Preface

Computer supported communication and infrastructure are integral parts of modern economy. Their security is of incredible importance to a wide variety of practical domains ranging from Internet service providers to the banking industry and e-commerce, from corporate networks to the intelligence community.

The CSI-KDD workshop focuses on novel knowledge discovery methods addressing CyberSecurity and intelligence issues as well as innovative applications demonstrating the effectiveness of data mining in solving real-world security problems. The challenge for novel methods originates from the emergence of new types of contents and protocols, and only an integrated view on all modes promises optimal results. Innovative applications are essential as IT-communication as well as computer-supported technical and social infrastructure have an extremely complex structure and require a comprehensive approach to prevent criminal activities.

As an invited speaker we welcome André Bergholz, Fraunhofer IAIS. He will report on the lessons learnt on phishing filtering in the specific targeted research project AntiPhish funded by the European Union. He reports on filter methodologies evaluated in a test laboratory setting, and describes the application of this technology to real world email streams, to be used to filter all email traffic online in real time. In the afternoon there will be an invited talk entitled “Data Security and Integrity: Developments and Directions” given by Bhavani Thuraisingham. The talk is concentrated ensuring that only authorized individuals have address to data and data is protected from malicious corruption.

The workshop is organized in two tracks. In the first track "Novel Knowledge Discovery Methods for the Security Domain" is targeted to advanced data mining approaches for CyberSecurity. The second track "Innovative Techniques and Applications in Intelligence Informatics" concentrates on large-scale security applications.

Despite the fact that various types of security mechanisms have been defined and are widely deployed to prevent malicious users from launching attacks, cyber criminals remain pretty successful in misusing useful protocols and applications for their own benefits. Spam and phishing campaigns are routinely carried out. Drive by download attacks are among the major threats facing normal users surfing the web. BGP hijacks corrupting internet routing tables are becoming known to the public as well. Etc.

The first track considers data mining approaches, which can help addressing security issues according to, at least, three distinct axes. First of all, thanks the very large amount of application logs of various kinds available, it can be a valid approach to better understand the attacks we are facing and to help designing better preventive and, or, detection mechanisms in order to respond to these attacks. On the other hand, by analyzing traffic and other attack related traces data mining can be helpful in getting a better picture of who is attacking us tackling the ”attack attribution problem” under the umbrella of e-forensics techniques. Finally data mining can be employed to detect attacks by classifying content, e.g. by filtering spam and phishing messages.

Still a number of research issues remain. Different types of content have to be analyzed (e.g. email, websites, embedded images, transmitted code, activity logs), and only an integrated view on all modes promises optimal results. This, for instance, involves mining the different media associated with an email and combining the results to improve accuracy.

Spammers, hackers and producers of fraudulent content continuously change their tactics, requiring adaptive and even anticipatory mining techniques. As intelligent opponents aim at thwarting analyses, specific approaches for analysis are required. In addition, many new forms of messaging (e.g., SMS, MMS), often anchored in a mobile environment, become a victim of malicious manipulations.

For the first track we accepted four very interesting oral presentations covering topics such as intrusion and malware detection, attack attribution and spam filtering. The training time of intrusion detection models is often computationally expensive, hence the interest in efficient models while still assuring a high predictive accuracy of the intrusion detection. Chen Yo-Shu and Chen Yi-Ming present this view in the paper "Combining Incremental Hidden Markov Model and Adaboost Algorithm for Anomaly Intrusion Detection".

In the paper entitled “Addressing the Attack Attribution Problem using Knowledge Discovery and Multi-criteria Fuzzy Decision-Making", O. Thonnard, W. Mees and M. Dacier propose an analysis framework to reason about the root causes of attacks observed on the Internet. They apply it to some large real world datasets and derive interesting insights on what they call armies of zombies.

"Malware Detection using Statistical Analysis of Byte-Level File Content" by S. Momina Tabish, M. Zubair Shafiq and Muddassar Farooq discusses non-signature based malware detection. The approach is successful and assumes that benign files are quite distinct from malware files considering a byte-level format. A very novel and interesting approach inspired by gaming theories is presented and applied on phishing mail filtering in an adversarial setting by Gaston L’Huillier, Richard Weber and Nicolas Figueroa (Online Phishing Classification Using Adversarial Data Mining and Signaling Games).

The second track is called "Innovative Techniques and Applications in Intelligence Informatics". Intelligence Informatics is concerned with the study of the development and use of advanced information technologies and systems for national, international, and societal security-related applications. The annual IEEE International Conference series on Intelligence and Security Informatics with over two hundred attendants was started in 2003. In addition, the Pacific Asia Workshop on Intelligence and Security Informatics with over eighty attendees has been started in 2006. These intelligence and security informatics events have brought together academic researchers, law enforcement and intelligence experts, information technology consultants and practitioners to discuss their research and practice related to various intelligence and security informatics topics. Among these research topics in intelligence security, there is a strong focus on data mining and knowledge discovery. It is the first attempt to introduce intelligence informatics to the ACM SIGKDD community. The four major topics of intelligence security include (a) information sharing and data/text/web mining, (b) infrastructure protection and emergency responses, (c) terrorism informatics, and (d) enterprise risk management and information system security.

In information sharing and data/text/Web mining, we focus on criminal data mining, criminal/intelligence information sharing and visualization, cyber crime detection and analysis, authorship analysis, deception detection and analysis, and information sharing governance. We investigate how to use advanced data sharing and mining techniques to support law enforcement and intelligent experts in their investigations so that effective results can be achieved efficiently.

In infrastructure protection and emergency responses, we explore several interesting infrastructure problems such as bioterrorism information infrastructure, transportation and communication infrastructure protection, cyber-infrastructure design and protection, border safety, disaster prevention, detection and management, and emergency response and management. As we can see in recent natural disasters and terror attacks, a good infrastructure protection and emergency response management will minimize damages and recover from devastation in a shorter amount of time.

In terrorism informatics, we investigate several terrorism related informatics problems. For instances, we investigate terrorism related analytical methodologies and software tools, terrorism knowledge portals and databases, terrorist incident chronology databases, terrorism root cause analysis, social network analysis, forecasting and countering terrorism and measuring the effectiveness of counter-terrorism campaigns.

In the recent years, we have also included enterprise risk management and information systems security, in which we examine information security management standards, information systems security policies, fraud detection, board activism and influence, corporate sentiment surveillance, market influence analytics and medial intelligence, and consumer-generated media and social media analytics.

The program committee has selected five papers in intelligence informatics for presentation. Park and Treglia developed a model and theory of intelligence information sharing through a literature review, experience and interviews with practitioners. Yang and Tang proposed a subgraph generalization approach to share and integrate terrorist or criminal social network data between different intelligence and law enforcement units and preserve the privacy of individuals in social networks. Such social network sharing and integration technique improves the performance of social network analysis such as centrality measurements. Bhavani et al. investigated the information management component for military stabilization and reconstruction operations. The temporal service oriented architecture system (TG-SOA), which utilized the temporal geosocial semantic web to manage the lifecycle of stabilization and reconstruction operations, was developed. Senator examined the common criticisms on data mining applications for security. These criticisms argued that the data mining applications were ineffective and threatening civil liberties. He analyzed these criticisms by modeling a phenomena and proposing alternative designs. Kwok et al. studied the security problems in public companies' web servers against cyber attacks. The study included ten Hong Kong Hang Send Index companies and ten Hong Kong China Enterprises Index companies. A pyramid risk analysis tool was also proposed.

This workshop would not be possible without the invited speakers, the authors and the 37 members of the program committee. We express our gratitude towards them. We would thank also Fraunhofer IAIS, who was in care of the CSI-KDD 2009 website as well of the submission system.

Hsinchun Chen
Marc Dacier
Marie-Francine Moens
Gerhard Paass
Christopher C. Yang

CSI-KDD 2009 Organizers and Program Committee

Organizers

Hsinchun Chen	The University of Arizona, Tucson, USA	Track 2
Marc Dacier	Symantec Research Labs Europe, France	Track 1
Marie-Francine Moens	K.U. Leuven, Belgium	Track 1
Gerhard Paass	Fraunhofer IAIS, St. Augustin, Germany (contact)	Track 1
Christopher C. Yang	Drexel University, Philadelphia, USA	Track 2

Program Committee

Adedeji B. Badiru	Air Force Institute of Technology, Dayton, OH, USA.
Yigal Arens	USC/ISI, USA
John Aycock	University of Calgary, Canada
Antonio Badia	University of Louisville, USA
Andre Bergholz	Fraunhofer IAIS, Germany
Ulf Brefeld	MPI Saarbrücken, Germany
Patrick S. Chen	Tatung University, Taiwan
Robert W.P. Chang	Criminal Investigation Bureau, Taiwan
Domenico Dato	Tiscali Services, Italy
Yuval Elovici	Ben-Gurion University, Israel
Uwe Glaesser	Simon Fraser University, Canada
Nazli Goharian	Illinois Institute of Technology, USA
Mark Goldberg	RPI, USA
Henrik Grosskreutz	Fraunhofer IAIS, Germany
David Hicks	Aalborg University Esbjerg, Denmark
Thorsten Holz	University of Mannheim, Germany
Patrick Horkan	Symantec, Ireland
Sotiris Ioannidis	Institute of Computer Science, Greece
Latifur Khan	University of Texas at Dallas, USA
Engin Kirda	Eurecom, France
Christopher Kruegel	UC Santa Barbara, USA
Sheau-Dong Lang	University of Central Florida, USA
Ee-Peng Lim	Singapore management University, Singapore
Evangelos Markatos	Institute of Computer Science, Greece
Robert Moskovitch	Ben-Gurion University, Israel
William Pottenger	Rutgers University, USA
Raghav Rao	State University of New York at Buffalo, USA
Elliot Rich	University at Albany, SUNY, USA
Stefan Rüping	Fraunhofer IAIS, Germany
Bracha Shapira	Ben-Gurion University, Israel
David Skillicorn	Queen's University, Canada
Randy Smith	University of Alabama, USA
Paul Thompson	Dartmouth College, USA
Cedric Ulmer	SAP Research, France
Nalini Venkatasubramanian	University of California, Irvine, USA
Zhao Xu	Fraunhofer IAIS, Germany
Urko Zurutuza	Mondragon University, Spain

Table of Contents

Workshop Program

09:00 -10:00 Invited Talk

- *09:00 -10:00:* *AntiPhish – Lessons Learnt*
  André Bergholz

10:00-10:30 Coffee Break

10:30-12:30 Track1: Novel Knowledge Discovery Methods for the Security Domain

- *10:30-11:00* *Combining Incremental Hidden Markov Model and Adaboost Algorithm for Anomaly Intrusion Detection,*
  Chen Yo-Shu and Chen Yi-Ming
- *11:00-11:30* *Addressing the Attack Attribution Problem using Knowledge Discovery and Multi-criteria Fuzzy Decision-Making,*
  O. Thonnard, W. Mees and M. Dacier
- *11:30 – 12:00* *Malware Detection Using Statistical Analysis of Byte-Level File Content*
  S. Momina Tabish, M. Zubair Shafiq and Muddassar Farooq
- *12:00 – 12:30* *Online Phishing Classification Using Adversarial Data Mining and Signaling Games*
  Gaston L'Huillier, Richard Weber and Nicolas Figueroa

12:30-14:00 Lunch

14:00-14:40 Invited Talk

- *14:00 -14:40:* *Data Security and Integrity: Developments and Directions*
  Bhavani Thuraisingham

14:40-15:30 Track 2A: Innovative Techniques and Applications in Intelligence Informatics

- *14:40-15:05:* *Towards Trusted Intelligence Information Sharing,*
  Joon Park and Joseph Treglia
- *15:05-15:30:* *Social Networks Integration and Privacy Preservation using Subgraph Generalization,*
  Christopher C. Yang and Xuning Tang

15:30-16:00 Coffee Break

16:00-17:00 Track 2B: Innovative Techniques and Applications in Intelligence Informatics

- *16:00-16:30:* *Design of a Temporal Geosocial Semantic Web for Military Stabilization and Reconstruction Operations,*
  Bhavani Thuraisingham, Latifur Khan, Murat Kantarcioglu, and Vaibhav Khadilkar
- *16:30-17:00* *On the Efficacy of Data Mining for Security Applications*
  Ted Senator
- *17:00-17:30* *A Study of Online Service and Information Exposure of Public Companies*
  Sai Ho Kwok, Cheuk Tung Lai and Jason Yeung