] Tucson Police Department, Tucson, AZ 85701
Tim.Petersen@tucsonaz.gov

Abstract. Identity resolution is central to fighting against crime and terrorist activities in various ways. Current information systems and technologies deployed in law enforcement agencies are neither adequate nor effective for identity resolution. In this research we conducted a case study in a local police department on problems that produce difficulties in retrieving identity information. We found that more than half (55.5%) of the suspects had either a deceptive or an erroneous counterpart existing in the police system. About 30% of the suspects had used a false identity (i.e., intentional deception), while 42% had records alike due to various types of unintentional errors. We built a taxonomy of identity problems based on our findings.

1 Introduction

Identity resolution is central to fighting against crime and terrorist activities in various ways. Identity information, in many cases, is unreliable due to intentional deception [7] or data entry errors. Commercial database systems search records mainly based on exact-matches. Records that have very minor changes may not be returned by searching on an exact-match. This causes problems for information retrieval which in law enforcement and intelligence investigations would cause severe consequences [3].

In this research we look into the problems that produce difficulties in retrieving identities. The paper is organized as follows: In section 2 we briefly introduce the definition of personal identity based on related research. In section 3 we describe a case study conducted in a local law enforcement agency. Results are analyzed and summarized to create a taxonomy of identity problems. In section 4 we summarize our findings and suggest techniques that improve identity information retrieval.

2 Background

An identity is a set of characteristic elements that distinguish a person from others [2, 5]. Since identity theft/fraud has become a serious problem, some research has been done specifically on identity issues.

In their report on identity fraud, the United Kingdom Home office [4] identified three basic identity components: attributed identity, biometric identity, and biographical identity. Clarke's identity model [1] gives a more detailed classification

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 368 – 373, 2005.

of identity information. He argues that identity information falls into one of the five categories: social behavior, names, codes, knowledge, tokens, and biometrics. These works mainly focus on the representation of identities, addressing what it takes to distinguish a person from others. However in the real world, it is impossible to collect all types of identity information for every person. And information collected in the real world is far from perfect. For example, criminal suspects might intentionally use false identities when being confronted by police officers. Those problems make identity identification a difficult job.

3 A Case Study on Identity Problems

A rich source for research into identity problems is the records management systems of local police departments. We chose Tucson Police Department (TPD) as our test bed. TPD serves a relatively large population that ranks 30th among US cities with populations of over 100,000, and Tucson's crime index ranked around 20th highest among US metropolitan areas. We hope that the results of the case study conducted at the TPD can be generalized to other law enforcement agencies.

An identity record in the TPD system consists of many attributes such as name, DOB (date of birth), ID numbers (e.g., SSN, Driver's License Number), gender, race, weight, height, address, and phone number. Biometrics attributes such as fingerprints are not available to this study due to privacy and security reasons. An identity record may not have values in all of its attributes. Name is a mandatory attribute and always has a value. Other attribute values are allowed to be empty or are assigned a default value when not available (e.g., the default value for height in the TPD is 1).

3.1 Data Collection

TPD has 2.4 million identity records for 1.3 million people. Some people may have more than one identity record associated with them. We suspect there might be deception and errors in duplicate records. We first collected identity records for people who had more than one identity in the TPD database. To our disappointment, we found that multiple identities associated to the same person were exact duplicates. Most of them had exactly the same values in such attributes as name, DOB, ID numbers, etc. According to a TPD detective, two identities were associated only when police investigation happened to catch the duplicates. However, duplicate records that differ too much might be less noticed than exact duplicates during police investigation. Therefore, they might not have been associated in the database.

We then randomly drew 200 unique identity records from the TPD database. We considered them to be a list of "suspects" that we were trying to find any matching identities for in the TPD database. Given the huge amount of identity records in the TPD, it is nearly impossible to manually examine every one of them. We used an automated technique [7] that computes a similarity score between a pair of identities. This technique examines only the attributes of name, address, DOB, and SSN. It first measures the similarity between values in each corresponding attribute of the two identities and then calculates an overall similarity score as an equally weighted sum of the attribute similarity measures. We used this technique to compare each suspect's

identity to all other identities in the database. For each suspect's identity, we chose the 10 identity records that had the highest similarity scores.

We verified the 10 possible matches for each of the 200 suspects by manually examining their key attributes. Each possible match was classified into one of the four categories defined in Table 1. The first two categories, D and E, imply a true match. A matching identity was considered to be an error when identical values were found in key attributes such as name and ID numbers. A matching identity was considered deceptive when key attribute values such as name, DOB and ID numbers were not identical but showed similar patterns. If a matching identity had very different values on those key attributes, it was considered a non-match. If a matching identity was not in any aforementioned category, it was categorized as U (uncertain).

Table 1. Categories into which possible matches are classified

Category	Description
D	Intentional Deception
E	Unintentional Errors
N	Non-match
U	Uncertain (too little information to make a call)

3.2 Preliminary Evaluation

We asked a TPD detective who has worked in law enforcement for 30 years to verify our categorization results. He agreed on most of our decisions. It surprised us that more than half (55.5%) of the suspects had either a deceptive or an erroneous counterpart existing in the TPD system. About 30% of the suspects had used a false identity (i.e., intentional deception), while 42% had records alike due to various types of unintentional errors. As the numbers imply, some suspects may have both deceptive and erroneous records in the TPD system.

3.3 A Taxonomy of Identity Problems

As shown in Figure 1, we built a taxonomy of identity problems based on our findings. Among others available in the TPD, attributes such as name, DOB, ID numbers, and address indicate deception or errors in most cases. Attributes such as weight, height and race are usually estimated through visual inspection by police officers, and their values are by nature indefinite. We do not consider value differences for these attributes deception or errors. Although gender is also visually inspected and is often definite, we did not find any evidence in our sampled data that officers were deceived or made a mistake on that.

Deceptive identities and erroneous identities exhibit quite different characteristics. We discuss the two types of identity problems respectively in the rest of this section.

Identity Errors. Erroneous identities were found to have discrepancy in only one attribute, or having no discrepancy at all (i.e., duplicates). There were 65.5% of erroneous identities that had slightly altered values in either name (50.0%) or DOB

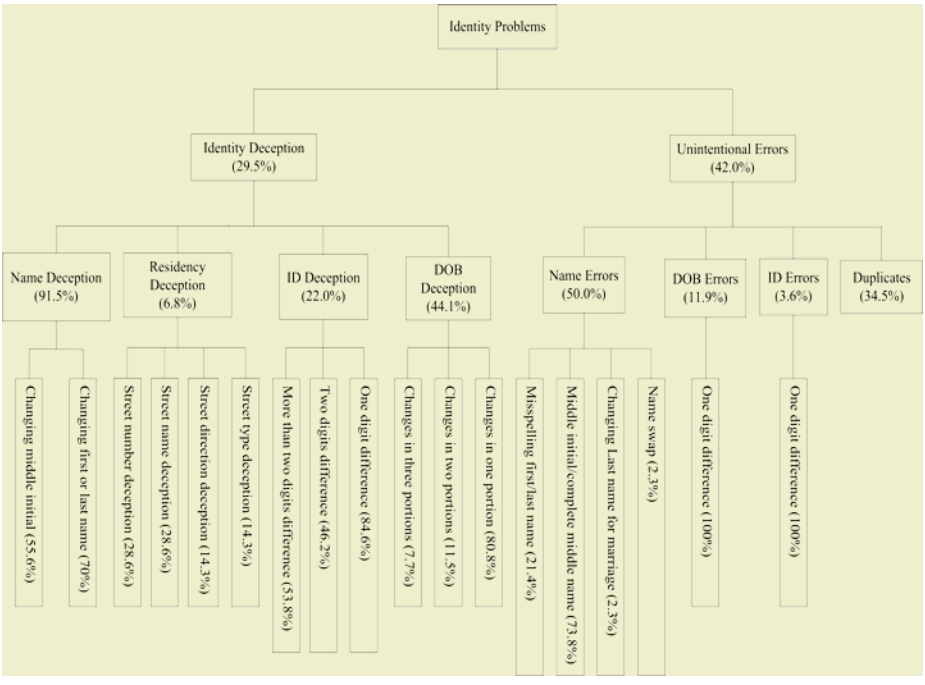


Fig. 1. A taxonomy of identity problems

(11.9%) or ID numbers (3.6%). They had errors in only one attribute and had other attribute values identical to those of the corresponding true identity. The rest of the erroneous identities (34.5%) were merely duplicates. Their attribute values were all identical to those of the corresponding true identity.

Values in name errors were altered in different ways. 73.8% of them either had their middle names missing or switched between using a middle initial and using a full middle name. Misspelling was found in 21.4% of name errors. There was only one instance where first name and middle name were swapped (2.3%). There was also another instance in which the name was altered by having a different last name after marriage (2.3%). Certainly, such an instance only applies to female identities.

A DOB value consists of three portions: a year, a month, and a day, e.g., 19730309. In our study most of the erroneous DOB values were only one-digit different from the corresponding true values. It is likely caused by mistyping or misreading from a police report. An erroneous DOB is not necessarily close to its corresponding true value in time. For example, one DOB was altered as "19351209," while its true value was "19551209." In this case one-digit difference made 20 years' difference in time.

Similar to DOB errors, most of the ID errors were also found to be different from the corresponding true values by only one digit.

Identity Deception. Deceptive Identities usually involves changing values in more than one attribute. The value changes are more drastic than those in erroneous identities. Name was found to be the attribute most often subject to deception (91.5%). Less than half of the deceptive identities (44.1%) had altered DOB values and 22% of them had altered ID numbers. There were also 6.8% of the deceptive identities with an altered residential address. We discuss each type of identity deception in details below.

Seventy percent of the deceptive names were altered by changing either first or last name, but not both of them. Also the names were not altered randomly. Popular name changing techniques we found include: using a name that is phonetically similar (e.g., Wendy being altered to Windy), using a nick name (e.g., Patricia being changed to Trish), using a name translated from other languages (e.g., a Spanish name Juan being changed to Johnny), or using someone else's name (e.g., brother's or sister's name). Changing first or last name was often accompanied by changing the middle name.

As opposed to DOB errors, deceptive DOB values can be made by changing more than one portion. We found that 7.7% of deceptive DOBs made changes in all three portions, 11.5% of them made changes in two portions, and 80.8% made changes in one portion. Deceptive DOB values were often altered in a way different from DOB errors. An erroneous DOB value often results in one-digit difference by, for example, mistyping a visually similar number (e.g., typing 0 for 8). One the contrary, a deceptive DOB value may have more than one-digit difference. For example, one subject made himself seem younger by reporting his DOB as "19630506," while his true DOB was "19580506." In deceptive DOB, although there is still only one altered portion, the altered value has a two-digit difference from the true value. There was also evidence that people alter DOB values using some patterns. One subject reported his DOB as "19730203," while his true DOB was "19730102." One can notice the pattern of adding one to both month and day values. Switching digits is another popular technique. For example, one subject reported "19380129," while his true DOB was "19390128."

Changes made in deceptive ID numbers were also more drastic than ID errors. 53.8% of deceptive ID numbers were altered more than three digits. 46.2% of them were altered by more than two. Making a one-digit change was also common in ID deception (84.6%). Those percentages do not sum up to one hundred percent because an identity record may have multiple ID numbers associated. One who deceives on one of his/her ID numbers will be very likely to deceive on other numbers associated with him/her. Some techniques used for DOB deception, such as switching digits, were also found in ID number deception.

Residential address is unreliable in determining a person's identity because many people move frequently. We found that 6.8% of the suspects deceived at least once on address information. An address consists of four main components: street number, street direction, street number, and street type. Sometimes there is also a suite/apartment number. 28.6% of address deception altered street numbers or street names, while 14.3% of address deception altered street direction or street type. It seems people are more used to altering textual values such as numbers and names than to changing nominal values such as street directions and types.

4 Conclusion and Discussion

In this study we examine the identity problems that result in difficulties for identity resolution. Two types of problems, including intentional deception and unintentional errors, were found in real law enforcement records. Attribute values were altered more drastically in deception than in errors. In most of the cases, altered values looked very similar to the corresponding true values.

Our future work is to develop an automated identity resolution technique that takes our findings on identity problems into account. Techniques that improve identity information retrieval should locate identity information in an approximate rather than exact manner. Similarity measures need to be defined for different attributes based on the characteristics presented by both deceptive and erroneous values. A string comparator such as Edit Distance [6] may be a good candidate in some cases. But in some other cases, the similarity measure can be more complex. For example, a DOB similarity measure needs to capture the pattern of adding one to both month and day values. A decision model is necessary to determine whether an identity should be retrieved given a set of similarity measures on attributes. Finally, the techniques have to be automated because the huge amount of data prohibits any possibility of manual operations.

Acknowledgement

This project has been primarily funded by the following grant: NSF, Digital Government Program, "COPLINK Center: Social Network Analysis and Identity Deception Detection for Law Enforcement and Homeland Security," #0429364, 2004-2006.

References

1. Clarke, R.: Human Identification in Information Systems: Management Challenges and Public Policy Issues. Information Technology & People 7, 4 (1994) 6-37
2. Donath, J. S.: Identity and Deception in the Virtual Community. In: M. a. K. Smith, P. (ed.): Communities in Cyberspace. Routledge, London, (1998)
3. GAO: Law Enforcement: Information on Timeliness of Criminal Fingerprint Submissions to the FBI. GAO-04-260, United States General Accounting Office (GAO) (2004)
4. HomeOffice, U. K.: Identity Fraud: A Study. United Kingdom HomeOffice (2002)
5. Jain, A. K., Prabhakar, S., Hong, L., and Pankanti, S.: FingerCode: A fingerbank for fingerprint presentation and matching. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (1999)
6. Levenshtein, V. L.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady 10, (1966) 707-710
7. Wang, G., Chen, H., and Atabakhsh, H.: Automatically Detecting Deceptive Criminal Identities. Communications of the ACM 47, 3 (2004) 71-76

Efficient Discovery of New Information in Large Text Databases

R.B. Bradford

SAIC, Reston, VA
bradfordr@saic.com

Abstract. Intelligence analysts are often faced with large data collections within which information relevant to their interests may be very sparse. Existing mechanisms for searching such data collections present difficulties even when the specific nature of the information being sought is known. Finding unknown information using these mechanisms is very inefficient. This paper presents an approach to this problem, based on iterative application of the technique of latent semantic indexing. In this approach, the body of existing knowledge on the analytic topic of interest is itself used as a query in discovering new relevant information. Performance of the approach is demonstrated on a collection of one million documents. The approach is shown to be highly efficient at discovering new information.

1 Introduction

Many of the most important sources of data for intelligence analysis are textual in nature. Locating relevant information in large volumes of text can be difficult, even when the analyst has a good idea of what he or she is looking for. The problem is much greater when the analyst does not know what he or she is missing. There is, thus, a great need for techniques that will allow an analyst to identify information that is potentially relevant to an intelligence topic without requiring the analyst to know what that missing information is in sufficient detail to formulate a directed query to find it.

The technique of latent semantic indexing (LSI) possesses unique properties applicable to this problem. This paper describes an iterative approach to information discovery in large text databases which exploits this technique.

2 Latent Semantic Indexing

The LSI technique can be applied to an arbitrary collection of documents. The technique automatically creates a vector space representation of both the documents of the collection and the terms that occur in those documents. Similarity of any two objects in that space is determined by the proximity of their representation vectors.

The technique of latent semantic indexing consists of the following primary steps [1]:

1. A matrix is formed, where each row corresponds to a term that appears in the documents of interest, and each column corresponds to a document. Each element

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 374 – 380, 2005.

(m,n) in the matrix corresponds to the number of times that the term m occurs in document n.
2. Local and global term weighting is applied to the entries in the term-document matrix.
3. Singular value composition (SVD) is used to reduce this matrix to a product of three matrices, one of which has non-zero values (the singular values) only on the diagonal.
4. Dimensionality is reduced by deleting all but the k largest values on this diagonal, together with the corresponding columns in the other two matrices. This truncation process is used to generate a k-dimensional vector space. Both terms and documents are represented by k-dimensional vectors in this vector space.
5. The relatedness of any two objects represented in the space is reflected by the proximity of their representation vectors, generally using a cosine measure.

The technique has been applied in a wide variety of information processing tasks, especially text retrieval [1] and document categorization [3]. A recent review by Dumais, one of the inventors of LSI, provides a good overview of the subject. [2] Singular value decomposition, in combination with automatic keyword expansion of queries has been applied to discovery of novel information by Vats and Skillicorn. [11] LSI has some particularly attractive features for information discovery:

- Conceptual similarity of documents as determined by LSI has been demonstrated to be remarkably well-correlated with conceptual similarity of those same documents as judged by human beings. [3,4]
- LSI vectors reflect subtle relationships derived from a holistic analysis of the totality of information indexed. [5]
- The technique is independent of language and can even be used in a cross-lingual fashion. [6]
- LSI is provably optimal for certain types of related information processing tasks. [8,9]
- An efficient and highly scalable commercial implementation of the technique is available for experimentation and application.

3 Proposed Information Discovery Approach

The approach proposed here for discovery of new information in text databases is an iterative one, with the following essential steps:

1. Generate an LSI representation space from the collection of documents that are to be examined in the discovery process.
2. Create a starting query using a textual representation of the analyst's initial knowledge on the topic of interest. (This could, for example, correspond to a draft intelligence report on the topic, or a collection of the analyst's notes. The LSI technique is not dependent upon text being expressed in proper sentences. Thus, raw notes such as phrases, incomplete sentences, and lists of names can be concatenated to create this initial query.)

3. Using the existing query, retrieve a ranked list of documents having the most similar representation vectors in the LSI space.
4. Review some number of the retrieved documents in rank order and extract textual items relevant to the analysis topic.
5. Concatenate the textual items extracted in this iteration with the text of the current query to create a new query.
6. Repeat steps 3 through 5 until a desired degree of completeness of information on the topic has been derived.

4 Testing

This paper reports the results of a test conducted to determine both the effectiveness and the efficiency of this proposed approach. The test was designed to emulate an intelligence analyst pursuing a counterterrorism analysis assignment.

The data employed consisted of 1.04 million English-language news articles from the period 2000 through 2004. These articles contain over 1.5 million unique terms and comprise 3.9 gigabytes of text.

The task chosen as a test case was to mine the collection of one million documents for information relevant to a specific terrorist organization. The organization chosen for the test was the Groupe Salafiste pour la Predication et le Combat, or Salafist Group for Call and Combat (GSPC). This organization, originating in Algeria, is closely linked to Al Qaeda.

For text retrieval systems, efficiency typically is measured in terms of *precision* and comprehensiveness in terms of *recall*. Ordinarily, these metrics are applied at the document level:

- Precision = number of relevant documents / number of retrieved documents
- Recall = number of retrieved relevant documents / total number of relevant documents

However, this definition did not appear appropriate for this study, as the emphasis here is on discovering *new information*, not just *relevant documents*. In the Text Retrieval Conferences (TREC), the novelty track has employed precision and recall measures at the sentence level. That is, human evaluators have judged each sentence in the documents under consideration with regard to relevance and novelty for the queries employed. [7] This approach was difficult to apply for the documents treated here. There are many sentences that are relevant only in an implicit sense in the overall context of a given article. Furthermore, many of the new items of information discovered are spread across multiple sentences. Thus, the use of a sentence-based measure did not appear to be appropriate here. The decision was made to apply these measures at the level of *information items*. During the testing, information regarding the GSPC was extracted (by the author) as small units (items) that capture a complete thought, for example; *Djamel Beghal is a member of the GSPC*. The measures applied here are analogous to nu-precision and nu-recall as defined by Allan et al [10], except that here we are dealing with *items* rather than *events* and *document*s rather than *sentences*.

This work is part of a larger effort that includes automatic generation of candidate RDF triples for inclusion in ontology on this terrorist group. The great majority of items extracted manually in the work reported here were such triples. The criterion for novelty was that the article contributed at least one novel entry for the ontology. In order to avoid confusion, a subscript *i* has been applied here to denote that the measures are applied at the information item level:

- Precision_i = number of documents containing relevant information items not seen in *any* previous document / number of documents retrieved
- Recall_i = number of unique relevant information items retrieved / total number of unique relevant information items in the database

The test program consisted of five steps. In Step 1, text relevant to the GSPC was used to discover articles containing information relevant to the GSPC among the 1.04 million articles indexed. In this test, a short (one page) report on the GSPC was used as an initial query. This seed report contained 39 items of information regarding the GSPC, including four names and one alias of members of the organization. This text was used as a query against the 1.04 million news articles indexed in the LSI space. As will be shown below, there are over 800 unique information items relevant to the GSPC in the data. Thus, this query corresponds to a case where the analyst's knowledge relative to the information in the overall database is very low (less than 5% of the total).

The LSI technique produces a ranking of documents within the space in terms of their conceptual similarity to the query. The most highly ranked 100 articles retrieved by this initial query were manually examined in search of *new* information items relevant to the GSPC. In order for document N from the 100 to be counted in the calculations of precision_i, that document had to contain at least one new relevant item not present in either the original query or in any of the N-1 documents reviewed up to that point. Of the top 100 documents retrieved in step 1, 76 met those criteria. In aggregate, these 76 articles provided 305 new items of information about the GSPC.

Documents containing information relevant to the GSPC are sparse in this database – roughly one in one thousand. Thus, this information discovery approach demonstrates a strong ability to discriminate against irrelevant information.

In Step 2 the proposed approach was applied iteratively. The text of the 305 newly identified information items from Step 1 was appended to the initial report and this resulting text was used as a new query. As in Step 1, the most highly-ranked documents produced by this query were then manually examined. An exclusion feature was applied to the result set, ensuring that none of the 100 documents examined in detail in step 1 were included in the documents to be examined in this round of the iteration.

In steps 3 and 4, the procedure was applied twice more. Results of the four iterations are shown in Table 1. After the four iterations, the analyst has accumulated a knowledge base on the GSPC which contains 819 unique information items. This includes 129 names of members of the GSPC plus 51 of the aliases that they have used. This is a major improvement in comprehensiveness in comparison to the initial knowledge base.

Table 1. Information Discovered using the Proposed Approach

	New Items	New Names	New Aliases	Cumulative Items	Cumulative Names	Cumulative Aliases
Initial Knowlede Base	39	4	1	39	4	1
Iteration 1	305	60	18	344	64	19
Iteration 2	263	28	16	607	92	35
Iteration 3	155	23	10	762	115	45
Iteration 4	57	14	6	819	129	51

A separate analysis effort indicated that there are approximately 1,000 unique information items relevant to the GSPC in the entirety of the database. This analysis was based in part on examination of articles selected on the basis of associated metadata. It also was based in part on examination of thousands of articles selected on the basis of criteria independent of this work (i.e., not selected using LSI and not selected as relevant to the GSPC). Although the approaches employed showed reasonably good agreement, this estimate should be considered as only a general approximation. Accordingly, the estimated recall numbers cited here should be considered only approximate.

Cumulative precision_i and recall_i numbers for Steps 1 through 4 of the testing are presented in the following table.

Table 2. Precision_i and Recall_i for each Step of the Iteration Testing

Test Step	Cumulative Precision_i of Discovery of New items	Estimated Cumulative Recall_i of Unique Items
1	76%	34%
2	69%	61%
3	60%	76%
4	53%	82%

As would be expected, precision is declining as the approach is repeatedly applied. However, it should be noted that by Step 4 we are dealing with a case where the analyst has developed a relatively comprehensive knowledge base on the GSPC. Even at that point, one document in three being examined contains *new* relevant data. With four iterations, the information discovery approach proposed here has achieved a recall_i of approximately 80% with a precision_i of over 50%. This is well beyond the performance of typical text retrieval techniques, even *without* the requirement of novelty.

Step 5 of the testing was designed to determine the effect of database size on performance of the approach. Subsets of the data were created that contained 606 thousand and 890 thousand of the documents. LSI spaces were then created for these subsets. The query from Step 1, containing 39 information items was applied against these two subsets of the data. The top 100 retrieved documents were then reviewed in the same manner as in Step 1. The results are shown in the following table.

Table 3. Effect of Database Size

Database Size (# of Documents)	$Precision_i$	# of New Items Found
606,000	62%	233
890,000	69%	288
1,047,000	76%	305

The most notable feature of these results is that performance improved as the size of the database increased. The results demonstrate that the approach is quite efficient independent of database size, at least within the general range of half a million to a million documents.

5 Summary

The test results presented here demonstrate that the approach proposed in this paper constitutes a highly efficient method for discovering new information relevant to a given topic in large textual databases. The approach discriminates very well against extraneous information. The approach is applicable over a broad range of comprehensiveness of the analyst's knowledge of the topic being researched. Efficiency declines as the comprehensiveness of the analyst's knowledge base increases. However, it does so at a slow rate. With the software employed[1], this approach can be used with text from any language that can be represented in Unicode. In fact, it can be applied to any type of data that can be partitioned into events with attendant observations.

The author wishes to acknowledge the contribution of Martin Hobson in creating the LSI spaces used in the testing.

References

1. Deerwester, S., Dumais, S., Landauer, T., Furnas, G., Beck, L.: Improving Information Retrieval with Latent Semantic Indexing. In: Proceedings of the 51st Annual Meeting of the American Society for Information Science **25** (1988) 36-40
2. Dumais, S. T.: Latent Semantic Analysis. In: Annual Review of Information Science and Technology, Vol. 38. Information Today Inc., Medford, New Jersey (2004) 189-230
3. Zukas, A., Price, R.J.: Document Categorization Using Latent Semantic Indexing. In: Proceedings, Symposium on Document Image Understanding Technology (2003) 87-91
4. Landauer, T. K., Laham, D., Foltz, P.: Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report. In: Advances in Neural Information Processing Systems 10, Cambridge: MIT Press, (1998) 45-51
5. Kontostathis, A., Pottenger, W. M.: A Mathematical View of Latent Semantic Indexing: Tracing Term Co-occurrences. Technical Report LU-CSE-02-006, Lehigh University (2002)

[1] The testing was carried out using version 2.2.1 of the Content Analyst™ text analytics software from Content Analyst Company, LLC.

6. Landauer, T., Littman, M.: Fully Automatic Cross-language Document Retrieval Using Latent Semantic Indexing. In: Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research (1990) 31-38
7. Soboroff, I., Harmon, D.: Overview of the TREC 2003 Novelty Track. In: The 12th Text Retrieval Conference (TREC 2003), NIST Special Publication SP-500-255 (2003)
8. Bartell, B.T., Cottrell, G.W., Belew, R. K.: Latent Semantic Indexing is an Optimal Special Case of Multidimensional Scaling. In: Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (1992) 161-167
9. Ding, C.H.Q.: A Similarity-based Probability Model for Latent Semantic Indexing. In: Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval. (1999) 59-65
10. Allan, J., Gupta, R., Khandelwal, V.: Temporal Summaries of News Topics. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2001) 10-18
11. Vats, N., Skillicorn, D.: The ATHENS System for Novel Information Discovery. Queens University External Technical Report, ISSN-0836-0227-2004-489, 13 October, 2004

Leveraging One-Class SVM and Semantic Analysis to Detect Anomalous Content

Ozgur Yilmazel, Svetlana Symonenko, Niranjan Balasubramanian, and Elizabeth D. Liddy

Center for Natural Language Processing,
School of Information Studies – Syracuse University,
Syracuse, NY 13244
{oyilmaz, ssymonen, nbalasub, liddy}@syr.edu

Abstract. Experiments were conducted to test several hypotheses on methods for improving document classification for the malicious insider threat problem within the Intelligence Community. Bag-of-words (BOW) representations of documents were compared to Natural Language Processing (NLP) based representations in both the typical and one-class classification problems using the Support Vector Machine algorithm. Results show that the NLP features significantly improved classifier performance over the BOW approach both in terms of precision and recall, while using many fewer features. The one-class algorithm using NLP features demonstrated robustness when tested on new domains.

1 Introduction

This paper reports on further developments in the research [1, 2] that leverages Natural Language Processing (NLP) and Machine Learning (ML) technologies to improve one aspect of security within the Intelligence Community (IC). This would be done by monitoring insiders' workflow documents and alerting the system assurance administrator if the content of the documents shifts away from what is expected, given the insiders' assignments. This capability is being implemented as one piece of a tripartite system prototype within the context of the ARDA-funded project, *A Context, Role and Semantic-based Approach for Countering Malicious Insider* [3]. We evaluate the applicability of a one-class categorization algorithm - Support Vector Machines (SVM) - which, unlike a regular classifier, is trained on 'typical' examples only and then used to detect 'typical' and 'atypical' data. This is warranted by the context of the problem where the subject domain of interest to the malicious insider is unknown in advance and, therefore, it is not feasible to provide 'off-topic' examples to train a classifier.

2 Problem Background

It is known from Subject Matter Experts (SMEs) from the IC that analysts operate within a mission-based context, focused mainly on specific topics of interest (TOIs)

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 381 – 388, 2005.

and geo-political areas of interest (AOIs). The information accessed by analysts ranges from news articles to analyst reports, official documents, emails, queries. The role and the task assigned to the analyst dictate the scope of their TOI/AOI. Within this mission-focused context, our hypothesis is that the ML-based categorization of documents using the features produced by the NLP-based semantic analysis will detect whether an insider's document workflow is within the TOI/AOI scope of their assigned task.

To illustrate the problem, consider the "Threat Scenario", based on a review of malicious insider cases and consultations with the SME. An analyst works on problems dealing with the Biological Weapons Program (TOI) in Iraq (AOI). The analyst begins collecting information on missiles in North Korea. Since the topic is beyond his task, these actions are covert, interspersed with 'normal', 'on-topic' communications. Now and then he would query a database and retrieve documents on North Korea's missiles; or he would send a question to an analyst from the North Korea shop and receive documents via email; he would copy data to a CD. As these actions involve textual artifacts (documents, queries, emails) analysis of their content should indicate which topics are of interest to the analyst. Comparison of these topics to what is *expected*, given the analyst's task, would reveal whether they are beyond the expected scope. It is important to note that the system will not replace human supervisors, but assist them by reducing the data to analyze to just the detected 'anomalies'.

In addition to monitoring insider's communications, semantic analysis can be run *ex-post-facto*, if an information assurance engineer grew suspicious of an individual. It can also help characterize large collections of documents by separating them into semantic-driven categories.

3 Related Work

The problem of detecting malicious insider has been mainly approached from the *cyber security* standpoint, with systems as the main object of attack [4, 5]. The 2003 and 2004 ISI Symposia demonstrated an increased appreciation of information as a factor of national security. As information is often represented through textual artifacts, linguistic analysis has been applied to the problems of cyber security. Sreenath [6] applied latent semantic analysis to reconstruct users' queries and to detect malicious intent. Studies looked at linguistic indicators of deception in interviews [7], emails [8], and online chat [9].

Another line of text classification research addresses situations when providing 'negative' examples for training is not feasible, for example, in intrusion detection [10], adaptive information filtering [11, 12], and spam filtering [13]. Research effort has focused on application of a one-class categorization algorithm, which is trained on positive examples only and then tested on the data that contain both positive and negative examples. Conceptually, the task is to acquire all possible knowledge about one class and then apply it to identify examples that do *not* belong to this class. The one-class SVM [14], shown to outperform other algorithms [11, 12, 15], was chosen for our experiments. The novelty of our approach is in evaluating its effectiveness on various sets of features selected to represent documents. In particular, we compared the BOW representation with different combinations of features generated using NLP techniques.

4 Proposed Solution

The task of identifying 'off–topic' documents is modeled as a text categorization problem. Models of expected topics are built from the semantic content of a set representing the analyst's 'normal' document flow. New documents are categorized as on- or off-topic based on their semantic similarity to the Expected Model. The effectiveness of the solution is dependent on the accuracy of the categorization model and its generalizability to new documents. The most commonly used document representation has been the BOW [16, 17]. It has been shown that the knowledge of statistical distribution of terms in texts is sufficient to achieve high classification performance. However, in situations where the available training data is limited classification performance on BOW suffers. Our hypothesis is that the use of fewer, more discriminative linguistic features can outperform the BOW representation.

The novelty of the proposed approach is in using linguistic features extracted or assigned by our NLP-based system [18]. Such features include entities, named entities, and their semantic categories (i.e. PERSON, ORGANIZATION). The system can also map these features into higher-level concepts from external knowledge sources. The resulting document vectors are well separated in the feature space.

The NLP analysis is performed by TextTagger™, a text processing system built at the Center for Natural Language Processing [19]. It employs a sequence of rule-based shallow parsing phases that use lexico-semantic and syntactic clues to identify and categorize entities, named entities, events, and relations among them. Next, individual topics and locations are mapped to appropriate categories from knowledge bases. The choice of knowledge bases was driven by the project context. TOI inference is supported by an ontology developed for the Center for Nonproliferation Studies' (CNS) [20] collection of documents from the weapons of mass destruction (WMD) domain. AOI is inferred based on the SPAWAR Gazetteer [21]. Given that analysts' tasks are usually at the country level, the AOI inference is set to this level as well, but other degrees of granularity are possible. The entity and event extractions are output as frames, with relation extractions as frame slots (Figure 1).

Authorities suspect the Bavarian Liberation Army, an extreme right-wing organization, may be responsible.

Bavarian Liberation Army
Country=Austria
CNS_Superclasses=Terrorist-Group

Fig. 1. A sample extraction and concept inference.

The NLP-extracted features are then used to generate document vectors for machine learning algorithms.

5 Experimentation

Experimentation Dataset

Experiments were run on a subset of the larger Insider Threat collection created for the project. Its core comes from the CNS collection and covers such topics as WMD

and Terrorism, and such genres as newswires, articles, analytic reports, treaties, and so on. Training and Testing document sets were drawn from the collection based on the project scenarios. The scenarios are synthetic datasets representing the insiders' workflow through atomic actions (e.g. 'search database', 'open document'). The scenarios include a baseline case (with no malicious activity) and six threat cases. The scenarios cover the workflow of hundreds of insiders with different roles and tasks; for our experiments, we focused on one analyst from the Iraq/Biological Weapon shop. The above described Threat Scenario set the base for the Training and Testing datasets.

The documents were retrieved in a manner simulating the analysts' work: manually constructed task-specific queries were run against the Insider Threat collection. Both sets included 'noise' (webpages on topics of general interest) as we assumed that analysts may use the Web for personal reasons as well. Documents retrieved by the 'North Korea' queries were labeled as OFF-topic. All other documents were labeled as ON-topic, since, for the purposes of the project, it will suffice if the classifier distinguishes the 'off-topic' documents from the rest. The Training set contained only ON-topic documents, whereas the Testing set also included OFF-topic documents.

Classification Experiments

For classification experiments, we used an SVM classifier not only because it has been shown to outperform kNN, Naïve Bayes, and other classifiers on the Reuters Collection [22, 23], but also because it can handle one-class categorization problems. Experiments were run in LibSVM [24], modified to handle file names in the feature vectors, and to compute a confusion matrix for evaluation.

We experimented with the following feature sets:

1. Bag-of-words representation (BOW): each unique word in the document is used as a feature in the document vector.
2. Categorized entities (CAT): includes only entities or named entities.
3. TOI/AOI extractions (TOI/AOI): includes only terms assigned TOI/AOI indicators
4. TOI/AOI extractions + important categories (TOI/AOI_cat): TOI/AOI features (as in 3) plus all entities categorized as domain-relevant concepts ('missile', 'WMD').

We applied stemming, a stop-word filter, and lower case conversion to all of the representations. The value for each term in the document vector is the term frequency. The experiments were run with the linear kernel SVM, all parameters set to default.

Classifier performance was assessed using standard metrics of precision and recall [25] and a weighted F-score, calculated for each class.

In mainstream text categorization research, the performance focus is usually on the 'positive' class, so the metrics are often reported for this class only. The context of our project, however, gives greater importance to detecting the 'negative' (potentially malicious) cases, while keeping the rate of 'false alarms' down. This set an uncommon task for training the classifier: to aim not only for higher precision on ON-topic, but also for greater recall of OFF-topic. In evaluating the classifier, we focused on the scores for the OFF-topic class, and the F-measure for the OFF-topic class was calculated with the weight $\beta=10$ for Recall. The F-score for the ON-topic class was calculated using the standard weight $\beta=1$. Figure 2 shows the F-measure formula used.

The actual value of β is not significant as long as it is greater than one, since F-score is not used to tune parameters of the learning algorithm.

$$\text{F-score} = \frac{(\beta+1) * \text{Precision} * \text{Recall}}{\beta * \text{Precision} + \text{Recall}} \quad (2)$$

Fig. 2. Weighted F-score

The results (Table 1) demonstrate that, similarly to what was observed in experiments with the regular SVM classifier [2], document representations using TOI/AOI features only or in combination with domain-important categories improve the classifier performance over the baseline (BOW), while using many fewer features. In particular, AOI/TOI shows over 5% improvement in Recall (OFF) while using forty-nine times fewer features. Using a combination of AOI/TOI and category information achieves 16% improvement on Recall (OFF) and over 12% improvement on the weighted F-OFF over the baseline with nine times fewer features than BOW.

Table 1. Experimental results

	Features	Prec ON	Rec ON	F ON, β=1	Prec OFF	Rec OFF	F OFF, β=10
BOW	19774	97.22	68.68	80.50	19.0	78.93	61.34
CAT	10682	96.07	57.20	71.71	14.0	74.84	53.65
AOI/TOI	403	97.27	53.25	68.82	14.32	83.96	58.22
AOI/TOI_cat	2149	99.34	70.61	82.55	23.12	94.97	74.05

Although the decision to switch from the regular to the one-class SVM was guided by the context of our project, it was supported by the significantly higher performance of the one-class SVM on the OFF-topic class (Table 2). Regular SVM suffered from training on a weakly representative set for the OFF-topic class. Considering that the one-class SVM was able to achieve up to 94% of Recall OFF with no prior knowledge of what constitutes 'off-topic', the improvement is impressive. The downside of such a high Recall OFF, however, was the deteriorated Recall ON. In other words, the one-class SVM errs in favor of the previously unknown 'negative' class, thus, causing 'false alarms'.

Table 2. Recall of the OFF-topic class: Regular vs. One-Class SVM

	Regular SVM		**One-Class SVM**	
	Recall OFF	F OFF, β=10	Recall OFF	F OFF, β=10
BOW	48.11	50.49	78.93	61.34
CAT	27.0	28.92	74.84	53.65
AOI/TOI	38.99	41.28	83.96	58.22
AOI/TOI_cat	38.68	40.96	94.97	74.05

Next, as in our experiments with the regular SVM [2], we wanted to assess how the one-class SVM will perform on a different 'off-topic' domain. We used the same Training set, and the ON-topic part of the Testing set. For the OFF- part of the Testing set, the documents were retrieved from the Insider Threat dataset with queries on the topic of 'China/Nuclear weapons'.

Experimental results (Table 3) support the trend observed in the prior experiments. One-class categorization using the NLP-enhanced features achieves superior performance, particularly on the 'off-topic' class, compared to the baseline (BOW). Besides, the domain change for the 'off-topic' documents does not impact significantly the classifier performance, which was the case with the regular SVM. Such robustness is quite reasonable, since the one-class SVM is not biased (via training) towards a particular kind of 'negative' data.

Table 3. Experimental results (OFF-topic documents drawn from the 'China/Nuclear' domain)

	Features	Prec ON	Rec ON	F ON, β=1	Prec OFF	Rec OFF	F OFF, β=10
BOW	19774	94.60	68.68	79.58	14.13	56.77	44.55
CAT	10682	95.93	57.20	71.67	13.44	73.23	52.14
AOI/TOI	403	94.99	53.25	68.24	11.82	69.03	47.94
AOI/TOI_cat	2149	96.83	70.61	81.67	18.70	74.52	58.61

Overall, the results show that the one-class SVM performs impressively well, especially, on recall of the OFF-topic class. Another important point is that the algorithm appears to be robust to handle different subject domains of 'negative' examples. We believe, therefore, that it can be effectively applied to categorization problems where only 'positive' examples are available. The results also demonstrate that the use of NLP-based features outperforms BOW with many fewer features.

6 Conclusion and Directions for Future Research

The experiments described herein show that leveraging one-class SVM with the NLP-extracted features for document representation improves classification effectiveness and efficiency. In future research we will seek to evaluate the impact of different combinations of linguistic features, extractions from text, and concepts inferred from external knowledge bases on categorization accuracy. To further explore the robustness of the one-class classifier, we plan to test it on a combination of different subject domains for the 'off-topic' class.

The one-class approach fits particularly well the situations where it is not feasible to provide 'atypical' examples. Overall, the research reported herein holds potential for providing the IC with the analytic tools to recognize anomalous insider activity; as well as to build content profiles of vast document collections when applied in a broader context.

Acknowledgements

This work was supported by the Advanced Research and Development Activity (ARDA).

References

1. S. Symonenko, E. D. Liddy, O. Yilmazel, R. DelZoppo, E. Brown, and M. Downey, "Semantic Analysis for Monitoring Insider Threats," ISI2004.
2. O. Yilmazel, S. Symonenko, E. D. Liddy, and N. Balasubramanian, "Improved Document Representation for Classification Tasks For The Intelligence Community (Forthcoming)," AAAI 2005 Spring Symposium.
3. R. DelZoppo, E. Brown, M. Downey, E. D. Liddy, S. Symonenko, J. S. Park, S. M. Ho, M. D'Eredita, and A. Natarajan, "A Multi-Disciplinary Approach for Countering Insider Threats," SKM Workshop, 2004.
4. J. Anderson, "Computer Security Threat Monitoring and Surveillance", 1980.
5. R. H. Lawrence and R. K. Bauer, "AINT misbehaving: A taxonomy of anti-intrusion techniques," 2000.
6. D. V. Sreenath, W. I. Grosky, and F. Fotouhi, "Emergent Semantics from Users' Browsing Paths," ISI 2003.
7. J. Burgoon, J. Blair, T. Qin, and J. Nunamaker, Jr., "Detecting Deception Through Linguistic Analysis," ISI 2003.
8. L. Zhou, J. K. Burgoon, and D. P. Twitchell, "A Longitudinal Analysis of Language Behavior of Deception in E-mail," ISI 2003.
9. D. P. Twitchell, J. F. Nunamaker Jr., and J. K. Burgoon, "Using Speech Act Profiling for Deception Detection," ISI 2004.
10. [K. A. Heller, K. M. Svore, A. Keromytis, D., and S. J. Stolfo, "One Class Support Vector Machines for Detecting Anomalous Windows Registry Accesses" DMSEC'03.
11. H. Yu, C. Zhai, and J. Han, "Text Classification from Positive and Unlabeled Documents," CIKM'03.
12. L. M. Manevitz and M. Yousef, "One-class SVMs for Document Classification," *The Journal of Machine Learning Research*, vol. 2, pp. 139-154, 2002.
13. K.-M. Schneider, "Learning to Filter Junk E-Mail from Positive and Unlabeled Examples," 2004.
14. B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," 1999.
15. B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, "Building Text Classifiers Using Positive and Unlabeled Examples," ICDM'03.
16. S. Dumais, P. John, D. Heckerman, and M. Sahami, "Inductive Learning Algorithms and Representations for Text Categorization," Seventh International Conference on Information and Knowledge Management, 1998.
17. F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, pp. 1-47, 2002.
18. E. D. Liddy, "Natural Language Processing," in *Encyclopedia of Library and Information Science*, 2nd ed. 2003.
19. Center for Natural Language Processing (CNLP), www.cnlp.org.
20. Center for Nonproliferation Studies (CNS), http://cns.miis.edu/.

21. B. Sundheim and R. Irie, "Gazetteer Exploitation for Question Answering: Project Summary," 2003.
22. Y. Yang and X. Liu, "A Re-Examination of Text Categorization Methods," SIGIR'99.
23. T. Joachims, *Learning to Classify Text using Support Vector Machines.* 2002.
24. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001.
25. C. J. van Rijsbergen, *Information Retrieval*. 1979.

LSI-Based Taxonomy Generation: The Taxonomist System

Janusz Wnek

Content Analyst Company, LLC,
11720 Sunrise Valley Drive,
Reston, VA 20191, USA
jwnek@ContentAnalyst.com

Abstract. The following presents a method for constructing taxonomies by utilizing the Latent Semantic Indexing (LSI) technique. The LSI technique enables representation of textual data in a vector space, facilitates access to all documents and terms by contextual queries, and allows for text comparisons. A taxonomy generator downloads collection of documents, creates document clusters, assigns titles to clusters, and organizes the clusters in a hierarchy. The nodes in the hierarchy are ordered from general to specific in the depth of the hierarchy, and from most similar to least similar in the breadth of the hierarchy. This method is capable of producing meaningful classifications in a short time.

1 Introduction

A taxonomy is a hierarchical classification of objects. At the root of the hierarchy is a single classification of all objects. Nodes below the root provide classifications of subsets of objects. The objects in the subsets are grouped according to some selected object properties. In constructing a taxonomy, these properties allow for grouping similar objects and distinguishing them from others. In applying a taxonomy to classify objects, the properties allow identification of proper groups to which the objects belong.

The best-known taxonomy is the taxonomy of living things that was originated by Carl Linnaeus in the 18th century. In his taxonomy of plants, Linnaeus focused on the properties of flower parts, which are least prone to changes within the category. This taxonomy enabled his students to place a plant in a particular category effortlessly.

Linnaean taxonomies came into use during a period when the abundance of the world's vegetation was being discovered at a rate that exceeded the regular means of analyzing and organizing the newly found species. In this *age of information*, we face a similar need for investigating, organizing and classifying information from a variety of media sources and formats. Content analysis has become critical for both human advancement and security. The rapid identification and classification of threats has become a priority for many agencies and, therefore, new taxonomies of security related information are sought in order to quickly recognize threats and prepare proper responses.

The challenge of analyzing large amounts of information is multiplied by a variety of circumstances, locations and changing identities among the entities involved. It is

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 389 – 394, 2005.

not feasible to build one classification system capable of meeting all current needs. Constant adaptation is required as soon as new information becomes available. Therefore, classification systems require automation for detecting new patterns and providing specific and understandable leads. Automation in this case means that the system learns patterns in an unsupervised fashion and organizes its knowledge in a comprehensive way. Such is the purpose of the Taxonomist System.

The Taxonomist System employs a well-known information retrieval technique called Latent Semantic Indexing (LSI) [1] to efficiently index all documents required for analysis. LSI was designed to overcome the problem of mismatching words of queries with words of documents, as evident in Boolean-query type retrieval engines. In fact, LSI can be used to find relevant documents that may not even include any of the search terms with a query. LSI uses a vector space model that transforms the problem of comparing textual data into a problem of comparing algebraic vectors in a multidimensional space. Once the transformation is done, the algebraic operations are used to calculate similarities among the original documents, terms, groups of documents and their combinations.

There have been numerous applications of Latent Semantic Indexing and analysis to document processing. They include information retrieval [1], cross-language information retrieval [2], text segmentation [3], and supervised text categorization [4]. Recently, the algebraic technique of Singular Value Decomposition (SVD) employed in LSI was used for clustering search results and assigning titles to groups of results in the LINGO method [5]. That method extracts frequent phrases from input documents, constructs a term-document matrix, and applies SVD. The vectors of the orthogonal basis are assumed to be abstract concepts detected in documents, and thus they become topics. The topics are then matched to the frequent phrases. Finally, documents are assigned to the discovered topics and presented as non-hierarchical categorizations. In another application of LSI, hierarchical taxonomies are constructed using LSI vector representation spaces [6]. This method generates useful taxonomies, but is not scaleable due to the complexity of the clustering algorithm.

2 Overview of the Taxonomist System

The input to the Taxonomist System is in the form of a repository of documents indexed by LSI and a set of high-level parameters. The output is in the form of a hierarchy of topics – represented in XML – with links to the original documents. A recursive process that constructs nodes at the consecutive levels of the hierarchy, employs three major algorithms: document clustering and cluster merging; topic generation; and topic-title-based cluster separation.

Before indexing, the source documents are preprocessed by a pipeline of filters. The pipeline may contain filters for stop-word and stop-phrase removal, HTML/XML tagging removal, word stemming, and a pre-construction of generalized entities. A generalized entity is a semantic unit of one or more stemmed words extracted from the source documents with the exclusion of stop-words. During the preprocessing, words and word pairs (bi-words) are collected and used in indexing the repository. The bi-words are later used in reconstructing longer phases in the process of topic title construction through a combination of bi-words that share the same word and have a

similar frequency of occurrence within a document. This enables both an efficient indexing of generalized entities and a reconstruction of original phrases of any length.

Document clustering is realized by grouping together documents that are alike in terms of cosine similarity measured among vectors representing the documents. For a given level in the hierarchy, the clustering algorithm inputs the number of required nodes and incrementally samples the cosine similarity level to generate the minimum number of required nodes. For simplicity, the similarity is mapped into a range between 0 and 100%. The sampling starts from 10% similarity and is incrementally increased. At this level, the clusters are rather general and may not satisfy the required number of nodes. A higher similarity induces more clusters, which are more specific and tend to create less overlap. Overlapping clusters generated with high similarities could occur when documents describe multiple topics.

Once the required similarity level has been established and the minimum number of clusters has been generated, the method determines which clusters are similar to each other. For this purpose, vectors representing all clusters are generated and similar clusters are merged if their representation vectors exceed some similarity threshold. That threshold is set at the same level as the similarity for which clusters were generated. After the merge phase, the resulting number of clusters is verified against the required minimum. If the minimum is not satisfied, the similarity threshold is increased and the cluster and merge phase repeated. Otherwise, the system enters the topic generation phase.

The topic title generation process consists of three stages: (1) collecting candidate terms for the topic titles; (2) evaluating and selecting the best candidates; and (3) constructing the title from the best candidate terms. One hundred candidate terms are primarily retrieved from the LSI index by constructing combined (centroid) vectors from the document vectors representing the topic and querying the term database. An auxiliary method for adding candidate terms is achieved by downloading samples of the source documents represented by topic then extracting terms from the collection.

The best terms are selected based on statistics that evaluate terms in the context of the whole document repository as well as on the intra- and inter-node evaluations. (The statistics pertaining to the whole repository are readily available as meta-data from the LSI indexing engine implementation.) The first evaluation is based on the term's frequency across the whole repository. It is a requirement that the term occurs in at least 40% of the node documents. This test is especially efficient in evaluating very specific terms because it is not expensive. Terms that pass the first test are then tested against the samples of documents from the node. In order to pass this intra-node test, the term must be present in at least 40% of the sampled documents. The third test is the most expensive because it involves all sibling nodes. It generally rejects those terms that occur in more than one node. However, a node may keep the term if that node dominates other nodes.

In the third stage of topic title generation, the best terms are combined into titles. Topic titles are generated based on the identification and utilization of generalized entities implied by the best terms. First, all of the top terms are sorted according to their frequencies within the node documents. If the term having the best coverage is a part of a multi-word generalized entity, we then examine individual words within the bi-word, searching for a fitting bi-word with the overlapping word. Those matching bi-words must have similar coverage in order to be re-constructed into a generalized

entity. Subsequently, the final title is reconstructed from the generalized entity that reflects the most common usage among the documents in the node. This way, the original letter cases and postfixes removed in the stemming process, as well as the stop-words, are restored.

The clusters resulting from the cluster-merge operation may overlap, especially when generated at the lower levels of similarity. We reduce such ambiguity by applying context of topic titles to decide in which clusters the overlapping documents should ultimately be placed. In order to perform disambiguation, each document shared by two or more clusters is compared to the titles of the shared documents. The documents are classified into two categories: (1) not similar to any topic; and (2) similar to at least one topic. Documents in the latter category are kept within the clusters. The unmatched documents in the former category are removed from the existing clusters and the complete clustering-merge-title-generation process is repeated until all documents are assigned clusters and titles that fit the document. At the end of each iteration, the newly produced topic titles are compared to the earlier topic titles. Those topics with titles that match with a high degree of similarity (above 40%) are combined and the remaining topics are added as separate clusters.

3 Example of an Automatically Generated Taxonomy

In this section, we illustrate certain features of the automatically generated taxonomy from the R-9133 collection, a subset of Reuters-21578 [7]. The documents in the Reuters-21578 collection are classified into 66 categories, with some documents belonging to more than one category. R-9133 contains 9,133 documents with only a single category assigned.

The Reuters-21578 documents are related to trade, acquisitions, earnings, money exchange and supply, and market indicators. The automatically generated taxonomy reflects closely human categorization. For example, the largest category, "earnings," is represented by the top four largest topics emphasizing different aspects of earnings reports: (1) reports of gains and losses in cents in comparable periods; (2) payments of quarterly dividends; (3) expected earnings as reported quarterly; and (4) board decisions for splitting stock.

Besides the grouping offered by the clustering algorithm, the topic titles are indicative of underlying relationships among objects described in the documents. Acronyms are often explained by full names (e.g., "Commodity Credit Corporation, CCC," "International Coffee Organization, ICO," "Soviet Union, USSR"). Correlated objects are grouped under one topic title (e.g. "Shipping, Port, Workers," "GENCORP, Wagner and Brown, AFG Industries," "General consensus on agriculture trade, GATT, Trade Representative Clayton Yeutter").

Table 1 shows topic #24 with its subtopics. These subtopics are ordered according to similarity between the represented documents and the topic title. For example, the first subtopic, which consists of 9 documents, is similar to the topic title at 69%.

The hierarchical arrangement of topics reveals certain types of relationships between represented objects as in the PART-OF relationships, or general to specific descriptions (Fig. 1). It also demonstrates an important functionality of the term disambiguation. For example, words "Minister" and "Ministry" are used several times

in the hierarchy. In the first layer of the hierarchy, three clusters capture separate topics related to these words. The word "Minister," which is placed in the context of "Gulf, Kuwait," is further instantiated to two "Oil Ministers." (Fig 2).

Table 1. Subtopics generated for the "Gulf, KUWAIT, Minister" topic

#	**Topic Title** ***SIM*** **| Subtopic title**	**Doc Cnt**	Human-assigned categories and document counts
24	**Gulf, KUWAIT, Minister**	**63**	crude.32 ship.24 money-fx.4 earn.1 acq.1 pet-chem.1
69	Saudi Arabia and the United Arab Emirates, Gulf Cooperation Council	**9**	money-fx.4 crude.3 pet-chem.1 ship.1
65	Shipping, OIL PLATFORM, ATTACKED	**41**	ship.22 crude.19
57	Oil Minister Gholamreza Aqazadeh, QASSEM AHMED TAQI, Iranian news agency	**6**	crude.6
54	OPEC, Prices, Oil Minister Sheikh Ali al-Khalifa al-Sabah	**10**	crude.10
50	Strategic Straits of Hormuz, Warships, Patrols	**4**	ship.4
45	Assets of nine community papers, Gulf Coast, SCRIPPS	**1**	acq.1
45	GULF STATES UTILITIES, Issued a qualified opinion, Auditor Coopers and Lybrand	**1**	earn.1
17	Decided to renew its one-year contract with Abu Dhabi, Supply of tonnes of Gulf of Suez	**1**	crude.1

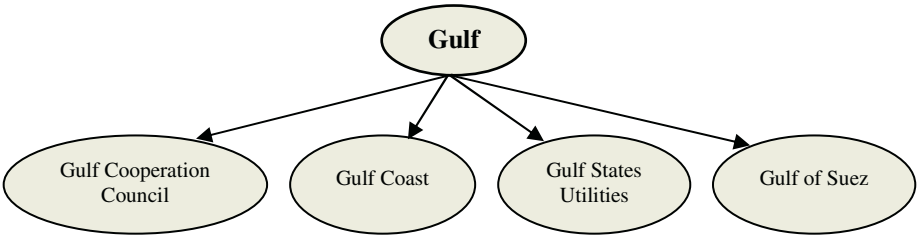


Fig. 1. Specific instances of a more general concept

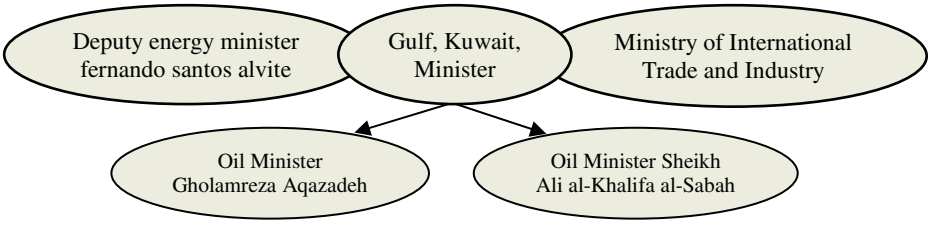


Fig. 2. Disambiguation of word meanings. Correlation of concepts

4 Experimental Results and Conclusion

Table 2 presents statistics characterizing scalability of the Taxonomist System as a function of: (1) the number of documents in a collection; (2) hierarchy depth; and (3) the number of nodes generated. The automatically generated taxonomies produced by the Taxonomist System matched human-produced categorizations in terms of accuracy and definitely exceeded efficiency in producing taxonomies. It took minutes for the system to produce a meaningful taxonomy of a sizable collection of documents, versus weeks required by humans.

Table 2. Scalability of the Taxonomist Method

Document Set & Size	Number of layers	Number of nodes	Time
R – 9,133	1	82	75 s
	2	133	90 s
Reuters - 21,578	1	88	160 s
	2	155	224 s
News – 164,095	1	98	30 m
	2	897	96 m
News – 804,414	1	145	2.5 h

Over the past three centuries, taxonomies have proven very useful in organizing large collections of data. The computer-generated taxonomies built from the textual data also demonstrated similar potential, as they exhibited many appealing features. They summarize, categorize and prioritize the textual material for efficient browsing from the general to more specific groupings. One important characteristic of the taxonomy structure is its effectiveness in resolving ambiguities. The same property can be supplemented by the context of events, locations and other circumstances that enable us to distinguish one John Smith from another. Transforming raw textual material into the rich knowledge structure creates new opportunities for analysts to rapidly investigate larger amounts of data in an organized manner.

References

1. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by Latent Semantic Analysis. J. Am. Soc. Inf. Sc. 46(4) (1990) 391-407
2. Landauer, T., Littman, M.: Fully Automatic Cross-language Document Retrieval Using LSI. In: Proc. 6th Conf. UW Centre New Oxford English Dict. and Text Research (1990) 31-38
3. Choi, F.Y.Y., Wiemer-Hastings, P., Moore, J.: Latent Semantic Analysis for Text Segmentation. In: Lee,L., Harman,D., (eds.) Proc. Conf. Emp. Meth in NLP (2001) 109-117
4. Dumais, S.: Using LSI for Information Retrieval, Information Filtering, and Other Things, In: Proc. of the Cognitive Technology Workshop, (1997)
5. Osinski, S., Stefanowski, J., Weiss, D.: Lingo - Search Results Clustering Algorithm Based on Singular Value Decomposition. In: Proc. IIS: IIPWM Conf. (2004) 359-368
6. Wnek, J.: Document Ontology Discovery Tool. In: Proc. IIS: IIPWM Conf. (2004) 475-480
7. Lewis, D.D.: Reuters-21578 Text Categorization Test Collection. Distribution 1.0 (1999)

Some Marginal Learning Algorithms for Unsupervised Problems

Qing Tao[1], Gao-Wei Wu[2], Fei-Yue Wang[1], and Jue Wang[1]

[1] The Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, P. R. China
qing.tao@mail.ia.ac.cn
[2] Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, P. R. China

Abstract. In this paper, we investigate one-class and clustering problems by using statistical learning theory. To establish a universal framework, a unsupervised learning problem with predefined threshold η is formally described and the intuitive margin is introduced. Then, one-class and clustering problems are formulated as two specific η-unsupervised problems. By defining a specific hypothesis space in η-one-class problems, the crucial minimal sphere algorithm for regular one-class problems is proved to be a maximum margin algorithm. Furthermore, some new one-class and clustering marginal algorithms can be achieved in terms of different hypothesis spaces. Since the nature in SVMs is employed successfully, the proposed algorithms have robustness, flexibility and high performance. Since the parameters in SVMs are interpretable, our unsupervised learning framework is clear and natural. To verify the reasonability of our formulation, some synthetic and real experiments are conducted. They demonstrate that the proposed framework is not only of theoretical interest, but they also has a legitimate place in the family of practical unsupervised learning techniques.

1 Introduction

In the last few years there have been very significant developments in the understanding of statistical learning theory and SVM (Support Vector Machine)([1], [2] and [3]). Statistical learning theory, which focuses on the induction and statistical inference from data to distribution, provides theoretical guarantees for good generalization ability of learning algorithms. As the first theoretically motivated statistical learning algorithm, SVM can minimize the expected risk in PAC

[1] This paper is supported by National Basic Research Program of China (2004CB318103). The first author is also supported by the Excellent Youth Science and Technology Foundation of Anhui Province of China (04042069).

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 395–401, 2005.

frame (Probably Approximately Correct, see [4]). The wonderful statistical learning nature in SVM inspires us to reconsider many learning problems in pattern recognition [5]. In this paper, we investigate one-class and clustering problems.

Within the last decade, both theory and practice have pointed to the concept of the margin of a classifier as being central to the success of a new generation of learning algorithms. This is explicitly true of SVMs, which in their simplest form implement maximal margin hyperplanes, but has also been shown to be the case for boosting algorithms such as Adaboost ([6]). Increasing the margin has been shown to improve the generalization performance ([7] and [3]) in PAC framework. The main idea in this paper is that we define a η-unsupervised learning problems with a intuitive margin and use the margin to design and analyze algorithms for general unsupervised learning, especially one-class and clustering problems. This may be the most important contribution of this paper.

The formal description of a unsupervised learning problem has been presented in [8] and [9]. But it seems that there are no available clustering algorithms like SVMs, in which the good generalization ability is obtained by the data-dependent algorithms. Can we generalize the idea in SVM to get some clustering algorithms with robustness, flexibility and high generalization performance? On the other hand, one-class problem can be viewed as a particular unsupervised problem. Its generalization follows directly from unsupervised problems. Can the available one-class learning algorithms minimize the expected risk in PAC frame? These are the main motivation of this paper.

2 The η-Unsupervised Learning Problems and Margin

Let η be a threshold, which represents the scale of the desired region for a unsupervised learning problem. Consider the following unsupervised problem

$$S = \{x_1, x_2, \ldots, x_l, \quad x_i \in R^N, i = 1, 2, \ldots, l.\} \tag{1}$$

where the samples are independently drawn and identified distributed according to a unknown density function $D(x)$. Let Z be an index set and $H = \{f : \quad f : Z \longrightarrow R^n\}$ denote the hypothesis space. For $p \in [1, \infty)$ and $x \in R^N$, $\|x\|_p$ represents the l^p norm of x.

Definition 2.1 (Loss function and the expected risk). Let $f \in H$. Let $\triangle(x, f(z))$ be a non-negative cost function to measure the closeness between x and $f(z)$. Let L be defined as follows:

$$L(\eta, x, f) = \begin{cases} 0, & if \quad min\{\triangle(x, f(z)) : z \in Z\} \leq \eta; \\ 1, & otherwise. \end{cases}$$

Then L is called the loss function of f about an η-unsupervised learning problem, and

$$errD(f, \eta) = \int L(\eta, x, f) D(x) dx \tag{2}$$

is called the expected risk of f about the η-unsupervised learning problem

Definition 2.2 (Empirical risk). Let $f(x) : R^N \to R$ and $f \in H$. Then

$$err_{emp}(f, \eta) = \sum_{i=1}^{l} L(\eta, x_i, f)$$

is called the empirical risk of f about the η-unsupervised learning problem.

Definition 2.3 (The optimal classifier and outlier of the η-unsupervised problem). Let $f_0 \in H$. If

$$err(f_0, \eta) = min\{\int L(\eta, x, f)D(x)dx, f \in H\}$$

Then f_0 is called the optimal classifier. The outlier of f_0 in $\{x_1, x_2, \ldots, x_l,\ x_i \in R^N, i = 1, 2, \ldots, l.\}$ is called an outlier of η-unsupervised learning problem.

Definition 2.4 (Margin). Let $f \in H$.The margin of a sample x_i $(i = 1, 2, \ldots, l.)$ is defined as $m(f, x_i, \eta) = \eta - min\{\triangle(x_i, f(z)) : z \in Z\}$. The margin of f is defined as $m(f, S, \eta) = min\{\eta - min\{\triangle(x_i, f(z)) : z \in Z\}, i = 1, 2, \ldots, l.\ \}$.

Definition 2.5 (Separable). If there exists a $f \in H$ such that $\eta - min\{\triangle(x_i, f(z)) : z \in Z\} \geq 0$ for $i = 1, 2, \ldots, l.$, the η-unsupervised learning problem is called separable.

Definition 2.6 (Maximum margin classifier). For a separable problem, if $f_0 \in H$ satisfies

$$m(f_0, S, \eta) = max\{m(f, S, \eta), f \in H\}$$

Then f_0 is called the maximum margin classifier.

3 Some Marginal Algorithms for One-Class Problems

Obviously, an η-one-class problem can be naturally obtained from η-unsupervised learning problems by setting $Z = \{1\}$. In this section, we discuss η-one-class problems under two different hypothesis spaces.

3.1 The Minimal Sphere One-Class problems

To find out the relationship between our marginal algorithms and the available one-class learning algorithms in [10], [12] and [13], we first specify H to be the set of all constant functions and define $\triangle(x_i, f) = \|x_i - f\|$. Now, it is easy to know that the expected risk (2) is the total probability that a point is outside a ball.

Let the η-unsupervised learning problem under the above assumptions be separable, the margin of a "classifier" x_0 now is $\gamma = min\{\eta - \|x_i - x_0\|, i = 1, 2, \ldots, l.\}$. Then the maximum margin algorithm is formulated as the following optimization problem:

$$\begin{cases} max_{\{x_0 \in R^n\}} \ \gamma \\ \eta - \|x_i - x_0\| \geq \gamma, \quad i = 1, 2, \ldots, l. \end{cases} \tag{3}$$

Obviously, problem (3) is equivalent to

$$\begin{cases} min_{\{R,x_0\}} & R^2 \\ \|x_i - x_0\|^2 \leq R^2, & i = 1, 2, \ldots, l. \end{cases} \tag{4}$$

Obviously, problem (4) is just the minimal sphere algorithm in [12] and [13]. However, this paper may be the first attempt to show that the minimal sphere optimization itself is a specific maximum margin algorithms.

3.2 The Minimal Slab One-Class Problems

To produce some new algorithms using our defined margins, a straightforward and the most natural way is to define some different hypothesis spaces. Motivated by the hypothesis spaces in SVMs, in this section, we set $H = \{w^T x + b : w \in R^n, b \in R^1, \|w\|_p = 1\}$ and define $\triangle(x_i, w^T x + b) = |w^T x_i + b|$. Now, it is easy to know that the expected risk (2) is the total probability that a point is outside a slab.

Let the one-class problem under the above assumptions be separable, the margin of x_i now is $\gamma = \eta - |w^T x_i + b|$.

Clearly, the maximum margin algorithm for a separable problem is to solve the following optimization problem:

$$\begin{cases} max_{\{w,b,\gamma\}} & \gamma \\ \|w\|_p = 1, & w^T x_i + b + \eta \geq \gamma, \quad w^T x_i + b - \eta \leq -\gamma, \quad i = 1, 2, \ldots, l. \end{cases} \tag{5}$$

Let $\rho = \eta - \gamma$, the optimization problem (5) becomes

$$\begin{cases} min_{\{w,b,\rho\}} & \rho \\ \|w\|_p = 1, & w^T x_i + b \leq \rho, \quad w^T x_i + b \geq -\rho, \quad i = 1, 2, \ldots, l. \end{cases} \tag{6}$$

The interpretation of (6) is very clear, i.e., it is to obtain a minimal width zone that contains all the samples. Contrast to problem (4), the zone in problem (6) is a slab other than a ball and the center of the slab is a line other than a point.

If we let $p = 1$ in problem (4) and employ the idea in [10] and [11], we can obtain a new algorithm for one-class problems. The details can be found in [14].

4 A New Algorithm of Clustering Problems

One of the most widely used algorithms for constructing locally optimal quantizers for distributions or data is the Generalized Lloyd (GL) algorithm (also known as the k-means algorithm. In this section, we denote $\triangle(x, y) = \|x - y\|^2$.

Since the margin for clustering can directly follow from that in η-unsupervised problems, it is easy to obtain some marginal algorithms. Unfortunately even in empirical loss cases, the global optimality is not feasible ([8]). So in this paper, we can only limit our discussion to a direct modification of k-means algorithms. In fact, our new clustering algorithms only replace the sample averages by the center of corresponding minimal sphere. We describe the details in the following.

The Minimal Sphere GL Algorithm

Step 0. Set $j = 0$ and set $C^{(0)} = \{v_1^0, v_2^0, \cdots, v_k^0\}$ to an initial codebook.
Step 1 (Partition). Construct $V^j = \{V_1^j, V_2^j, \cdots, V_k^j\}$ by setting

$$V_l^{(j)} = \{x : \triangle(x, v_l^j) = \|x - v_l^j\|^2 \leq \triangle(x, v_m^j), m = 1, 2, \cdots, k\} \text{ for } l = 1, 2, \cdots, k.$$

Step 2 (Expectation). Construct $C^{(j+1)} = \{v_1^{j+1}, v_2^{j+1}, \cdots, v_k^{j+1}\}$. v_l^{j+1} is obtained by solving optimization problem (4) on the training set V_l^j for $l = 1, 2, \cdots, k$.
Step 3. Stop if $1 - \dfrac{\triangle_l(q^{j+1})}{\triangle_l(q^j)}$ is less than or equal to a certain threshold, where $\triangle_l(q^j) = \dfrac{1}{l}\sum_{i=1}^{k}\sum\{\|v_i - x\|^2 : x \in V_i\}$. Otherwise, let $j = j+1$ and go to Step 1.

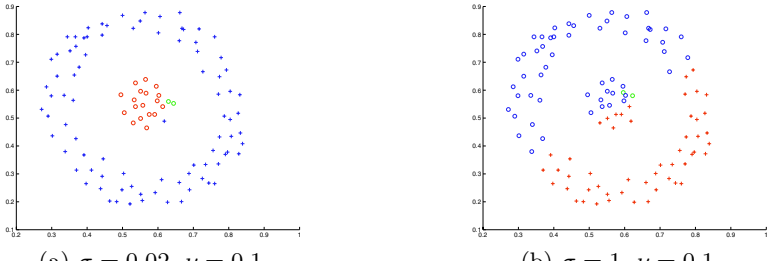
Obviously, if $k(x, y)$ is used instead of $\|x - y\|^2$ in Step 1 and Step 3, we can obtain kernel-based minimal sphere clustering algorithms. Compared with the work in [15] and [16], our contribution is that we establish a complete SVM formulation for clustering problems. Then, the robustness and flexibility can be achieved by using the soft idea and kernel techniques in SVMs.

5 Experiments

Example 5.1. To illustrate the performance of our kernel minimal sphere GL algorithms intuitively, we design an synthetic example. Please see Figure 1.

From Figure 1, it is easy to find that the different kernel parameters have different effects on the clustering. They demonstrate to some extent the performance of our kernel algorithms.

Example 5.2. Some clustering results on a real data set. The initial clusters are selected using centroid plus some random permutations. The kernel minimal sphere GL algorithm is run on Iris data set for 30 times. $\sigma = 0.1$ and $\nu = 0.01$. The averaged error rate is 9.73%.



(a) $\sigma = 0.02$, $\nu = 0.1$ (b) $\sigma = 1$, $\nu = 0.1$

Fig. 1. The kernel minimal sphere GL algorithm applied to the ring data, where the green points are outliers

As a summary of our clustering experiments, we indicate that our new algorithms can achieve the same effect as the other available clustering algorithms. But here, the soft margin parameter can determine the number of outliers and kernel parameter can decide the shape of clusters.

6 Conclusions

Inspired by the statistical nature of statistical learning theory in SVMs, we define an η-unsupervised learning problem and introduce the concept of margin. As a result, the available one-class optimization problems are essentially margin algorithms. Moreover, our new framework of unsupervised learning problems enable us to achieve some new margin algorithms for one-class and clustering problems.

References

1. V. Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
2. V. Vapnik. *Statistical Learning Theory*. Addison-Wiley, 1998.
3. N. Cristianini and J. Schawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge: Cambridge Univ Press. 2000.
4. L. G. Valiant. A Theory of the Learnable. Communications of the ACM, 1984, 27(11): 1134-1142.
5. R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification*. Second Edition, John Wiley & Sons. 2001.
6. R. Schapire, Y. Freund, P. Bartlett, and W. Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. Ann. Statist. 1998, 26(5): 1651-1686.
7. G. Rätsch. *Robust boosting via convex optimization*. Ph. D. thesis, University of Posdam. 2001.
8. B. Kégl. *Principal curves: learning, design, and applications*. Ph. D. Thesis, Concordia University, 1999.
9. A. J. Smola, S. Mika, B. Schölkopf and R. C. Williamson. Regularized principal manifolds. Journal of Machine Learning Research. 2001, 1: 179-200.
10. B. Schölkopf, J. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson. Estimating the Support of a High-Dimensional Distribution. Neural Computation. 2001, 13(7): 1443-1471.
11. G. Rätsch, S. Mika, B. Schölkopf and K. R. Müller. Constructing boosting algorithms from SVMs: an application to one-class classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002, 9(4): 1184-1199.
12. D. Tax and R. Duin. Support vector domain description. Pattern Recognition Letters. 1999, 20: 1191-1199.
13. D. Tax and R. Duin. Support vector data description. Machine Learning, 2004, 54: 45-66.
14. Qing Tao, Gaowei Wu and Jue Wang. A new maximum margin algorithm for one-class problems and its boosting implementation. Pattern Recognition. 2005. (accepted).

15. M. Girolami. Mercer Kernel-Based Clustering in Feature Space. IEEE Trans Neural Networks. 2002, 13(3): 780-784.
16. A. Ben-Hur, D. Horn, H. T. Siegelmann and V. Vapnik. Support Vector Clustering. Journal of Machine Learning Research. 2001, 2: 135-137.
17. B. Schölkopf, A. J. Smola, R. Williamson and P. Bartlett. New support vector algorithms. Neural Computation, 2000, 12: 1083-1121. 225-254.

Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites

Edna Reid[1], Jialun Qin[1], Yilu Zhou[1], Guanpi Lai[2], Marc Sageman[3], Gabriel Weimann[4], and Hsinchun Chen[1]

[1] Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721, USA
{ednareid, qin, yiluz, hchen}@bpa.arizona.edu
[2] Department of Systems and Industry Engineering, The University of Arizona, Tucson, AZ 85721, USA
guanpi@email.arizona.edu
[3] The Solomon Asch Center For Study of Ethnopolitical Conflict, University of Pennsylvania, St. Leonard's Court, Suite 305, 3819-33 Chestmut Street, Philadelphia, PA 19104, USA
sageman@sas.upenn.edu
[4] Department of Communication, Haifa University, Haifa 31905, Israel
weimann@soc.haifa.ac.il

Abstract. The Internet which has enabled global businesses to flourish has become the very same channel for mushrooming 'terrorist news networks.' Terrorist organizations and their sympathizers have found a cost-effective resource to advance their courses by posting high-impact Websites with short shelf-lives. Because of their evanescent nature, terrorism research communities require unrestrained access to digitally archived Websites to mine their contents and pursue various types of analyses. However, organizations that specialize in capturing, archiving, and analyzing Jihad terrorist Websites employ different, manual-based analyses techniques that are inefficient and not scalable. This study proposes the development of automated or semi-automated procedures and systematic methodologies for capturing Jihad terrorist Website data and its subsequent analyses. By analyzing the content of hyperlinked terrorist Websites and constructing visual social network maps, our study is able to generate an integrated approach to the study of Jihad terrorism, their network structure, component clusters, and cluster affinity.

1 Introduction

Nowadays, the Internet has allowed terrorist groups to easily acquire sensitive intelligence information and control their operations [19]. Some research showed that terrorists use the Internet to develop a world-wide command, control, communication and intelligence system (C3I). For example, Jenkins posited that terrorists have used the Internet as a broadcast platform for the "terrorist news network" [12] which is an effective tactic because they can reach a broad audience with relatively little chance of detection.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 402 – 411, 2005.

Although this alternate side of the Internet, referred to as the Dark Web has recently received extensive government and media attention, our systematic understanding of how terrorists use the Internet for their campaign of terror is limited. According to studies by the Institute for Security and Technology Studies (ISTS) at Dartmouth College [11] and Anderson [2], there is a need to address this under-researched issue. In this study, we explore an integrated approach for harvesting Jihad terrorist Websites to construct a high quality collection of Website data that can be used to validate a methodology for analyzing and visualizing how Jihad terrorists use the Internet, especially the World Wide Web, in their terror campaigns. Jihad terrorist Websites are Websites produced or maintained by Islamic extremist groups or their sympathizers.

In this study, we answer 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 information on the Web so as to identify usage and relationships among groups? How do you conduct a content analysis of the Jihad terrorists' collection?

2 Previous Research

In this section, we briefly review related studies on collection and analyzing terrorist Websites.

2.1 Terrorist Websites

The Web has been intensively used by terrorist organizations for their advantages. Arquilla and Ronfeldt [3] described this trend as netwar, an emerging model of conflict in which the protagonists use network forms of organization and exploit information technology. Many studies have been conducted on analyzing the terrorists' use of the Web. Examples include Elison [9], Tsfati and Weimann [21], ISTS [11], and Weimann [22]. All of them used terrorists' and their sympathizers' Websites as their primary data sources and provided brief descriptions of their methodologies.

To ensure the researchers and experts have access to terrorist Websites for research and intelligence analysis, several organizations are collecting, archiving, and analyzing Jihad terrorist Websites. These organizations include: the Internet Archive, the Project for Research of Islamist Movements, (PRISM) at the Interdisciplinary Center Herzliya, the Jihad and Terrorism Studies Project at the Middle East Research Institute (MEMRI), the Search for International Terrorist Entities (SITE Institute), and Professor Gabriel Weimann's collection at the University of Haifa, Israel [17]. Although all of them manually capture and analyze terrorist Websites to publish research reports, none publish their specific collection building and analytical approaches. Except for using search engines to identify terrorist Websites, none of the organizations seem to use any other automated methodologies for capturing and analyzing terrorist Websites.

2.2 Automated Web Harvesting

Previous research from the digital library community suggested automatic approaches to harvesting WebPages in particular domains. Web harvesting is the process of gathering and organizing unstructured information from pages and data on the Web [13]. Albertsen [1] 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.

In this 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 [16]. Domain experts provide seed URLs as well as typologies for constructing metadata that can be used in the spidering process. Their project's goal is to develop a methodology for constructing an archive of broad-spectrum political communications over the Web. In contrast, for the September 11 and Election 2002 Web Archives projects, the Library of Congress' approach was to manually collect seed URLs for a given theme [18]. The seeds and their immediate neighbors (distance 1) are then crawled.

2.3 Web Link and Content Analysis

Web link analysis has been previously used to discover hidden relationships among commercial companies [10]. Borgman [4] defines two classes of web link analysis studies: 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. Gibson [10] describes a methodology for discerning Web communities on the WWW. His work is based on Hyperlink-Induced Topic Search (HITS), a tool that searches for authoritative hypermedia on a given broad topic. In contrast, we construct a Website topology from a high quality Jihad Terrorism collection.

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 ignored existing methodologies for conducting a systematic content analysis of Website data [21,11]. In Bunt's [5] overview of Jihadi movements' presence on the Web, he described the reaction of the global Muslim community to the content of terrorist Websites. Tsfati and Weimann's [21] study of terrorism on the Internet acknowledges the value of conducting a systematic and objective investigation of the characteristics of terrorist groups' communications. All the studies mentioned above are qualitative studies. 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.

Demchak and Friis' [8] work focused on measuring "openness" of government Websites using a Website Attribute System 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 Proposed Approach and Preliminary Results

We propose an integrated approach to the study of Jihad Terrorism Web infrastructure. We combined a sound methodology for constructing a high-quality Jihad terrorism collection, a hyperlink analysis for the study of Jihad terrorism group relationships, and a systematic content analysis to study the details on the terrorists' Web usage. In the following sub-sections, we will describe the details of our approach and report the preliminary results from our analysis.

3.1 Jihad Collection Building

A systematic and sound methodology for collecting the Jihad terrorism Websites guarantees that our collection, which is the cornerstone of the study, is comprehensive and representative. We take a three step systematic approach to construct the 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 search major search engines (google.com, yahoo.com …etc) using information such as the group names as queries to find Websites of these groups. Three Jihad terrorist Websites were identified: www.qudsway.com of the Palestinian Islamic Jihad, www.hizbollah.com of Hizbollah, and www.ezzedine.net which is a Website of the Izzedine-Al-Qassam, the military wing of Hamas. Then, we used Google back-link 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 adopts 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 the collection by means of manually search 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, 16 more Websites were identified and our final Jihad collection contains 39 terrorist Websites.

After identifying the Jihad terrorist Websites, we download all the Web pages within the identified sites. Our final collection contains more than 300,000 high-quality Web pages created by Jihad terrorists.

3.2 Link Analysis

Our goal here is to shed light on the infrastructure of Jihad Websites and to provide the necessary tools for further analysis of Jihad terrorist group relationships. 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 multivariable 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 multidimensional scaling (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 1. 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. Hizbollah community and the Palestinian militant groups' community were connected through hyperlink. Hizbollah has traditionally sympathized 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 over the Middle-East and in Europe. Although groups in this cluster are not officially recognized as terrorist groups, they do have links pointing to the Hizballah 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 Hezbollah's community as they follow radically different ideologies.

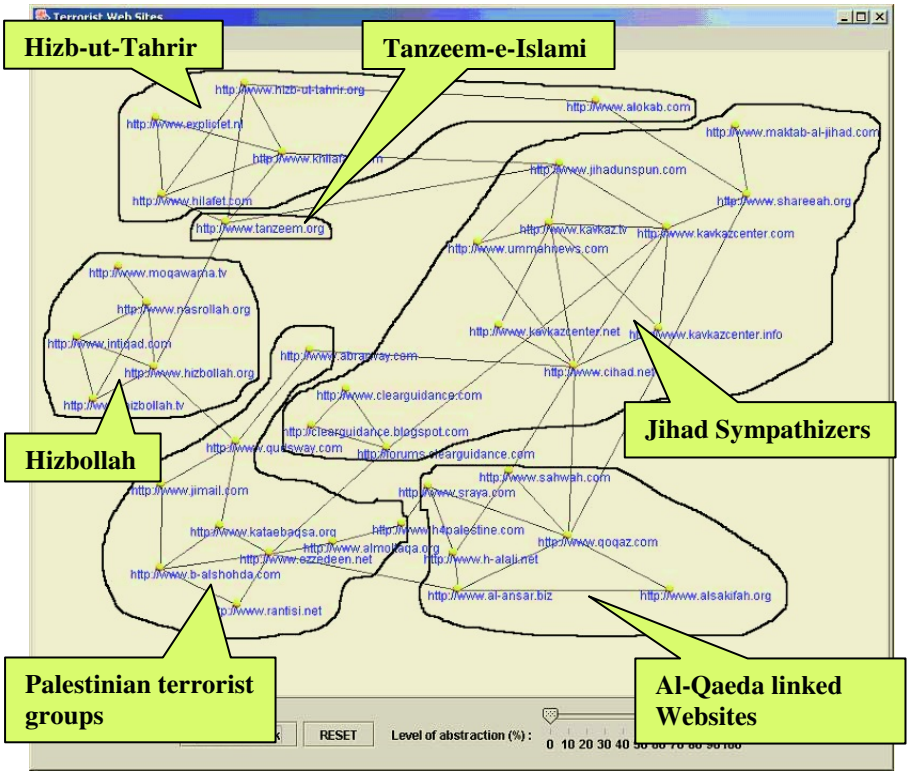


Fig. 1. The Jihad Terrorism Network with Automatically Generated Hyperlinked Communities

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.

3.3 Content Analysis

To complete our analysis of Jihad terrorism on the Web we propose a framework for content analysis. The framework consists of high level attributes, each of which is composed of multiple fine grained low level attributes. This approach is similar to what is presented in Demchak and Friis' study [8]. Table 1 shows the high level and associated low level attributes used in this study.

Table 1. Attributes used in the study

High Level Attribute	Low Level Attribute
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

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 record the presence of low-level attributes. The manual coding of the attributes in a Websites takes around 45 minutes of work. After completing the coding scheme for 32 Websites in the collection, we then compared the content of the clusters or hyperlinked communities in the network shown in Figure 2. We aggregated data from all Websites belonging to a cluster and displayed the result in snowflake diagrams. Figure 3 shows two such diagrams.

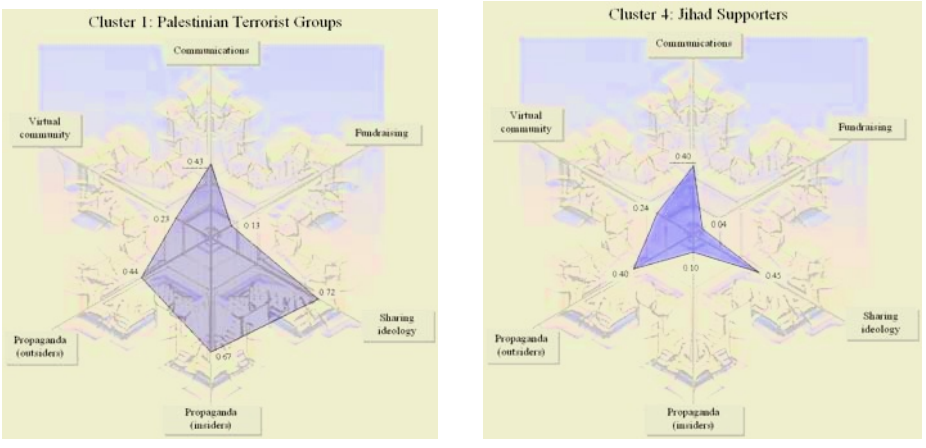


Fig. 2. Snowflake Diagrams for Palestinian terrorist groups and Jihad Supporters Web Communities

An interesting observation in these snowflake diagrams is the discrepancy in 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 an effective tool for influencing youth and possibly recruiting new members. Conversely, the sympathizers try to explain their views to outsiders (Westerners) and try to justify terrorist actions.

4 Discussion and Future Work

We have developed an integrated approach to the study of the Jihad terrorism Web Infrastructure. Hyperlinked communities' analysis brings an overall view of the terror web infrastructure. Visualizing hyperlinked communities facilitates the analysis of Web infrastructures and paves the way for more refined microscopic content analysis of the Websites. 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 Web pages. 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.

Acknowledgements

This research has been supported in part by the following grants:

- NSF, "COPLINK Center: Information & Knowledge Management for Law Enforcement," July 2000-September 2005.

- NSF/ITR, "COPLINK Center for Intelligence and Security Informatics Research – A Crime Data Mining Approach to Developing Border Safe Research," September 2003-August 2005.
- DHS/CNRI, "BorderSafe Initiative," October 2003-March 2005.

We would like to thank Dr. Joshua Sinai from the Department of Homeland Security, Dr. Rex A. Hudson from the Library of Congress, and Dr. Chip Ellis from the MIPT organization for their insightful comments and suggestions on our project. 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 Homa Atabakhsh, Cathy Larson, Chun-Ju Tseng, and Shing Ka Wu.

References

1. Albertsen, K.: The Paradigma Web Harvesting Environment. 3rd ECDL Workshop on Web Archives, Trondheim, Norway (2003)
2. Anderson, A.: Risk, Terrorism, and the Internet. Knowledge, Technology & Policy 16(2) (2003) 24-33
3. Arquilla, J., Ronfeldt, D.F.: Advent of Netwar. Rand Report (1996) http://www.rand.org/
4. Borgman, C. L., Furner, J.: Scholarly Communication and Bibliometrics. Annual Review of Information Science and Technology, ed. B. Cronin. Information Today, Inc (2002)
5. Bunt, G. R.: Islam In The Digital Age: E-Jihad, Online Fatwas and Cyber Islamic Environments. Pluto Press, London (2003)
6. Carley, K. M., Reminga, J., Kamneva, N.: Destabilizing Terrorist Networks. NAACSOS Conference Proceedings, Pittsburgh, PA (2003)
7. Carmon, Y.: Assessing Islamist Web Site Reports Of Imminent Terror Attacks In The U.S. MEMRI Inquiry & Analysis Series #156 (2003)
8. Demchak, C. C., Friis, C., La Porte, T. M.: Webbing Governance: National Differences in Constructing the Face of Public Organizations. Handbook of Public Information Systems, G. David Garson, ed., New York: Marcel Dekker Publishers (2000)
9. Elison, W.: Netwar: Studying Rebels on the Internet. The Social Studies 91 (2000) 127-131
10. Gibson, D., Kleinberg, J., Raghavan, P.: Inferring Web Communities from Link Topology: Proceedings of the 9th ACM Conference on Hypertext and Hypermedia (1998)
11. Institute for Security Technology Studies: Examining the Cyber Capabilities of Islamic Terrorist Groups. Report, ISTS (2004) http://www.ists.dartmouth.edu/
12. Jenkins, B. M.: World Becomes the Hostage of Media-Savvy Terrorists: Commentary. USA Today (2004) http://www.rand.org/
13. Kay, R.: Web Harvesting. Computerworld (2004) http://www.computerworld.com.
14. Kenney, A R., McGovern, N.Y., Botticelli, P., Entlich, R., Lagoze, C., Payette, S.: Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism. D-Lib Magazine 8(1) (2002)
15. Reid, E. O. F.: Identifying a Company's Non-Customer Online Communities: a Prototypology. Proceedings of the 36th Hawaii International Conference on System Sciences, (2003)
16. Reilly, B., Tuchel, G., Simon, J., Palaima, C., Norsworthy, K., Myrick, L.: Political Communications Web Archiving: Addressing Typology and Timing for Selection, Preservation and Access. 3rd ECDL Workshop on Web Archives, Trondheim, Norway (2003)

17. Research Community PRISM.: The Project for the Research of Islamist Movements. http://www.e-prism.org.MEMRI: Jihad and Terrorism Studies Project (2003)
18. SITE Institut : Report (2003) http://www.siteinstitute.org/mission.html.
19. Schneider, S. M., Foot, K., Kimpton, M., Jones, G.: Building thematic web collections: challenges and experiences from the September 11 Web Archive and the Election 2002 Web Archive. 3rd ECDL Workshop on Web Archives, Trondheim, Norway (2003)
20. Tekwani, S.: Cyberterrorism: Threat and Response. Institute of Defence and Strategic Studies, Workshop on the New Dimensions of Terrorism, Singapore (2002)
21. The 9/11 commission report (2004) http://www.gpoaccess.gov/911/
22. Tsfati, Y., Weimann, G.: www.terrorism.com: Terror on the Internet. Studies in Conflict & Terrorism 25 (2002) 317-332
23. Weimann, G.: www.terrorism.net: How Modern Terrorism Uses the Internet. Special Report 116, U.S. Institute of Peace (2004) http://usip.org/pubs/

Evaluating an Infectious Disease Information Sharing and Analysis System[1]

Paul J.-H. Hu[1], Daniel Zeng[2], Hsinchun Chen[2], Catherine Larson[2], Wei Chang[2], and Chunju Tseng[2]

[1] School of Accounting and Information Systems,
University of Utah
actph@business.utah.edu
[2] Department of Management Information Systems,
University of Arizona, Tucson, Arizona
{zeng, hchen,cal, weich, chunju}@eller.arizona.edu

Abstract. Infectious disease informatics is a subfield of security informatics that focuses on information analysis and management issues critical to the prevention, detection, and management of naturally occurring or terrorist-engineered infectious disease outbreaks. We have developed a research prototype called BioPortal which provides an integrated environment to support cross-jurisdictional and cross-species infectious disease information sharing, integration, analysis, and reporting. This paper reports a pilot study evaluating BioPortal's usability, user satisfaction, and potential impact on practice.

1 Introduction

Infectious disease informatics (IDI) systematically studies information management and analysis issues critical to infectious disease prevention, detection, and management. IDI has direct relevance to intelligence and security informatics because of the growing concern that terrorists may attack by deliberately transmitting infectious disease using biological agents [13]. From a perspective of both technology and task analysis support, IDI faces a similar set of challenges concerning cross-jurisdictional data sharing, data integration, data ownership and privacy, information overload, data summarization and visualization, as most other security informatics subfields.

We have established an interdisciplinary research team (consisting of IT researchers, public health researchers, and state public health departments) to address various data sharing and analysis challenges facing IDI. A research prototype called BioPortal has been developed to provide distributed, cross-jurisdictional access to datasets concerning several major infectious diseases, together with spatio-temporal data analysis methods and visualization capabilities.

The technical aspects of BioPortal have been discussed previously [14] [15]. This paper reports a pilot study evaluating BioPortal's usability, user satisfaction, and potential impact on public health and bioterrorism monitoring practice. This and

[1]. Reported research has been supported in part by the NSF through Digital Government Grant #EIA-9983304 and Information Technology Research Grant #IIS-0428241.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 412 – 417, 2005.

subsequent user studies not only provide valuable feedback to our system design and development but also shed light on broader technology adoption and user satisfaction issues critical to system usage and sustainable system success [10].

2 BioPortal System Overview and Main Functionality

BioPortal (www.bioportal.org) provides an integrated, cross-jurisdictional infectious disease information infrastructure which has been available for research and testing purposes since early 2004. Although not yet operational, BioPortal contains a number of real-world datasets. Table 1 summarizes the current data coverage of BioPortal.

Table 1. BioPortal Datasets

Disease	Related Datasets
West Nile Virus	Human (NY, CA '03); Captive Animal (NY '03); Bird Sightings (NY '01-'03, CA '03, USGS '99-'03); Mosquito Pool (NY '03, CA '00); Mosquito Treatment (CA '04) Chicken Sera (CA '03)
Botulism	Adult (NY, CA '01-'02); Infant Botulism (national '04); Avian Botulism (USGS '99-'03)
Foot-and-Mouth Disease	Middle Eastern Countries (Iran '87-'03, Iraq '85-'02, Afghanistan '96-'03, Pakistan '85-'03, Turkey '85-'03); South America (Argentina '01)

BioPortal is loosely-coupled with underlying public-health information sources which transmit disease information through secure links to BioPortal via the Public Health Information Network Messaging System (PHINMS) or similar systems based on XML and HL7 (www.hl7.org). The transmitted information is then normalized and stored in BioPortal's internal data store. BioPortal encompasses a role-based user access control module to ensure secure and proper use of data.

The infectious disease data query and analysis functions of BioPortal are most relevant to the current study and are summarized as follow. (a) ***Query***. BioPortal provides customized query interfaces to the infectious disease datasets available. For each dataset, spatial and temporal coordinates of the disease cases and sightings/test results, among others, are essential. Both coordinates can be presented at different granularities (e.g., for location, specific street address/county/state; for time, specific day and time/weekday/week/month/year). BioPortal provides a flexible tabular tool that allows users to select a preferred granularity level and presents the summary data accordingly. (b) ***Analysis and Visualization***. BioPortal supports hotspot analysis using various methods for detecting *unusual* spatial and temporal clusters of events. Hot spot analysis facilitates disease outbreak detection and predictive modeling. BioPortal includes Spatial-Temporal Visualizer (STV), a visualization tool which allows users to effectively explore spatial and temporal patterns, based on an integrated tool set consisting of a GIS tool, a timeline tool, and a periodic pattern tool.

3 Hypotheses and Evaluation Design

Based on extant literature, we evaluated BioPortal using different but complementary system success metrics pertinent to individual task performance [4], system usability [7], perceived system usefulness and ease of use [3][9], and user satisfaction [1]. Choice of our evaluation focus is congruent with the discussion in [6][11], which emphasized user satisfaction and user impact when evaluating information systems in the public sector. We explicitly distinguished user information satisfaction and end-user satisfaction, in part because of the criticality of information support by BioPortal and similar systems. User information satisfaction emphasizes on the user's information requirements and, in this study, refers to the extent to which an individual believes an information system meets his or her information needs [8]. On the other hand, end-user satisfaction has a prominent end-user computing orientation and often encompasses different fundamental aspects of user satisfaction. In this study, end-user satisfaction embraces a system's content, accuracy, format, ease of use, and timeliness [5]. For benchmarking purposes, our evaluation included an Excel-based program which approximated a typical analysis method common to public health professionals.

Specifically, we tested the following hypotheses.

H1: Use of BioPortal will generate analyses more accurate than those by the spreadsheet program.
H2: Use of BioPortal will result in user information satisfaction higher than that by the spreadsheet program.
H3: Use of BioPortal will result in end-user satisfaction higher than that by the spreadsheet program.
H4: BioPortal is more usable than the spreadsheet program.
H5: BioPortal is perceived to be more useful than the spreadsheet program.
H6: BioPortal is perceived to be easier to use than the spreadsheet program.

We tested the hypotheses using a controlled experiment in which subjects were randomly assigned to either the treatment or the control group. Our subjects were graduate students of a major state university who voluntarily participated in the study. Subjects in the treatment group used BioPortal to complete the assigned tasks, whereas their control counterparts used the spreadsheet program. To assess the technology's impact on individual task performance, we collaborated with several domain experts to design 6 analysis tasks common to public health professionals. These tasks range from simple frequency count or summation to nontrivial pattern characterization or trend analysis. With the assistance of these domain experts, we also developed a "gold-standard" analysis for each task and used it to evaluate each subject's performance, based on a 5-Point performance scale with 1 being highly unsatisfactory and 5 being highly satisfactory. To evaluate system usability, we used the User Interaction Satisfaction (QUIS) instrument by [2], based on the Object-Action Interface model by [12]. Congruent with QUIS, we examined each system's usability in terms of overall reactions toward the system, the system's screen layout and sequence, the system's capability, the terminology/information used in the system, and subjects' learning to use the system, based on a 9-Point Likert scale. We adopted the previously validated instruments to measure each system's usefulness and ease of use perceived by subjects [3], based on a 7-Point Likert scale with 1 being strongly disagreed and 7 being strongly agreed. We measured user information satisfaction and end-user

satisfaction using the respective instruments developed by [8] and [5], based on the same 7-Point Likert scale.

Using a scripted document, we explicitly informed all subjects the study's purpose and our intended use and management of the data to be collected. We addressed potential concerns about information privacy by assuring subjects our analysis of the data at an aggregate level; i.e., not in a personally identifiable manner. We provided warm-up exercises to familiarize subjects with the particular technology to be used in the experiment; i.e., BioPortal or the spreadsheet program. Each subject began the assigned tasks after explicitly signaled his or her readiness for doing so. A total of 6 analysis tasks were presented to the subject who was asked to provide an answer to or an analysis of each question included in a task. When completing all 6 tasks, each subject was asked to provide specific demographic information as well as to complete the question items for assessing system usability, perceived usefulness and ease of use, user information satisfaction, and end-user satisfaction.

4 Experimental Findings and Analysis

A total of 10 subjects participated in the study, 5 in each group. We aggregated the subjects' responses to assess the magnitude and statistical significance of the between-subjects difference. Based on the gold-standard analysis by domain experts, we examined a subject's performance in each analysis task. We then aggregated subjects' analysis performances by group (i.e., treatment versus control) and then performed a between-group comparison. We used the same approach to comparatively evaluate BioPortal and the spreadsheet program in terms of system usability, perceived usefulness and ease of use, user information satisfaction, and end-user satisfaction. Analysis showed that subjects in the treatment group were comparable to those in the control group in terms of age, gender, affiliated school (business or public health), self-assessed competence in using Excel and Internet, and general familiarity of the public health domain, and knowledge about the tasks included in the experiment.

As summarized in Table 2, subjects appeared to have produced more accurate analyses using BioPortal than using the spreadsheet program, though the difference was not significant statistically. (We suspect that the observed insignificance can be partly attributed to the small sample of the study). Further analysis suggests the use of BioPortal resulting in considerable accuracy improvements in complex (such as pattern characterization or trend analysis) as opposed to simple tasks (such as frequency count or summation). Our subjects exhibited significantly higher user satisfaction with BioPortal than with the spreadsheet program. The difference in both user information satisfaction and end-user satisfaction was statistically significant; i.e., respective p-values less than 0.05 and 0.01. Based on the subjects' assessments, BioPortal appeared to be more usable than the spreadsheet program. At various statistical significance levels (ranging from 0.05 to 0.0001), subjects considered the usability of BioPortal to be higher than that of the spreadsheet in terms of overall reactions toward the system, the system's screen layout and sequence, the terminology/information used or available in the system, and their learning to use the system. Our subjects also considered BioPortal to have greater capabilities than the spreadsheet program but the differential was not of statistical significance. As a group, our subjects perceived

BioPortal to be more useful and easier to use than the spreadsheet program, statistically significant at a 0.0001. We also examined the amount of time a subject needed to complete the assigned tasks. On average, subjects in the treatment group were able to complete the tasks within 1 hour, while their control counterparts needed 1 hour and 15 minutes to do so[1].

Table 2. Summary of Evaluation Results – BioPortal versus Spreadsheet Program

	BioPortal		Spreadsheet			
	Mean	STD	Mean	STD	Difference	P-Value
Aggregate Analysis Accuracy	4.01	1.65	3.62	1.89	0.39	> 0.05
Task 1	4.64	0.94	5.00	0.00	-0.36	N/A
Task 2	3.86	1.52	2.50	0.91	1.36	N/A
Task 3	4.64	0.94	5.00	0.00	-0.36	N/A
Task 4	4.00	1.91	2.12	2.53	1.88	N/A
Task 5	2.50	2.43	2.50	3.54	0.00	N/A
Task 6	4.50	0.84	4.50	0.71	0.00	N/A
User Information Satisfaction	5.17	1.51	4.75	1.46	0.42	< 0.05
End-User Satisfaction	5.19	1.32	4.28	1.34	0.86	< 0.01
Aggregate System Usability	2.83	3.21	3.93	3.64	1.10	<0.0001
Overall Reactions towards System	2.53	1.25	4.62	1.53	2.10	< 0.0001
Screen Layout and Sequence	2.04	1.04	3.44	2.22	1.40	< 0.05
Terminology & System Information	2.90	1.67	3.75	1.62	0.85	< 0.05
Learning to use the System	2.86	1.94	4.00	1.98	1.14	< 0.05
Capabilities of the System	3.73	2.36	3.60	2.16	0.13	> 0.05
Perceived Usefulness	5.52	1.25	4.07	1.33	1.45	< 0.0001
Perceived Ease of Use	5.67	1.10	4.36	1.45	1.31	< 0.0001

Our experiment results are encouraging and support all the hypotheses except that stating the use of BioPortal generating more accurate analyses than those by the benchmark spreadsheet program. Both user information satisfaction and end-user satisfaction are significantly higher with BioPortal than the spreadsheet program. As a system, BioPortal is considered more usable than the benchmark program by our subjects who also perceived BioPortal to be more useful and easier to use than the benchmark program.

5 Summary

Sustainable system success is critical to security informatics and requires comprehensive evaluations. In this study, we experimentally evaluated BioPortal along such essential system success dimensions as system usability, user satisfaction, perceived usefulness and ease of use, and impacts on individual task performance. While encouraging, our results should be interpreted in light of the following limitations. First, the reported results are based on an experiment study involving limited tasks and subjects. Second, the diversity of our experiment tasks has to be enhanced to mimic the

[1] The amount of time required by the control-group subjects varied greatly with their experience with Excel, ranging from 45 minutes to 100 minutes. The dispersion was considerably smaller in the treatment group, ranging from 40 minutes to 60 minutes.

task complexity and scenarios in public health. Sources of our subjects represent another limitation. While the use of graduate student subjects is appropriate for preliminary evaluations and the range of investigated tasks, continued evaluations should include public health professionals, preferably from different institutions and regions. Our future research will address these limitations.

References

1. Bailey, E., and Pearson, S. (1983). "Development of a Tool for Measuring and Analyzing Computer User Satisfaction," *Management Science*, Vol. 29. No.5. pp. 530–544.
2. Chin, J.P., Diehl, V.A., and Norman, K.L. (1988). "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface," in *Proceedings of the ACM CHI '88*, Washington, DC, pp. 213-218.
3. Davis, F.D. (1989). "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, Vol.13, No. 3, pp. 319-339.
4. DeLone, W.H., and McLean, E.R. (1992). "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research*, Vol.3, No.1, pp.60-95.
5. Doll, W.J., and Torkzadeh, G. (1988). "The Measurement of End-user Computing Satisfaction," *MIS Quarterly*, Vol. 12. No.2, pp. 259-274.
6. Hu, P. J. (2003). "Evaluating Telemedicine Systems Success: A Revised Model," *the Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, Hawaii, January 3-6, 2003.
7. Grudin, J. (1992). "Utility and Usabilities: Research Issues and Development Contexts," *Interacting with Computers*, Vol. 4, No. 2, pp. 209-217.
8. Ives, B., Olson, M.H., and Baroudi, J.J. (1983). "The Measurement of User Information Satisfaction," *Communications of the ACM*, Vol. 26, No. 10, pp. 785–793.
9. Ives, B., and Olson, M. (1984). "User Involvement and MIS success: A Review of Research," *Management Science*, Vol. 30, No. 5, pp. 586-603.
10. Liu-Sheng, O. R., Hu, P.J., Wei, C., Higa, K., and Au, G. (1998). "Organizational Adoption and Diffusion of Telemedicine Technology: A Comparative Case Study in Hong Kong," *Journal of Organizational Computing and Electronic Commerce*, Vol. 8, No. 4, pp. 247-275.
11. Rocheleau, B. (1993). "Evaluating Public Sector Information Systems: Satisfaction versus Impact," *Evaluation and Program Planning*, Vol. 16, pp. 119-129.
12. Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd Ed. Reading, MA: Addison-Wesley
13. Siegrist, D. (1999). "The Threat of Biological Attack: Why Concern Now?" *Emerging Infectious Diseases*, Vol. 5, No. 4, pp. 505-508.
14. Zeng, D., Chang, W., and Chen, H. (2004). "A Comparative Study of Spatio-Temporal Data Analysis Techniques in Security Informatics," in *Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems*, pp. 106-111.
15. Zeng, D., Chen, H., Tseng, L., Larson, C., Eidson, M., Gotham, I., Lynch, C., and Ascher, M. (2004). "West Nile Virus and Botulism Portal: A Case Study in Infectious Disease Informatics," in *Intelligence and Security Informatics*, Proceedings of ISI-2004, Lecture Notes in Computer Science, pp. 28-41, Vol. 3073, Chen, H., Moore, R., Zeng, D., and Leavitt, J. (eds), Springer.

How Question Answering Technology Helps to Locate Malevolent Online Content

Dmitri Roussinov[1] and Jose Antonio Robles-Flores[1,2]

[1] Dept of Information Systems, W.P. Carey School of Business,
Arizona State University, P.O. Box 874606, Tempe, AZ 85287-4606, USA
{Dmitri.Roussinov, Jose.Robles}@asu.edu
[2] ESAN University, Apartado 1846, Lima 100, Peru
jrobles@esan.edu.pe

Abstract. The inherent lack of control over the Internet content resulted in proliferation of online material that can be potentially detrimental. For example, the infamous "Anarchist Cookbook" teaching how to make weapons, home made bombs, and poisons, keeps re-appearing in various places. Some websites teach how to break into computer networks to steal passwords and credit card information. Law enforcement, security experts, and public watchdogs started to locate, monitor, and act when such malevolent content surfaces on the Internet. Since the resources of law enforcement are limited, it may take some time before potentially malevolent content is located, enough for it to disseminate and cause harm. The only practical way for searching the content of the Internet, available for law enforcement, security experts, and public watchdogs is by using a search engine, such as Google, AOL, MSN, etc. We have suggested and empirically evaluated an alternative technology (automated *question answering* or *QA*) capable of locating potentially malevolent online content. We have implemented a proof-of-concept prototype that is capable of finding web pages that provide the answers to given questions (e.g. "How to build a pipe bomb?"). Using students as subjects in a controlled experiment, we have empirically established that our QA prototype finds web pages that are more likely to provide answers to given questions than simple keyword search using Google. This suggests that QA technology can be a good replacement or an addition to the traditional keyword searching for the task of locating malevolent online content and, possibly, for a more general task of interactive online information exploration.

1 Introduction

After the September 11 attacks, the world started to pay close attention to its vulnerable assets, one of which is undoubtedly the Internet -- the backbone of modern information superhighway. Making cyber infrastructure resilient to any attacks or misuse became a priority for scientists and funding agencies [4]. However, the media still reports numerous cases of government web sites "defaced" or shut down by hackers [7]. In addition, the proliferation of illegal spamming, computer viruses, identity theft, software piracy and fraudulent schemes has threatened the trust behind electronic

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 418 – 421, 2005.

means of communication to the degree of becoming a threat to the national cyber infrastructure [9].

The algorithms that search engines use are based on lexical match (so called "bag of words" approach) in which the pages are represented as sets (bags) of words. This approach is results in a very well known problem of information overload on the Web [3][5][8]. Considering that the web has more than 4 billion pages, the vast majority of them are legitimate and harmless. Performing a Google (or other search engine) search on the topic of "hacking" and "phishing" results in the thousands of pages from the news media (e.g. cnn.com) or political discussion forums (e.g. soc.culture.usa) since the search engines' algorithm locate the content based on the lexical match and the popularity of the web sites [1], thus mostly overlooking "shady" portions of the web. It is the level of technical detail that can distinguish innocuous pages from harmful ones (e.g. news from "how-to" manuals) but this level of depth cannot be picked by lexical matching or link analysis.

A typical QA system would take the question in a natural language form such as "How can I guess a password?" and return an answer such as "You can use a password dictionary to guess passwords." and/or a link to a source page that provides the answer. Recent breakthroughs in the QA technologies have been reported [10]. In our study, we used the approach by [6] who expanded work by [2] by automated identification and training of patterns, triangulation, and using trainable semantic filters instead of manually created ones. Pattern based approach has additional advantages over deep NLP approaches for locating content on the Web since it can look for grammatically irregular sentences or combinations of headings followed by answer paragraphs. For example, when the system is given the question "How to hack into computer networks," it be looking for such patterns as "Re: How to hack into computer networks, " "How to hack into computer networks Tutorial," "Introduction into hacking into computer networks," etc.

2 Empirical Evaluation

Our main objective was to compare two different approaches to locating web content (keyword search vs. pattern based question answering) exemplified by the following two tools: 1) our QA tool and 2) Google search portal. We have performed our study in two phases. During the first ("task level") phase, the study volunteers came up with their own questions and attempted to find the answers using one of the tools, once for each question. During the second (blind evaluation) phase, the volunteers only evaluated the results retrieved by each of the tools without knowing which particular tool had retrieved them.

Figure 2 (posted at http://www.public.asu.edu/~droussi/IEEE2005-short-tables.html) shows a typical QA session. The interface is very straightforward: the user enters the question and receives a set of answers along with the links to the pages where the answers were found. Since we only needed Google's functionality to provide keyword search, we had implemented a front end to Google that limits its capabilities to keyword search only, disabling all the other potentially distracting features at the portal such as image search, toolbar icons, shopping, news, advertisement, etc. Our interface (posted Figure 3) redirected the query to Google without any modifica-

tions and presented the snippets returned by Google to the user, in the same order and without any modifications either.

We involved 9 volunteers in our (task level) phase. We asked them to put themselves in the shoes of a "cyber-criminal" e.g. as if they were trying to commit illicit actions (e.g. "hack" into computer networks) and come up with 6 questions, answering to which would help to learn the criminal techniques. In order to get familiar with the topic and get inspired, we suggested spending 10 minutes searching Google News for topics related to cyber crime and skimming through news articles. Then, we asked the users to find answers to their questions using Google for 3 questions and QA tool for the other 3 questions, switching turns to minimize learning effects. The volunteers followed the instructions online, on their own (not monitored), at the time of their choice. The volunteers were (all but one) undergraduate students in a school of business in a major US research university, majoring in Information Systems, familiar with web searching and the domain of study (cyber crime).

In order to evaluate the relevance of the retrieved pages we analyzed the search logs and computed the very popular metric of *reciprocal rank* of the first web page that can be considered an answer [10]. Users decided themselves if the found page could be considered an "answer page" based on our criteria explained in the instructions: as long as the page helps to answer the question even "a little bit" it can be considered an answer. The users did not have to find an exact answer nor to compile the answers based on multiple sources. Each time, when the user re-formulated Google queries, we increased the rank by 10 (the number of pages returned by Google as response to a query). The assumption behind this was that the user has at least glanced at all the top 10 snippets (or pages), did not find the answer, and moved along to a different query.

The details of the results of this phase can be found in Table 1 (posted on the web). There was no statistically significant difference in the relevance of the retrieved pages as measured by the reciprocal answer ranks. Thus, *we were not able to conclusively answer research question Q1* (improvement at the task level) and proceeded to the second phase, which promised more statistical power due to a more efficient design.

The second phase compared only the relevance of the pages returned by the tools. The results can be found on the web in Table 2. The relative improvement was quite substantial: by using Question Answering technology, the average reciprocal ranks were increased by up to 25%, which we believe is a practically important result.

3 Conclusions

We have established that pattern based Question Answering (QA) technology is more effective at locating web pages that may provide answers to the set of indicative questions (such as "How do I crack passwords?" "How do I steal a credit card number?" etc.) and by this delivering potentially malevolent content. Our direct implications are to the law enforcement officers and to the law-enforcement IT systems designers. We deliberately did not embed any domain specific decisions while designing and implementing our QA system, so from a technical perspective, it remains an open domain system. This allows us to believe that our results may be extended to the open domain question answering in general, including many other important applications where

locating online content quickly is vital. This includes business intelligence, intellectual property protection, digital forensics, and preventing acts of terrorism. Investing into new emerging technologies, as the one studied here, makes our world safer and more prosperous.

References

1. Brin, S., and Page, L.. The Anatomy of a Large Scale Hypertextual Web Search Engine. Stanford technical report. Stanford Database Group Publication Server (1998) http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1998-8&format=pdf&compression=
2. Dumais, S., Banko, M., Brill, E., Lin, J., and Ng, A. Web Question Answering: Is More Always Better? In Proceedings of ACM Conference on Information Retrieval, ACM (2002)
3. Lyman, P., and Varian, H.R. How Much Information?, School of Information Management and Systems, at the University of California at Berkeley (2000) [WWW] http://www.sims.berkeley.edu/research/projects/how-much-info/ (February, 2005)
4. National Science Foundation. NSF Announces $30 Million Program in "Cyber Trust." NSF Web site (2003) [WWW] http://www.nsf.gov/od/lpa/news/03/pr03133.htm (February, 2004)
5. Roussinov, D., & Chen, H. Information navigation on the web by clustering and summarizing query results. Information Processing and Management, 37, 6, (2001) 789-816
6. Roussinov, D. and Robles-Flores, J.A. Web Question Answering: Technology and Business Applications, in Proceedings of the Tenth AMCIS, Aug 6-8, NY, USA, (2004) 3248-3254
7. Swartz, J. Hackers hijack federal computers, USA Today, (2004) http://www.usatoday.com/tech/news/computersecurity/2004-08-30-cyber-crime_x.htm
8. Turetken, O., Sharda, R. Development Of A Fisheye-Based Information Search Processing Aid (FISPA) For Managing Information Overload In The Web Environment, Decision Support Systems, 37, 3, (2004) 415-434
9. Verton, D. and Verton, D. Black Ice: The Invisible Threat of Cyber-Terrorism, McGraw-Hill Osborne Media, Emeryville (2003)
10. Voorhees, E. and Buckland, L., Eds. Proceedings of the Twelfth Text REtrieval Conference TREC, November 18-21, Gaithersburg, Maryland, USA, NIST (2003).

Information Supply Chain: A Unified Framework for Information-Sharing

Shuang Sun and John Yen

324 Information Sciences and Technology Building, University Park, PA 16802
{ssun, jyen}@ist.psu.edu

Abstract. To balance demand and supply of information, we propose a framework called "*information supply chain*" (ISC). This framework is based on supply chain management (SCM), which has been used in business management science. Both ISC and SCM aim to satisfy demand with high responsiveness and efficiency. ISC uses an information requirement planning (IRP) algorithm to reason, plan, and satisfy needers with useful information. We believe that ISC can not only unify existing information-sharing methods, but also produce new solutions that enable the right *information* to be delivered to the right *recipients* in the right *way* and at the right *time*.

1 The Information Supply Chain Framework

Information-sharing refers to activities that distribute useful information among multiple entities (people, systems, or organizational units) in an open environment. Sharing information should consider four questions: 1) *what* to share, 2) *whom* to share with, 3) *how* to share, and 4) *when* to share. Better answering these questions can greatly improve information-sharing results: avoiding overload or deficiency, reducing sharing cost, and being more responsive. To address those questions and achieve better information-sharing results, we propose a framework called "information supply chain" or ISC.

The ISC framework is based on studies of supply chain management (SCM), which has been widely used in management science. A supply chain fulfills its customer's demand by a network of companies, mainly including suppliers, manufactures, and distributors. Fig. 1a shows a typical supply chain. A supply chain has two primary targets: to balance demand and supply and to improve efficiency and responsiveness. These are also the primary goals for sharing information. We, therefore, envision that the well studied SCM framework can work for information-sharing.

Similar to a supply chain, an information supply chain [1] (ISC) fulfills users' information requirements by a network of information-sharing agents (ISA) that gather, interpret, and satisfy the requirements with proper information. Fig. 1b shows an information supply chain.

[1] An ISC is different from the information flow of a supply chain.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 422 – 428, 2005.

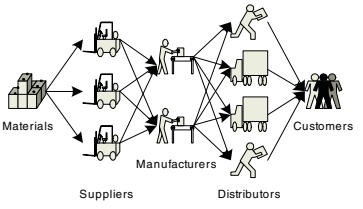


Fig. 1a. A typical supply chain

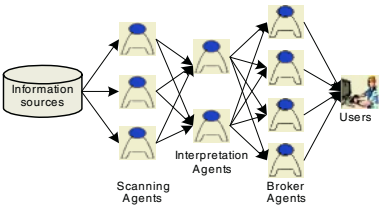


Fig. 1b. An information supply chain

Develop ISC from SCM

Developing the ISC framework requires understanding information-sharing problems, defining terms, developing methods, and choosing evaluation criteria. First, we examine if the SCM problems fit the problems of information-sharing. SCM always faces a problem to balance demand and supply. Unbalanced demand and supply leads to poor supply chain performances: either high cost due to over supplies or poor customer service due to stock outs. Information-sharing has the same problem: unbalanced demand and supply can cause either information overload or deficiency. The ultimate goal of both ISC and SCM is to make the demand and supply balanced.

Second, we collect terms that can be used to describe concepts in the ISC framework. Basic activities or objects in SCM such as purchase, sales, product, supplier, customer, or warehouse are comparable to those in ISC: query, inform, information, supplier, requester, or knowledge-base. Some concepts in SCM even suggest complex ideas for handling information-sharing problems. For example, bill of materials (BOM) lists the components needed to produce one unit of a product. Checking each component's availability can reveal the shortage for desired productions. Fig. 2a shows a BOM for a computer in a tree structure. If we have a main-board, a CPU, a monitor, and a keyboard, we need a hard-disk to assemble a computer.

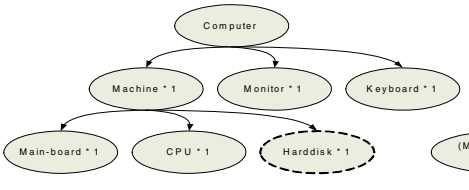


Fig. 2a. A BOM tree

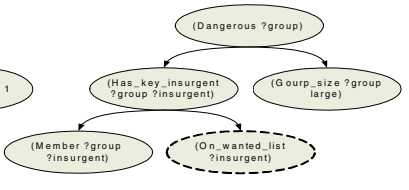


Fig. 2b. An IDR tree

We can find similar composition or dependency relationships among information. For example, implied information depends on one or more sets of supporting information or evidences, each of which may further depend on their own evidences. Such dependency relationship is called information dependency relation (IDR), which can be represented as a goal tree. Fig. 2b shows an IDR tree that illustrates intelligence analysis. Each node in the tree corresponds to the application of an

antecedent-consequent rule[2]. Suppose a group is large and its members are known. Diagnosing the IDR can identify the missing information— "if the members are on the wanted list".

Furthermore, warehouses, machines, or vehicles have capacity limitations for material storage, production, or transportation. Information-sharing agents have capacity limitations in a similar way: they have limited memory, reasoning power, or communication bandwidth. Besides, human users have grater cognitive limitations than agents have. People can only process very limited amount of information, thus they are easily overloaded by information from poorly designed systems.

After developing the concepts, we can adopt business models to handle information-sharing problems. For example, vender managed inventory (VMI) is a business model that specifies venders to manage their customers' inventories. After a customer sets its demands over a period of time, the vendor monitors the customer's stock and decides to refill when the stock level is low. It is an effective model that enables a company to focus its attention to customer service. We can adopt VMI model to information subscription, in which a provider updates its subscribers about any new or changed information. By subscription, a user can save time on queries and spend the time on processing information. We call subscription a counterpart of VMI. Other business models that have no counterparts for information-sharing can suggest new ways of sharing information.

Finally, we evaluate information supply chains by two criteria that are used to evaluate material supply chains: fill rate and total cost. Fill rates measure responsiveness— the more demands is fulfilled, the better the performance. Total cost measures efficiency by considering numbers of actions for seeking or sharing information. Fill rate and total cost are different from precision and recall, two criteria that are used for evaluating information retrieval systems. In the ISC framework, precision and recall can be used as "quality control" — to evaluate how well the provided information satisfies requirements.

ISC Differs from SCM

ISC differs from SCM. SCM deals with the flow of materials, while ISC deals with the flow of information. When we borrow strategies and methods from SCM to ISC, the difference between material and information should be considered. Quantity is used to measure material requirements. One material cannot fulfill demands from two. In contrast, we cannot use quantity to describe information. A piece of information can fulfill all demands about this kind of information, no matter how many requests are about it. Furthermore, material handling such as ordering, producing, packing, and shipping, differs from information processing such as query, observing, and transforming. Finally, materials have values, which can be determined in a market, whereas no market exists for information exchange. Although material differs from information in many ways, we believe that concepts, goals, methods, and philosophy of SCM can greatly improve information-sharing results.

[2] We use logical rules as an example for IDR. However, IDR can also be used to capture other dependences such as the aggregative or selective relations among views and data sources.

Information Requirement

To balance demand and supply, an ISC should manage well information requirements. An information requirement (r) specifies what is required (p), who need it (a), how to respond (m), and when it is needed (t) — formally denoted as $r=<p,a,m,t>$. What is required (p) specifies both information type and size limit of expected results. An agent can fulfill a requirement if it knows information i that can satisfy p. How to represent i or p is relevant to problem domains. For example, p can be either a logical condition or a SQL query statement. Likewise, i can be either a logical proposition or a database record. In addition, p also includes a size limit, which specifies a maximum number of results that the needer can process. An agent should be clear about what is required so that it can satisfy the requirements with both the right type and the right amount of information.

Who need it (a) specifies a requestor and a needer. The requestor may be a different agent from the needer. Agent a_1 may request certain information from a_2 for agent a_3. a_1 is the requestor; a_3 is the needer. The difference between requestor and needer has been incorporated in current business practice for a long time: a sales order normally specifies a sold-to party who placed the order and a ship-to party who get the products. Yet current communication methods have overlooked the deference. They simply specify a requestor or an initiator without explicitly specify who the needer is. This makes an agent unable to identify duplication of information requirements that are from different requestors.

How to respond (m) specifies a protocol (m) such as “one-time query”, “third-party subscribe” or “protell” [4]. Each protocol specifies how a provider interprets requirements, and how the provider interacts with other agents such as needers or requestors. For example, if an agent subscribes certain information, the provider should update the information regarding to changes. Including protocols in requirements allows information to be shared in the right way.

When it is needed (t) specifies a time condition such as “before certain time”, “as soon as possible”, “at certain time”, or “periodically”. Most information requirements may choose “before certain time” or “as soon as possible”. Nevertheless, if a requester chooses subscribe, the requestor should choose “periodically” as its time condition. Whether or not a provider satisfies the time conditions can be used to evaluate its service quality. In that way, we can improve an information supply chain’s performance by satisfying time conditions.

Information requirements come from three sources: direct requests, collaborative sharing of requests, and anticipations. An agent can request certain information by directly asking or subscribing. Such requests generate information requirements immediately. After the agent created a new request, it may forward it to other agents who can provide the information or seek information by investigating evidences, which may generate new requirements. Additionally, an agent can anticipate other’s needs according to their mutual beliefs [1]. The total size limits of all information requirements should be less than the needer’s information processing capacity so that the needer will not be overloaded.

The ISC framework can increase demand visibility. Information demands, in current research, are often implicit or incomplete. Demands are often hidden in assumptions, queries, or protocols. This makes it difficult to address the four questions: what to share, whom to share with, how to share, or when to share. With the ISC framework, it is easy to organize, analyze, plan, and fulfill information requirements, because it makes the requirements explicit and complete. We believe that better demand visibility can make information-sharing systems more responsive without causing information overload.

Information Requirement Planning (IRP)

In an ISC, planning and satisfying information requirements should be collaborative. Material requirement planning (MRP) proposes how to satisfy material requirements by considering type, quantity, and time of the requirements. According to the BOM and available materials, MRP can determine any shortages and creates the appropriate procurement or production plans. On the basis of MRP, we developed information requirement planning (IRP) for proposing plans to satisfy information requirements. IRP determines missing information according to the information dependency relation (IDR) and available information. IRP, then, creates information seeking plans accordingly. In addition, SCM uses collaborative planning methods to prevent unstable demand forecast and supply problems, known as "bullwhip" effects. Through collaboration, business partners create a common plan on how to satisfy consumers' demand across the supply chain. This avoids redundant or deficient supplies. We can apply the same principle in ISC management. Agents can avoid duplications on anticipating, finding, or sending information through collaborations.

2 ISC Unifies Existing Methods

The ISC framework serves as an information-sharing platform regardless of complex information contents. The framework is general enough to manage various information-sharing activities, from scanning and interpretation to information delivery. Many existing information-sharing methods can be unified and incorporated into ISC by matching a counterpart method in the SCM framework. For example, FIPA "Query Interaction Protocol" [2] specifies how to handle a query between an

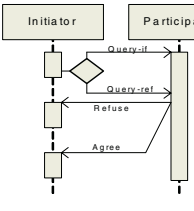


Fig. 3a. FIPA Query

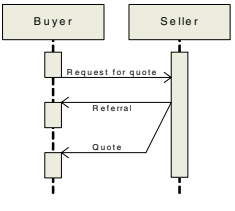


Fig. 3b. PIP FRQ

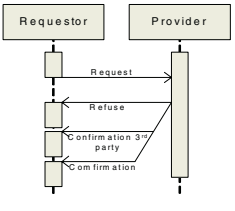


Fig. 3c. ISC Query

initiator and a participant as shown in Fig 3a. We can find a counterpart in SCM, such as PIP-3A1 (Request Quote as shown in Fig. 3b) from Rosettanet [3] in which a seller can choose to confirm a request or refer other suppliers if it cannot satisfy the request. It is easy to notice that the referral option is ignored in FIPA's specification. We thus can extend the current query protocol to incorporate the choice of referring alternative suppliers as 3rd party query confirmation (shown in Fig. 3c). Similar to the query interaction protocol, ISC is capable of unifying many other existing methods such as subscription, third-party subscribe" and "protell" [4].

The ISC framework can also lead to new solutions for information-sharing because years of research and practice in supply chain management can suggest overlooked problems, concepts, methods in the field of information-sharing. For example, Just-in-Time (JIT) is an efficient supply mode that aims to maintain high volume productions with minimum inventories (raw materials, work-in-process, or finish goods). Successful JIT productions depend on close collaborations with suppliers. The JIT philosophy, "to avoid over supply", reflects a goal of information-sharing — to avoid information overload. We thus propose using the JIT method to handle situations when agents/users are overloaded with frequently changed information. For example, suppose Tom lives in New York. He needs go to London for a conference in one month. He wants to check weather conditions before he leaves. It would be appropriate to pass the local weather forecast of London to Tom just before he leaves, as specified in JIT. Other methods are inappropriate. If Tom requests a forecast now, the information will become obsolete at the time when he leaves. A JIT request is the most appropriate approach because a) the information will not become obsolete and b) Tom will not get overloaded by irrelevant forecasts. The JIT method is suitable for requesting changing information such as weather forecasts, locations of moving objects, exchange rates, prices, and so on.

3 Conclusion

The goal of ISC is not to give a complete set of solutions regarding all aspects of information-sharing. Instead, we aim to create ISC as a SCM metaphor: a set of concepts, methods, theories, and technologies that can help to organize concepts, unify existing methods, and discover new solutions that have been neglected. Sharing information requires clear understanding about what to share, whom to share with, how to share, and when to share. The information supply chain framework explicitly captures these questions as information requirements, so we expect that the systems developed under the framework will enable the right *information* to be delivered to the right *recipients* in the right *way* and at the right *time*.

References

1. Cannon-Bowers, J.A., E. Salas, and S.A. Converse: Shared mental models in expert team decision making, in Individual and group decision making, N. Castellan, Editor (1993). p. 221-246

2. FIPA: FIPA Query Interaction Protocol Specification, (2002) Foundation for Intelligent Physical Agents Geneva, Switzerland
3. RosettaNet: PIP 3A1 Business Process Model, (2003) Uniform Code Council
4. Yen, J., X. Fan, and R.A. Volz, eds: Proactive Communications in Agent Teamwork. Advances in Agent Communication, ed. F. Dignum. Vol. LNAI-2922 (2004) Springer. 271-290

Map-Mediated GeoCollaborative Crisis Management

Guoray Cai, Alan M. MacEachren, Isaac Brewer, Mike McNeese,
Rajeev Sharma, and Sven Fuhrmann

GeoVISTA Center,
Pennsylvania State University,
University Park, PA,
(1+) 814-865-4448
{gxc26, maceachren, ixb117, mdm25, rxs51, suf4}@psu.edu

Abstract. Managing crises requires collecting geographical intelligence and making spatial decisions through collaborative efforts among multiple, distributed agencies and task groups. Crisis management also requires close coordination among individuals and groups of individuals who need to collaboratively derive information from geospatial data and use that information in coordinated ways. However, geospatial information systems do not currently support group work and can not meet the information needs of crisis managers. This paper describes a group interface for geographical information system, featuring multimodal human input, conversational dialogues, and same-time, different place communications among teams.

1 Introduction

Crisis events, like the 9.11 attack and the recent tsunami devastation in South Asia, have dramatic impact to human society, economy and our environment. Crisis management activities, involving immediate response, recovery, mitigation, and preparedness, present large scale and complex problems where government in all levels plays a key role. Geographical information systems have been indispensable in all stages of crisis management, where computers are used to map out evolving crisis events, affected human and infrastructure assets, as well as actions taken and resources applied). Their use, however, has been mostly confined to single users within single agency. The potential for maps and related geospatial technologies to be the media for collaborative activities among distributed agencies and teams have been discussed (MacEachren 2000, 2001, Muntz et al. 2003, MacEachren and Brewer 2004), but feasible technological infrastructure and tools are not yet available. An interdisciplinary team from Penn State University (comprised of GIScientists, information Scientists and computer scientists) has joined efforts with collaborators from federal, state, and local agencies to develop an approach to and technology to support GeoCollaborative Crisis Management (GCCM). The project faces two broad challenges: (1) *we have little understanding on the roles of geographical information in distributed crisis management activities*; and (2) *we have no existing computational models and experiences in developing geospatial information technologies and human-computer systems to facilitate geocollaborative crisis management*. We

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 429 – 435, 2005.

address this chicken-and-egg problem by human-centered information systems design approach (Flanagan et al. 1997).

This paper presents our progress towards supporting geocollaborative activities through innovations on two aspects (namely *group work with GIS*, *multimodal dialogue interfaces*) and their integration (see section 3 and 4). We briefly describe our initial implementation of GCCM_Connection, a dialogue-enabled multimodal, multi-parties interface to geographical information systems for crisis response. Before we introduce technical advances, it is necessary to understand the nature of collaborative activities in crisis management.

2 GeoCollaborative Crisis Management

Crisis management includes both strategic assessment (work to prepare for and avert crises) and emergency response (activities designed to minimize loss of life and property). Managing crises requires collecting geographical intelligence and making spatial decisions through collaborative efforts among multiple, distributed agencies and task groups. Typically, one or more emergency operation centers (EOC) works in cooperation with teams of field responders through communication of the situation and coordination of actions. In such collaborative processes, maps encourage efficient communication of knowledge, perceptions, judgment, and actions. This is best explained by a scenario below:

> *The Crystal River nuclear power plant has notified officials that an accident occurred, resulting in a potential radioactive particulate release within 9 hours. Response professionals with a range of expertise, work to determine the impact area, order and carry out evacuations, and deploy RAD health teams to identify 'hot zones' in residential and agricultural areas. Based on available information, immediate decisions must be made about where and how to evacuate or quarantine residents, establishing decontamination checkpoints, deploying rescue and RAD health teams, ordering in-place sheltering, and prioritizing situations. As field personnel deploy, the command Center focuses on coordination of the distributed activity among many participants who are using a range of devices and who have a wide range of geospatial information needs.*

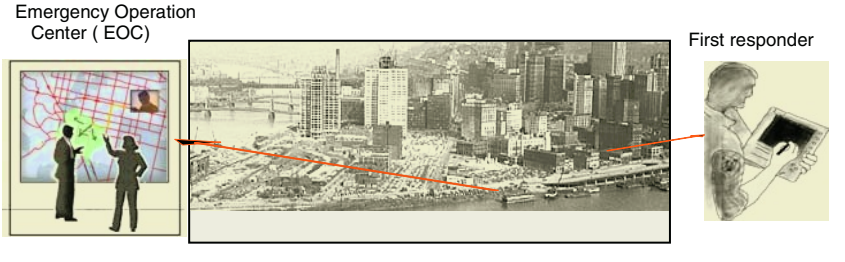


Fig. 1. A scenario of *GeoCollaborative Crisis Management (GCCM):* a crisis management center with the large-screen display collaborating with a first responder in the field

We have been collaborating with the Florida hurricane response center, New York / New Jersey Port Authority, the Pennsylvania Incident Response System, and multiple U.S. EPA units. We worked directly with agency experts to produce explicit representations crisis management components, their interrelations, and changes in both through stages of a crisis event.

3 Group Work and GIS

Neither geospatial information technologies nor groupware (Roseman and Greenberg 1992) and Co-opWare (Darnton 1995) technologies support group work with geographical information (MacEachren 2000). Both the CSCW community and the GIScience community have recognized the importance of supporting collaborative work with geographical information in the last decade, but the communities seems to have been taken two distinct approaches:

(1) *Extending a GIS with some ad-hoc features of group support*. This is the approach taken by most geographical information scientists, especially those in spatial decision-making context. Some interesting extensions include: *shared graphics* (Armstrong 1994), *argumentation map* (Rinner 2001), and shared interaction and annotation on the map view (Shiffer 1998, Rogers et al. 2002, MacEachren et al. 2005).

(2) *Extending groupware toolkits by selected GIS functions*. This is the approach taken by those from CSCW community. For example, *GroupArc* (Churcher and Churcher 1999) is an extension of GroupKit (Roseman and Greenberg 1992) to support video conferencing with geographic information. *Toucan Navigate* (www.infopatterns.net) is an extension of Groove (www.groove.net) to support 'virtual map room' functions.

Our approach to enable group work with geographical information in geocollaborative crisis management is fundamentally different from existing ones. The central idea is to use a collaboration agent to combine selected GIS and collaborative functions dynamically (at run-time) according to the need of the ongoing activity. This is based on the belief that a geocollaborative application is likely to need a small subset of functions from GIS and groupware, but what this subset is depends on the kinds of collaborative activities. This view is inline with the *task-technology fit* theory (Zigurs and Buckland 1998) that was developed in the domain of computer-supported co-operative work (CSCW).

4 Multimodal Dialogue Interfaces

Toward the multimodal, multi-user human-computer-human dialogue-enabled environment envisioned in GCCM, we developed methods for both capturing and understanding individual modalities for interaction, as well as the fusion of information at various levels. Algorithms for tracking multiple people and recognizing continuous gestures have been developed and integrated with speech recognition (Sharma et al. 2003). For quick deployment of multimodal interfaces in

EOC environment, we created a Multimodal Interface Platform for Geographic Information (GeoMIP) (Agrawal et al. 2004). GeoMIP is both an interface and application development tool. It supports speech and free-hand gesture inputs for interacting with a map-based interface to GIS. A close integration of speech recognition, dialog management and database further improves the usability of the system.

The design of mobile field devices for first responders takes advantage of pen-based gesture capturing and speech technologies as natural input modalities. Besides technological issues of multimodal fusion, the system adapts to the mobile contexts by explicitly reasoning on the role, task and goals of the device user. Each step within crisis management generates different tasks and goals. However, the overall goal of mobile system design should be to provide useful devices to the whole activity of the user, supporting easy and error-free operation even during stressful situations (Nielsen 1993).

Our approach to mediate human-GIS dialogues is to incorporate computational theories of contexts, intentions, and collaborative plans to support efficient dialogue strategies with a heterogeneous set of devices in the geospatial domain (Cai et al. 2005). Our design of GCCM is also informed by the cognitive-semiotic conceptual framework sketched by MacEachren (MacEachren 1995) and collaborative discourse theory (Grosz and Sidner 1986, Grosz and Kraus 1996, Lochbaum 1998).

5 The System: GCCM_Connection

We have developed a map-enabled groupware environment called GCCM_Connection (see figure 2). GCCM_Connection is a distributed multi-agent system that is designed to mediate collaborative activities among emergency managers in emergency operation centers (EOCs) and first responders in the field. Here we assume that the EOCs are equipped with a large-screen display together with microphones and cameras to capture human speech and free-hand gestures and support human-system dialogue. The EOC coordinates with hand-held device clients that support user-tool dialogue with natural speech and pen-based gestures. All communications are through XML-based web service protocols. Mobile devices use wireless connections, while the EOC system(s) use high-speed network connections.

6 Discussion and Conclusion

The overall goal of GCCM project has been to deepen our understanding on the (existing and potential) roles of geospatial information technologies as well to develop advanced human-computer systems supporting geocollaborative crisis management. This goal has been partially addressed through our analysis of work domains and through the integration of groupware, multimodal interfaces, GIS, semantic modeling and dialogue management into a suite of software modules,

GCCM-Connection. Future work will extend both the framework and implementation to support geocollaboration across multiple levels of government agencies, better design of visual mediation, and adapt to device capabilities and local contexts for field users.

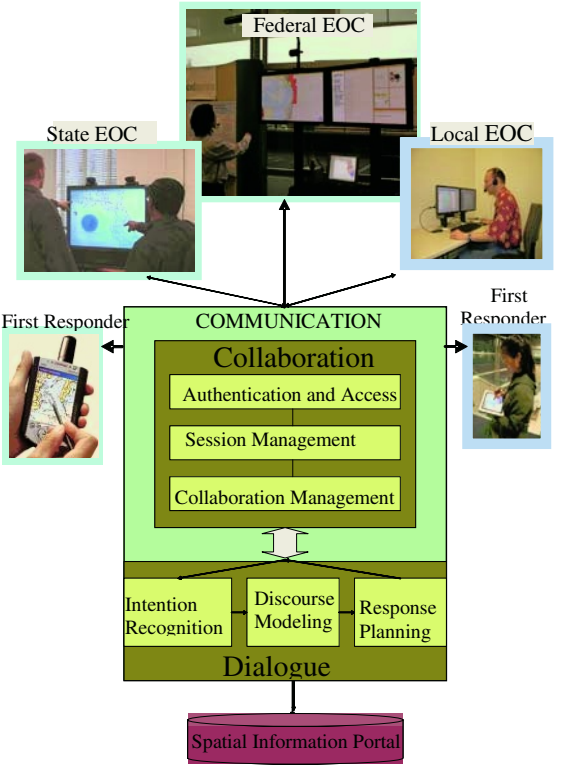


Fig. 2. GCCM_Connection Environment

To ensure the relevancy of solution development, we carefully couple two aspects of our work following the principles of human-centered systems (Flanagan et al. 1997). On one hand, we employ cognitive systems engineering methods (Brewer 2002, Hoffman et al. 2002) to seek deep understanding of the crisis management domain and the roles of geospatial information. The other component of our work is on developing advanced technologies that meets the special human-computer interaction requirements for crisis response activities (Cai et al. 2005).

Acknowledgement

This work is supported by a grant from NSF (NSF-EIA-0306845).

References

1. Agrawal P, Rauschert I, Inochanon K, Bolelli L, Fuhrmann S, Brewer I, Cai G, MacEachren A and Sharma R 2004 Multimodal Interface Platform for Geographical Information Systems (GeoMIP) in Crisis Management. In *International Conference on Multimodal Interfaces*. State College, PA
2. Armstrong M P 1994 Requirements for the development of GIS-based group decison-support systems. *Journal of the American Society of Information Science* 45: 669-77
3. Brewer I 2002 Cognitive Systems Engineering and GIScience: Lessons learned from a work domain analysis for the design of a collaborative, multimodal emergency management GIS. In *International Conference on Geographical Information Science 2002*. Boulder, CO: 22-25
4. Cai G, Sharma R, MacEachren A M and Brewer I 2005 Human-GIS Interaction Issues in Crisis Response. *International Journal of Risk Assessment and Management* special issue on GIS in Crisis management(In Press)
5. Cai G, Wang H, MacEachren A M and Fuhrmann S 2005 Natural Conversational Interfaces to Geospatial Databases. *Transactions in GIS* 9(2): 199-221
6. Churcher C and Churcher N 1999 Realtime Conferencing in GIS. *Transactions in GIS* 3(1): 23-30
7. Darnton G 1995 Working together: a management summary of CSCW. *Computing & Control Engineering Journal* 6(1): 37 -42
8. Flanagan J, Huang T, Jones P and Kasif S 1997 *Human-Centered Systems: Information, Interactivity, and Intelligence*, National Science Foundation and University of Illinois at Urbana-Champaign, Arlington, VA
9. Flanagan J, Huang T, Jones P and Kasif S 1997 National Science Foundation Workshop on Human-Centered Systems: Information, Interactivity, and Intelligence,
10. Grosz B J and Kraus S 1996 Collaborative plans for complex group action. *Artificial Intelligence* 86: 269-357
11. Grosz B J and Sidner C L 1986 Attention, intentions, and the structure of discourse. *Computational Linguistics* 12: 175-204
12. Hoffman R R, Klein G and Laughery K R 2002 The state of cognitive systems engineering. *IEEE Intelligent Systems* 17(1): 73-75
13. Lochbaum K E 1998 A collaborative planning model of intentional structure. *Computational Linguistics* 24(4): 525-72
14. MacEachren A M 1995 *How maps work: representation, visualization and design*. New York, Guilford Press
15. MacEachren A M 2000 Cartography and GIS: facilitating collaboration. *Progress in Human Geography* 24(3): 445-56
16. MacEachren A M 2001 Cartography and GIS: extending collaborative tools to support virtual teams. *Progress in Human Geography* 25(3): 431-44
17. MacEachren A M and Brewer I 2004 Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographical Information Science* 18(1): 1-34
18. MacEachren A M, Cai G, Sharma R, Brewer I and Rauschert I 2005 Enabling collaborative geoinformation access and decision-making through a natural, multimodal interface. *Internaional Journal of Geographical Information Science* 19(1): 1-26

19. Muntz R R, Barclay T, Dozier J, Faloutsos C, Maceachren A M, Martin J L, Pancake C M and Satyanarayanan M 2003 *IT Roadmap to a Geospatial Future, report of the Committee on Intersections Between Geospatial Information and Information Technology*. Washington, DC, National Academy of Sciences Press
20. Nielsen J 1993 *Usability Engineering*. Boston, AP Professional
21. Rinner C 2001 Argumentation maps: GIS-based discussion support for online planning. *Environment and Planning B-Planning & Design* 28: 847-63
22. Rogers Y, Brignull H and Scaife M 2002 Designing Dynamic Interactive Visualisations to Support Collaboration and Cognition. In *First International Symposium on Collaborative Information Visualization Environments, IV 2002, London, July 10-12, 2002*: 39-50
23. Roseman M and Greenberg S 1992 GroupKit: A groupware toolkit for building real-time conferencing applications. In *Proceedings of the ACM CSCW Conference on Computer Supported Cooperative Work*. Toronto, Canada: 43-50
24. Sharma R, Yeasin M, Krahnstoever N, Rauschert, Cai G, Brewer I, MacEachren A and Sengupta K 2003 Speech-gesture driven multimodal interfaces for crisis management. *Proceedings of the IEEE* 91(9): 1327-54
25. Shiffer M J 1998 Multimedia GIS for planning support and public discourse. *Cartography and Geographic Information Systems* 25(2): 89-94
26. Zigurs I and Buckland B K 1998 A Theory of Task/Technology Fit and Group Support Systems Effectiveness. *MIS Quarterly* 22: 313-34

Thematic Indicators Derived from World News Reports

Clive Best, Erik Van der Goot, and Monica de Paola

IPSC, Joint Research Centre, Italy
Clive.best@jrc.it

Abstract. A method for deriving statistical indicators from the Europe Media Monitor (EMM) is described. EMM monitors world news in real time from the Internet and various News Agencies. The new method measures the intensity of news reporting for any country concerning a particular theme. Two normalised indicators are defined for each theme (j) and for each country (c). The first (I_{cj}) is a measure of the relative importance for a given theme to that country. The second (I_{jc})is a measure of the relative importance placed on that country with respect to the given theme by the world's media. The method has then been applied to news articles processed by EMM for each day during August 2003. This month was characterized by a number of serious terrorist bomb attacks visible both in the EMM data and in the derived indicators. The calculated indicators for a selection of countries are presented. Their interpretation and possible biases in the data are discussed. The data are then applied to identify candidate countries for "forgotten conflicts". These are countries with high levels of conflict but poorly reported in the world's media.

1 Introduction

Political science research in the area of international conflicts has often been based on the analysis of world events as reported by news media. Event data research was developed in the 70's - 90's by defining coding schemes such as WEIS (World Event Interaction Survey)[1] and more recently IDEAS (Integrated Data for Event Analysis)[2]. A recent JRC technical note gives a broad summary of the method and the state of play regarding Event Analysis [3]. Since the 90's automatic coding software has been developed which can handle large quantities of reports. Another advantage of automatic event coding systems as opposed to human event coding, is that subjective influences can be avoided. The main software packages available for event coding are KEDS [4], TABARI [5] and VRA [6]. Traditionally, event coding software is based on analysis of the leading sentences in Reuters News Reports. This has the advantage that each report is already country related and concerns a single "event" that has happened somewhere in the world. Event coding software normally uses a natural language parsing of the sentence which identifies the verb (event) and where possible the two actors (subject and object). The Events are then compared against an extensive list of similar verbs to identify an event code. The event codes themselves are classified according to the severity of the action on a standard such as the Goldstein scale [7]. Likewise dictionaries of actors are used to identify classes of actors, be they civil or governmental persons. Event coding software is only fully available in English. Some drawbacks of just using Reuters news sources is that the perspective

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 436 – 447, 2005.

on news tends to be Anglo-American and coverage in some parts of the world is fairly limited. The widespread growth in on-line news reporting on the web means that global coverage in many languages is becoming possible. A recent study has used the TABARI event coder to benchmark news gathered by the Europe Media Monitor [8], but again only in English.

The work described in this paper takes a new broader approach to deriving statistical indicators which reflect the current political state within a country with regard to world news reporting. Advantages of this new approach are that it is fully automatic, multilingual and can cover all countries in the world simultaneously. The method builds on the Alert detection system developed within EMM. EMM itself has been described elsewhere [9,10]. The EMM Alert system filters articles gathered from the Internet and/or from News Agency wires according to topic specific criteria. It thus falls into the category of Topic Tracking within the DARPA TIDES program [11]. An alert definition consists of groups of multilingual keywords (and stems) which can either be weighted (positive and negative) or be combined into Boolean combinations. The alert system extracts the pure article text from the web page and then processes each word against all keywords defined for all alerts, in a single pass. An article triggers an alert if either the sum of detected keyword weights passes a threshold, or if a Boolean keyword combination is valid, or both. An alert can consist of one or more combinations and/or a weighted list of keywords. The alert system thus provides a very flexible and tunable 'topic definition' for automatically classifying articles in different languages. Recent alert definitions include keywords in Chinese, Farsi and Arabic.

Profile-based approaches to topic tracking [12] identify a topic as a vector of (sometimes stemmed) words and then apply statistical methods to measure the relevance of a new document to the topic profile. These methods require a collection of manually pre-categorised documents. The advantage of the search word-based EMM alert definition is that the users, who are widely spread over many different disciplines, can use their intuition as subject domain specialists to define efficient alert queries across multiple languages. A further advantage of EMM's keyword based alert system is it's speed, which is essential for the real time operation of the service. The full text of a single news article is processed against 600 pre-defined alerts containing 10,000 multilingual keywords in < 0.1s. About 30,000 articles from 700 sources in 30 languages are processed each day by EMM, and 4000 email and SMS messages are sent automatically to subscribers. EMM is a real time system which depends on alerts for pre-defined topic tracking and a separate "breaking news detection system" for new event detection. The breaking news detection system adopts the keyword clustering approach [13], which is more suitable for unforeseen events.

Recently a "World News" section has been added to EMM with the aim of geolocating news stories by content. It consists of 218 new alerts, one for each country in the world. An individual country alert simply consists of variations of that country name in all 20 European Languages plus usually the capitol city. Capitol cities are avoided where false triggers occur due to name clashes. An individual article will therefore trigger each country to which it explicitly refers. This means that articles on any subject which refer to a given country will be flagged for that country. In addition a number of global subject "themes" were defined such as "conflict", "food aid" etc. Likewise these theme alerts consist of groups of characteristic keywords, which occur

frequently in articles about the particular theme. These are standard EMM alerts and clearly again the quality of classification depends on fine tuning these multilingual keywords.

A feature of the EMM alert system is the recording for each article of which (multiple) alerts it has triggered. This information is stored in XML files held for each alert. At the same time the ALERT system also maintains statistics on the number of articles triggering for each ALERT per hour and per day. The results are coded in XML as shown below.

```
<alert id="Israel">
   <count>1085</count>
   <stats>64,18,11,19,22,25,21,37,55,69,41,47,84,135,98,76,34,33,36,38,35,
          37,28,22</stats>
</alert>
```

Finally for this new work on indicators a statistical combination measurement was added to the alert system specifically to cross-correlate Countries and Themes. Combinations occur when a single article triggers BOTH alerts referenced in the combination. A combination statistic coded in XML automatically by the EMM Alert processor is shown below.

```
<combination>
   <alertRef id="Conflict" />
   <alertRef id="Israel" />
   <count>112</count>
   <stats>5,2,1,8,2,8,4,7,0,4,3,4,14,16,8,5,3,2,3,5,3,2,3,0</stats>
</combination>
```

For both XML snippets, the <count> tag gives the total number of articles published in a single 24 hour period and the <stats> tag gives the hourly statistics. One full XML file containing statistics for all EMM alerts and combinations between countries and themes is generated for each day. These files, therefore record time series of theme and country statistics, and are available for trend studies.

2 Definitions of Socio-political Indicators

The objective of this work is to define a numerical measurement of the relative importance of a given theme for an individual country, as reported by the world's media. To be useful the indicator must be normalized to allow country comparisons and trend studies.

For this study two normalised Indicators for a given theme and a given country can be defined as follows :

Definition 1:

$$I_{cj} = N_{cj} / N_c \quad \left| \begin{array}{l} \text{where } N_{cj} = \text{article <count> for Combination of country c} \\ \text{and theme j and } N_c = \text{total <count> for country c} \end{array} \right. \tag{1}$$

I_{cj} is therefore a measure of the fraction of all articles written about that country which also refer to the particular theme. In some sense it measures the relevance of a given theme to a particular country as sampled from the world's media and varies between 0 (no relevance) and 1 (100% relevance). This standard definition becomes particularly relevant for trend studies. In practice we renormalise the indicator to be a percentage by multiplying by 100.

Definition 2:

$$I_{jc} = N_{cj} / N_j \quad \left| \begin{array}{l} \text{where } N_{cj} = \text{article <count> for Combination of country} \\ \text{C and theme j and } N_i = \text{total <count> for theme j.} \end{array} \right. \tag{2}$$

I_{jc} thus measures the fraction of articles written on a given theme which also refer to a given country. There is a subtle difference between I_{jc} and I_{cj}. I_{jc} measures the focus of the world's media attention when writing about a given subject. I_{jc} is useful for finding forgotten crises. Some countries are rarely in the news, but when they are the media does just report troubles. Therefore I_{cj} can have a very high value, but this does not mean that the world's attention is focused on that countries problems. I_{jc} will identify this because it will have a low value, reflecting the fact that the crisis is rarely reported in the media priorities. The signature for a forgotten crisis therefore is a country with a high I_{cj} and a low I_{jc}

Both indicators are measured automatically on a daily basis by EMM, and all values are archived. Note that EMM actually measures them on an hourly basis, but in practice this is usually too stochastic. For similar reasons it is sometimes necessary to aggregate some indicator measurement over longer time periods than one day such as a week or n-days. This is usually the case for countries which are rarely in the news but where there is still the need to measure the indicator.

3 Pilot Study

The study period was August 2003, which was used to determine the effectiveness of the new method. For this trial the following themes have been defined for global indicators. The actual keywords used for each theme can be seen on-line at http://emm.jrc.org - select World News/Themes.

Table 1. List of themes used by EMM for this analysis

Communicable Diseases	Man Made Disasters
Conflict	Natural Disasters
Development	Militant Islam
Ecology	Human Rights (Society)
Food Aid	Terrorism

Some themes were taken because EMM already had alerts defined in these areas. In other cases they were added especially for this study. Note that any other theme can be so defined and added to the monitoring system. The quality of the article filtering can always be improved by careful tuning of the keywords, and in particular extended in language cover. Currently the main theme effected by lack of language cover is Conflict, whose keywords for this initial study were defined mainly in English. The other themes however covered all the main languages used in EMM. The actual numerical value of I_{cj} for Conflict is therefore not yet fully accurate, since it is normalised to the total counts for each country across all 12 languages. I_{jc} however is a fair and objective measure of the country focus for each of the themes at least for those concerning articles written in the main European languages (English mainly for Conflict). Since for this first study, time trends were important, it was decided not to tune the keywords during the trial, and no changes were made to the definitions during the period of one month. It is also a simple matter to add more themes but an elapse time for data collection is needed before analyzing results. Articles were filtered through the statistics package and derived indicators per day for each theme and for every world country generated.

EMM is currently producing statistics on all 218 countries, which allow indicators to be calculated for each one. Figure 2 shows the daily indicators for Terrorist Attack for four selected countries during August 2003. Such trend graphs can be produced for any countries and any theme.

4 Discussion

August 2003 was a bad month for terrorist bomb attacks. Figure 1 shows a geographic distribution of raw totals of news articles derived from EMM on the days of 4 major attacks. These images were produced by the CommonGis [12] visualization software and show the total numbers of articles published per country for the EMM TerroristAttack alert. The most widely reported incident was the Baghdad bombing of the UN-HQ on 19th August. The colour intensity and size of circle represent the relative numbers of articles for each country detected by EMM that day for the Alert: Terrorist Attack. The highest intensity of the month was Iraq on 19th August.

Figure 2 shows the normalized indicators for the TerroristAttack alert. It is I_{jc} that follows the time variation of these events, whereas I_{cj} shows a slower trend variation. This can be understood, because the focus of world attention is measured by I_{jc}, whereas I_{cj} measures the relative importance of a theme within a country. $I_{cj,}$ however should be a better measure of long term trends i.e. as to whether a situation is improving or deteriorating. In other words, if the level of conflict is falling, this should be reflected in a falling trend for $I_{cj.}$

Figure 2 also illustrates how media coverage decays after a serious incident. The underlying coverage of terrorism in Indonesia is relatively low at under 5% of global coverage. The bomb attack in Jakarta on 5th August increased coverage to 60%, which thereafter decays slowly. After about 20 days coverage is back to pre-attack levels.

There are a number of advantages, and some drawbacks to this approach for monitoring thematic indicators using EMM. The main advantage is the fully automatic world coverage, allowing time and location dependent comparisons of

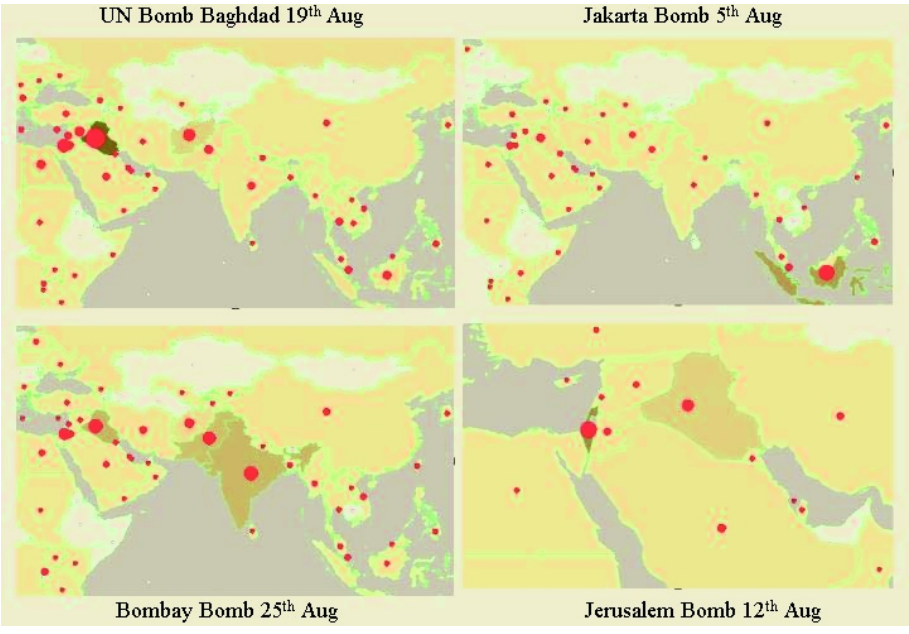


Fig. 1. In August 2003 there were several serious terrorist bomb attacks. The images show the distribution of the number of articles detected in EMM world news, which also triggered the TeroristAttack Alert. The colour coding is absolute with the highest number of articles the darkest brown. The bombing of the UN building in Baghdad, generated the highest number of news articles during the same day as the bombing

thematic indicators to be made. Particularly relevant is the time dependence within a given country, thematic indicators to be made. Particularly relevant is the time dependence within a given country, because any biases, as discussed below, remain invariant with time. Thus trends within a given country as to whether a given situation is improving or worsening are in general reflective. Another advantage is the fact that derived EMM indicators are current and available the very day of the events.

To understand the numerical values and cross-country comparisons a number of biases and their resolution need to be considered.

1. Firstly the quality of theme definitions need to be improved. In particular some definitions need to be expanded to other languages. The Conflict study has concentrated mainly on English keywords.
2. Some countries are always in the news e.g. USA and Iraq. Therefore there is a large quantity of total number of articles, which by not identifying all language variants can result in artificially low values.
3. Some areas of the world e.g. Africa have low global coverage and most articles originate from the local websites. These tend to be in English, but the number per day is too low to show accurate trends. Therefore, it is important to aggregate statistics over at least one week.

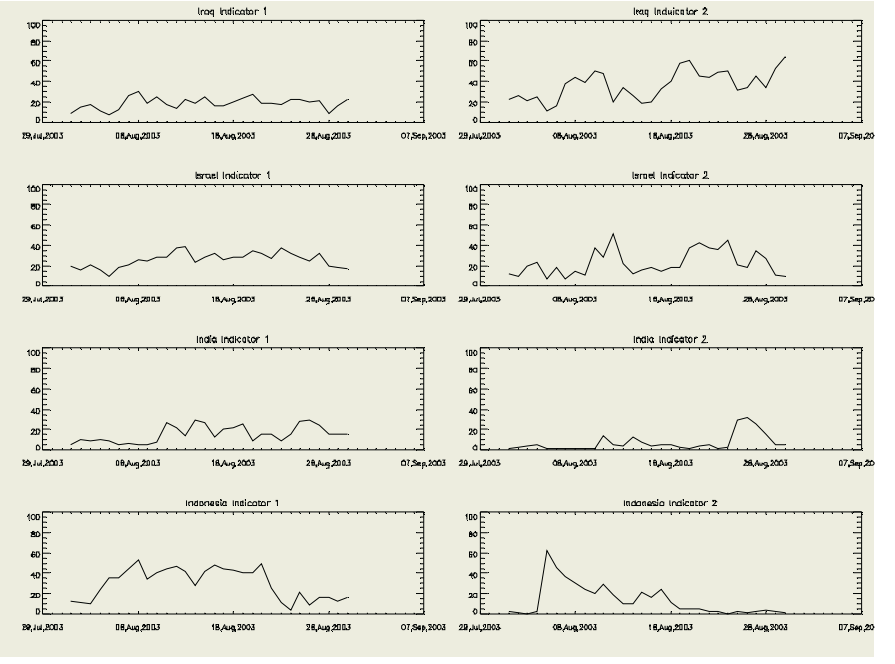


Fig. 2. TerroristAttack indictators for Irag, India, Israel and Indoinesia for August 2003. During this period each country suffered terrorist bomb attacks. Indicator 1 (Ijc) gives the percentage of media coverage on terrorism for each country whereas Indicator 2 (Icj) gives the percentage of media coverage for each country concerning terrorism. Indicator 2 shows the main incidents clearly

Steps can be taken to overcome these biases, mainly by improving the alert definitions, increasing news coverage and tuning aggregation time periods. In general the results of this pilot study show that this method of deriving indicators from world news using EMM is able to accurately reflect time dependent trends. Improvements in the multi-lingual alert definitions should allow more accurate absolute measurements. Encouraged by these results, the indicators were then applied to so-called “forgotten conflicts”.

5 Forgotten Crises

The European Commission’s Humanitarian Aid Office (ECHO) has highlighted the existence of “forgotten crises”. These are countries with potential humanitarian emergencies, which the world’s media are not reporting adequately. As a result, aid and assistance is sometimes lacking to these countries, and they are therefore in this sense forgotten. To investigate whether EMM can detect these forgotten crises using the criteria described above, the ECHO list of watch countries in this category have been

studied. Figure 4 shows 3 of these countries. On the left are the indicators I_{cj} and on the right the indicator I_{jc}. It is clearly evident that for the countries identified by ECHO the two indicators are out of balance as predicted. Therefore this method seems to be effective at identifying and continuously monitoring candidate countries for so-called forgotten crises.

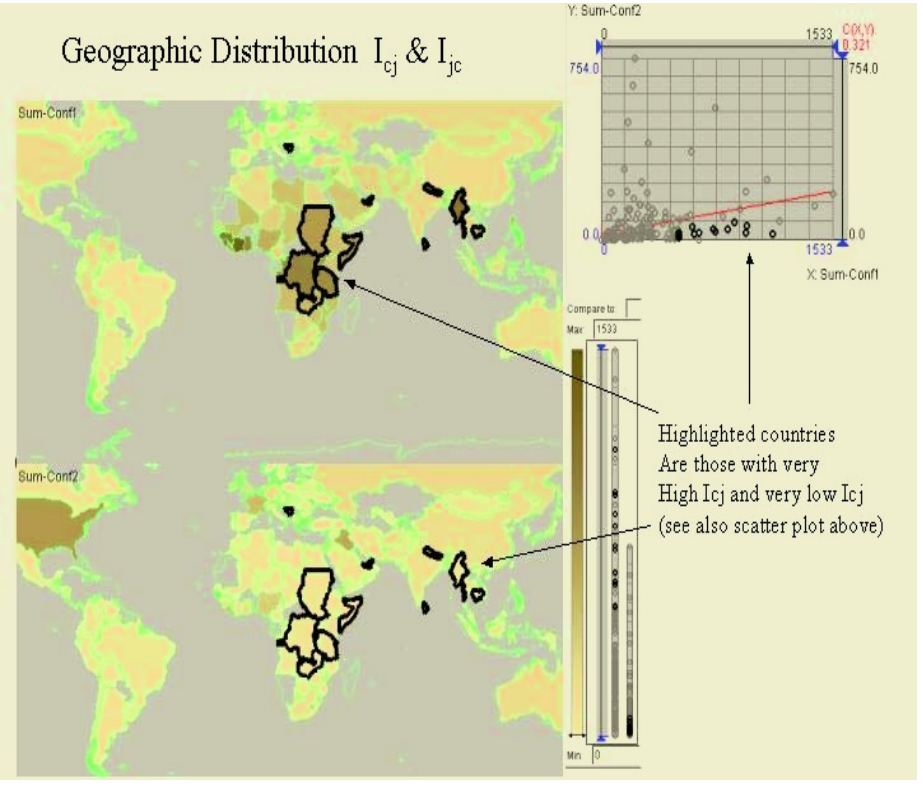


Fig. 3. Candidate Countries for "Forgotten Crises". The figure shows highlighted countries where the Conflict indicator I_{cj} is high but the second Conflict Indicator I_{jc} is relatively low. The selection criteria is highlighted in the scatter plot. These countries are those where the relative level of reporting globally is low compared to other countries, but when they are in the news it is usually about internal conflicts. A list of the 19 most "candidate" countries is given in table 2

The CommonGis [14] software was again used to study all world countries for the two indicators during August. The indicator values were averaged over the entire month and displayed as a choropleth map as shown in Figure 3. Highlighted on the map are those countries falling into the "forgotten" category. The selection criteria, used is highlighted on the scatter plot. It can be seen that several countries in Africa and Asia fit this category. A list of candidates for forgotten crises is given in Table 2.

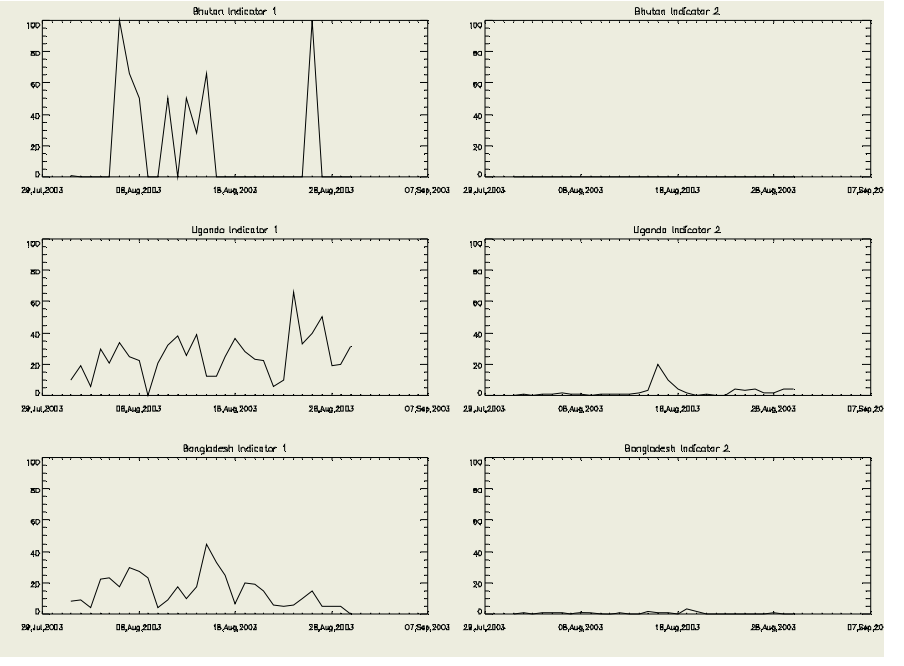


Fig. 4. Time comparisons for conflict articles for three "Forgotten Crisis" candidate countries. They show large conflict related indicators within the country but negligible values relative to worldwide reporting

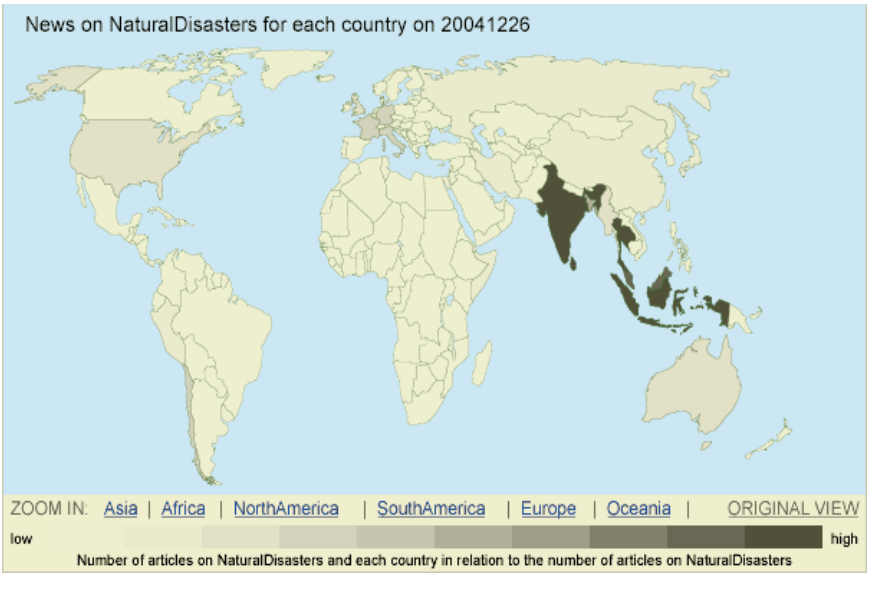


Fig. 5. Indicator I_{jc} "News Map" for the theme "Natural Disasters" for 26th December 2004. The High level all have values > 0.3

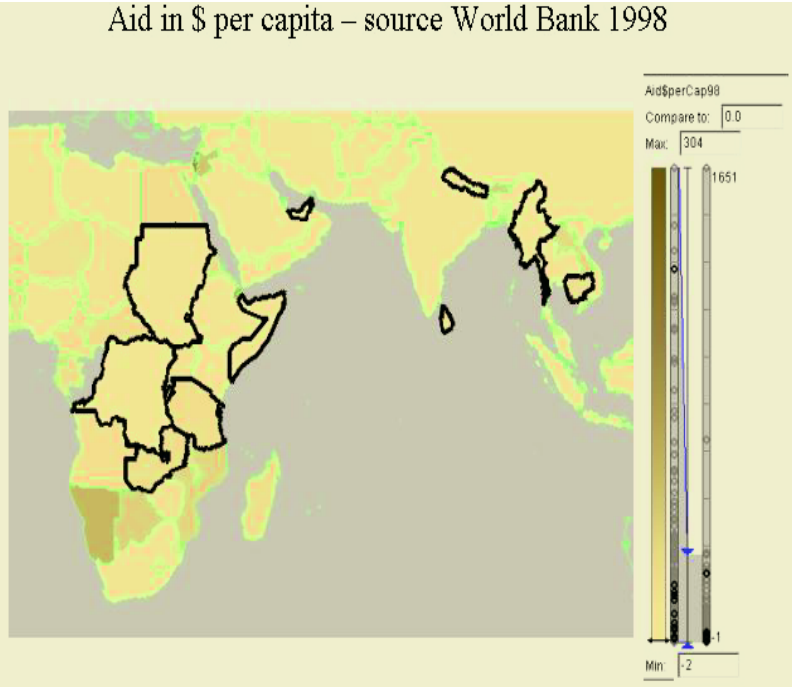


Fig. 6. World Bank data (1998) showing the relative level of aid per capita (colour weighted). The highlighted countries are the same as those shown in Figure 13. The data tends to support the view that "forgotten areas" in conflict received relatively poor amounts of AID compared to other countries

It is often felt that these countries are perhaps the one's receiving least aid from the developed world. CommonGis was again used to display data from the World Bank on Aid in dollars per capita for 1998 and compare this to those countries identified in Figure 3. Figure 6 shows the choropleth map displaying aid for the region. The data support the impression that forgotten crises are in countries receiving proportionately less aid.

Table 2. Candidate countries for forgotten crises identified by comparing the two conflict indicators described in the text

Sudan	Dem.Rep.Congo	Zambia	Tanzania
Somalia	Ethiopia	Djibouti	Mozambique
Angola	Sri Lanka	Bosnia	Nepal
Eritrea	Myanmar	Cambodia	Guinea
Ivory Coast	UA Emirates	Haiti	

6 Live Indicator Monitoring

EMM is a real time news monitor similar to Google News. For the last year EMM has also provided a live "news map" service based on these thematic indicators. Users can view current status on a world map coloured according to indicator values calculated from 0am up to that moment in the day. The maps are generated in SVG (Scalable Vector Graphics) using data from EMM statistics files. An XSLT transform is applied to the statistics XML file to generate the SVG shading values for each named country. The system also allows users to generate any previous results and animations through time. The animations are particularly effective in highlighting major events over long periods of 1-3 months. Figure 5 shows the "Natural Disaster" I_{jc} map for 26th December 2004, which was the day of the Tsunami disaster in South East Asia. News maps give a fast and effective overview of what is happening where. Indicator time series and animations provide an impression of how events develop.[1]

7 Conclusions

Two statistical indicators for measuring country specific "themes" derived from world news monitoring by EMM have been introduced. These indicators are generated automatically by the EMM alert system on an hourly and daily basis for all world countries. First results for August 2003 are very encouraging, and provide statistical measures of world events. In particular the possibility to identify and monitor forgotten crises has been demonstrated. The accuracy of these indicators will improve with better theme definitions in multiple languages. A live monitoring service has been operational for nearly a year, with positive user reaction from the external relations department of the European Commission.

References

[1] World Event/Interaction Survey (WEIS), 1966-1978. Charles McClelland, University of Southern California.
[2] Integrated Data for Event Analysis (IDEA), 1998-2002 IDEA Project. http://vranet.com/idea/
[3] Event-Based Conflict Alert/ Early Warning Tools, Delilah Al Khudhairy, 29/11/2002. Link to Report
[4] Kansas Event Data Project (KEDS), Philip A. Schrodt, Dept. of Political Science, University of Kansas. http://www.ku.edu/~keds/project.html
[5] Text Analysis By Augmented Replacement (TABARI), Instructions), Philip A. Schrodt, Dept. of Political Science, University of Kansas. http://www.ku.edu/~keds/tabari.html
[6] Virtual Research Associates http://www.vranet.com/
[7] A Conflict-Cooperation scale for WEIS Events data, Joshua Goldstein Journal of Conflict Resolution, 36:369-385
[8] Automatic Event Coding in EMM, Clive Best, Kathryn Coldwell, David Horby, Teofilo Garcia, Delilah Al Khudhairy 2003 JRC TN

[1] A public overview can be seen at http://press.jrc.it/Alert/sotw/index.jsp

[9] EMM Technical Report , Clive Best, Erik van der Goot Monica de Paola, Teo Garcia, David Horby, 21/10/2002. Link to Report
[10] EMM - Europe Media Monitor, EC Bulletin Informatique, Avril 2003
[11] DARPA Translingual Information Detection, Extraction, and Summarization (TIDES) program http://www.nist.gov/speech/tests/tdt/
[12] Allan James, Ron Papka & Victor Lavrenko, On-line New Event detection and tracking. Proceedings of 21st Annual International ACM SIGIR Conference on R & D in Information Retrieval. Melbourne Australia (1998).
[13] Schultz J. Michael & Mark Liberman (1999) Topic detection and tracking using idf-weighted Cosine Coefficient. DARPA Broadcast News Workshop Proceedings.
[14] Andrienko, G and Andrienko, N: Interactive Maps for Visual Data Exploration. International Journal Geographic Information Science Special Issue on Visualization for exploration of Spatial Data, 1999, v.13(4), pp 355-374.

A Novel Watermarking Algorithm Based on SVD and Zernike Moments

Haifeng Li, Shuxun Wang, Weiwei Song, and Quan Wen

[1] Institute of Communication Engineering, Jilin University, Changchun 130025, China
lhfvip_2000@163.com

Abstract. A robust image watermarking technique is proposed in this paper. The watermarked image is obtained by modifying the maximum singular value in each image block. The robustness of the proposed algorithm is achieved from two aspects: the stability of the maximum singular values and preprocessing before watermark extraction. Zernike moments are used to estimate the rotation angle, and the translation and scaling distortions are corrected by geometric moment methods. Experimental results show that this algorithm makes a trade-off among the imperceptibility, robustness and capacity.

1 Introduction

Digital watermarking is an important technique for intellectual property protection of digital multimedia. The basic principle of watermarking methods is applying small, pseudorandom changes to the selected coefficients in spatial or transform domain [1].

Many of the existing watermarking algorithms fail to survive even very small geometric distortions, such as rotation, scaling and translation etc. Such attacks are effective in that they can destroy the synchronization in a watermarked bit-steam, which is vital for most of the watermarking techniques. Ruanaidh and Pun [2] presented a RST (rotation, scaling and translation) resilient watermarking scheme based on the Fourier-Mellin transform (FMT). The weakness of this approach is that practical implementation suffers from numerical instability resulting from inverting log-polar mapping to get the watermarked image. In [7], Kutter proposed a watermarking scheme that can tolerate geometric attacks by embedding a reference watermark. The main weakness is that the other non-central eight peaks are inherently less robust to attacks. Bas et al [3] proposed a scheme that used robust feature points in an image to construct a triangular tessellation where to embed the watermark. The robustness of this approach is limited by the ability of the feature points detector to extract features on highly textured images after the attacks are performed.

To remedy the synchronization in the watermarked bit-steam, this paper proposes a novel solution as follows: Before embedding the watermark, the original image is translated to its centroid and scaled to a predetermined size; Prior to watermark extraction, rotation distortion is corrected firstly by estimating the rotation angle ,which is resorted to the phase information of the Zernike moments. Then the image is translated to its centroid to achieve translation invariance. Last the image is scaled to a standard image to correct scaling distortion.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 448 – 453, 2005.

A robust image watermarking algorithm based on block SVD (Singular Value Decomposition) is proposed. Each image block is embedded one bit information by quantization modulation. The watermark can be extracted successfully without the original image. Experimental results show that this proposed watermarking algorithm makes a trade-off among the imperceptibility, robustness and capacity.

2 Estimation of the Rotation Angle Based on Zernike Moments

Let $f^{'}(x,y)$ denotes the image $f(x,y)$ that has rotated by α degrees. The Zernike moments of the rotated image are given by [4]:

$$Z_{pq}^{'} = Z_{pq} \cdot \exp(-jq\alpha) \tag{1}$$

where Z_{pq}, $Z_{pq}^{'}$ denote Zernike moments with order p repetition q of the original and rotated image respectively.

The rotation angle can be easily estimated by means of the phase information of the Zernike moments. Supposing the phase $angle(Z_{pq})$ of the original image is known to us, the rotation angle α can be computed as following:

$$\alpha = \frac{angle(Z_{pq}^{'}) - angle(Z_{pq})}{q} \quad q \neq 0 \tag{2}$$

Experimental results show that the accuracy of the estimated rotation angle is rather high when the pair (p,q) takes the following values: (2,2), (4,2), (5,1), (6,2), (8,2), (9,1). Hence, we calculate the rotation angle α employing the average in equation (8).

$$\alpha = \frac{1}{6}\sum_{i=1}^{6}\alpha_i \tag{3}$$

where, $\alpha_i (i = 1,\cdots,6)$ are respectively the estimated angles when (p,q) is (2,2), (4,2), (5,1), (6,2), (8,2), (9,1).

3 Watermarking Algorithm

Watermark embedding algorithm adopts the quantization modulation after block SVD. To embed a watermark into an image:

1. Center the image $f(x,y)$. The centered image $g(x_t, y_t)$ has supporting coordinate set of (x_t, y_t),

$$\begin{cases} x_t = x - \bar{x} \\ y_t = y - \bar{y} \end{cases} \tag{4}$$

where $(\bar{x}, \bar{y})$ is the centroid of the image $f(x,y)$.

2. Apply a scaling transform to the image $g(x_t, y_t)$. The resulted image $h(x_s, y_s)$ shall have a prescribed standard size

$$\begin{cases} x_s = x_t / s \\ y_s = y_t / s \end{cases} \tag{5}$$

where $s = \sqrt{\delta / m_{00}^{'}}$, with a predetermined value δ and its zero-order moment $m_{00}^{'}$.

3. Divide the image $h(x_s, y_s)$ into non-lapped blocks. The size of blocks shall be altered according to the size of embedding data. The extracted watermark shall be more accuracy with larger block size because the singular values are more stationary. Then compute the standard variance of each image block.

$$\sigma(i) = \sqrt{\sum_{k=0}^{15} (h_k(i) - \overline{h}(i))^2 \Big/ (4 \times 4)} \tag{6}$$

where $\sigma(i)$ is the standard variance of this block, $h_k(i)$ is the k th pixel of the i th block, $\overline{h}(i)$ is the average of the i th block, and Th is a predetermined threshold. If $\sigma(i) > Th$, embed the corresponding watermark bit; else, skip this block because this block is too smooth to be suit for embedding the watermark. The imperceptibility is better with larger threshold Th, while the capacity of embedded data decreases.

4. Apply SVD to each image block. Construct the new matrix P by randomly selecting the maximum singular values of the blocks to be embedded the watermark. $P(m,n) = \lambda_{\max}^{(i)}$, where $\lambda_{\max}^{(i)}$ is the maximum singular values of the i th block.

5. Choose a proper quantization step J and quantize the maximum singular values matrix P. Firstly calculate the matrix G, $G(m,n) = P(m,n) \bmod J$. Then create the integer matrix Q, $Q(m,n) = \dfrac{P(m,n) - G(m,n)}{J}$. When the embedding watermark bit is 1, if $Q(m,n)$ is an odd number, $Q(m,n) = Q(m,n) - 1$; else, $Q(m,n)$ maintains. When the watermark bit is 0, if $Q(m,n)$ is an even number, $Q(m,n) = Q(m,n) - 1$; else, $Q(m,n)$ is unaltered.

6. Reconstruct the maximum singular values matrix $P^{'}$.

$$P^{'}(m,n) = Q(m,n) + \tilde{Q}(m,n) \tag{7}$$

where $\tilde{Q}(m,n)$ is a 2-D pseudo-random sequence with normal distribution having mean 0 and variance $J/8$. $\tilde{Q}(m,n)$ is used to reduce the error due to the quantization effect. Each block embedded watermark can be denoted by:

$B^{'}(m,n)=\sum_{l=1}^{L}\tilde{\lambda}_l^{(m,n)}u_l v_l^H$, where L is the rank of the block, $\tilde{\lambda}_1^{(m,n)}=P^{'}(m,n)$, the other singular values unchanged.

7. Obtain the watermarked image F_w by reorganizing the image blocks.

4 Watermarking Extraction

The watermark extraction is just a reverse process of the watermark embedding procedure. The basic algorithm for watermark decoding proceeds as follows.

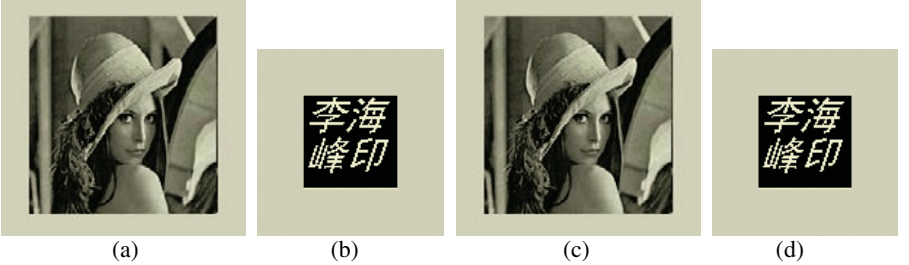
First, correct the rotation, scaling and translation distortion for the suspicious images to remedy the synchronization.

(1) Adopt the method mentioned in section 2 to estimate the rotation angle α. Then rotate the image by $-\alpha$ degree to rectify rotation distortion.
(2) Obtain the standard image using the approach referred to in section 3.2.

Second, partition the preprocessed image into non-lapped blocks and apply SVD to those blocks. Generate the 2-D pseudo-random sequence $\tilde{Q}(m,n)$, and calculate $Q(m,n)=\left[\frac{\lambda_{\max}^{(m,n)}-\tilde{Q}(m,n)}{step}\right]$, where $\lambda_{\max}^{(m,n)}$ is the maximum singular values of the corresponding block and $[\bullet]$ denotes the round operation. If $Q(m,n)$ is an even.number, the extracted watermark bit is 1; otherwise, the extracted watermark bit is 0.

5 Experimental Results

The Lena image (figure 1(a)) of size 256×256 is used as the host image. The binary image (figure 1(b)) of size 64×60 is used as the watermark. Figure 1(c) shows the watermarked image.



(a) (b) (c) (d)

Fig. 1. (a) the original image (b) the original watermark (c) the watermarked image (d) the extracted watermark

We adopt WPSNR (Weighted Peak Signal-to-Noise Ratio) to measure the visual quality of the watermarked image. NC (Normalization Correlation) is used to evaluate the similarity of the original and extracted watermark.

The WPSNR (Weighted Peak Signal-to-Noise Ratio) of the watermarked image is $46.5619\,dB$. The NC (Normalization Correlation) of the extracted watermark in figure 1(d) is 0.9987.

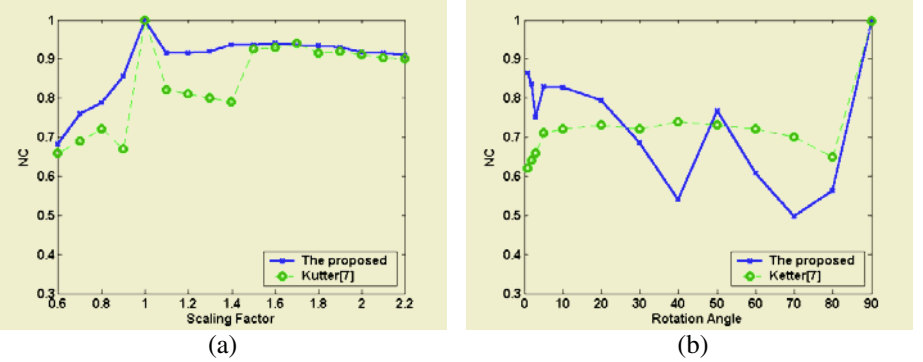
Pixel removal attacks are imposed on the watermarked image. The test results are shown in table 1. For pixel removal in the table, we randomly removed 10 columns or 10 rows for all 4 test cases.

Table 1. Pixel removal attacks

Pixel removal	ours	Zheng [6]
Case 1	0.9971	0.9580
Case 2	0.9957	0.9540
Case 3	0.9932	0.9647
Case 4	0.9978	0.9676

Figure 2(a) shows the behavior of the watermark extraction performance against scaling with factor from 0.5 to 2.2. Only when the scaling factor becomes very small the extraction performance degrades sharply because the original pixels are very few.

Figure 2(b) represents the NC after rotation different angle. We can conclude that the scheme is robust to rotation even when the rotation angle is larger.



(a) (b)

Fig. 2. (a) The results after scaling (b) The results after rotating

Combined attacks are also applied. The test results are listed in table 1, from which we see that the NC in terms of the rotation and scaling are all greater than 0.6. The stability of the singular values contributes the robustness to the additive noise, and the preprocessing prior to watermark extraction increases the robustness to geometric distortions.

Table 2. Combined attacks

Angle/scale	NC	Angle/noise	NC
3°/0.8	0.6141	3°/0.01	0.5040
3°/1.5	0.7446	3°/0.02	0.4053
10°/0.8	0.6205	10°/0.01	0.5474
10°/1.5	0.7101	10°/0.02	0.4157

The primary factor that causes degradation of the extracted watermark is the interpolation and quantization error caused by rotation and scaling. So many artifacts are introduced that perfect recovery of synchronization cannot be achieved. This limitation can be improved to some extent by adopting higher-order interpolation, however the error cannot be wholly avoided.

6 Conclusions

This paper proposed a novel image watermarking algorithm robust to rotation, scaling and translation. The cover image is preprocessed into a standard image. The watermark is embedded into the maximum singular values bit by bit through quantization modulation. Before watermark extraction, the geometric distortions are remedied. It is turned out that our algorithm is not only robust to common signal processing, but also geometric attacks. Further work will be concentrated on solving the local geometric attacks.

References

1. F. Hartung, and Martin Kutter.: Multimedia watermarking techniques. Proceedings of the IEEE, Vol. 87, No. 7, (1999) 1079-1106
2. J. O'Ruanaidh and T. Pun.: Rotation, scale and translation invariant spread spectrum digital watermarking. Signal Processing, Vol. 66, No. 3, (1998) 303-317
3. P. Bas, J. M. Chassery, and B. Macq.: Geometrically invariant watermarking using feature points. IEEE trans. on Image Processing, Vol. 11, No. 9, (2002) 1014-1028
4. Ye Bin, and Peng Jiaxiong,: Invariance analysis of improved Zernike moments. Journal of Optics A: Pure and Applied Optics, Vol. 4, No. 9, (2002) 606-614
5. H. C. Andrews, C. L. Patterson III.: Singular value decomposition (SVD) image coding. IEEE transactions on Communications, Vol. 42, No. 4, (1976) 425-432
6. Dong Zheng, Jiying Zhao, and A. El Saddik.: RST invariant digital image watermarking based on log-polar mapping and phase correlation. IEEE trans. on. Circuits and Systems for Video Technology, Vol. 13, No. 8, (2003) 753-765
7. M. Kutter.: Watermarking resistant to translation, rotation, and scaling. Proceedings SPIE Int. Conf. on Multimedia Systems and Applications, Vol. 3528, (1998) 423-431

A Survey of Software Watermarking

William Zhu[1], Clark Thomborson[1,⋆], and Fei-Yue Wang[2,3]

[1] Department of Computer Sciences,
The University of Auckland, Auckland, New Zealand
[2] Systems and Industrial Engineering Department,
The University of Arizona, Tucson, AZ 85721, USA
[3] The Key Laboratory of Complex Systems and Intelligent Science,
Institute of Automation, the Chinese Academy of Sciences, Beijing 100080, China
fzhu009@ec.auckland.ac.nz, cthombor@cs.auckland.ac.nz,
feiyue@sie.arizona.edu

Abstract. In the Internet age, software is one of the core components for the operation of network and it penetrates almost all aspects of industry, commerce, and daily life. Since digital documents and objects can be duplicated and distributed easily and economically cheaply and software is also a type of digital objects, software security and piracy becomes a more and more important issue. In order to prevent software from piracy and unauthorized modification, various techniques have been developed. Among them is software watermarking which protects software through embedding some secret information into software as an identifier of the ownership of copyright for this software. This paper gives a brief overview of software watermarking. It describes the taxonomy, attack models, and algorithms of software watermarking.

Keyword: Software Security, Software Watermarking.

1 Introduction

Software watermarking [3, 5] is a process for embedding secret information into the text of software. This information may identify the ownership of the software, so when an unauthorized use of this software occurs the copyright holders of this software can have evidence of piracy by extracting this secret message from an unauthorized copy.

This paper is structured as follows. After the introduction is Section 2 which details the taxonomy of software watermarks according to different situations. Section 3 describes the software watermark attack models. In Section 4, we give a brief description of the software watermarking algorithms currently available. This paper concludes in section 5.

⋆ Research supported in part by the New Economy Research Fund of New Zealand.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 454–458, 2005.

2 Taxonomy of Software Watermarks

Software watermarks can be classified in different ways by their functions and properties. The following are some classification schemes in published literatures.

Software watermarks are classified by their functional goals [10] as prevention marks, assertion marks, permission marks, and affirmation marks. Prevention marks prevent unauthorized uses of a software. Assertion marks make a public claim to ownership of a software. Permission marks allow a (limited) change or copy operating a software. Affirmation marks ensure an end-user of a software's authenticity.

Software watermarks can also be classified by their extracting techniques as static or dynamic [2]. A static software watermark is one inserted in the data area or the text of codes. The extraction of such watermark needs not run the software. A dynamic software watermark is inserted in the execution state of a software object. More precisely, in dynamic software watermarking, what has been embedded is not the watermark itself but some codes which cause the watermark to be expressed, or extracted, when the software is run. An example is the CT algorithm.

Robust software watermarks and fragile software watermarks: A robust software watermark can be extracted even if it has been subjected to adversarial or casual semantics-preserving or near-semantics-preserving code translation. Such watermarks are used in systems to prevent unauthorized uses(permission), and in systems that make public claims to software ownership(assertion). A fragile software watermark will (ideally) always be destroyed when the software has been changed. Such watermarks are used in integrity verification of software (affirmations) and in systems that allow limited change and copy(permission).

According to the features that a user of software can experience, software watermarks can be categorized as visible software watermarks and invisible software watermarks. If a visible software watermark is embedded, the watermarked software will generate some legible image like a logo, etc. upon some special input. An invisible software watermark is one that will not appear as a legible image to the end-user, but can be extracted by some algorithm not in the end-user's direct control.

According to whether the original program and the watermark are the inputs to the watermark extractor, software watermark can be categorized as either "blind" or "informed".

3 Attacks on Software Watermarks

Attacks on software can occur in two ways: malicious client attacks or malicious host attacks. Generally, software watermarking aims to protect software from a malicious host attack. There are four main ways to attack a watermark in a software.

Additive attacks: Embed a new watermark into the watermarked software, so the original copyright owners of the software cannot prove their ownership by their original watermark inserted in the software.

Subtractive attacks: Remove the watermark of the watermarked software without affecting the functionality of the watermarked software.

Distortive attacks: Modify watermark to prevent it from being extracted by the copyright owners and still keep the usability of the software.

Recognition attacks: Modify or disable the watermark detector, or its inputs, so that it gives a misleading result. For example, an adversary may assert that "his" watermark detector is the one that should be used to prove ownership in a courtroom test.

4 Software Watermarking Algorithms

We will describe in the section the main software watermarking algorithms currently available.

Basic block reordering algorithm: In 1996, Davidson and Myhrvold [8] published the first software watermarking algorithm. It watermarks a program by reordering its basic blocks.

Register allocation algorithm [12]: It is proposed by Qu and Potkonjak. This method inserts a watermark into the interference graph of a program.

Spread-spectrum algorithms: Stern et al. [13] proposed a spread-spectrum algorithm which represents a program as a vector and modifies each component of the vector with a small random amount. Curran et al. [7] also proposed a spread-spectrum software watermarking method.

Opaque predicate algorithms: Monden et al. [9] and Arboit [1] proposed methods to insert a watermark into a dummy method and opaque predicate.

Threading algorithm: Nagra and Thomborson [11] proposed a threading software watermarking algorithm based on the intrinsic randomness for a thread to run in a multithreaded program.

Abstract interpretation algorithm: Cousot and Cousot [6] embedded the watermark in values assigned to designated integer local variables during program execution. These values can be determined by analyzing the program under an abstract interpretation framework, enabling the watermark to be detected even if only part of the watermarked program is present.

Dynamic path algorithm: It was proposed by Collberg et al. [4]. This algorithm inserts a watermark in the runtime branch structure of a program to be watermarked. It is based on the observation that the branch structure is an essential part of a program and that it is difficult to analyse such a structure completely because it captures so much of the semantics of the program.

Graph-based algorithms: Venkatesan, Vazirani and Sinha [15] proposed the first graph-based software watermarking, called the VVS algorithm. It is a static software watermarking algorithm. Collberg and Thomborson [2] proposed the first dynamic graph algorithm, the CT algorithm. It embeds the watermark in a graph data structure which is built during the execution of the program,

so it is a dynamic software watermarking algorithm. Thomborson et al. [14] developed a variant of the CT algorithm, the constant-encoding algorithm. It tries to transform some numeric or non-numeric constants in the program text into function calls and to establish some dependencies of the values of these functions on the watermark data structures.

5 Conclusions

Software piracy is a worldwide issue and becomes more and more important for software developers and vendors. Software watermarking is one of the many mechanisms to protect the copyright of software.

References

1. G. Arboit. *A Method for Watermarking Java Programs via Opaque Predicates*, In The Fifth International Conference on Electronic Commerce Research(ICECR-5), 2002
2. C. Collberg and C. Thomborson. *Software Watermarking: Models and Dynamic Embeddings*, POPL'99, 1999
3. C. Collberg and C. Thomborson. *Watermarking, tamper-proofing, and obfuscation - tools for software protection,* IEEE Transactions on Software Engineering, vol. 28, 2002, pp. 735-746
4. C. Collberg, E. Carter, S. Debray, A. Huntwork, J. Kececioglu, C. Linn and M. Stepp. *Dynamic path-based software watermarking*, ACM SIGPLAN Notices , Proceedings of the ACM SIGPLAN 2004 conference on Programming language design and implementation, Vol. 39, Iss. 6, June 2004
5. C. Collberg, S. Jha, D. Tomko and H. Wang. *UWStego: A General Architecture for Software Watermarking*, Technical Report(Aug. 31, 2001), available on http://www.cs.wisc.edu/ hbwang/watermark/TR.ps on Nov. 20 , 2004
6. P. Cousot and R. Cousot. *An abstract interpretation-based framework for software watermarking*, Principles of Programming Languages 2003, 2003, pp. 311-324
7. D. Curran, N. Hurley and M. Cinneide. *Securing Java through Software Watermarking*, PPPJ 2003, Kilkenny City, Irenland, pp. 145-148
8. R. Davidson and N. Myhrvold. *Method and system for generating and auditing a signature for a computer program*, US Patent 5,559,884, September 1996. Assignee: Microsoft Corporation.
9. A. Monden, H. Iida, K. Matsumoto, K. Inoue and K. Torii. *A Practical Method for Watermarking Java Programs*, T he 24th Computer Software and Applications Conference (compsac2000), Taipei, Taiwan, Oct. 2000
10. J. Nagra, C. Thomborson and C. Collberg. *A functional taxonomy for software watermarking*, In Proc. 25th Australasian Computer Science Conference 2002, ed. MJ Oudshoorn, ACS, January 2002, pp. 177-186
11. J. Nagra and C. Thomborson. *Threading Software Watermarks*, Proc. Information Hiding Workshop, 2004
12. G. Qu and M. Potkonjak. *Analysis of Watermarking Techniques for Graph Coloring Problem*, Proceding of 1998 IEEE/ACM International Conference on Computer Aided Design, ACM Press, 1998, pp. 190-193

13. J. Stern, G. Hachez, F. Koeune, and J. Quisquater. *Robust Object Watermarking: Application to Code*, In Information Hiding, 1999, pp. 368-378
14. C. Thomborson, J. Nagra, R. Somaraju, and C. He. *Tamper-proofing Software Watermarks*. In Proc. Second Australasian Information Security Workshop(AISW2004), ed. P. Montague and C. Steketee, ACS, CRPIT, Vol. 32, 2004, pp. 27-36
15. R. Venkatesan, V. Vazirani, and S. Sinha. *A Graph Theoretic Approach to software watermarking*, presented at 4th International Information Hiding Workshop, Pittsburgh, PA, USA, Apr. 2001

Data Distortion for Privacy Protection in a Terrorist Analysis System

Shuting Xu, Jun Zhang, Dianwei Han, and Jie Wang

Department of Computer Science, University of Kentucky,
Lexington KY 40506-0046, USA
jzhang@cs.uky.edu

Abstract. Data distortion is a critical component to preserve privacy in security-related data mining applications, such as in data mining-based terrorist analysis systems. We propose a sparsified Singular Value Decomposition (SVD) method for data distortion. We also put forth a few metrics to measure the difference between the distorted dataset and the original dataset. Our experimental results using synthetic and real world datasets show that the sparsified SVD method works well in preserving privacy as well as maintaining utility of the datasets.

1 Introduction

The use of data mining technologies in counterterrorism and homeland security has been flourishing since the U.S. Government encouraged the use of such technologies. However, recent privacy criticism from libertarians on DARPA's Terrorism Information Awareness Program led to the defunding of DARPA's Information Awareness Office. Thus, it is necessary that data mining technologies designed for counterterrorism and security purpose have sufficient privacy awareness to protect the privacy of innocent people.

In this paper, we will discuss several data distortion methods that can be used in protecting privacy in some terrorist analysis systems. We propose a sparsified Singular Value Decomposition (SVD) method for data distortion. There are some publications about using SVD-related methods in counterterrorism data mining techniques, such as in detecting local correlation [6], social network analysis, novel information discovery and information extraction, etc. However, to the best of our knowledge, there has been no work on using SVD-related methods in data distortion. We also propose some metrics to measure the difference between the distorted dataset and the original dataset.

2 Analysis System and Data Distortion

2.1 A Simplified Model Terrorist Analysis System

A simplified model terrorist analysis system can be consisted of two parts, the data manipulation part and the data analysis part. Only the data owner or

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 459–464, 2005.

authorized users can manipulate the original data. After the data distortion process, the original dataset is transformed into a completely different data matrix and is provided to the analysts. All actions in the data analysis part are operated on the distorted data matrix. For example, the analysts can apply data mining techniques such as classification, relationship analysis, or clustering, on the distorted data. As the data analysts have no access to the original database without the authorization of the data owner, the privacy contained in the original data is protected. k-anonymity protection [7] and its variance have been used in similar scenarios, but they do not work for data distortion.

2.2 Data Distortion

Data distortion is one of the most important parts in the proposed model terrorist analysis system. We will review two of the commonly used random value data distortion methods, as well as propose a class of SVD-based methods for data distortion in this section.

Uniformly Distributed Noise. In this method, the original data matrix A is added by a uniformly distributed noise matrix N_u [2]. N_u is of the same size as A, and its elements are random numbers chosen from the continuous uniform distribution on the interval from C_1 to C_2.

Normally Distributed Noise. Similarly as the previous method, here the original data matrix A is added by a normally distributed noise matrix N_n, which has the same size as A [2]. The elements of N_n are random numbers chosen from the normal distribution with parameters mean μ and standard deviation σ.

SVD. Singular Value Decomposition (SVD) is a popular method in data mining and information retrieval [3]. It is usually used to reduce the dimensionality of the original dataset A. Here we use it as a data distortion method.

Let A be a sparse matrix of dimension $n \times m$ representing the original dataset. The rows of the matrix correspond to data objects and the columns to attributes. The singular value decomposition of the matrix A is $A = U\Sigma V^T$, where U is an $n \times n$ orthonormal matrix, $\Sigma = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_s]$ $(s = \min\{m, n\})$ is an $n \times m$ diagonal matrix whose nonnegative diagonal entries are in a descending order, and V^T is an $m \times m$ orthonormal matrix. We define $A_k = U_k \Sigma_k V_k^T$, where U_k contains the first k columns of U, Σ_k contains the first k nonzero diagonals of Σ, and V_k^T contains the first k rows of V^T. It has been proven that A_k is the best k dimensional approximation of A in the sense of the Frobenius norm. A_k can be seen as a distorted copy of A, and it may keep the utility of A as it can faithfully represent the original data. We define $E_k = A - A_k$.

Sparsified SVD. We propose a data distortion method based on SVD: a sparsified SVD. After reducing the rank of the SVD matrices, we set some small size entries, which are smaller than a certain threshold ϵ, in U_k and V_k^T, to zero. We refer to this operation as the dropping operation [5]. For example, given a threshold value ϵ, we drop u_{ij} in U_k if $|u_{ij}| < \epsilon$. Similarly, an element v_{ij} in V_k^T is also dropped if $|v_{ij}| < \epsilon$. Let $\overline{U}_k$ denote U_k with dropped elements and $\overline{V}_k^T$

denote V_k^T with dropped elements, we can represent the distorted data matrix $\overline{A}_k$, with $\overline{A}_k = \overline{U}_k \Sigma_k \overline{V}_k^T$. The sparsified SVD method is equivalent to further distorting the dataset A_k. Denote $E_\epsilon = A_k - \overline{A}_k$, we have $A = \overline{A}_k + E_k + E_\epsilon$. The data provided to the analysts is $\overline{A}_k$ which is twice distorted in the sparsified SVD method.

The SVD sparsification concept was proposed by Gao and Zhang in [5], among other strategies, for reducing the storage cost and enhancing the performance of SVD in text retrieval applications.

3 Data Distortion Measures

We propose some data distortion measures assessing the degree of data distortion which only depend on the original matrix A and its distorted counterpart, $\overline{A}$.

3.1 Value Difference

After a data matrix is distorted, the value of its elements changes. The value difference (VD) of the datasets is represented by the relative value difference in the Frobenius norm. Thus VD is the ratio of the Frobenius norm of the difference of A and $\overline{A}$ to the Frobenius norm of A: $VD = \| A - \overline{A} \|_F \; / \; \| A \|_F$.

3.2 Position Difference

After a data distortion, the order of the value of the data elements changes, too. We use several metrics to measure the position difference of the data elements.

RP is used to denote the average change of rank for all the attributes. After the elements of an attribute are distorted, the rank of each element in ascending order of its value changes. Assume dataset A has n data objects and m attributes. $Rank_j^i$ denotes the rank of the jth element in attribute i, and $\overline{Rand_j^i}$ denotes the rank of the distorted element A_{ji}. Then RP is defined as: $RP = (\sum_{i=1}^{m} \sum_{j=1}^{n} |Rank_j^i - \overline{Rank_j^i}|)/(m * n)$. If two elements have the same value, we define the element with the lower row index to have the higher rank. RK represents the percentage of elements that keep their ranks of value in each column after the distortion. It is computed as: $RK = (\sum_{i=1}^{m} \sum_{j=1}^{n} Rk_j^i)/(m*n)$. If an element keeps its position in the order of values, $Rk_j^i = 1$, otherwise, $Rk_j^i = 0$.

One may infer the content of an attribute from its relative value difference compared with the other attributes. Thus it is desirable that the order of the average value of each attribute varies after the data distortion. Here we use the metric CP to define the change of rank of the average value of the attributes: $CP = (\sum_{i=1}^{m} |RAV_i - \overline{RAV}_i|)/m$, where RAV_i is the rank of the average value of attribute i, while $\overline{RAV}_i$ denotes its rank after the distortion. Similarly as RK, we define CK to measure the percentage of the attributes that keep their ranks of average value after the distortion. So it is calculated as: $CK = (\sum_{i=1}^{m} Ck^i)/m$. If $RAV_i = \overline{RAV}_i$, $Ck^i = 1$. Otherwise, $Ck^i = 0$.

The higher the value of RP and CP, and the lower the value of RK and CK, the more the data is distorted, and hence the more the privacy is preserved. Some privacy metrics have been proposed in literature [1, 4]. We will relate the data distortion measures to the privacy metrics in our later work.

4 Utility Measure

The data utility measures assess whether a dataset keeps the performance of data mining techniques after the data distortion, e.g., whether the distorted data can maintain the accuracy of classification, clustering, etc. In this paper, we choose the accuracy in Support Vector Machine (SVM) classification as the data utility measure. SVM is based on structural risk minimization theory [9]. In SVM classification, the goal is to find a hyperplane that separates the examples with maximum margin. Given l examples (x_1, y_1), ..., (x_l, y_l), with $x_i \in R^n$ and $y_i \in \{-1, 1\}$ for all i, SVM classification can be stated as a quadratic programming problem: minimize $\frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{l} \xi_i$, subject to (1) $y_i(< w, x_i > + b) \leq 1 - \xi_i$, (2) $\xi_i \geq 0$, (3) $C > 0$, where C is a user-selected regularization parameter, and ξ_i is a slack variable accounting for errors. After solving the quadratic programming problem, we can get the following decision function: $f(x) = \sum_{i=1}^{l} \alpha_i y_i < x_i, x > + b$, where $0 \leq \alpha_i \leq C$.

5 Experiments and Results

We conduct some experiments to test the performance of the data distortion methods: SVD, sparsified SVD (SSVD), adding uniformly distributed noise (UD) and adding normally distributed (ND) noise.

5.1 Synthetic Dataset

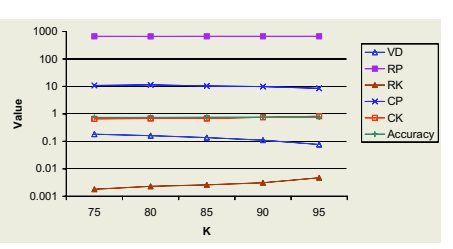
First, we compare the performance of the four data distortion methods using a synthetic dataset. The dataset is a 2000 by 100 matrix (Org), whose entries are randomly generated numbers within the interval [1,10] obeying a uniform distribution. We classify the dataset into two classes using a randomly chosen rule. The uniformly distributed noise is generated from the interval [0, 0.8]. The normally distributed noise is generated with $\mu = 0$ and $\sigma = 0.46$. The parameters of UD and ND are chosen so that the VD value of UD, ND, and SVD is approximately equal. For SVD and SSVD, the rank k is chosen to be 95, and in SSVD, the dropping threshold value ϵ is 10^{-3}.

We can see in Table 1 that the SVD-based methods achieve a higher degree of data distortion. And SSVD is better than SVD in all the position distortion measures. The Accuracy column in Table 1 shows the percentage of the correctly classified data records. Here all the distorted methods obtain the same accuracy as using the original data.

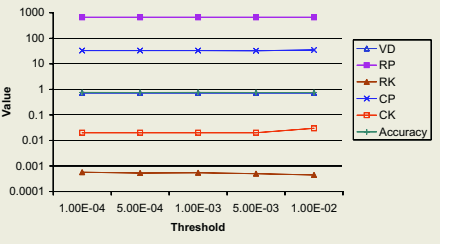
Figure 1 illustrates the influence of the parameters in the SVD-based methods. With the increase of k in SVD, VD and CP decrease while RK, CK and

Table 1. Comparison of distortion methods for the synthetic dataset

Data	VD	RP	RK	CP	CK	Accuracy
Org	-	-	-	-	-	76%
UD	0.0760	662.8	0.0058	0	1	76%
ND	0.0763	661.6	0.0067	0	1	76%
SVD	0.0766	664.0	0.0047	8.5	0.82	76%
SSVD	0.7269	666.6	0.0005	33.2	0.02	76%



(a) The influence of k in SVD



(b) The influence of threshold ϵ

Fig. 1. The influence of the parameters in the SVD-based methods

Table 2. Comparison of the distortion methods using a real world dataset

	Classification 1						Classification 2					
Data	VD	RP	RK	CP	CK	Accuracy	VD	RP	RK	CP	CK	Accuracy
Org	-	-	-	-	-	67%	-	-	-	-	-	67%
UD	0.0566	0	1	11.4	0.15	67%	0.0575	31.9	0.0166	9.5	0.07	66%
ND	0.0537	31.9	0.0298	12.2	0.27	66%	0.0566	34.1	0.0390	12.0	0.07	64%
SVD	0.0525	31.2	0.0251	12.2	0.12	70%	0.0525	31.2	0.0251	12.2	0.12	70%
SSVD	1.0422	37.5	0.0066	13.1	0.05	69%	1.3829	35.0	0.0090	11.5	0.02	65%

Accuracy increase. But with the increase of ϵ in SSVD ($k = 95$), there is no observable trend in data distortion or utility measures.

5.2 Real World Dataset

For a real world dataset, we download information about 100 terrorists from a terrorist analysis web site [8]. We selected 42 attributes, such as their nationality, pilot training, locations of temporary residency, meeting attendance, etc. To test the real world dataset, the uniformly distributed noise is chosen from the interval [0, 0.09]. The normally distributed noise is generated with $\mu = 0$ and $\sigma = 0.05$. The rank k for SVD and SSVD is chosen to be 25, and ϵ in SSVD is 10^{-3}.

In Classification 1, we classify the terrorists into two groups, those are related with Bin Laden and those are not. Here the SVD-based methods improve the accuracy a little bit. For data distortion, SSVD is the best for all the measures.

In Classification 2, the terrorists are grouped according to whether or not they have relationship with the terrorist organization Al Qaeda. Here SVD is the best for the classification accuracy, the other three methods decrease the accuracy slightly. For the data distortion measures, SSVD again works best.

6 Concluding Remarks

We proposed to use the sparsified SVD method for data distortion in a simplified model terrorist analysis system. The experimental results show that the sparsified SVD method works well in preserving privacy as well as maintaining utility of the datasets.

References

1. Agrawal, D., Aggarwal, C.C.: On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of the 20th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems.*, Santa Barbara, California, USA, (2001)
2. Agrawal, R., Srikant, R: Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD*, Dallas, Texas, (2000)
3. Deewester, S., Dumais, S., *et al.*: Indexing by latent semantic analysis, *J. Amer. Soc. Infor. Sci.*, **41** (1990) 391–407
4. Evfimievski, A., Gehrke, J., Srikant, R.: Limiting privacy breaches in privacy preserving data mining. In *Proceedings of PODS 2003*, San Diego, CA, June, (2003)
5. Gao, J., Zhang, J.: Sparsification strategies in latent semantic indexing. In *Proceedings of the 2003 Text Mining Workshop*, San Francisco, CA, (2003) 93–103
6. Skillicorn, D.B.: Clusters within clusters: SVD and counterterrorism. In *Proceedings of 2003 Workshop on Data Mining for Counter Terrorism and Security*, 12 pages, San Francisco, CA, May 3, (2003)
7. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, **10** (2002) 557–570
8. www.trackingthethreat.com
9. Vapnik, V.N.: Statistical Learning Theory. John Wiley & Sons, New York, (1998)

Deception Across Cultures: Bottom-Up and Top-Down Approaches

Lina Zhou[1] and Simon Lutterbie[2]

[1] University of Maryland, Baltimore County, Baltimore, MD 21250
zhoul@umbc.edu
[2] St. Mary's College of Maryland,
18952 E. Fisher Rd, St. Mary's City,
MD 20686-3001
sjlutterbie@smcm.edu

Abstract. This paper examines various approaches to analyzing differences in patterns of deception and how deception is evaluated across cultures. The approaches are divided into bottom-up approaches, which examine the foundations of culture and how they affect deception, and top-down approaches, which refer to models of deception and how their dynamics change across cultures. Considerations of the various approaches have led to a conclusion that the most comprehensive method for modeling deception across cultures would be to synthesize the two approaches, rather than consider them as opposing schools of thought.

1 Introduction

With the increasing trend of globalization and internationalization, interaction across national borders has become a common thing of work and daily life. As a result, the issue of deception across cultures arises and starts to attract attention. It concerns a broad scope of applications ranging from business and political negotiations, online communications, to military operations. However, there is a dearth of literature on deception across cultures. Deception in a single culture is a complex issue, and it becomes more challenging to address deception across cultures. By analyzing the dynamics of deception in different cultures, we provide insights into the difference in the perception and motivation of deception between members of different cultural groups and provide suggestions for deception detection across cultures.

Based on a review of the literature on deception [1, 2], especially deception across cultures [3-6], we proposed a classification of approaches consisting of two categories: bottom-up and top-down approaches. The former uses the characteristics of the culture(s) being studied to explain the differences in patterns of deceit. The latter is based on the general theories of deception, exploring how differences in the patterns correlate with cultures. This research highlights influential methods of viewing deception across cultures from both perspectives. Moreover, based on the analysis of the strengths and weaknesses of the two approaches, we suggested a comprehensive approach, leading to a better understanding of the dynamics of culture in deception.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 465 – 470, 2005.

2 The Bottom-Up Approach

The bottom-up approach to cross-cultural deception research is grounded on analyzing how members of the culture interact with each other. By gaining insight into the values of the culture, we can understand how information is perceived and conceptualized differently.

Individualism vs. Collectivism: The collectivism-individualism dynamic is seen as one of the fundamental differences between cultures [7, 8]. Chinese and Americans were considered as typical examples of collectivist and individualist, respectively.

Classic Chinese culture has its foundations in Confucian values, which emphasize honesty and modesty as a means to form harmonious interpersonal relationships, the cornerstone of the collective society [7]. More recently, communist influence has strengthened the collectivist foundations of the culture [5]. In general, Chinese tend to behave in ways that are expected of them by others, rather than directly for their personal benefit [7, 8], creating a greater focus on upholding the social rule and avoiding disruptive conflicts, even if it means deceiving to avoid such disturbances. Thus, collectivists tend to be more accepting of lies that support the social-good [3].

Individualistic cultures stress self-assertion and promotion, as well as competition [7]. Individualistic cultures accept small lies to avoid hurt, and self-aggrandizement is an accepted part of personality [5]. Individualists are more likely to accept lies that are self-serving [3]. People in individualistic cultures developed different, more absolute systems of moral judgment than commonly found in collectivist cultures, which creates a different dynamics for evaluating deception [3].

The impact of collectivism-individualism continuum on patterns of deception was generally supported. Compared with than Americans, Chinese children were found to be more likely to accept lying in pro-social situations [5] and more likely to accept deception designed to maintain harmonious social relationships [3]. It was found that Canadian children were more likely to be accepting of lies designed to promote or protect themselves [5], because they might view relating good deeds to others as bragging, and therefore gave telling the truth about bad deeds a lower ranking [9]. This is in line with the notion that people from individualistic cultures are more likely to deceive for self-protection and aggrandizement than people in collectivist cultures. In Canadian culture, which values individual accomplishment, bragging about one's good deeds and lying to cover misdeeds is an acceptable, and even sometimes encouraged behavior, whereas Chinese children are taught to value social harmony over individual happiness. The finding that individualistic cultures view self-promoting lies as more acceptable was not supported in a follow-up study [3]. This was partly due to difference in the development of moral judgment in two types of cultures.

Moral Development: Many researchers have found children of different ages viewed and evaluated deception differently. In particular, younger children were more likely to judge the statements based on the extent of the deceit and the severity of the resulting punishment [5]. As children got older, however, they started incorporating the protagonist's intent when deceiving in their judgments. This change suggests that the morals that determine right from wrong change over time, and that this moral development is critical in perceptions of deception.

Theory of Moral Development [10] has been used to examine deception. It stated that individuals go through a series of developmental stages, including pre-conventional (avoiding punishment or gaining rewards), conventional (focusing on individual's role in society and duties that must be fulfilled to meet the expectations of others), and post-conventional (basing behaviors on higher-level principles such as personal philosophies) stages. Individuals would progress through the stages when they are cognitively stimulated with situations that challenge their moral judgment. Each of the stages is associated with a different set of criteria for evaluating and modeling behavior. They were formed through the cognitive development of individuals, thereby theorized to be more or less universal [10] across cultures. However, Research on the universality of proposed stages produced mixed results. While the progression of stages remains relatively the same across cultures, the stages experienced varies across cultures [10]. In China, the differences between conventional level stages are not as distinct [11]. Chinese culture and education focus on good interpersonal relationships [11]. Rather than focus on the rights of the self, Chinese culture focuses on the requirements of society, which may lead to different moral judgments.

The difference in moral development and judgment between cultures further explained the difference in patterns of deception. Honesty and modesty are an integral part of the Chinese education system, and it is deemed valuable that one makes sacrifices for the common good. Their moral judgments are thus shaped to give higher priority to social harmony than personal accomplishment [9]. Therefore, Chinese children were more accepting of lies in pro-social situations [5]. Moreover, as children's moral judgments mature, they gain a greater sense of the importance of social harmony through continued socialization and education, and thus are more likely to view them as acceptable, when promoting social harmony.

An examination of the ability of Australian and Italian children to discern truths from lies further confirmed that the criteria for moral judgments fostered by various cultures have an impact on how deception is viewed, even if the cultural differences responsible for the difference do not seem directly aimed at influencing moral judgments [12]. It was found that Australian and Italian children performed similarly when identifying truths and lies; however Italian children tended to judge lies that were deemed to be "mistakes" more harshly than Australian children.

The cross-cultural difference in the acceptability and motivation of deception within specific relationship types is not limited to children but can be generalized to college students and adults. It was found that, in general, Chinese rated deception as more acceptable than Americans [3], although the acceptance of specific types of deception depended on the culture. For example, in interacting with a stranger, American participants were more likely to accept deception to maintain secrecy and protect the self, whereas Chinese participants were more acceptable of deception aimed to benefit the other. However, Chinese participants gave surprisingly high ratings to spouse deception aimed at benefiting the self or even malice towards the other. In sum, Chinese participants generally rated deception higher when it was designed to benefit others, promote social affiliation, or to improve social harmony. The variation in ratings of deception with the relationship and the motivation suggested that moral judgment plays an important role in evaluating deception, however, such judgments are made on a specific basis, rather than general moral absolutes.

3 The Top-Down Approach

The top-down approach starts with the general theories and models of deception, which explains how and why deception occurs. The main objective is to find generic deception patterns to guide broad deception research and practice. There are two theories that have been frequently used in the deception research: Interpersonal Manipulation Theory and Interpersonal Deception Theory.

Information Manipulation Theory (IMT): IMT treats deception as a monitoring and modification of information by the deceiver, however, rather than deal with the motivation of the deceiver, IMT deals with the types of information modifications made by the deceiver [13]. The deceiver can create a deceptive message by modifying the amount, the veracity, the relevance, and/or the clarity of the information [13]. Information manipulations along any one of these dimensions can result in a different type of deception [4]. When deception was attempted by manipulating the quality of information, it results in a falsification of the information. A manipulation of quantity will result in "lies of omission" [4], relevance manipulations result in evasion, and a lack of clarity will result in deception by equivocation [4].

An initial study of IMT found that different information manipulations tended to be used in different situations [13]. IMT was applied to evaluate deception in a Western culture [3]. It was found that all the methods of manipulation were viewed as deceptive, among which manipulation of quality was rated as most deceptive. The study also found that being completely honest and not engaging in information manipulation was the most useful course of action when considering deception within relationships. However, very different results were obtained when IMT was tested with participants in Hong Kong, China [4]. For example, only statements involved with quality and/or relevance manipulations were rated as more deceptive than regular statements. The study also found less of a connection between honesty and an absence of manipulations in Chinese culture. They revealed that deception was evaluated differently across cultures for every form of information manipulation.

The above cultural difference can be explained by referring to the foundation of IMT – conversation maxims in individual interpretations. The receiver expects that conversations follow a series of rational assumptions about how information is presented [13]. It is this assumption that allows the sender to manipulate information and create successful deception. It was argued [4] that the manipulations would therefore not hold in cultures that have a different conversational construction. For example, Chinese culture may view it as acceptable, even expected, that one would manipulate the quantity of information presented in order to avoid upsetting the other. According to IMT, this manipulation should be viewed as deceptive and evaluated negatively.

Interpersonal Deception Theory (IDT): IDT holds that deception should not be examined as separate for senders and receivers, but rather as an integral part of the overall process [1]. Deception occurs when the sender monitors the information they are transmitting and controls it to transmit information that differs from the truth [1]. IDT holds that deception consists of three main components, the central deceptive message, ancillary messages designed to increase the believability of the main message, and inadvertent behaviors that can indicate deception is taking place. There are three common motivations for deception: the desire to avoid hurting another, to avoid trauma, and to protect or promote the deceiver's image [1]. More importantly, IDT

would predict that different motivations for deception will be viewed as more acceptable in different situations. Culture variations entail situational change, thus the evaluations of deception would vary across cultures.

The predictions of IDT were supported by many follow-up studies of deception in mono-culture, especially Americans ([2]). The theory has yet to be tested widely in other cultures. However, the dynamics of deception in interpersonal interaction, manifested in strategic control and non-strategic behavioral leakage, is expected to be generalized to other cultural contexts. One proposition of IDT is that the degree of strategic action is influenced by initial expectation of honesty, goals of deceptions, and deception apprehension. As stated before, culture was found to matter to perception of motivations for deception, acceptance of deception, and moral development in various situations. They highlight the importance of situational analysis.

4 Conclusion

Individualism-collectivism dimension represents a fundamental variation between cultures. The differences in moral development across cultures allow moral evaluations to be based somewhat on situational factors. Thus, the bottom-up approach is an effective method for examining the overall patterns of deception across cultures, however it does not account for situational variations that take place in all cultures. The top-down approach can account for much of the situational variation observed in patterns of deception by dealing with motivations and methods of deception. However, both IMT and IDT fall short when it comes to explaining more general trends. A comparison of the two approaches is shown in Figure 1.

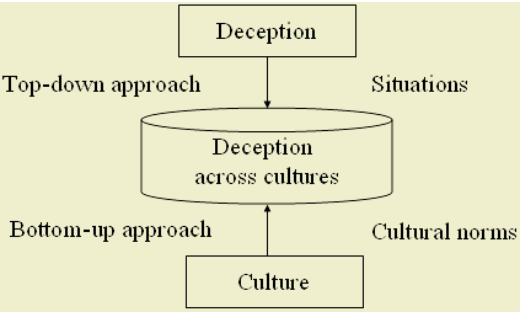


Fig. 1. A comparison between the two approaches

Based on the above analysis, it becomes apparent that the two approaches should not be taken as opposing views, but rather as complimentary methods that can be integrated to form a more comprehensive view of patterns of deception across cultures in our world today. In creating a cross-cultural model for deceptive behavior one would be well advised to consider a bottom-up approach when explaining the general trends of deceptive behaviors for different cultures. Once the general trends are established, switching to a top-down approach to explain the specific variations in

deceptive behavior for various cultures might be an effective approach. We believe that they provide a model to guide deception research in an international context in future.

Acknowledgements

Portions of this research were supported by funding from the National Science Foundation (NSF) under Grant# 0328391 and from the NSF Research Experiences for Undergraduates Site award EIA-0244131. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

1. Buller, D. B. and Burgoon, J. K.: Interpersonal Deception Theory. Communication Theory, **6** (1996) 203-242
2. Zhou, L., Burgoon, J. K., Nunamaker, J. F., and Twitchell, D.: Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: An empirical investigation. Group Decision & Negotiation, **13** (2004) 81-106
3. Seiter, J. S., Bruschke, J., and Bai, C.: The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. Western Journal of Communication, **66** (2002) 158-180
4. Yeung, L. N. T., Levine, T. R., and Nishiyama, K.: Information manipulation theory and perceptions of deception in Hong Kong. Communication Report, **12** (1999)
5. Lee, K., Cameron, C. A., Fen, X., Genyue, F., and Board, J.: Chinese and Canadian children's evaluations of lying and truth telling: Similarities and differences in the context of pro and antisocial behavior. Child Development, **68** (1997) 924-934
6. Bond, C. F., Omar, A., Mahmoud, A., and Bonser, R. N.: Lie detection across cultures, Journal of Nonverbal Behavior, **14** (1990) 189-204
7. Hofstede, G.: Culture's consequences: international differences in work-related values. Beverly Hills: Sage Publications (1980)
8. Hofstede, G.: Cultures and Organizations: Software of the Mind. Berkshire, England: McGraw-Hill Book Company Europe (1991)
9. Lee, K., Fen, X., Genyue, F., and Cameron, C. A.: Taiwan and Mainland Chinese and Canadian children's categorization and evaluation of lie- and truth-telling: A modesty effect. British Journal of Developmental Psychology, **19** (2001) 525-542
10. Eckensberger, L. H.: Moral development and its measurement across cultures. in Psychology and Culture, Lonner, W. J. and Malpass, R. S. Eds. Needam Heights, MA: Allyn and Bacon (1994)
11. Hau, K. and Lew, W., Moral development of Chinese students in Hong Kong. International Journal of Psychology, **24** (1989) 561-569
12. Gilli, G., Marchetti, A., Siegal, M., and Peterson, C.: Children's incipient ability to distinguish mistakes from lies: An Italian investigation. International Journal of Behavioral Development, **25** (2001) 88-92
13. McCornak, S. A., Levine, T. R., Solowczuk, K. A., Torres, H. I., and Campbell, D. M.: When the alteration of information is viewed as deception: An empirical test of information manipulation theory. Communication Monographs, **59** (1992)

Detecting Deception in Synchronous Computer-Mediated Communication Using Speech Act Profiling

Douglas P. Twitchell, Nicole Forsgren, Karl Wiers, Judee K. Burgoon, and Jay F. Nunamaker Jr.

Center for the Management of Information, University of Arizona, 114 McClelland Hall, Tucson, Arizona 85748
{dtwitchell, nforsgren, kwiers, jburgoon, jnunamaker}@cmi.arizona.edu

Abstract. Detecting deception is a complicated endeavor. Previous attempts at deception detection in computer-mediated communication have met with some success. This study shows how speech act profiling [1] can be used to aid deception detection in synchronous computer-mediated communication (S-CMC). Chat logs from an online group game where deception was introduced were subjected to speech act profiling analysis. The results provide some support to previous research showing greater uncertainty in deceptive S-CMC. Also shown is that deceivers in the specific task tend to engage in less strategizing than non-deceivers.

1 Introduction

Studies have shown deception detection in any media to be a complicated task [2, 3]. While there are a number of cues to deception that seem to be accurate, they seem to be very sensitive to medium, context, culture, and individual differences. For example, a relatively small number of words in a message seems to be indicative of deception, but only when the deception is some form of concealment. So, in addition to the sensitivities listed above, deception type likely plays a role as to which deception cues emerge and how pronounced they are. Therefore, until this complexity is reduced or resolved, it is useful to narrow any particular deception detection method to a relatively narrow domain. The study presented in this paper focuses on concealment and persuasive deception in synchronous computer-mediated communication (S-CMC).

Because of the lack of nonverbal cues, deception detection may be even more difficult with computer-mediated communication (CMC) than face-to-face communication. Nevertheless, several recent studies have shown some success in detecting deception in CMC. For example, Zhou et. al. [4] showed that several stylistic cues are significantly different between deceptive and truthful messages. Additional studies using those stylistic indicators as a basis for classifying messages as deceptive or truthful achieved a 60% - 80% success rate [5] and showed that deception in CMC seems to occur in the "middle" of a conversation consisting of a number of interactive messages [6]. Others have also recently begun to look at deception in CMC where more than two people are involved [7]. Results have indicated that most people who are not alerted to the presence of deception are poor detectors.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 471 – 478, 2005.

It should be noted that while the above-mentioned studies focused on automation attempts at deception detection in CMC, this paper looks at automating deception detection in S-CMC specifically. Little has been done in this area. Zhou and Zhang [8] found that dominance plays a role, and Twitchell et. al. demonstrated how speech act profiling can be used to find uncertainty, a correlate of deception, in S-CMC [9]. This study attempts to confirm that uncertainty can be found using speech act profiling to analyze S-CMC conversations. It also shows that narrowing the domain of the model should result in better deception detection.

1.1 Deception

Deception is defined as the active transmission of messages and information to create a false conclusion [10]. Messages unknowingly sent are not considered deceptive, as there is no intention to deceive. Most people are poor at detecting deception even when presented with all of the verbal and non-verbal information conveyed in a face-to-face discussion. Detection becomes even more difficult when the deception is conveyed in text (e.g., written legal depositions, everyday email, or instant messaging), and it is nearly impossible when there are large amounts of text to sift through. Furthermore, deception strategies may change with every situation as the deceiver attempts to fool possible detectors. Therefore, automated methods for deception detection are desirable.

2 Speech Act Profiling

Speech act profiling is a method of automatically analyzing and visualizing S-CMC (e.g., chat and instant messaging), aimed at making the search for deception in large amounts of conversational data easier than searching by keywords and/or reading whole conversations. It is based on the work of Stolcke et. al. [11] on dialog act modeling, which utilizes n-gram language modeling and hidden Markov models to classify conversational utterances into 42 dialog act categories. Speech act profiling takes the probabilities (not the classifications) created by the combination of the language model and the hidden Markov model and sums them for the entire conversation, giving an estimate of the number of each of the dialog acts uttered by the participants. The probabilities for each participant can be separated and displayed on a graph. The resulting conversation profiles are useful in a number of situations, including visualizing multiple S-CMC conversations, testing hypotheses about the conversation's participants, and, of course, the post-hoc analysis of persistent conversations for deception. A full introduction to speech act profiling, including a guide to the speech acts and abbreviations, can be found in [1].

2.1 Speech Act Profiling for Deception Detection

Twitchell et. al. [9] demonstrated the promise of speech act profiling in finding deception in online conversations. Deceptive participants in three-person online conversations showed a significantly greater proportion of speech acts that express uncertainty than did their partners.

The study did have at least one shortcoming. The training corpus, which was used to create the speech act profiling model for detecting uncertainty, was the SwitchBoard corpus of telephone conversations. Though this is the largest corpus to be manually annotated with speech acts, it is nevertheless a collection of telephone conversations, not online conversations, which were the focus of the study. The authors argue that even though there are differences in language use between telephone and online conversations, the conversations must still be done in English that is understood by all parties of the conversation and is therefore manageably different for the purposes of speech act classification. That might be true, but using a corpus that is more similar in language use should produce better results. Furthermore, in the SwitchBoard corpus, participants are dyads discussing a number of general topics, whereas the data used in the study were chat logs from a three-person online game. Such differences could cause problems with the results of the study.

To alleviate this shortcoming, we undertook the current investigation, which employed an annotated corpus (called the StrikeCom corpus) of online conversations from the chat logs of the three-person games in the previous study. The StrikeCom corpus' "speech act" annotations themselves were created for a different study, and therefore are not the optimal codings for studying deception detection nor are they speech acts in the classical sense, but, as with the SwitchBoard corpus, given that the coding had already been completed, we sought to take advantage of a potentially fruitful opportunity.

The acts annotated in StrikeCom corpus include two kinds of acts. The first are what will be referred to as dialog acts. These acts are those that express some intent for uttering the message. They include the following:

- **Statement.** Assertions or opinions about the state of the world that are not direct responses to questions. *I got an X on D4.*
- **Direction.** Requests for others to do something during the game, but not requests for information. *Put your optical satellite on D4.*
- **Question.** Requests for information. *What was the result on D4?*
- **Response.** Information given in response to a question. *I got an X.*
- **Frustration.** Expressing frustration or anger toward the game or another player. *I don't get it!*

In the previous research, uncertainty was greater in deceptive participants than in their partners. Unfortunately, the only act in the current set that relates to any of the uncertain acts from the previous research is the question. Therefore, attempting to confirm the previous work leads us to the following hypothesis.

H1: Deceivers will have a greater proportion of questions during their online conversations than non-deceivers.

The second kind of act deals with the content of the utterance as it relates to the flow of the game. These acts are specific to the game played by the subjects. Hence, they lack generalizability, but may still inform us on the usefulness of speech act profiling for deception detection. They will be referred to as communicative acts and include the following:

- **Strategy.** Utterances about how to organize game play and team structure.
- **Asset Placement.** Utterances advocating placing assets on specific game board locations.
- **Result.** Reports of the results of asset placement.
- **Off-topic.** Utterances that have no relevance to the game.

These more specific, game-focused communicative acts are less informed by theory than the dialog acts above or in [9]. It is therefore more difficult to create hypotheses based on these data. There is, however, at least one idea that can be considered. First, deception has been posited to be cognitively difficult. Because of this difficulty, deceivers tend to lie using fairly simple techniques such as using variants of names for aliases [12]. The deceivers in this situation were told to steer their partners in the game to attack the incorrect positions on a game board. The simplest way to accomplish this would be to sit back and allow other group members to play the game then simply report erroneous results to throw the team off. This approach would be consistent with other research showing deceivers often opt for a more subservient or nondominant stand in their communication, shifting responsibility for the conversational "ball" to other interlocutors [8, 13]. Therefore we can ask the following research questions:

RQ1: Will deceivers have a smaller proportion of utterances classified as strategy and asset placement than non-deceivers?

RQ2: Will deceivers have fewer total utterances than non-deceivers?

It would be interesting to explore other questions, but the small sample size does not allow for much additional exploratory work.

3 Methodology

The StrikeCom corpus originated in a study where subjects played a three-player military decision-making game named StrikeCom [14]. The game required teams to search a grid-based game board for enemy camps. Each player had two assets with which to search the board. During each of five turns, the players reviewed the board and submitted their search. At the end of each turn, the game returned one of three results: likely nothing found, uncertain, or likely something found. After the end of the five searching turns, the teams used the information to place bombs for destroying the enemy camps. During some games, one player was told to deceptively steer the group away from bombing what they thought were the correct locations. The game was designed to foster communication and allow experimenters to assign different players to various treatments such as deceptive and truthful. All communication was conducted through text-based chat, which was captured and now comprises the StrikeCom corpus.

As noted earlier, a portion of the corpus was annotated with the acts described in Section 2.3. This portion is comprised of 47 games containing a total of 7112 annotated utterances. This portion was used to train the speech act profiling model. The

resulting model was then applied to 33 games, 16 of which included a participant who was instructed to be deceptive. Participants in the other 17 games were not given any instructions related to deception. Running speech act profiling on these conversations resulted in estimates of the number of each speech acts uttered by a participant during the game. These were then divided by the total number of utterances produced by that participant during the game, resulting in the proportion of each speech act used during the game.

3.1 Findings

The hypothesis and research questions were tested two ways, within group and between groups. The within group test compares the proportion of acts a deceiver makes with those of his or her two partners. A paired t-test is used to test statistical significance. The between groups test compares the deceptive participant with his or her equivalent participant in a group with no deception. A two-sample t-test assuming equal variances is used to test significance. Employing two comparisons allows us to observe any interactional effect occurring within the group.

Table 1. Results of comparison between deceivers and non-deceivers both within group (using paired t-test) and between groups (using standard t-test). * indicates significance at .1 level, ** .05 level, and *** at .01 level

	Deceivers	*Non-deceivers*			
		Within-group		*Between-groups*	
	Mean (stdev.)	Mean (stdev.)	p	Mean (stdev.)	p
Questions (H1)	0.17 (0.12)	0.12 (0.04)	0.08*	0.14 (0.05)	0.16
Strategy (RQ1)	0.03 (0.04)	0.03 (0.02)	0.25	0.08 (0.05)	<0.001***
Total Utterances (RQ2)	60.19 (38.48)	48.91 (27.34)	0.05**	52.64 (32.24)	0.54

As shown in Table 1, H1 was only somewhat supported ($p = 0.08$, one-tailed) within the group, which indicates that deceivers have a higher number of questions than their partners in the group. H1 was not supported when comparing the deceiver with an equivalent player in a group in which no one was instructed to deceive. RQ1, on the other hand, was strongly supported given that deceivers had a smaller proportion of utterances labeled as strategy than their equivalent players in other games with no deception ($p > 0.00$, one-tailed). RQ2 was significant in the opposite direction than predicted ($p = 0.05$, two-tailed).

3.2 Discussion

H1's lack of strong support is not surprising given that we only have questions as a crude measure of uncertainty. In the previous research that did show a significant difference with uncertainty, it was measured with a number of speech acts including questions, but also including a distinction between statements and opinions, certain

backchannels, and hedges. That the crude measure of uncertainty using only questions did show some difference reinforces the conclusion that deceivers express more uncertainty and that uncertainty should be automatically detectable using speech act profiling.

The strong support for RQ1 between the groups seems to confirm the premise expressed in Section 2.3 that deceivers were being cognitively lazy in their choice of how to deceive. Rather than attempting to change the strategy of the group, deceivers simply inserted misinformation into the results or didn't follow the strategy of the group when placing assets. The reverse conclusion from RQ2, however, seems to indicate that deceivers were full participants in the conversation. That participation, however, did not occur during the strategy phase of the game.

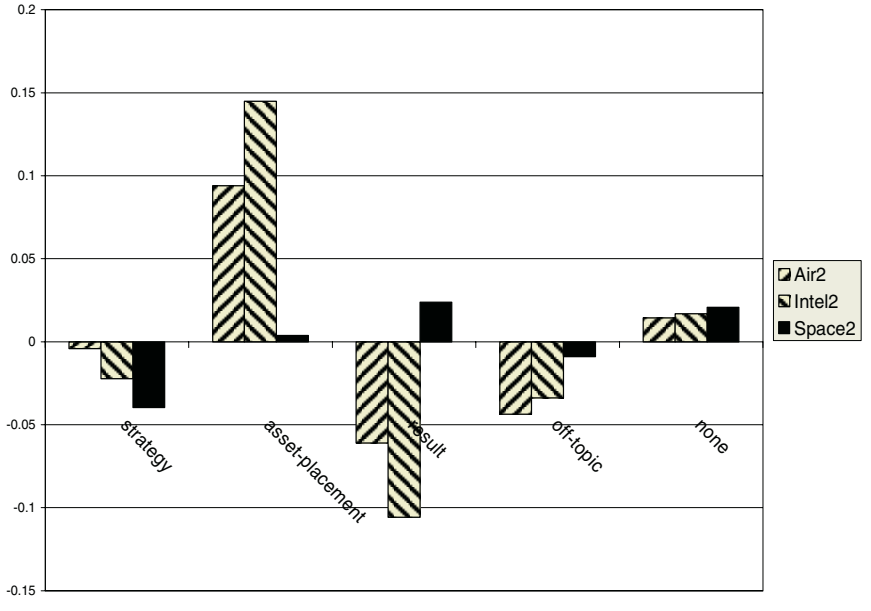


Fig. 1. Speech act profile of the example StrikeCom chat log

This behavior can be seen in an example game whose speech act profile is shown in Fig. 1 and from which the excerpt in Table 2 was taken. The profile shows that Space1, the deceiver in this case, has a smaller proportion of strategy utterances than normal (the zero line represents a perfectly "normal" conversation based on the averages of the testing corpus). Looking at the conversation itself in Table 2, we can see Space1 strategizing some in the beginning (17-18), but quickly deferring to the others (21-22). During the middle of the game Space1 again defers to the others by asking them what they think should happen (65 and 109). Finally, at the end of the game, Space1 again defers to the others (143), but in the end attempts to make a quick exit (151). These behaviors seem to indicate that Space1 is working assiduously to evade detection. Terminating the activity is one means of preventing his or her unmasking.

Table 2. An excerpt from a StrikeCom chat log where Space1 was instructed to be deceptive. Ellipses indicate where portions of the conversation were cut out. The total conversation length was 179 utterances. Fig. 1 is a speech act profile of the entire conversation

Seq. Num.	Player	Utterance
17	Space1	ok, ill take te bottom right. someone take the center
18	Space1	and someone take the top left
19	Intel1	in three turns we can get all the spaces on the board checked and then the next two turns we can double check
20	Air1	top left
21	Space1	good point
22	Intel1	ok, so space, are you taking columns 1 and 2?
…	...	…
63	Intel1	I'm going to go back and look over the first 3 and say which ones I got strikes on and which ones I got questionmarks on again
64	Air1	Its different for each command that is what I saying
65	Space1	ok... so what now?
…	…	….
108	Space1	so what do you guys think?
109	Intel1	ok, definitely f4 then
…	…	…
141	Intel1	ok, I have bombs on those 4, want me to submit it like that?
142	Air1	Because I got two strikes there
143	Space1	ok, so which do you guys want, D6 or F4. i know its not both
144	Intel1	who got stuff on f4 and f6?
145	Air1	I had both strikes
146	Space1	i didnt get f4
147	Intel1	I got f4
148	Space1	F3, F6, E6
149	Intel1	and I got a check on e6
150	Air1	not F6 actually
151	Space1	ok im submitting mine... i have a paper to go home and work on. F3, E6, F6

4 Conclusion

The most promising conclusion that can be made from this study is the apparent ability to detect uncertainty in S-CMC. The detected uncertainty could be used along with other indicators from previous studies to attempt to classify S-CMC messages and participants as having deceptive intent. Also shown was that the use of domain-

specific communicative acts (e.g., strategy utterances in the StrikeCom domain) may be useful for detecting some types of deception in the context of that domain. In other S-CMC domains, other communicative acts might be important, but using speech act profiling as a tool for finding communicative acts that indicate deception shows promise. These advances lend support to the recent proposition that the detection of deception in CMC can be improved with automated tools.

References

1. Twitchell, D.P. and J.F. Nunamaker Jr. Speech Act Profiling: A probabilistic method for analyzing persistent conversations and their participants. in Thirty-Seventh Annual Hawaii International Conference on System Sciences (CD/ROM). 2004. Big Island, Hawaii: IEEE Computer Society Press.
2. DePaulo, B.M., et al., Cues to Deception. Psychology Bulletin, 2003. 129(1): p. 75-118.
3. Zuckerman, M. and R.E. Driver, Telling Lies: Verbal and Nonverbal Correlates of Deception, in Multichannel Integrations of Nonverbal Behavior, A.W. Siegman and S. Feldstein, Editors. 1985, Lawrence Erlbaum Associates: Hillsdale, New Jersey.
4. Zhou, L., et al., Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: An empirical investigation. Group Decision and Negotiation, 2004. 13(1): p. 81-106.
5. Zhou, L., et al., Toward the Automatic Prediction of Deception - An empirical comparison of classification methods. Journal of Management Information Systems, 2004. 20(4): p. 139-166.
6. Zhou, L., J.K. Burgoon, and D.P. Twitchell. A longitudinal analysis of language behavior of deception in e-mail. in Intelligence and Security Informatics. 2003. Tucson, Arizona: Springer-Verlag.
7. Marett, L.K. and J.F. George, Deception in the Case of One Sender and Multiple Receivers. Group Decision and Negotiation, 2004. 13(1): p. 29-44.
8. Zhou, L., et al., Language dominance in interpersonal deception in computer-mediated communication. Computers and Human Behavior, 2004. 20(3): p. 381-402.
9. Twitchell, D.P., J.F. Nunamaker Jr., and J.K. Burgoon. Using Speech Act Profiling for Deception Detection. in Lecture Notes in Computer Science: Intelligence and Security Informatics: Proceedings of the Second NSF/NIJ Symposium on Intelligence and Security Informatics. 2004. Tucson, Arizona.
10. Burgoon, J.K., et al., The role of conversational involvement in deceptive interpersonal interactions. Personality \& Social Psychology Bulletin, 1999. 25(6): p. 669-685.
11. Stolcke, A., et al., Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. Computational Linguistics, 2000. 26(3): p. 339-373.
12. Wang, G., H. Chen, and H. Atabakhsh, Automatically detecting deceptive criminal identities. Communications of the ACM, 2004. 47(3): p. 70-76.
13. Burgoon, J.K., et al., Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. Communication Research, 1996. 23: p. 724-748.
14. Twitchell, D.P., et al. StrikeCOM: A Multi-Player Online Strategy Game for Researching and Teaching Group Dynamics. in Hawaii International Conference on System Sciences (CD/ROM). 2005. Big Island, Hawaii: IEEE Computer Society.

Active Automation of the DITSCAP

Seok Won Lee, Robin A. Gandhi, Gail-Joon Ahn, and Deepak S. Yavagal

Department of Software and Information Systems,
The University of North Carolina at Charlotte, Charlotte, NC 28223
{seoklee, rgandhi, gahn, dsyavaga}@uncc.edu

Abstract. The Defense Information Infrastructure (DII) connects Department of Defense (DoD) mission support, command and control, and intelligence computers and users through voice, data, imagery, video, and multimedia services, and provides information processing and value-added services. For such a critical infrastructure to effectively mitigate risk, optimize its security posture and evaluate its information assurance practices, we identify the need for a structured and comprehensive certification and accreditation (C&A) framework with appropriate tool support. In this paper, we present an active approach to provide effective tool support that automates the DoD Information Technology Security C&A Process (DITSCAP) for information networks in the DII.

1 Introduction

The DoD increasingly relies on software information systems, irrespective of their level of classification, in order to perform a variety of functions to accomplish their missions. The DITSCAP provides an excellent platform to assess the security of software information systems from organizational, business, technical and personnel aspects while supporting an infrastructure-centric approach. However, the lack of an integrated C&A framework and tool support often diminishes its effectiveness. DITSCAP itself can be quite overwhelming due to its long and exhaustive process of cross-checks and analysis which requires sifting through a multitude of DITSCAP polices and requirements. The complex interdependencies that exist between information from such large and diverse sources, significantly restricts human ability to effectively comprehend, develop, configure, manage and protect these systems.

To address these shortcomings and enhance the effectiveness of DITSCAP, we discuss our design principles, modeling techniques and supporting theoretical foundations that lead to the conceptual design of the DITSCAP Automation Tool (DITSCAP-AT). DITSCAP-AT aggregates C&A related information from various sources using a uniform representation scheme, and transforms static record keeping repositories into active ones that link to each other from different perspectives, allowing for their reuse and evolution through all stages of the system C&A lifecycle. DITSCAP-AT combines novel techniques from software requirements engineering and knowledge engineering to leverage the power of ontologies [10] for representing, modeling and analyzing DITSCAP-oriented requirements, while actively assisting the discovery of missing, conflicting and interdependent pieces of information that are critical to assess DITSCAP compliance.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 479 – 485, 2005.

2 DITSCAP Overview and Objectives for Its Automation

DITSCAP is a standard DoD process for identifying information security requirements, providing security solutions, and managing information systems security activities [3] for systems in the DII. DITSCAP certification is a "*comprehensive evaluation of the technical and non-technical security features of an information system and other safeguards made in support of the accreditation process, to establish the extent to which a particular design and implementation meets a set of specified security requirements"* [3]. Ensuing certification, the accreditation statement is an approval to operate the information system in a particular security mode using a prescribed set of safeguards at an acceptable level of risk by a designated approving authority. DITSCAP distributes its activities over four phases that range from the initiation of the C&A activities to its maintenance and reaccreditations. The level of rigor in each phase depends on the certification level chosen for the information system among the four levels available [2]. The security plan for DITSCAP is documented in the Software Security Authorization Agreement (SSAA) to "*guide actions, document decisions, specify IA requirements, document certification tailoring and level-of-effort, identify potential solutions, and maintain operational systems security*" [3].

Although the DITSCAP application manual [2] outlines the C&A tasks and activities along with associated roles and responsibilities of C&A personnel, they are expressed at an abstract level to maintain general applicability. Such abstractness makes it hard to ensure objectivity, predictability and repeatability in interpreting and enforcing DITSCAP requirements and policies. Furthermore, an entirely manual approach to cross-reference a multitude of DITSCAP-oriented directives, security requisites and policies in the light of user/system criteria to determine applicable security requirements raises serious concerns about the accuracy and comprehensiveness of such assessments. A structured and comprehensive method to assess and monitor the operational risk of information systems is also missing in the current approach.

To address the above shortcomings, the first and foremost objective of DITSCAP automation is to effectively assess the extent to which an information system meets the DITSCAP-oriented security requirements by supporting the process of identifying, interpreting and enforcing the applicable requirements based on user/system criteria. To reduce the amount of long and exhaustive documentation, carefully designed interfaces need to be developed that guide the user interactions through the DITSCAP tasks and activities. These interfaces should leverage thoroughly designed questionnaires and criteria, extracted from DITSCAP related C&A goals, directives, security requisites and other widely accepted best practices. These questionnaires along with predefined answers become the basis for building well defined metrics and measures that encompass the scope of the C&A goals addressed by them. The DITSCAP automation also demands structured, justifiable and repeatable methods to have for a comprehensive risk assessment, providing a firm basis to create cost versus risk measures. To actively discover and monitor network vulnerabilities, DITSCAP automation requires network self-discovery capabilities that allow comparison between the intended and the actual operational environment. Currently we limit the scope of DITSCAP-AT to level one DITSCAP certification as applied to Local Area Network (LAN) systems only. In the following section, we present the DITSCAP-AT

conceptual architecture conceived through our analysis to accomplish the aforementioned objectives.

3 DITSCAP-AT Conceptual Architecture

The conceptual architecture of DITSCAP-AT is shown in Fig. 1. The Process-driven Workflow module guides the DITSCAP through a well-defined course of action that results in the elicitation of required user criteria and generation of the SSAA. The tasks contained in each process component (P_1, P_2…P_n) are extracted from the DITSCAP application manual [2] and homogenously grouped based on their interdependent goals/objectives. Each task is then further expressed using carefully designed questionnaires/forms embedded in wizard-based interfaces to gather and establish well-defined C&A metrics and measures.

The Requirements Repository module provides a complete ontological engineering support for DITSCAP-AT. It provides utilities to support representation of security requirements, meta-knowledge creation, ability to query pre-classified and categorized information structures and other browsing and inference functionalities. The requirements repository is a specialized module built upon the GENeric Object Model (GenOM) [6], an integrated development environment for ontological engineering processes with functionalities to create, browse, access, query and visualize associated knowledge-bases (Ontology + Rules).

The Multi-strategy Machine Discovery module supports network self-discovery capabilities that allow the comparison of intended and operational environments. A set of network tools are selected on the basis of the information required to assess DITSCAP compliance, such as hardware, software and firmware inventories, configurations of network devices and services, and vulnerability assessment using penetration testing. A combination of network discovery tools and scripts enables to gather and fuse aggregated information as meta-knowledge in the requirements repository, which is then suitably transformed for inclusion in the SSAA.

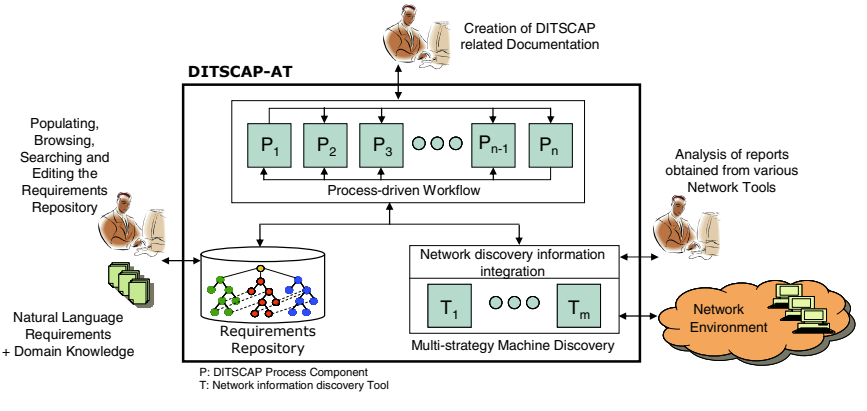


Fig. 1. DITSCAP-AT Conceptual Architecture

The process components of the Process-driven Workflow module retrieve applicable requirements/policies/meta-knowledge from the Requirements Repository and network discovery/monitoring information from the Multi-strategy Machine Discovery module, to actively assist the user in the C&A process.

In the following section, we discuss the use of information aggregated by DITSCAP-AT to achieve the objectives of DITSCAP automation.

4 The DITSCAP Automation Framework

In order to actively support the C&A process, uniformly across the DII, we create a DITSCAP Problem Domain Ontology (PDO) that provides the definition of a common language and understanding of the DITSCAP domain at various levels of abstractions through the application domain concepts, properties and relationships between them. The PDO is a machine understandable, structured representation of the DITSCAP domain captured using an object oriented ontological representation in the Requirements Repository. We elaborate more on methods and features for deriving the PDO in [7]. To satisfy the objectives of DITSCAP automation, the PDO specifically includes structured and well defined representations of: 1) A requirements hierarchy based on DITSCAP-oriented directives, security requisites and policies; 2) A risk assessment taxonomy that includes links between related risk sources and leaf node questionnaires with predictable answers that have risk weights and priorities assigned to them; 3) Overall DITSCAP process aspect knowledge that includes C&A goals/objectives; 4) Meta-knowledge about information learned from network discovery/monitoring tools; and 5) Interdependencies between entities in the PDO.

One of the objectives of DITSCAP-AT is to assess the extent to which an information system meets the DITSCAP-oriented security requirements by supporting the process of identifying, interpreting and enforcing the applicable requirements based on user criteria. The PDO supports such features through a requirements hierarchy that is constructed by extracting requirements from DITSCAP-oriented security directives, instructions, requisites and policies. A hierarchical representation includes high-level Federal laws, mid-level DoD/DoN policies, and site-specific requisites in the leaf nodes, which naturally corresponds to generic requirements, domain spanning requirements and sub-domain requirements in the requirements hierarchy. Also, there exists several non-taxonomic links that represent relationships within the requirements hierarchy as well as with other entities in the PDO.

A requirements hierarchy, therefore, allows the determination of applicable security requirements by successively decomposing the high-level generic requirements into a set of specific applicable requirements in the leaf nodes based on user criteria. Furthermore, the non-taxonomic links can be utilized to effectively interpret and enforce requirements by identifying the related requirements in other categories as well as relationships with entities in various dimensions from the PDO to ensure a comprehensive coverage of the C&A process.

To address the needs for a structured, justifiable and repeatable method for a comprehensive risk assessment, the PDO includes a risk assessment taxonomy which aggregates a broad spectrum of all possible categories and classification of risk related information. The risk assessment goals expressed in the higher level non-leaf nodes of

this taxonomy can be achieved using specific criteria addressed in the leaf nodes. For example, the risk taxonomy in the upper level non-leaf nodes consists of threat, vulnerabilities, countermeasures, mission criticality, asset value and other categories related to risk assessment. Each non-leaf node is then further decomposed into more specific categories. In addition, several non-taxonomic links identify relationships with other risk categories as well as with other entities in the PDO. Current scope of the risk related categorization is mainly based on the National Information Assurance Glossary [1] as well as other sources such as the DITSCAP Application Manual [2] and the DITSCAP Minimal Security Checklist [2]. We also utilize the information security metrics that have been established by the National Institute of Standards and Technology [8], [9].

A predictable and quantitative risk assessment is carried out using weights assigned to pre-classified answers for specific questions/criteria in the leaf nodes. These answers can be elicited from a variety of sources such as DITSCAP-AT users, network self-discovered information, or other sources. Furthermore, the questions/criteria in the leaf nodes of the risk assessment taxonomy naturally relate to various security requirements in the requirements hierarchy by expressing their testability in the form of criteria to measure their level of compliance. Such relationships along with the priorities/weights/criticalities associated with answers to questions/criteria in the leaf nodes of the risk assessment taxonomy can be used to develop complex risk calculation algorithms and establish metrics and measures that enable further articulation of critical weakest points in the system. The risk assessment taxonomy also promotes a uniform and comprehensive interpretation of different risk categories that are established through a common understanding of the concepts, properties and relationships that exist in the DITSCAP PDO. Such a shared understanding is inevitable to effectively estimate the collective impact of residual risk from all supporting systems on the overall critical infrastructure.

To populate the models discussed here, we have designed several core mock interfaces for DITSCAP-AT to realize a complete course of action for gathering and analyzing the required information [7]. Such mock interfaces provide a thorough understanding of the important aspects of DITSCAP-AT user interaction and offer valuable insight and assurance in realizing the theoretical aspects of DITSCAP automation.

5 Multi-dimensional Link Analysis

The root of Multi-Dimensional Link Analysis (MDLA) lies in the concept of proxy viewpoints model from the PVRD methodology proposed by Lee [5] to discover missing requirements and relationships. Lee suggests that "*Individual pieces of information finally become valuable knowledge when they establish 'links' with each other from various aspects/dimensions based on a certain set of goals*". Following this paradigm, MDLA can be carried out from different dimensions such as user criteria, viewpoints [4], system goals, business/mission requirements, regulatory requirements, specific operational concepts, and risk categories based on the DITSCAP C&A goals which can help understand various interdependencies between DITSCAP-oriented requirements, facilitating their interpretation and enforcement. The DITSCAP PDO that resides in the requirements repository fosters such analysis due to its ontological

characteristics that provides inherent properties for an active approach to link requirements and other entities from different perspectives and dimensions. MDLA's integrated framework for analytical analysis promotes assurance for a comprehensive coverage of the certification compliance space by actively assisting the process of discovering missing, conflicting, and interdependent pieces of information as well as establishing C&A metrics and measures based on common understanding and the reflected language from various dimensions.

6 Conclusion and Future Work

DITSCAP-AT contributes to the automation of DITSCAP in several ways. Firstly, it provides an effective tool support to identify, interpret and enforce DITSCAP polices and requirements. Secondly, it provides a structured and comprehensive approach to risk assessment from a broad spectrum of categories contributing to risk and finally, the ability to perform multi-dimensional link analysis provides the opportunity to reveal the "emergent" or "missing" information pieces that in-turn provides the assurance of a comprehensive coverage of the certification compliance space.

Our future work includes the software realization of DITSCAP-AT mock interfaces while systematically realizing all its core functional components. Although we limit the current scope of DITSCAP-AT to include DoD directives, security requisites and best practices for secure software development, it can be easily scaled to accommodate general security requirements, policies and practices in any domain of interests. We also realize that development of appropriate metrics and measures for a comprehensive and uniform risk assessment in the DITSCAP domain is an area that requires significant attention for the success of DITSCAP-AT.

Acknowledgements

This work is partially supported by the grant (Contract: N65236-05-P-0597) from the Critical Infrastructure Protection Center (CIPC), Space and Naval Warfare (SPAWAR) Systems Center, Charleston, SC. USA. We acknowledge the support and encouragement from Scott West, John Linden, Bill Bolick, and Bill Chu. Finally, we thank Divya Muthurajan and Vikram Parekh for their contributions to this research.

References

1. Committee on National Security Systems (CNSS) Instruction No. 4009.: National Information Assurance (IA) Glossary. (2003)
2. DoD 8510.1-M: DITSCAP Application Manual (2000)
3. DoD Instruction 5200.40.: DITSCAP (1997)
4. Kotonya, G. and Sommerville, I.: Requirements Engineering with Viewpoints. BCS/IEEE Software Engineering Journal, Vol. 11, Issue 1 (1996) 5-18
5. Lee, S.W. and, Rine D.C.: Missing Requirements and Relationship Discovery through Proxy Viewpoints Model. Studia Informatica Universalis: International Journal on Informatics, December (2004)

6. Lee, S.W. and, Yavagal, D.: GenOM User's Guide. Technical Report: Dept. of Software and Information Systems, UNC Charlotte (2004)
7. Lee, S.W., Ahn, G. and Gandhi, R.A.: Engineering Information Assurance for Critical Infrastructures: The DITSCAP Automation Study. To apprear in: Proceedings of the Fifteenth Annual International Symposium of the International Council on Systems Engineering (INCOSE '05), Rochester New York July (2005)
8. Swanson, M., Nadya, B., Sabato, J., Hash, J., Graffo, L.: Security Metrics Guide for information Technology Systems. NIST #800-55 (2003)
9. Swanson, M.: Security Self-Assessment Guide for Information Technology Systems. NIST #800-26 (2001)
10. Swartout, W. and Tate, A.: Ontologies. In: Intelligent Systems, IEEE, Vol. 14(1) (1999)

An Ontological Approach to the Document Access Problem of Insider Threat

Boanerges Aleman-Meza[1], Phillip Burns[2], Matthew Eavenson[1], Devanand Palaniswami[1], and Amit Sheth[1]

[1] LSDIS Lab, Department of Computer Science, University of Georgia, Athens, GA 30602
{boanerg, amit}@cs.uga.edu
{durandal, devp}@uga.edu
[2] Computer Technology Associates, 7150 Campus Drive, Ste 100, Colorado Springs, CO 80920
phillip.burns@cta.com

Abstract. Verification of legitimate access of documents, which is one aspect of the umbrella of problems in the Insider Threat category, is a challenging problem. This paper describes the research and prototyping of a system that takes an ontological approach, and is primarily targeted for use by the *intelligence community*. Our approach utilizes the notion of *semantic associations* and their discovery among a collection of heterogeneous documents. We highlight our contributions in (graphically) capturing the scope of the investigation assignment of an intelligence analyst by referring to classes and relationships of an ontology; in computing a measure of the relevance of documents accessed by an analyst with respect to his/her assignment; and by describing the components of our system that have provided early yet promising results, and which will be further evaluated more extensively based on domain experts and sponsor inputs.

1 Introduction

Insider Threat refers to the potential malevolent actions by employees within an organization, a specific type of which relates to legitimate access of documents. In the context of the intelligence community, one of the goals is to ensure that an analyst accesses documents that are relevant to his/her assigned investigation objective, i.e., accesses the data on a "need to know" basis.

In this paper we discuss our work as part of an Advanced Research and Development Activity (ARDA) funded project, in which we have developed an ontological approach to address the *legitimate document access* problem of Insider Threat. There is a range of techniques that support determining if a collection of documents is relevant to a particular domain. Such techniques can be applied to determine if documents accessed by an intelligence analyst are relevant to his/her job assignment. Examples include statistical, NLP, and machine learning techniques such as those leading to

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 486 – 491, 2005.

document clustering and/or automatic document classification that exploit implicit semantics[1]. A concern with these approaches is that they generally do not support an ability to clearly understand the reasons behind why an accessed document is relevant (or not relevant) to the investigation objective of the intelligence analyst. Most of these techniques have also focused on mapping documents to a predefined taxonomy, which is found to be a rather limited method of representing knowledge when named relationships between concepts (e.g., a person *works-for* an organization) represent an important part of the domain knowledge. In this context, we pursue a strategy that uses ontology to capture domain semantics and semantic metadata to capture semantics of heterogeneous domains.

In our approach, we utilize *semantic associations*, which aim to capture meaningful and possibly complex relationships between entities (in a large dataset of metadata based on a graph model) [3]. Initially we sought to leverage our previous experience where we have applied such associations to a class of national security and homeland security applications (e.g., Passenger Threat Assessment [7]). The need to represent the scope of the investigative assignment given to an analyst required us to take a fresh look at our previous work in capturing a user's interest with respect to an ontology (or subset thereof) [1]. Additional technical challenges include the need to compute a large number of semantic associations per document. Scalability becomes an issue given the potentially large collection of documents to be analyzed. For our ontological approach, a starting point was the building of a populated ontology. In doing so, we have built upon our significant experience in the development of large populated ontologies (e.g., [2], Glycomics Ontology[2]).

This paper presents the following novel conceptual and technical contributions:

- A practical yet flexible notion of capturing the scope of the investigation assignment of an analyst in terms of semantic constraints over an ontology. We call it the *context of investigation*, and we specify it using a graphical user interface to be used by the supervisor or investigator associated with an analyst's assignment.
- A computational measure that exploits *semantic associations* in a novel way to determine the relevance of a document with respect to a context of investigation.
- A prototype tested with a small-to-medium but representative document set.

Since we have not completed a comprehensive evaluation and have not fully evaluated scalability challenges, we present this work as a short paper. A comprehensive literature overview is also not presented for brevity.

2 Our Ontological Approach to the Legitimate Access Problem

Figure 1 provides a schematic of our approach. We use a large ontology populated from trusted sources to semantically annotate a collection of documents (viewed by

[1] Implicit semantics (as used here) capture possible relationships between concepts, but cannot or do not name specific relationships between the concepts. Explicit semantics use named relationships between concepts, and in the context of recent Semantic Web approaches, often use ontologies represented using a formal language; for further discussion, see [8].

[2] http://lsdis.cs.uga.edu/Projects/Glycomics/

an intelligence analyst). The system provides a means to define a *context of investigation* that aims to capture, in ontological terms, the scope of an investigation assignment given to an intelligence analyst. Hence, the goal is to measure the relevance of each document (using the annotations), with respect to the context of investigation. The documents are then grouped based on that measure (using a user-customizable threshold). Additionally, each document can be inspected by a supervisor to gain insight on the purpose of access by the analyst (beyond the "need to know"). The system supports this task by graphically displaying the *semantic associations* that interconnect entities in a document to those that form part of the context of investigation.

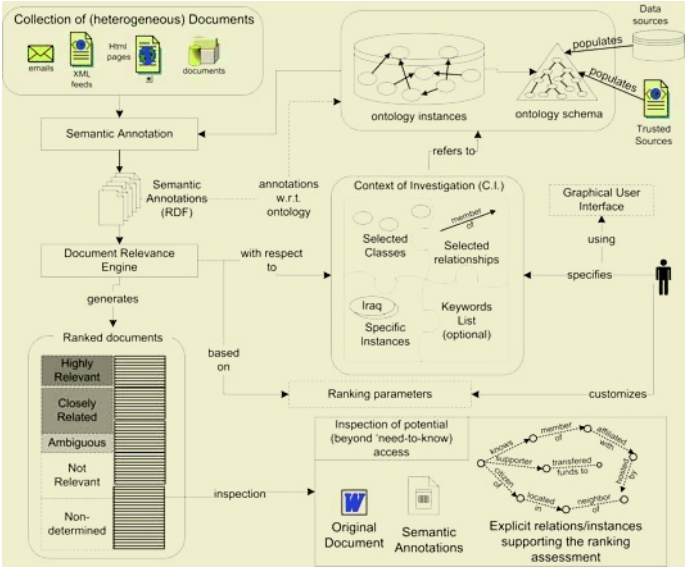


Fig. 1. Schematic of Ontological Approach to the 'Legitimate Document Access' Problem

a) Ontology Specification and Development

The ongoing Semantic Discovery project at the LSDIS lab has created (and maintains) a test-bed (SWETO) for evaluating semantic technologies [2]. We used and refined a subset of SWETO focusing on the domain of National Security and Terrorism. It was populated with real-world publicly available data maintained by international organizations. For ontology design and population, we used Semagix's Freedom[3], a commercial software based on earlier research developed at and licensed from the LSDIS lab [6]. The ontology consists of about 40 classes, populated with about 32,000 entities and about 35,000 explicit relationships among them.

b) Context of Investigation

The intuition behind a context of investigation lies in capturing, at an ontology level, the types of entities and relationships that are to be considered important. The context can contain semantic constraints. For example, it can be specified that a relation 'af-

[3] http://www.semagix.com

filiated with' is part of the context only when it is connected with an entity that belongs to a specific class, say, 'Terror Organization'. The *context of investigation* is a combination of (i) entity classes; (ii) entity instances; (iii) named relationships between entity classes. Our prototype supports a graph-based user interface for defining a *context of investigation* (using TouchGraph[4]).

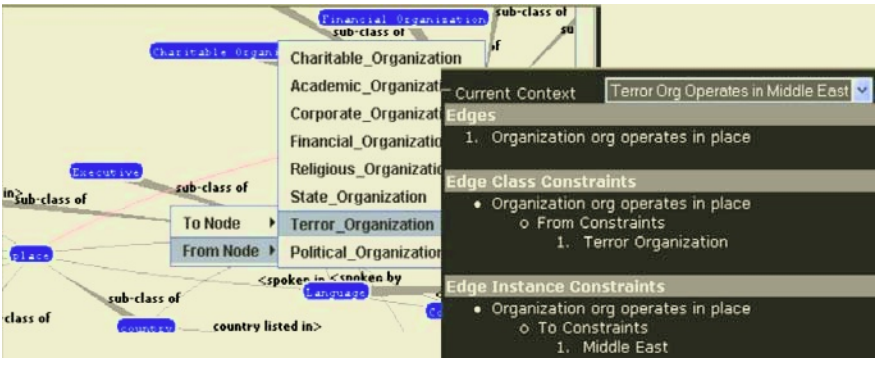


Fig. 2. Specifying Context of Investigation

Figure 2 displays an example of a context of investigation where *Middle Eastern Terrorism* is specified using the relationship "*Organization*" *operates_in* "*Place*" constrained into: "*Terror Organization*" *operates_in* "*Middle East*".

c) Semantic Annotation

The documents viewed by the analyst are processed to produce *semantically annotated documents*. Semantic annotation is metadata (data that describes data) added to a document, and also the process of generating such metadata[5]. Semagix's Freedom software was used to semantically enhance the documents that an analyst accessed as part of the assignment. The Freedom software searches the document and picks out entity names or synonyms within the document that are contained in the ontology.

d) Relevance Measure for Documents

The *Documents Relevance Engine* measures the relevance of annotated documents with respect to (w.r.t.) the context of investigation. The engine takes as input the set of semantically annotated documents, the context of investigation for the assignment, the ontology schema represented in RDF[6], and the ontology instances represented in RDF. The goal is to provide a ranked list of the documents based on their relevance to the assignment (represented using context of investigation described in Section 2.2). The documents relevance engine measures the relevance of the entity annotations in an annotated document w.r.t. the context of investigation. The relevance score of each document to the context of investigation is computed using semantic associations. A

[4] http://www.touchgraph.com
[5] For example, the KIM Platform http://www.ontotext.com
[6] Resource Description Framework, http://www.w3.org/RDF/

formalization of Semantic Associations over metadata represented in RDF was presented in [3]. Here we provide an adapted definition.

Definition 1 (ρ-*Semantic Association*): Two entities e_1 and e_n are semantically associated if there exists a sequence $e_1, P_1, e_2, P_2, e_3,..., e_{n-1}, P_{n-1}, e_n$ in an RDF graph where e_i, $1 \leq i \leq n$, are entities and P_j, $1 \leq j < n$, are relationships.

The relevance measure of a document d considers four components as follows:

$$Relevance(d) = C_{CI} + R_{CI} + E_{CI} + K_{CI} \quad (1)$$

where, C_{CI} is the component of matching classes with respect to *CI*. Similarly, R, E, and K are the components for matching relations, entities, and keywords, respectively. In our system, we pre-compute semantic associations for each document up to a (fixed) association length n. That is, a neighborhood of n hops from the entities on the document (similar to the intuition of a 'semantic neighborhood' described in [5]).

C_{CI} is computed based on whether there is a match of the types of the entities of the document and its neighborhood with respect to the context of investigation,

$$C_{CI} = \frac{\sum_{e_j \in d} \left[\sum_{i=1}^{|ng(e_j)|} \frac{1}{dist(e_j, v_i)+1} \right]}{|d|} \quad (2)$$

where, $ng(e)$ is the set of nodes and relationships in the neighborhood of entity e; and the function $dist(e, v)$ computes the distance between e and v. Computing components R_{CI}, and E_{CI} proceeds in similar fashion. In the component for keywords, K_{CI}, the formula differs by considering all attributes of each entity v_i with those keywords specified in the context of investigation. We plan to incorporate into the formula for K_{CI} a simplified version of the ideas presented in [4].

3 Initial Results and Conclusions

Our initial experiments were conducted on a collection of 1000 documents. A few example results in the context of Middle-Eastern Terrorism discussed in Section 2.2 are provided here. A high score of 0.91 was calculated for a document on 'Ansar al-Islam' where the semantic association *Ansar al-Islam –operates in→ Middle East* relates it to the context. A score of 0.735 for a document on Abu Sayyaf was the result of the (longer) semantic association *Abu Sayyaf Group –affiliated with→ Al Qaeda –operates in→ Middle East*. A low score of 0.425 was calculated for a document on the Sri Lankan group 'LTTE' due to the long semantic association *Sriperumbudur –located in→ India ←national of– Dawood Ibrahim –affiliated with→ Al Qaeda –operates in→ Middle East*.

We acknowledge that further evaluations are needed, but early results are promising and provide useful insights for our future work. An online demo is available[7].

[7] http://lsdis.cs.uga.edu/Projects/SemDis/NeedToKnow/

Our approach has several advantages, including: (a) capability to keep the ontology updated (this becomes particularly important in dealing with changing and/or new information, e.g., new data being posted in watch-lists); (b) a means to support inspection of the explicit relationships on why a document is relevant to the context of investigation. Thus the supervisor of the intelligence analyst is able to gain insight on the need-to-know reason for access to the document. Our next steps in this project include conducting extensive evaluations, and addressing quality and scalability issues.

Acknowledgements. This work is conducted as part of the Advanced Research Development Activity (ARDA) Insider Threat Initiative, contracted through the Department of the Interior, Ft. Huachuca, contract # NBCHC030083. The larger projects at the LSDIS lab which provide the basis for research in Semantic Association Discovery are funded by the National Science Foundation through Awards 0219649 ("Semantic Association Identification and Knowledge Discovery for National Security Applications"), and IIS-0325464 ("SemDis: Discovering Complex Relationships in Semantic Web"). We also acknowledge our collaboration with Semagix, Inc, which enabled our use of Semagix Freedom.

References

1. B. Aleman-Meza, C. Halaschek, I.B. Arpinar, A. Sheth, Context-Aware Semantic Association Ranking. Proceedings of Semantic Web and Databases Workshop, Berlin, September 7-8, 2003, pp. 33-50
2. B. Aleman-Meza, C. Halaschek, A. Sheth, I.B. Arpinar, and G. Sannapareddy. SWETO: Large-Scale Semantic Web Test-bed. Proceedings of the 16th International Conference on Software Engineering and Knowledge Engineering (SEKE2004): Workshop on Ontology in Action, Banff, Canada, June 21-24, 2004, pp. 490-493
3. K. Anyanwu and A. Sheth ρ-Queries: Enabling Querying for Semantic Associations on the Semantic Web The Twelfth International World Wide Web Conference, Budapest, Hungary, 2003, pp. 690-699
4. C. Rocha, D. Schwabe, M.P. Aragao. A Hybrid Approach for Searching in the Semantic Web, In Proceedings of the 13th International World Wide Web, Conference, New York, May 2004, pp. 374-383.
5. M.A. Rodriguez, M.J. Egenhofer, Determining Semantic Similarity Among Entity Classes from Different Ontologies, IEEE Transactions on Knowledge and Data Engineering 2003 15(2):442-456
6. A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke. Managing Semantic Content for the Web. IEEE Internet Computing, 2002. 6(4):80-87
7. A. Sheth, B. Aleman-Meza, I.B. Arpinar, C. Halaschek, C. Ramakrishnan, C. Bertram, Y. Warke, D. Avant, F.S. Arpinar, K. Anyanwu, and K. Kochut. Semantic Association Identification and Knowledge Discovery for National Security Applications. Journal of Database Management, Jan-Mar 2005, 16 (1):33-53
8. A. Sheth, C. Ramakrishnan, and C. Thomas, Semantics for the Semantic Web: the Implicit, the Formal and the Powerful, International Journal on Semantic Web and Information Systems, 2005, 1(1):1-18

Filtering, Fusion and Dynamic Information Presentation: Towards a General Information Firewall

Gregory Conti, Mustaque Ahamad, and Robert Norback

Georgia Tech Information Security Center, Georgia Institute of Technology,
801 Atlantic Avenue, Atlanta, Georgia, 30332-0280 USA
conti@acm.org

Abstract. Intelligence analysts are flooded with massive amounts of data from a multitude of sources and in many formats. From this raw data they attempt to gain insight that will provide decision makers with the right information at the right time. Data quality varies from very high quality data generated by reputable sources to misleading and very low quality data generated by malicious entities. Disparate organizations and databases, global collection networks and international language differences further hamper the analyst's job. We present a web based information firewall to help counter these problems. It allows analysts to collaboratively customize web content by the creation and sharing of dynamic knowledge-based user interfaces that greatly improve data quality, and hence analyst effectiveness, through filtering, fusion and dynamic transformation techniques. Our results indicate that this approach is not only effective, but will scale to support large entities within the Intelligence Community.

1 Introduction

Intelligence analysts are besieged with data from legitimate sources and the problem is compounded by active malicious entities attempting to subvert their work by injecting misleading or incorrect data into their information space. From this sea of data, analysts attempt to glean useful information that meets the intelligence needs of their customers. The rate at which data is being produced, combined with the immense amount of existing data, sets the stage for denial of information attacks against both analysts and their customers. Denial of Information (DoI) attacks are similar to Denial of Service (DoS) attacks against machines. While DoS attacks attempt to deny users access to system resources by consuming machine resources, DoI attacks target the human by exceeding their perceptual, cognitive and motor capabilities. In most cases, a small amount of malicious information is all that is required to overwhelm or deceive the human. A successful DoI attack occurs when the human does or does not take action they otherwise would have [1]. Denial of Information attacks are of critical importance to intelligence analysts. Every bit of time, albeit small, wasted on a false lead or due to information overload reduces the probability of timely and accurate intelligence. To counter Denial of Information attacks against analysts we employed collaborative, knowledge-based user interfaces that improve data quality. These interfaces, based upon filtering, fusion and dynamic transformation techniques, reduce the amount of irrelevant data (noise) and increase the useful information

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 492 – 497, 2005.

(signal) presented to the analyst. The following example demonstrates the operation of the system. In step one, an information producer generates a large web page of moderate quality data with a poorly designed interface. Note that the page would require the analyst to scroll through about eight screens of information. The page is rife with links to irrelevant information (such as advertisements) and is constructed with poorly chosen foreground and background colors. In step two, an analyst creates a transform for the page which dramatically improves the quality of information and interface. This transform is then shared via a centralized transform server in step three. In step four, other users, both intelligence customers and other analysts, can then browse/search the server for relevant transforms. These results contain ratings based on previous users' experiences as well as descriptions of the transforms. After a consumer selects a transform, step five, they see a dramatically altered version of the page. The result is a significantly more usable page with much higher information gain. The interface problems have been corrected and the page shows just the relevant information, but there is a link to the original source information, if required. The end user may then create a new transform or modify the existing transform and submit it to the server. In addition, the user may help collaboratively filter the transform itself by voting on its quality and usefulness. The cycle continues with new consumers utilizing and improving upon the transforms.

For this work, we assumed that analysts will operate primarily within private intranets. This assumption greatly reduces the legal implications of filtering content without the permission of the information producer. In particular, the act of removing paid advertisements from content is of questionable legality, even by intelligence analysts employed by government organizations. To avoid classification concerns, we tested the system using unmodified open source web pages. As we explore the broader applicability of the system, it is important to consider the threat model. Malicious meta-information producers could generate transforms designed to mislead or filter information from legitimate customers. In our work, we assumed that transforms would only be generated by trusted members of the organization. While not all transforms meet the needs of all customers, we believe, that given this assumption, the combination of filter descriptions, direct links to original unmodified source material and collaborative rankings of transforms will allow end users to find and utilize the transforms that they need.

2 Related Work

The uniqueness of this work springs from the distributed and collaborative approach to increase the quality of information accessed from data sources in order to provide better products to analysts and intelligence customers. We facilitate this human assisted collaborative analysis by incorporating the insights of both analysts *and* customers in the analytic loop. While the underlying technologies will change, the collaborative fusion, filtering and interface transforming approaches we present will be far more enduring. Communities of analysts and their customers can iteratively improve the quality of intelligence by creating and sharing dynamic user interfaces and information transforms based upon their tasks and needs.

The Galaxy of News system designed by Rennison is most directly applicable to our work [2]. The system employs visualization techniques and a relationship construction engine to build implicit links between related, but independently authored news articles. While we were influenced by this work, our work differs significantly. The primary differences are in the following areas. Galaxy of news relies upon the relationship construction engine to build links between news articles. Our work focuses instead upon using human moderated, collaborative transforms to increase the information quality of individual articles and to fuse together disparate information sources to meet customized user needs. In addition, we incorporate the notion of trust through the use of user rankings and links to original source content. Also, there exists a large body of work in the domain of web usability. See the work of Nielsen for excellent examples [3]. These works describe best practices and design techniques to create more usable and information rich web pages, but were designed for individual web designers to apply to their *own* content. Our system provides the mechanism for users to apply, share and improve upon the *work of others* to meet their own specific information needs. Given our assumption that transforms are only created by trusted members of the intelligence organization, we believe that our collaborative ranking system and link to the source document is sufficient to protect users from unintentional, badly designed transforms.

There have been several approaches, both centralized and decentralized, to filtering web content. The primary difference with our work is in the ability to collaboratively share, rate and improve upon filters created by others. BugMeNot [4] is an interesting centralized approach. The website seeks to increase information quality and access by bypassing login procedures through communally shared user ID's and passwords. While the overlap is minimal with our work, the sharing of access credentials to increase information access, avoid advertisements and protect privacy is worthy of examination. DOM Inspector [5] illustrates the current state of the art in the decentralized, browser, plug-in approach. Real Simple Syndication (RSS) is a push technology, incorporating eXtensible Markup Language (XML), that distributes news and other new content from websites to virtually any client platform. Clients can subscribe to various streams of interest and tools exist to fuse together multiple streams to form a single consolidated picture. It is important to consider the domain of intelligent information retrieval such as distributed retrieval, search strategies, semantic filtering, content indexing and information discovery. We believe that current and projected techniques in this area complement our work and can be incorporated, as applicable, into the information firewall architecture that we propose. Finally, two other classes of technology merit discussion: HTTP proxy content rewriting and inline content transforming devices. Representive of proxy content rewriting is Privoxy [6]. Its primary focus differs from ours in that it transforms content to remove advertisements and protect privacy of individual users. WebWasher [7] offers a suite of tools and inline appliance that transforms information streams between organizational users and external data sources. WebWasher is designed to protect enterprise networks from frivolous use and malware. It is typically used in an adversarial manner to enforce organizational policy on users.

3 Information Firewall Design and Implementation

The design of the system is based upon the analyst's need to acquire the highest quality information with a minimum of effort. Our core users are intelligence analysts, but they may also be any consumers of information. Hence, the primary design goal was to maximize the valuable information contained in web content while reducing unwanted information. We include in this definition the ability to modify the information interface and navigation structure to assist in improving the task-specific interaction and presentation of information. In other words, we wish to increase signal while decreasing noise such that the information is presented in a format and information architecture desired by users. It is important to note that the definition of valuable information will vary from analyst to analyst. Because of this need, our second design goal was to allow information consumers and producers to create, share and collaboratively rank information transforms. This capability affords information producers the opportunity to create initial transforms based on their best interpretation of customer requirements. Their end users could then use these transforms as is, or modify and layer them to meet their specific needs as well as share the resulting transform with other users. To allow efficient and effective sharing, our third design goal was to create centralized communities of trust where users could easily seek out and contribute transforms. To support decentralized sharing we wished to reduce the routine usage complexity of information producer-to-analyst and analyst-to-analyst sharing to the order of emailing a hyperlink or bookmarking the transform in a common browser. Transforming the content of information used by analysts runs the risk of masking important information. To counter this effect, our fourth design goal was to provide the analyst with easy access to the original source content.

Use of the system begins with analytic organizations generating and publishing content to network accessible web servers. This content can be in any web accessible format, but for purposes of our experimentation we focused exclusively on HTML documents. Optionally, information producers will publish a variety of suggested transforms based on their perception of analyst and other customer requirements. In order to gain the benefits provided by the information firewall, analysts typically do not access content directly. Instead, they perform a search of available information transforms on a web server hosted by the information firewall. This search is conducted based on the website name, URL and transform name. The search index returns transforms that match any of these values and are ordered alphabetically and by user rating. Alternatively, the analyst may perform a general search of the Internet using the Google search engine. The transform engine, acting as a proxy, sends the search to Google and receives the unmodified results. These results are modified by the transform engine to include applicable transform links associated with each search result before presentation to the user. For example, if the search was for newswebsite.com, the transform engine would add transform links associated with that URL to the Google search results returned to the user. These might include transforms such as "top stories," "international news" or "ad free news." If an analyst is unable to find a transform providing the desired functionality they have the option of creating a new transform using the center button. Upon submission, this filter is added to the transform database with an initial feedback rating of neutral. In general, transforms take *any* information object as input and convert it to another information object. The

rules to make the conversion are created by the user with the aid of the tools provided by the information firewall. After creation, the rules are stored in the information firewall's database. The user selects options they would like in a transform and registers the transform with the information firewall which then stores the parameters in its database. In this case, it is a web page filter that removes undesired information from the page as well as changes interface parameters to present information in the desired way. While, in our current prototype, we implemented a limited number of transforms, we believe this to be a powerful concept. In any instance where there exists a reasonable algorithmic solution to convert one information object to another this algorithm may be included in the firewall as the basis of a transform. Human end users may then interact with these algorithms to create transforms which customize the information they receive.

Transforms are executed by selecting a hypertext link that calls a CGI script on the transform engine. This link includes a numeric parameter representing the transform to execute as well as the URL of the source website. First, the transform engine queries the database for details of the transform using the numeric parameter as an index. The transform engine then contacts the URL and requests the content in question. As the content is returned from the URL it strips or adds content and dynamically generates the resulting page for the user. To facilitate trust, all transformed pages include a clearly delineated header that includes a link to the unaltered document and the author of the transform. A small, one-line form in the header allows the user to vote on the usefulness of the transform. Deliberately designed with simplicity in mind, the use of hypertext links to execute transforms is a key aspect of the system. As a result, users can share transforms easily via email and store them quickly as bookmarks.

4 Conclusions and Future Work

The information firewall we present is viable and useful within the Intelligence Community. It improves the efficiency and effectiveness of analysts by dramatically increasing the signal to noise ratio of intelligence data thereby easing the cognitive burden placed upon intelligence analysts as well as intelligence consumers. While the information firewall concept proved to be effective in the constrained environment of intelligence organization intranets, we plan to explore decentralized approaches to improve scalability. In addition to intelligence community applications, we wish to apply the approach to the larger commercial Internet. For this to be feasible, we must explore the legal ramifications of filtering, fusing and transforming data with and without the permission of the information source. Finally, we plan on extending our work to include a generic information firewall for all digital content. For this to be feasible, we must explore ways to better access the embedded knowledge within available data using such technologies as XML. We envision that the information transformation techniques presented in this paper will work extremely well in a variety of additional applications including increasing accessibility, e.g. via custom presentation to color-blind or vision impaired users, conversion from text to audio, streamlined language translation and web-based intelligence monitoring and analysis. We also believe that the transformation of information and interfaces will allow use on a wide variety of computing platforms, including very small devices such as per-

sonal digital assistants and those with severe bandwidth constraints. The ultimate strength of the system lies in the ease with which individual analysts may create robust, high-resolution information transforms. In the future we plan to investigate intuitive tools to support transform construction, perhaps using the visual web page editing paradigm.

Acknowlegment. This work was supported in part by the National Science Foundation Information Technology Research award 0121643.

References

1. Conti, Gregory and Ahamad, Mustaque. Countering Denial of Information Attacks. IEEE Security and Privacy. (to be published)
2. Rennison, Earl. Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes. Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology, 1994, pp. 3 - 12.
3. Nielsen, Jakob. Designing Web Usability. New Riders, 2000.
4. Bug Me Not. Frequently Asked Questions, http://www.bugmenot.com/faq.php, last accessed on 3 January 2005.
5. The Mozilla Organization. DOM Inspector. http://www.mozilla.org/projects/inspector/, last accessed on 3 January 2005.
6. Privoxy Project Homepage. http://www.privoxy.org/, last accessed on 3 January 2005.
7. Webwasher. Webwasher CSM Appliance and Suite. http://www.webwasher.com/, last accessed on 3 January 2005.

Intrusion Detection System Using Sequence and Set Preserving Metric

Pradeep Kumar[1, 2], M. Venkateswara Rao[1,2], P. Radha Krishna[1], Raju S. Bapi[2], and Arijit Laha[1]

[1] Institute for Development and Research in Banking Technology, IDRBT, Castle Hills, Masab Tank, Hyderabad, India -500057

[2] University of Hyderabad, Gochibowli, Hyderabad, India -500046

{pradeepkumar, mvrao, prkrishna, alaha}ibrbt.ac.in, bapics@uohyd.ernet.in

Abstract. Intrusion detection systems rely on a wide variety of observable data to distinguish between legitimate and illegitimate activities. In this paper we investigate the use of sequences of system calls for classifying intrusions and faults induced by privileged processes in Unix Operating system. In our work we applied sequence-data mining approach in the context of intrusion detection system (IDS). This paper introduces a new similarity measure that considers both sequence as well as set similarity among sessions. Considering both order of occurrences as well as content in a session enhances the capabilities of kNN classifier significantly, especially in the context of intrusion detection. From our experiments on DARPA 1998 IDS dataset we infer that the order of occurrences plays a major role in determining the nature of the session. The objective of this work is to construct concise and accurate classifiers to detect anomalies based on sequence as well as set similarity.

1 Introduction

Intrusion detection is the process of monitoring and analyzing the events occurring in a computer system in order to detect signs of security problems[1,2]. Computer security can be achieved by maintaining audit data. Cryptographic techniques, authentication means and firewalls have gained importance with the advent of new technologies. With the ever-increasing size of audit data logs it becomes crucial for network administrators and security analysts to use some efficient and automatic Intrusion Detection System (IDS), in order to reduce the monitoring activity. Machine learning and data mining are being investigated for IDS [3, 4, 5, 6]. The key idea to use data mining techniques is to discover consistent and useful patterns of system features that describe program and user behavior, and use the set of relevant system features to compute (inductively learned) classifiers that can recognize anomalies and known intrusions. Classification of IDS data stream is crucial, since the training dataset changes often. Building a new classifier each time is very costly with most techniques. There are various techniques for classification such as Decision Tree Induction, Bayesian Classification, and Neural Networks [10].

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 498 – 504, 2005.

Intrusion detection systems rely on a wide variety of observable data to distinguish between legitimate and illegitimate activities. In this paper we study one such observable sequences of system calls into the kernel of an operating system. Several different programs, to represent normal behavior accurately and to recognize intrusions, generate system call data sets. There are many ways in which system call data could be used to characterize normal behavior of programs, each of which involves building or training a model using traces of normal process. The methods described in [3,4] define normal behavior in terms of short sequences of system calls in a running process. In this work they demonstrated the usefulness of monitoring sequences of system calls for detecting anomalies induced by processes. This work depends on enumerating sequences that occur empirically in traces of normal behavior and subsequently monitoring for unknown patterns. Lookahead pairs were also used in [3]. The later paper reported that contiguous sequences of some fixed length gave better discrimination than lookahead pairs [4].

A scheme based on the k-Nearest Neighbor (kNN) Classifier has been proposed by Liao and Vemuri[8], in which each process is treated as a document and each system call as a word in that document. Process containing sequence of system calls are converted into vectors and cosine similarity measurement is used to calculate similarity among processes. A similar approach followed by [9], in which they introduced a new similarity measure, termed Binary Weighted Cosine (BWC) metric. In this paper we propose a new sequence similarity metric that contains both sequence related as well as set related information. We use the new metric in conjunction with kNN classifier. This work aims to develop an efficient metric, that considers both number of shared sequences of system calls, and the order of occurrences of these system calls(position) that enhances the capabilities of simple kNN classifier significantly especially in the context of intrusion detection. We propose a new similarity metric, which takes into account both the orders of occurrences as well as commonly shared sequences of system calls in a session.

The rest of the paper is organized as follows. We explained our proposed similarity measure S^3M in section 2. We outline kNN classification algorithm with proposed similarity measure in section 3. Experimental results and discussion have been presented in section 4.

2 Sequence And Set Similarity Measure (S^3M)

One common way of measuring distance between sequences (strings) is using Levenshtein distance[7]. Another approach has been to convert sequences of symbols into vector form by considering frequencies of occurrences of symbols and transforming every sequence into a vector of finite length. Distance measure such as Euclidian and cosine similarity are applicable on such vectors [8,9]. One drawback in all these metrics is that the sequence information is lost in the process of transformation. To overcome this difficulty, we propose a new metric called sequence and set similarity measure (S^3M). S^3M considers both set similarity indicating similarity of content and sequence similarity capturing similarity in order of occurrences of symbols.

Set similarity is measured in a standard way using Jaccard's method as follows.

$$\text{SetSim(A,B)} = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Sequence similarity is given as the length of longest common subsequence between two sequences. We define sequence similarity as

$$\text{SeqSim(A,B)} = \frac{|LCS(A,B)|}{\max(|A|,|B|)} \quad (2)$$

We combine set similarity as well as sequence similarity, that is, eq (1) and eq (2) by attaching proper weightage

$$\text{Sim(A,B)} = p\frac{|LCS(A,B)|}{\max(|A|,|B|)} + q\frac{|A \cap B|}{|A \cup B|} \quad (3)$$

where $p + q = 1$ and $p, q \geq 0$. Here, p determines weightage given to sequence similarity and q determines weightage to set similarity.

3 k-NN Classifier with S^3M Scheme

k-Nearest Neighbor (kNN) is a predictive technique suitable for classification models. In the k-Nearest Neighbor classification, the k closest patterns are found and a voting scheme is used to determine the outcome. Unlike other common classifiers, a kNN classifier does not build a classifier in advance [10]. That is what makes it suitable for IDS data streams. When a new sample arrives, kNN finds the k neighbors nearest to the new sample from the training space based on some suitable similarity or distance metric. Thus, kNN classification is a very good choice, since no residual classifier needs to be built ahead of time.

Instead of considering only frequency of system calls, we look at the both order of occurrences as well as content of a system calls within a session. One distinguishing feature of our proposed IDS is the simultaneous use of local positioning information of a system call along with the actual information of system calls in a session. We have applied kNN classification algorithm with the S^3M measure.

Let F be the normal set of system calls that consists of set of unique process under normal execution and M be the total number of process in F, that is, $F=\{a_1,a_2,\ldots..,a_M\}$.For all new incoming process P, we compute LCS(P,a_j), UNION(P,a_j), INTERSECTION(P, a_j), Max_Len(P, a_j), where j=1,.. M. Then, we calculate the sequence similarity score $\lambda(P, a_j)$ and set similarity score $\mu(P, a_j)$. Using our proposed similarity measure as described in eq (3), we calculate sim(P, a_j) for a fixed value of p and q such that p+ q =1. If for any a_j and new process P, sim(P,a_j) =1 then we classify P as a normal session. If for any a_j and new process P Sim(P,a_j) is not equal to 1 then from among similarity values, we choose the k nearest neighbors. Then, we calculate the average similarity of these k neighbors. If the average similarity of k neighbors is greater than user specified threshold then kNN classifier classifies the new process as normal else abnormal. The experimental results using the kNN classification algorithm with S^3M is described in section 4.

4 Experimental Results and Discussion

We applied the *k*-Nearest Neighbor classifier with S^3M similarity measure to the 1998 DARPA data. A large sample of computer attacks embedded in normal background traffic is provided in 1998 DARPA Intrusion Detection System Evaluation program (www.ll.mit.edu/IST/ideval). The network traffic of an Air Force Local Area Network was simulated to collect TCPDUMP and BSM audit data. There were 38 types of network-based attacks and several realistic intrusion scenarios conducted in the midst of normal background data.

We used the Basic Security Module (BSM) audit data collected from a victim Solaris machine inside the simulation network. The BSM audit logs contain information on system calls produced by programs running on the Solaris machine. Names of the system call were only recorded. There are around 2000 normal sessions reported in the four days of data. We extract the processes occurring during these days and our normal data set consists of 606 unique processes. There are 412 normal sessions on the fifth day and we extract 5285 normal processes for the testing data. In order to test the detection capability of our method, we incorporate 55 intrusive sessions into our testing data. From 55 abnormal sessions the session

Table 1. False Positive Rate vs. Detection Rate

Threshold value	p = 0.5 and q = 0.5			
	k=5		k=10	
	False positive Rate	Detection Rate	False positive Rate	Detection Rate
0.89	0.010028	1	0.010974	1
0.88	0.010028	1	0.010974	1
0.87	0.008325	1	0.010974	1
0.86	0.008325	0.94545	0.010785	1
0.84	0.00756	0.94545	0.009839	0.94545
0.8	0.00491	0.94545	0.008136	0.94545
0.78	0.004162	0.92727	0.00719	0.92727
0.75	0.00245	0.92727	0.00359	0.92727
0.7	0.00189	0.8909	0.00227	0.8909
0.65	0.000946	0.76363	0.0017	0.85454
0.6	0.000189	0.36363	0.00089	0.4
0.55	0	0.27272	0	0.2909

test.1.5_processtable is completely similar to one of the process as among the 606 test dataset. Hence, we removed this process from test data and finally we have only 605 unique test processes. We performed the experiments with k = 5, and k = 10, that is, the number of nearest neighbors. Table 1 shows the results for k= 5 and k = 10 for p = 0.5 and q = 0.5. As described in the algorithm, threshold values determine how close the given process is to the training dataset containing all normal sessions. If the average similarity score obtained for the new process is below the threshold value, it will be classified as abnormal. Detection rate(DR) is the ratio of number of intrusive sessions (abnormal) detected correctly to the total number of intrusive sessions. False positive rate(FPR) is defined as the number of normal processes detected as abnormal divided by the total number of normal processes. Receiver operating characteristics (ROC) curve depicts the relationship between FPR and DR. ROC curve gives an idea of the trade off between FPR and DR achieved by an classifier. An ideal ROC curve would be parallel to FPR axis at DR equal to 1.

Pairs of FPR and DR are typically obtained by varying an internal parameter of the classifier. In our case, this parameter is threshold value. ROC curve has been drawn in fig 1 at k =5 and k= 10 at various threshold values. Fig 1 shows kNN classification with S^3M measure seems to work quite satisfactory as indicated by high DR of 1 at low FPR of 0.008 kNN classifier with k=5 seems to work better than with k =10. as a result, further experiments are carried out using k =5.

Fig 2 shows the effect of varying sequence weightage (p) on DR at various threshold values. Graph shows that as p value increases the DR value tends to reach to the ideal value of 1 faster. That means the order of occurrences of system calls plays significant role in determining the nature of a session, whether it is normal or abnormal. For DARPA 1998 dataset we found that the order of occurrences of system call plays role in determining the nature of the session. We compared our proposed approach with [9] where they used a metric called BWC with kNN classification. We used p =0.5 and q =0.5 for the experiments. Using the S^3M measure we have high DR

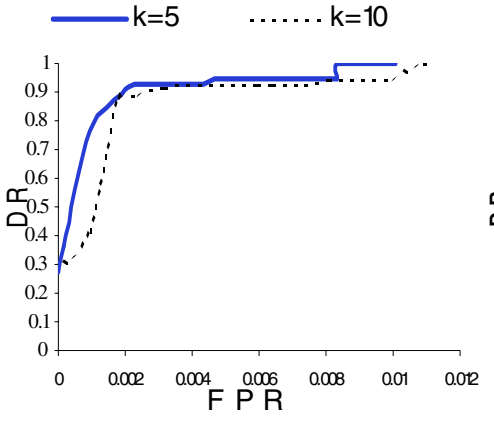


Fig. 1. ROC curve for the S^3M scheme at k=5,10

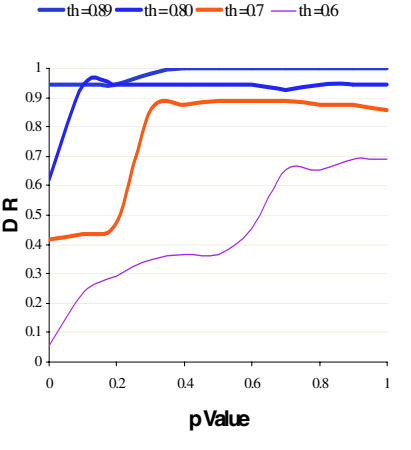


Fig. 2. Graph between p and detection rate at various threshold values

achieved at lower FPR as compared to BWC scheme. The S^3M scheme achieves a DR of 100% at a FPR of 0.8% BWC scheme achieves 100% DR at 8% FPR.

Experiments in [9] were carried out by considering 604 sessions in test dataset on removing two sessions namely. *4.5_it_162228loadmodule* and *5.5_it_fdformat_ chmod.* Hence, we also removed these two sessions and carried out our experiments. Fig 3 shows the comparative ROC curves for KNN classification using S^3M measure and BWC measure. ROC curve due to S^3M measure has high DR than BWC measure at low FPR.

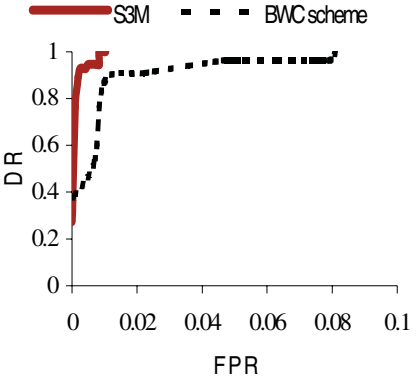


Fig. 3. ROC curve for the S^3M scheme and BWC scheme for k = 5

With this reduced test data set and using our proposed S^3M scheme we were able to achieve higher DR at low FPR than the BWC scheme. From Table 2 we observed that as DR reaches 100% the corresponding FPR as low as 0.8%.

Table 2. FPR vs DR after removing two attacks

Threshold value	p =0.5, q = 0.5			
	k=5		k=10	
	False positive Rate	Detection Rate	False positive Rate	Detection Rate
0.89	0.010028	1	0.010974	1
0.88	0.010028	1	0.010974	1
0.87	0.008325	1	0.010974	1
0.86	0.008325	0.9811	0.010785	1
0.84	0.00756	0.9811	0.009839	0.9811
0.8	0.00491	0.9811	0.008136	0.9811
0.78	0.004162	0.9622	0.00719	0.9622
0.75	0.00245	0.9622	0.00359	0.9622
0.7	0.00189	0.9245	0.00227	0.9245
0.65	0.000946	0.7924	0.0017	0.8867
0.6	0.000189	0.3773	0.00089	0.415
0.55	0	0.283	0	0.3018

From the results presented in this section, we can conclude that the proposed similarity measure is well suited for developing the intrusion detection system where the order of occurrences of system calls plays a role in determining the nature of the session. This similarity measure would also have potential application in web mining as well as for comparing amino acid and DNA sequences in computational biology.

References

1. Bace R.: Intrusion Detection. Macmillan Technical Publishing (2000).
2. Base, R., Mell, P.: NIST special publication on intrusion detection system. NIST, (2001) 800-831.
3. Forrest, S, Hofmeyr, S. A., Somayaji, A. and Longstaff ,T..A.: *A Sense of self for UNIX process.* In Proceedings of the IEEE Symposium on Security and Privacy(1996) 120-128.
4. Hofmeyr, S. A., Forrest, S., and Somayaji, A.: Intrusion Detection Using Sequences of System calls. Journal of Computer Security vol. 6 (1996) 151-180.
5. Kumar S,.Sppafford, E.H. A pattern matching model for misuse intrusion detection. In 17th National Computer Security Conference (1994).11-21,.
6. Lee, W. and Stolfo, S. J.: Data mining approaches for intrusion detection. In Proceedings of the 7th USENIX Security Symposium(1998).
7. Levenshtein, L.I.: Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics–Doklady*, vol. 10, no. 7(1966) 707–710.
8. Liao, Y., Vemuri, V.R.: Using Text Categorization Techniques for Intrusion Detection. In: Proceedings USENIX Security (2002) 51-59.
9. Rawat, S, Pujari A.K., Gulati, V. P. and Vemuri Rao V.: Intrusion Detection using Text Processing Techniques with a Binary-Weighted Cosine Metric, International Journal of Information Security, Springer-Verlag (2004).
10. Tom. M. Mitchell, Machine learning, Mc Graw Hill (1997).

The Multi-fractal Nature of Worm and Normal Traffic at Individual Source Level*

Yufeng Chen[1], Yabo Dong[1], Dongming Lu[1], and Yunhe Pan[1]

College of Computer Science and Technology,
Zhejiang University,
Hangzhou 310027, P. R. China
{xztcyfnew, dongyb, ldm, panyh}@zju.edu.cn

Abstract. Worms have been becoming a serious threat in web age because worms can cause huge loss due to the fast-spread property. To detect worms effectively, it is important to investigate the characteristics of worm traffic at individual source level. We model worm traffic with the multi-fractal process, and compare the multi-fractal property of worm and normal traffics at individual source level. The results show that the worm traffic possesses less multi-fractal property.

1 Introduction

Data security is very important in web age because we should assure the availability of Internet and web-based information systems. Worms have been becoming a serious threat because worms can spread in short time and cause huge loss. Last year, two notorious worms, the "Blaster" worm [1] and "Welchia" worm [2] infected a lot of computers and the losses are heavy. Early warning is an effective method to prevent the spread of worms. The infectious computers should be located for eradicating worms from systems. Thus, we investigated the traffic characteristics of worm and normal traffic at individual source level to compare the diversity of the traffic characteristics.

The self-similar [3, 4, 5, 6, 7] and multi-fractal [8, 9] models have been proposed to depict the characteristics of network traffic from the angle of traffic engineering [3]. However, because the fractal characteristic is the nature of network traffic, we tried to investigate the fractal characteristics of worm traffic find the abnormity of worm traffic. Because worm detection is a competition against worm propagation, the short-range characters is more important. Thus we payed our attentions to the multi-fractal nature of traffics because multi-fractal model possesses the capability of describing the short-range property of time series as

* This work is supported by a grant from Zhejiang Provincial Natural Science Foundation (No.Y104437), Hubei Provincial Natural Science Foundation (No. 2004ABA018), Science and Technology Program of Hubei Provincial Department of Education (No. 2004D005), and Science and Technology Program of Hubei Provincial Department of Education (No. D200523007).

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 505–510, 2005.

well as long-range property. We studied the worm and normal traffic at individual source level, i.e., traffics are generated by individual computers. We found that the worm traffic owns less multi-fractal characteristic than normal traffic does. The data set was collected on a 100Mbps link connecting one dormitory building to our campus network. And as an example of worm traffic, the traffic generated by "Welchia" worm was collected and analyzed.

The paper is organized as follows. The related works are introduced briefly in Sect. 2 and the mathematical background is introduced in Sect. 3. In Sect. 4, we give an overview of our data set. The diversity of multi-fractal properties of normal and worm traffics at individual source level are compared in Sect. 5. In Sect. 6, the conclusions and future work are presented.

2 Related Works

Some models and methods of worm traffic have been proposed, which mostly focus on the aggregated traffic characters of worm and normal traffics. Cowie et al. described the idea of "worm induced traffic diversity" that is the primary cause of the BGP instabilities[10]. However, they just propose the preliminary conclusions without deeper investigation. [11] presented an statistical-based approach that utilizes application specific knowledge of the network services, which is used to find Remote-to-Local attacks. [12] presented the ideas of exploiting the property of self-similarity in network traffic to detect the faulty behavior. [13] took advantage of multi-fractal model to detect fault in network traffic based on the fact that faults in a self-similar traffic destroy the structure at the time points they occur. However, the meaning of errors and faults in [12] and [13] is rather broad, maybe including the abnormal traffic generated by worms.

The proposed statistical models of worm behaviors and fractal models of abnormal traffic are investigated at the aggregated traffic level. However, they can not provide information to identify the infectious computers. Thus, we try to provide the idea for detecting the infectious computers at individual source level from the angle of multi-fractal nature of network traffic.

3 Background

Firstly, we give the basis of multi-fractal analysis briefly. For detailed discussion, please refer to [8, 9, 14]. Consider a probability measure μ on the unit interval $[0, 1]$ and random variables

$$Y_n = \log \mu(I_K^{(n)}) \ , \tag{1}$$

where $I_K^{(n)}$ denotes the partition into 2^n equal subintervals

$$I_K^{(n)} := [k2^{-n}, (k+1)2^{-n}] \ , \tag{2}$$

and K is a random number from $\{0, \ldots, 2^n - 1\}$ with uniform distribution P_n. If the rate function, also called "partition function" or "free energy"

$$\tau(q) := \lim_{n\to\infty} \frac{-1}{n} \log_2 \sum_{k=1}^{2^n} \mu(I_k^{(n)})^q \tag{3}$$

exists and is differentiable on $\Re$, then the double limit

$$f_G(\alpha) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log_2 2^n P_n[\alpha(I_K^{(n)}) \in (\alpha - \varepsilon, \alpha + \varepsilon)] \tag{4}$$

with

$$\alpha(I_K^{(n)}) := \frac{-1}{n \log 2} Y_n = \frac{\log \mu(I_K^{(n)})}{\log |I_K^{(n)}|} \tag{5}$$

exists, where $\alpha(I_K^{(n)})$ is termed Holder exponent. And

$$f_G(\alpha) = f_L(\alpha) := \tau^*(\alpha) := \inf_{q\in\Re}(q\alpha - \tau(q)) \ . \tag{6}$$

And f_L is often referred to as Legendre spectrum. Because of the robustness and simplicity of f_L, we will analyze the multi-fractal property of our data series through f_L. And the typical shape of f_L is a $\cap$. The parameter α quantifies the degree of regularity in a point x: here, the measure of an interval $[x, x + \Delta x]$ behaves as $(\Delta x)^\alpha$. In traffic measurements, $(\Delta x)^\alpha$ can be interpreted as the number of packets or bytes in this interval. Consequently, $\alpha < 1$ indicates a burst of events around x "on all levels", while $\alpha > 1$ is found in regions where events occur sparsely. Thus, for a process, if the interval of $\alpha < 1$ is larger, the process possesses more multi-fractal property.

The scaling of "sample moments" can also be studied through the partition function

$$\tau(q) := \liminf_{n\to\infty} \frac{\log S^{(n)}(q)}{-n \log 2} \ , \tag{7}$$

where the partition sum

$$S^{(n)}(q) := \sum_{k=0}^{2^n-1} |Y((k+1)2^{-n}) - Y(k2^{-n})|^q \ . \tag{8}$$

When we inspect the log-log plots of partition sum against q, if the plot appears to be linear, the observed process is multi-fractal. To get the partition sum and Legendre spectrum of the worm and normal traffic, we make use of a tool named Fraclab [15].

4 Data Set

Our data sets were collected from a 100Mbps link connecting one dormitory building to our campus network, in December 26, 2003, when the "Welchia" worm broke out. The data sets contains information of source and destination

Table 1. Summary of our data set

Item	Value
Packet Counts	25,436,630
Number of Distinct Source Addresses	2,889
Number of Distinct Destination Addresses	300,053
Number of Total Distinct IP Addresses	300,133

IP address, timestamp, packet length. The packet count of the data set is about 26 million. Table 1 shows the summaries of the data sets.

We picked up two packet streams generated by two sources, one for normal traffic and the other for worm traffic. The summaries of the two packet streams are shown in Table 2. The source addresses are renumbered for privacy reasons, which are 2 and 72. The source 2 generates normal traffic, and the source 72 generates worm traffic. The source 72 marked with "worm" exhibits the character of "Welchia" worm because the volume of the corresponding destination addresses is rather vast.

5 Diversity of Multi-fractal Properties at Individual Source Level

We analyze the number of packets of normal and worm traffic generated by individual source with the time unit of 100 milliseconds. And because we want to investigate the characteristics of short term, we choose 4096 observations for each time series, and each observation represents the number of packets sent over the Ethernet by the corresponding source every 100 milliseconds. Fig. 1 shows the plots of the partition sum and Legendre spectrum of the two streams.

Fig. 1 depicts the multi-fractal characteristics of the two packet streams. The top plots show the partition sum against the scale on log-log scale. It's obvious that $S^{(n)}(q)$ of the worm traffic is much different from that of the normal traffic. And the bottom plots show Legendre spectrum $f_L(\alpha)$ against α. And we can learn from the plots that the worm traffic possesses less multi-fractal property because the interval of $\alpha > 1$ is much larger than that of normal traffic, which

Table 2. Summary of the two packet streams. The corresponding destination address means the destination address of the packet whose source address is the corresponding renumbered source address, and the corresponding packet has the similar meaning

Traffic	Normal	Worm
Renumbered Source Address	2	72
Number of Corresponding Destination Addresses	220	129,437
Number of Corresponding Packets	141,092	129,528

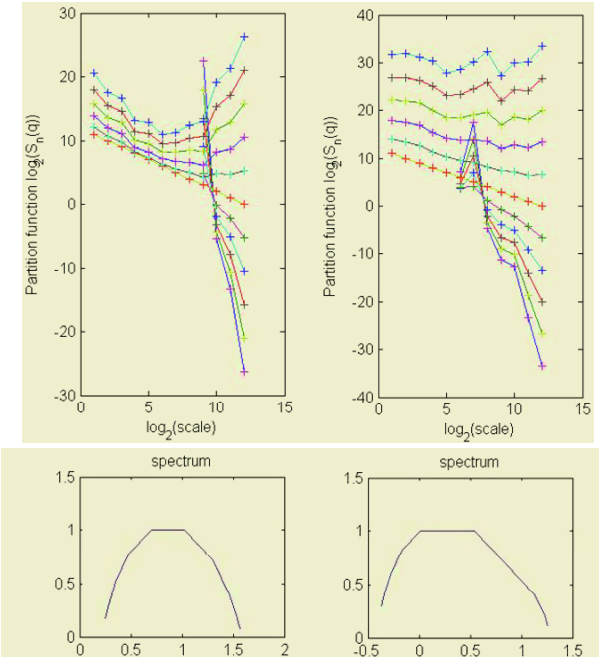


Fig. 1. the plots of partition sum $S^{(n)}(q)$ and Legendre spectrum $f_L(\alpha)$ based on Discrete Wavelet Transform. The number of q ranging from -5 to 5, and the total number of q is 11. The left part is corresponding to the infectious source 72, and the right part is corresponding to the normal source 2

means that the proportion of burstiness of events is less for worm traffic. The reason of less multi-fractal for worm traffic can be interpreted intuitively. When a computer is infected by "Welchia" worm, this source will send ICMP echo request packets continuously, and the traffic generated by this source exhibits less burstiness on all levels.

6 Conclusions and Future Works

In this paper, we presented the idea to investigate the diversity of multi-fractal nature of worm and normal traffics for further application, such as worm detection. And the results show that worm traffic possesses less multi-fractal property than that of normal traffic. The further work includes: 1) investigate the multi-fractal characteristics of traffic of other worms; 2) analyze the factors that influence the multi-fractal property and how they influence; 3) implement the algorithm of detecting infectious sources based on the multi-fractal characteristics of traffic.

Acknowledgements. The authors acknowledge the continuing support from the Ningbo Network Access Point and Networking Center of Zhejiang University.

References

1. CERT, 2003, "CERT advisory CA-2003-20 w32-blaster worm" http://www.cert.org/advisories/CA-2003-20.html
2. CCERT, 2003, "CCERT advisory of security" http://www.ccert.edu.cn/notice/advisory.htm
3. Leland, W.E., Willinger, W., Taqqu, M.S. and Wilson, D.V., 1995, "On the self-similar nature of ethernet traffic", *ACM SIGCOMM Computer Communication Review*, Vol. 25, No. 1, pp. 202-213
4. Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D.V., 1997, "Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 1, pp. 71-86
5. Crovella, M.E. and Bestavros, A., 1997, "Self-similarity in world wide web traffic: Evidence and possible causes", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, pp. 835-846
6. Garrett, M.W. and Willinger, W., 1994, "Analysis, modeling and generation of self-similar VBR video traffic", *ACM SIGCOMM Computer Communication Review*, Vol. 24, No. 4, pp. 269-280
7. Paxson, V. and Floyd, S., 1995, "Wide area traffic: The failure of poisson modeling", *IEEE/ACM Trans. Networking*, Vol. 3, No. 3, pp. 226-244
8. Mannersalo, P. and Norros, I., 1997, "Multifractal analysis of real ATM traffic: a first look", http://www.vtt.fi/tte/tte21/cost257/multifaster.ps.gz
9. Riedi, R.H. and Vehel, J.L., 1997, "Multifractal properties of TCP traffic: a numerical study", http://www.stat.rice.edu/∼riedi/Publ/ts.ps.gz
10. Cowie, J., Ogielski, A.T., Premore, B. and Yuan, Y., 2001, "Global routing instabilities triggered by codered II and nimda worm attacks", http://www.renesys.com/projects/bgp_instability
11. Krugel, C., Toth, T. and Kirda, E., 2002, "Service specific anomaly detection for network intrusion detection", In *Proc. of ACM Symposium on Applied Computing*, pp. 201-208, March 2002
12. Schleifer, W. and Mannle, M., 2001, "Online error detection through observation of traffic self-similarity", In *Proc. of IEE on Communications*, Vol. 148, pp. 38-42, February 2001
13. Tang, Y., Luo, X. and Yang, Z., 2002, "Fault detection through multi-fractal nature of traffic", In *Proc. of IEEE on Communications, Circuits and Systems and West Sino Expositions*, Vol. 1, pp. 695-699, June 2002
14. Riedi, R.H., 2002, "Multifractal processes", http://www.stat.rice.edu/∼riedi/Publ/mp.ps.gz
15. FRACTALES group, 2001, "Fraclab: A fractal analysis toolbox for signal and image processing", http://fractales.inria.fr/index.php?page=download_fraclab

Learning Classifiers for Misuse Detection Using a Bag of System Calls Representation*

Dae-Ki Kang[1], Doug Fuller[2], and Vasant Honavar[1]

[1] Artificial Intelligence Lab, Department of Computer Science, Iowa State University
{dkkang, honavar}@iastate.edu
[2] Scalable Computing Lab., Iowa State University and U.S. Department of Energy
dfuller@scl.ameslab.gov

Abstract. In this paper, we propose a "bag of system calls" representation for intrusion detection of system call sequences and describe misuse detection results with widely used machine learning techniques on University of New Mexico (UNM) and MIT Lincoln Lab (MIT LL) system call sequences with the proposed representation. With the feature representation as input, we compare the performance of several machine learning techniques and show experimental results. The results show that the machine learning techniques on simple "bag of system calls" representation of system call sequences is effective and often perform better than those approaches that use foreign contiguous subsequences for detecting intrusive behaviors of compromised processes.

1 Introduction

In most intrusion detection systems (IDS) that model the behavior of processes, intrusions are detected by observing fixed-length, contiguous subsequences of system calls. For example, in anomaly detection, subsequences of input traces are matched against normal sequences in database so that foreign sequences [1, 2] are detected. One potential drawback of this approach is that the size of the database that contains fixed-length contiguous subsequences increases exponentially with the length of the subsequences. In this paper, we explore an alternative representation of system call traces for intrusion detection. We demonstrate that simple *bag of system calls* representation of system call sequences is surprisingly effective in constructing classifiers for intrusion detection of system call traces.

2 Alternative Representations of System Call Sequences

Let $\Sigma = \{s_1, s_2, s_3, \ldots, s_m\}$ be a set of system calls where $m = |\Sigma|$ is the number of system calls. Data set D can be defined as a set of labeled sequences $\{< Z_i, c_i > | Z_i \in \Sigma^*, c_i \in \{0, 1\}\}$ where Z_i is an input sequence and c_i is a

* Supported by NSF grant IIS 0219699.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 511–516, 2005.

corresponding class label with 0 denoting a "normal" activity and 1 denoting a "intrusive" activity. Given the data set D, the goal of the learning algorithm is to find a classifier $h : \Sigma^* \rightarrow \{0, 1\}$ that maximizes given criteria. Such criteria are accuracy, detection rate and false positive rate. Each sequence $Z \in \Sigma^*$ is mapped into a finite dimensional feature vector by a feature representation $\Phi : \Sigma^* \rightarrow \mathbf{X}$. Thus, the classifier is defined as $h : \mathbf{X} \rightarrow \{0, 1\}$ for data set $\{< X_j, c_j > | X \in \mathbf{X}, c_j \in \{0, 1\}\}$. This allows us to use a broad range of machine learning algorithms to train classification for intrusion detection.

2.1 Contiguous Foreign Subsequences

In this approach, a feature is defined as $X_j = x_1 x_2 x_3 \ldots x_l$, a substring of Z_i, where $x_k \in \Sigma$ and l is a constant. The number of possible features is $\left|\Sigma^l\right| \geq j$ and each feature X_j is assigned a class label c_i according to the original sequence Z_i. *STIDE* [3] uses sliding windows with length l over an original input trace to generate fixed-length substrings as features and constructs a database of the features in the training stage, and decides a test sequence is anomalous if the number of mismatches in the user-specified locality frame (locality frame count), which is composed of adjacent features in the frame, is more than the user-specified threshold.

2.2 Bag of System Calls

"Bag of system calls" representation is inspired by "bag of words" representation that has been demonstrated to be effective in text classification problems. In our approach, a sequence is represented by an ordered list $X_i = \langle c_1, c_2, c_3, \ldots, c_m \rangle$ where $m = |\Sigma|$ and c_j is the number of occurrence of system call s_j in the input sequence Z_i. Note that this representation of system call traces does not preserve information about relative order of system calls in the sequence.

3 Data Sets

3.1 UNM System System Call Sequences

The University of New Mexico (UNM) provides a number of system call data sets. The data sets we tested are "live lpr", "live lpr MIT", "synthetic sendmail", "synthetic sendmail CERT", and "denial of service"(DoS).

In UNM system call traces, each trace is an output of one program. Most traces involve only one process and usually one sequence is created for each trace. Sometimes, one trace has multiple processes. In such cases, we have extracted one sequence per process in the original trace. Thus, each system call trace can yield multiple sequences of system calls if the trace has multiple processes. Table 1 shows the number of original traces and the number of sequences for each program.

3.2 MIT Lincoln Lab Data Sets

We used data sets provided by the MIT Lincoln Lab [4]. The fourth week (starting at 6/22/98) training data set of year 1998 is used for the experiments in this

Table 1. The number of original traces and generated sequences in UNM data sets

Program	# of original traces	# of sequences
live lpr (normal)	1232	1232
live lpr (exploit)	1001	1001
live lpr MIT (normal)	2704	2704
live lpr MIT (exploit)	1001	1001
synthetic sendmail (normal)	7	346
synthetic sendmail (exploit)	10	25
synthetic sendmail CERT (normal)	2	294
synthetic sendmail CERT (exploit)	6	34
denial of service (normal)	13726	13726
denial of service (exploit)	1	105

paper. MIT Lincoln Labs datasets include omnibus files containing all system call traces. For each omnibus file, there is a separate, network traffic analysis data file that indicates inbound network connections to the system. Attack attempts are logged with the network data, so labeling of the training data requires cross-indexing this file with the system call trace file. The system call trace file identifies the source of each call using the process ID. Therefore, cross-indexing requires tracking the argument to the 'exec' system call identifying the binary to be executed. Additionally, the timestamps from the network traffic analyzer do not exactly correspond to the execution timestamps from the operating system kernel. A tolerance of one second was chosen and seems to permit the matching of a large majority of connection attempts with their corresponding server processes run on the target system. All processes detected that do not correspond to some network connection attempt identified in the trace are removed from consideration (since they cannot be classified), as are all calls attributed to a process ID for which an 'exec' system call is not found. The resulting data are available at `http://www.cs.iastate.edu/~dkkang/IDS_Bag/`.

4 Experiments and Results

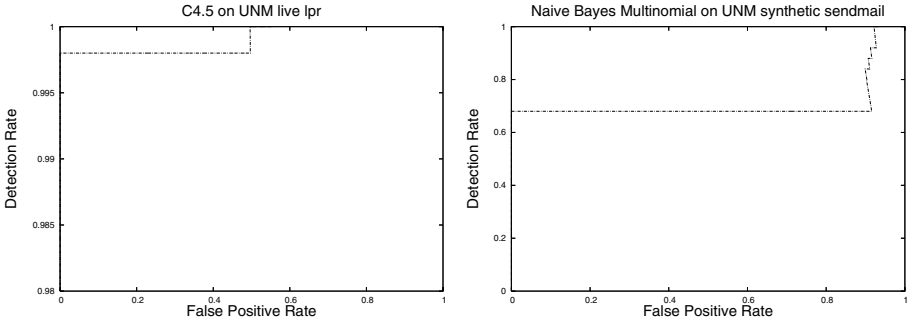
For the evaluation of classifiers generated in the experiment, ten-fold cross validation is used for rigorous statistical evaluation of the trained classifiers. Thus, in each experiment, the data set is divided into ten disjoint subsets, nine of which are used for training the classifier and the tenth part used for evaluating the classifier. The reported the results represent averages over ten such runs. Table 2 shows the accuracy, detection rate, and false positive rate [5] of the data sets we tested. The detection rate is a fraction of the intrusions identified and the false positive rate is a fraction of normal data mis-identified as intrusion.

Figure 1 shows the Receiver Operating Characteristic (ROC) Curve of "UNM live lpr" and "UNM synthetic sendmail" data sets using C4.5 and Naive Bayes Multinomial algorithms respectively.

Table 2. Percentage of misuse detection based on 10 fold cross-validation

Program	Naive Bayes Multinomial	C4.5	RIPPER	SVM	Logistic Regression
UNM live lpr					
accuracy	83.43	99.91	99.91	100.00	99.91
detection rate	100.00	99.80	99.80	100.00	100.00
false positive rate	30.03	0.00	0.00	0.00	0.16
UNM live lpr MIT					
accuracy	54.52	99.89	99.86	99.83	99.97
detection rate	100.00	99.90	99.80	99.80	99.90
false positive rate	62.31	0.11	0.11	0.14	0.00
UNM synthetic sendmail					
accuracy	20.21	94.87	94.33	95.68	95.41
detection rate	92.00	40.00	48.00	40.00	64.00
false positive rate	84.97	1.15	2.31	0.28	2.31
UNM synthetic sendmail CERT					
accuracy	24.39	96.64	95.42	96.03	96.03
detection rate	100.00	85.29	82.35	64.70	82.35
false positive rate	84.35	2.04	3.06	0.34	2.38
UNM denial of service					
accuracy	98.70	99.97	99.96	99.98	99.97
detection rate	44.76	99.04	98.09	100.00	99.04
false positive rate	0.88	0.02	0.02	0.01	0.01
MIT LL 1998 4^{th} Week					
Monday					
accuracy	100.00	100.00	100.00	100.00	100.00
detection rate	100.00	100.00	100.00	100.00	100.00
false positive rate	0.00	0.00	0.00	0.00	0.00
Tuesday					
accuracy	99.55	99.55	99.55	99.55	99.55
detection rate	98.60	98.60	98.60	98.60	98.60
false positive rate	0.00	0.00	0.00	0.00	0.00
Thursday					
accuracy	99.73	99.73	99.73	99.73	99.73
detection rate	100.00	100.00	100.00	100.00	100.00
false positive rate	0.04	0.04	0.04	0.04	0.04
Friday					
accuracy	98.80	98.80	98.80	98.80	98.80
detection rate	89.28	89.28	89.28	89.28	89.28
false positive rate	0.00	0.00	0.00	0.00	0.00

The results in table 2 show that standard machine learning techniques are surprisingly effective in misuse detection when they are used to train misuse detectors using simple bag of system calls representation. For example, with SMO (a widely used algorithm for training SVM) using a linear kernel, an SVM can perfectly detect both normal and intrusion sequences in the "UNM live lpr" data set.



(a) C4.5 decision tree induction on UNM live lpr

(b) Naive Bayes Multinomial on UNM synthetic sendmail

Fig. 1. ROC Curve of "UNM live lpr" and "UNM synthetic sendmail" data sets in misuse detection

5 Summary and Discussion

Results of our experiments using widely used benchmark data sets - the University of New Mexico (UNM) and MIT Lincoln Lab (MIT LL) system call sequences show that the performance of the proposed approach in terms of detection rate and false positive rate is comparable or superior to that of previously reported data mining approaches to misuse detection. In particular, as shown in table 2, the proposed methods achieve nearly 100% detection rate with almost 0% false positive rate on all the data sets studied with the exception of two synthetic data sets ('UNM synthetic sendmail' and 'UNM synthetic sendmail CERT').

When compared with the widely used fixed-length contiguous subsequence models, the *bag of system calls* representation explored in this paper may seem somewhat simple. It may be argued that much more sophisticated models that take into account the identity of the user or perhaps the order in which the calls were made. But our experiments show that a much simpler approach may be adequate in many scenarios. The results of experiments described in this paper show that it is possible to achieve nearly perfect detection rates and false positive rates using a data representation that discards the relationship between system call and originating process as well as the sequence structure of the calls within the traces.

Forrest et. al. [1, 3] showed that it is possible to achieve accurate anomaly detection using fixed-length contiguous subsequence representation of input data. In their approach, the detector will find anomalous subsequences right after they are executed depending on user-specified thresholds. The proposed '*bag of system calls* representation has advantage of fast learning, low memory requirement for training classifiers. A simple counter program can be used to discriminate normal sequences and abnormal sequences very quickly, before the process is terminated.

We limit our discussion for misuse detection in this paper. Additional experimental results and detailed discussions including an application to anomaly detection can be found in [5].

References

1. Forrest, S., Hofmeyr, S.A., Somayaji, A., Longstaff, T.A.: A sense of self for unix processes. In: Proceedings of the 1996 IEEE Symposium on Security and Privacy, IEEE Computer Society (1996) 120
2. Hofmeyr, S.A., Forrest, S., Somayaji, A.: Intrusion detection using sequences of system calls. Journal of Computer Security **6** (1998) 151–180
3. Warrender, C., Forrest, S., Pearlmutter, B.A.: Detecting intrusions using system calls: Alternative data models. In: IEEE Symposium on Security and Privacy. (1999) 133–145
4. Lippmann, R., Cunningham, R.K., Fried, D.J., Graf, I., Kendall, K.R., Webster, S.E., Zissman, M.A.: Results of the darpa 1998 offline intrusion detection evaluation. In: Recent Advances in Intrusion Detection. (1999)
5. Kang, D.K., Fuller, D., Honavar, V.: Learning classifiers for misuse and anomaly detection using a bag of system calls representation. Technical Report 05-06, Iowa State University (2005)

A Jackson Network-Based Model for Quantitative Analysis of Network Security*

Zhengtao Xiang[1], Yufeng Chen[2], Wei Jian[3], and Fei Yan[1]

[1] Computer Center, Hubei Automotive Industrial Institute,
Shiyan 442002, Hubei, P. R. China
{xztcyf, yanfei131}@163.com
[2] College of Computer Science and Technology,
Zhejiang University,
Hangzhou 310027, P. R. China
xztcyfnew@zju.edu.cn
[3] Deparment of R & D, Hubei Automotive Industrial Institute,
Shiyan 442002, Hubei, P. R. China
qyky@dfminfo.com.cn

Abstract. It is important for trusted intranets to focus on network security as a whole with dynamic and formalized analysis. The qualitative and current quantitative methods have difficulties to reach the requirements. After analyzing the attacking process, a Jackson network-based model with absorbing states is proposed, where the absorbing states mean the attacks succeed or fail. We compute the steady-state joint probability distribution of network nodes, the mean time of attack data spent in network, and the probabilities from the network entry node to absorbing states. According to the analysis of the above measures, we analyze the relationship between network security and performance.

1 Introduction

Network security is an increasing priority with the boost of e-Government, e-Business, etc. The precondition of enhancing network security is the effective theory and methods of evaluating network security threats or risks. To evaluate network security effectively, the followed problems should be considered. First, the dynamic behaviors in the context of actual network should be reflected because the real environment is operational. The behaviors consist of attack and defense behaviors. Second, the network security should be considered as a whole. Third, formalized analyzing models and methods should be proposed to ensure the consistency of evaluating network security.

* This work is supported by a grant from Hubei Provincial Natural Science Foundation (No. 2004ABA018), and Science and Technology Program of Hubei Provincial Department of Education (No. 2004D005), and Science and Technology Program of Hubei Provincial Department of Education (No. D200523007).

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 517–522, 2005.

The current models and methods either are unprecise and have difficulties of quantification, or focus on analysis of individual system security without considerations of network security as a whole.

The widely used qualitative methods based on the security evaluation criteria [1] may introduce significant differences in evaluation results due to the subjectivity of experts. The results of [2] show that a typical attacker behavior consists of three phases: the learning phase, the standard attack phase, and the innovative attack phase. During standard attack phase, the probability of successful attacks is much higher than probabilities during the other two phases. To evaluate various security solutions, a multi-attribute risk evaluation method is proposed in [3]. A semi-Markov process model is presented in [4], which is based on the hypothesis that a security intrusion and the system response to the attack can all be modeled as random processes. A security evaluation framework is proposed in [1], which interprets how to obtain the quantified security by summing the weighting factors.

In this paper, we propose a Jackson network-based model with absorbing states for quantitative analyses of network security. According to the functions of security components in network, this model considers them as various security boundaries, and depicts the relationships between them by utilizing queueing theory. By modeling the states of attack data and transitions between the states, this model gives a dynamic, formalized analysis of network security.

2 Jackson Network-Based Security Analysis Model

Security defenses includes firewalls, IDSs, application security measures, information encryption, and so on. We model the defenses as security boundaries, which consists of network boundary, system boundary, application boundary and information boundary. Network boundary consists of network firewall and network IDS; system security boundary consists of security defense of OS, personal firewalls, host IDS, and anti-virus software; application security boundary consists of authentication, authorization, and other security measures of applications; information security boundary means attackers can not crack the encrypted data. Security boundaries intercepting attacks means attacks are prevented from succeeding, such as detecting and discarding packets. Based on the above analysis, we develop a Jackson network-based model, as shown in Fig. 1. Before analyzing the mode, we give some assumptions as follows: the arrival of the attack data follows a Poisson process with an average arrival rate of λ; the service time distribution at each node, which represents the security system, is postulated to be exponential; the service policy of each node is FCFS(Fist Come, First Served).

For representing the success or failure of attacks, two absorbing states, the failure state, denoted as state 5, and the success state, state 6, are introduced, which also mean attack data leave the network. The number of states appearing in Fig. 1 depends on the attack goals. If the goal is to breach the system boundary, such as scanning attacks for identify the existence of some hosts, the states on the right side of arc 2 are unreachable. Arc 1 and 3 have Similar meanings.

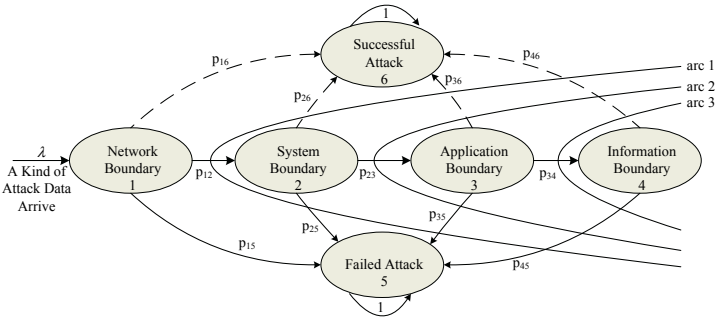


Fig. 1. A state transition diagram for attack data. This shows states of attacks and the transitions between them

Accordingly, only one dashed arrow line, which is corresponding to the attack goal, can exist in Fig. 1, that is, if the attack target is aimed at the network boundary, only one dashed arrow line from state 1 to 6 exists. Attacks are considered fail if they enter state 5 from states 1, 2, 3, or 4. Once attack data enter states 5 or 6, the probability of leaving these states is 0.

3 Model Analysis

Now, we discuss and derive some interested measures, the steady-state joint probability distribution of network nodes, denoted by η, the mean time of attack data spent in network, denoted by T_q, and the probabilities from the network entry node to absorbing states, π_5 and π_6 respectively. As an example, we concern our analysis with the attack exploiting RPC vulnerability, whose target is to breach the system boundary. After eliminating the states on the right side of system boundary in Fig. 1 and splitting the network boundary into network firewall "1F" and network IDS "1I", and the system boundary into personal firewall "2F" and host IDS "2I", the resulting model is depicted in Fig. 2.

First, we classify the states in Fig. 2 into transient set $TS = \{1F, 1I, 2F, 2I\}$, and absorbing set $AS = \{5, 6\}$. The parameters in the figure are listed below:

- λ: the average arrival rate of attack data;
- $\mu_{1F}, \mu_{1I}, \mu_{2F}, \mu_{2I}$: the exponential mean service rate of network firewall, network IDS, personal firewall, host IDS, respectively;
- p_{1FI}, p_{2FI}: the transition probabilities of the attack data entering network IDS and host IDS from network firewall and personal firewall, respectively;
- $p_{1F5}, p_{1I5}, p_{2F5}, p_{2I5}$: the transition probabilities of attack data from transient states to failure state 5;
- p_{26}: the transition probability of attack data entering the success state 6.

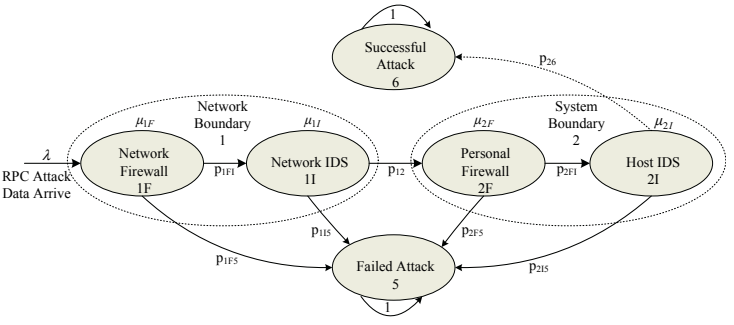


Fig. 2. A state transition diagram for attacks exploiting RPC vulnerability. The network and system boundaries are all split into two components: firewall and IDS

Based on the model, the transition probability matrix can be written as:

$$P = \begin{vmatrix} 0 & p_{1FI} & 0 & 0 & p_{1F5} & 0 \\ 0 & 0 & p_{12} & 0 & p_{1I5} & 0 \\ 0 & 0 & 0 & p_{2FI} & p_{2F5} & 0 \\ 0 & 0 & 0 & 0 & p_{2I5} & p_{26} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} Q & C \\ 0 & I \end{vmatrix}, \quad \text{where} \quad \begin{matrix} p_{1F5} = 1 - p_{1FI} \\ p_{1I5} = 1 - p_{12} \\ p_{2F5} = 1 - p_{2FI} \\ p_{2I5} = 1 - p_{26} \end{matrix} . \quad (1)$$

Matrix Q gives the transition probabilities between transient states, and matrix C gives the transition probabilities from transient states to absorbing states.

3.1 Steady-State Joint Probability Distribution of Network Nodes η

According to Jackson's theory, the steady-state joint probability distribution has a simple expression,

$$\eta(n_{1F}, n_{1I}, n_{2F}, n_{2I}) = \prod_{j \in TS} \eta(n_j) = \prod_{j \in TS} (1-\rho_j)\rho_j^{n_j}, \rho_j = \lambda_j/\mu_j, j \in TS \ , \quad (2)$$

where state $n = (n_{1F}, n_{1I}, n_{2F}, n_{2I})$, n_j is the queue length of node j, ρ_j denotes the utilization of j, μ_j is the mean service rate of node j, and λ_j is the average arrival rate of attack data to node j. We have $\lambda_j = \lambda V_j$, where V_j is the average number of times the state j is visited before the data enter one of the absorbing states. To compute V_j, we need the matrix $M = (1 - Q)^{-1}$, whose first row of elements is correspond to each V_j. Now, we give the results as followed:

$$M = \begin{vmatrix} 1 & p_{1FI} & p_{1FI}p_{12} & p_{1FI}p_{12}p_{2FI} \\ 0 & 1 & p_{12} & p_{12}p_{2FI} \\ 0 & 0 & 1 & p_{2FI} \\ 0 & 0 & 0 & 1 \end{vmatrix}, V_j = \begin{cases} 1 & j = 1F \\ p_{1FI} & j = 1I \\ p_{1FI}p_{12} & j = 2F \\ p_{1FI}p_{12}p_{2FI} & j = 2I \end{cases} . \quad (3)$$

Finally, we get

$$\eta(n_j) = (1-\rho_j)\rho_j^{n_j} = \begin{cases} (1-\frac{\lambda}{\mu_{1F}})(\frac{\lambda}{\mu_{1F}})^{n_{1F}} & j = 1F \\ (1-\frac{\lambda p_{1FI}}{\mu_{1I}})(\frac{\lambda p_{1FI}}{\mu_{1I}})^{n_{1I}} & j = 1I \\ (1-\frac{\lambda p_{1FI}p_{12}}{\mu_{2F}})(\frac{\lambda p_{1FI}p_{12}}{\mu_{2F}})^{n_{2F}} & j = 2F \\ (1-\frac{\lambda p_{1FI}p_{12}p_{2FI}}{\mu_{2I}})(\frac{\lambda p_{1FI}p_{12}p_{2FI}}{\mu_{2I}})^{n_{2I}} & j = 2I \end{cases} . \quad (4)$$

3.2 Mean Time of Attack Data Spent in Network T_q

We can apply Little's formula because Jackson's theory tells us that each node is an independent queueing system in Jackson Networks. As a result, the mean time of attack data spent in network is given by,

$$T_q = \sum_{j \in TS} q_j/\lambda, \quad \text{where} \quad q_j = \rho_j/(1-\rho_j) \ . \quad (5)$$

q_j is the average queue length of the node j. Then, the mean time of attack data spent in each node can be written as

$$T_{qj} = \frac{q_j}{\lambda} = \begin{cases} \frac{1}{\mu_{1F}-\lambda} & j = 1F \\ \frac{p_{1FI}}{\mu_{1I}-\lambda p_{1FI}} & j = 1I \\ \frac{p_{1FI}p_{12}}{\mu_{2F}-\lambda p_{1FI}p_{12}} & j = 2F \\ \frac{p_{1FI}p_{12}p_{2FI}}{\mu_{2I}-\lambda p_{1FI}p_{12}p_{2FI}} & j = 2I \end{cases} \quad \text{and} \quad T_q = \sum T_{qj} = \sum_{j \in TS} \frac{\frac{V_j}{\mu_j}}{1-\lambda\frac{V_j}{\mu_j}} \ . \quad (6)$$

3.3 Probabilities From the Network Entry Node to Absorbing States, π_5 and π_6

π_5 and π_6 mean the probabilities of attack fail and succeed, respectively. Therefore, the two probabilities denote the degree of network security. Because all attack data enter the network from the entry, i.e., network firewall in the network boundary, we regard π_5 and π_6 as the probabilities from transient state 1F to absorbing state 5 and 6, respectively. To compute π_5 and π_6, we define the matrix

$$B = [b_{ij}], i \in TS, j \in AS \ , \quad (7)$$

where b_{ij} is the probability from transient state to absorbing state. Hence, the first row of elements is the answer we want, which mean the probability from 1F to absorbing states. We can use the result denoted in [5],

$$\begin{aligned} &B = MC, \quad \text{which gives,} \\ &\pi_5 = b_{1F5} = p_{1F5} + p_{1FI}p_{1I5} + p_{1FI}p_{12}p_{2F5} + p_{1FI}p_{12}p_{2FI}p_{2I5} \ . \\ &\text{and} \quad \pi_6 = b_{1F6} = p_{1FI}p_{12}p_{2FI}p_{26} \end{aligned} \quad (8)$$

3.4 Analysis of Network Security and Performance

The four measures, η, T_q, π_5 and π_6 are important in evaluating network security and performance, where π_5 and π_6 give the degree of network security, T_q indicates the network latency caused by security systems, η denotes the processing

capability of each node. However, the four measures are not consistent. Consider a given network, if we hope the degree of security to meets the specific requirements, such as $\pi_5 \geq 0.99$ and $\pi_6 \leq 0.01$, the network security solution may be proposed to satisfy the conditions. However, because the security solution induces service time at every check point, we should compute T_q and compare with the circumstance without the security solution, to ensure the increment of T_q is controlled in an acceptable range.

We now come to the impact of individual security component on network security and performance. By comparing p_{1F5}, p_{1I5}, p_{2F5}, and p_{2I5}, we can estimate the effectiveness of each security component to certain attacks. On the other hand, by comparing the factors of η, we can conclude which node is the processing bottleneck from the point of performance. When we want to improve the security effectiveness of a certain node, we also face problems of balancing security and performance.

4 Conclusions and Future Works

In this paper, we proposed a Jackson network-based quantitative model for network security after analyzing the attacking process. This model gives a dynamic, formalized analysis as well as considerations of network security as a whole. Through analysis of this model, we obtained four important measures related to network security and performance, to imply the result of confrontation between attacks and security boundaries. Our model can be extended to analyze security in the circumstance of hybrid attacks and complex network, such as intranet with hierarchy structure. Therefore, our future work is to extend our model in two ways. One is to model the states of one kind of attack data and transitions in the case of hierarchy network security boundaries, and the other is to model multiple kind of attacks. And further more, we should investigate the relationship between network security and performance, which will indicate us to find a balance.

References

1. Zhang, Y.R., Xian, M., Zhao, Z.C., Xiao, S.P. and Wang, G.Y., 2002 "A study on the evaluation technology of the attack effect of computer networks", *Guofang Keji Daxue Xue-bao/Journal of National Journal of National University of Defense Technology(Chinese)*, Vol. 24, No. 5, pp. 24-28
2. Jonsson, E. and Olovsson, T., 1997, "A quantitative model of the security intrusion process based on attacker behavior", *IEEE Transactions on Software Engineering*, Vol. 23, No. 4, pp. 235-245
3. Butler, S.A., 2002, "Security attribute evaluation method: a cost-benefit approach", In *Proc. of International Conference on Software Engineering*, pp. 232–240, may 2002
4. Madan, B.B., Goseva-Popstojanova, K., Vaidyanathan, K. and Trivedi, K.S., 2002, "Modeling and quantification of security attributes of software systems", In *Proc. of International Conference on Dependable Systems and Networks*, pp. 505-514, jun 2002
5. Medhi, J.. *Stochastic processes.* Wiley, New York, 1994.

Biomonitoring, Phylogenetics and Anomaly Aggregation Systems

David R.B. Stockwell[1] and Jason T.L. Wang[2]

[1] San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093
davids@sdsc.edu
[2] Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102
wangj@njit.edu

Abstract. While some researchers have exploited the similarity between cyber attacks and epidemics we believe there is also potential to leverage considerable experience gained in other biological domains: phylogenetics, ecological niche modeling, and biomonitoring. Here we describe some new ideas for threat detection from biomonitoring, and approximate graph searching and matching for cross network aggregation. Generic *anomaly aggregation* systems using these methods could detect and model the inheritance and evolution of vulnerability and threats across multiple domains and time scales.

1 Introduction

Threats such as surreptitious worms could arguably subvert upwards of 10,000,000 Internet hosts when launching a simultaneous attack on nation's critical infrastructure [1]. We are interested in applying lessons from biomonitoring and phylogenetic domains to the design of *anomaly aggregation* systems for detecting threats. Some research areas of relevance include biomonitoring [2, 3], invasive pest science and characterizing invasions [4, 5], predicting potential extents and threats [6, 7], infectious diseases [8-12], and phylogenetic tree inference and aggregation [13]. Several recently developed techniques for finding cousin 'species', consensus patterns, and paths in trees [14, 15] and graphs [16] are particularly applicable to aggregation of threat information. Integration of information from diverse sources will also be important for making use of background knowledge of the world 'environment', e.g. correlations with world events, characteristic source locations and target locations of traffic and the character of the particular attack [17].

1.1 Threat Detection (Short and Long Term)

Some success in recognizing threats has been achieved with 'signature' methods such as SNORT [18], and vulnerability assessment benchmarking programs developed by the Center for Internet Security [19]. While fast and informative, signature systems must be kept up to date with new threats. Anomaly detection systems such as MINDS [20] are more generic but have a reputation for high levels of false positives. In comprehensive testing of anomaly recognition systems, Local Outlier Factor methods (LOF) gave a lower rate of false positives [21].

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 523 – 529, 2005.

Biomonitoring of river systems for pollution has also discovered the limitations of signature methods based on recognition of signs of impact (e.g. eutrophication, turbidity), and is tending towards anomaly detection. Programs such as RIVPACS (River In-Vertebrate Prediction and Classification System) use observed data from a large set of 'pristine' river reference sites to predict expected species at monitoring sites and compare with observed species for anomalies [22]. Such systems have been successful in both detecting and quantifying subtle impacts, and programs using these analytical tools have led to long term improvements in river quality both in Europe and Australia [23]. It is not necessary for reference sites to be 'pristine' -- simply higher quality. New approaches have generalized localized reference set designs (like LOFs) [24, 25], and added rigorous *F* tests for level of impact and confidence [26]. Comprehensive evaluation of datasets from Australia and New Zealand demonstrated improvements over previous methods [3] and further development continues to yield improvements [2].

1.2 Aggregation (Inference and Evolution)

A graph can be used to model a social network in which each node represents a person and each edge (or link) represents the connection between two persons. Each person (node) may be associated with attributes such as name, age, gender, job title, education background, while each edge (link) between two persons may be associated with attributes indicating levels of connection such as frequency of meetings, or shared bank accounts. Social networks are widely used in economics and social informatics [27] but are also being used to model activities of terrorists, and for the analysis of hubs and authorities on the Internet [28]. Of particular interest to cyber-security are extensions of tree-matching [29-31] and *graphgrep* [16] techniques to find consensuses of graphs. We see as critical efficient solutions to three types of problems for cyber-security applications.

1. Recognizing Differences Between Graphs. Integrating intelligence such as terrorist networks or customer links requires new techniques for graph manipulation such as the *graphdiff* technique [32, 33] to reconcile the difference between networks. Building supergraphs is a more challenging problem than creating the supertree of trees because a graph may contain a circle. Techniques are available for graph matching whereby a graph is transformed to a tree first by removing the cross edges and back edges [15]. Building supergraphs for directed graphs is equally important in tracking software vulnerability. For example, when systems are upgraded, multiple versions of software are merged, and potentially unforeseen security threats may emerge.

2. Consensus of Graphs. Can we detect common topology, with or without considering node attributes or edge attributes, in a set of graphs? This is an analogy to sequence alignment where one can get a consensus sequence when aligning two or multiple sequences. Current graph mining tools such as SUBDUE do not really do this, because a structure might be infrequent yet still be isomorphic [34]. Detecting the consensus of multiple graphs may help to discover regional activities of terrorists, identify patterns in these activities and infer potential groups or individuals involved in a particular terror event. Identifying common topology of virus code or sequences of behavior is a precursor to more informative studies.

3. Graph Clustering and Mining. We can also partition a graph into regions and then cluster the regions. Another useful operation is to discover a hierarchy in a given graph, allowing one to find supervisor-subordinate relationships in a social network or in a terrorist network.

In solving these problems we may also want to consider the concept of equivalent classes of attribute values on nodes and edges, used in protein sequence alignment where substitution matrixes such as BLOSUM and PAM [35] are used to score amino acid pairs.

2 Applications

These approaches hold at many levels: whether the data is listing of processes on a machine (e.g. via *ps* command), a feature vector of connections arriving at a port (e.g. via *tcpscan*), or an inventory of vulnerability of software installed machines on a network (e.g. via *rpm* files). The following case studies illustrate the potential range of application domains.

1. Threat Detection. To describe how the biomonitoring approach might be applied to a cyber-security domain, we performed a preliminary application of E-ball biomonitoring software to a *ps* listing of a UNIX machine [3]. Unlike LOF approaches where all variables used a single feature vector, E-ball inputs two sets of variables, *environment* and *impact.* A real world example would use potentially hundreds of variables in both the environment and impact sets. The variables we extracted from a *ps* listing were:

Environment Variables *e*: NI: nice, UID: user ID, PID: process ID, PGID: group ID, VSZ: virtual memory size, F: flags.

Impact Variables *i*: %CPU: percentage processor time, %MEM: percentage memory used.

Table 1. Results of E-ball monitoring system illustrating reduction of false positive results

Run 1			**Run 2**		
Site	n	Pe		n	Pe
X	75	0.02	csh	2	0.05
mozilla-bin	75	0.58	X	2	0.52
nautilus	75	0.89	mozilla-bin	2	0.88
gnome-panel	75	0.91	nautilus	2	0.95
gnome-terminal	75	0.94	gnome-panel	5	0.96
rhn-applet-gui	75	0.95	gnome-terminal	5	0.97
emacs	75	0.96	rhn-applet-gui	5	0.97
apmd	75	0.97	emacs	43	0.97
atd	75	0.97	apmd	43	0.97
…	…	…	atd	43	0.97

Table 1 shows two runs of the anomaly detection system (Table 1). In Run 1 each process is compared with all others running on the machine (n=75), thus novel processes such as X windows using significantly more %MEM and %CPU have low probability (*Pe*) (i.e. a false positive). In the second run (Run 2), each process is compared with closest neighbors (n<75) according to the metrics defined on the environmental variables. This leads to a significant increase in probability of X based processes from 0.02 to 0.52 and 0.58 to 0.88, as each is only compared with similar processes according to the *e* feature vector. If a process were to act uncharacteristically, e.g. a Csh with high %MEM or %CPU it would however be flagged as anomalous.

2. Vulnerability Assessment. With adequate data of normal process information and parameterization of the *e* and *i* vectors, deviation from normality in potentially any domain can be detected and quantified. Where the rows are connection or machine information (for example using *rpm* files to identify versions of software, patching, and known security holes) we can quantify improvement of machines in a quality improvement program. There are two main ways of defining 'reference' and 'monitoring' sets: two physically different *production* and *outback* systems with strict and lax security policies, and secondly identifying anomalous units in a single network to iteratively improve the security by prioritizing and improving the worst units first.

3. Threat Diagnosis. Threats come in many forms, all however amenable to phylogenetic analysis for inherent tree and graph structures. Potential data sources include versions of virus codes, trace syslogs of incidents, historic web log files, and identified incidents. Applications of aggregation of information include: inferring the genealogy of viruses and other threats permitting inference across incidents, to identify common hubs and authorities shared by different search engines, and to evaluate the trust level of websites.

The following is a simple example of diagnosis of a Distributed Denial of Service (DDoS) attack known and SYN flood in which an attacker creates a number of half open TCP/IP connections with spoofed non-existent source addresses (i.e. spoofed.net.1191), held open as the final ACK message is never sent to the victim server. Below is example *tcpdump* output from a victim machine during a SYN flood attack:

```
20:03:42.105589 spoofed.net.1191 > 192.168.20.10.23: S
   70894115:70894115(0) win 8192 <mss 1460>
20:03:42.313064 192.168.20.10.23 > spoofed.net.1191: S
   1737393897:1737393897(0) ack 70894116 win 4288 <mss 1460>
20:03:42.105589 spoofed.net.1192 > 192.168.20.10.23: S
   70897870:70897870(0) win 8192 <mss 1460>
20:03:42.313064 192.168.20.10.23 > spoofed.net.1191: S
   1741139606:1741139606 (0) ack 70897871 win 4288 <mss 1460>
…
```

Aggregation into a grep-like string reduces data transfer and provides structured instances for anomaly detection. Further aggregation across machines on a network results in a motif representing the root node of a tree of specializations that could potentially be used to query other machines on the network for the threat.

```
20:03:42.* spoofed.net.*: S *(0) * <mss 1460> ?
```

3 Challenges

The major challenges implementing solutions from biological systems to cyber-security are improving background knowledge, extracting useful information, and increasing speed and rigor. Specific tools for increasing background knowledge are returning the geographic location of IP addresses, constructing AS-level graphs from the IP addresses in raw traces, use of large address space monitoring (a.k.a. Internet Telescope) to track and understand global network security events such as global Denial-of-Service attacks and Internet worms [19]. Three approaches to increasing speed are fast algorithms, identifying an 'ecology' of cyber-space defining where the useful information lies, and integrated database, analysis, and interface functions. One such system we have found useful is an efficient high level functional scripting language called K [36] with integrated entirely in memory database system, an 'inverted' database design storing information as columns not rows, functional syntax using vector operations, and low-level graphics functions. Originally developed for high throughput financial applications K has demonstrated superior speed in a range of areas e.g. over 1000 times faster at certain database operations than an Oracle database. The systems and techniques originating in the biological sciences described in this paper could allow the development of anomaly integration systems that both respond in real time with diagnostic information alerting users to the seriousness of a threat, and support system wide quality improvement programs.

References

1. Staniford, S., Paxson, V., Weaver, N.: How to 0wn the Internet in Your Spare Time. Proceedings of the 11th USENIX Security Symposium (Security '02) (2002)
2. Linke, S., Norris, R.H., Faith, D.P., Stockwell, D.: ANNA: A new prediction method for bioassessment programs. Freshw. Biol. (in press)
3. Linke, S., Norris, R., Faith, D.P.: Australian River Assessment System: Improving AusRivAS Analytical Methods DDRAM and E-Ball (Phase I Final Report). Commonwealth of Australia, Canberra and University of Canberra, Canberra, (2002)
4. Stockwell, D.R.B.Noble, I.R.: Induction Of Sets Of Rules From Animal Distribution Data - A Robust And Informative Method Of Data-Analysis. Mathematics And Computers In Simulation 33 (1992) 385-390
5. Peterson, A.T.: Predicting the geography of species' invasions via ecological niche modeling. Q Rev Biol 78 (2003) 419-33
6. Erasmus, B., Van Jaarsveld, A., Chown, S., Kshatriya, M., Wessels, K.: Vulnerability of South African animal taxa to climate change. Glob. Ch. Biol. 8 (2002) 679-693
7. Peterson, A.T., Vieglais, D.A., Andreasen, J.K.: Migratory birds modeled as critical transport agents for West Nile Virus in North America. Vector Borne Zoonotic Dis 3 (2003) 27-37
8. Costa, J., Peterson, A.T., Beard, C.B.: Ecologic niche modeling and differentiation of populations of Triatoma brasiliensis neiva, 1911, the most important Chagas' disease vector in northeastern Brazil (hemiptera, reduviidae, triatominae). Am J Trop Med Hyg 67 (2002) 516-20
9. Peterson, A.T., Bauer, J.T., Mills, J.N.: Ecologic and geographic distribution of filovirus disease. Emerg Infect Dis 10 (2004) 40-7

10. Levine, R.S., Peterson, A.T., Benedict, M.Q.: Distribution of members of Anopheles quadrimaculatus say s.l. (Diptera: Culicidae) and implications for their roles in malaria transmission in the United States. J Med Entomol 41 (2004) 607-13
11. Levine, R.S., Peterson, A.T., Benedict, M.Q.: Geographic and ecologic distributions of the Anopheles gambiae complex predicted using a genetic algorithm. Am J Trop Med Hyg 70 (2004) 105-9
12. Beard, C., Pye, G., Steurer, F., Rodriguez, R., Campman, R., Peterson, A., Ramsey, J., Wirtz, R., Robinson, L.: Chagas disease in a domestic transmission cycle in southern Texas, USA. Emerg. Infect. Dis. 9 (2003) 103-105
13. Shasha, D., Wang, J.T.L., Zhang, S.: Unordered Tree Mining with Applications to Phylogeny. Proceedings of the 20th International Conference on Data Engineering, Boston, Massachusetts (2004)
14. Shasha, D., Wang, J.T.L., Shan, H., Zhang, K.: ATreeGrep: Approximate Searching in Unordered Trees. Proceedings of the 14th International Conference on Scientific and Statistical Database Management, Edinburgh, Scotland (2002)
15. Wang, J.T.L., Zhang, K., Chang, G., Shasha, D.: Finding Approximate Patterns in Undirected Acyclic Graphs. Pattern Recogn. 35 (2002) 473-483
16. Shasha, D., Wang, J.T.L., Giugno, R.: Algorithmics and Applications of Tree and Graph Searching. Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Madison, Wisconsin (2002)
17. Vatis, M.: Cyber Attacks During the War on Terrorism: A Predictive Analysis. Dartmouth College, (2001)
18. Caswell, B.Roesch, M.: SNORT Intrusion Detection System. (2004)
19. CAIDA: Center for Internet Security. (2003)
20. Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Srivastava, J., Kumar, V., Dokas, P.: The MINDS - Minnesota Intrusion Detection System: Next Generation Data Mining: MIT Press (2004)
21. Lazarevic, A., Ertoz, L., Ozgur, A, Srivastava, J., Kumar, V.: A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection. Proceedings of Third SIAM Conference on Data Mining, San Francisco (2003)
22. Clarke, R., Furse, M., Wright, J., Moss, D.: Derivation of a biological quality index for river sites: Comparison of the observed with the expected fauna. J. Appl. Stats. 23 (1996) 311-332
23. Clarke, R., Wright, J., Furse, M.: RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecol. Model. 160 (2003) 219-233
24. Faith, D., Dostine, P., Humphrey, C.: Detection of moning impacts of aquatic macroinvertebrate communities - results of a disturbance experiment and the design of a multivariate BACIP monitoring program at Coronation Hill, Northern Territory. Aust. J. Ecol. 20 (1995) 167-180
25. Humphrey, C., Faith, D., Dostine, P.: Base-line rquirements for assessment of moining impact using biological monitoring. Aust. J. Ecol. 20 (1995) 150-166
26. Stockwell, D.R.B.Faith, D.P.: Investigation of alternative approaches to linking habitat variables with site classification in a RIVPACS model - Final Report., (1996)
27. Eubank, S., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Wang, N.: Structural and algorithmic aspects of massive social networks. Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, New Orleans, Louisiana (2004)

28. Moret, B.M.E., Nakhleh, L., Warnow, T., Linder, C.R., Tholse, A., Padolina, A., Sun, J., Timme, R.: Phylogenetic networks: modeling, reconstructibility, and accuracy. IEEE/ACM Transactions on Computational Biology and Bioinformatics 1 (2004) 13-23
29. Wang, J.T.L., Shapiro, B.A., Shasha, D., Zhang, K., Currey, K.M.: An Algorithm for Finding the Largest Approximately Common Substructures of Two Trees. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 889-895
30. Wang, J.T.L.Zhang, K.: Information Sciences 126 (2000) 165-189
31. Wang, J.T.L.Zhang, K.: Finding Similar Consensus between Trees: An Algorithm and a Distance Hierarchy. Pattern Recogn. 34 (2001) 127-137
32. Wang, J.T.L., Zhang, K., Chirn, G.-W.: Algorithms for Approximate Graph Matching. Information Sciences 82 (1995) 45-74
33. Zhang, K., J. T. L. Wang and D. Shasha: On the Editing Distance between Undirected Acyclic Graphs. International Journal of Foundations of Computer Science 7 (1996) 43-57
34. Cook, D.J., Holder, L.B.: Graph-Based Data Mining. IEEE Intelligent Systems 15 (2000) 32-41
35. Wang, J.T.L., Zaki, M.J., Toivonen, H.T.T., Shasha, D.: Data Mining in Bioinformatics. London/New York: Springer, (2004)
36. Whitney, A.: K programming language. (2004)

CODESSEAL: Compiler/FPGA Approach to Secure Applications*

Olga Gelbart[1], Paul Ott[1], Bhagirath Narahari[1], Rahul Simha[1], Alok Choudhary[2], and Joseph Zambreno[2]

[1] The George Washington University, Washington, DC 20052 USA
[2] Northwesten University, Evanston, IL 60208 USA

Abstract. The science of security informatics has become a rapidly growing field involving different branches of computer science and information technologies. Software protection, particularly for security applications, has become an important area in computer security. This paper proposes a joint compiler/hardware infrastructure - CODESSEAL - for software protection for fully encrypted execution in which both program and data are in encrypted form in memory. The processor is supplemented with an FPGA-based secure hardware component that is capable of fast encryption and decryption, and performs code integrity verification, authentication, and provides protection of the execution control flow. This paper outlines the CODESSEAL approach, the architecture, and presents preliminary performance results.

1 Introduction

With the growing cost of hacker attacks and information loss, it is becoming increasingly important for computer systems to function reliably and securely. Because attackers are able to breach into systems in operation, it is becoming necessary not only to verify a program's integrity before execution starts, but also during runtime. Attackers exploit software vulnerabilities caused by programming errors, system or programming language flaws. Sophisticated attackers attempt to tamper directly with the hardware in order to alter execution. A number of software and software-hardware tools have been proposed to prevent or detect these kinds of attacks [1, 2, 3, 8]. Most of the tools focus on a specific area of software security, such as static code analysis or dynamic code checking. While they secure the system against specific types of attacks, current methods do not provide code integrity, authentication, and control flow protection methods that address attacks using injection of malicious code.

We propose a software/hardware tool - CODESSEAL - that combines static and dynamic verification methods with compiler techniques and a processor supplemented with a secure hardware component in the form of an FPGA (Field Programmable Gate Array) in order to provide a secure execution environment

* The research is supported in part by NSF grant CCR-0325207.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 530–535, 2005.

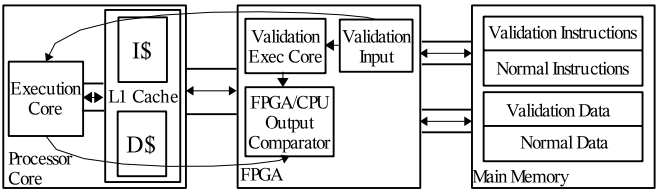


Fig. 1. Processor Core Supplemented with an FPGA

for fully encrypted execution. Figure 1 shows the architecture of our hardware. The main objective of the tool is to provide techniques to help proper authorization of users, to prevent code tampering and several types of replay, data, and structural attacks (such as control flow attacks), and to make it considerably harder for attackers to extract any information that could point to a potential vulnerability. CODESSEAL is also designed to support fully-encrypted executables to ensure confidentiality. Our preliminary experimental results reveal low performance overheads incurred by our software protection methods for many applications.

Several software techniques have been proposed for code security; these range from tamper resistant packaging,copyright notices, guards, code obfuscation [1, 2, 3, 4] Software techniques typically focus on a specific type of vulnerability and are susceptible to code tampering and code injection by sophisticated attackers. Hardware techniques, including secure coprocessors and use of FPGAs as hardware accelerators [9, 10], are attractive because of the quality of service they provide but at the same time require substantial buy-in from hardware manufacturers and can considerably slow down the execution. Of greater relevance to our approach are the combined hardware-software approaches such as the XOM project [8], the Secure program execution framework (SPEF) [7], and hardware assisted control flow obfuscation [11]. In addition to other advantages over these techniques, such as the use of reconfigurable hardware in the form of FPGAs and working on the entire compiler tool chain, our approach addresses new problems arising from attacks on encrypted executables.

2 Our Approach: CODESSEAL

CODESSEAL (COmpiler DEvelopment Suite for SEcure AppLications) is a project focused on joint compiler/hardware techniques for fully encrypted execution, in which the program and data are always in encrypted form in memory. Encrypted executables do not prevent all forms of attack. Several types of replay, data and structural attacks, such as control flow attacks, are possible and can uncover program behavior. We term such attacks as EED attacks – attacks on *Encrypted Executables and Data*. EED attacks are based on exploiting structure in encrypted instruction streams and data that can be uncovered by direct manipulation of hardware (such as address bus manipulation) in a well-equipped

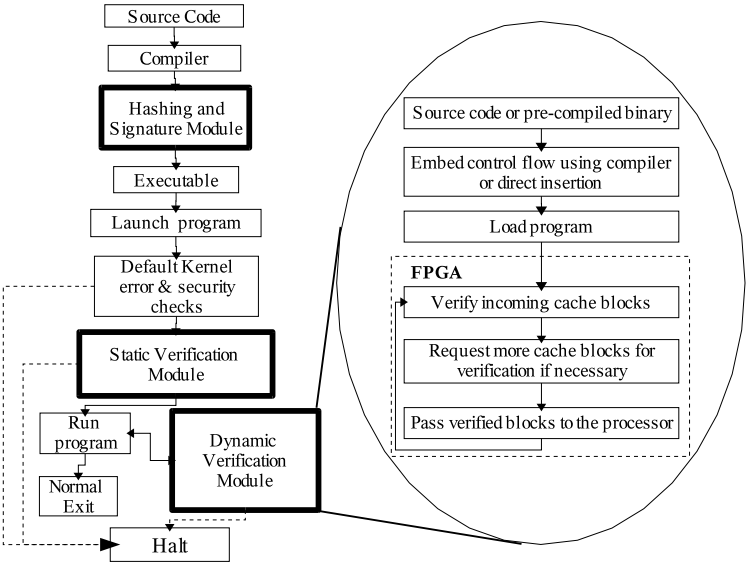


Fig. 2. CODESSEAL Framework

laboratory. To help detect such attacks, compilers will need to play a key role in extracting structural information for use by supporting hardware.

Fig. 2 describes the overall CODESSEAL framework. It has two main components – (1) static verification and (2) dynamic verification. At compile-time, static integrity and control flow information is embedded into the executable. The static verification module checks the overall integrity of the executable and also checks its signature. Upon success, the executable is launched and each block is dynamically verified in the FPGA for integrity and control flow. The CODESSEAL hardware architecture is shown in Figure 1. The architecture reconfigurability provided by an FPGA provides us with the ability to change the verification and cryptographic algorithms embedded in the hardware by simply reprogramming the FPGA. A detailed exposition of the static verification and authentication components of our framework, and experimental results, are provided in [6]. In this paper we focus on the dynamic verification component and present an implementation and preliminary performance study.

2.1 Dynamic Verification

If the static verification is performed successfully, the system allows launching of the executable. After this, the dynamic verification module is responsible for preventing run-time attacks on the program. Our dynamic verification module has two components: (i) checking that code and data blocks have not been modified at run-time by an attacker and (ii) asserting the legal control flow in the program, i.e., any changes made to the control flow graph of the program is considered equivalent to code tampering and the program is halted.

Our technique involves the use of an FPGA placed between main memory and the cache that is closest to main memory (L1 or L2, depending on the system) – see Figure 2. The instructions and data are loaded into the FPGA in blocks and decrypted by keys that are exclusive to the FPGA. Thus, the decrypted code and data are visible "below" the FPGA, typically inside a chip, thereby preventing an attack that sniffs the address/data lines between processor and memory. The original code and data are encrypted by a compiler that uses the same keys. The assumption is that both FPGA loading and compilation occur at a safe site prior to system operation.

If full encryption is not used, instruction and data block hashes (using SHA-1, for example) can be maintained inside the FPGA and verified each time a new block is loaded. If the hash does not match the stored hash, the processor is halted. Even when full encryption is used, it is desirable to perform a hash check because a tampering attack can be used to disrupt execution without decryption. While this technique maintains code integrity it does not prevent structural (control flow) attacks which is the subject of our ongoing work.Our approach to preventing control-flow attacks embeds the control flow information, as captured by a control flow graph of the program, into the code. The signature of each block contains a hash of itself as well as control flow information denoting parent and child blocks. During the execution, only blocks whose hash and parent information is verified are permitted for execution. If a malicious block is introduced, its hash or control flow verification (performed in the secure FPGA component) will fail, thereby halting the program. Note that by using additional hardware to verify the program at runtime, we avoid adding additional code to the executable, thus preventing code analysis attacks.

Data tampering in encrypted systems is more complicated because a write operation necessitates a change in the encryption: data needs to be reencrypted on write-back to RAM. Also, because data can get significantly larger than code, a large set of keys might be needed to encrypt data, resulting in a key management problem.

2.2 Preliminary Experimental Results

Experimental Setup. For the dynamic verification technique, we have currently implemented a scheme in which each instruction or data block of the executable is hashed using SHA1 algorithm. (Ongoing work explores the use of other cryptographic algorithms.) The hashes are stored in each block as well as in the FPGA. The FPGA performs hash verification as each block loads. Each block verification involves three steps: on L1 cache miss, a block is brought in, (a) its hash is calculated, (b) the "correct" hash is fetched from the FPGA memory and (c) the two hashes are compared.

The experimental setup was as follows. We used SimpleScalar version 3.0 processor simulator for the ARM processor and a gcc 3.3 cross-compiler. The ARM processor chosen to be represented by SimpleScalar had an ARM1020E core and ran at 300 MHz. The FPGA chosen was modeled after the Virtex-II XC2V800 and ran at 150 MHz and had at most 3 MB on onboard memory. The

Table 1. Performance results for dynamic verification

Benchmark	Comment	No. instr	%Penalty (instr)	%Penalty (data)	%Penalty (both)
djpeg (MiBench)	136KB, 720x611	67.52M	.0389	2.2368	2.2757
	162KB,1265x2035	232.28M	.0117	4.3126	4.3242
rijndael(MiBench)	pdf 116.7KB	9.55M	.1238	.1100	.2337
	jpg 455KB	39.90M	.0297	.0263	.0560
susan(MiBench)	256KB, 512x512	71.16M	.0160	.7749	.7949
blowfish(MiBench)	pdf 116.7KB	20.21M	.0301	.0350	.0650
	jpg 455KB	83.86M	.0072	.0084	.0157
go(SPECINT2000)	6x6 board	26.22M	10.1072	7.7642	17.8724
	8x8 board	75.23M	9.5524	7.0244	16.5768
Transitive closure (DIS)	16-64 vertices 123-2048 edges	15.24M	0.8	3.36	4.16
	256-512 vertices 16384-196608 ed	7.22B	0	42.89	42.89
field (DIS)	Average for 5 runs	2.9B	0.02	0.08	0.08
pointer (DIS)	Average for 11 runs	1.09B	0.02	1.24	1.24

performance penalty for each of the three steps described above was: 6 processor cycles(3 FPGA cycles)for step (a) to calculate SHA-1 (maximum clock rate is 66 Mhz), 2 processor cycles(1 FPGA cycle) for step (b) (maximum clock rate for memory is 280 Mhz) and 2 processor cycles(1 FPGA cycle) for step (c) to compare hashes. The memory requirement on the FPGA was 22 bytes per cache block(20 per hash and 2 to map addresses to hashes).

Results and Analysis. Dynamic verification was implemented for both code and data blocks and tested for a number of applications from the MiBench, DIS (Data intensive systems), and SPECINT2000 benchmarks. The control flow protection scheme is currently being implemented and thus not included in the results presented in this paper. Our experiments show an average of 3.97% performance penalty. The experimental results are summarized in Table 1, and reveal that our dynamic verification techniques result in very low performance degradation (overheads) in most cases. Some of the data intensive systems benchmarks, such as transitive closure, result in high overheads (of 42%) thereby motivating the need to study tradeoffs between security and performance.

3 Conclusions and Future Work

This paper proposed a tool - CODESSEAL - that combines compiler and FPGA techniques to provide a trusted computing environment while incurring low performance overhead in many benchmarks. The goal of our tool is to provide additional security to a computer system, where software authentication, integrity verification and control flow protection are particularly important. The tool is

also designed to protected mission-critical applications against a hands-on attack from a resourceful adversary.

References

1. Cowan, C.: Software Security for Open-Source Systems. IEEE Security and Privacy (2003)
2. Chang, H., Attallah, M.J.: Protecting Software Code by Guards. Proceedings of the 1st International Workshop on Security and Privacy in Digital Rights Management (2000) 160-175
3. Colberg, C., Thomborson, C., Low, D.: A taxonomy of obfuscating transformations. Technical Report. Dept of Computer Science, Univ. of Auckland (1997)
4. Fisher, M: Protecting binary executables. Embedded Systems Programming (2000) Vol.13(2).
5. Actel: Design security with Actel FPGAs. http://www.actel.com (2003)
6. Gelbart, O., Narahari, B., Simha, R.: SPEE: A Secure Program Execution environment tool using static and dynamic code verification. Proc. the 3rd Trusted Internet Workshop. International High Performance Computing Conference. Bangalore, India (2004)
7. Kirovski, D., Drinic, M., Potkonjak, M.: Enabling trusted software integrity. Proceedings of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems (2002) 108-120
8. Lie, D., Thekkath, C., Mitchell, M., Lincoln, P., Boneh, D., Mitchell, J., Horowitz, M.: Architectural support for copy and tamper resistant software. Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems (2000) 168-177
9. Smith, S., Austel, V.: Thrusting trusted software: towards a formal model of programmable secure coprocessors. Proceedings of the 3rd USENIX Workshop on Electronic Commerce (1998) 83-98
10. Taylor, R., Goldstein, S.: A high-preformance flexible architecture for cryptography. Proceedings of the Workshop on Cryptographic Hardware and Software Systems (1999)
11. X. Zhuang, T. Zhang, S. Pande: Hardware assisted control flow obfuscation for embedded processors. Proc. of Int. Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES) 2004.

Computational Tool in Infrastructure Emergency Total Evacuation Analysis

Kelvin H.L. Wong and Mingchun Luo

ArupFire, Ove Arup & Partners HK Limited,
Level 5, Festival Walk, 80 Tat Chee Avenue, Kowloon Tong,
Hong Kong SAR, China
kelvin.wong@arup.com

Abstract. Investigation has been made in the total evacuation of high profile infrastructures like airport terminal, super-highrise building, racecourse and tunnels. With the recent advancement of computer technologies, a number of evacuation modelling techniques has been developed to visualize the evacuation pattern and optimize the evacuation provisions. Computer simulations enable the integration of individual human factors like age, gender, percentage of crowd, mobility impairment, walking speed and patience level into evacuation model. Other behavioural factors like shortest distance, quickest time, adjacent movement and personal space can also be computed. The simulation results can be more realistic and reliable then the traditional hand calculations or code compliance design which cannot consider the actual performance and human factors. The simulation results can be used to characterize the efficiency of total evacuation and to maximize the life safety protection, which is a major concern by general public and authorities for high-profile infrastructures.

1 Introduction

The catastrophic terrorist attack on September 11, 2001 to the prominent World Trade Center in New York has demonstrated that an efficient emergency total evacuation is extremely important in ensuring the building occupants safety [1]. Following the incident, the Occupational Safety & Health Administration of US Department of Labor developed an Evacuation Planning Matrix to provide employers with planning considerations to reduce their vulnerability to a terrorist act or the impact of a terrorist release. The matrix divided into three zones with different procedures of notification, isolation, evacuation and training in emergency planning. Since the Matrix is limited to the individual employer, the development of emergency evacuation plan can only be limited to the identification of additional or different evacuation routes, exits, staging locations and shelter areas within individual rental area.

The traditional egress provisions of buildings are primarily designed for fire scenarios in which zoned/phased evacuation is suggested and considered effective to cater for most of the fire scenarios. The phase evacuation is initiated via the use of building public address system to evacuate the floor/zone of fire origin and adjacent floors/zones. Occupants on the other floors/zones are remote from the fire and assumed to be safe for some time. Therefore the evacuation provisions like stairs and

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 536 – 542, 2005.

exits have been optimized. However, in the situation of extreme emergencies such as bomb threat and NBC attack, total evacuation is necessary. The traditional provisions in terms of egress are not sufficient to handle these new situations. For the simultaneous evacuation of building from multiple floors, blockage of staircases and exits by occupants from upper floors can easily be ignored. Therefore, a total building evacuation analysis instead of individual tenants is necessary to ensure the occupants safety, especially in the existing high-profile, high-capacity infrastructures to characterized the evacuation efficiency and maximize the life safety protection.

2 Tools for Evacuation Analysis

Computer simulation is a modern tool in evacuation analysis. The evolution of computer models in evacuation has been discussed in literature [2]. With the advancement in computational power, a number of powerful evacuation models have been developed [3]. These models can be classified into following two major categories.

The first type of evacuation simulation model is flow-based model. It treats the movement of occupants as a continuous flow like fluid rather than an aggregate of persons. Each room/floor in the building is regarded as a node with connections to another nodes by corridors, stairs and lobbies as links. The usable area in each node is specified to calculate the maximum number of occupants inside each node, and the width of links is specified to calculate the flow rate between nodes. The nodes and links form a network which occupants flow from node through links to the ultimate node (ultimate safety). The advantage of this model is the capability in handling multiple in and out flows of complex multi-compartment buildings. Treating the floor plan as a single node can eliminate the detail modeling of floor plan layout. Therefore the computational time can be reduced. No individual human behaviors such as physical abilities, individual positioning and direction of movement can be considered. For detail evacuation pattern analysis, another model type has been developed.

The second type of evacuation model is the agent-based model. It individualized the movement of occupants by dividing the floor space into cells. Each occupant can occupy one or multiple cells and moves around the floor space. Furniture modeled as blockages can be inserted which occupant cannot walk though. This feature is important for the visualization of evacuation pattern in the space densely packed with furniture like tables and chairs. By modeling the occupants as individuals, certain set of attributes can be assigned such as gender, age, mobility impairment, walking speed, body size, familiarity of exits, patience level, etc. More realistic evacuation behavior such as overtaking of slow walking occupants, avoidance of physical obstruction and searching of shortest evacuation path can be modeled in the agent-based model.

The calculation of agent-based model is repeated in each time step for each individual to simulate the movement of occupants on floor space. Therefore the computational time is much longer than the flow-based model. The simulation of a hundred floor super-highrise building with 20,000 occupants can be five hours for a 3GHz processor. Some of the agent-based evacuation models include a post-processor which can generate 3-D moving human shape to visualize the evacuation movement. Different occupants types can be highlighted for easy recognition and further analysis.

3 STEPS Evacuation Model

Since the agent-based model can take into account the human factors and floor plan layout, an agent-based evacuation simulation model STEPS (Simulation of Transient Evacuation and Pedestrian movementS) has been selected in the emergency total evacuation analysis. STEPS is designed to simulate how people move in both normal and evacuation situations within complex building structures. Other agent-based models are also available for evacuation modeling, but STEPS can model a large-scale evacuation in terms of tenths of thousand occupants with multi floors 3-D visualization. Human factors can also be inputted in STEPS for more realistic simulation.

The calculation algorithm of STEPS model is based on a grid system, where walls and furniture are modeled as obstruction. People are added to the system in available predefined cells. Each person calculates a score for every exit on their current plane based upon four cumulative criteria: the shortest distance to an exit, familiarity with an exit, crowding around an exit and the service rate of each exit. The total score is calculated for each person on every time step according to Eq. 1, where D is the distance to an exit, v is the walking speed, N is the number of persons queuing at the exit, f is the exit flow rate, P_{ahead} is the potential number of persons in front of the calculated ones, P_{target} is the potential number of persons reaching the exit, and C_1, C_2, C_3 are constants to determine the occupants' patience, queuing and walking characteristics respectively. The score S_{Total} determines the exit that the occupant is heading to on every time step.

$$S_{Total} = C_1\left(\frac{D}{v} - P_{ahead}\right) + C_2C_3\left(\frac{N}{f} - P_{target}\right) \tag{1}$$

Three interconnecting components in the model are considered: the plane and path network, the description of the human characteristics, and the movement of the people within the system. The algorithm for a person to select the travel path is based on a combination of decision and network-based models. Planes that represent the actual floor space consist of a grid configuration on which people can walk, the spacing of which is dependant on the maximum specified population density. Alternatively, predefined paths or planes are used to represent stairways, upon which deviations of the walking directions are not possible until another path or plane is reached.

The specification of occupants consists of people types, body dimensions and their associated walking speeds. The walking speed values are adopted with reference to the SFPE Handbook [4] for emergency evacuation design. Parameters for occupants' familiarity and patience are according to software default values. The choice of appropriate parameters allows convenience in modeling different scenarios.

This software visualize the evacuation in 3-D animation, users could rotate the model and using the zoom function for a detail inspection. The simulation results are generated in the form of spreadsheet, snapshot and video files. These can assist the user for a better understanding of evacuation pattern.

The accuracy of the STEPS model was compared [5] with two hand calculation examples demonstrated in NFPA 130 [6] and STEPS gave more conservative values. Unlike hand calculations, STEPS simulates an uneven population distribution in using stairs during evacuations, which leads to more realistic and conservative results.

4 Discussion on Key Infrastructures Total Evacuation

Total evacuation of public facilities has been studied here due to the uniqueness in shape and function that a major incident in these buildings would cause significant life lost. Some of them are high-profile infrastructures that would draw the general public and international community great attention for any large-scale incident.

Airport Terminal

For the airport terminal buildings zone evacuation is effective to handle most of the emergency scenarios like accidental fires, and can minimize the interruption of airport operation. But there is a concern by the public and officials that it is necessary to have a total evacuation analysis to access the egress provisions in extreme emergency since the airport is sometimes a symbolic building for a city or country. With the actual flight schedules and statistical dwell time of passengers in terminal buildings, the total number of occupants for evacuation can be estimated for a performance based analysis instead of adherence to prescriptive code values. When comparing the total occupant number with the floor area of terminal building, the occupant density is much lower than office space as shown in Fig. 1. The large number of staircases to provide sufficient discharge capacity for peak population in each zone is far enough to handle the peak occupant load of the whole building. In addition, the simulation also showed occupants in circulation area can be cleared in a relative short period of time since they do not need to pass through rooms and corridors to reach the exit staircases.

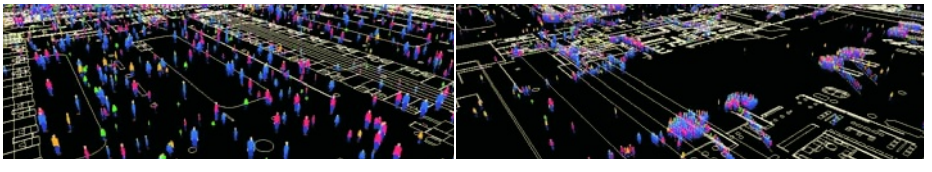


Fig. 1. The check-in hall before evacuation (*left*) and M&G hall during evacuation (*right*)

Super-Highrise Building

Highrise and super-highrise buildings are still major concern to the life safety of occupants due to the elevated height and extended travel distance for the egress and means of access. Although most of super-highrise buildings consist of different type of facilities like observation deck, hotel, restaurant, clubhouse, office and retail, most of them would be separated by the fire rated floor slab into different fire compartments. Phased evacuation is effective to handle most of the fire scenarios. The occupants on the most dangerous floor can be prioritized to evacuate first using phased evacuation and queuing time into staircase can be reduced. Therefore the egress provisions are designed to handle the evacuation of a few floors only. In case of total building evacuation, all occupants approach the exits at the same time. Extensive queuing observed in the exit entrances and protected lobbies after staircases and landings were filled with people as shown in Fig. 2. Some cities require the provision of designated refuge floors for temporary stabling of occupants [7]. Refuge floor can provide a temporary place of safety for a short rest before occupants to continue their journey. Refuge floors as reservoirs allowed the continuous flow of upper staircases

occupants in the simulations, which was shown to be an enhancement for super-highrise building evacuation. Simulation also showed a 35% decrease in total evacuation time with the assistance of elevators [8].

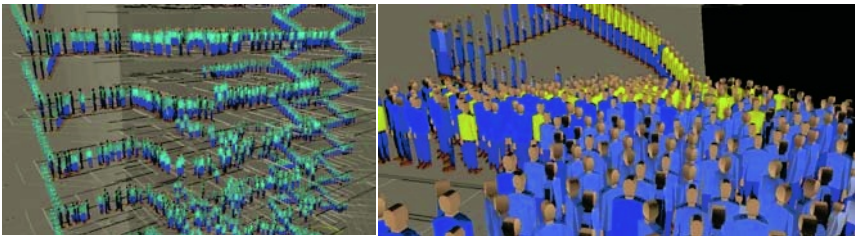


Fig. 2. The typical office floors (*left*) and refuge floor (*right*) during evacuation

Racecourse

Racecourses consist of inclined grand viewing stand and multiple floors structure to maximize the viewing angle. The seats and stairs on viewing stand create further difficulty for large number of occupants in evacuation. Sufficient handrails must be provided. Simulation showed that most of the occupants from the viewing stand discharged onto the forecourt to reach the final exit as shown in Fig. 3. The occupants cannot reach the forecourt directly used the staircases for evacuation. The exit width on forecourt and total staircase width can be optimized using evacuation modeling.

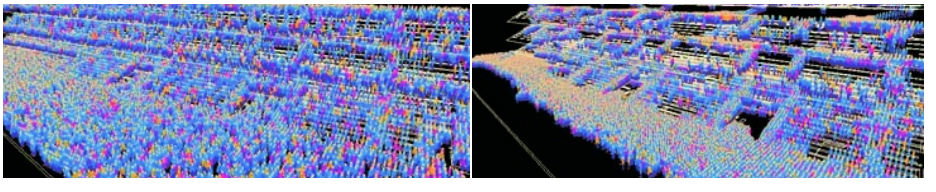


Fig. 3. The racecourse before evacuation (*left*) and during evacuation (*right*)

Metro/Railway Tunnel

For extreme scenario, train may lose power and stop in the middle of underground tunnel. Sufficient egress provisions for crush loading train passengers are important since means of access is challenging for narrow tunnel tube. Continuous walkway connected to cross-passages (CP) and adjacent tunnel or smooth in-track walkway leading to the next station is provided in railway or metro system respectively. With

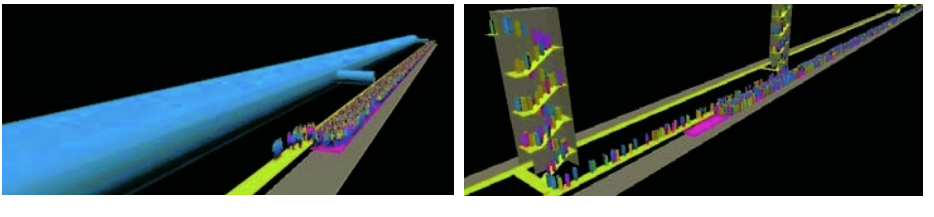


Fig. 4. The detrainment from one end (*left*) and both ends (*right*) of train inside tunnel

the combinations of different egress provisions, fire locations, emergency procedures and operation of smoke extraction system, evacuation modeling is the best tool to visualize and optimize the evacuation strategy as shown in Fig. 4.

Road Tunnel

There are two general type of construction methods for road tunnel tubes: immersed tube (IMT) and shield driven (SD). The construction of IMT tunnels allows a closer separation distance between two CP while SD tunnels required the separate construction of CP which represent a significant cost implication. With the consideration of different pre-movement time for various vehicle locations, the evacuation of tunnel occupants have been simulated to justify the means of escape provisions before tunnel environment become untenable for a performance based design approach.

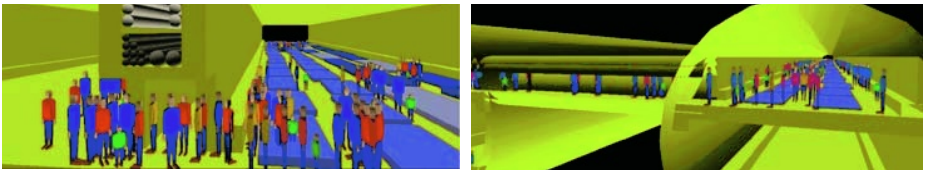


Fig. 5. The IMT tunnel (*left*) and SD tunnel (*right*) during evacuation

6 Conclusions

Total evacuation of a few key infrastructures has been studied and discussed using computer simulation tool STEPS. With the consideration of human factors, more realistic results than the traditional hand calculation method can be obtained. Since the traditional egress provisions are mainly designed for phased evacuation only, total building evacuations have been simulated to address the critical areas for further improvements. With the visualization of complex evacuation process, areas of concern can be highlighted to improve the evacuation efficiency and optimize the egress provisions to enhance the life safety protection of occupants during extremely emergencies, which require total building evacuation.

References

1. Fahy, R.F., Proulx, G.: A Comparison of the 1993 and 2001 Evacuations of the WTC. Proc. of Fire Risk and Hazard Assessment Symposium, Baltimore MD, (2002) 111-7
2. Gwynne, S., et al.: A Review of the Methodology Used in the Computer Simulation of Evacuation from the Built Environment. Building and Environment 34 (1999) 741-9
3. Olenick, S.M., Carpenter, D.J.: An Updated International Survey of Computer Models for Fire and Smoke. J. of Fire Protection Eng. V.13 No.2 (2003) 87-110
4. Nelson, H.E. Mowrer, F.W.: Emergency Movement. The SFPE Handbook of Fire Protection Engineering (3rd ed), Dinenno P.J. (ed), NFPA, Quincy MA, (2002) 3/370
5. Wall, J.M., Waterson, N.P.: Predicting Evacuation Time - A Comparison of the STEPS Simulation Approach with NFPA 130. Fire Command Studies

6. Standard for Fixed Guideway Transit and Passenger Rail Systems. NFPA 130, (2003)
7. Lo, S.M., et al.: A View to the Requirement of Designated Refuge Floors in High-rise Buildings in HK. Fire Safety Science - Proc. of the 5th Int. Sym., IAFSS, (1997) 737-45
8. Guo, Dagang, Wong, H.L.K., Luo, M.C., Kang, L., Shi, B.B.: Lift Evacuation Design of Ultra-Highrise Building. Proc. of Fire Conf. 2004 - Total Fire Safety Concept, (2004) 141-8

Performance Study of a Compiler/Hardware Approach to Embedded Systems Security*

Kripashankar Mohan, Bhagi Narahari, Rahul Simha, Paul Ott[1], Alok Choudhary, and Joe Zambreno[2]

[1] The George Washington University, Washington, DC 20052
[2] Northwestern University, Evanston, IL 60208

Abstract. Trusted software execution, prevention of code and data tampering, authentication, and providing a secure environment for software are some of the most important security challenges in the design of embedded systems. This short paper evaluates the performance of a hardware/software co-design methodology for embedded software protection. Secure software is created using a secure compiler that inserts hidden codes into the executable code which are then validated dynamically during execution by a reconfigurable hardware component constructed from Field Programmable Gate Array (FPGA) technology. While the overall approach has been described in other papers, this paper focuses on security-performance tradeoffs and the effect of using compiler optimizations in such an approach. Our results show that the approach provides software protection with modest performance penalty and hardware overhead.

1 Introduction

The primary goal of software protection is to reduce the risk from hackers who compromise software applications or the execution environment that runs applications. Our approach to this problem has been described in an earlier paper [1]. In this approach, bit sequences are inserted into the executable by the compiler that are then checked by supporting hardware during execution. The idea is that, if the code has been tampered, the sequence will be affected, thereby enabling the hardware to detect a modification. At this point, the hardware component can halt the processor.

Observe that the compiler can instrument executables with hidden codes in several ways. The most direct approach is to simply add codewords to the executable. However, this has the disadvantage that the resulting executable may not execute on processors without the supporting hardware, and may be easily detected by an attacker. Our approach is to employ the freedom that the compiler has in allocating registers. Because there are many choices in allocating

* This work is supported in part by Grant CCR 0325207 from the National Science Foundation.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 543–548, 2005.

registers, compilers can use these choices to represent binary codes that can then be extracted by the hardware. And because the register allocation is a valid one, the executable will run on processors that do not perform any checking.

The hardware support is needed so that the checking is itself not compromised, as is possible with software checking methods [10]. While it is generally expensive to build custom hardware support for this type of run-time checking, Field Programming Gate Array (FPGA) technology provides an attractive alternative. These programmable fabrics are today available with many commercial processors, and can easily be configured to perform the kind of checking during runtime. Because FPGA's are usually on-chip, and because they can be optimized to perform simple computations, they are also space and time efficient. Thus, the combination of compiler and FPGA technology makes the whole approach worthy of investigation. The purpose of this paper is to explore the performance of this approach.

Several factors can affect performance when actively checking an executable at runtime. The computation time in the FPGA depends on the lengths of the codes, their spread across the executable and how often the checking is performed. In particular, we consider the lengths of basic blocks and simple compiler techniques such as loop unrolling. We find that, overall, the approach imposes a modest penalty. At the same time, we find that loop-unrolling provides negligible performance benefit, thereby suggesting that other compiler techniques will be needed to further reduce the performance overhead.

Because a complete survey of software protection and security-driven compilation is presented in some of our earlier work [1, 9], we refer the reader to these papers for reviews of related work.

2 The Compiler/Hardware Approach

Figure 1 depicts the overall system architecture using our earlier approach [1]. The right side of Figure 1 shows a processor and FGPA on a single chip. As instructions stream into the chip, the FGPA secure component extracts the register information and uses the stream of registers to extract the hidden sequences, which are then checked. All the checking is performed by the FPGA itself.

To see how this works, consider a sample program as shown in Figure 2. The code on the left shows the output of standard compilation, before our register encoding is performed. The compiler approach replaces the standard register allocation algorithm with one that embeds keys. We use an even-numbered register to encode a 0 (zero) from the binary key, and an odd-numbered register to encode a 1 (one). This process continues until all the available "definition" registers of the basic block are assigned according to the key. The reverse process is performed in the FPGA.

In this paper, we consider two fundamental ways in which an embedded bit-sequence can be used. In the first method, the extracted bit sequence is compared against a pre-stored key in the FPGA. This, however, requires addressing the problem of key distribution. For systems in which such distribution or storage is

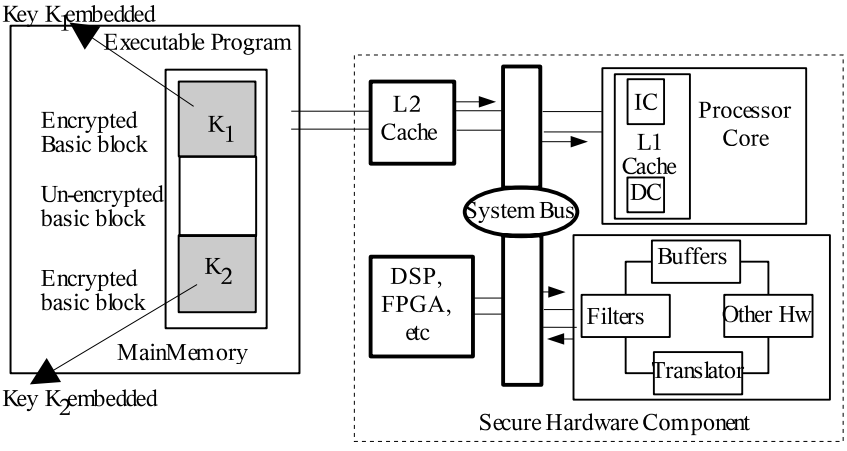


Fig. 1. System architecture

not possible, a simpler (albeit less secure) approach can be used. In this case, the bit-sequence is compared against a specific (say, the first) opcode in the basic block. The overall technique relies on the fact that, when the code is tampered with, the bit sequence will be disrupted with high probability.

Consider, for example, the second flavor mentioned above. In Figure 2, the opcode of the first instruction of a basic block yields a key of '11010000011000000000'. Since the first bit of the key is '1' the register assigned to the first instruction is 'R1'. From the use-def and def-use chain of the registers, the instructions that depend on the first instruction's register 'R1' are updated. In the second instruction, register 'R1' is once again chosen, as the bit value is '1' and the register is also redefined. An even register 'R0' is chosen for the fourth instruction from the available even registers, since the third bit of the key is '0'. This register's "use" locations are then updated. This process continues until all the registers are modified based on the key chosen. A modified program with a new register sequence is thus obtained. For the above example, Figure 2 shows the result.

How is tampered code detected? Consider an insertion of instructions. Clearly, any instruction that uses registers is likely to disrupt the bit sequence with high

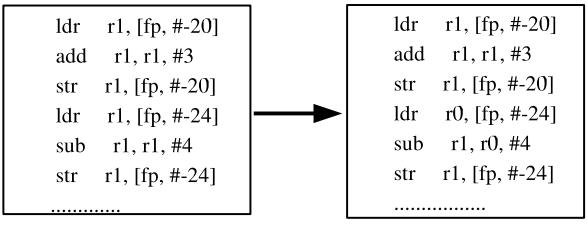


Fig. 2. Register encoding example with key 11010000011

probability, and will therefore be detected. Although not discussed here, it is easy to select opcodes that are at a fixed distance away, so that any insertion or deletion of instructions will result in a different opcode being checked, and therefore in failing the test. The security of a program can be increased by using private keys instead of opcodes of the instructions. This includes additional performance overhead of storing the keys in FPGA but can be tuned to meet the security objectives. Note that processors with caches will need to allow FPGA to probe the cache to fetch instructions of the encrypted basic blocks that span across multiple cache blocks.

3 Experimental Analysis

3.1 Experimental Framework

Our experimental framework uses the gcc cross-compiler targeting the ARM instruction set. The register allocation pass, the data flow analysis, loop unrolling and code generation pass of GCC were modified to incorporate the register-encoding scheme. The output of the compiler consists of the encrypted ARM executable and a description file for the FPGA.

The FPGA is simulated using the *Simplescalar* toolset [8]. The FPGA is placed between L1 cache and the main memory. It has access to the cache controller and can probe the cache. We assume that there is no L2 Cache. The FPGA is configured like a Virtex-II XC2V8000 component which is an ARM1020E type processor running at a chosen clock speed of 150 Mhz but with a processor clock rate of 300 MHz. The size of the on-board memory is 3MB with a memory access delay of 2 processor cycle (1 FPGA Cycle) and delay for comparing values is 2 processor cycles. The L1 cache is configured with 16 sets with the block size and associativity of cache being 32 and 64 respectively. The cache follows the Least Recently Used policy with a cache hit latency of one cycle. Main memory access latency is 15 and 2 cycles for the first and rest of the data respectively. The width of the FPGA operator is 8 bytes, nonsequential access latency being 10 cycles and sequential access latency being 10 cycles respectively. The FPGA validates each cache block that is being fetched from the main memory and placed into the L1 cache.

3.2 Benchmarks

Diverse benchmarks such as Susan (an image recognition package for Magnetic Resonance Images), Dijkstra (the well-known shortest path algorithm), Fast Fourier Transform, Rijndael (the AES encryption/decryption algorithm) and Patricia (a trie data structure) are used. We study the effect on overall system performance using basic blocks of varying lengths for purposes of encoding the bit sequences. The results are normalized to the performance of unencrypted case as shown in Figure 3 (left side). The overall performance for most of the benchmarks is within 78% of the unencrypted case. Benchmarks Patricia and

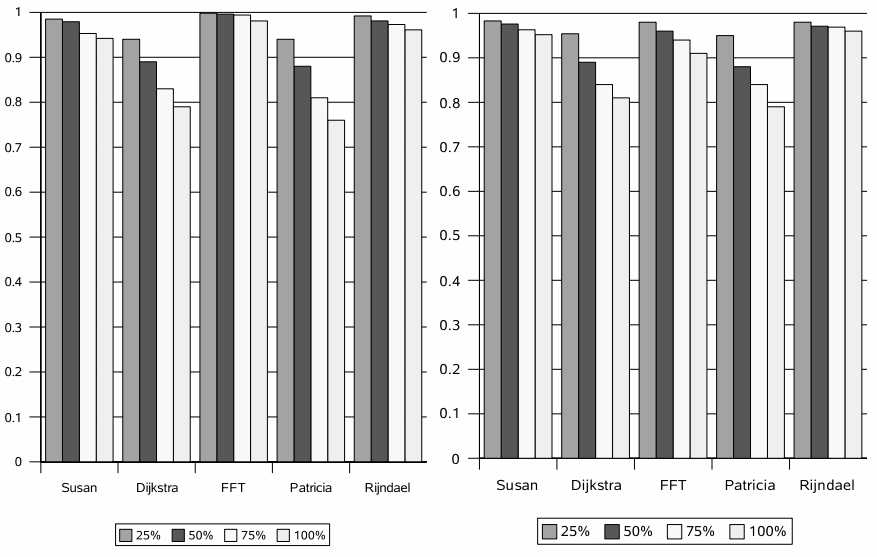


Fig. 3. Performance as a function of encryption of varying basic block lengths before(left side) and after(right side) performing loop unrolling on the benchmarks. (25%, 50%, 75% and 100% basic blocks are encrypted)

Dijkstra suffered a penalty of above 23%. This is because of instruction cache misses that occur in the FPGA that incur high cache miss penalty. Benchmarks Rijndael and FFT performed the best. Dijkstra suffered higher penalty due to large number of looping instructions being committed and it also runs longer.

The effect of loop unrolling on the performance of the benchmarks was studied by selecting the basic blocks with longer instruction count. The main purpose of performing loop unrolling on the benchmarks is to increase the code size thereby increasing the number of instructions in the basic block, so that keys of greater length can be embedded into them. Further loop unrolling reduces loop overheads such as index variable maintenance and control hazards in pipelines, increases the number of statements in basic block to optimize and also improves the effectiveness of other optimizations such as common-subexpression elimination, software pipelining, etc. But its disadvantage is that the unrolled version of the code is larger than the rolled version, thereby having a negative impact on the performance on effectiveness of instruction cache. The effect of loop unrolling on the benchmarks is shown in Figure 3 (right side). The optimization had little effect on Susan and Rijndael benchmarks. FFT suffered a 0.1% decrease in performance but performance of both computationally intensive Dijkstra and Patricia benchmarks is increased by nearly 3%. This shows that loop-unrolling provides negligible performance benefit, thereby suggesting that other compiler techniques will be needed to reduce the performance overhead. The security of the program is increased by loop unrolling, as the key length is increased since

more number of registers are now available to embed keys due to the increase in code size.

4 Conclusions and Future work

Because embedded processors constitute an overwhelming share (above 90%) of the processor market, and because embedded devices are easily accessible to hackers, security has now become an important objective in the design of embedded systems. Pure-software approaches may not stop the determined hacker and pure-hardware approaches require expensive custom hardware. A combined compiler-FPGA approach offers the advantages of both at a fraction of the cost, while also offering backward-compatibility. The purpose of this paper was to study the performance of this approach using some well-known benchmarks, and to examine the effect of some compiler optimizations.

References

1. Zambreno, J., Choudhary, A., Simha, R., Narahari, B., Memon, A.: SAFE-OPS: A Compiler/ Architecture Approach to Embedded Software Security. ACM Transactions on Embedded Computing.
2. Chang, H., Atallah, M.: Protecting software code by guards. Proceedings of the ACM Workshop on Security and Privacy in Digital Rights Management. (2000) 160-175.
3. Collberg, C., Thomborson, C., and Low, D.: A taxonomy of obfuscating transformations. Dept of Computer Science, University of Auckland. Tech. Rep. 148 (1997).
4. Daeman, J. Rijmen, V.: The block cipher Rijndael. In Smart Card Research and Applications. Lecture Notes in Computer Science, vol. 1820. Springer-Verlag (2000) 288-296.
5. Dallas Semiconductor: Features, advantages, and benefits of button-based security. available at http://www.ibutton.com.
6. Gobioff, H., Smith, S., Tygar, D., Yee, B.: Smart cards in hostile environments. Proceedings of 2nd USENIX Workshop on Electronic Commerce. (1996) 23-28.
7. Necula, G.: Proof-carrying code. Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languagers. (1997) 106-119.
8. Todd Austin, Eric Larson, Dan Ernst.: SimpleScalar: An Infrastructure for Computer System Modeling. IEEE Computer. (February 2002).
9. Simha, R., Narahari, B., Choudhary, A., Zambreno, J.: An overview of security-driven compilation. Workshop on New Horizons in Compiler Analysis and Optimizations, Bangalore, India, (Dec 2004).
10. Wurster, G., Oorschot, P., Somayaji, A.: A generic attack on checksumming-based software tamper resistance. IEEE Symp. Security and Privacy, Oakland, CA (2005).

A Secured Mobile Phone Based on Embedded Fingerprint Recognition Systems

Xinjian Chen, Jie Tian[1], Qi Su, Xin Yang, Feiyue Wang

The Center for Biometrics and Security Research, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation Chinese Academy of Science,
Graduate School of the Chinese Academy of Science,
P.O.Box 2728, Beijing, 100080, China
tian@doctor.com, jie.tian@mail.ia.ac.cn

Abstract. This paper presents a prototype design and implementation of secured mobile phones based on embedded fingerprint recognition systems. One is a front-end fingerprint capture sub-system and the other is a back-end fingerprint recognition system based on smart phones. The fingerprint capture subsystem is an external module which contains two parts: an ARM-Core processor LPC2106 and an Atmel Finger Sensor AT77C101B. The LPC2106 processor controls the AT77C101B sensor to capture the fingerprint image. In the fingerprint recognition system, a new fingerprint verification algorithm was implemented on internal hardwares. The performance of the proposed system, with 4.16% equal error rate (EER) was examined on Atmel fingerprints database. The average computation time on a 13 MHz CPU S1C33 (by Epson) is about 5.0 sec.

1 Introduction

In the vast majority of the current mobile phones, the only security measure to prevent unauthorized use of the mobile phones is the digital Personal Identification Number (PIN). As the amount of sensitive information stored in mobile phones increases, the need for better security increases as well. Personal identification using biometrics, i.e. personal physiological or behavioral traits, is one of the most promising "real" applications of pattern recognition [1]. Fingerprint sensing is often deemed the most "practical" biometric option for mobile phones because of low cost and accuracy.

The embedded system based on biometric authentication is applied as the platform of personal identification. Many new methods have been developed in the field of the hardware and software components designs. Ahyoung Sung etc. [2] provided a test data selection technique using a fault injection method for hardware and software interaction. El-Kharashi, M.W. etc. [3] proposed a flow for hardware/software co-design including coverification, profiling, partitioning, and co-synthesis.

[1] Corresponding author: Jie Tian, Telephone: 8610-62532105; Fax: 8610-62527995.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 549 – 553, 2005.

This paper presented a mobile security system based on fingerprint recognition. Section 2 presents the details of fingerprint recognition algorithm. Section 3 describes the proposed mobile security system. Experimental results with the proposed system are shown in Section 4. Section 5 concludes our work.

2 Fingerprint Recognition Algorithms

The block diagram of our fingerprint verification system is shown in Figure 1. It is composed of 4 stages: first, read the fingerprint image (128x128pixels); second, apply image filters based on frequency domain; third, extract minutiae from the fingerprints; and finally match the input template and enroll template. The proposed algorithm can be easily implemented on hardwares such as mobile phones.

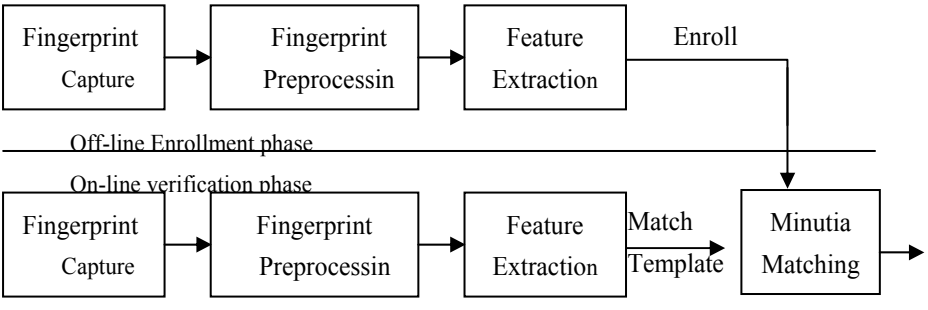


Fig. 1. The flowchart of fingerprint verification system

2.1 Frequency Domain Enhancement

Fingerprints exhibit a well defined local ridge orientation and ridge spacing in spatial domain. Fingerprint features are characterized by certain frequency components in the spectrum. We call this "clouds" as high energy annulus.

Using polar coordinates, the filter function [4] of the proposed algorithm is expressed as:

$$Filter(\rho,\theta) = F(energy_{(\rho,\theta)}) \bullet F_{radial}(\rho) \bullet F_{angle}(\theta) \quad (1)$$

When the ridge map of each filtered image is obtained, the next step is to combine them to generate an enhanced fingerprint and convert it into a binary image. The binarizing and minutiae extraction algorithm are given by Lin Hong et. al [5].

2.2 Fingerprint Matching Algorithm

The proposed matching algorithm has two steps. First, X.P.Luo's matching method [6] has been used to process the coarse match. A changeable bounding box is applied during the matching process which makes it more robust to nonlinear deformations

between the fingerprint images. Then X.J. Chen's method [7] is used to compute the similarity between the template and input fingerprints.

3 The Mobile Security System

The mobile security system is composed of two sub-systems, front-end fingerprint capture sub-system and back-end fingerprint recognition sub-system based on BIRD smart phone E868 [8]. The structure of the whole system is shown in Figure 2.

The fingerprint capture sub-system is an external module. The main parts of the sub-system are an ARM-Core processor LPC2106 and an Atmel Finger Sensor AT77C101B. The LPC2106 processor receives the commands from the smart phone via UART interface, controls the AT77C101B sensor to capture the fingerprint image, and sends it to the smart phone.

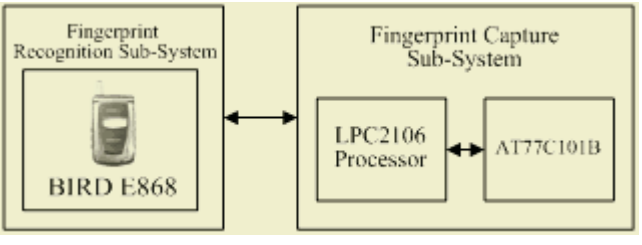


Fig. 2. The mobile security system block diagram

3.1 ARM Core Processor

The LPC2106 ARM-Core processor is manufactured by Philips Company [9]. It is a powerful processor with an ARMTDMI-S core, working at 60 MHz. It has a 128 kilobyte on-chip Flash and a 64 kilobyte on-chip Static RAM. Moreover, the processor is as compact as 7mm×7 mm in size. It can be operated in two low power working modes which make it suitable for mobile applications.

3.2 Atmel Finger Sensor

Atmel's AT77C101B FingerChip IC for fingerprint image capture combines detection and data conversion circuitry in a single rectangular CMOS die [10]. It captures the image of a fingerprint as the finger is swept vertically over the sensor window. It requires no external heat, light or radio source.

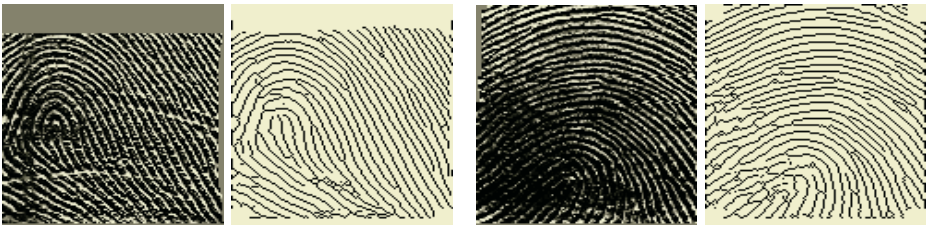
The AT77C101B sensor is divided into two sections: sensor array and data conversion. The sensor array comprises an array of 8 rows by 280 columns, giving 2240 temperature-sensitive pixels and one column selection circuit. The data conversion consists of an analog signal amplifier, two 4-bits Analog-to-Digital Converter (ADC) and a digital signal output circuit.

4 Experimental Results

We have developed a fingerprint mobile phone prototype based on BIRD smart phone E868 as shown in Figure 3. The central processing unit of the E868 is a 16-bit embedded processor S1C33. The processor is produced by Epson Company and its working frequency is 13 MHz.

Fig. 3. The prototype of fingerprint mobile phone

Figure 4 shows some examples of original captured fingerprints and enhanced thinned fingerprint. It can be seen from which a significant enhancement is seen in captured fingerprints.



(a)origin image (b) enhanced image of (a) (c)origin image(d) enhanced image of (c)

Fig. 4. Four examples of enhanced fingerprints

Reliability is a critical factor in biometrics system. To measure the accuracy of our embedded fingerprint verification system, we have developed a small fingerprints database using the Atmel fingerprint capture sensor. The database contains 300 fingerprints from 60 different fingers, 5 samples for every finger. The proposed algorithms were evaluated on this database according to FVC rules [11]. EER of the proposed algorithm is 4.16%. False non match rate (FNMR) equals 5.85% while false match rate (FMR) equals 1%. The average computational time.is calculated as well. It is 5.0 sec on a 13 MHz CPU S1C33 by Epson.

5 Conclusions

The mobile phone security system is quickly becoming an important area of technical development. This paper presented a mobile security system based on fingerprint recognition. The performance of the proposed algorithm on Atmel fingerprints database was evaluated with the EER at 4.16%. The average computation time on a 13 MHz CPU S1C33 by Epson is 5.0 sec. Future investigations will be focused on developing new methods to reduce the computation time.

Acknowledgments

This paper is supported by the Project of National Science Fund for Distinguished Young Scholars of China under Grant No. 60225008, the Key Project of National Natural Science Foundation of China under Grant No. 60332010, the Project for Young Scientists' Fund of National Natural Science Foundation of China under Grant No.60303022, Shandong Vehicular Electronics project (2004GG1104001), 973 Project from the Ministry of Science and Technology of China (2002CB312200), the National 863 Project of China (2004AA1Z2360), the Outstanding Young Scientist Research Fund (60125310) and the National Science Foundation of China (60334020).

References

[1] K. Uchida, Fingerprint-Based User-Friendly Interface and Pocket-PID for Mobile Authentication, ICPR2000, vol. 4, pp. 205–209, Barcelona, 2000.

[2] A. Sung; B. Choi, An Interaction Testing Technique between Hardware and Software in Embedded Systems, Ninth Asia-Pacific Software Engineering Conference, pp. 457-464, Dec. 2002,.

[3] M. W. El-Kharashi, M. H. El-Malaki, S. Hammad, etc., Towards Automating Hardware/Software Co-Design, 4th IEEE International Workshop on System-on-Chip for Real-Time Applications, pp.189-192, 19-21 Jul. 2004.

[4] X.J. Chen, J. Tian, Y.L. He, Low Quality Fingerprint Enhancement Algorithm Based on Filtering in Frequency Domain, *The 4th Chinese Conference on Biometric Recognition*, pp.145-150, 2003.

[5] L. Hong, Y. f. Wan, and A. Jain, Fingerprint image Enhancement: Algorithm and Performance Evaluation, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no.8, pp777-789, 1998.

[6] [X.P. Luo, J. Tian and Y. Wu, A Minutia Matching algorithm in Fingerprint Verification, 15th ICPR, Vol.4, P833-836, Barcelona, 2000

[7] X.J. Chen, J. Tian, X. Yang, A Matching Algorithm Based on Local Topologic Structure, Proceedings of ICIAR2004, LNCS 3211, pp. 360-367, 2004.

[8] BIRD DOEASY E868 Mobile Business Elite Introduce, http://doeasy.net.cn/index_2.htm.

[9] LPC2106/2105/2104 User Manual, Philips Semiconductors.

[10] FCD4B14 FingerChip Datasheet, Atmel Corporation.

[11] Biometric Systems Lab, Pattern Recognition and Image Processing Laboratory, Biometric Test Center, http://bias.csr.unibo.it/fvc2000/.

Connections in the World of International Terrorism

Yael Shahar

Senior Researcher, International Policy Institute for Counter-Terrorism,
Interdisciplinary Center Herzliya
webmaster@ict.org.il

Abstract. This paper gives an introduction to an ambitious database project currently running at the International Policy Institute for Counter-Terrorism. The project builds on an extensive database of terrorist incidents dating back to 1968 and adds both content and functionality to make these databases more accessible to researchers. The information on terrorist incidents has been supplemented with information on the organizations responsible, the individual perpetrators, front companies, and monetary sources. The content now being added to the database includes raw historical data from interviews and court documents. This information can provide valuable sociological data for researchers, including how perpetrators were recruited for each attack; their stated motivation; their socio-economic background; what influenced them to join the terrorist organization; etc.

1 Introduction

Terrorism is fast becoming the domain of a new kind of globalization. Terrorism is now the province of widespread networks built along the lines of multi-national corporations.

This paper gives an introduction to an ambitious database project at the International Policy Institute for Counter-Terrorism. The project builds on an extensive database of terrorist incidents dating back to 1968 and adds both content and functionality to make these databases more accessible to researchers. The records of terrorist incidents have been supplemented with information on the organizations responsible, the individual perpetrators, front companies, and monetary sources. The result is a far more comprehensive database, allowing researchers to track connections in the world of international terrorism.

The content now being added to the database can be divided into two main types: the first being raw historical data, and the second consisting of analysis of operational capabilities and relative risk posed by the different organizations in the database. Data of the first kind includes interviews and court documents. This information can provide valuable sociological data for researchers, including how perpetrators were recruited for each attack; their stated motivation; their socio-economic background; what influenced them to join the terrorist organization; etc.

1.1 Open Source Intelligence

Much of today's data-mining as applied to counter-terrorism revolves around the idea of building a "mega database" of information that will serve as the collective "brain"

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 554 – 560, 2005.

of counter-terrorism agencies.[1] The need for some sort of collective database was recognized in the aftermath of the September 11 attacks, when it was shown that significant facts had not been made accessible across agency borders or even within agencies.[2]

Most counter-terrorism agencies have also come to recognize the World Wide Web as a primary resource for open source intelligence.[3] Terrorists rely on the World Wide Web in the same ways as do corporations and businessmen. In fact the web is of even greater importance to terrorists, due to its anonymous nature and lack of censorship. Thus, it seems reasonable to assume that the World Wide Web will be integrated into a comprehensive database of terrorist incidents, individuals, and organizations, with the goal of forming the sort of World-wide Counter-Terrorism Brain that so many people envisage.

The discussion that follows is slanted toward the viewpoint of the counter-terrorism security and decision-makers—not professional data-miners or knowledge network people. This is because, regardless of the technical capabilities of our computational systems, the output must satisfy first and foremost, the professionals in the field.

This model assumes the following working assumptions:

- Terrorism can be "modeled" so that a knowledge network can be built to serve as the "brains" of the international counter-terrorism community.
- This network will need input data from diverse sectors, including security agencies, academia, the so-called "dark web," and the media.

Uses for the database include analyzing the "evolutionary cycle" of terrorist organizations, using some of the newer "organizational dynamics" models to map the growth and stasis point of the Global Jihad. Other researches might deal with how the Islamist organizations recruit new members, and how they inculcate loyalty among their followers. All these researches could deliver practical recommendations for security forces, intelligence agencies, and decision-makers.

2 The Terrorist Chronology Database

The ICT Incidents Database is one of the most comprehensive non-governmental resources on terrorist incidents in the world. Based on work done by Professor Ariel Merari at Tel-Aviv University between 1975 and 1995, the current database holds over 40,000 terrorist incidents, including foiled attacks and counter-terrorist operations.

---

[1] Travers, Russell E. Statement to the National Commission on Terrorist Attacks Upon The United States. Seventh public hearing of the National Commission on Terrorist Attacks Upon the United States. January 26, 2004

[2] See for example, *Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001*. p. 54.

[3] For two different viewpoints on this, see Channell, Ralph Norman. "Intelligence and the Department of Homeland Security" *Strategic Insights*, Volume I, Issue 6 (August 2002). Center for Contemporary Conflict; and Tenet, George J., Director of Central Intelligence. Testimony Before the Senate Select Committee on Intelligence. February 6, 2002.

The Incident Database is currently housed in a Microsoft Access database, though it will soon be ported to SQL Server. Information includes data on terrorist incidents, plus Background and Follow-up information for selected incidents. While not all-inclusive, this database provides a valuable resource for the study of the changing methods and targets of terrorists around the world. The database is updated on a regular basis, including continual follow-ups to older attacks as further information becomes available.

3 The Connections Database

The Terrorist Connections Project combines the terrorist chronology database described above with the results of decades of research into terrorist groups and individuals to create a database for tracking connections between individuals and organizations.

Upon completion, this database will be useful not only to researchers, but also to those tasked with fighting terrorism locally. It will enable governments, law enforcement and immigration agencies to track the connections between individuals and organizations by adding proprietary information available to them to the open-source information already entered into the database. The goal is a database that is not only user-friendly, but that will make real-time data available to those who need it most.

3.1 Content

The Connections Database contains information on terrorist groups, individuals, and front companies worldwide. This information is derived from open sources, including news media, academic articles, research analysis, and government documents. All of this background material is linked to the relevant organizations, individuals, and incidents via many-to-many link-tables. Table 1 shows the categories of data.

Table 1. Data Categories

Terrorist Incidents Failed Attacks Aborted by planners, Thwarted by security forces, Failed due to technical problems. Successful Attacks	*Terrorist Organizations* Profiles of Organizations, Profiles of Umbrella groups *Institutions* Front groups, Charity organizations, Straw companies.
Individuals Leaders, Ideologues, Supporters, Perpetrators, Technical personnel	*Counter-Terrorist measures* Laws Policy decisions Police Actions Intelligence Operations

The data is organized horizontally rather than vertically, with background information linked to relevant entities at a number of levels. Individuals are linked to incidents, to organizations and front groups, and to one another. The same is true of all the other entities in the database.

3.2 Example of Data on Individuals

The table "People" contains personal information on individuals, including aliases, date of birth, Country of origin, Religion, Current whereabouts, current status, etc. Below is an example of how this information looks to the user. Most of these fields are selected from relational tables, such as the table "Countries," and "Status." This not only makes data entry easier and less prone to error, but also facilitates detailed search, categorizations, and statistical analysis of the data.

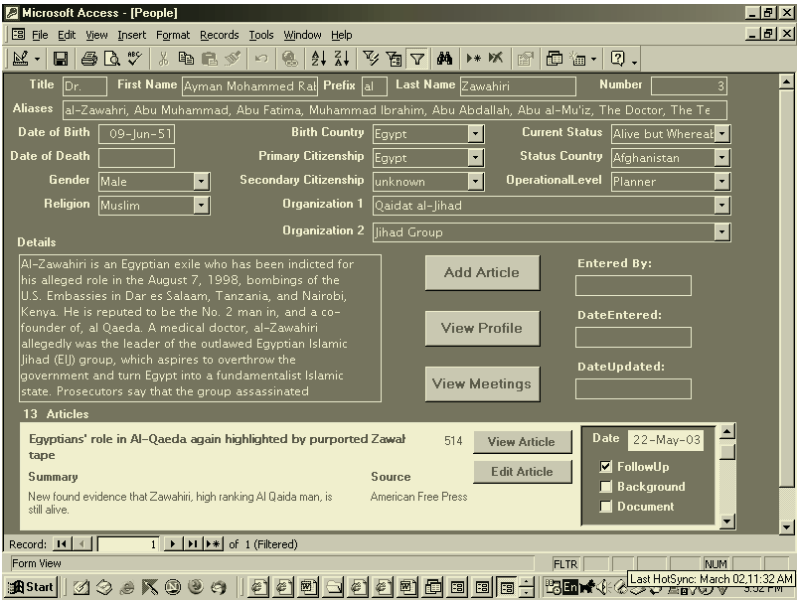


Fig. 1. Example profile screen

This screen also shows data from yet another many-to-many table—this one linking individuals to documents and news write-ups, with each document classified in relation to the individual, eg. Follow-up, Official Document, Background Article, etc.

While this information is useful to the researcher, where the relational database really comes into its own is in the connections between entities in the database.

3.3 Many-to-Many Connections

One of the more useful features of the connections database is the way in which entities are connected to one another. The relational database is particularly useful in

its ability to form many-to-many relationships, such as those between individuals in the database. This relationship connects individuals to one another and specifies the nature of the relationship, which itself comes from yet another table, called "RelationshipType." Examples of relationship are "Hierarchy of command," "Family relationship," "Money Trail," etc.

The Individuals are affiliated with the different organizations via a similar many-to-many table, the "PeopleOrgLinks" table. This relationship is expressed to the user via interface elements such as the one below, which allows the user to flip through virtual "index cards" with the different organizations and their members (Figure 11).

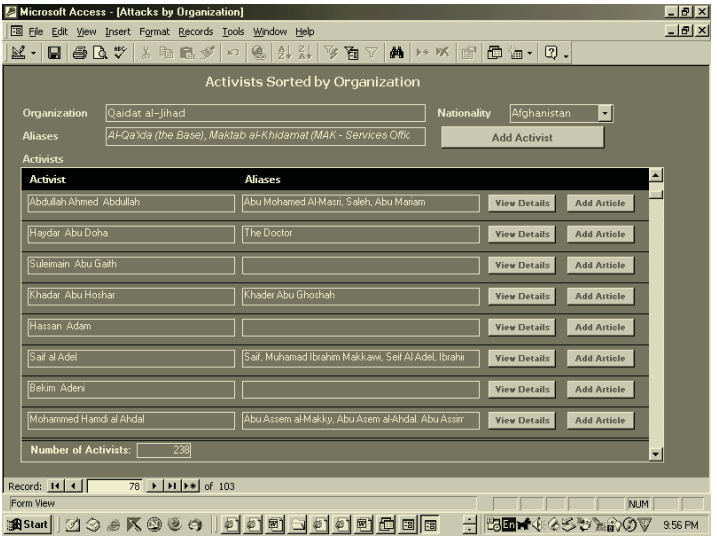


Fig. 2. Individuals sorted by organization

Because the database relies on many-to-many tables, each individual can have membership in an unlimited number of organizations. Ayman al-Zawahiri, for example, will appear listed under "Qaidat al-Jihad" and under "Egyptian Jihad Group."

A similar many-to-many table labeled "Meetings" connects individuals to one another and to a particular venue, thus allowing the user to few meetings attended by any particular individual, and look up information on other individuals present at a particular meeting.

3.4 User Interface

The user interface allows users to search for individuals from more than 100 organizations, and to access background articles and/or follow-up information on individuals. Other information related to individuals includes meetings between individuals and chronologies & travel records for individuals.

Users can go from profiles of terrorist organizations to connections between any entities in the database, whether individuals, organizations, or institutions.

4 Further Enhancements – International Online Database

In the past, ICT has often made use of the expertise of local institutes, for example, the IDSA in India, for profiles of country-specific terrorist organizations. With the formation of the International Counter-Terrorism Academic Community, it is now possible take this a step further, with data-entry responsibility for specific areas allocated to institutions with geographical, linguistic, or other relevancy to the specific area. For example, a reputable research institute in South America might be responsible for entering into the database information on organizations, incidents, and terrorist activists in South America. An institute like the IDSA in India would enter incidents, activists, front groups, etc. for the Kashmiri and Indian terrorist organizations, while a similar institute in Spain might enter data relevant to ETA. Ideally, we would want at least one institution or well-reputed scholar to cover each local conflict and its organizations. In some regions, it would be helpful to have multiple institutes entering data.

Each institute would be held to a high standard of accuracy for the information that it enters. Each datum entered into the database would require the citation of sources, and the entering institution would electronically "sign" each entry. This procedure would ensure that researchers using the data could track any particular entry back to its origin if additional information were needed. The fact that each institution would be "personally" held responsible for the information it enters would accomplish three main goals:

- Information would have a "father and mother;" meaning that the information would not be an isolated factoid afloat in a sea of similar bits of information.
- Information would benefit from the local expertise of participating institutes. Local institutions (or scholars) have expertise in specific areas, often far in excess of institutions elsewhere in the world. An institute that deals specifically with Kashmiri separatist groups would have access to far more accurate and in depth information on its local conflict than would a similar institute in, say, Canada.
- Information could be retrieved from local sources in the original languages, allowing access to a greater variety of sources and richer detail; translation would be performed by native speakers of these languages.
- Institutions would be held responsible for adhering to a high standard of accuracy by the fact that the institution's name would appear as source for information approved and entered by its staff.
- Mechanisms can be created to permit (or even require) vetting/proofreading of newly-entered information before it is made available online. These mechanisms can encompass varying levels of "signing authority" for different participating institutions.

5 Technical Issues

The infrastructure for the database would include a method by which each institute would be given a particular "electronic signature" which accompanies each datum entered by that institute. This feature not only serves to ensure accountability, but is also a guarantor of authenticity, to ensure that bogus information is not entered into the database by non-authorized users.

Other security measures would need to be implemented to ensure that the system is not penetrated from the outside. Even though the information is not technically "sensitive," since it will be derived from open sources, the aggregate of information is worth far more that the sum of its component data, and thus merits appropriate security to ensure against theft or corruption of data.

One of the major failings of previous terrorist-incident and law-enforcement databases lay in the difficulties in answering specific questions and in correlating data from different categories. These failing were highlighted by documents relating to the failure of U.S. intelligence agencies prior to the September 11 Attacks.[4] This is perhaps the most important factor of all in designing a database of this kind: a database is only useful if end-users can easily retrieve answers to the questions that interest them.

[4] Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001. p. 54.

Forecasting Terrorism: Indicators and Proven Analytic Techniques

Sundri K. Khalsa

PO BOX 5124, Alameda, CA, 94501
SundriKK@hotmail.com

Abstract. This forecasting methodology identifies 68 indicators of terrorism and employs proven analytic techniques in a systematic process that safeguards against 36 of the 42 common warning pitfalls that experts have identified throughout history. The complete version of this research provides: 1) a step-by-step explanation of how to forecast terrorism, 2) an evaluation of the forecasting system against the 42 common warning pitfalls that have caused warning failures in the past, and 3) recommendations for implementation. The associated CD has the website interface to this methodology to forecast terrorist attacks. This methodology could be applied to any intelligence topic (not just terrorism) by simply changing the list of indicators. The complete version of this research is available in *Forecasting Terrorism: Indicators and Proven Analytic Techniques*, Scarecrow Press, Inc., ISBN 0-8108-5017-6.

1 Introduction: Correcting Misconceptions

Important lessons have arisen from the study of intelligence warning failures, but some common misconceptions have prevented the Intelligence Community from recognizing and incorporating these lessons. **Analysis, Rather Than Collection, Is the Most Effective Way to Improve Warning.** The focus to improve warning normally turns to intelligence collection, rather than analysis. That trend continues after September 11th. However, warning failures are rarely due to inadequate intelligence collection, are more frequently due to weak analysis, and are most often due to decision makers ignoring intelligence. Decision makers, however, ignore intelligence largely because analytical product is weak. **Hiring Smart People Does Not Necessarily Lead to Good Analysis.** Studies show that, "frequently groups of smart, well-motivated people . . . agree . . . on the wrong solution. . . . They didn't fail because they were stupid. They failed because *they followed a poor process in arriving at their decisions*." **A Systematic Process Is the Most Effective Way to Facilitate Good Analysis.** The nonstructured approach has become the norm in the Intelligence Community. A key misunderstanding in the debate over intuition versus structured technique is that an analyst must choose either intuition or structured technique. In fact, both intuition and structured technique can be used together in a systematic process. "Anything that is qualitative can be assigned meaningful numerical values. These values can then be manipulated to help us achieve greater insight into the meaning of the data and to help us examine specific hypotheses." It is

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 561 – 566, 2005.

not only possible to combine intuition and structure in a system, research shows the combination is more effective than intuition alone. "Doing something systematic is better in almost all cases than seat-of-the-pants prediction." Moreover, decision makers have called on the Intelligence Community to use methodology. "The Rumsfeld Commission noted that, '. . . an expansion of the methodology used by the IC [Intelligence Community] is needed.' . . . These techniques are the heavy lifting of analysis, but this is what analysts are supposed to do. If decision makers only needed talking heads, those are readily available elsewhere."

2 How to Forecast Terrorism: Abbreviated Step-by-Step Explanation of the Methodology

The explanation of this methodology begins at the lowest level of *indicators* and then builds up to the big picture of countries within a region. The forecasting assessments of this methodology are maintained on a website display, which is available on the associated CD. Figure 1 shows a breakdown of the 3 primary types of warning picture views from the web homepage: 1) country list view, 2) target list view, and 3) indicator list view.

The methodology consists of 23 tasks and 6 phases of warning analysis, shown in Table 1 (Overview of Tasks in Methodology). The 23 tasks include 14 daily tasks, 3 monthly tasks, 4 annual tasks, and 2 as-required tasks. The 14 daily tasks can be completed in 1 day because tasks have been automated wherever possible. Three types of analysts are required to operate this methodology: *Raw Reporting Profilers*, *Indicator Specialists*, and *Senior Warning Officers*.

Indicators are the building blocks of this warning system. Indicators are "those [collectable] things that would have to happen and those that would likely happen as [a] scenario unfolded." For a terrorist attack, those would be things like: terrorist travel, weapons movement, terrorist training, target surveillance, and tests of security. This project research has identified 68 indicators of terrorism encompassing terrorist intentions, terrorist capability, and target vulnerability, which are the three components of risk.

Each potential terrorist target is evaluated in a webpage hypothesis matrix using the 68 indicators of terrorism. The indicators are updated near-real-time with incoming raw intelligence reports/evidence. Analysts use *Indicator Key Questions, Answers, & Evidence Logs* (shown in Figure 2) to rate the status of the indicators on a five level scale of: 1) *Critical (~90%)*, about 90 percent probability, color coded red on the website, 2) *Significant (~70%)*, color coded orange, 3) *Minor (~30%)*, color coded yellow, 4) *Slight (~10%)*, color coded gray, and 5) *Unknown (or ~50%)*, color coded black.

3 Conclusion

Rather than face the War on Terrorism with the traditional intuition-dominated approach, this methodology offers a systematic forecasting tool that:

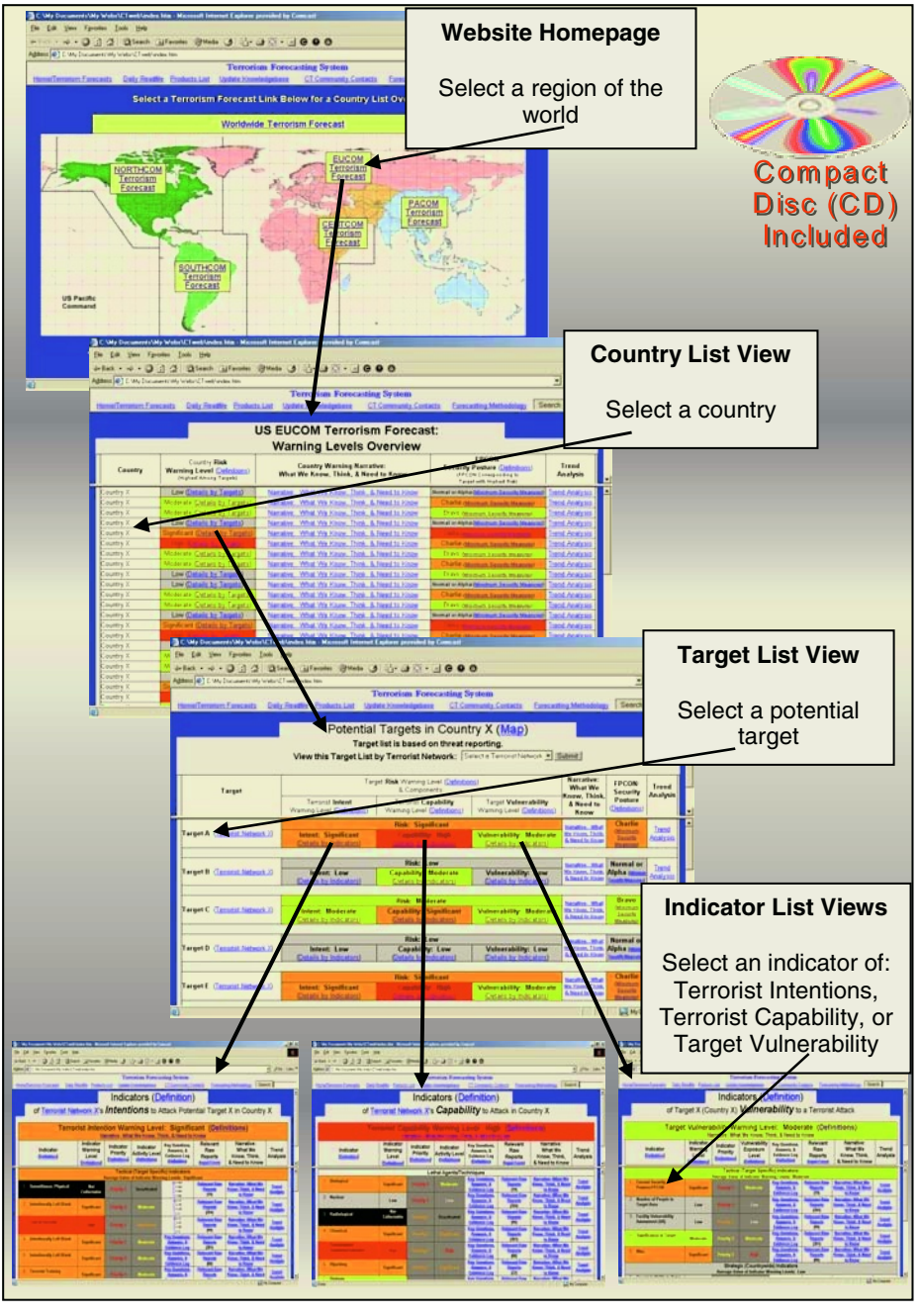


Fig. 1. The 3 Primary Warning Picture Views. Map from "Unified Command Plan 2002 (UCP 02)," *Proposed EUCOM Area of Responsibility Change (AOR),* 1 October 2002, *United States European Command*, www.eucom.mil/AOR/index.htm (12 June 2002)

Table 1. Overview of Tasks in Methodology

Task Set	Phase I: Define/Validate Key Elements of the Intelligence Problem (Using Indicators)
1	Identify/Validate Indicators **(Annually)**
2	Determine/Validate Priorities of Indicators **(Annually)**
3	Develop/Validate Key Question Sets on Each Indicator **(Annually)**
4	Determine/Validate Priorities of Questions in Key Question Sets **(Annually)**
Task Set	Phase II: Consolidate Information (Using Master Database)
5	Intelligence Community Master Database Receives *All* Raw Intelligence Reports from Intelligence Collectors **(Daily)**
Task Set	Phase III: Sort Information (Using Hypothesis Matrices)
6	Enter *All Terrorism Related* Raw Intelligence Reports into Terrorism Forecasting Database Under Appropriate Indicators, Key Questions, Targets, Countries, Terrorist Groups, and Other Data Profile Elements **(Daily)**
7	Terrorism Forecasting Database Creates *Potential Target Hypothesis Matrices* with Raw Intelligence Reports Filed by Indicators **(Daily)**
8	Terrorism Forecasting Database Feeds Raw Intelligence Reports into Appropriate *Indicator Key Questions, Answers, & Evidence Logs* within the Hypothesis Matrices **(Daily)**
9	Assess Raw Intelligence Reports' *Information Validity* **(Daily)**
Task Set	Phase IV: Draw Conclusions (Using Intuitive and Structured Techniques)
10	Assess *Indicator Warning Levels* **(Daily)**
11	Assess *Terrorist Intention Warning Level* **(Daily)**
12	Assess *Terrorist Capability Warning Level* **(Daily)**
13	Assess *Target Vulnerability Warning Level* **(Daily)**
14	Assess *Target Risk Warning Level* **(Daily)**
15	Assess *Country Risk Warning Level* **(Daily)**
16	**Update/Study Trend Analysis of Indicator Warning Levels** (Monthly)
17	Update/Study Trend Analysis of Target Risk Warning Levels **(Monthly)**
18	Update/Study Trend Analysis of Country Risk Warning Levels **(Monthly)**
Task Set	Phase V: Focus Collectors On Intelligence Gaps to Refine/Update Conclusions (Using Narratives that Describe What We Know, Think, and Need to Know)
19	Write/Update *Indicator Warning Narrative: What We Know, Think, & Need to Know* **(Daily)**
20	Write/Update Executive Summary for *Target Warning Narrative: What We Know, Think, & Need to Know* **(Daily)**
21	Write/Update Executive Summary for *Country Warning Narrative: What We Know, Think, & Need to Know* **(Daily)**
Task Set	Phase VI: Communicate Conclusions/Give Warning (Using Templates in Website)
22	**Brief Decision maker with Website Templates** (As Required)
23	Rebrief Decision maker with New Evidence in Templates **(As Required)**

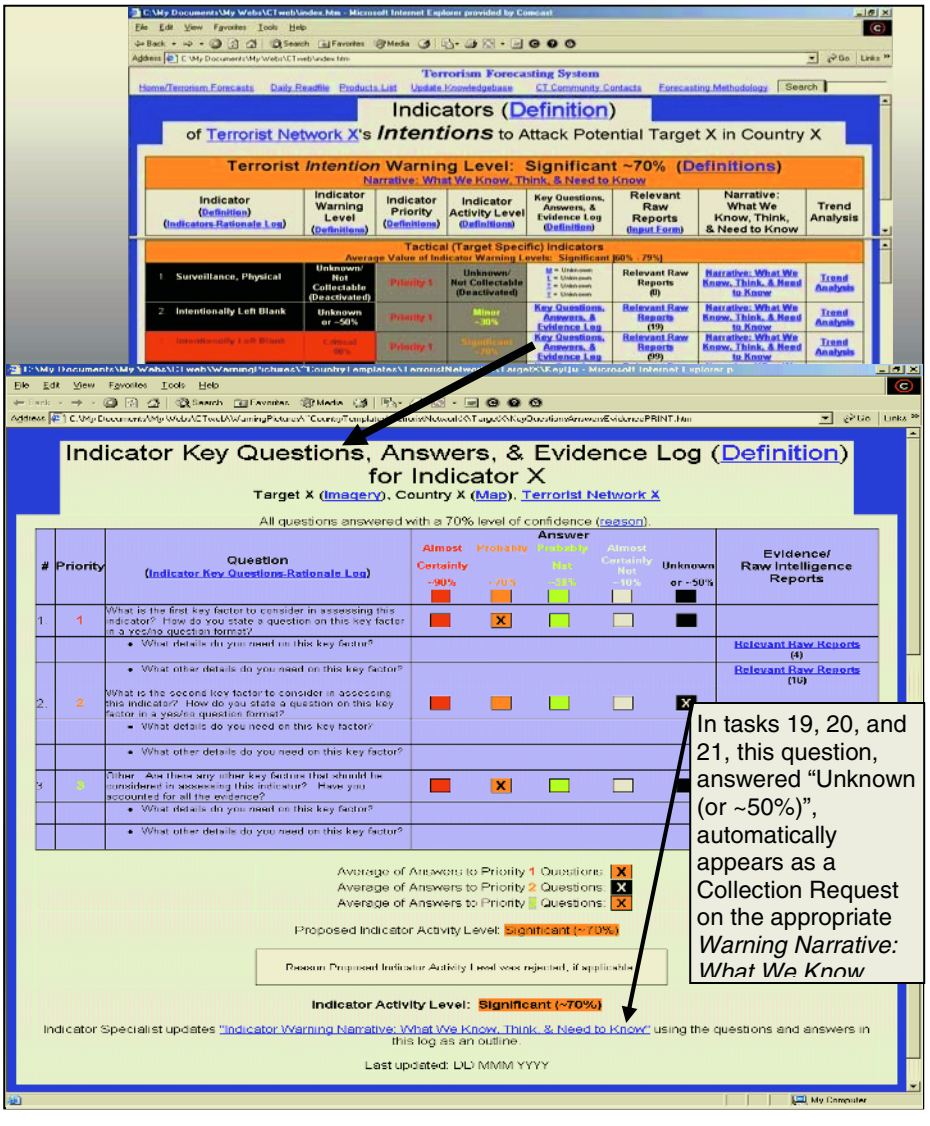


Fig. 2. Indicator Key Questions, Answers, & Evidence Log (in Hypothesis Matrix)

- Guards against nearly 81 percent of common warning pitfalls, and ultimately, improves the terrorism warning process.
- Coordinates analysts in a comprehensive, systematic effort.
- Automates many proven analytic techniques into a comprehensive system, which is near-real-time, saves time, saves manpower, and ensures accuracy in calculations and consistency in necessary, recurring judgments.
- Enables collection to feed analysis, and analysis to also feed collection, which is the way the intelligence cycle is supposed to work.

- Fuses interagency intelligence into a meaningful warning picture while still allowing for compartmenting necessary to protect sensitive sources and methods.
- Provides a continuously updated analysis of competing hypotheses for each potential terrorist target based on the status of the 68 indicators of terrorism.
- Is the first target specific terrorism warning system; thus far, systems have only been country specific.
- Is the first terrorism warning system with built in trend analysis.
- Combines threat (adversary intentions and adversary capability) with friendly vulnerability to determine risk and provide a truer risk assessment than typical intelligence analysis.
- Includes a CD that is the tool to implement this terrorism forecasting system.

Officials in the FBI and the Defense Intelligence Agency (DIA) characterized this terrorism forecasting system as "light-years ahead," "the bedrock for the evolving approach to terrorism analysis," and an "unprecedented forecasting model."

References

1. Ephraim Kam, Surprise Attack: The Victim's Perspective (Cambridge, MA: Harvard University Press, 1988), 53.
2. A source, mid-level intelligence professional at a national intelligence organization, who wishes to remain anonymous, interview by author, 10 July 2002.
3. Ronald D. Garst, "Fundamentals of Intelligence Analysis," in Intelligence Analysis ANA 630, no. 1, ed. Joint Military Intelligence College (Washington, D.C.: Joint Military Intelligence College, 2000): 7.
4. Hans Heymann Jr., "The Intelligence—Policy Relationship," in Intelligence Analysis ANA 630, no. 1 ed. Joint Military Intelligence College (Washington, D.C.: Joint Military Intelligence College, 2000): 55.
5. J. Edward Russo and Paul J. H. Schoemaker, Decision Traps: The Ten Barriers to Brilliant Decision-Making and How to Overcome Them (New York: Rockefeller Center, 1989), 146.
6. Robert D. Folker, Jr., Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods, Occasional Paper, no. 7 (Washington, D.C.: Joint Military Intelligence College, 2000), 1.
7. William M.K. Trochim, "The Qualitative Debate," Cornell University: Research Methods Knowledge Base 2002, trochim.human.cornell.edu/kb/qualdeb.htm (31 May 2002).
8. Russo, 136.
9. Donald Rumsfeld, press conference, quoted in Mary O. McCarthy, "The Mission to Warn: Disaster Looms," Defense Intelligence Journal 7 no. 2 (Fall 1998): 21.
10. James J. McDevitt, "Summary of Indicator-Based-Methodology," unpublished handout, n.p., n.d. provided in January 2002 at the Joint Military Intelligence College.

Forecasting Terrorist Groups' Warfare: 'Conventional' to CBRN

Joshua Sinai

ANSER (Analytic Services)
joshua.sinai@verizon.net

To assess the type or spectrum of warfare that a terrorist group is likely to conduct in its operations, this paper proposes an indications and warning (I&W) methodology to comprehensively and systematically map all the significant indicators (as well as sub-indicators and observables) that need to be examined.

A key proposition is that most of the current threat assessments fail to adequately assess the likelihood and magnitude of the types of threats posed by contemporary terrorism because of three fundamental problems in their approaches. First, many threat assessments focus exclusively on two types of warfare, "conventional" or CBRN, but not on three types of warfare that are more prevalent in terms of their actual or potential impact: conventional low impact (CLI), conventional high impact (CHI), and CBRN.[1]

Second, most of the current approaches do not adequately differentiate between the characteristics that define the types of warfare that terrorist groups are likely to resort to achieve their objectives, particularly in terms of the leadership, motivation, strategy, and other factors as well as accelerators, triggers, and the numerous constraints involved in the transition by terrorist groups from CLI to CHI and to CBRN warfare.

Third, most of the current approaches fail to address the crucial factor of the spectrum of disincentives and constraints that are likely to deter most terrorist groups away from CBRN warfare, particularly when they can resort to conventional explosives which have become more lethal and "catastrophic" in their impact. As a result, most of the current approaches tend to lump all terrorist groups currently operating on the international scene as potential CBRN actors, without an adequate understanding of which terrorist groups are likely or unlikely to transition not only from CLI to CHI, but from CHI to CBRN warfare.

To forecast the type of warfare that a group is likely to choose, we must first attain an understanding of terrorism. Terrorism is a form of physical and psychological warfare by non-state actors against their more powerful state adversaries. In the terrorists' decision-making calculus, a localized incident is intended not only to cause casualties and physical damage, but to spread fear and anxiety throughout the larger society. Terrorists generally calibrate their warfare to maximize these objectives.

[1] In this approach, CLI refers to 'low impact' attacks involving relatively 'few' casualties or physical damage, such as detonating explosives or shootings; CHI refers to the use of conventional means to inflict massive casualties, such as 9/11 type attacks; while CBRN refers to the use of chemical, biological, radiological, nuclear weapons and devices.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 567 – 570, 2005.

In this approach, a trichotomous outcome (CLI, CHI, or CBRN warfare) is hypothesized to predict the types of warfare most likely to be engaged in by terrorist groups. The pre-incident activities by terrorist groups are the **independent variables (X)**, while the trichotomous outcome is the **dependent variable (Y)**. As discussed later on, the independent variables are related to each other, so they need to be correlated in combination. The premise is that if one can collect this type of intelligence information about terrorist groups it will be possible to attain a sufficient predictive accuracy, while recognizing that there will always remain a certain percent of "real world" uncertainty built into any I&W forecasting system.

This analysis is based on 31 I&W indicator categories generally considered to influence and shape a terrorist group's warfare proclivity.

The Pre-Incident's Four Phases of Terrorism's 31 I&W Indicator Categories

Group Formation	Plan	Develop	Execute
(1) Societal Conditions	(14) Decisive Meeting	(17) Acquisition	(22) Tactics
(2) Radical Subcultures	(15) Recruitment	(18) Development/ Production	(23) Security
(3) Types of groups	(16) Training	(19) Testing	(24) Communication
(4) Leadership		(20) Weaponization	(25) Logistics
(5) Motivation		(21) Storage Facilities	(26) Surveillance
(6) Strategy			(27) Targeting
(7) Agenda			(28) Accelerators
(8) Front Organizations (political, economic, religious/charity)			(29) Triggers
(9) Organization			(30) Internal Hurdles
(10) Funding			(31) External Hurdles
(11) Constituency			
(12) Foreign Group Linkages			
(13) State Sponsor			

Developing an accurate, timely, and actionable I&W indicators system to forecast a terrorist group's warfare proclivity involves formulating hypotheses, which are broad explanatory statements that generate factors, which are more specific indicators suggesting a type of terrorist warfare proclivity. Indicators need to be considered in combination because no single factor is likely to indicate a particular warfare proclivity.

To correlate the pre-incident's processes, paths and links involved in the likelihood that a terrorist group will resort to CLI/CHI/or CBRN warfare, this analysis groups the I&W indicators into four phases (group formation, plan, develop, and execution). These four phases are distinguished for analytical purposes and, in reality, some of the indicators or phases could be carried out contemporaneously or bypassed altogether.

Moreover, within these four phases, the I&W indicators should be considered in combination because no single factor has sufficient independent explanatory value.

- These four phases of the pre-incident process can be further broken down into ten levels of analysis:

- First, which **geographic areas/states** require monitoring for precipitating **societal conditions (#1)** and the **proliferation of radical subcultures (#2)**, from which terrorist groups emerge?
- Second, which **particular terrorist groups (#3)** are inclined, in terms of their **leadership (#4), motivation (#5), strategy (#6)**, and **agenda ((#7)** to transition from conventional to CBRN warfare, and why would they would choose CBRN "catastrophic" warfare when "conventional" warfare might be sufficiently lethal?
- Third, what is the nature of the terrorist group's core **constituency (#11)** and how would it react to mass casualty/CBRN-type attacks as opposed to "low casualty" attacks with conventional weapons (although certain CBRN weapons might result in few casualties but mass panic throughout society)?
- Fourth, what kinds of **accelerators (#14) and triggers (#15)** are likely to drive terrorist leaders to plan a "high casualty" as opposed to a "low casualty" type attack?
- Fifth, what kinds of **organizational (#8), funding (#10), recruitment (#17), acquisition (#19), development (#20), and logistical (#27)** capabilities will a terrorist group need to attain the operational capability to execute either a "conventional" or CBRN attack?
- Sixth, in terms of the terrorist group's targeting options, what **vulnerable 'key targeting (#29) points'** are they most interested in attacking and are there certain key anniversaries around which they are likely to attack?
- Seventh, how does a terrorist group's decision to attack a particular target affect their **choice of CBRN weapons, devices and delivery systems**?
- Eighth, can the terrorist group embark on terrorist warfare **on its own** or with the **support of a state sponsor (#13)**?
- Ninth, what **internal (#30) and external (#31) hurdles** must terrorist groups overcome in order to execute a CBRN attack?
- Finally, what can the targeted adversary do during the pre-incident process to **harden its defenses** against the spectrum of terrorist attacks?

Applying these 31 indicator categories to currently operating or emerging terrorist groups will reveal whether a group is planning an attack, its motivation and strategy, the type (or types) of weapons it plans to use (particularly "conventional" or CBRN), and its likely targeting. In such a way, this methodology enables an analyst to correlate along the pre-incident's four phases a terrorist group's internal factors (such as the nature of its leadership, motivation and strategy, organization, funding, recruitment, training, front organizations, constituency, logistical network, surveillance of potential targets), with external factors (such as linkages with foreign groups, state sponsors, and foreign suppliers), as well as potential accelerators and triggers (such as access on the grey or black markets to needed components or devices, or a dramatic event that would precipitate a group to take drastic action), and a group's capability to overcome various internal and external hurdles (such as defections, breakdowns in

security, testing failures or accidents with devices, or monitoring or penetration of a group by external intelligence or counterterrorism organizations), in order to ascertain a group's attack potential. Thus, if these indicator categories and their sub-indicators and observables could be correlated—recognizing that some indicators are more significant and have higher quantifiable weighting properties than others—such analysis might indicate increasing threat possibilities, including the possible resort to conventional or single or multiple CBRN weapons and devices, and their likely targeting.

This analysis is intended to highlight some of the internal and external factors, requirements and hurdles that need to be considered in assessing a terrorist group's current and future development status and operational capability to conduct CBRN warfare. Correlating these internal and external factors and hurdles would make it possible to assess which terrorist groups and state sponsors are likely to embark on CBRN warfare, the types of adaptations and changes they would require to transition to such warfare, the types of weapons and targeting they are likely to pursue (including the possible resort to single or multiple CBRN weapons and devices), the timelines for such attacks, and vulnerabilities that could be exploited by foreign intelligence and counterterrorism agencies to constrain terrorist groups—and, when applicable, state sponsors—from embarking on such warfare.

Hopefully, such a conceptual approach will make it possible for the counterterrorism community, whether policy makers, warfighters, or analysts to efficiently calibrate their resources to intervene at the earliest possible phases to influence, preempt, deter, prevent and defeat terrorist actions, whether CLI, CHI, or CBRN.

The Qualitative Challenge of Insurgency Informatics

Scott Tousley

Logos Technologies, 3811 North Fairfax Street., Suite 100, Arlington, VA 22203
stousley@logostech.net

Abstract. Terrorism and insurgency analysis depends critically on qualitative understanding, to ensure that quantitative work is on target. Key concerns include: qualitative analysis distorted by policy expectation, obstructed by compartmentalization and classification, and mistaken when drawing from the wrong history and experience. Open presentation of analytic processes allows more effective judgment of the credibility of analysis and interpretations. Data itself will not be the problem, but rather the constructs and methodologies used to assemble the data into conclusions. These challenges are at the heart of terrorist informatics.

1 Introduction

The American national security complex is improving analysis of the terrorist and Islamist insurgent environment, especially through use and visualization of large sets of quantitative data. This analysis depends critically on qualitative understanding of data context, far more on qualitative than quantitative analysis of information. We must focus on qualitative understanding so that data analysis is not misdirected or wrongly suggests excessive precision. What lessons should informatics draw from the qualitative judgments of past experts like Bernard Fall and Mao Zedong?

Social science works like Fukayama's "The Great Disruption," Sageman's "Understanding Terror Networks," and Roy's "The Globalization of Islam," all build on qualitative analysis. Studying terrorist and insurgent networks operating within social networks, what are the social network structures and how do we gauge insurgent and defensive strengths in parent populations and complex demographics? We must understand dynamic network interactions between the Iraqi insurgency, Al Qaeda and the global Salafi Jihad, and American presence in the Middle East and prosecution of the "War of Ideas." Qualitative analysis enables understanding of the dynamic character of these interactions, to understand the intelligence we draw from our data.

2 Nature of the Environment

Spiritual insurgency is a key driver influencing the post cold war world[1]. Fueled by modernization, it interacts with the transformation of culture and society and contributes to conflicts. At the psychological level, modernization involves shifting values,

[1] Steven Metz, "The Future of Insurgency," Strategic Studies Institute, December 1993, p. 24.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 571 – 578, 2005.

attitudes and expectations, all inherently qualitative measurement challenges. A business case study approach can be used to understand commercial insurgencies like the Columbian narco-insurgency or international criminal organizations. However, "….little evidence that U.S. policymakers and strategists fully grasp the motives, fears, and hopes driving emerging forms of (spiritual) insurgency…Americans are particularly likely to fail against insurgents driven by intangible motives like justice, dignity and the attainment of personal meaning and identity"[2]. Intangible motives require qualitative measures.

Both France and the U.S. discovered too late their weakness against the political appeal of Vietnamese independence offered by Ho and Giap[3]. The U.S. today finds it similarly difficult to judge the strength and nature of the Iraqi insurgency, to establish calm against Islamist influences, Shia/Sunni rivalries and tribal conflicts. Stephen Metz has discussed the Iraq insurgency and historical examples from Israel and South Africa, highlighting their qualitative nature.[4] Insurgents were able to "psychologically cast or shape the conflict to be widely perceived as a liberation struggle…the essence of any insurgency and its most decisive battle space is the psychological." In the 1960s, insurgency was referred to as armed theater, with protagonists on stage sending messages to wider audiences. Insurgency is about perceptions, beliefs, expectations, legitimacy, and will…altering the psychological factors that are most relevant." This nature requires more qualitative estimate than quantitative measurement.

3 American Operations

A recent report discusses networked force advantages in contingencies other than war: finding, distinguishing, and destroying resistance; pursuing distributed objectives, controlling areas, and seizing critical points; evacuating and protecting noncombatants; eliminating threats and restoring order; and minimizing damage and casualties.[5] These strongly resemble local unit missions of the Marine Corps' Vietnam era Combined Action Program, used successfully to[6]: destroy local communist infrastructure; protect public security and maintaining law and order; organize local intelligence nets; participate in civic action and anti-insurgent propaganda; motivate militia pride, patriotism and aggressiveness; and train militia. Evaluating these missions requires largely qualitative judgments and measurements: degree of insurgent (human) infrastructure destroyed, local population perceptions of security, law & order, propaganda effectiveness, militia morale and likely long-term effectiveness, etc.

---

[2] Metz, "The Future of Insurgency," and "Counterinsurgency: Strategy and the Phoenix of American Capability," Strategic Studies Institute, February 1995, pp. 26 and 31, respectively.

[3] John Shy & Thomas Collier, "Revolutionary War," 1986, p. 848.

[4] Metz, Killebrew, Linn and Sepp, "Relearning Counterinsurgency," AEI presentation, January 10, 2005.

[5] Gompert, Pung, O'Brien and Peterson, "Stretching the Network," RAND, 2004, p. 23.

[6] Keith Kopets, "The Combined Action Program: Vietnam," Military Review, July-August 2002, p. 14, and Frank Pelli, "Insurgency, Counterinsurgency, and the Marines in Vietnam," USMC Command & Staff College, 1990.

French insurgency authority Roger Trinquier cited an American officer in identifying the two different objectives of counter-guerrilla operations, destruction of guerrilla forces and eradication of their influence on the population.[7] While numbers can describe the first objective (note also doubtful Vietnam-era body counts and a current version, the size of the Iraqi insurgency), judgments of influence are inherently uncertain, difficult and qualitative. Trinquier identified components of "internal warfare" as police operations, propaganda efforts, social programs and military programs.[8] This is a challenge in Iraq because each is a different responsibility: Iraqi and American military police and the FBI; the Defense and State Departments (complicated by commercial/embedded media); military civil affairs units and various US government and international organizations; and military efforts of intelligence organizations, special forces and military services. Consistent evaluation across different components and organizations is a significant challenge; each will judge success on different qualitative metrics and agreement will not come easily. Decision makers may assess parts of the situation correctly but fail to understand context or anticipate reactions to actions (like "occupation"). Systemic, qualitative civil-military data and assessment is needed to gauge effectiveness. The U.S. intelligence community has made complex civil-military operations harder in recent years by focusing on force protection at the expense of mission execution, diverting human intelligence collection away from situational understanding towards threat identification.[9] In other words, it favored a quantitative focus on the specific over the qualitative understanding of the general.

Gentry concluded U.S. institutional characteristics have significant and often adverse implications for effectiveness of U.S. forces in complex civil-military operations.[10] He cited general stages of civil military operations as problem identification, mission determination, planning and execution, and the processes of measurement of effectiveness and of learning, similar to the military "OODA loop" (for Observe, Orient, Decide and Act). To make key decisions following our orientation, we must understand "the prevailing authoritative-social structures (governmental, tribal, and religious) and personalities in various localities; make a cultural 'story board;' continuously assess the tribal, rivalries, jealousies and ethno-religious fault lines affecting the local communities."[11] These are inherently qualitative processes. Unfortunately, we tend to focus on the physical or kinetic level of war, to the virtual exclusion of the more powerful mental & moral levels; the easier quantitative over the more powerful qualitative. And what works at the physical or kinetic level often works against us at the mental & moral levels. So if we allow ourselves to be drawn to the easier quantitative over the harder qualitative judgments, our decisions grow riskier.

---

[7] Roger Trinquier, "A French View of Counterinsurgency," London, 1964, p65.

[8] Trinquier, p43.

[9] John A. Gentry, "Complex Civil-Military Operations," Naval War College Review, Autumn 2000, p. 61.

[10] Gentry, p. 65.

[11] G. I. Wilson, Greg Wilcox and Chet Richards, "Fourth Generation Warfare and OODA Loop Implications of the Iraqi Insurgency" presentation, Washington, DC, 2004.

4 Implications

In 1966 Army Chief of Staff Harold Johnson commissioned a study of the war.[12] The PROVN study was conducted by ten officers with a specified backgrounds: historian, political scientist, economist, cultural anthropologist, and specialists in intelligence, military operations, psychological operations, and economic assistance. The study charge stated that "Vietnamese and American people, individually and collectively, constitute strategic determinants of today's conflict…" The group and its work was nothing like the quantitative analysis favored by Secretary McNamara's "whiz kids" of the time. The PROVN report mirrored Vann's assessment that the U.S. must shift to "a strategy of pacification, at the village level where the war must be fought and won."[13] Considering our current and future urban battlefields in Iraq, or someday perhaps in Karachi or Cairo or Riyadh, we must be able to determine the situation with respect to the local economy, public safety, religious/social/ethnic trends, public health conditions, and physical infrastructures. Two particular directions of qualitative analysis are important: (A) operations "down spectrum" from classic counterinsurgency: peacekeeping, humanitarian relief, and support of local security forces, and (B) how crafted, evolving military organizations face conflict networks with networks of their own.[14] American insurgency challenges involve complex four way judgments involving American and insurgent forces and the supporting civilian infrastructures supporting each side. Measuring the strength of insurgent "bonds" between their military forces and civilian sympathizers, to compare against the often-weaker bond between American forces and the supported government, these important calculations are predominantly qualitative. Numbers alone will not tell the complete story, in fact qualitative and trend judgments will be the numbers that matter.

Mao Zedong cited the 1934-1935 Long March as the formative experience of the Chinese communist movement, the climax of "the crucible years."[15] But by quantitative measure an outside observer would have judged the Long March a horrible and fatal blow to the movement. Only through thoughtful qualitative judgment could one conclude that Long March strengthened the Chinese communist movement. This is a very relevant point to our challenge today of judging the strength and breadth of global Islamic extremism and its connections with the Iraqi insurgency. Unlike our failure to adequately understand the Soviet/Afghan crucible, we must qualitatively and accurately judge the influence and impact of today's conflict in Iraq.

5 Qualitative Analysis

Insurgency is a complex social and human factors problem and must be understood in qualitative ways, which introduces assumptions, principles, and values about truth &

[12] Lewis Sorley, "To Change a War: General Harold K. Johnson and the PROVN study," Parameters, 1998, p. 96.
[13] Buzzanco, Bob, "Vietnam and the Transformation of American Life," Houston, 1999.
[14] Gompert, Pung, O'Brien and Sepp, p. 29.
[15] Paul Godwin, "Excerpts from 'The Chinese Communist Armed Forces,'" Air University Press, June 1988, p. 8.

reality. Human experience takes place in subjective terms, social context, and historical time; we search for interpretive terrorism understanding against a world-wide background of dynamic Islamic history and pressure. Kvale suggests that qualitative analysis can be thought of as craftsmanship, suggesting susceptibility to political and bureaucratic distortion. Analysts often move directly from conclusions to program and policy recommendations, which should fit context, findings and understanding.[16] Qualitative analysis of current terrorism and insurgency problems must be done in a consistent, disciplined and rigorous manner, to ensure the crafted outcome is of high quality and not distorted by expectations. Intelligence is not always actionable.

Qualitative data analysis involves coding or categorization, an interpretive process requiring judgment in treating full, partial, and inapplicable data.[17] Corpus-based methods take as input a mass of textual material that is analyzed as a whole, while case-based methods view the data as a set of comparable cases that replicate each other.[18] Corpus-based methods are normally more interpretive, less formal, less replicable, and more subject to bias. For example, how might American intelligence evaluate local human intelligence generated in the Baghdad area over the past few months? Is the quality of the intelligence increasing? What qualitative analysis can be drawn from insurgent and terrorist public statements? Much of the public discussion process consists of pro and con arguments drawn from the same corpus of data, but presented in opposite directions depending on the presumption of outcome.

Judgments drawn from qualitative data must demonstrate validity of several kinds.[19] Assuming that descriptive validity (factual accuracy) is clear, other forms must still be confirmed. Interpretive validity requires accurate understanding and reporting of participant viewpoints and thoughts; theoretical validity requires that derived theory fits the data and hence is credible and defensible; internal validity is demonstrated to the degree that causality is demonstrated or proven; and external validity shown in the applicability or transferability of conclusion. Does analysis of Al Qaeda and Zarqawi statements and various Islamic commentary on them demonstrate interpretive validity? How applicable is our current understanding of network models and social network analysis to the presumed network-like groups of insurgents and terrorists?

Our challenge of evaluation crosses the two dimensions of data and analysis.[20] Qualitative data and analysis (Region I) is often the domain of literary criticism, interpretation, and theory. Region II, quantitative data and qualitative analysis, involves factor and cluster analysis; Region III, qualitative data and quantitative analysis, involves the statistical analysis of text frequencies, code occurrences, etc. Quantitative data and analysis (Region IV) is the domain of multivariate methods. Our analysis too often focuses on product and quantity instead of insight and quality; we are too often in Region IV and needs to move towards Region I. Our data systems too often generate precision that is not there; a volume of gathered data presumes certain insight.

---

[16] M.B.Miles & A.M. Huberman, "Qualitative Data Analysis, 2nd Edition," 1994, pp. 249-251.
[17] Sally Thorne, "Data analysis in qualitative research," 2000, p. 69.
[18] Steven Borgatti, "Introduction to Grounded Theory," Analytictech, January 2005.
[19] R. Burke Johnson, "Examining the Validity Structure of Qualitative Research," Education, 1997, p. 284.
[20] John Seidel, "Qualitative Data Analysis," Qualis Research, 1998.

Various techniques are used to identify themes in qualitative data, such indigenous categories, key words in context, comparison & contrast, metaphors and analogies, and others.[21] But many of these techniques are difficult to use against our current challenge. Word repetition is difficult to use well against the challenging Islamic religious, political and linguistic environment. Metaphors and analogies are popular current tools, but blending historical or cyclical insurgency trends with historical shifts and trends in Islamic religion, culture, and politics, remains a progress in work as the last few years have seen substantial thematic re-examination of qualitative information about terrorism and the global Salafi jihad.

Good qualitative analysis is both systematic and disciplined, and arguably replicable by "walking through" analyst thought processes and assumptions.[22] Unfortunately, this process is not done well across the boundaries of our bureaucracy. Qualitative analysis of the same problem by Defense and State Departments and Intelligence community will show substantial differences. Each organization applies somewhat different data and bring different editorial perspectives to the analysis. Morse cites that qualitative analysis involves comprehension, synthesis of relations and linkages, theory of the nature of relations, and re-contextualizing, or putting new knowledge back out for consideration. Miles and Huberman cite stepping back and systematically examining and re-examining the data, using "tactics for generating meaning:" noting patterns and themes, clustering cases, making contrasts and comparisons, partitioning variables, subsuming particulars in the general, etc.[23] Qualitative analysts employ some or all of these tools, simultaneously and iteratively. However, this process is often done poorly across the American national security community, across its compartmentalized and classified structure. Our current focus on information sharing is not only important to discover unknown information, but perhaps more so to open up qualitative analysis and interpretation across the analytical community.

6 Conclusion

The nature of qualitative data analysis in today's terrorism, insurgency and national security domain suggests several key concerns. Qualitative data analysis fails when pressured by policy expectations and interpretations. It also breaks down across bureaucratic separation of security classification and compartmentalization. And qualitative data analysis fails when drawing on analysis constructs from the wrong history and experience. Qualitative analysis must be honest and reflective about analytic processes and not just product. Clear and open presentation of analysis allows independent judgment of whether analysis and interpretation are credible and accurate.

We require skill with insurgency and terrorist informatics so that quantitative analysis builds on strong qualitative foundations. We must properly understand, anticipate and diagnose complex problems such as insurgency trends in Saudi Arabia and Karachi and terrorist support networks in the homeland. The key problem will

[21] R. Burke Johnson, p. 285.
[22] Thorne, p. 70.
[23] Seidel, "Qualitative Data Analysis."

not be the data but rather constructs and methodologies used to assemble the present (and missing) data into conclusions. This qualitative challenge is the heart of terrorist informatics.

References

1. Stephen P. Rosen, "Vietnam and the American Theory of Limited War", *International Security* 7:2 (Autumn 1982), pp. 83-113.
2. Sun Tzu, The Art of War, translated by Ralph D. Sawyer, Boulder, Colorado, Westview Press, 1994, pp. 165-193.
3. Mao Tse Tung, "Problems of Strategy in China's Revolutionary War," from *Selected Military Writings of Mao Zedong*, Peking: Foreign Language Press, 1972, CSI reprint.
4. Paul Godwin, The Chinese Communist Armed Forces, Maxwell Air Force Base, Alabama, Air University Press, June 1988, pp. 3-16.
5. Steven Metz (December 1993), "The Future of Insurgency," Strategic Studies Institute monograph, US Army War College, Carlisle Barracks, Pennsylvania.
6. Steven Metz (February 1995), "Counterinsurgency: Strategy and the Phoenix of American Capability," Strategic Studies Institute monograph, US Army War College, Carlisle Barracks, Pennsylvania.
7. John Shy & Thomas Collier, "Revolutionary War," in *Makers of Modern Strategy*, edited by Peter Paret, Princeton University Press, Princeton, New Jersey (1986), pp. 815-862.
8. Roger Trinquier, *Modern Warfare: A French View of Counterinsurgency*, translated by Daniel Lee, Pall Mall Press, London, 1964, CSI reprint.
9. Max Boot, *The Savage Wars of Peace*, Basic Books, New York, NY (2002).
10. Gentry, John A. "Complex Civil-Military Operations: A U.S. Military-Centric Perspective." Naval War College Review 53, no. 4 (Autumn, 2000): 57-76.
11. "Columbia Insurgency," maintained by John Pike of Global Security, accessed January 2005 at http://www.globalsecurity.org/military/world/war/colombia.htm.
12. Keith F. Kopets, "The Combined Action Program: Vietnam," Military Review, Combined Arms Center, Ft. Leavenworth, Kansas, July-August 2002.
13. Bob Buzzanco, "Vietnam and the Transformation of American Life," Blackwell Publishers, 1999, accessed through (http://vi.uh.edu/pages/buzzmat/fbb3.htm)
14. Sorley, Lewis, "To Change a War: General Harold K. Johnson and the PROVN Study," Kansas University Press, Lawrence, Kansas, 1998, pp. 93-109.
15. Robert M. Cassidy, "Back to the Street Without Joy: Counterinsurgency Lessons from Vietnam and Other Small Wars," Parameters, US Army War College, Carlisle, Pennsylvania, 2004, pp. 73-83.
16. Whitecross, Chris, "Winning the Nation's Hearts and Minds - Combating Asymmetric Warfare," Canadian Forces College, 2002.
17. David C. Gompert, Hans Pung, Kevin A. O'Brien and Jeffrey Peterson, "Stretching the Network," RAND, Santa Monica, CA, 2004.
18. Frank Pelli, "Insurgency, Counterinsurgency, and the Marines in Vietnam," USMC Command & Staff College, Quantico, VA, 1990.
19. Robin Navarro Montgomery, "Psychological Warfare and the Latin American Crisis," Air University Review, Montgomery, AL, August 1982.
20. Brian M. Linn, "Provincial Pacification in the Philippines, 1900-1901," Military Affairs (April 1987): 62-66.

21. Col G. I. Wilson, LTC Greg Wilcox and Col Chet Richards, presentation titled "Fourth Generation Warfare and OODA Loop Implications of the Iraqi Insurgency," distributed by Inside Washington Publishers.
22. Steven Borgatti, "Introduction to Grounded Theory," accessed January 2005 at www.analytictech.com/mb870/introtoGT.htm
23. John V. Seidel, "Qualitative Data Analysis," 1998, accessed at www.qualisresearch.com.
24. Trisha Greenhalgh and Rod Taylor, "How to read a paper: Papers that go beyond numbers (qualitative research)," British Medical Journal 1997; 315:740-743 (20 September, accessed at http://bmj.bmjjournals.com
25. Sally Thorne, "Data analysis in qualitative research," EBN notebook; 3:68-70 (2000), accessed at http://bmj.bmjjournals.com
26. Anne Marshall & Suzanne Batten, "Researching Across Cultures: Issues of Ethics and Power," Forum: Qualitative Social Research, www.qualitative-research.net/fqs, volume 5, number 3, article 39 – September 2004.
27. M. B. Miles and A. M. Huberman, Qualitative Data Analysis, 2nd edition" Sage Publishing, Thousand Oaks, CA., 1994, pp 245-262
28. R. Burke Johnson, "Examining the Validity Structure of Qualitative Research," Education, 1997, #118, 282-293.
29. Bobbi Kerlin, "Qualitative Research: Analysis Without Numbers," 1999, accessed at http://kerlins.net/bobbi/research.
30. Steven Metz, Robert Killebrew, Brian Linn and Kalev Sepp, "Relearning Counterinsurgency: History Lessons for Iraq, Afganistan, and the Global War on Terror," transcript from American Enterprise Institute presentation, January 10, 2005, accessed at http://www.aei.org/events/filter.,eventID.982/transcript.asp

The Application of PROACT® RCA to Terrorism/Counter Terrorism Related Events

Robert J. Latino

Executive Vice President, Reliability Center, Inc., Hopewell, VA 23860
blatino@reliability.com

Abstract. Field proven Root Cause Analysis (RCA) from the industrial sector can assist the terrorism community in decompiling terrorist acts to further understand the mentalities that trigger such events to escalate. RCA is a disciplined thought process that is not specific to any industry or given situation, but specific to the human being. We will focus on how to logically breakdown a seemly complex event into it more manageable sub-components.

1 What Is Root Cause Analysis (RCA)?

RCA has been around for centuries in some form or fashion. Some would say dating back to Socrates and Plato. However, it became a recognized and disciplined Reliability Engineering science at the turn of the 20th Century with the advent of the aerospace industry. For years the aerospace and U.S. military have used evolving variations of what was deemed Reliability Engineering and RCA at the time.

In the latter part of the 1960's such technologies were migrated into the heavy manufacturing sectors. In 1972 Allied Chemical Corporation (now Honeywell) chartered their Corporate R&D Reliability Center to explore the areas of equipment, process and human reliability as it pertained to the Chemical industry. As a result of this R&D, they created the first known Reliability Engineering Departments in the Chemical sector. Technologies such as RCA were a part of this emerging field of engineering.

From that point on, formal RCA spread to other continuous process and discrete product (or batch) manufacturing operations. The nuclear field also picked up the science and integrated into the way they did business. As of recent the healthcare and service industries are now applying such technologies to improve patient safety and increase customer satisfaction.

To date, as a provider in the Root Cause community, a standard definition of RCA has not been adopted due to an inability to gain consensus among the major providers and practitioners. However, of the attempts made to do so in this community, this rendition is an acceptable one to this author:

Root Cause Analysis (RCA)
To be classified as RCA, a process must answer all of the following questions satisfactorily and in the order stated:

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 579 – 589, 2005.

1. "What is the problem to be analyzed (problem definition)?"
2. "What is its significance to the stakeholders (problem significance)?"
3. "What are the causal factors that combined to cause the defined problem (problem analysis)?"
4. "How are the causal factors interrelated (causal chart)?"
5. "What evidence exists to support each causal factor (cause validation)?"
6. “What are the recommended actions to solve the same and similar problems (recommendations)?"
7. "What assurance exists that the recommended actions are justified (solution validation)?
8. "What are the lessons to be learned from the problem analysis (lessons)?"

2 The PROACT® Approach to Root Cause Analysis

The original Allied Chemical Reliability Center R & D Group developed the PROACT® Approach to RCA. It has been applied in most every industry type within the Fortune 500. What is important to remember is that RCA is not industry or problem specific. RCA is a skill developed and refined by the human mind in how to accurately determine why things go wrong and why people make the decisions that they do. The nature of the undesirable event is irrelevant when considering the thought processes to solve them is inherently the same.

Once a qualified candidate for RCA has been selected, it must be subjected to a formal investigative process. Here we will explore the PROACT® Approach in more detail. This is an acronym for the following:

PReserving Event Data
Ordering the Analysis Team
Analyzing the Data
Communicating Findings and Recommendations
Tracking for Bottom Line Results

In terrorism we experience undesirable outcomes. These undesirable outcomes are a result of a connected series of cause-and-effect relationships, whether intentional or not, that line up in a particular sequence on a particular day. Dr. James Reasons of the University of Manchester coined the term Swiss Cheese Model to describe this sequence.

This is an effective mental model because it easily depicts that the conditions for undesirable outcomes are always present. It is the ability for those conditions to line up in a particular sequence that allows them to cause these adverse events. Each slice of the cheese represents a defense mechanism or a barrier to ensuring that the conditions do not line up in such a fashion to allow a cause-and-effect relationship to form. We will discuss these details in the STEP 3 – Analysis.

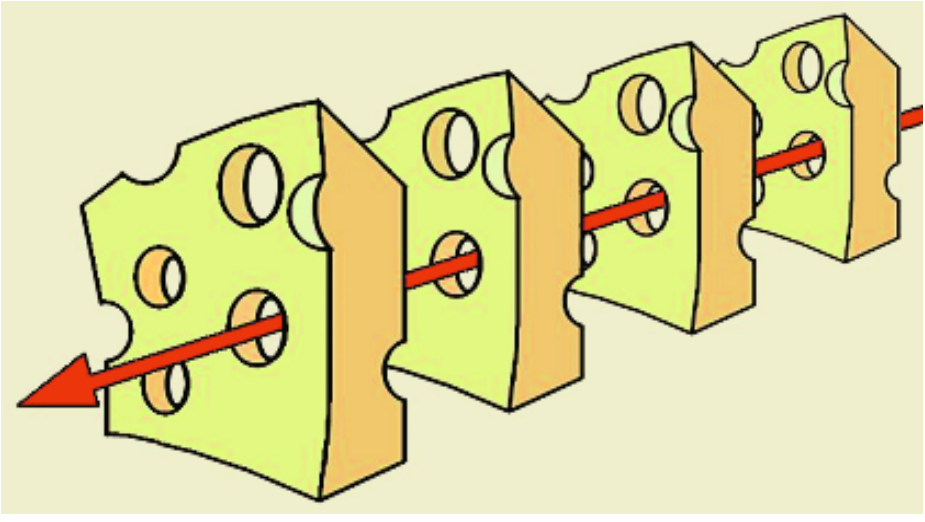


Fig. 1. James Reason's Swiss Cheese Model

3 PROACT STEP 1 - PReserving Event Data

Whether a Terrorism analyst, an NTSB investigator, a police detective or a doctor for that matter, the first step in any analysis is the collection of data. When we watch detective shows on the television, what are the first actions that take place at a crime scene? We see the area being roped off, all the evidence being "bagged and tagged" and interviews taking place. This is a matter of standard practice for investigative agencies.

The PROACT® Approach employs the use of the 5-P's data gathering process. The 5-P's are as follows:

Parts
Position
People
Paper
Paradigms

Any necessary data from an undesirable outcome can be categorized in one of these categories. This is simply a job aide to assist in organizing data collection efforts. It helps to manage the information and organize documentation efforts. Mainly, it obtains the data in a formalized manner and makes it available for the analysts to use it is needed to prove and disprove certain hypotheses.

Without the preservation of data, we are like a detective trying to solve a crime with no evidence and no leads.

4 PROACT STEP 2 - Ordering the Analysis Team

It is our contention that the role of an RCA team leader is purely facilitative in the RCA methodology. In other words, facilitator's facilitate and do not participate. The expertise on the team will provide the talent to generate and validate appropriate hypotheses.

By not being an expert in the nature of the undesirable being analyzed, we can ask any question we like because we are not expected to know the answer. An expert leading the team is not afforded this luxury. They are expected to know and therefore must act like they know all, even if they do not. It has been our experience that when such "experts" lead these RCA teams, they tend to know their conclusions before they start with the team, and subsequently will drive the team to their conclusions. Such an expert also tends to intimidate the team members because of their perceived expertise even if they do not mean to.

Noted author Eli Goldratt in his best selling book The Goal, cited "an expert is not someone that gives you the answer, but someone that asks you the right questions". This accurately describes the role of an RCA facilitator. If we give people the answer, they have not learned how to deduce it for themselves to come to the same conclusions. That deduction process is vital to RCA and critical to the thought pattern development of the analyst.

5 PROACT STEP 3 - Analyzing the Data

Ideally at this point we have collected our initial data and have organized a specific team to analyze the event. The data at this point is like the pieces of the proverbial puzzle. Now we must take the pieces of the puzzle and make sense of them by constructing an effective manner is which to reconstruct the sequence of events that led to an undesirable outcome. The PROACT process utilizes what we call a logic tree.

A logic tree is a graphical representation of a cause-and-effect sequence. No matter the nature of an event, there will always be a sequence of "chain links" that led to the adverse outcome. It is the RCA team's responsibility to determine the sequence of events leading to an undesirable outcome and to base their findings on as solid facts as our available.

To construct a logic tree, the team must start off with facts. At this point, we are dealing with what HAS occurred, not with what might occur. If we can revert back to our detective analogy, we are starting with the evidence and leads from the crime scene, or the FACTS! We call this first step in the logic tree development the Top Box. This is usually the first two levels of the logic tree.

Most do not realize that when conducting true RCA, we are really analyzing an event because of its consequences as opposed to the event itself. If an adverse outcome occurs, and its consequences are not significant, then a true RCA will not likely be commissioned. It is important to accept this reality and understand its ramifications.

If we investigate an "incident" as opposed to its "consequences" we are likely to miss the time frame between the incident and its consequences. This is key because

as part of the learning of the RCA, we want to understand what we could have done better to minimize the consequences and how we may have been able to respond more appropriately to the event. By just analyzing the event itself, we would miss the opportunity to analyze why the consequences were permitted to be so severe. With this differentiation being made, the Top Block is the description of the consequences and the Modes (2nd level) describe the event(s) that led to the consequences.

For instance, based on the 9/11 Investigation Report, the "Event" may be characterized as "An Attack on the U.S. Homeland Resulting in 2973 Deaths".

The level beneath the event is what we call the mode level. This is often the symptoms or manifestations of the event that occurred. In the case of the 9/11 attacks, the "Modes" may have been characterized as "Clinton Administration Deficiencies in Response to Threats", "Bush Administration Deficiencies in Response to Threats", "9/11 Response Deficiencies During Attacks", "9/11 Post-Attack Response Deficiencies". This Top Box might look like Figure 2.

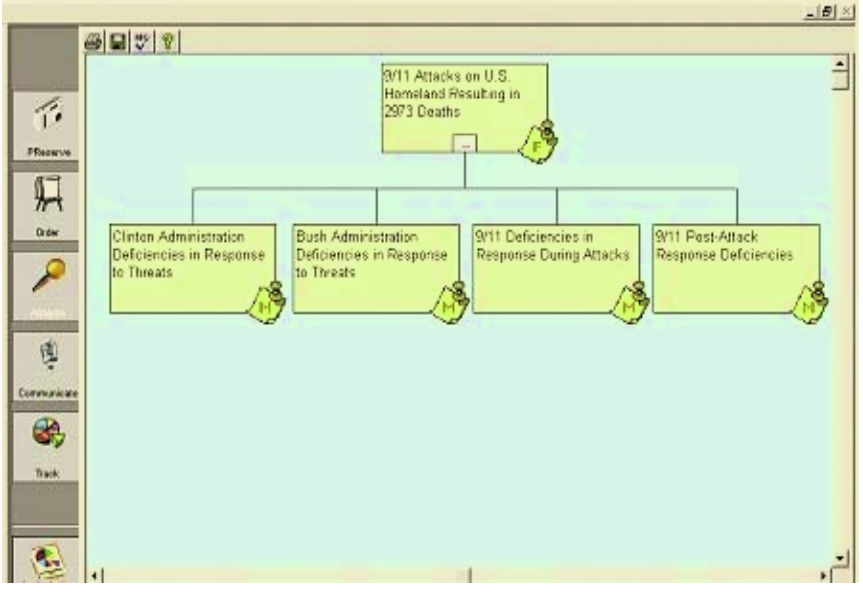


Fig. 2. Sample TOP BOX of Logic Tree

This is being used for example's sake and is expressing a desire to analyze why the terrorist act was permitted to permeate the defenses that we had in place. We could just as easily add a mode to the effect of "Terrorist Desire to Attach U.S." and focus on the human decision-making processes that involve fundamental and critical driving forces such as culture, perception, and competing institutional and personal agendas.

At this point in the tree, we must revert to hypothetical questioning. If we first explored the Mode about the post-attack during the attacks, all we knew at the time was that commercial airliners struck the World Trade Center (WTC) North and South Towers and the Pentagon.

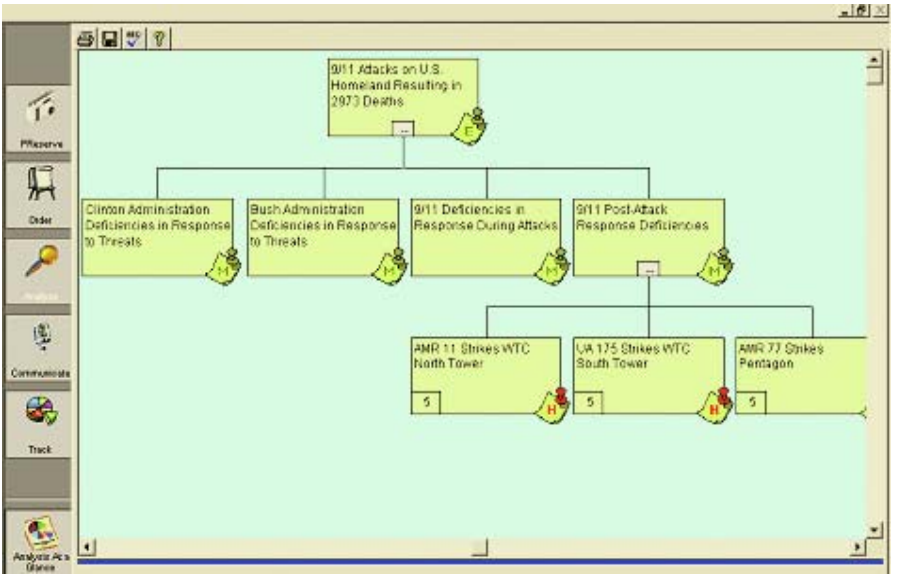


Fig. 3. Begin Hypothesizing from the TOP BOX and verifying

We do not know why, we just know that they were struck. Now we must start to hypothesize about, "How could they have been allowed to be struck?" The team of experts will provide the hypotheses to fill in the blanks. But it is the facilitator's role to ensure that each hypothesis is either proven or disproved with some essence of science or direct observation and evidence.

The readers will notice in Figure 3 that numbers are located in the lower left hand corner of each node. This is referred to as a Confidence Factor. This is a subjective assessment of the strength that specific evidence brings to validating a particular hypothesis. The scale ranges from 0 to 5, where a "0" indicates that with the evidence collected, we are 100% sure that the hypothesis is NOT true. Conversely, a "5" would indicate that with the evidence collected, we are 100% sure that the hypothesis IS true. The rest of the spectrum provides a weighing system for inconclusive evidence that swings the pendulum more in one direction than another.

Certainly when we get into the Human and Latent areas of such analyses, evidence becomes more difficult to analyze as we are now dealing with "soft" issues as opposed to physical issues. We must recognize that people's paradigms, whether reality or not, drive their decision-making processes. These paradigms include their attitudes, perceptions and basically how and why they see the world as they do. While validating such esoteric hypotheses is difficult, it is not insurmountable.

Undesirable outcomes cannot occur unless people trigger certain conditions to escalate. People "trigger" these conditions with their decisions to act or not act (errors of commission or omission). People make such decision based on their paradigms that we discussed earlier. So from a data validation standpoint, hypotheses about our attitudes, beliefs, etc. will eventually be able to be validation based on our behavior in the form of decision-making. Our decision-making triggers a physical

sequences of events to occur until our defenses either stop the chain, or allow it to progress to the point that the undesirable outcome is permitted to occur.

As we reiterate this "How Can" questioning process down the logic tree, we hypothesize and then we validate continuously until we reach various root cause levels. The various root cause levels comprise of Physical, Human and Latent root causes. All events that we have ever analyzed have included these three levels of cause. However, we cannot say the same for the some of the analyses we have seen from our clients. Let us explain.

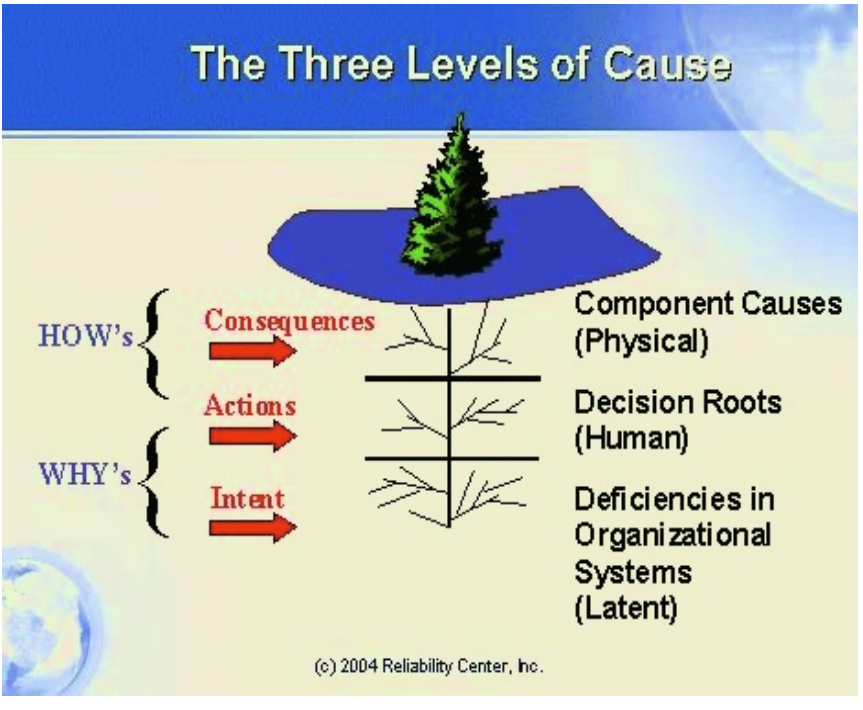


Fig. 4. The Three Levels of Roots

Figure 4 depicts our comments earlier about the triggering mechanisms that result in undesirable outcomes. Dr. Reasons uses the terms Intent, Action and Consequences, which are synonymous with our use of Latent Roots, Human Roots and Physical Roots. Generally physical root are associated with tangibles and deal with "how" something occurred. When delving into the human and latent levels we are dealing with the "whys". This is why popular shows like CSI depict forensic engineers using high technology to determine how things occur using physical evidence, but it is the detective and prosecutors responsibility to determine the "whys".

The Physical Roots are usually where most organizations that claim to do Root Cause Analysis, stop. This is the tangible or component level. In our example of the 9/11 Response perhaps we determine that some of the radios used by first responders

to the WTC did not have proper signal strength to reach the upper floors, or they did not pick up certain frequencies at all. If we stop at this level, "Have we solved the root causes of what allowed the event to occur and be as bad as it was?" Not likely as we have merely identified a physical condition that existed and not considered how such radios did not have the signal strength needed and were not able to pick up needed frequencies. This is especially in light of the 1993 World Trade Center bombing where such deficiencies were noted as well. How come we did not learn from the past terrorist attack and ensure that these noted deficiencies would be corrected?

This same phenomenon is true when contrasting the Challenger disaster to that of the Columbia. The "culture" at NASA had a great deal to do with these tragedies and how come we did not learn from our mistakes with Challenger so that we could have prevented Columbia? Boston University Sociologist Diana Vaughn put it into proper perspective when she coined the term "Normalization of Deviance".

As mentioned earlier, Human Roots are errors of omission or commission by the human being – decision errors. Errors of commission are like the 911 operators telling stranded people on floors above the fires in the WTC to go to the roof. The decision to tell them this triggered the response of these people's attempts to get to the roof when it was actually deemed that a roof rescue was impossible and that there was no Standard Operating Procedures (SOP) for a rescue above the fire at that level. This is a sensitive cause level and one that is often perceived as "witch hunting" if handled incorrectly. This is definitely NOT the level to stop at when performing an RCA.

The last level is truly the ROOT CAUSE level. This is the level beneath the Human Root and it is termed the Latent Root Cause. This is level that delves into the mind of the people that made decision errors above in the human root. We call these latent roots because they are dormant and hidden in the daily routine of business. They are not earth shattering to find out. We all know that we have these problems and we accept it until something fails to perform as desired. We often use the term latent root and management system root synonymously. Management systems are the rules and regulations of any organization. They are the policies, procedures, practices, etc. that govern an organization.

This is even true of terrorist organizations, as they must possess an infrastructure to be successful. The 9/11 Commission Report cited some example of these terrorist organizations that violated their own rules like 1) using cell phones that could be traced and 2) making unscheduled travel plans to visit girlfriends in different countries. These types of activities provide a path that can be traced and violate the internal rules of some terrorist organizations. Latent Root Causes describe the flawed systems put in place to help people make better decisions.

When we continue to use the same example as we have been using, we may find that the event that occurred was not anticipated therefore plans for its occurrence were not in place. Possible Latent Root Causes may be "lack of proper foresight in planning for a catastrophic fire on the top floors", "lack of foresight in planning for a coordinated communication system for all first responders (where they could all have access to the same frequencies)" and "lack of SOP for informing 911 responders how to advise people stranded above a fire in the WTC towers". These are just some possible examples.

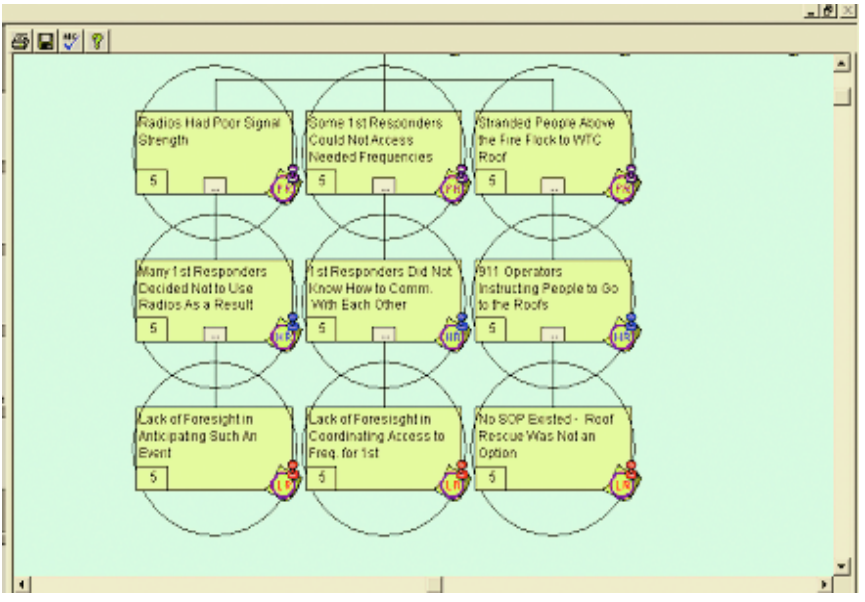


Fig. 5. Example of Root Cause Sequences Using 9/11 Attacks Example

6 PROACT STEP 4 - Communicating Recommendations and Findings

While the focal point of the paper is on determining Root Causes, a key part of the process is to ensure that the recommendations that we develop to counter the root causes identified, are implemented.

Effective strategies for doing this will vary from organization to organization. However, in the end, if we do not implement the countermeasures to identified root causes, then we can plan on the event occurring again.

Examples of this have been cited in this paper where we did not learn from past events, and were not able to prevent similar future events. Many of the root causes identified in the Columbia Investigation were the same as identified in the Challenger investigation. Most of the Challenger recommendations dealt with cultural issues that were apparently not overcome because Columbia suffered a similar fate.

The same can be said for the 1993 WTC bombing. First responders should have learned a great deal from that event that would have helped minimize the consequences of the 9/11 attacks.

7 PROACT STEP 5 - Tracking For Bottom-Line Results

What is the true definition of success of a RCA? Is it determining accurate causes? Is it developing plausible recommendations? Is it gaining approval of the

recommendations? Is it implementing recommendations? The answer to all of the above should be NO.

The true definition of a successful RCA is that some bottom-line performance measurement has improved as a result of recommendations implemented from the RCA. Have we averted attacks because our defense mechanisms have been sharpened? Have we sharpened our ability to respond more efficiently and effectively to an attack? Are we seeing fewer threats as a result of our new global strategies on terrorism? Are wee seeing terrorist using different tactics because our defenses are preventing them from using conventional tactics? Are we seeing less movement of the terrorist because their conventional funding routes have been disrupted?

Remember that a true RCA is not successful unless we have improved future performance in some manner. We cannot determine our success unless we demonstrate the positive effects of the analysis.

8 Automating Root Cause Analysis

The PROACT® Approach described throughout this feature can, and obviously has, been performed manually for decades. With the recent advances in technology and the availability of computers in the workplace, automation has proven to be a sound alternative.

Automation in terms of software does not mean that we utilize software that does the thinking for us. Automation should be introduced when humans must perform repetitive tasks in the same manner. These types of situations are classic for the introduction of human error due to the monotony of the task. Such tasks can be automated thus reducing the risk of human error.

In conducting a RCA, the repetitive tasks are the double handling of data. When utilizing a manual method of analysis, we often must first write on an easel, a marking board, drafting paper, etc. when we work in teams. This requires someone to transcribe what was on the paper to a software program. Once this is done, the information is then disseminated to the remaining team members. This process usually involves a lapse of a day or two before all the team members have the same information.

The PROACT Software allows all documentation for a given analysis to reside in the one file on one server. When using the software, we are able to project a real time logic tree that allows all team members to focus on a certain string of logic and to be able to present evidence to validate or refute certain hypotheses. Therefore, we are developing a collective thought process that everyone agrees upon.

Because this information is stored in database, this information can be manipulated for data-mining purposes and also shared from remote locations. When completed, these analyses can be used for lessons learned purposes used to train analysts in the field to understand in details why these events occurred and what were the triggering mechanisms.

9 Conclusion

The PROACT® Approach is a time and results proven tool that assists analysts in breaking down complex events into their manageable components for further learning and understanding. This evidence-based approach promotes a non-personal, non-threatening tool where data drives the analysis and uncovers the truth. Utilizing the automation technologies associated with such tools allows for the instant dissemination of such vital information to those that need it most. Such tools are also designed to be used for lessons learned when educating others in the thought processes involved that lead to the undesirable outcome.

References

1. A registered trademark of Reliability Center, Inc., www.reliability.com Root Cause Analysis Improving Performance for Bottom Line Results (2nd Ed., 2002, c. 246 pp., ISBN: 0:8493-1318-X, CRC Press)
2. Reasons, James. Managing the Risks of Organizational Accidents. Page 9. England: Ashgate Publishing, 1998
3. Goldratt, E. M. and J. Cox (1993): The Goal, Gower Publishing, Croftroad.
4. The 9/11 Commission Report, Official Government Edition, (2004), ISBN 0-16-072304-3
5. Challenger: Disaster and Investigation. Cananta Communications Corp.1987
6. Lewis, Richard S. Challenger: The Final Voyage. New York: Columbia University Press, 1988.

A Group Decision-Support Method for Search and Rescue Based on Markov Chain

Huizhang Shen[1], Jidi Zhao[2], and Ying Peng[3]

Institute of System Engineering, Shanghai Jiao Tong University,
Shanghai, 200052, China
{hzshen, judyzhao33, pengying_1127}@sjtu.edu.cn

After perils of sea such as shipwreck and tsunami, airplane disaster or terrorist raid befell, the rescuing departments will receive omnifarious alarms, asking for help, orders and requests. One of the most important tasks of the work is spreading the search for the lost people and rescuing the people. As the conditions limitation, the rescuing departments can not respond to every request, and can not search and rescue all over the districts at the same time. Thus the decision-makers should make choice among several kinds of search and rescue action schemes, which forms the group decision-making problem of the search and rescue schemes that we will discuss about in this paper.

The premise of the search and rescue problems is that the group decision-makers are all rational man. Let DMr be the r^{th} decision-maker, G be the decision-making group, and let x^i be the i^{th} scheme, X be the scheme set. Prior research experiences on psychological theories show that people can carve up their preference into grades, such as "equal, a little better, better, much better, absolutely better". Let $\theta_r(x^i, x^j)$ be the quantificational difference of the DMr's preference degree on the two schemes x^i and x^j. Choose a random scheme x^0 as the benchmark, and require DMr to present his preference utility vector $(\theta_r(x^1,x^0) \quad \theta_r(x^2,x^0) \quad \cdots \quad \theta_r(x^s,x^0))^T$ on the scheme set. Therefore, the preference utility vectors of l decision-makers can construct the preference utility value matrix $\wedge$.

Evaluation Model. The evaluation model for search and rescue decision-making based on preference utility value quantitatively analyzes the decision group **G**'s preference values of each scheme on the scheme set **X** and solves the sorting sequence of the schemes. If $\theta(x^i) \geq \theta(x^j)(i, j = 1,2,\cdots,s)$ where $x^j, x^j \in X$ or $x^i R x^j, (i, j = 1, \cdots, s, i \neq j)$, it shows that the decision-making group prefers to choose x^i between x^i and x^j. If $\theta(x^1) \geq \theta(x^2) \geq \cdots \geq \theta(x^s)$, the group decision-making preference sequence is $x^1 R x^2 R \cdots R x^s$. When the decision-making problem is to choose the best scheme from the scheme set, the scheme selected by the decision-making group is x^1.

Equilibrium Factor. Assume that the decision-makers have equal weights of making decision. In order to avoid the great excursion caused by extreme preference utility value given by one or two decision makers, we introduce an equilibrium factor to

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 590 – 592, 2005.

assure the group decisions' non-autarchy and balance the departure of the decision-maker. The chosen normal equilibrium factor satisfies normal distribution. Let $\varphi(\theta_r(x^i,x^0))=(\sqrt{2\pi}\sigma(x^i,x^0))^{-1}\exp(-(\theta_r(x^i,x^0)-\mu(x^i,x^0))^2/2\sigma^2(x^i,x^0))$. The farther the decision-maker is from the average preference utility value, the smaller the corresponding equilibrium factor is, vice versa. We multiply each element of $\wedge$ by corresponding factor and get the preference utility value matrix $\wedge$'.

All the elements of $\wedge$' are positive numbers. The process of group decision-making preference convergence is a Markov stochastic process according to its definition. Thus if only we can construct the probability matrix of the preference utility value according to the preference utility value matrix $\wedge$', regard it as the one-step Markov transition matrix, and show that the matrix is a normal random matrix, we can make out the solution of the metasynthesis problem for search and rescue by the property of normal random matrix.

Construct the Markov Transition Matrix. The preference utility value matrix $\wedge$' is a $l\times s$ matrix while the transition matrix P is a $n\times n$ matrix. In the actual group decision-making process, there are few instances that the number of the decision-makers is equal to the number of the schemes, so we introduce virtual decision-makers and virtual schemes to construct the Markov transition matrix. Let $n=\max(l,s)$. There are three instances.

(1) If $l=s$, then $n=l=s$, let $p_{ri}=\dfrac{\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0)}{(\sum_{i=1}^{s}\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0))}$

where $r=1,\cdots,l \quad i=1,\cdots s \quad x^i\in X$

(2) if $l<s$, then $n=s$, here we add $s-l$ virtual decision-makers, DMr ($r=l+1,\cdots,n$), let

$$p_{ri}=\begin{cases}\dfrac{\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0)}{\sum_{i=1}^{s}(\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0))} & r=1,\cdots,l\\ 1/n, & r=l+1,\cdots,n\end{cases}$$

where $i=1,\cdots s \quad x^i\in X$

(3) if $l>s$, then $n=l$, here we add $l-s$ virtual schemes $x^i(i=s+1,\cdots,n)$, let

$$p_{ri}=\begin{cases}(1-\varepsilon)\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0)/\sum_{i=1}^{s}\varphi(\theta_r(x^i,x^0))\cdot\theta_r(x^i,x^0), & i=1,\cdots,s\\ \varepsilon/(l-s), & i=s+1,\cdots,n\end{cases}$$

where $r=1,\cdots,l \quad x^i\in X$, $\varepsilon\in(0,1)$.

Obviously, every element p_{ri} of matrix $P=[p_{ri}]_{n\times n}$ is positive number. It can be easily show that the matrix P is a normal random matrix.

The meta-synthesis method based on Markov chain for search and rescue group decision-making problems follows the steps: (1) According to the rule of that the minority submit to the majority, choose the benchmark scheme; (2) Each decision-maker provides his preference utility value between each scheme and the benchmark scheme independently; (3) Calculate the equilibrium factors, and publish these factors to the decision-maker group. If any decision-maker is not satisfied with the equilibrium factor he gets, then back to the step (2) to revise his preference utility value, otherwise go to step (4); (4) Construct the corresponding Markov one-step

transition matrix; (5) Get the solution of the equations group according to the property of the normal random matrix; (6) Sort the group preference utility values.

A numerical study based on the theoretical analysis is then conducted to provide detailed guidelines of the meta-synthesis method application. Further research should include some experimental data to support the conclusion that this method can accelerate the convergence speed of the group decision-making result of search and rescue.

A New Relationship Form in Data Mining

Suwimon Kooptiwoot[1] and Muhammad Abdus Salam[2]

[1] Computer Department,
Suan Sunandha Rajabhat University,
1 Uthong Nok, Dusit, Bangkok 10300, Thailand
suwimonktw@yahoo.com
[2] School of Information Technologies,
The University of Sydney, NSW 2006, Australia
msalam@cs.usyd.edu.au

In this paper we study the problems pertaining to the rules mined from the data that do not always hold in the real world. We argue that the cause-effect relationships are more complicated in the real world than those can be presented by the rules mined using the current data mining techniques. Inspired by the concept of cause-effect relationships, the characteristic of the catalyst in Chemistry, and the theory of Net Force in Physics, we propose a new form of representation of rules by introducing a complex cause-effect relationships, the importance of the factors, and force unit. This form of relationship among attributes consists of all of the related attributes and the number of force units of each attribute and also the degree of importance that is like weight of each attribute to the target attribute. The target attribute of interest results from both the change in direction and the number of force units of the change. We have to consider the net force calculated from all of the related attributes including their degree of importance and the number of force unit on the target attribute.

The new form of the relationship we propose is shown follow.

$$\sum_{i=1}^{n} \delta A_i \int A_i = \triangle B \tag{1}$$

where
B is target attribute of interest

$\triangle B$: Outcome of B consists of the change direction and the force unit of the change

A_i : attribute i which effects on the change of B

δA_i : Degree of importance of A_i effecting on the change of B

$\int A_i$: Number of force unit of A_i acting on the change of B

We can use the relationship in this form to explain the results in the data mining experiments with real life data set and also the events seen in real life which do not follow the rules from data mining. This finding is very useful for both data mining users and data mining researchers. For users, this relationship can give users better understanding of the relationships among the attributes and show the reason why certain things do not follow the normal rules sometimes. For the researchers in data

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 593 – 594, 2005.

mining, it is an important key idea that the researchers have to find out the data mining algorithms in order to be able to extract the relationships in this form from real life data set.

Full paper can be obtained directly from the authors.

A Study of "Root Causes of Conflict" Using Latent Semantic Analysis

Mihaela Bobeica[1], Jean-Paul Jéral[2], Teofilo Garcia[2], and Clive Best[2]

[1] University of Nice, CRDL, BP 3209, 98 Bvd. Herriot, 06204 Nice cedex, France TNO, PO BOX 155, 2600 AD Delft, The Netherlands
mihaela.bobeica@wanadoo.fr
[2] IPSC, Joint Research Centre, European Commission, I-21020 Ispra, Italy
{jean-paul.jeral, teofilo.garcia, clive.best}@jrc.it

This paper describes a method for the measurement of root causes of conflicts, as defined in a checklist drawn up by the European Commission's External Relations Directorate General (DG RELEX) and used for monitoring and early warning. Our approach uses Latent Semantic Analysis (LSA) to measure these conflict indicators on a corpus composed of news articles extracted from the Europe Media Monitor (EMM) archive.

DG RELEX have defined a methodology for assessing the risks of conflict within a given country and these assessments are known as Country Conflict Assessments (CCAs). They are a valuable source both for detecting long-term trends (improving or worsening situation) and for maintaining a current watch list of countries assessed as being at high risk of conflict. By taking a purely numerical approach to estimating these indicators in news data, we have produced results that avoid manual interpretation. These results can be used as objective input to foreign policy analysis and conflict assessment tasks and could also provide timely alerts to policy-makers and analysts.

The corpus used as a basis for our study was extracted from the EMM archive and represents news reports gathered over a two-month period, from April until May 2004. EMM is a media monitoring system for the European Commission, which runs 24/7 scanning about 600 news websites in 30 languages and 15 news agencies. EMM detects new articles as they are published in the world's media. All open source article texts are analyzed and the alerts they triggered are recorded. EMM has processed about 10 million articles spanning over 30 months of news.

The technique used for the measurement of conflict indicators in texts was Latent Semantic Analysis. LSA is a statistical technique based on the analysis of word distributions across a corpus of documents. Each word and each document is represented as a vector within a normalized vector space. The semantics of a document or a word represents a direction within this vector space. Documents or words pointing approximately in the same direction are semantically related. *Artificial topics* (corresponding to the listed conflict indicators) were created for each conflict indicator (e.g. "economic dependence"). Each list contains a few hundred words which were themselves derived using LSA in an iterative

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 595–596, 2005.

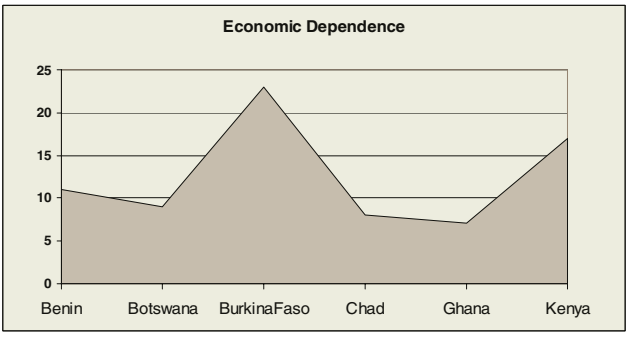


Fig. 1. Countries with the highest absolute value of the indicator "Economic Dependence"

process. The assumption is that these words represent a semantic direction for that conflict indicator within the normalized vector space.

Using LSA, close to 300000 EMM articles and the 24 artificial topics have been analyzed, in order to calculate the component vectors for each of them. The similarity value between an article and an artificial topic represented the dot product between the normalized vector of a given article and the normalized vector of the conflict indicator. A similarity threshold has been determined, after analyzing the relevancy of the results.

For each indicator, the articles were further divided by country alert. Countries mentioned most with respect to a given cause of conflict score higher. Therefore two normalized indicators were calculated: the importance of an indicator for the given countries and the overall reporting level for that country, on any subject.

We have used this fully-automatic method to process an important amount of data and the results hereafter can be used for cross-country comparisons. As an example, the results concerning some African countries - the Democratic Republic of Congo, Chad, Sudan, Rwanda, Uganda, Burundi, Benin - show that these countries have high scores for geopolitical instability, economic dependence and weak judicial system. These results correspond to the analysis presented in a variety of national and international risk assessment reports concerning these conflict-prone African countries. Figure 1 shows the results for several African countries, concerning economic dependence.

This initial LSA study is only in English and therefore it could be argued that a fairer coverage of languages and of news sources would allow for a more balanced and accurate distribution of results across countries and indicators. This analysis assumes that news reports reflect the real situation in each country, and that the relative importance depends on the relative numbers of articles as a fraction of the total for each category.

Full paper is available online at `http://media.jrc.it/Root_causes.pdf`.

An Empirical Study on Dynamic Effects on Deception Detection

Tiantian Qin and Judee K. Burgoon

University of Arizona
{Tqin, jburgoon}@cmi.arizona.edu

A threat to accurate deception detection is the dynamic nature of deceptive behavior. Deceivers tend to adapt their communication style over time by continuously monitoring their targets for signs of suspiciousness. As a result, deceivers manage to tell lies that sound more and more like truth. Such trends imply that deception detection in later phases of an interaction would be more difficult and thus less accurate than detection in earlier phases. This paper studies dynamic effects that influence deception detection and provides empirical evidence supporting the prediction.

Deceptive Behavior as a Dynamic Process. Buller and Burgoon [1] argued that to predict the behavior of deception, it is important to consider not just individual psychological variables but also interpersonal communicative processes. People tend to adapt their communication style over time. Deceivers usually engage in several tasks simultaneously: create messages, manage conversation, manage their self presentation, continuously monitor their targets for signs of suspiciousness, and adapt their behavior accordingly. Interpersonal deception theory predicts that deceivers will perform better over time as they become accustomed to the situation, setting, task, and communication style of their target and as they work to allay any suspicions revealed by their target [2, 3]. If this is true, then classifying deception in later phases of an interaction should be less accurate than in the early stages due to deceivers increasingly approximating a truthful communication pattern over time [4]. *The objective of this paper was to test the hypothesis that deception detection accuracy diminishes over time.*

Verbal and Nonverbal Cues. Both verbal and nonverbal cues are considered in the study. There are two major reasons for selecting these cues. First, differences between truth-tellers and deceivers can be captured by the cues because deceivers experience greater cognitive load, need longer processing time, may have different emotional states, and engage in different communication strategies. Second, the selected cues are ones that are relatively easy to implement and calculate by programs. The verbal classes of cues and respective indicators were as follows:

1. Quantity (number of words, number of verbs, number of sentences)
2. Complexity (Average sentence length, average word length, pausality)
3. Uncertainty (Modal verbs, modifiers)
4. Nonimmediacy (Passive voice, reference (second person pronounce, self-, group- reference and third person pronounce)

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 597 – 599, 2005.

5. Diversity (lexical diversity, content diversity, redundancy)
6. Specificity (Temporal immediacy, temporal nonimmediacy, spatial-far details, spatial-close details, and sensory)
7. Affect (affect, pleasantness, activation, imagery, emotiveness).

The nonverbal cues selected for examination were:

Talk and silence durations (talking time, turn switch pauses, other-talking time)
Nonfluencies (vocalized pauses, nonvocalized pauses, other nonfluencies)

Major Findings and Discussion. This investigation analyzed data from two segments from one experiment, called Deceptive Interviews. Three classification methods (discriminant analysis, logistic regression and neural networks) were considered to measure the performances of deception detection in the first and second segments of the experiment. Our results were consistent with expectations: detecting deception in the early phase of communication (Segment 1 of the Deceptive Interviews) was better than in the later phase (Segment 2). Table 1 is a summary of classification results.

Table 1. Classifications for Deceptive Interviews Experiment

	Method	**1st segment**	**2nd segment**
Classification Results	Discriminant Analysis	82%	59%
	Logistic Regression	61%	41%
	Neural Network	51%	46%

Deceivers' behaviors were more likely to betray them in the first segment than in the second, consistent with interpersonal deception theory 's postulate that deceivers deliberately adapt their language style and nonverbal behavior over time, thereby decreasing the difference between deception and truthful communication. This trend is also supported by the more reliable and powerful (in terms of Wilk's lambda) discriminant functions constructed from the Segment 1 data than Segment 2 data.

The other evidence supporting our expectation is the fact that far fewer significant cues appear in segment 2 compared to segment 1. The reduction in significant cues suggests that deceivers are improving their techniques so that the gap between truth and lies becomes smaller. In terms of specific dynamic language style, deceivers used more sentences over time, implying that deceivers became more comfortable in talking. More pauses in the early phase also suggests that deceivers were more nervous and experiencing larger cognitive load early on but that over time, they became more comfortable and their communication became more natural.

References

1. Buller, D. B. and Burgoon, J. K. (1996). "Interpersonal Deception Theory." Communication Theory, 6, 203-242.

2. Buller, D. B. and Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. In J. A. Daly & J. M. Wiemann (Eds.), Strategic interpersonal communication (pp. 191-223). Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Burgoon, Buller, Floyd and Grandpre, (1996). Strategic behavior during deceptive conversation. Journal of Language and Social Psychology.
4. White, C. H., and Burgoon, J. K. (2001). Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations. Human Communication Research, 27, 9-37.

Anti Money Laundering Reporting and Investigation – Sorting the Wheat from the Chaff

Ana Isabel Canhoto and James Backhouse

London School of Economics
james.backhouse@lse.ac.uk

The collection and analysis of financial data, referred to as financial intelligence, is gaining recognition has a key tool in the war on crime in general and terrorism in particular. Money in electronic form leaves a trail which means that individuals cannot easily disappear. There is a burgeoning industry providing sophisticated computer technology and complex mathematical models to mine financial data and single out unusual patterns of transactions. The use of automated monitoring systems is often seen as a powerful ally in the fight against money laundering and terrorist financing, justified by the increase in size of the typical transactional database, and by a desire to keep compliance costs under control. However, the power of automated profiling can result in negative outcomes such as over-reporting and increased expenditure on manual compliance checking.

This paper examines some of these problems using a semiotic framework that draws attention to the delicate interplay of technical and social factors in the search for both efficiency and effectiveness.

We compare the intelligence building process to an act of communication between three levels that are very different in nature: 1) the technical level that captures and manipulates the data, 2) the informal level of human interactions in the process of giving meaning to the data provided by the technical level and acting upon this knowledge and 3) the formal level of the organisation where these agents participate.

We argue that the process of creating and disseminating intelligence is not the discovery of an 'objective' truth, in which the 'methods' are neutral techniques separated from the personal positioning of the investigator. Rather, we argue, 'subjectivity matters'. The personal and organizational circumstances surrounding the agent, including perceptions and organizational values may influence how specific intelligence is perceived and communicated. When organizations make decisions, they must take into account that different perceptions are brought in by different interested persons.

Additionally, we argue that instead of the prevailing assumption about a simple flow of information from financial institution to Financial Intelligence Unit and prosecuting agency, this approach suggests that the identification and communication of suspicion is a more complex and social process whose nature should be better understood.

The paper addresses the social aspects of the creation and use of information, in the context of the fight against money laundering and terrorist financing. We suggest that, instead of merely adding to the battery of technology and regulation that already exists in this field, financial crime investigation efforts might find profit in

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 600 – 601, 2005.

considering social and pragmatic aspects of intelligence building. The present approach is informed by organizational semiotics, a theoretical framework that addresses the creation of meaning in an organizational setting. We adopt the subjectivist paradigm, which assumes that reality is created subjectively and socially, leading to subtle differences between groups of knowing agents. Additionally, we consider that 'meaning' is the relationship between a sign and some pattern of action, and that this relationship is established as a norm within a given group.

The implication for the researcher in the fields of communication and information systems is that language and linguistic labels are not only descriptive tools, but also constructive ones: the use of language can have the effect of constructing or altering the social world, and can create new states of affairs perhaps just through the deployment of a sign. To study the social world, the researcher must first understand the way in which the members of a society create, modify and interpret the world to which they belong. The agent is responsible for the existence of a sign and its meaning.

A case study of collection, analysis and dissemination of information regarding a suspected case of money laundering underpins an analysis of how social factors interact with technology, in the decision making process regarding the classification of specific patterns of financial behaviour as suspicious of money laundering. We describe the application of the semiotic framework to the development of the identity of a criminal, drawing on interviews with financial investigators operating in financial intelligence units and law enforcement agencies in Europe.

Application of Latent Semantic Indexing to Processing of Noisy Text

R.J. Price[1] and A.E. Zukas[2]

[1] Content Analyst LLC, Reston, VA
[2] Science Applications International Corporation, Reston, VA
rprice@contentanalyst.com, zukasa@saic.com

Latent semantic indexing (LSI) is a robust dimensionality-reduction technique for the processing of textual data. The technique can be applied to collections of documents independent of subject matter or language. Given a collection of documents, LSI indexing can be employed to create a vector space in which both the documents and their constituent terms can be represented. In practice, spaces of several hundred dimensions typically are employed. The resulting spaces possess some unique properties that make them well suited to a range of information-processing problems. Of particular interest for this conference is the fact that the technique is highly resistant to noise. Many sources of classified text are still in hardcopy. Conversion of degraded documents to electronic form through optical character recognition (OCR) processing results in noisy text and poor retrieval performance when indexed by conventional information retrieval (IR) systems. The most salient feature of an LSI space is that proximity of document vectors in that space is a remarkably good surrogate for proximity of the respective documents in a conceptual sense. This fact has been demonstrated in a large number of tests involving a wide variety of subject matter, complexity, and languages. This feature enables the implementation of high-volume, high-accuracy automatic document categorization systems. In fact, the largest existing government and commercial applications of LSI are for automated document categorization. Previous work [1], has demonstrated the high performance of LSI on the Reuters-21578 [2] test set in comparison to other techniques. In more recent work, we have examined the ability of LSI to categorize documents that contain corrupted text. Testing using the Reuters-21578 test set demonstrated the robustness of LSI in conditions of increasing document degradation. We wrote a Java class that degraded text in the test documents by inserting, deleting, and substituting characters randomly at specified error rates. Although true OCR errors are not random, the intent here was simply to show to what extent the text of the documents could be degraded and still retain useful categorization results. Moreover, the nature of comparisons in the LSI space is such that random errors and systematic errors will have essentially the same effects. These results are extremely encouraging. They indicate that the categorization accuracy of LSI falls off very slowly, even at high levels of text errors. Thus, the categorization performance of LSI can be used to compensate for weaknesses in optical character recognition accuracy. In this poster session we present results of applying this process to the much newer (and larger) Reuters RCV1-v2 categorization test set [3]. Initial results indicate that the technique provides robust noise immunity in large collections.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 602 – 603, 2005.

References

1. Zukas, A., and Price, R. *Document Categorization Using Latent Semantic Indexing*. In Proceedings: 2003 Symposium on Document Image Understanding Technology, Greenbelt MD, April 2003 87-91.
2. Lewis, D. Reuters-21578 Text Categorization Test Collection. Distribution 1.0. README file (version1.2). Manuscript, September 26, 1997. http://www.daviddlewis.com/resources/testcollection/reuters21578/readme.html.
3. Lewis, D., Yang, Y., Rose, T., Li, F. RCV1: A New Benchmark Collection for Text Categorization Research. Journal of Machine Learning Research, 5(2004):361-397, 2004. http://www.jmlr.org/papers/volume5/lewis04a/lewis04a.pdf.

Detecting Misuse of Information Retrieval Systems Using Data Mining Techniques

Nazli Goharian, Ling Ma, and Chris Meyers

Information Retrieval Laboratory, Illinois Institute of Technology
goharian@iit.edu

Misuse detection is often based on file permissions. That is, each authorized user can only access certain files. Predetermining the mapping of documents to allowable users, however, is highly difficult in large document collections. Initially [1], we utilized information retrieval techniques to warn of potential misuse. Here, we describe some data mining extensions used in our detection approach.

Initially, for each user, we obtain a profile. A system administrator assigns profiles in cases where allowable task vocabularies are known a priori. Otherwise, profiles are generated via relevance feedback recording schemes during an initial proper use period. Any potential misuse is then detected by comparing the new user queries against the user profile. The existing system requires a manual adjustment of the weights emphasizing various components of the user profile and the user query in this detection process. The manual human adjustment to the parameters is a cumbersome process. Our hypothesis is: **Data mining techniques can eliminate the need for the manual adjustment of weights without affecting the ability of the system to detect misuse.** The classifier learns the weights to be placed on the various components using the training data. Experimental results demonstrate that using classifiers to detect misuse of an information retrieval system achieves a high recall and acceptable precision without the manual tuning.

Our test data contained 1300 instances, each assessed by four Computer Science graduate students. We ran a 10-fold cross validation using the commonly available freeware tool, WEKA on classifiers such as support vector machine (SMO), neural network (MLP), Naïve Bayes Multinomial (NB), and decision tree (C4.5). The misuse detection systems used throughout our experimentation are based on the nature of the user query length. That is, in different applications the user queries may be short (Title) or longer (Descriptive). Thus, we considered the following systems: 1) short queries are used for building profile and detection (T/T); 2) long queries are used for building profile and detection (D/D); and 3) long queries are used for building profile and short queries are used for detection (D/T). For each system setup, we chose top M=10, 20, 30 feedback terms from top N=5, 10, 20 documents, based on BM25 term weighting. The distribution of the a priori known class labels are 40.9% "Misuse", 49.3% "Normal Use", and 8.7% "Undecided". "Undecided" cases are the cases that the human evaluators were unable to determine otherwise. The pool of queries creating the instances contains 100 TREC 6-7 Title and Descriptive ad hoc topics. The

[1] Ling Ma, Nazli Goharian, *Query Length Impact on Misuse Detection in Information Retrieval Systems*, ACM Symposium on Applied Computing, Santa Fe, NM, March 2005.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 604 – 605, 2005.

disks 4-5 2GB collection was used. Unfortunately, there is no standard benchmark to use in evaluating misuse detection systems. Thus, we had to build our own benchmark. We evaluate the accuracy of misuse detection using both *Precision* (correctly detected misuse/detected as misuse); and *Recall* (correctly detected misuse/total misuse).

Precision (%)		**Title Build & Title Detect (T/T)**					**Desc. Build & Desc. Detect (D/D)**					**Desc. Build & Title Detect (D/T)**				
N	***M***	**MA**	**SMO**	**MLP**	**NB**	**C4.5**	**MA**	**SMO**	**MLP**	**NB**	**C4.5**	**MA**	**SMO**	**MLP**	**NB**	**C4.5**
5	***10***	68	70	71	67	73*	69	69	70	68	73*	69	64	71	62~	72*
	20	69	70	71	67	72+	69	69	71	67	71	69	70	71	62~	73
	30	69	69	70	69	72	68	69	71	67	71	70	70	70	62~	73
10	***10***	70	71+	72+	68	73+	69	69	71	67-	70	70	70	70	63~	71
	20	70	69	72	69	73	70	67	70	67	70	71	71	72	62~	72
	30	70	69	73	69	72	71	67~	71	67~	72	71	71	71	62~	71+
20	***10***	70	71	72	69	73	70	67~	71	67-	72	71	70	73	63~	74
	20	71	69	72	69	72	71	67~	71	69~	71	72	71	72	63~	73
	30	71	70	73	69	72	72	67~	71	68~	70	72	71	72	64~	72

Recall (%)		**Title Build & Title Detect (T/T)**					**Desc. Build & Desc. Detect (D/D)**					**Desc. Build & Title Detect (D/T)**				
N	***M***	**MA**	**SMO**	**MLP**	**NB**	**C4.5**	**MA**	**SMO**	**MLP**	**NB**	**C4.5**	**MA**	**SMO**	**MLP**	**NB**	**C4.5**
5	***10***	98	97	96	99	95~	97	98	96	98	94	95	97	96	99*	95
	20	98	97	95~	99	94~	97	98	96	98	94	95	96	95	99*	93
	30	97	98	95	98	95	97	98	96	98	95	94	97+	95	99*	94
10	***10***	98	95~	95-	98	94~	95	98	96	96	94	95	97	95	99*	94
	20	98	97	95~	97	96	96	99+	95	97	95	93	95+	95	99*	95
	30	98	98	95-	98	95-	95	99+	95	98	93	92	98*	95	99*	94
20	***10***	98	96	95~	97	95-	94	99*	95	97+	94	94	98	94	99*	91
	20	97	98	95	97	94	93	99*	94	98*	93	92	97*	94	99*	94
	30	95	98	94	97	93	91	99*	93	97*	93	90	97*	94	99*	93+

We illustrate the precision and recall of four classifier based (SMO, MLP, NB and C4.5) and our baseline, manually adjusted (MA) detection system. To systematically compare, 10 trials of 10-fold cross-validated paired T-test of classifiers versus our MA baseline were conducted over the precision and recall of the "Misuse" class. In the tables shown, statistically significant entries at the 0.01 and 0.05 significance level are designated +/- and */~, respectively. Markers - and ~ indicate that the manual adjustment performed better than the classifiers. All entries without a marker are statistically equivalent. As the results demonstrate for each of the systems T/T, D/D, D/T, there is always a classifier that performs statistically equivalent to or better than the manual adjustment approach, eliminating the need for manual intervention. Examples of such are SMO and NB for T/T, MLP and C4.5 for D/D and D/T in regards to both Precision and Recall. Furthermore, some classifiers such as NB for D/T favor recall over precision and vice versa in the case of C4.5 in T/T. Hence, depending on the application and organization, a classifier can be chosen that optimizes either recall or precision over the other.

Mining Schemas in Semistructured Data Using Fuzzy Decision Trees

Wei Sun and Da-xin Liu

College of Computer Science and Technology, HARBIN Engineering University,
HARBIN Heilongjiang Province, China
sunwei78@hrbeu.edu.cn

1 Summary

As WWW has become a huge information resource, it is very important for us to utilize this kind of information effectively. However, the information on WWW can't be queried and manipulated in a general way.

Semi-structured data differ from structured data in traditional databases in that the data are not confined by a rigid schema which is defined in advance, and often exhibit irregularities. To cope with the irregularities of semi-structured data, regular path expressions are widely used in query languages for semi-structured systems. However, the lack of schema poses great difficulties in query optimization. Based on automata equivalence, DataGuide is proposed [1]. It is a concise and accurate graph schema for the data graph in terms of label paths. In [2], several heuristic-based strategies are proposed to build approximate DataGuide by merging "similar" portions of the DataGuide during construction. However these strategies are situation specific. [3] and [4] aim to type objects in the data set approximately by combining objects with similar type definitions, instead of constructing an approximate structural summary for the data graph.

Our method proposed to construct an approximate graph schema by clustering objects. Our approach has the following unique features. No predetermined parameters are required to approximate the schema. It is cheap to construct and maintain the resultant schema and it is small in size.

2 Mining Schemas by Fuzzy Decision Tree

The goal of this work is to be able to approximately type a large collection of semi-structured data efficiently. First, the schema of semi-structured data is not precision and it is a character of semi-structured data. So the schema is not precision that we are mining from semi-structured data. We use fuzzy decision tree to decide the schema of the clusters of data. In the second, the K-cluster is widely used in mining schema and the number of clusters K requires to be appointed according to persons' experience, and it is difficult to determine the proper type. The third and final, the defect of the mining schema includes two patterns of excess and deficit. In the case of relational and object data that are very regular and with the proper typing program, we obtain a perfect classification of the objects. We proposed three definitions to use fuzzy decision tree method. They are Including-type, Included-type and comparability. Includ-

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 606 – 607, 2005.

ing-type, Cluster K_i includes type set of $\tau_{i_1}, \tau_{i_2}, \ldots, \tau_{i_n}$. If type τ includes label l and label l is present in every type τ_{i_j} ($1 \le j \le n$), type τ is a including-type of K_i. Denoted by including-type (K_i)= τ. Included-type, Cluster K_i includes type set of $\tau_{i_1}, \tau_{i_2}, \ldots, \tau_{i_n}$. If type τ includes label l and label l is present in not less than one of type τ_{i_j} ($1 \le j \le n$), type τ is an including-type of K_i. Denoted by included-type (K_i)= τ. Comparability, The comparability can be defined by a ratio, The change from an instance to type τ is a× the number of adding labels + b× the number of reducing labels, a and b are power parameters. The ratio $=1-\frac{change(ins\tan ce_i)}{totaloflabel(ins\tan ce_i)}$.

3 Discussion

In this paper, we proposed an approximate schema based on an incremental clustering method and fuzzy decision tree method. Our method showed that for different data graphs the sizes of approximate schemas were smaller than that of accurate schemas. Also, for regular path expression queries, query optimization with the approximate graph schema outperformed the accurate graph schema. In the near future, we plan to explore various approximate schemas and conduct an extensive performance study in order to evaluate their effectiveness.

References

1. R. Goldman and J. Widom. Dataguides: Enabling query formulation and optimization in semistructured databases. In Proceedings of the 23rd International Conference on Very Large Data Bases,(1997)
2. R. Goldman and J. Widom. Approximate dataguides. Technical report, Stanford University, (1998)
3. S. Nestorov, S. Abiteboul, and R. Motwani. Inferring structure in semistructured data. In Proceedings of the Workshop on Management of Semistructured Data,(1997)
4. S. Nestorov, S. Abiteboul, and R.Motwani. Extracting schema from semistructured data. In Proceedings of ACM SIGMOD International Conference on Management of Data, (1998)

More Than a Summary: Stance-Shift Analysis

Boyd Davis[1],Vivian Lord[1], and Peyton Mason[2]

[1] University of North Carolina-Charlotte
bdavis@email.uncc.edu
[2] Linguistic Insights, Inc., Charlotte, NC

1 Introduction

Our corpus-based, multivariate approach to the analysis of text is keyed to the interaction of two dozen language features that have shown potential for assessing affect, agency, evaluation and intention. They include public and private or perceptual verbs; several categories of adverbs, modals, pronouns and discourse markers. The language features, which we use as variables, make up what text and corpus linguists define as *stance*. Stance is how people use their words to signal confidence or doubt, appraisal or judgment [1] about topics, values, audiences, situations, or viewpoints. Speakers construct different personae out of the ways they use language features that signal modality, evidentiality, hedging, attribution, concession, or consequentiality. In depositions, interviews or interrogations, those personae often include:

- The one who claims s/he has no recall or knowledge of a situation or its details;
- The one who is strongly committed to what he or she is saying;
- The one who is not confident about what s/he is saying, and is back-pedaling;
- The one who can appropriate agency to self or ascribe it or actions to others.

We detect the speakers' positioning among these personae by tracking their shifts in stance.

Speakers constantly shift verbal footing, positioning, and style as they talk. We identify, track and measure the language patterns an individual speaker uses to express evidentiality, hedging, attribution, or concession, relative to the ways that person talks throughout a particular interaction. Stance-shift analysis shows where speakers in an interrogation, interview or deposition use words to wince, blink, stall or wiggle. It identifies where questioners not only *could* ask more questions, they probably *should.* We have applied it to depositions and pre-trial interviews in civil and criminal cases, and to online chats, e-conferences, and focus groups.

2 Stance-Shifting: The Technique

Corpus linguists often work with 6 to 7 dozen variables to characterize spoken and written registers; elsewhere, we discuss our testing and selection of the two dozen that identify stance[2]. We divide the transcript into questions and answers, standardize the transcript of answers into successive 100-word sections for transcripts over 1000 words (50-word sections for shorter transcripts), and code sequentially:

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 608 – 609, 2005.

1. Obtain frequency counts for language features in each segment, weighted by the factor scores for type or genre of text being analyzed.
2. Convert the results of the coding into scale scores: the factor score coefficients are used as the scales to measure stance.
3. Use significant scale scores, which are 1 or more standard deviations above the mean scale score for the transcript under analysis, to identify and interpret the significant sections of heightened interest within the transcript.
4. Reinsert into question sequence; reinterpret.

3 An Example from a Deposition

In depositions, the answers are contingent, the interaction between the lawyer and the deponent always emergent. This excerpt is from an insurance case. Errors had been made on policies for property which suddenly burned: to what extent were the errors due to paper-shuffles as opposed to something less accidental? The agent seeks to present himself as competent and no longer socially connected to the claimants. Section 9 was scored above the speaker's mean for both Scale 1 and Scale 2: that is, it coupled more than the agent's usual amount of elaborative details (Scale 2) with his claims about thinking and knowing (Scale 1). As he talks, he shifts stance to signal in Section 9 both his increasing discomfort (signaled by his continued efforts to monitor the listener's understanding) and an underlying concern about arson. We have reinserted the sections into the question answer sequence, and boldfaced Section 9.

Q: Do you recall what he said about that?
A: No, I don't, I don't.

Q: What else, if anything, do you remember about that conversation with Husband on the 28th of December, 2000?
A: *I just remember thinking the whole thing from the very beginning was -- you know, was strange, I mean, and very bizarre, because like I say, I had knowledge that it was going, you know, in a hostile manner. From talking -- you know, you know, with both of them, I recall them making all these other different changes and things, you know, and wondering and really kind of trying to disseminate from talking with him if -- you know, where he was coming from, you know -- you know, what his –* (9) *you know, if he was guilty of anything, you know, in the process. I mean, I was -- and to this day, I mean, there again, there was a separate rental house that burned and within a day or two of when this one burned with no insurance on it.*

References

1. Biber, D. and E. Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text 9:93-125.
2. Mason, P., B. Davis and D. Bosley. 2005 in press. Stance analysis: when people talk online. In Innovations in E-Marketing, II, ed. S. Krishnamurthy. Hershey: PA: Idea Group.

Principal Component Analysis (PCA) for Data Fusion and Navigation of Mobile Robots

Zeng-Guang Hou

Laboratory of Complex Systems and Intelligence Science, Institute of Automation, The Chinese Academy of Sciences, Beijing 100080, P.R. China
hou@compsys.ia.ac.cn

1 Introduction

A mobile robot system usually has multiple sensors of various types. In a dynamic and unstructured environment, information processing and decision making using the data acquired by these sensors pose a significant challenge. Kalman filter- based methods have been developed for fusing data from various sensors for mobile robots. However, the Kalman filter methods are computationally intensive. Markov and Monte Carlo methods are even less efficient than Kalman filter methods. In this paper, we present an alternative method based on principal component analysis (PCA) for processing the data acquired with multiple sensors.

Principal component analysis (PCA) is a procedure that projects input data of larger dimensionality onto a smaller dimensionality space while maximally preserving the intrinsic information in the input data vectors. It transforms correlated input data into a set of statistically independent features or components, which are usually ordered by decreasing information content. The learning rule for PCA is basically non-supervised and was first proposed by Oja in 1982. More advanced learning algorithms were also proposed for PCA including a neural network based approach called PCA network (PCANN). In this paper, we present a PCA network scheme for the sensor information processing of a mobile robot.

2 Principal Component Analysis (PCA) Based Data Fusion and Navigation Scheme

A core step of PCA analysis is to identify a linear orthogonal transformation ($y = Wx$) for a given input vector $x \in R^n$ with zero-mean and covariance C, and a specified dimension of the output vector $y \in R^m$ such that the covariance matrix of y is given by $\Lambda = \mathrm{diag}\{\lambda_1, \cdots, \lambda_m\}$, where $\lambda_1 \geq \cdots \geq \lambda_m$ are the eigenvalues of C and the rows of $W = \left[w_1, \cdots, w_m\right]^T \in R^{m \times n}$ are the corresponding eigenvectors. Some algorithms can analytically compute PCA but they have to solve matrix equations. An alternative, more efficient method using neural networks was proposed for implementing PCA on-line with local learning rules. In this approach, the PCA network (PCANN) is given by $y_i(k) = \sum_{j=1}^{n} w_{ij}(k) x_j(k)$, $i = 1, \ldots, m$, and the weights are updated according to the following $\Delta w_{ij}(k) = \eta y_i(k) \left[x_j(k) - \sum_{l=1}^{i} w_{lj}(k) y_l(k)\right]$, $\eta > 0$.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 610–611, 2005.

We use the PCA network to process the data acquired with the ultrasonic and infrared sensors, and further incorporate the processed results with the information acquired with CCD, odometer and compass to facilitate final decisions on the robot navigation strategy. We define the ultrasonic data as $x_{ultasonic} = \left[s_{u1}^T, s_{u1}^T, \cdots, s_{u16}^T\right]^T$, where s_{ui} denotes the data array acquired with the ith ultrasonic sensors. Similarly, we can define the data arrays acquired with the infrared sensors. To remove or attenuate disturbances and noises, the Kalman filter is used to preprocess the original data. The data arrays of the ultrasonic and infrared sensors constitute the high dimensional input vector to the PCA neural network, and the output vector is a lower dimensional one. The first principal component indicates a feasible direction for the mobile robot. Based on the PCA network, we developed a hierarchical architecture for data fusion and navigation of the mobile robot with multiple sensors.

Acknowledgments: This research is supported by the National Natural Science Foundation of China and the National Basic Research Program of China.

References

1. Cichocki, A., Unbehauen, R.: Neural Networks for Optimization and Signal Processing. John Wiley & Sons, Chichester (1993)
2. Gutmann, J.S., Fox, D.: An Experimental Comparison of Localization Methods Continued. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. (2002) 454-459

BioPortal: Sharing and Analyzing Infectious Disease Information[1]

Daniel Zeng[1], Hsinchun Chen[1], Chunju Tseng[1], Catherine Larson[1], Wei Chang[1], Millicent Eidson[2], Ivan Gotham[2], Cecil Lynch[3], and Michael Ascher[4]

[1] Department of Management Information Systems,
University of Arizona, Tucson, Arizona
{zeng,hchen,chunju,cal,weich}@eller.arizona.edu
[2] New York State Department of Health, SUNY, Albany
{mxe04,ijg01}@health.state.ny.us
[3] California Department of Health Services, UC Davis, Sacramento
clynch@dhs.ca.gov
[4] Lawrence Livermore National Laboratory
ascher1@llnl.gov

1 Introduction

Infectious disease informatics (IDI) is an emerging field of study that systematically examines information management and analysis issues related to infectious disease prevention, detection, and management. IDI research is inherently interdisciplinary, drawing expertise from a number of fields including but not limited to various branches of information technologies such as data integration, data security, GIS, digital library, data mining and visualization, and other fields such as biostatistics and bioinformatics. Funded by the NSF through its Digital Government and Information Technology Research programs with support from the Intelligence Technology Innovation Center, we have been developing scalable technologies and related standards and protocols, implemented in the BioPortal system, to deal with various data sharing and analysis challenges arising in the field of IDI. BioPortal provides distributed, cross-jurisdictional access to datasets concerning several major infectious diseases (of both humans and animals) *West Nile Virus* (WNV), *Botulism*, *Foot-and-Mouth Disease* (FMD), and *Chronic Wasting Disease* (CWD). It also provides access to test data collected from the BioWatch system, an air pathogen sensing system developed to detect airborne hazards. In addition to data access, BioPortal provides advanced spatial-temporal data visualization and analysis capabilities, which are critically needed for infectious disease outbreak detection and can provide valuable information to facilitate disease outbreak management and response. The intended users of BioPortal include public health researchers and practitioners at all levels of the government including international partners, analysts and policy makers; the general public and students in public health or related fields; and law enforcement and national security personnel involved in counter bioterrorism efforts and emergency preparedness and response.

[1] Reported research has been supported in part by the NSF through Digital Government Grant #EIA-9983304 and Information Technology Research Grant #IIS-0428241.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 612 – 613, 2005.

2 BioPortal System Overview

From a systems perspective, BioPortal is loosely-coupled with state health information systems in that the state systems will transmit disease information through secured links to the portal system using the Public Health Information Network Messaging System (PHINMS) or other similar systems based on XML and HL7. Such information, in turn, is normalized and stored in the internal data store maintained by BioPortal. The data store is used to support various query, reporting, analysis, and visualization functions. BioPortal implements a role-based user access control module to ensure secure and proper use of data. This module is particularly important in IDI applications because of the sensitivity of disease-related data and related privacy regulations and practice, especially in a cross-jurisdictional context. The BioPortal data store is implemented largely using MS SQL server. The majority of data items are internally stored in customized XML formats for flexibility, while a small number of selected data elements are extracted from XML records and stored in indexed tables for access efficiency. All the query and analysis functions are implemented in the Java and JSP environments.

3 Demo Plan

We plan to demonstrate several complete BioPortal use case scenarios to illustrate the Web-enabled process of accessing, querying, and analyzing infectious disease datasets collected from distributed sources. These scenarios will include WNV, FMD, and BioWatch datasets. We also plan to demonstrate the BioPortal's spatial-temporal data analysis and visualization capabilities using real-world datasets. These capabilities are implemented through two modules of BioPortal: (a) a hotspot analysis module that implements existing methods (e.g., scan statistics-based methods) that can detect "unusual" spatial and temporal clusters of events as well as several new methods using Machine Learning techniques, and (b) a visualization tool called Spatial-Temporal Visualizer (STV), which can help the user effectively explore spatial and temporal patterns through an integrated set of tools including a GIS tool, a timeline tool, and a periodic pattern tool.

DIANE: Revolutionizing the Way We Collect, Analyze, and Share Information

Jin Zhu

The Arlington Institute, 1501 Lee Highway, Suite 204, Arlington, VA 22201
jin@arlingtoninstitute.org

Abstract. The international intelligence community is in urgent need of specialized knowledge, tools, models, and strategies to track knowledge of terrorist-related individuals, groups, and activities that could make a significant difference in being able to anticipate, detect, prevent, respond, and recover from major threats and terrorist events. At this demo, we feature a specific suite of tools—The Digital Analysis Environment (DIANE)—and show how this powerful collection and analysis tool set provides intelligence analysts, researchers, and industry consultants and practitioners with the ability to extract open source information, conduct information and data analysis, and identify linkages and relationships between current events and potential future events. This suite of tools falls into what is now considered Terrorism Informatics, a relatively new stream of research using the latest advances in social science methodologies, technologies, and tools.

1 Introduction

Over the last few years, the development of tools that apply information technology to terrorism research as well as intelligence collection and analysis has been flourishing in the field of Informatics research. These tools provide local, national, and international government officials with the means for critical law enforcement and security-related public policy, including the prediction, timely detection, and prevention of terrorist attacks and terrorist behavior. The Digital Analysis Environment (DIANE) is a powerful analysis toolset that provides intelligence analysts, researchers, and industry consultants and practitioners the ability to extract information from large amounts of data and make sense of it. This suite of tools falls into what is now considered Terrorism Informatics, a relatively new stream of research using the latest advances in social science methodologies, technologies, and tools.

2 A Focus on The Digital Analysis Environment (DIANE)

DIANE enables analysts and practitioners to FIND information, PICTURE the relationships, REALIZE the situation, FORESEE the outcomes, and MODEL the possibilities.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 614 – 616, 2005.

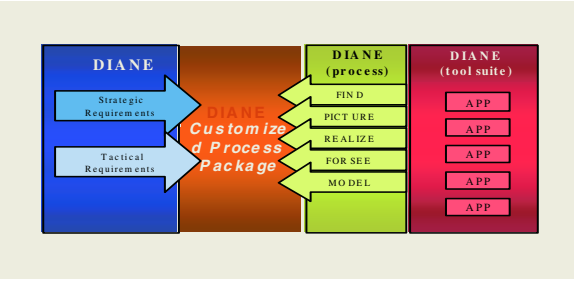


Fig. 1. DIANE's Ability to Find, Picture, Realize, Foresee, and Model

This innovative information and knowledge planning system enables organizations to anticipate futures rather than respond to events. The unique properties of DIANE provide breakthrough capabilities:

- Search, monitor, and linguistically process large amounts of structured and unstructured information.
- Provide enterprise-wide knowledge discovery through sophisticated analytical processes and system and user interaction management
- Provide a diverse array of advanced information visualization aids
- Assess the significance of events extracted from data sets as they are captured at the earliest point of "consumability"
- Transform analytical processes throughout the enterprise from reactive analysis to anticipatory analysis, therefore reducing surprises
- Facilitate organization-wide analytical collaboration via advanced process-driven, secure, decision management
- Seamlessly output intelligence product through Microsoft Office, i2's Analyst Notebook and other commonly used analytical and presentation tools

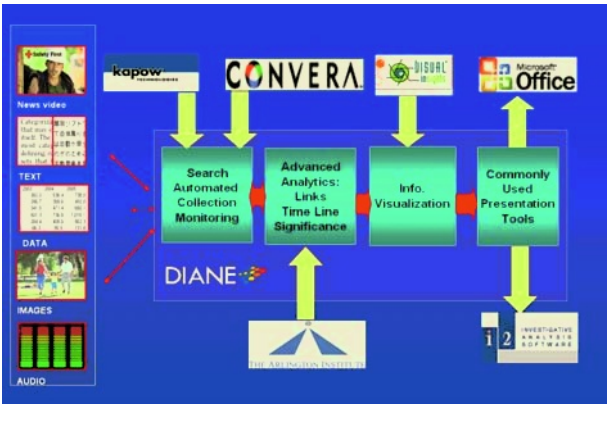


Fig. 2. DIANE's Collection & Analysis Capabilities and Tool Suite

The DIANE system consists of an Integration Engine (DIANE Integration Engine) and a suite of best-of-breed commercial-off-the-shelf (COTS) Web Services components transparently integrated to support advanced, large-data set analysis. The

DIANE Integration Engine adopts Service-Oriented Architecture (SOA) that fully complies with industry-wide and E-Gov initiative Web Services standards. The components selected for the Tool Suite are the result of carefully evaluating leading edge COTS information acquisition and knowledge generation applications, each of which provides best-of-breed functionality.

3 Conclusion

The conceptual and structural design of DIANE places it at the leading edge of tools that are defining the next generation of advanced IT solutions to be utilized in a variety of areas and sectors, including in the area of Terrorism Informatics. These kinds of tools are critical in our ability to understand and track knowledge of terrorist-related individuals and activities as well as enhance our ability to respond to pressing national security issues.

Processing High-Speed Intelligence Feeds in Real-Time

Alan Demers, Johannes Gehrke, Mingsheng Hong, and Mirek Riedewald

Department of Computer Science, Cornell University
{ademers, johannes, mshong, mirek}@cs.cornell.edu

Intelligence organizations face the daunting task of collecting all relevant pieces of information and to draw conclusions about potential threats in a timely manner. Typical information sources range from news tickers, financial transaction logs and message logs to satellite images and speech recordings. This wealth of data is continuously updated and arrives in high-speed data streams; it needs to be analyzed both in real-time (e.g., to estimate the importance of the information and to generate early threat alerts) and offline by sophisticated data mining tools. This work focuses on the real-time aspects of processing these massive streams of intelligence data. We also show how real-time and data mining components can interact effectively.

Current database and data warehousing technology supports expressive queries, but is not designed for real-time processing of data streams [1]. At the other end of the spectrum are publish/subscribe (pub/sub) systems. They can support very high throughput even for millions of active subscriptions [2], but have only limited query languages. It is not possible to search for patterns involving multiple incoming messages and across streams.

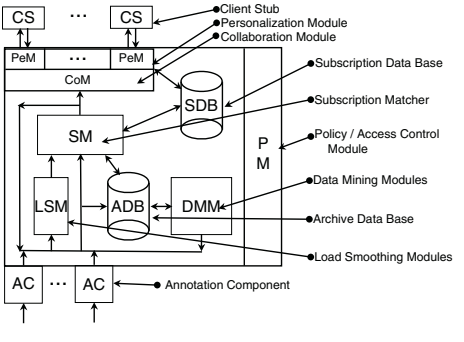


Fig. 1. Cornell Knowledge Broker

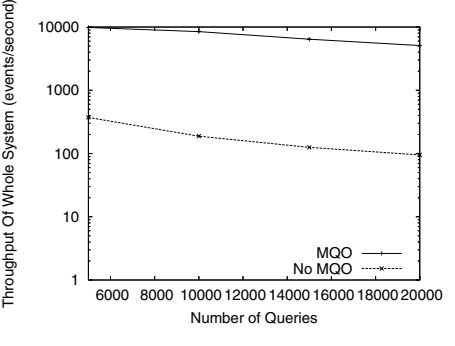


Fig. 2. Throughput vs. number of active monitoring queries

We are building the Cornell Knowledge Broker (CKB), a system for processing high-speed streams of intelligence data. The architecture of the CKB provides a unique combination of new technology for data stream processing, data mining, and privacy/access control (see Figure 1). The *Subscription Matcher (SM)*

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 617–618, 2005.

component lies at the heart of the CKB.[1] The SM module manages expressive monitoring queries, which are continuously evaluated for all incoming messages. As soon as a relevant pattern is detected, it generates an alert message, which in turn is either processed by another query or directly forwarded to an analyst. We have developed a powerful query language, which enables analysts to search for complex patterns in the incoming data.

Let us illustrate the functionality of the SM module through one example query: "Notify me whenever Alice passes sensitive information to Bob, either directly or through several intermediate parties". The SM module monitors all incoming messages, building up internal data structures that capture the relevant network of information flow. As soon as a path with sensitive information flow between Alice and Bob has been established, an analyst is notified. The sensitivity of a message could be determined based on simple keywords or by using a sophisticated text classification approach [4]. Note that it is easy to see through this query that the expressiveness of the SM module goes far beyond the simple predicate filtering of current pub/sub technology. The SM module also supports queries that can search for patterns like a sequence of phenomena that indicate the outbreak of an epidemic and many others.

We have built a running prototype of the SM module, which can effectively interact with text analysis modules. The system has an easy-to-learn user interface, enabling users to formulate sophisticated monitoring queries in a high-level language. Our system automatically detects and exploits commonalities between different monitoring queries. Figure 2 shows the throughput (number of relevant events or messages per second) with varying number of concurrently active monitoring queries. Notice the logarithmic scale of the y-axis. The MQO curve is for our system, while "no MQO" is for a traditional implementation where the query functionality is located in an application server.

References

1. Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Stonebraker, M., Tatbul, N., Zdonik, S.: Monitoring streams — a new class of data management applications. In: Proc. Int. Conf. on Very Large Databases (VLDB). (2002)
2. Fabret, F., Jacobsen, H.A., Llirbat, F., Pereira, J., Ross, K.A., Shasha, D.: Filtering algorithms and implementation for very fast publish/subscribe. In: Proc. ACM SIGMOD Int. Conf. on Management of Data. (2001) 115–126
3. Demers, A., Gehrke, J., Riedewald, M.: The architecture of the cornell knowledge broker. In: Proc. Second Symposium on Intelligence and Security Informatics (ISI-2004). (2004)
4. Joachims, T.: Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms. Kluwer (2002)

[1] The functionality of the other components is indicated by their names, details can be found in our previous work [3].

Question Answer TARA: A Terrorism Activity Resource Application

Rob Schumaker and Hsinchun Chen

Artificial Intelligence Lab,
Department of Management of Information Systems,
The University of Arizona,
1130 E. Helen St. Tucson, AZ 85721
{rschumak, hchen}@email.arizona.edu

1 Introduction

Terrorism research has lately become a national priority. Researchers and citizens alike are coming to grips with obtaining relevant and pertinent information from vast storehouses of information gateways, dictionaries, self-authoritative websites, and sometimes obscure government information. Specific queries need to be manually sought after, or left to the mercy of search engines that are generally scoped.

In our poster we demonstrate a possible solution in the form of a narrowly focused Question Answer system, TARA, which is based upon successful ALICE chatterbot which has won the Loebner contest for most human computer in 2000, 2001, and 2004.

2 Terrorism Activity Resource Application (TARA)

TARA is a type of shallow fact-driven Question Answer system that performs syntactic parsing of user input and can convey an 'expert appearance' in narrow knowledge domains [2]. Terrorism-specific definitional information is collected from various reputable websites and incorporated into the knowledge banks of TARA. Thus asking the system about Zyklon B, or Ricin will return valuable information to the user.

3 Testing TARA: Experimental Results

From experiments on the TARA system, we discovered that interrogative-style input is most prevalently used, which coincides with the findings of Moore and Gibbs whom discovered that chatterbots are often used as simple search engines [1]. In particular, participants were most interested in obtaining definitional information, however, they desired such information in a conversational context. User input beginning with 'What' was the most used interrogative, and surprisingly enough, sentences beginning with 'Are' provided the most satisfactory chatterbot responses. It was also discovered that providing terrorism domain information along with general conversational knowledge increased user satisfaction levels as well.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 619 – 620, 2005.

4 Conclusions

The main contribution of the TARA system is to provide a way to mitigate many of the current information overload problems present in field of terrorism research. Providing definitional responses to natural language queries could prove to be a valuable aid in contrast to existing search strategies.

Acknowledgements

This work was supported in part by the NSF, ITR: "COPLINK Center for Intelligence and Security Informatics Research" Sept. 1, 2003 – Aug. 31, 2005.

References

1. Moore, R. and Gibbs, G.., Emile: Using a chatbot conversation to enhance the learning of social theory, Univ. of Huddersfield, Huddersfield, England, 2002.
2. Wallace, R.S. The Anatomy of A.L.I.C.E. in A.L.I.C.E. Artificial Intelligence Foundation, Inc., 2004.

Template Based Semantic Similarity for Security Applications

Boanerges Aleman-Meza, Christian Halaschek-Wiener[1], Satya Sanket Sahoo, Amit Sheth, and I. Budak Arpinar

Large Scale Distributed Information Systems (LSDIS) Lab,
Department of Computer Science,
University of Georgia, Athens, GA. 30602-7404
halasche@cs.umd.edu,
{boanerg, sahoo, amit, budak}@cs.uga.edu
http://lsdis.cs.uga.edu/

Today's search technology delivers impressive results in finding relevant documents for given keywords. However many applications in various fields including genetics, pharmacy, social networks, etc. as well as national security need more than what traditional search can provide. Users need to query a very large knowledge base (KB) using semantic similarity, to discover its relevant subsets. One approach is to use templates that support semantic similarity-based discovery of suspicious activities, that can be exploited to support applications such as money laundering, insider threat and terrorist activities. Such discovery that relies on a semantic similarity notion will tolerate syntactic differences between templates and KB using ontologies. We address the problem of identifying known scenarios using a notion of template-based similarity performed as part of the SemDIS project [1, 3]. This approach is prototyped in a system named TRAKS (Terrorism Related Assessment using Knowledge Similarity) and tested using scenarios involving potential money laundering.

A *template* provides a means to represent a specific manner in which collection of entities are interconnected thus capturing a scenario or a set of circumstances of interest in security applications. The template is defined using classes and relationships of an ontology, forming a 'typed' directed graph. In terms of information retrieval, a template can be viewed as a query. Querying requires data to match the classes and the interconnections of the named relationships of the template. However, our approach exploits inheritance hierarchies in ontologies to detect similarities semantically. Computing similarity involves looking at syntactical, structural, and semantic properties of instance data with respect to the template.

Our work is aligned with the current semantic Web vision where ontologies play a central role. We used SWETO [2] as our dataset because it includes entities and relationships of relevance to security applications (e.g., banks, organizations, persons, watch lists). Known money laundering scenarios were described as templates to evaluate our approach. The results are ranked based on how close the types of entities or relationships are to those in the template. A graph-based visualization, based on TouchGraph, provides support better understanding of the results.

[1] Currently a Ph.D. student in the Computer Science Dept. at University of Maryland.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 621 – 622, 2005.

It is possible to test our prototype with available datasets and ontologies (using W3C's OWL recommendation). Figure 1 illustrates the architecture of TRAKS. Both the Web application and a technical report are available online[2].

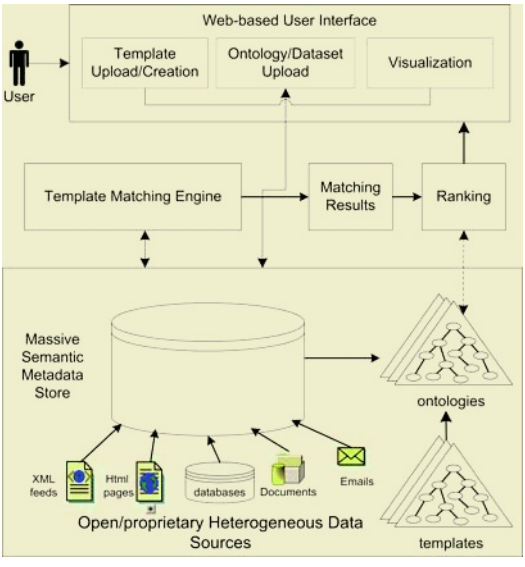


Fig. 1. TRAKS System Architecture

Acknowledgements. This work is funded by NSF-ITR-IDM Award#0325464 titled 'SemDIS: Discovering Complex Relationships in the Semantic Web' and NSF-ITR-IDM Award#0219649 titled 'Semantic Association Identification and Knowledge Discovery for National Security Applications.'

References

1. B. Aleman-Meza, C. Halaschek, I.B. Arpinar, A. Sheth, Context-Aware Semantic Association Ranking. Proceedings of Semantic Web and DB Workshop, Berlin, 2003, pp. 33-50
2. B. Aleman-Meza, C. Halaschek, A. Sheth, I.B. Arpinar, and G. Sannapareddy. SWETO: Large-Scale Semantic Web Test-bed. Proceedings of the 16th International Conference on Software Engineering and Knowledge Engineering (SEKE2004): Workshop on Ontology in Action, Banff, Canada, June 21-24, 2004, pp. 490-493
3. A. Sheth, B. Aleman-Meza, I.B. Arpinar, C. Halaschek, C. Ramakrishnan, C. Bertram, Y. Warke, D. Avant, F.S. Arpinar, K. Anyanwu, and K. Kochut. Semantic Association Identification and Knowledge Discovery for National Security Applications. Journal of Database Management, Jan-Mar 2005, 16 (1):33-53

[2] http://lsdis.cs.uga.edu/proj/traks/

The Dark Web Portal Project: Collecting and Analyzing the Presence of Terrorist Groups on the Web

Jialun Qin[1], Yilu Zhou[1], Guanpi Lai[2], Edna Reid[1], Marc Sageman[3], and Hsinchun Chen[1]

[1] Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721, USA
{ednareid, qin, yiluz, hchen}@bpa.arizona.edu
[2] Department of Systems and Industry Engineering, The University of Arizona, Tucson, AZ 85721, USA
guanpi@email.arizona.edu
[3] The Solomon Asch Center For Study of Ethnopolitical Conflict, University of Pennsylvania, St. Leonard's Court, Suite 305, 3819-33 Chestmut Street, Philadelphia, PA 19104, USA
sageman@sas.upenn.edu

1 Introduction

While the Web has evolved to be a global information platform for anyone to use, terrorists are also using the Web to their own advantages. Many terrorist organizations and their sympathizers are using Web sites and online bulletin boards for propaganda, recruitment and communication purposes. This alternative side of the Web, which we call the Dark Web, could be analyzed to enable better understanding and analysis of the terrorism phenomena. However, due to problems such as information overload and language barrier, there has been no general methodology developed for collecting and analyzing Dark Web information. To address these problems, we developed a Web-based knowledge portal, called the Dark Web Portal, to support the discovery and analysis of Dark Web information. Specifically, the Dark Web Portal integrates terrorist-generated multilingual datasets on the Web and uses them to study advanced and new methodologies for predictive modeling, terrorist (social) network analysis, and visualization of terrorists' activities, linkages, and relationships.

2 System Components

In this work, we demonstrate the three major components of our project: 1) a comprehensive and high-quality Dark Web collection built through a systematic Web crawling process; 2) the results from in-depth Web content and link analysis conducted on the Dark Web collection; and 3) a intelligent Web portal with advanced post-retrieval analysis and multilingual support functionalities to help experts easily access and understand the information in the Dark Web collection.

2.1 Dark Web Collection Building

We have developed a systematic procedure to collect multimedia and multilingual Web contents that were created by terrorist groups from all over the world. Following

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 623 – 624, 2005.

this procedure, we have collected three batches of Dark Web contents since April 2004. Our collection is in multiple languages (English, Arabic, and Spanish) and contains textual documents (HTML, TXT, MS Word, PDF, etc) as well as multimedia documents (images, audio/video files, etc). Our collection provides a unique and valuable data source for terrorism experts and researchers to pursue various types of analyses.

2.2 Content and Link Analysis on the Dark Web Collection

Analyzing terrorist Websites could help us gain valuable knowledge on the terrorists' ideology, their propaganda plans, and the relationship between different terrorism organizations. We conducted a quantitative content analysis on the Dark Web collection in which domain experts examined the Dark Websites and identified the appearances of a set of attributes in the contents. These attributes reflect various facets of terrorists' Web usage such as ideology sharing and fund raising. We also studied the hyperlinks between the terrorist Websites, which would let us learn the relationships and communications between the terrorist groups. Both the content and link analysis results were visualized for easy access and understanding.

2.3 Dark Web Portal

A Web based intelligent knowledge portal called the Dark Web Portal was built. This portal was integrated with advanced post-retrieval analysis functionalities such as document summarization, document categorization, and visualization. These functionalities could help experts quickly and easily locate the information they want. Cross-language information retrieval and machine translation components were also added to the portal to help English-speaking experts to access the Arabic and Spanish content in the Dark Web collection.

Toward an ITS Specific Knowledge Engine

Guanpi Lai[1] and Fei-Yue Wang[1, 2]

[1] Department of Systems and Industrial Engineering,
University of Arizona,
Tucson, AZ 85721, USA
{guanpi, feiyue}@email.arizona.edu
[2] Chinese Academy of Sciences,
Beijing, 100080, China
feiyue@mail.ia.ac.cn

1 Introduction

New technologies and researches are being developed every day for Intelligent Transportation Systems. How to recognize and maximize the potentials of ITS technologies becomes a big challenge for ITS researchers. Usually people would rely on general search engines like Yahoo!, Google to retrieve related information. The direct problem of these search engines is information overload [1]. Another issue with the search engines is that it's difficult to keep the web pages up-to-date.

To address the problems above, it is highly desirable to develop new integrated, ITS specific knowledge engines which support more effective information retrieval and analysis. In this paper, we describe the process of building an ITS specific knowledge engines using the ASKE approach [2].

2 The ITS Specific Knowledge Engine

Following the system architecture of ASKE [2], we developed an ITS specific knowledge engine. In this knowledge engine, users would not get a long list of Web documents but an organized report that generated based on the preset KCF.

2.1 Data Collection

We adopted vertical searching [5] approach to build the data collection. We identified 112 authoritative ITS related web sites, and collected about 72K web pages.

Another approach to collect ITS related Web pages is meta-searching. We developed a Meta-Searcher module which is responsible for retrieving documents from the other data sources besides the Web sites we mentioned before.

As a result, a collection of around 86,300 documents was built. Afterwards, we used SWISH-E which is an open source tool to index each document and stored the searchable index to our data repository. With high-quality, comprehensive resources, we found that we can easily update the whole collection in a reasonable short period, e.g., 3 to 4 days.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 625 – 626, 2005.

2.2 KCF and Report Generation

The Knowledge Configuration File (KCF) is used to specify topics, keywords, searching sequences and schedules for query processing.

In our system, KCF is user dependent. Users can configure the topics, keywords, searching sequences and schedules easily via the KCF configuration interface.

To present an organized report to users, several functional components were implemented. The Aggregation component is to categorize the search results and organize them as a brief report providing a content overview of the entire result set. The Document List component was designed as usual web search engines. The Keyword Suggestion component was applied in order to provide ITS-related search terms rather than general keywords for researchers to refine their KCF configuration. The Summarization component was provided to users to save their time when they read Web documents (web pages or papers).

3 Conclusion

In this paper, we explored the detail of building an ITS specific knowledge engine, using ASKE approach. ITS researchers will efficiently locate information, analyze the current research trend and generate new ideas with the knowledge engine.

Acknowledgement

This work is supported in part by the National Outstanding Young Scientist Research Fund (6025310) from the NNSF of China.

References

1. Oyama, S., Kokubo, T., Ishida, T.: Domain-Specific Web Search with Keyword Spices. IEEE Transactions on Knowledge and Data Engineering, vol.16, no. 1, pp. 17-27, 2004.
2. Wang, Fei-Yue, Lai, Guanpi, Tang, Shuming: An Application Specific Knowledge Engine for ITS Research. Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems, Washington D.C., USA, October 3-6, 2004.
3. Selberg, E., Etzioni, O.: Multi-service Search and Comparison using the MetaCrawler, Proceedings of the 4th World Wide Web Conference, 1995.
4. Chen, H., Houston, A., Sewell, R., Schatz, B.: Internet Browsing and Searching: User Evaluation of Category Map and Concept Space Techniques. Journal of the American Society for Information Science, Special Issue on AI Techniques for Emerging Information Systems Applications, vol. 49, No. 7, pp. 582-603, 1998.
5. Chau, M., Chen, H., Qin, J., Zhou, Y., Qin, Y., Sung, W. K., McDonald, D.: Comparison of Two Approaches to Building a Vertical Search Tool: A Case Study in the Nanotechnology Domain. In Proceedings of JCDL'02, Portland, OR, USA, July 2002.

A Blind Image Watermarking Using for Copyright Protection and Tracing

Haifeng Li, Shuxun Wang, Weiwei Song, and Quan Wen

Institute of Communication Engineering, Jilin University,
Changchun 130025, China
lhfvip_2000@163.com

With the development and mature of the Internet and wireless communication techniques the copy and distribution of the digital production becomes easier and faster. The copyright protection is an urgent problem to be resolved. The watermark is a digital code unremovably, robustly, and imperceptibly embedded in the host data and typically contains information about origin, status, and/or destination of the data. Most of present watermarking algorithms embed only one watermark, however one watermark is not sufficient under some circumstances. One of the reasons of privacy is that we cannot trace the responsibility of the pirates. It would come true by means of embedding the different exclusive watermarks belong to the issuer. As the different watermarks are needed at the different time, the multiple watermarks algorithm is required. The reports about multiple watermarks scheme are rather few. Cox et al. extend the single watermark algorithm to embed multiple orthogonal watermarks. The disadvantage is that the original image is needed at the watermark detector, and the watermark capacity is small. Stankovic et al. proposed a scheme utilizing the two-dimensional Radon–Wigner distribution. The lack is that the watermark capacity is small and we cannot judge the validity of the extracted watermark directly. Tao et al. present a multiple watermark algorithm in the DWT (Discrete Wavelet Transform) domain. The binary image is used for the watermark, and is embedded into all frequencies of DWT. The shortage is that the algorithm is not blind.

This paper proposed a novel image multiple watermarks algorithm using for the copyright protection and tracing. Different watermarks are embedded into the image at the phase of image production and release to the aim of copyright tracing. Based on the comparison of the characteristics HT (Hadamard Transform), DCT (Discrete Cosine Transform) and DFT (Discrete Fourier Transform), HT domain is selected for watermarking. The watermark adopts the binary image and is visually distinguishable after extraction. Of course the watermark may be the unique number of the copy, i. e. the image copy ID. The superiority of the HT (Hadamard Transformation) offers the feasibility of embedding multiple watermarks. The original image is divided into non-overlapped blocks and HT is applied to each block. To reduce the computation time, we adopt FHT (Fast Hadamard Transform) instead of HT. The sub-blocks are pseudo-randomly selected for watermark embedding under control of the key in order to increase the security of the watermarking system. Scanning the AC coefficients of the selected sub-blocks in Zigzag order and we can obtain a 1D-sequence. Watermark data inserted in high frequencies is less robust because the high frequency coefficients

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 627 – 628, 2005.

are small and change a lot when signal processing is performed. As the low and middle frequency components survive in most occasions, we select these as the characteristic collection. Watermark embedding method adopts the quantization index modulation (QIM). Multiple keys make the scheme more security. The preprocess of the watermark enhanced the security and robustness of the system; The embedding modes are very flexible and we can embed multiple watermarks at the same time or embed different watermark at the different time. The watermarks extraction does not require the original image and the detection false probability curve is depicted. The obtained meaningful watermark is obtained, which can be judged by human eyes. For simplicity, we adopt the Normalize Coefficient (NC) to evaluate the visual quality of the extracted watermark. Given the false probability, we can select the corresponding detection threshold. The experimental results indicate that embedded watermarks using our algorithm can give good quality and robust to common signal processing operations and geometrical distortions.

As far as the proposed algorithm, the factors affecting the imperceptibility are as following: the watermark size, the quantization step, the embedding positions, the expanding length of the watermark and the characteristics of the original image et al. Above all, the key factors are the watermark size and the quantization step. When the watermark size increases, the numbers of the changed coefficients also increase, which induces more visual distortions. Likewise, when increasing the quantization step, the alter range of the transform coefficients increases so the robustness is strengthened, which bring on the worse imperceptibility. We can utilize the characteristics of the human visual system (HVS) and adjust the quantization step adaptively to enhance the imperceptibility. The factors related to the robustness are the quantization step, the embedding positions, the expanding length of the watermark, the characteristics of the original image and so on. Experimenal results showed that the quantization step was the most crucial factor in the robustness. Therefore choosing the appropriate step is the key part of the proposed algorithm.

Towards an Effective Wireless Security Policy for Sensitive Organizations[1]

Michael Manley, Cheri McEntee, Anthony Molet, and Joon S. Park

School of Information Studies, Syracuse University, Syracuse, New York, 13244, USA
{memanley, CAMcente, ammolet, jspark}@syr.edu

1 Introduction

Wireless networks are becoming increasingly popular with organizations and corporations around the world. With the prospect of increased productivity and convenience at a reduced cost through the use of wireless, many organizations have introduced wireless networks into their infrastructures in a hope to reap its benefits. However, the adoption of wireless technologies brings with it new security concerns. The possibility that the signals from a wireless local area network (WLAN) being transmitted beyond the physical boundaries of an office make it easier for cyber criminals to monitor network traffic, disrupt data flows, and break into networks. The prospect of a breach of security becomes even more dangerous given that 70 percent of the data transmitted through wireless access points is unencrypted. These risks have elevated the importance of wireless security. With this increased concern for wireless security issues, a well thought-out and implemented wireless security policy is paramount. The goal of this work is to make the reader aware of the weakness in current wireless security models and to lay a framework of a reliable wireless security policy for sensitive organizations. We also examine a case study, the Department of Defense, of real world implementation of wireless security policies, analyzing their deficiencies based on our proposed framework.

2 Enhancement: What Are Currently Missing?

It's been widely accepted that information security policies are a requirement in today's information age and a lot of sources will include what they believe are key components that should be included in a policy. Most sources will also provide guidelines for the policy developers to keep in mind when developing the policy. However, after examining many references, there seems to be key elements that are missing from the templates and guidelines that don't get mentioned that would be beneficial to the policy and the development process.

[1] This work was a final team project of IST623 (Introduction to Information Security), taught by Prof. Joon S. Park, at Syracuse University in Fall 2004. We would like to thank our fellow students for their valuable feedback, insight and encouragement as we researched and developed this project during the semester.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 629 – 630, 2005.

Readability – The developers should know the intended audience and write to that audience. In a wireless security policy, including technical information is unavoidable but, in most cases, this information can be described in a way that non-technical users, or users with little technical background, can understand.
How are things really done? – It's important to know how the organization functions in reality versus how it is "supposed" to function according to Standard Operating Procedures (SOPs).
Management buy-in – Many templates discuss the importance of getting management to agree to the security policy but, very few discuss how to accomplish this. Management shouldn't be surprised by a policy that they will be required to insist that their subordinates follow.
User buy-in – It's common to find "obtaining management buy-in" as a key component of a security policy but, rare to find anything that mentions "user or employee buy-in."
Who is making the decisions? – Before developing a policy the policy developers should know who makes the decisions at the organization. It would be helpful for policy developers to know who "really" makes the decisions at an organization.
Scalability – Business needs and technology needs change. A policy has to be flexible enough to evolve with the advances of technology and the changing needs of the business.
Policy review – No matter how specific or general the developed policy is, it should be reviewed periodically. The policy should state how often the policy is to be reviewed and who should be included in the review process.
Disaster recovery –The policy should discuss the need for redundancy, backups and retention policies in the event that the information systems needed to be restored. The disaster recovery component of a policy will have to be developed with implementation cost, acceptable risk of data loss, and consequences of data loss all factored into the process.
Can we overdo it? – Finally, policy developers have to realize that there's a recognized tradeoff between security and convenience and the possibility that policy developers could overdo it when it comes to establishing security practices.

3 Case Study: Wireless Security Policy in DoD

The Department of Defense (DoD) is a very large, sensitive, and diverse organization. Not only do its policies have to address the hundreds of different governmental agencies that it interacts with but it also must address the hundreds of government contractors that work on and with its systems. The DoD has the unique task of implementing, managing, and enforcing their policies within their own organization, but must also provide controls to enforce policies with those outside its immediate control. The DoD's 8100.2 directive is a good match up to our ideal policy. The main weakness of the DoD's directive is the need to address a very large audience with diverse needs. In our case study, we discuss their deficiencies based on our proposed framework.

A Taxonomy of Cyber Attacks on 3G Networks

Kameswari Kotapati, Peng Liu, Yan Sun, and Thomas F. LaPorta

The Pennsylvania State University, University Park, PA 16802
kotapati@cse.psu.edu

Early first and second generation (1G and 2G, respectively) wireless telecommunication networks were *isolated* in the sense that their signaling and control infrastructure was not directly accessible to end subscribers. *The vision of the next generation 3G wireless telecommunication network is to use IP technologies for control and transport.* The introduction of IP technologies has opened up a new generation of IP-based services that must interwork with traditional 3G wireless telecommunication networks. *Cross Network Services* will use a combination of *Internet-based data* and *data from the wireless telecommunication network* to provide services to the wireless subscriber. They will be multi-vendor, multi-domain, and will cater to a wide variety of needs. An example of such a Cross Network Service is the *Email Based Call Forwarding Service (CFS)*, where the *status of the subscriber's email inbox* is used to *trigger call forwarding* in the wireless telecommunication network.

A security risk is introduced by providing Internet connectivity to the 3G networks in that certain *attacks* can be *easily enforced* on the *wireless telecommunication network indirectly from the IP networks*. These *Cross Infrastructure Cyber Attacks* are

Table 1. Summary of Attack Taxonomy

Dimension I: Physical access	**Dimension II**: Attack type	**Dimension III**: Attack Means
Level I: **Access to air interface with physical device**: Intruders have access to standard inexpensive 'off-the-shelf' equipment that could be used to impersonate parts of the network.	**Interception:** Passive attack- Intruder intercepts information but does not modify or delete them.	**Messages:** Signaling messages are compromised.
Level II: Access to Cables connecting 3G entities: Intruder may cause damage by disrupting normal transmission of signaling messages.	**Fabrication:** Intruder may insert spurious objects (data, messages and service logic) into the system.	
Level III: Access to 3G core network entities: Intruder can cause damage by editing the service logic or modifying subscriber data (profile, security and services) stored in the network entity.	**Modification of Resources:** The intruder causes damage by modifying system resources.	**Data:** The data stored in the system is compromised.
Level IV: Access to Links connecting the Internet based Cross Network Services and the 3G core network: This is a Cross Infrastructure Cyber Attack.	**Denial Of Service:** Intruder causes an overload or a disruption in the system.	**Service Logic:** The service logic running on the network is compromised.
Level V: Access to Internet Cross Network Servers: This is a Cross Infrastructure Cyber Attack. Intruder can cause damage by editing the service logic, modifying subscriber data (profile, security) stored in the Cross Network Servers.	**Interruption:** The intruder causes an Interruption by destroying resources	

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 631 – 633, 2005.

simple to execute and yet have serious effects. In this paper we present a *unique attack taxonomy* in which we consider the Cross Infrastructure Cyber attacks in addition to the standard Single Infrastructure attacks, some of which have already been studied. In presenting the taxonomy of attacks on the 3G Network, we classify the attacks into three dimensions summarized in Table 1.

To understand such attacks, we have developed a detailed abstract model of the 3G telecommunications network infrastructure. Each entity is modeled into atomic processing units (agents) and data stores (permanent and cached) as shown in Figure 1. We map attacks to these specific modules in each entity as discussed below.

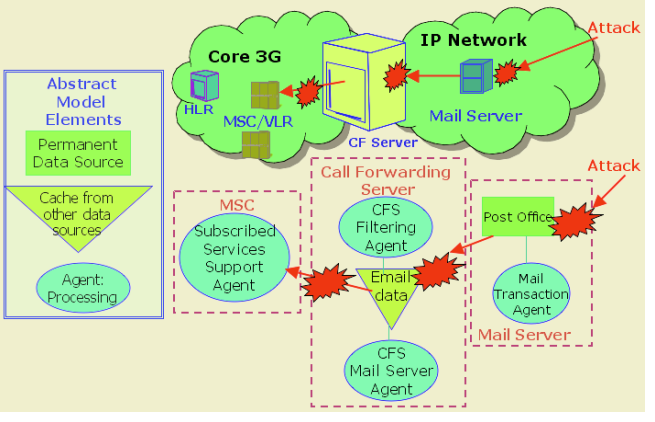


Fig. 1. Attack Propagation in CFS with simplified abstract model

The CFS forwards a call to a subscriber's voice mail if there is no email from the caller in its inbox of over a certain age; otherwise the call is forwarded to the subscriber's cellular phone. The *CFS Mail Server Agent* in the CFS periodically fetches email stored in the *Post Office* data source of the Mail Server. This data is stored in the *Email data* cache of the CFS. When there is an incoming call for the CF subscriber, the *Subscribed Services Support Agent* in the MSC will query the CF Server on how to forward the call. The *CFS Filtering Agent* will check its *Email data* cache, and if there is appropriate email from the caller, the call is forwarded to the subscriber's cellular phone.

The propagation of the attack from the Mail Server to the CFS, and finally the 3G-network entity, is illustrated in Fig 1. Using any well-known Mail Server vulnerabilities the attacker may compromise the Mail Server and corrupt the *Post Office data* source by deleting emails from certain people from whom the victim is expecting calls. The *CFS Mail Server Agent* queries the *Mail Transaction Agent* for emails from the *Post Office data* source and the *Mail Transaction Agent* will pass on the corrupted email data to the *CFS Mail Server Agent*. The *CFS Mail Server Agent* will cache the email data in its *Email data* cache. When the *Subscribed Services Support Agent* in

the MSC entity of the 3G network sends out a 'How to forward the call?' query to the CF Server, the CF Server will check its corrupt *Email* data cache, find that there are no emails from the caller, and route the call incorrectly. Thus the effect of the attack on the Internet Mail Server has propagated to the 3G network. This is a classic example of a *Dimension: I-Level V Cross Infrastructure Cyber Attack*, where the attacker gains access to the *Cross Network Server* and attacks by modifying data in the data source of the *Cross Network Server*.

The full paper presents a model for 3G networks, a set of *Cross Network Services*, and a full attack taxonomy including *Cross Infrastructure Cyber attacks*.

An Adaptive Approach to Handle DoS Attack for Web Services

Eul Gyu Im and Yong Ho Song

College of Information and Communications,
Hanyang University,
Seoul, 133-791, Republic of Korea
{imeg, yhsong}@hanyang.ac.kr

1 Introduction

Recently web services become an important business tool in e-commerce. The emergence of intelligent, sophisticated attack techniques makes web services more vulnerable than ever. One of the most common attacks against web services is a denial of service attack.

Christoph L. Schuba et al. [1] proposed a defense mechanism against SYN flooding attack. They developed a software monitoring tool that resides in a firewall. The program examines all TCP packets, and categorizes source IP addresses into the following states: *null, good, bad, new, perfect*, and *evil*. The incoming TCP packets are filtered if the source addresses of the packets are in *bad* or *evil* states.

There are some limitations of the Schuba's approach. 1) Since most attackers uses spoofed IP addresses and these IP addresses probably are not in the *bad* state, the SYN packets with spoofed source IP addresses have same chances to be processed with SYN packets from legitimate users. 2) If a SYN packet from a legitimate user is accidently dropped, the source IP address goes to the *bad* state, and the subsequent packets from the source IP address are filtered even though the server has enough free resources.

This paper proposes an adaptive mechanism to detect DoS attacks and to mitigate the effects. SYN flooding attacks are simulated as a case study to evaluate the performance of the proposed defensive mechanism.

2 Our Defense Mechanism

The design goal of our proposed defense mechanism is to detect DoS attack and to mitigate the effects of the attack through adaptive rule updates. The proposed defense mechanism augments the Schuba's approach to solve the above limitations. The major characteristics of our approach is summarized below:

Prioritization: To solve limitations of Schuba's approach mentioned above, we assign different priorities to different states. In addition, packets from IP addresses with higher priorities have more chance to be processed, which is similar to a priority queue mechanism. We assign a higher priority to known good

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 634–635, 2005.

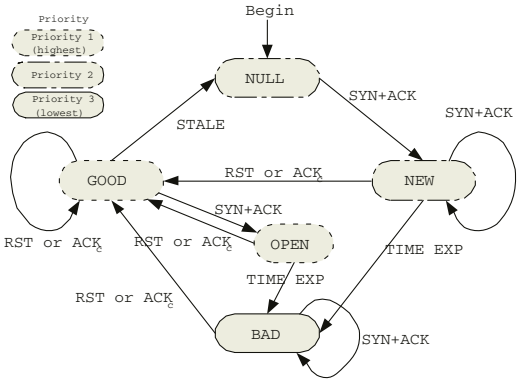


Fig. 1. Finite State Machine for our DoS Defense Mechanism

IP addresses, so packets from known good IP addresses have more chances to be processed than those from other IP addresses. Since packets from IP addresses with lower priorities still have chances to be processed, an IP address can get a higher priority later if an IP address is assigned with a lower priority due to errors or temporary harmful activities.

Examination of Response Packets: Our mechanism examines incoming packets as well as outgoing packets to/from web servers. Here, outgoing packets from web servers are examined to detect unknown attacks on web servers. Statistics show that when a web server is compromised through an attack, the amount of outgoing packets from the web server increases.

Figure 1 shows the finite state machine diagram for the proposed algorithm. The source IP address of each TCP packet is examined to find out which state the IP address belongs to. Incoming packets are assigned to different priority levels according to the state they belongs to.

Our proposed approach is specific to the SYN flooding attack, and we think the algorithm can be extended to cover other DoS attacks to web services. The proposed mechanism can be integrated into a coordinated defense mechanism by dynamically providing *bad* IP addresses to router based defense mechanisms.

References

1. Christoph L. Schuba, Ivan V. Krsul, and Markus G. Kuhn. Analysis of a denial of service attack on TCP. In *Proceedings of the 1997 IEEE Symposium on Security and Privacy*, pages 208–223, May 1997.

An Architecture for Network Security Using Feedback Control

Ram Dantu[1] and João W. Cangussu[2]

[1] Department of Computer Science,
University of North Texas
rdantu@unt.edu
[2] Department of Computer Science,
University of Texas at Dallas
cangussu@utdallas.edu

In the past active worms have taken hours if not days to spread effectively. This gives sufficient time for humans to recognize the threat and limit the potential damage. This is not the case anymore. Modern viruses spread very quickly. Damage caused by modern computer viruses (example - Code red, sapphire and Nimda) is greatly enhanced by the rate at which they spread. Most of these viruses have an exponential spreading pattern. Future worms will exploit vulnerabilities in software systems that are not known prior to the attack. Neither the worm nor the vulnerabilities they exploit will be known before the attack and thus we cannot prevent the spread of these viruses by software patches or antiviral signatures. Hence there is a need to control fast spreading viruses automatically since they cannot be curtailed only by human initiated control. Some of the automatic approaches like quarantining the systems and shutting down the systems reduce the performance of the network. False positives are one more area of concern. Feedback control strategy is desirable in such systems because well-established techniques exist to handle and control such systems. Our technique is based on the fact that an infected machine tries to make connections at a faster rate than the machine that is not infected. The idea is to implement a filter, which restricts the rate at which a computer makes connection to other machines. The delay introduced by such an approach for normal traffic is very low (0.5-1 Hz). This rate can severely restrict the spread of high-speed worm spreading at rates of at least 200 Hz. As a first step, we apply feedback control to the first level of hierarchy (i.e., host). We will then expand the model to further levels (e.g., firewalls, IDS) as shown next in the description of the system architecture.

Architecture: It is assumed a secured network consists of firewalls, sensors, analyzers, honey pots, and various scanners and probes. These components are either separate elements or collocated with hosts, servers, routers and gateways. In this architecture, a (centralized or distributed) controller is responsible for collection and monitoring of all the events in the network. This controller is knowledgeable about the network topology, firewall configurations, security policies, intrusion detections and individual events in the network elements. This controller is logical function and can be deployed anywhere in the network. As shown in Figure 1, the controller communicates with clients located in different network elements.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 636–637, 2005.

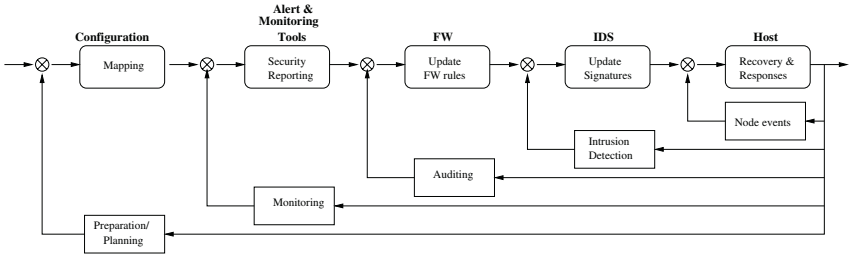


Fig. 1. Controller architecture for end-to-end security engineering (FW: Fire Wall, IDS: Intrusion Detection System

Clients are responsible for detection and collection of the events in the node and communicate to the controller. Subsequently, controller will run through the algorithms, rules, policies and mathematical formulas (transfer functions) for next course of action. These actions are communicated to the clients. As described in Figure 1, the architecture evolves from a concept of closed loop control. Changes regarding the security behavior are captured and mixed with the incoming network signals. This piece of information is used to formulate the next course of action. The final result is outcome from multiple loops and integration of multiple actions. The response times within each loop (we call them defender-loops) varies from few milliseconds to several tens of minutes. For example, nodal events like buffer overflows, performance degradation can be detected in matter of milliseconds. On the other hand, it may take several seconds to detect failed logins, changes to system privileges and improper file access.

The anomalies and the measurements from the hosts and servers are fed back to the intrusion detection systems and firewalls. We have represented velocity of newly arriving requests using a state space model. In particular, we have used velocity of arrival of new connections for detecting abnormalities. These new connections are categorized as either safe or suspect connections and enqueued in two different queues. Suspected connections are delayed for further processing. These suspected connections are dropped either because of higher rate of arrival or application layer protocol timeouts.

The size of both the queues follows an S-shaped behavior similar to the input function. However, safe queue saturates far earlier than the delay queue. This is because of the entries in the delay queue are dropped due to application timeouts. The velocity increases initially until an inflection point is reached and it starts to decrease until the saturation level in the safe queue is achieved and finally velocity goes to zero. To regulate the number of connection requests, a PI (proportional and integral) controller has been used. Due to PI control the detection of abnormality at different points on the S-curve will reach the same saturation point. This behavior is due to the fact that the system is stable, i.e., it will converge to the same equilibrium point even when starting with different initial conditions. We found that these results are similar when the controller is applied at the firewall or at the host in the feedback loop.

Defending a Web Browser Against Spying with Browser Helper Objects

Beomsoo Park, Sungjin Hong, Jaewook Oh, and Heejo Lee

Department of Computer Science and Engineering,
Korea University,
Seoul 136-701, Korea

Microsoft's Internet Explorer (IE) is the most widely used web browser, and the IE's global usage is reported as 93.9% share in May 2004 according to OneStat.com. The dominant web browser supports an extensible framework with a Browser Helper Object (BHO), which is a small program that runs automatically everytime starting IE. However, malicious BHOs abuse this feature to manipulate the browser events and gather private information, which are also known as *adwares* or *spywares*

Malicious BHOs have been used mainly for adwares which change the start page of IE or insert ads to web pages. Furthermore, it is possible to gather private information by spying on a web browser with a BHO for logging all inputs typed on the browser. This means that a malicious BHO can capture the passwords behind the asterisks and the credit card numbers copied by cut-and-paste mouse operations; thus, spying with BHOs is more powerful than conventional "keystroke" loggers. Nonetheless, proper countermeasures have not been studied extensively. One trivial defense is to disable BHOs on the browser, but disabling BHOs implies that users cannot make use of normal BHOs such as Google Toolbar.

While there are many detection and protection mechanisms for defending against keyloggers, malicious BHOs are another stream of threats insufficiently handled by anti-keylogging techniques. Therefore, we need to find another way to defend against malicious BHOs, while keeping normal BHOs working for good jobs.

In order to defend against malicious BHOs, we propose a secure automatic sign-in (SAS) architecture. The design goals of SAS are described as follows.

Securing sign-in information on web pages: If some information is entered on a web page through a browser, the information still remains until the page is submitted. And when submitting the page to the web server, subsequent events are incurred by the browser IE. At that time, a BHO can detect the events and obtain the contents of the web page. In order to protect from stealing sign-in information, the one and only way is making sign-in information inaccessible to the web page.

Preventing from keystroke logging: Protection mechanisms for keystroke logging have been proposed in many ways. One obvious way is the use of an alternative input method instead of a keyboard. Virtual on-screen keyboard comes under this category. In order to apply transparently to the protection of malicious BHOs, there are two conditions: 1) no use of keyboard and 2) no modification of web pages. We propose a defensive mechanism using a virtual keyboard in order to prevent from tracking keystrokes.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 638–639, 2005.

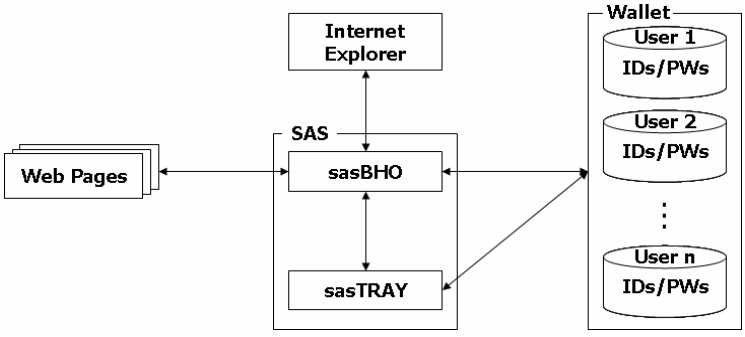


Fig. 1. Secure automatic sign-in (SAS) architecture

The avoidance technique for securing sign-in information works in two phases. First, fill the sign-in form of the current site with a fake information, and submit the page. Next, before the page is sent to the web server, intercept the HTTP request message and replace the fake information with the valid sign-in information stored at Wallet.

The SAS architecture, shown in Fig. 1, consists of three components as follows.

- **sasBHO** : A guardian BHO, called sasBHO, detects a sign-in form if exists in a web page. As well, this BHO program is responsible for invoking a virtual keyboard to register IDs and passwords, and for executing the automatic sign-in procedure using the registered IDs and passwords in Wallet.
- **sasTRAY** : An application program sasTRAY runs separately from IE or BHOs and resides in the system tray. And, sasTRAY is responsible for sustaining the information of master authentication even after terminating IE and sasBHO.
- **Wallet** : The access-controllable storage for maintaining registered IDs and passwords is called Wallet, which stores per-user IDs and passwords for registered sites. The IDs and passwords are stored after being encrypted by using a symmetric-key cryptography.

The proposed SAS architecture is to protect sign-in information from known and unknown malicious BHOs. We have implemented the SAS architecture on Windows systems. The reference implementation shows that the current implementation works properly for 83% of web sites among the most popular 100 web sites. Furthermore, we can increase the effective range by adapting the detection algorithm to other exceptional sites.

Full paper is available online at http://ccs.korea.ac.kr under the “publication” area. This work was supported in part by the ITRC program of the Korea Ministry of Information & Communications. For further correspondence, please contact with Prof. Heejo Lee by email `heejo@korea.ac.kr`.

Dynamic Security Service Negotiation to Ensure Security for Information Sharing on the Internet

ZhengYou Xia[1], YiChuan Jiang[2], and Jian Wang[1]

[1] Department of computer, Nanjing University of Aeronautics and Astronautics
zhengyou_xia@yahoo.com
[2] Department of Computing & Information Technology, Fudan University

1 Introduction

The term "quality of security service" is first presented by Cynthia Irvine [4]. The original definition is: “quality of security service refers to the use of security as a quality of service dimension and has the potential to provide administrators and users with more flexibility and potentially better service, without compromise of network and system security policies.” The original definition is focused on the quality of security service from the point of view of system administrators and users. We refine and define the term "quality of security service" in relation to security service negotiation among senders and receivers in a network, i.e. we focus on the quality of security service in the network.

Definition: Quality of Security Service refers to security service multi dimension spaces that are composed of strength of cryptographic algorithms, length of cryptographic key and Robustness of authentication mechanisms, etc., and is negotiated among the senders and receivers in the network.

Security is very important for an Internet application. The users don’t want to expose their message to others or be forged by others. They make extensive use of cryptography and integrity algorithms to achieve security. Although lots of cryptography and integrity algorithms have been suggested for Internet, if the users want to use a different security configuration for their application, they need to use dynamic mechanisms to negotiate quality of security service with the receivers. The sender can achieve the high quality of security service (high security level), only if the receivers and routers along path to receivers can support or satisfy the security service level requested by the sender. At this time the networking community has yet to develop a generic mechanism to solve the negotiation process for quality of security service. The traditional session mechanism [9][10][11][12] between the sender and receiver is only suited to the Point-to-Point case, because one obvious security service negotiation paradigm would have the sender transmit a negotiation request towards the receiver. However, the point-to-multipoint and multipoint –to-multipoint case [13] [14] is very difficult to solve using the session mechanism. In particular, one must not assume that all the receivers of a multicast group possess the same security capacity for processing incoming data, nor even necessarily desire or require the same quality of security service from the Internet. At the same time, the membership in a multicast

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 640 – 641, 2005.

group can be dynamic. To solve the above problems, we propose an extended RSVP [1] [2] [3] protocol called SSRSVP (security service RSVP) to provide the needed mechanism for quality of security service, to dynamically negotiate the quality of security service among the senders and receivers of multicast on the Internet. It provides different quality of security service resolutions to different receiver nodes in a multicast group with different security service needs. However, the SSRSVP is different to [15][16][17] ,which describe the format and use of RSVP's INTEGRITY object to provide hop-by-hop integrity and authentication of RSVP messages, and can support the IPSEC protocols[17]. SSRSVP is not to enhance or support RSVP security function but to provide different security service negotiation among the senders and receivers.

Enhancing Spatial Database Access Control by Eliminating the Covert Topology Channel

Young-Hwan Oh[1] and Hae-Young Bae[2]

[1] Dept. of Information and Communication, Korea Nazarene University
yhoh@kornu.ac.kr
[2] Dept. of Computer Science and Engineering, Inha University
hybae@inha.ac.kr

This paper presents a method to preserve security in GIS databases and describes a polyinstantiation approach to prevent covert channel in a multilevel secure spatial database. Usually, security is handled by giving users permissions/privileges and database objects security levels. In the case of GIS objects, the complex relationships among them (spatial, topological, temporal, etc.) mean that extra care must be taken when giving levels; otherwise users may see something they are not authorized to see or may be denied access to needed information. Currently, database security is combined with the spatial database system on the fields of facility management, air-traffic control and military area. Although much research in the field of database security have been made, there has been no study in the field of secure spatial database[1,2].

In the case of support topological information between spatial objects in spatial database systems, there is spatial information flow from a higher-user to a lower-user. Spatial objects have positional information and a topological relationship between neighboring objects. It can be inferred to a lower-user through topological information of neighboring objects. It is called covert topology channel.

To overcome these problems, a polyinstantiation for the spatial objects is proposed in this paper. Generally, the polyinstantiation method, which is one of the access control methods in the previous database security field, is used and applied[3,4]. The proposed method expands the polyinstantiation method of the previous relational database system. Also, it prevents service denial and information flow on the spatial objects. This method is divided into two steps: security level conversion test and polyinstance generation. In the step of security level conversion test, the security level of the spatial object and the security level of the user who has requested access on the object is compared to judge for either forward modification or downward modification. Afterward the step of polyinstance generation follows the order of copying the security level, the aspatial data and the spatial object identifier. In the main chapter, we investigate covert channel and covert channel with topological relationship. Then we explain polyinstantiation method for spatial information and propose the algorithms of polyinstantiation for spatial objects.

1 Covert Topology Channel

Spatial objects have positional information and a topological relationship between neighboring objects. It can be inferred to a higher-object class through topological

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 642 – 643, 2005.

information of neighboring objects by a lower-object. In that, indirect flow of spatial information occurs by topological relationship and, moreover, most GIS applications also use graphic user interface(GUI). In addressing these problems, the polyinstantiation of spatial data for solving spatial information flow that occurs by topological relationship of spatial data was proposed.

2 Polyinstantiation for Covert Topology Channel

We propose polyinstantiation method for security management of spatial objects. This method extends the existent polyinstantiation method to support multilevel security, and considers the characteristic of spatial objects. We suggest the polyinstantiation management policy to prevent an information leakage and denial of service. And we implement the security subsystem and define spatial data operation for multi-level security of spatial object in the spatial database systems.

3 Algorithms for Secure Spatial Objects

The 9-intersection model has been the most popular mathematical framework for formalizing topological spatial relations. This model considers whether the value of the 9-intersections, a range of binary topological relations, can be identified. The algorithms of polyinstantiation for operations on the spatial object, such as contain, intersect, overlap and disjoint, are described.

Acknowledgements

This work was supported by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program.

References

1. Young Sup Jeon, Young Hwan Oh, Soon Jo Lee, Ki Wook Lim, Hye Young Bae, “Controlling information flow through the spatial view that has a multi-level security” pp.93-96, Number 1, Volume 8, Springtime collection of scholastic papers, Information Processing Society, 2001
2. Wan Soo Cho, “System design of the expanded relational database for multi-level security.” Doctorate paper, Inha University, 1996.
3. Jajodia, S. and Sandhu, R., “Polyinstantiation Integrity in Multilevel Relations,” *Proc.of IEEE Computer Society Symposium on Research in Security and Privacy*, pp.104-115, 1990.
4. Jajodia, S. and Kogan, B., “Integrating an Object-Oriented Data Model with Multilevel Security,” *Proc. of IEEE Symposium on Research in Security and Privacy*, 1990.

Gathering Digital Evidence in Response to Information Security Incidents*

Shiuh-Jeng Wang+ and Cheng-Hsing Yang

+ Information Management Dept., Central Police University,
Taoyuan, Taiwan, 333
sjwang@mail.cpu.edu.tw
Computer Science Dept., National Pingtung Teachers College,
Pingtung County, Taiwan

1 Introduction

To effectively fight computer crime, it is necessay to locate criminal evidence, from within the computer and the network; this necessitates forensic enquiry so that the evidence will be secured in an appropriate manner, and will be acceptable in a court of law as proof of criminal behavior. Digital evidence is data in computer storage that can be used to prove criminal behavior [1]. The digital evidence, however, is easily copied and modified, is not easy to prove in source and integrity, cannot be well perceived by human senses in the presentation of digital information.

Our goal is to enable an enterprise unit, when dealing with such contingencies, to retain and secure reliable digital evidence to facilitate investigation and prosecution of criminal behavior.

2 Procedural Response to Information Security Incident

Gathering information about an information security incident can be an extremely complex procedure. We further enriched incident response procedures to achieve the stated objectives for information security incidents. The proposed responses to an information security incident are presented below. *1. Incident inspection, 2. Initial incident response, 3. Incident response strategy planning, 4. Network monitoring, 5. Investigation, 6. Making a duplicate copy of the suspect case, 7. Security Assessment, 8. Restoration, 9. Report.*

These procedures above-mentioned, in facing unpredictable invasive attacks, proper preparation before the event can provide strong system protection, and is the first step in guarding against information security incidents. Being prepared can reduce the risk of incidents, yet a good system information inspection record can provide relevant records for law enforcement and become a source of evidence for future investigations. System

* This work was supported in part by National Science Council in R.O.C. under Grant No. NSC-42186F (Visiting Florida State University, 2004) and NSC 93-2213-E-015-001.

+ Whom Correspondence.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 644 – 645, 2005.

management should establish a set of clear and definite incident response procedures in order to solve the problem as quickly and effectively as possible.

3 Digital Evidence Exposure

When dealing with incident response procedures, the majority of enterprises are more concerned with restoring the system to normal operation than tracing the invader and investigating the crime. Yet with any computer operation, there is a possibility of damaging or altering the evidence of the invasion, leaving the system exposed to even more harm, while waiting for the law enforcement unit to come and investigate. Because of this, the best way to proceed is to let the incident response team search for digital evidence at the earliest possible moment after a system break-in has occurred, in order to allow the law enforcement unit to proceed with the investigation.

Nowadays, with computer forensics being available, some enterprises will attempt to use a program composed in-house to proceed with computer forensics [2]; this may not be a wise thing to do, however. A self-composed program does not always have the public's trust and its validity can be brought into question. Because it is not seen to be independent, its evidence gathering ability may be compromised when the time comes to present that evidence in court. As for the small programs targeting computer forensics found on the Internet, they are not able to achieve the in-depth analysis required for data inspection. Accordingly, for a digital evidence gathering tool to be reliable for the task of computer forensics, it is appropriate that it possess both stability and integrity. During the computer forensics process, procedures should be periodically recorded for the safekeeping, testing, analysis and presentation of digital media evidence in order to prove that the process is above board. If a stable and reliable forensic software program is used to gather evidence, based on standard operating procedures (SOP), there should be no question as to its validity.

4 Remarks and Conclusions

In this study, we have laid out the recommended steps and procedures to be followed during an information security incident response. When an organization, faced with an information security incident, is unable to proceed with systematic incident response procedures at the most opportune, or earliest possible, moment, it could result in greater damage to the information system and hinder the follow-up investigation; consequently the criminal may remain out of the reach of the law. Therefore, a functional forensic procedure is required to clarify the investigation in the security incident.

References

1. E. Casey, Handbook of Computer Crime Investigation: Forensic Tools & Technology, Academic Press, pp. 1-16, 2002.
2. A. Yasinsac, R.F. Erbacher, D.G. Marks, M.M. Pollitt, and P.M. Sommer, "Computer Forensics Education," IEEE Security and Privacy, pp. 15-23, Aug. 2003.

On the QP Algorithm in Software Watermarking

William Zhu and Clark Thomborson*

Department of Computer Sciences, University of Auckland, New Zealand
fzhu009@ec.auckland.ac.nz, cthombor@cs.auckland.ac.nz

Software security is a key issue in an Internet-enabled economy. In order to prevent software from piracy and unauthorized modification, many techniques have been developed. Software watermarking [1, 2] is such a technique, which can be used to protect software by embedding some secret information into the software to identify its copyright owner. In this paper, we discuss algorithms of software watermarking through register allocation.

The QP Algorithm [4, 5] was proposed by Qu and Potkonjak to watermark a software through register allocation. This algorithm is based on the graph coloring problem which colors the vertices of a graph with the fewest number of colors such that no vertices connected by an edge receive the same color. The QP algorithm adds edges between chosen vertices in a graph according to the value of the message to be inserted. Qu and Potkonjak have not published any implementation for the QP algorithm; they mainly considered the credibility and overhead and even have not considered any attacks in their analysis, however, resistance to attack is of vital importance in software watermarking. This paper gives an example to show that inserting two different messages into an original graph respectively, we can get the same watermarked graph, so the outline of the extracting algorithm for the QP algorithm is not correct; in fact, the QP algorithm is not extractable.

Myles and Collberg [3] implemented the QP algorithm, and conducted an excellent and thorough empirical evaluation of this algorithm. They had tried various attacks on the QP algorithm to analyze its robustness. Furthermore, they proposed the QPS algorithm to compensate for the flaws they discovered in the QP algorithm. However, for the reasons shown in this paper, the QPS algorithm is also incorrect, for there are some contradictory statements in this algorithm. Furthermore, the QPS algorithm involves a complicated and restrictive concept which we show how to avoid in this paper. In addition, they supplied an incorrect example when arguing that the QP extraction algorithm is unreliable. They concluded that the QP algorithm is susceptible to a large number of distortive attacks, with a very low bit-rate, and is not fit for watermarking architecture-neutral code but could probably be used to watermark native code with some tamper-proofing techniques.

This paper gives three potential corrections to the QPS algorithm. Through examples, we show that a message embedded into a graph by two of these

* Research supported in part by the New Economy Research Fund of New Zealand. The full paper is available at http://www.cs.auckland.ac.nz/~fzhu009/Publications/OntheQPAlgorithminSoftwareWatermarking.pdf.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 646–647, 2005.

corrected QPS embedding algorithms also can not be recognized reliably. The soundness of the third correction of the QPS algorithm is proved in this paper.

We propose an improvement on the QP algorithm, the QPI algorithm as we call it, through adding two additional vertices to a graph to be watermarked. This algorithm realizes what the QP algorithm really wants to do.

Through examples and analyses, we reach our following conclusions about the QP algorithm and its variant.

1. The message embedded into a graph by the QP embedding algorithm can not be recognized reliably.
2. The QP extraction algorithm is not correct. It tries to recognize a message just by the unwatermarked graph; it does not use the watermarked graph.
3. The QPS watermarking algorithm proposed by Myles and Collberg contains contradictions and involves a very complicated and restrictive concept. It does not solve the robustness problem in the QP algorithm.
4. The QPI algorithm proposed by us can correctly realize Qu and Potkonjak's idea of software watermarking through register allocation.

From the paper [4, 5], we think it is important to distinguish an extraction algorithm and a recognition algorithm in software watermarking. An extraction algorithm tries to extract all bits of the message inserted in a software, while a recognition algorithm decides whether a watermark exists in a software. A good work to define these concepts is not as easy as it seems. We will explore this problem in our future research.

References

1. C. Collberg and C. Thomborson: *Software Watermarking: Models and Dynamic Embeddings*. POPL'99, (1999)
2. C. Collberg and C. Thomborson: *Watermarking, tamper-proofing, and obfuscation - tools for software protection*. IEEE Transactions on Software Engineering. vol. 28 (2002) 735-746
3. G. Myles and C. Collberg: *Software Watermarking Through Register Allocation: Implementation, Analysis, and Attacks*. LNCS 2971 (2004) 274-293
4. G. Qu and M. Potkonjak: *Analysis of Watermarking Techniques for Graph Coloring Problem*. Proceeding of 1998 IEEE/ACM International Conference on Computer Aided Design. ACM Press. (1998) 190-193
5. G. Qu and M. Potkonjak: *Hiding Signatures in Graph Coloring Solutions*. Information Hiding. (1999) 348-367

On the Use of Opaque Predicates in Mobile Agent Code Obfuscation

Anirban Majumdar and Clark Thomborson

Department of Computer Science,
The University of Auckland,
Private Bag 92019, Auckland, New Zealand
{anirban, cthombor}@cs.auckland.ac.nz

Mobile agent technology is an evolving paradigm that combines the inherent characteristics of intelligent agents, namely, adaptability, reactivity and autonomy with mobility. These characteristics of mobile agents provide an excellent means of meeting the distributed and heterogeneous requirements of many military applications that involve low bandwidth and intermittently connected networks. In typical military applications, mobile agents can be used to perform *information push*, *information pull*, and *sentinel monitoring* [1].

In spite of its tremendous potential, several technical requirements must be met in order to support the widespread transition of agent technology to the military domain. Confidentiality of agent code is of foremost concern. The countermeasures directed toward agent protection are radically different from those used for host protection [2]. Host protection mechanisms are a direct evolution of traditional mechanisms employed by trusted hosts and traditional mechanisms are not devised to address threats originating on agents from the execution environment. Agents executing in military applications cannot trust the platforms they are executing on and this problem stems from the inability to effectively extend the trusted environment of an agent's host platform to other agent platforms visited by the agent. Previous works [3] on provable mobile agent security have proposed cryptographic techniques to protect agents from unauthorised code interception; however, since any information belonging to a mobile agent is completely available to its host system, it cannot possibly keep the cryptographic key secret from the system on which it is running.

Obfuscation is the technique of transforming a program into a form that is more difficult to understand for either a human adversary or for an automated one or both, depending upon the transformation applied [4]. An obfuscated program should have "identical" behaviours with respect to the original unobfuscated one.

This extended abstract focuses on obfuscating agent code by introducing opaque predicates to guard the control-flow. An opaque predicate is a conditional expression whose value is known to the obfuscator, but is difficult for the adversary to deduce. A predicate P is defined to be *opaque* at a certain program point p if its outcome is only known at obfuscation time. We write P^F_p (P^T_p) if predicate P always evaluates to False (True) at program point p. The opaqueness of such predicates determines the resilience of control-flow transformations.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 648 – 649, 2005.

Obfuscation using opaque predicates that use the inherent concurrency associated with mobile agent systems and aliasing are resilient against well known static analysis attacks. It will be very difficult for an adversary to understand the dynamic structure of the predicate and mount attacks that could statically analyse the associated values for each of the terms present in the predicate. The flexibility of this scheme can be further enhanced by making the predicate structure arbitrarily complex by incorporating numerous guard agents dynamically. In an already existing message-passing scenario between agents, the probe messages sent by the obfuscated agents to the guards will contribute to a negligible amount of extra overhead. The technique also does not depend of any particular language feature and is therefore applicable to generic mobile agent platforms.

Our technique is not an alternative to Sakabe's [5] mobile agent obfuscation technique, which takes advantage of polymorphism and exception handling mechanism of Java. Sakabe et al. established that their obfuscation technique can be reduced to an instance of the NP-hard problem of *precisely* determining if an instance of a class points-to an overloaded method during an execution of a program. We observe that obfuscation using method aliasing as a standalone technique may not be sufficiently resilient since the smaller size of agent programs compared to that of commercial Java software may result in a fewer number of methods to be considered for overloading in Sakabe's technique. Hence, using our technique in conjunction to the one already proposed by Sakabe et al will substantially strengthen the resilience of obfuscated agents.

References

1. McGrath, S., Chacón, D., and Whitebread, K.*: Intelligent Mobile Agents in Military Command and Control*. In Autonomous Agents 2000 Workshop/Agents in Industry, Barcelona, Spain. 2000.
2. Sander, T., and Tschudin, C.F.: *Protecting mobile agents against malicious hosts.* In Vigna G., ed.: Mobile Agents and Security. Volume 1419 of Lecture Notes in Computer Science. Springer-Verlag. 1998
3. Hohl, F.: *Time limited blackbox security: Protecting mobile agents from malicious hosts.* In Vigna G., ed.: Mobile Agents and Security. Volume 1419 of Lecture Notes in Computer Science. Springer-Verlag. 1998
4. Collberg, C., Thomborson, C., and Low, D.: *Manufacturing Cheap, Resilient, and Stealthy Opaque Constructs*. In Proceedings of 1998 ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'98). 1998.
5. Sakabe, Y., Masakazu, S., and Miyaji, A.: *Java obfuscation with a theoretical basis for building secure mobile agents*: . In Lioy A., Mazzocchi D. eds.: Communications and Multimedia Security (CMS 2003). Volume 1419 of Lecture Notes in Computer Science. Springer-Verlag. 2003.

Secure Contents Distribution Using Flash Memory Technology

Yong Ho Song and Eul Gyu Im

College of Information and Communications,
Hanyang University, Seoul, Republic of Korea
{yhsong, imeg}@hanyang.ac.kr

The use of flash memories, built upon EEPROM technology, increases at an explosive rate due to the high capacity, low cost, and non-volatileness. Many battery-powered mobile devices such as digital cameras, cell phones, and MP3 players use flash memories to implement cost-effective solid-state storage.

In fact, a flash memory device is constituted as an array of blocks each containing a number of pages. Each page consists of data area for holding one to four hard disk sectors and spare area for storing meta data such as error correction codes and bad block indicators.

When used for implementing solid-state storage, the flash memory devices are accessed through a software layer, called *Flash Translation Layer* (FTL), which sits between a disk driver and physical devices. This layer is used to provide the same interface as a hard disk to operating systems for the storage [1]. To this end, the FTL translates disk operations from operating systems into corresponding flash memory operations, and maps a target location specified in logical sector number to the offset within a memory device represented by a pair of physical block and page number.

Depending on applications, it is required that the data contents in storage should be prevented from being duplicated through ordinary access methods, accessible only by legal users and destroyed after the use. One example of such applications is the software contents distribution to legal license holders. In the past, copy-protected installation floppy disks were often used in such a way that the disks purchased by legal users are invalidated after the installation process ends. Nowadays, however, this approach is less appropriate because the distribution contents are often too big to fit into floppy disks.

This paper presents a novel mechanism to distribute software contents to legal users using flash memories. The proposed mechanism provides a way to prevent the media contents in flash memory from being duplicated through ordinary access methods, and to destroy the contents after the legal use. It is implemented as yet another FTL, called *Security-enhanced FTL* (in short, SFTL), as shown in Figure 1. In the flash-based storage, two software components exist: an application and software contents. The application could be an installation program of the contents or a dedicated player for them.

The SFTL augments the functions of FTL in the following ways. It implements a secure read operation. The contents stored in the flashed memory are in an encrypted form, $E\{C, K_{ENC}\}$, where C is the actual content to distribute,

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 650–651, 2005.

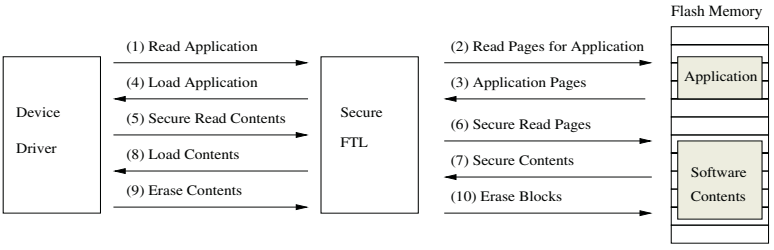


Fig. 1. Secure Contents Delivery

K_{ENC} is the key used for encryption, and E is an encryption function. The secure read operation differs from the ordinary read operation in that the former applies a decryption function, $D\{E\{C, K_{ENC}\}, K_{DEC}\}$, before returning the data to operating systems, where K_{DEC} is the key used for decryption, and D is a decryption function.

The decryption key, K_{DEC}, can be provided in one of two ways: first, it is stored in the first block of the flash memory at the manufacturing time so that the SFTL retrieves the key before executing a decryption function, and second, the key is passed by the application at run-time. In the former, the key is not associated with the user in any way. That is, anybody can access the contents as long as (s)he has the media. On the other hand, in the latter, the application accepts a key from the user during execution, and writes it into a pre-defined location of flash memory so that the key is available for the SFTL's secure read operations.

The SFTL invalidates the contents upon the receipt of an erase request from the application by issuing explicit erase commands to flash memory devices. This explicit erase request from the application can be implemented as an ordinary read request from the device driver with a pre-determined set of parameters.

The operation procedure of the proposed mechanism is illustrated in Figure 1. The application is first loaded into main memory (Step 1-4). During the execution, it issues secure read contents operations (Step 5). After SFTL reads out the encrypted memory contents (Step 6-7), it decrypts and returns them to operating systems (Step 8). Once the application is about to complete, it issues the erase requests to SFTL which are then translated to actual erase commands to flash memory (Step 9-10).

In summary, this paper presents a new mechanism for the implementation of secure contents delivery. The proposed mechanism can be also used for secure message delivery in military applications where it is necessary to destroy secret information after the delivery.

References

1. Jesung Kim and et al. A space-efficient flash translation layer for CompactFlash systems. *IEEE Transactions on Consumer Electronics*, 48:366–375, May 2002.

Background Use of Sensitive Information to Aid in Analysis of Non-sensitive Data on Threats and Vulnerabilities

R.A. Smith

Content Analyst LLC, Reston, VA
rsmith@contentanalyst.com

One of the 9-11 commission's recommendations on a different way of organizing intelligence activities of the United States was to unify the effort in information sharing across the Intelligence Community. Challenges include the need to deal with information that is geographically distributed and held in compartmented repositories having restricted access. A demonstrated 'need to know' is required before the data can be shared, and that assumes that one knows it exists and where to ask for it. Each intelligence agency has its own data security practices that restrict out-right sharing of the data within the Intelligence Community at large. Commercial off-the-shelf solutions exist for securely sharing highly sensitive 'need to know' data between cooperating agencies via digitally cosigned contracts, even using the Internet. Once data has been exchanged it still needs to be protected with the same 'need to know' restrictions by the receiving agency. In the post 9-11 world the question is how do we securely share information in a manner that protects the data but enables its value to be discovered by others having a 'need to know'. This poster session proposes a demonstration of a secure data sharing technique that allows sharing of sensitive documents to influence collections of documents available to first responders and others; without exposing the contents of the sensitive documents. The underlying technology that makes this demonstration possible is latent semantic indexing (LSI). LSI is a robust dimensionality-reduction technique for the processing of textual data. The technique can be applied to collections of documents independent of subject matter or language. Given a collection of documents, LSI indexing can be employed to create a vector space in which both the documents and their constituent terms can be represented. Sensitive documents are employed as part of the training data. The relationship information implicit in the sensitive documents is smoothly blended with the relationship information implicit in the non-sensitive documents. This has the effect of slightly perturbing the representation vectors for tens to hundreds of thousands of term vectors in the resulting LSI space. The sensitive documents can be completely protected in this process – there is no way, even in principle, for the text of sensitive documents to be reconstructed. However, the subtle changes in the term representation vectors can yield dramatic improvements in analysis activities carried out using the non-sensitive documents. Non-sensitive documents are made available to first responders and others who need to be made aware of threats and vulnerabilities. The influence of sensitive documents on the non-sensitive documents

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 652 – 653, 2005.

will produce a re-ordering of non-sensitive documents as the result of a first responder query. For example, first responder queries of similar semantic context to the sensitive training documents will return non-sensitive documents highly influenced by the sensitive training documents. In a real implementation of the system, first responders would be able to infer the implications of the sensitive training documents through the non-sensitive searchable documents. The richness of both document sets (sensitive and non-sensitive) has implications for performance of the system. This technique directly addresses the recent GAO report on recommendations for improving the sharing of information between the federal government and the private sector on incidents, threats, and vulnerabilities.

Securing Grid-Based Critical Infrastructures*

Syed Naqvi and Michel Riguidel

Computer Sciences and Networks Department,
Graduate School of Telecommunications,
46 Rue Barrault, 75634 Paris Cedex 13, France
{naqvi, riguidel}@enst.fr

The emerging Grid Computing Technologies are enabling the conception of heavily ICT-based critical infrastructures (CIs). The nature, volume and sensitivity of information that may be exchanged across the computing resources will be expanded substantially. As a result of increasing processing power, data storage capacity and interconnectivity, these Grid-based CIs will be exposed to a much larger number and a wider variety of threats and vulnerabilities. This raises new security issues for the CIs. In this paper, we have discussed the new paradigm of Grid-based CIs; the inevitability of their adoption in the near future; the risks associated with them; and a global framework to tackle the security related issues to gain the confidence of an already skeptical public.

The term *Grid* refers to systems and applications that integrate and manage resources and services distributed across multiple control domains [1]. It suggests that the resources of many computers can be cooperatively and perhaps synergistically harnessed and managed as a collaboration toward a common objective. The Grid intends to aggregate all kinds of heterogeneous resources that are geographically distributed. Pioneered in an e-science context, Grid technologies are also generating interest in industry, as a result of their apparent relevance to commercial distributed applications [2].

Today, with IT-based CIs, the governments entrust their information and business operations to their information systems, and often the Internet. With so much value online, the need for information systems security (ISS) becomes obvious. Computational Grids are the best candidate to meet the increasing demand of more sophisticated systems for CIs. They provide enormous computing power and storage space to handle the bulk of data and its processing in the most efficient way. However, the security services of Grids are still at the verge of development. The evolution of these services require a detailed application-specific requirements analysis. To our knowledge, no previous paper has been published that provides such analysis for the Grid-based CIs. This situation has overwhelmingly motivated us to take this initiative.

Grid computing technology has been widely used in areas such as high-energy physics, defense research, medicine discovery, decision-making, and collaborative

* This research is supported by the European Commission funded project SEINIT (Security Expert Initiative) under reference number IST-2002-001929-SEINIT. The overall objective of the SEINIT project is to ensure a trusted security framework for ubiquitous environments, working across multiple devices and heterogeneous networks, in a way that is organization independent (inter-operable) and centered around an end-user. Project webpage is located at www.seinit.org

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 654 – 655, 2005.

design. Currently, Grid computing has started to leverage web-services to define standard interfaces for business services like business process outsourcing, a higher level outsourcing mode of e-business on demand. The Grid can provide people from different organizations and locations to work together to solve a specific problem, such as design collaboration. This is a typical dynamic resource sharing and information exchange. The Grid computing platform allows resource discovery, resource sharing, and collaboration in a distributed environment.

Critical Infrastructures are large scale distributed systems that are highly interdependent, both physically and in their greater reliance on the information and communication technologies (ICT) infrastructures, which logically introduce vulnerabilities that make them increasingly complex and fragile. Failures, accidents, physical or cyber attacks can provoke major damages which can proliferate by cascading effects and then can severely affect a part or the whole society. Because of their interdependencies and their increasing reliance on open systems, critical infrastructures constitute an unbounded system where faults may occur and proliferate in a severe way and where security represents a real challenge and requires new methodologies and tools. Securing the communications is an essential task. However, it is necessary to protect the infrastructures themselves (especially critical infrastructures) so that they become self-healing, fault tolerant, fault resistant, and resilient architectures.

As the computing world has grown more dependent on the communications networks, the Grid computing is increasing the visibility of computer systems in the running of businesses, boosting the cost of system downtime; even short interruptions in the functioning of the Internet and other networks have become unacceptable. Consequently, Denial of Service (DOS) attacks that prevent access to on-line services are one of the greatest threats to the information infrastructure.

When regarding the protection of the essential information infrastructures (and especially critical infrastructures), most of the time one concentrates on the availability subject. However, we put emphasis on not forgetting to protect integrity of provided services as well. Moreover, Service availability may conflict with other security goals that can be more fundamental in some infrastructure cases; when integrity and confidentiality are the main goals, the most secure system is often one that does nothing. Therefore, protection against DoS often requires architectural changes to the system, which may prove expensive.

Another challenge for securing infrastructures is to make a trade-off between security and privacy. Technological developments and the needs of law enforcement provide increased opportunities for surveillance in cyberspace. Better managing and strengthening the infrastructure would make it more efficient and resilient without the need for unnecessary surveillance. A typical aspect of this issue is the problem of attack trace-back in Internet between the security (detecting the attacker) and the privacy (protecting the anonymity of Internet users).

References

1. Foster I., Kesselman C., *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999. ISBN 1558604758
2. Foster I., Kesselman C., Nick J., Tuecke S., The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, Globus Project, 2002.

The Safety Alliance of Cushing – A Model Example of Cooperation as an Effective Counterterrorism Tool

David Zimmermann

Federal Bureau of Investigation,
3301 West Memorial, Oklahoma City, Oklahoma 73134
dzimmerm@leo.gov

Oklahoma, known for its oil, is home to the Cushing Pipeline Junction (CPJ), one of the nation's most critical infrastructure assets. In wake of the modern terrorism threat that has specifically identified the petroleum industry as a target, CPJ stakeholders have implemented the most basic of tools - cooperation - to counter this threat; in doing so, they have sharpened this tool to a degree that their cooperation has become a model for the nation.

The CPJ's nine major oil companies and their respective tank farms, pipeline operations, and transport operations, span two counties in central Oklahoma, spilling into the city limits of Cushing, a town of 8,500 that is located one hour northeast of Oklahoma City.

The CPJ is the largest oil pipeline gathering facility in the United States (U.S.). On a daily basis approximately 40% of the crude oil from the Gulf Coast region flowing to the Midwest region (which is between 2% and 5% of the nation's crude oil) travels through the CPJ. Approximately 20% of the Midwest region's crude oil is stored at the CPJ.

In April 2001, a CPJ oil company's transport truck overturned and a small product spill ensued. Numerous CPJ oil companies offered to provide mutual aid but local emergency services single-handedly worked the crisis. After the spill, safety officers from each of the oil companies met with local emergency services leadership and discussed how the oil companies' equipment and training could be used to assist each other, and local emergency services, during a future crisis. They proposed the creation of the Safety Alliance of Cushing (S.A.C.) and a model for public-private cooperation was born.

To provide for more resources, the S.A.C. quickly expanded to include other local, as well as county and state, government agencies.

Initially the S.A.C. was only a response-based organization, formed simply for the purpose of providing mutual aid to member companies and agencies in times of need. However, when airplanes flew into buildings in September 2001, the mindset of this organization changed. Because intelligence reports repeatedly stated that terrorists want to cripple the U.S. economically, the S.A.C. realized that a new threat existed and reacting to CPJ incidents was not enough to keep the CPJ safe. With its resources in people and equipment, preventing terrorist acts against the CPJ became a necessary mission that fell into the lap of the S.A.C.

After September 11, 2001, the Oklahoma City Federal Bureau of Investigation (FBI) realized that the S.A.C. provided an existing terrorism prevention backbone that

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 656 – 657, 2005.

could help protect the CPJ. The FBI joined the S.A.C., provided education on terrorism, and held two tabletop exercises to train the S.A.C. in reacting to both physical and cyber threats. In addition, in 2004 the FBI executed a full-field exercise that simulated multiple terrorist attacks on CPJ facilities. This nine-hour exercise was an enormous undertaking that involved 100 different agencies and companies. 62 controllers trained 388 participants while 162 observers from numerous organizations watched. The exercise successfully trained the incident response of all participating companies and agencies.

Since its inception, the S.A.C. has grown to 23 member organizations representing the oil industry, law enforcement, emergency management, emergency services, and local government.

Open communication amongst S.A.C. members is one of the keys to its success. As an example, the S.A.C. has created and maintained a call circle that is activated regularly in emergency training and real life scenarios. In addition, the S.A.C. member organizations have a common radio frequency, with a repeater tower, that is used for emergency radio communications when needed.

The S.A.C. meets monthly to discuss safety, security, and emergency response issues and stays abreast of pertinent unclassified intelligence through regular email contact with, and personal visits from, the FBI.

The S.A.C. realizes that technology will play a large part in the organization's continued success and future projects include the formation of a listserv to enhance communication and the offering of online safety and security training.

Civic organizations, along with government and industry leaders, have asked S.A.C. representatives to speak at meetings and provide information on how the S.A.C. functions. In Midland, Texas, where similar oil assets are located, oil companies have begun to imitate the S.A.C. and work closer with industry and government partners.

The most useful aspects of the S.A.C. – open communications, joint training, public-private cooperation, and a common mission – can be used by similar organizations in any infrastructure sector. It is not necessary that the organizations wishing to use the S.A.C. as a model be response-oriented; what is necessary is only that organizations realize that cooperation, whether it is with the government or with competitive industry partners, is the key to taking on a terrorism prevention role.

Post September 11, 2001, the S.A.C. has emerged as one of the best public-private cooperative efforts in the nation, efficiently combining the resources of government and private industry to ensure a common goal – the protection of a national critical infrastructure asset. Built on cooperation and strengthened by technology, the S.A.C. should be viewed both as a tool for countering terrorism and as a model to be emulated by other communities who live and work near our nation's critical infrastructure assets.

A Framework for Global Monitoring and Security Assistance Based on IPv6 and Multimedia Data Mining Techniques

Xiaoyan Gong and Haijun Gao

The Key Lab for Complexity Systems and Intelligence Sciences,
Institute of Automation, Chinese Academy of Sciences, 100080, Beijing, China
gxylsh@hotmail.com

1 Introduction and Related Work

Nowadays people often find them lost in the ocean of textual data. However, with development and application of the Next Generation Network (NGN) based on IPv6, people will find them in the ocean of multimedia data. With large number of IP address resources and wide bandwidth available, NGN based on IPv6 makes online multimedia data transmission possible.

The large amount of multimedia data brings a new challenge. it is difficult for people to deal with the data only by sitting before monitors and keeping their eyes open all the time. To solve the problem, multimedia data mining techniques can be a good help. Moreover, the current anti-terrorism situations are becoming more and more urgent and austere. Some work such as global pursuit of persons at large and tracking suspicious objects brings great challenges to government offices. Easyliving[1] is such an example that sets focus on personal home.

In order to help people complete public security jobs, this paper proposes a framework called Global Monitoring and Security Assistances System (GMSAS) based on IPv6 and multimedia data mining techniques.

2 Framework and Functions of GMSAS

As shown in Fig. 1, framework of GMSAS is a hierarchical and multi-layered system which includes global layer, continent layer, nation layer, and city layer. With lots of cameras on almost any place in the world, wherever you are, whenever you go, you are under protection and monitoring of governments. Each layer can offer monitoring and public services within its precinct, receive orders from upper layers and distributes orders to down layers.

Global layer is the topmost layer in which inter-continent and inter-country public security tasks such as suspects pursuits can be easily implemented by sending orders to lower layers and letting them work together through distributed calculation. It can also make security evaluation for each country by analyzing multimedia data from them. If a bad security evaluation result for a country occurs, Global layer will send alarms to relative country centers immediately.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 658 – 659, 2005.

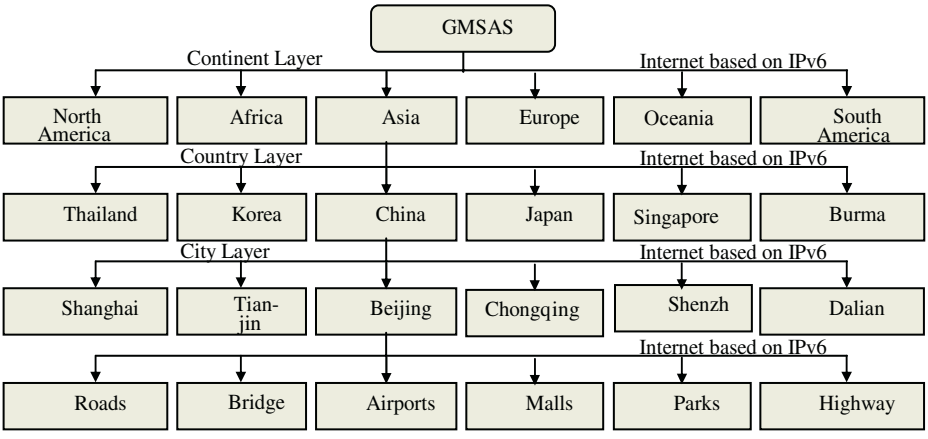


Fig. 1. Framework of Global Monitoring and Security Assistant System (GMSAS)

System functions of GMSAS can be divided into eight classes. a. Monitoring, b. Orders' distribution, c. Objects searching which means searching missing property or lost persons global-wide, continent-wide, country-wide, and city-wide. d. Doubtful objects identification which means with offline multimedia data mining procedure suspicious objects can be located through their daily tracks. e. Detection of urgent events which means through online analysis of video and audio data, urgent events such as theft, traffic accident, snowstorm, fire events and etc can be detected and located, much more quickly than traditional method. f. Evaluation of public security and pre-alarming. g. Private services.

In conclusion, based on IPv6 and multimedia data mining techniques, the proposed GMSAS framework integrates NGN based on IPv6 and distributed online and offline multimedia data mining techniques. Not only can it provide monitoring and security services for pubic places, it also offers private services for individuals.

Acknowledgement

This work is supported by the National Outstanding Young Scientist Research Grant 6025310 from the NNSF of China.

References

1. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., and Shafer, S.: Multi-Camera Multi-Person Tracking for EasyLiving, 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland, (2000) 3-10.

An Agent-Based Framework for a Traffic Security Management System

Shuming Tang[1,2] and Haijun Gao[2]

[1] Institute of Automation, Shandong Academy of Sciences, Jinan, Shandong, China
[2] The Key Laboratory of Complex Systems and Intelligence Science,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
sharron@ieee.org, zkyghj@vip.sina.com

1 Introduction

In the first 11 months of 2003, China had an increase of more than 35 percent on the number of vehicles [1]. If an incident lasted one more minute, traffic delay time would amount to 4~5min at a non-rush hour [2]. Statistical analysis also shows that traffic delay caused by little primary incident can be avoided if unprepared approaching drivers could be warned in time [3].

Under such background, here we propose to use agent-based control methods (ABC) and the principle of "local simple, remote complex (LSRC)" for networked systems [4] to design a framework for traffic security management systems (TSMS).

2 An Agent-Based Framework for the TSMS

The traffic security management system to manage traffic incidents should be composed of eight parts as shown in figure 1. We use mobile agents (MAs) to design its work flow chart in a TSMS (see Fig.1).

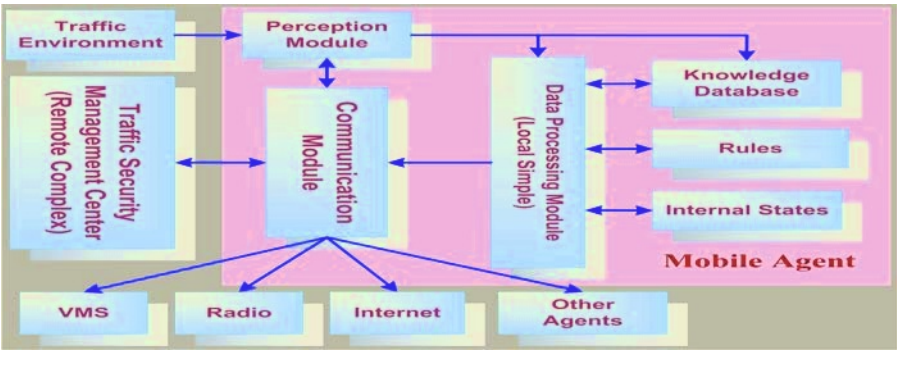


Fig. 1. The Framework of an Agent in the Traffic Security Management System

A major objective of a MA in the TSMS is to respond for an incident via broadcast incident message to available medias. The work procedure of the MA can be specified

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 660 – 661, 2005.

as: 1) Detection. The perception module perceives the variances of available parameters of traffic environment and sends the variance information to the knowledge database (KD) to backup traffic data, the data processing module (DPM) to verify the existence of a possible incident, and the communication module (CM) to inform the traffic management and control center (TMCC). 2) Verification. DPM is to verify an incident existence via local simple rules, for instance, simple fuzzy logic rules. 3) Response. The MA sends out the incident message via communication module to VMS, Radio, and Internet and the TMCC.

For an urban traffic area, the framework prototype of the traffic security management system based on ABC and LSRC is shown in Fig.2. In Fig.2, there are two types of mobile agents: OMA (Organizational Mobile Agent) and RMA (Regular Mobile Agent). Each RMA and OMA can freely roam in the whole urban traffic area.

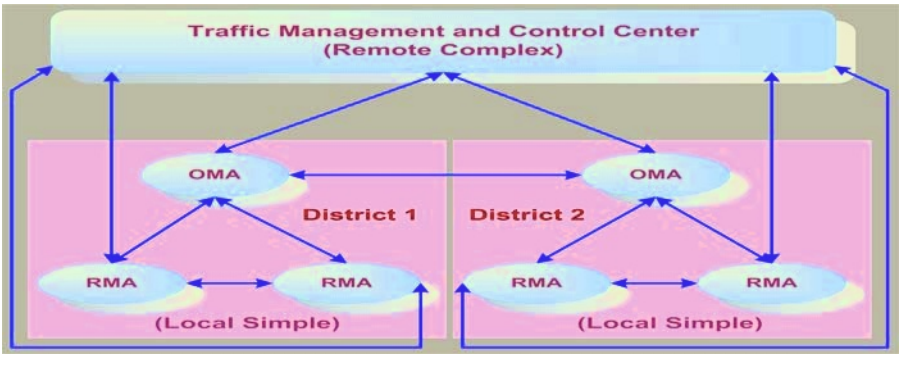


Fig. 2. A Framework Prototype of the TSMS Based on ABC and LSRC

Acknowledgements

The work is supported in part by the National Outstanding Young Scientist Research Awards for National Natural Science Foundation of China under Grant 6025310.

References

1. Fei-Yue Wang, Shuming Tang, Artificial Societies for Integrated and Sustainable Development of Metropolitan Systems, IEEE Intelligent Systems, 2004, Vol.19, No.4, pp.82-87.
2. Shuming Tang, Haijun Gao, A Traffic Incident Detection Algorithm Based on Nonparametric Regression, IEEE Transactions on Intelligent Transportation Systems, March 2005, Vol.6, No.1.
3. J. Versavel, Road safety through video detection, Proceedings of IEEE International Conference on Intelligent Transportation Systems, 1999, pp.753-757.
4. Fei-Yue Wang, Intelligent Control and Management for Networked Systems in a Connected World, Pattern Recognition and Artificial Intelligence, 2004, Vol.17, No.1, pp.1-6.

Application of a Decomposed Support Vector Machine Algorithm in Pedestrian Detection from a Moving Vehicle

Hong Qiao[2], Fei-Yue Wang[2], and Xianbin Cao[1]

[1] Department of Computer Science and Technology,
University of Science and Technology of China, Hefei, 230026, China
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, China
xbcao@ustc.edu.cn, hong.qiao@mail.ia.ac.cn, feiyue@gmail.com

1 Introduction

For a shape-based pedestrian detection system [1], the critical requirement for pedestrian detection from a moving vehicle is to both quickly and reliably determine if a moving figure is a pedestrian. This can be achieved by comparing the candidate pedestrian figure with the given pedestrian templates. However, due to the vast number of templates stored, it is difficult to make the matching process fast and reliable. Therefore many pedestrian detection systems [2, 3, 4] re developed to help the matching process. In this paper, we apply a decomposed SVM algorithm in the matching process which can fulfill the recognition task efficiently.

2 The Decomposition SVM Algorithm

Consider a group of training data. The purpose of SVM is to find out a hyper-plane which can maximally separate the group. The dual optimization problem of the SVM can be defined as [5]:

$$\min(L_D) = \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j x_i . x_j - \sum_{i=1}^{l}\alpha_i = \tfrac{1}{2}\alpha^T Q\alpha - e^T \tag{2a}$$

$$\text{Subject to: } 0 \le \alpha_i \le C \tag{2b}$$

$$y^T\alpha = 0$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots, \alpha_l)^T$, $i = 1, \ldots, l$, $Q = (Q_{ij})_{i,j=1,\ldots,l}$ $Q_{i,j} = y_i y_j x_i x_j$ is the Hessian matrix.

In this paper we will use Joachims' decomposition algorithm [6] to solve (2). The general steps in a decomposition algorithm can be reviewed as follows:

1) Initialize $\alpha = \alpha^1$ and set $k = 1$, where k is the iteration step.
2) If α^k is the solution of Equation (2), stop. Otherwise, the whole training set is separated into 2 sets: the +working set B with size q, and the non-

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 662 – 663, 2005.

working set N with the size $l-q$ where l is the number of all training samples. Accordingly, the vector α^k is divided into 2 sets: α_B^k and α_N^k.

3) Solve the following sub-quadratic problem with variable α_B^k:

$$\begin{aligned} &\min\left(\frac{1}{2}\left(\alpha_B^k\right)^T Q_{BB}^k\left(\alpha_B^k\right)-\left(e_B-Q_{BN}^k\alpha_N^k\right)^T\alpha_B^k\right) \\ &\text{subject to: } 0\le\left(\alpha_B^k\right)_i\le C,\quad i=1,\cdots,q \\ &\qquad\left(y_B^k\right)^T\alpha_B^k+\left(y_N^k\right)^T\alpha_N=0 \end{aligned} \tag{3}$$

where e_B and e_N are vectors with all elements being 1.

4) Set α_B^{k+1} to be the solution of step 3) and if $\alpha_N^{k+1}=\alpha_N^k$ then set $k\leftarrow k+1$ and go to step 2).

Compared with the usual SVM algorithm, the decomposed SVM algorithm proposed in this paper has the following features:

(a) The proposed algorithm is stable and convergent.
(b) The algorithm is suitable to the case where the amount of training samples is large.
(c) The algorithm is suitable for a dynamic training process.
(d) The training process is faster.

Acknowledgement

This work is supported in part by Grants (2004AA1Z2360 and 2004GG1104001) and an open research project from KLCSIS.

References

1. A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi, Shape-based pedestrian detection, Proc. IEEE Intell. Veh. Symp., pp.215-220, 2000.
2. D. M. Gavrila, Pedestrian detection from a moving vehicle, Proc. Eur. Conf. Comp., vol.2, pp. 37-49, 2000.
3. L.Zhao and C.Thorpe, Stereo-and neural network-based pedestrian detection, Procs. IEEE Intl. Conf. on Intelligent Transportation System'99, (Tokyo, Japan), pp.298-303, Oct. 1999.
4. C. Papageorgiou, T. Evgeniou, and T. Poggio, A trainable pedestrian detection system, Proc. IEEE Intelligent Vehicles Symposium'98, (Stuttgart, Germany), pp.241-246, Oct. 1998.
5. C.J.C. Bugres, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.
6. T. Joachims, "Making large-scale SVM learning practical," in: Advances in Kernel Methods-Support Vector Learning (B. Schölkopf, C.J.C. Burges, and A.J. Smola, Eds.), Cambridge, MA: MIT Press, 1998.

Application of Cooperative Co-evolution in Pedestrian Detection Systems

Xianbin Cao[1], Hong Qiao[2], Fei-Yue Wang[2], and Xinzheng Zhang[1]

[1] Department of Computer Science and Technology,
University of Science and Technology of China, Hefei, 230026, China
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, 10080, China
xbcao@ustc.edu.cn, hong.qiao@mail.ia.ac.cn, feiyue@gmail.com

1 Introduction and Related Work

In general, a shape-based[1] pedestrian detection system includes the following two steps:

(a) finding out and tracking a possible pedestrian figure, and
(b) determining if the candidate pedestrian figure is really a pedestrian figure by checking if it matches with any of the pedestrian templates.

Since there are a large number of templates, it is necessary to build up a search tree for the match process [2,3]. Each node in the tree is one feature of the corresponding templates that can be used for classification and where each branch is one pedestrian template. Usually, the search tree is not adjustable during the matching process.

2 Procedure of the Design

In this paper, we use a new cooperative co-evolutionary algorithm (CCEA) to setup a template search tree which can be adjusted during the matching process. This algorithm has the following advantages over the normal cooperative co-evolution method.

(a) The algorithm is suitable for the cases where the search space is large.
(b) The cooperative co-evolutionary algorithm has low requirement on the knowledge of the samples in the search space.
(c) The reduction of the population size reduces the computational load.
(d) The new algorithm possesses the ability of dynamically controlling the size of the sub-communities.

The searching tree is built by using the new CCEA. It can be described as follows:

(1) Once a set of templates is put into the new CCEA, it will cluster the templates that are highly similar into one group. In each clustered templates group, a representative template will be selected and can be seen as the father node of the whole group members.
(2) Each run of the new algorithm will generate a layer of the searching tree. In order to add the searching tree with self-learning and adjusting abilities, the system may select and save a recognized pedestrian figure as a new template.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 664 – 665, 2005.

3 Experimental Results

The experimental results showed that for a templates library which contains 5000 templates, the system set up a 7-layer tree and the average matching time was 15. the recognition of a candidate pedestrian is very fast and accurate.

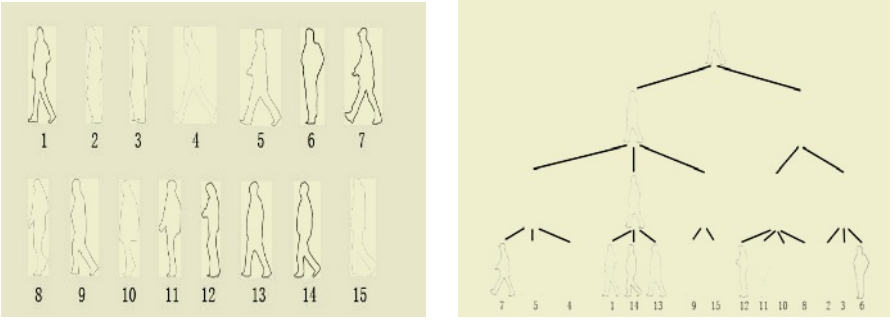


Fig. 1. Template searching tree

In conclusion, this paper proposed a new CCEA to build and adjust a template searching tree efficiently and adaptively for the matching process of the pedestrian detection system.

Acknowledgement

This work is supported in part by Grants (2004AA1Z2360 and 2004GG1104001) and an open research project from KLCSIS.

References

1. A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi, Shape-based pedestrian detection, Proc. IEEE Intell. Veh. Symp., pp.215-220, 2000.
2. D. M. Gavrila, Pedestrian detection from a moving vehicle, Proc. Eur. Conf. Comp., vol.2, pp. 37-49, 2000.
3. Gavrila, D.M., Giebel, J., Munder, S. Vision-based pedestrian detection: the PROTECTOR system, IEEE IVS04, 2004.
4. J M Emlen, Population biology: the co-evolution of population dynamics and behaviors. Macmillan Publishing Company,New York,1984.
5. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.J.; Comparing images using the Hausdorff distance, IEEE Trans on Pattern Analysis and Machine Intelligence 15 (9) , pp.850 – 863, Sept. 1993.

Biometric Fingerprints Based Radio Frequency Identification

S. Jayakumar[1] and C. Senthilkumar[2]

[1] Senior Lecturer, Department of Electronics and Communication,
Amrita Vishwa Vidyapeetham, Coimbatore,
TamilNadu - 641105, India
s_jayakumar2@ettimadai.amrita.edu
[2] Associate Software Engineer, Software Engineering Services,
Torry Harris Business Solutions, Bangalore,
Karnataka – 560025, India
csenthilkumar_ece@rediffmail.com

1 Introduction

In recent years, Radio Frequency Identification procedures have become very popular in various aspects of life. Radio frequency identification, or RFID, is a generic term for technologies that use radio waves to automatically identify people or objects. There are several methods of identification, but the most common is to store a serial number that identifies a person or object. In most of the cases the serial number is usually the roll number of the person or the serial number of the associated object. The most notable disadvantage of such an automated identification system is their inability to avoid the miss use of RFID tags.

In this write-up we propose a method to incorporate the Biometric Fingerprint Authentication technology with the existing Radio Frequency Identification systems. The goal of our work is to develop a more reliable Radio Frequency Identification system, which can prevent the misuse of tags. In the proposed system we replace the serial number with codes similar to EPC. We use the data obtained by processing the user fingerprint for generating the mentioned codes.

2 Radio Frequency Identification

At its most basic level, RFID is a short-range radio communication to uniquely identify objects or people. RFID systems include electronic devices called transponders or tags, and reader electronics to communicate with the tags.

An RFID tag consists of a microchip attached to an antenna. Tags are either active or passive. An RFID reader serves the same purpose as a barcode scanner. The reader captures the RF waves from tags and converts into digital data. The RFID reader handles the communication between the Information System and the RFID tag. An RFID antenna activates the RFID tag and transfers data by emitting wireless pulses.

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 666 – 668, 2005.

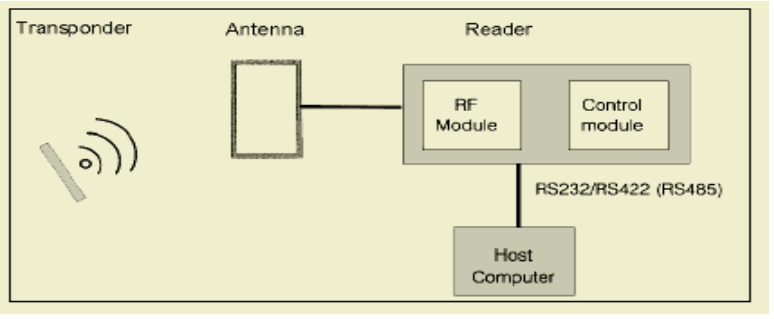


Fig. 1. Radio Frequency Identification System. This figure shows the various RFID system components

3 Biometric Fingerprint Authentication

Ridges and valleys form fingerprints. These ridges and valleys take unique form one person to other. Fingerprint authentication systems use features like ridge endings, ridge bifurcation, islands, enclosures etc. These features of the fingerprint are called as the fingerprint minutiae. Biometric authentication system on scanning and processing the fingerprint extracts the minutiae features from the fingerprint. Minutiae extracted are stored as templates in the database. Whenever authentication is required minutiae are extracted from the scanned fingerprint and compared with the stored templates in the database. The notable disadvantage in this highly effective and secured authentication system is the storage space required for storing the templates.

In the proposed system the fingerprint minutiae templates are converted into unique integers and these integers are stored in the database instead of fingerprint minutiae templates. During authentication the same algorithm is used to process the scanned fingerprint and the unique integer representing the scanned fingerprint is compared with those in the database with an allowed margin of deviation. For converting the minutiae templates to unique integers we plot the minutiae position in a two dimensional plane and generate a relative common reference point, which remains same irrespective of rotation of the fingerprint. The distance between this reference point and all the minutiae points are obtained and summed up. This summed up value, on scaling linearly results in unique integers to represent fingerprints.

4 Proposed EPC Format for Biometric – RFID

Biometric RFID tags stores data obtained using the unique integers generated by fingerprint authentication system. Proposed data format to be used in biometric RFID system environment is similar to the EPC format. The data to be stored in the RFID tag to represent a human is proposed to have continent code, country code, state code and scaled unique integer from fingerprint processing. The scaling can be reader specific giving way for encryption and privacy for the user. This system by using the unique integers obtained from fingerprint processing incorporates the reliability of fingerprint authentication with the RFID technology.

5 Is Biometric RFID a More Reliable Option?

Biometric RFID system proposed above is more reliable compared to the existing RFID technology. This greatly avoids the misuse of Radio Frequency Identification Tags. Automated entry points to highly restricted environments can use both the RFID and fingerprint authentication for authentication where us the other access points in same the environment can use RFID technology. By using Biometric RFID we avoid misuse of RFID tags and also enjoy using the contact less RFID authentication system at most of the access points. Since tags store data with specific scaling, privacy is ensured for the users. This makes RFID technology a more reliable one to be used and extends usage of RFID technology at highly secured environments and for human tracking.

6 Enhanced Applications

This approach introduces a fingerprint authentication, which uses fingerprint templates in integer formats rather than collection of minutiae points. This drastically reduces the memory needed to store a fingerprint template and also results in a faster fingerprint authentication algorithm. Apart from contributing to the biometric fingerprint authentication by incorporating the results from fingerprint processing in RFID technology, RFID system is made a more reliable technology. Applications involving human such as human tracking and monitoring systems are made more secured increasing the privacy of the user.

BorderSafe: Cross-Jurisdictional Information Sharing, Analysis, and Visualization

Siddharth Kaza[1], Byron Marshall[1], Jennifer Xu[1], Alan G. Wang[1], Hemanth Gowda[1], Homa Atabakhsh[1], Tim Petersen[2], Chuck Violette[2], and Hsinchun Chen[1]

[1] Department of Management Information Systems, University of Arizona, Tucson, Arizona
{skaza, byronm, jxu, gang, homa, hchen}@eller.arizona.edu
heman@email.arizona.edu
[2] Tucson Police Department, Tucson, Arizona
{tim.petersen, chuck.violette}@tucsonaz.gov

1 Project Background

The BorderSafe project funded by Department of Homeland Security (DHS) and the Corporation for National Research Initiatives (CNRI) aims to develop, foster, and leverage information sharing between law enforcement agencies for border safety and national security. The partners in the project include the Artificial Intelligence (AI) Lab at the University of Arizona, Tucson Police Department (TPD), Pima County Sheriff's Department (PCSD), Tucson Customs and Border Protection (CBP), San Diego Automated Regional Justice Information System (ARJIS), and the San Diego Supercomputer Center (SDSC). We describe the three major areas of research in the BorderSafe project at the AI Lab, University of Arizona.

2 Criminal Activity Network Analysis

A criminal activity network (CAN) is a network of interconnected people (often known criminals), vehicles, and locations based on law enforcement records. These networks aid in identifying suspicious individuals, vehicles, and locations based on data from multiple tiers of law enforcement agencies. Cross-jurisdictional information sharing and triangulation can help generate better investigative leads and strengthen legal cases against criminals. In the BorderSafe project, CANs are used to explore the criminal links of individuals and vehicles based on local police and border crossing records. The analysis has provided valuable results for law enforcement.

3 Critical Infrastructure Protection

Homeland security concerns include protecting critical infrastructures like power plants, water treatment plants, airports etc. Incidents that might pose threat to infrastructures are recorded in local law enforcement datasets. Analysis of these incidents can be used to set up alerts for individuals and vehicles involved in

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 669 – 670, 2005.

suspicious activity around critical infrastructures. The locations of critical infrastructures and police incidents are geo-coded using the state plane coordinate system used by ESRI . The AI Lab's Spatio-Temporal visualizer is used to analyze and plot the incidents around the critical infrastructure.

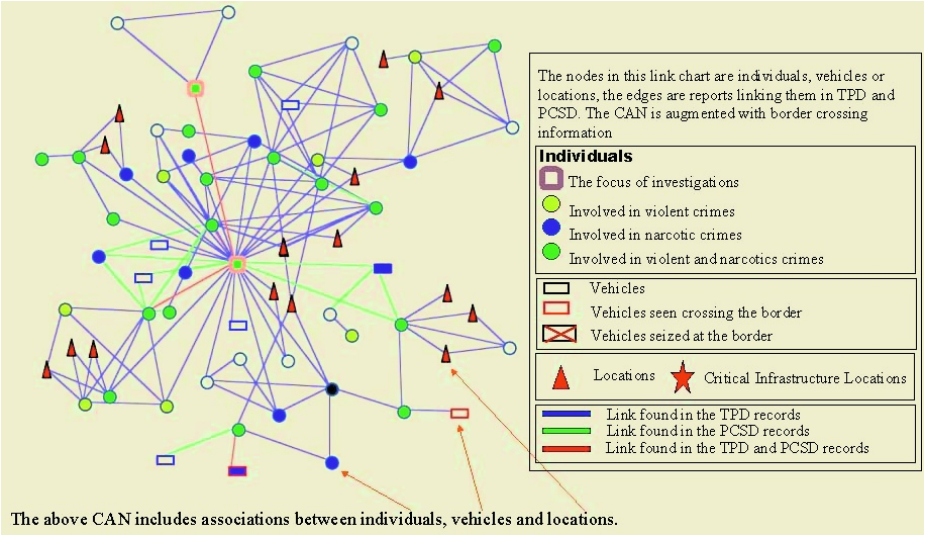


Fig. 1. An example Criminal Activity Network

4 Deception Detection

In law enforcement it is critical to penetrate identity concealment attempts as quickly and as effectively as possible. Our research goal is to create algorithms that automatically analyze law enforcement datasets to produce a ranked list of different individual identity entries that are likely to represent the same individual, and to tag each identity record with an estimate of the probability that it represents an active attempt at identity concealment. We have conducted a case study on real concealment cases identified by a police detective and developed a taxonomy of identity concealment. Probability-based detection algorithms, such as record linkage and Bayesian network, are currently being developed.

Author Index

Abbasi, Ahmed 183
Acar, Evrim 256
Adam, Nabil R. 1
Ahamad, Mustaque 492
Ahn, Gail-Joon 479
Aleman-Meza, Boanerges 486, 621
Arpinar, I. Budak 621
Ascher Michael 612
Atabakhsh, Homa 368, 669
Atluri, Vijayalakshmi 1
Atwood, Michael 60

Backhouse, James 600
Badia, Antonio 49
Bae, Hae-Young 642
Balasubramanian, Niranjan 381
Bapi, Raju S. 498
Baumes, Jeffrey 27
Best, Clive 436, 595
Bobeica, Mihaela 595
Bradford, R.B. 374
Brewer, Isaac 429
Burgoon, Judee K. 198, 471, 597
Burns, Phillip 486

Cai, Guoray 429
Çamtepe, Seyit A. 256
Cangussu, João W. 127, 636
Canhoto, Ana I. 600
Cao, Xianbin 662, 664
Chang, Kuiyu 37
Chang, LiWu 72
Chang, Wei 412, 612
Chen, Hsinchun 183, 287, 322, 368, 402, 412, 612, 619, 623, 669
Chen, Rui 81
Chen, Xinjian 549
Chen, Yufeng 505, 517
Choudhary, Alok 530, 543
Conti, Gregory 492

Dantu, Ram 115, 127, 636
Davis, Boyd 608
Davis, David 305
Demers, Alan 617
de Paola, Monica 436
Dong, Guozhu 171
Dong, Yabo 505
Dugan, Laura 340

Eavenson, Matthew 486
Eidson, Millicent 612
Elovici, Yuval 244

Famtepe, Aseyit 256
Forsgren, Nicole 471
Frendak-Blume, Allison 305
Friedman, Menahem 244
Fuhrmann, Sven 429
Fuller, Doug 511
Fung, Benjamin C.M. 171

Galloway, John 14
Gandhi, Robin A. 479
Gao, Haijun 658, 660
Garcia, Teofilo 595
Gehrke, Johannes 617
Gelbart, Olga 530
Goharian, Nazli 604
Goldberg, Mark 27
Gong, Xiaoyan 658
Gotham, Ivan 612
Gowda, Hemanth 669
Gunaratna, Rohan K. 37

Halaschek-Wiener, Christian 621
Han, Dianwei 459
Honavar, Vasant 511
Hong, Mingsheng 617
Hong, Sungjin 638
Hou, Zeng-Guang 610
Hu, Daning 287

Hu, Paul J.-H. 412
Hu, Xiaohua 60

Im, Eul Gyu 634, 650

Janeja, Vandana P. 1
Jayakumar, Sundaram 666
Jensen, Matthew L. 198
Jéral, Jean-Paul 595
Jian, Wei 517
Jiang, YiChuan 640

Kandel, Abraham 244
Kang, Dae-Ki 511
Kantardzic, Mehmed 49
Kayacik, H. Günes 362
Kaza, Siddharth 669
Khalsa, Sundri K. 561
Kolan, Prakash 115
Kooptiwoot, Suwimon 593
Koppel, Moshe 209, 269
Kotapati, Kameswari 631
Krishnamoorthy, Mukkai S. 256
Kruse, John 198
Kumar, Pradeep 498
Kwon, Taekyoung 99

LaFree, Gary 340
Laha, Arijit 498
Lai, Guanpi 402, 623, 625
LaPorta, Thomas F. 631
Larson, Catherine 412, 612
Last, Mark 244
Latino, Robert J. 579
Lee, Heejo 638
Lee, Seok Won 479
Li, Haifeng 448, 627
Liddy, Elizabeth D. 381
Lim, Ee-Peng 37
Liu, Da-xin 606
Liu, Peng 631
Lord, Vivian 608
Lu, Dongming 505
Lui, Chunju 412
Luo, Mingchun 536

Lutterbie, Simon 465
Lynch, Cecil 612

Ma, Ling 604
MacEachren, Alan M. 429
Magdon-Ismail, Malik 27
Majumdar, Anirban 648
Manley, Michael 629
Marshall, Byron 669
Mason, Peyton 608
Matwin, Stan 72
McEntee, Cheri 629
McNeese, Mike 429
Meservy, Thomas O. 198
Meyers, Chris 604
Mohan, Kripashankar 543
Molet, Anthony 629
Moon, Hyeonjoon 99

Naqvi, Syed 654
Narahari, Bhagirath 530, 543
Norback, Robert 492
Nunamaker, Jr., Jay F. 198, 471

Oh, Jaewook 638
Oh, Young-Hwan 642
Ong, Teng-Kwee 37
Ott, Paul 530, 543

Palaniswami, Devanand 486
Pan, Yunhe 505
Park, Beomsoo 638
Park, Joon S. 629
Peng, Ying 590
Petersen, Tim 368, 669
Piquero, Alex R. 340
Price, Robert 602
Probst, Peter S. 316

Qiao, Hong 662, 664
Qin, Jialun 287, 402, 623
Qin, Tiantian 597

Radha Krishna, P. 498
Radlauer, Don 153
Rao, H. Raghav 81

Rao, M.Venkateswara 498
Reid, Edna 322, 402, 623
Riedewald, Mirek 617
Riguidel, Michel 654
Robles-Flores, Jose A. 418
Roussinov, Dmitri 418
Rumm, Peter 60

Sageman, Marc 287, 402, 623
Sahoo, Satya Sanket 621
Salam, Muhammad Abdus 593
Schler, Jonathan 209
Schneider, Moti 244
Schumaker, Rob 619
Senthilkumar, Chandramohan 666
Shahar, Yael 139, 554
Shapira, Bracha 244
Sharma, Rajeev 429
Sharman, Raj 81
Shen, Huizhang 590
Sheth, Amit 486, 621
Simha, Rahul 530, 543
Simoff, Simeon J. 14
Sinai, Joshua 280, 567
Siraj, Ambareen 218
Skillicorn, David B. 231
Smith, Richard A. 652
Solewicz, Yosef A. 269
Song, Weiwei 448, 627
Song, Yong Ho 634, 650
Stockwell, David R.B. 523
Su, Qi 549
Sun, Shuang 422
Sun, Wei 606
Sun, Yan 631
Sun, Zhen 37
Symonenko, Svetlana 381

Tang, Shuming 660
Tao, Qing 395
Thomborson, Clark 454, 646, 648
Tian, Jie 549
Tousley, Scott 571
Tseng, Chunju 412, 612
Turi, Janos 127
Twitchell, Douglas P. 471

Upadhyaya, Shambhu J. 81

Vaidya, Jaideep 1
Van Der Goot, Erik 436
Vaughn, Rayford B. 218
Violette, Chuck 669

Wang, Alan G. 368, 669
Wang, Fei-Yue 395, 454, 549, 625, 662, 664
Wang, Jason T.L. 523
Wang, Jian 640
Wang, Jie 459
Wang, Jue 395
Wang, Ke 171
Wang, Shiuh-Jeng 644
Wang, Shuxun 448, 627
Weimann, Gabriel 402
Wen, Quan 448, 627
Wheeler, Jennifer 305
Wiers, Karl 471
Wnek, Janusz 389
Wong, Kelvin H.L. 536
Woodcock, Alexander E.R. 305
Worrell III, Clarence 305
Wu, Gao-Wei 395

Xia, ZhengYou 640
Xiang, Zhengtao 517
Xu, Jennifer J. 287, 669
Xu, Shuting 459

Yan, Fei 517
Yang, Cheng-Hsing 644
Yang, Xin 549
Yavagal, Deepak S. 479
Yen, John 422
Yener, Bülent 256
Yilmazel, Ozgur 381
Yoo, Illhoi 60

Zaafrany, Omer 244
Zambreno, Joseph 530, 543
Zeng, Daniel 412, 612

Zhan, Justin 72
Zhang, Jun 459
Zhang, Xinzheng 664
Zhao, Jidi 590
Zhou, Lina 465
Zhou, Yilu 402, 623
Zhu, Jin 614
Zhu, William 454, 646
Zigdon, Kfir 209
Zimmermann, David 656
Zincir-Heywood, Nur 362
Zukas, Anthony 602