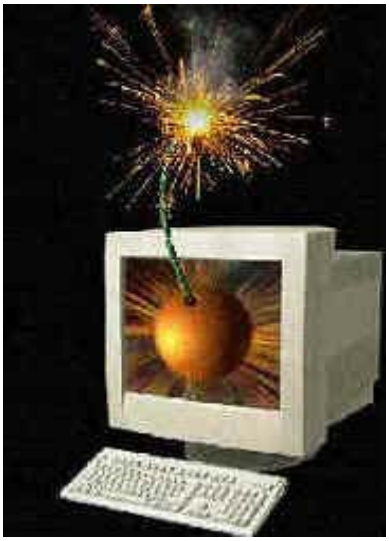
Workshop on

Data Mining for Cyber Threat Analysis

in conjunction with
IEEE International Conference on Data Mining

December 9-12, 2002

Maebashi TERRSA, Maebashi City, Japan



Workshop Organizers:

Vipin Kumar, University of Minnesota
Aleksandar Lazarevic, University of Minnesota
Jaideep Srivastava, University of Minnesota

Workshop Objectives

People have always depended upon information technology of some type, beginning with smoke signals in ancient days and turning into network-based computer systems today. Information technology becomes an essential part of the way various organizations function. Nowadays, the computers control power, oil and gas delivery, communications, transportation, banking and financial services. They are used to store and exchange vital information, from publicly know facts to well kept secrets.

Notwithstanding the tremendous benefits that the emergence of this technology brings, there is inevitably an escalation of "dark side of the force" in the form of cyber terrorism. As a form of convergence between cyberspace and terrorism, cyber terrorism "refers to unlawful attacks and threats of attack against computers, networks, and the information stored therein when done to intimidate or coerce a government or its people …".

As the cost of the information processing and Internet accessibility falls, more and more organizations will be vulnerable by potential cyber threats. According to a recent research survey, cyber attacks have increased by almost 80 percent over the last six months. This indicates that there is an urgent need to expand efforts in the battle against cyber terrorism. The key question is whether contemporary computer technologies such as artificial intelligence and data mining can contribute to this battle and further enhance defense mechanisms.

The main aim of the workshop on cyber threat analysis is to bring together leading figures from academia, military, government and industry to assess the state-of-the-art in the area as well as to explore the applications of data mining to address the problem of cyber threat analysis.

Topics of Interest:

- Methods to identify the most critical infrastructures
- Methods to detect cyber terrorist attacks
- Methods to protect against cyber terrorism
- Information assurance
- Intrusion detection and analysis via data mining
- Data mining in forensic
- Credit card fraud analysis
- Economic espionage

Workshop Schedule, December 9, 2002 (updated December 4, 2002)

8:30 – 9:00	*Breakfast*	
9:00 – 9:10	*Vipin Kumar* *Aleksandar Lazarevic*	Welcome & Introductory Notes
9:10 – 10:00	*Robert Grossman*	Keynote talk
10:00 – 10:30	*Mamoru Takahashi* *Toshihiko Kamon*	Cyberspace and the Police
10:30 – 11:00	*Coffee Break*	
11:00 – 11:30	*Jaideep Srivastava*	Data Mining for Information Vigilance
11:30 – 12:00	*Jau-Hwang Wang*	Cyber Forensics – An Introduction
12:00 – 1:30	*Lunch*	
1:30 – 2:00	*Vipin Kumar*	Detecting and Characterizing Novel Network Intrusions using Anomaly Detection
2:00 – 2:30	*Jeff Undercoffer*	Data Mining for Cyber Threat Analysis: What are we digging for and where are we digging?
2:30 – 3:00	*Dave DeBarr*	Using Association Discovery and Cluster Analysis to Summarize Network Security Data
3:00 – 3:30	*Coffee Break*	

3:30– 4:00	*Parthasarathy Srinivasan*	NIC-based Intrusion Detection: A Feasibility Study
4:00 – 4:30	*Kenji Yamanishi*	Detecting Anomalies and Change-points for Cyber Threat Analysis
4:30 – 5:30	*Panel Discussion*	

Presenter: Robert Grossman
Department of Electrical Engineering and Computer Science
Science and Engineering Offices
851 South Morgan Street (M/C 154)
Chicago, Illinois 60607-7053
E-mail: grossman@uic.edu

Keynote Talk

Bio of the presenter: Robert Grossman is the Founder and CEO of the Two Cultures Group. The Two Cultures Group provides consulting and services focused on data which combine the best technology practices with the best business practices.

Grossman is also the Director of the Laboratory for Advanced Computing at the University of Illinois at Chicago (UIC), which he founded in 1989. The Laboratory is an acknowledged leader in data mining, high performance networking, and internet technologies. He led the development of new software tools for data warehousing, distributed computing and high performance networking, introduced standards in data mining, and ran a testbed for a next generation global data network. Grossman currently holds a part time appointment at UIC where he teachers a course in data mining and e-business.

Prior to founding the Two Cultures Group, he founded Magnify, Inc., where he is currently the Chairman. Magnify's software, services and hosted solutions provide real-time, one-to-one personalization so that businesses can increase customer response and loyalty. The privately-held, venture-backed company is headquartered in Chicago.

Grossman is currently the spokesperson for the Data Mining Group (DMG), an industry consortium responsible for the Predictive Model Markup Language (PMML), an XML language for data mining and predictive modeling.

Grossman is on the board of InfoBlox, a company providing network appliances and the scientific advisory board of the Global Information Networking Institute (GINI).

Before starting Magnify in 1994, Grossman led two technology consortia. He was co-founder and co-director of the National Scalable Cluster Project, a consortium of three universities and four industrial partners which pioneered the use of cluster computing. He was also co-founder and co-director of the PASS project, a consortium of two universities and three national laboratories which developed next-generation data warehousing and data mining technology for scientific data.

Grossman is a frequent speaker and often participates on panels at conferences and shows about data mining, personalization, e-commerce, data warehousing, knowledge discovery and web-based computing. He has written over 75 papers and edited four books on these subjects. He earned his A.B. degree in mathematics from Harvard University, Cambridge, Mass., and his doctorate in mathematics from Princeton University, Princeton, N.J.

Presenters:
Mamoru Takahashi
Senior Engineer of Police
Chief of Forensic Center,
High-Tech Crime Technology Division of NPA
phone +81-3-3581-0141 ex 6255
E-mail: Mtakahashi00@npa.go.jp

Toshihiko Kamon
Senior Engineer of Police
Assistant Director of Division,
High-Tech Crime Technology Division of NPA
Phone: +81-3-3581-0141 ex. 6256
Email: Tkamon97@npa.go.jp

Title: **Cyberspace and the Police**

Abstract:
Preface:
To build the safe and stable e-government in the future, Japanese administration are trying to find the way of establishing the high level information security measures. Among these activities, information security measures discussing in the Cabinet Secretariat of Government are giving wide range of beneficial framework involving the partnership between private sectors and governmental agencies so far. On the other hand, the National Police Agency carries counter cyber crime and cyber terrorism in practical; High-level technical support section, which has been set up lately, takes responsibility of dealing with emergency response.

At this workshop, I will introduce main activities for information security posed by the Cabinet Secretariat and National Police Agency, and then I'd like to propose a kind of emerging issues and technologies in the near future.

Note of presentation:
1. Cyber security measures of Japanese administration
   A Security measures for establishing the e-government
   B Information security policy and emergency response
   C Framework for information sharing between private sectors and government agencies.

2. Countermeasures against cyber crime and cyber terrorism posed by National Police Agency.
   A Current Status of Cyber Crime
   B Countermeasures against cyber crime
   C Countermeasures against cyber terrorism
   D Statistics of Cyber Attacks

Bio of the presenter: Mamoru Takahashi joined National Police Agency in 1980, and successively held various posts in information communication bureau and security bureau, which are responsible for countermeasures against cyber crime. National Police Agency is one of leading agencies in Japanese administration and will be playing an important part in the E-government regarding its cyber security. Since 2000, he heads the forensic center where is for supporting cyber crime investigations at the scene and laboratory. He is now the member of Expert Study Team as well, which is for evaluating the security level of every governmental agency, because he has great knowledge and experiences in practical. He's been giving a presentation in many communities and conferences on behalf of the NPA so far.

Presenter: Jaideep Srivastava
CS Department, University of Minnesota
200 Union Street SE, 4-192, EE/CSci Building
Minneapolis, MN 55455
Phone: 612 - 626 - 8107
E-mail: srivasta@cs.umn.edu

Title: **Data Mining for Information Vigilance**

Abstract:
In the present world, we depend critically on our information infrastructure to maintain economic progress. The emerging networked society will increase this dependence, making us increasingly vulnerable to attacks against the infrastructure. The escalating magnitude of this threat is evident from the increasing rate of cyber attacks against our computers in the past few years. According to a recent survey by CERT/CC (Computer Emergency Response Team/ Coordination Center), the rate of cyber attacks has been more than doubling every year in recent times. Thus, the very same information infrastructure that has brought a high degree of agility, has also created a degree of fragility - which if not remedied can cause serious damage to our security and progress. Information assurance is a broad range of techniques whose goal is to ensure that the information infrastructure continues to operate smoothly even in the presence of dire and continuous threats. Broadly speaking, information assurance (IA) is the range of techniques (technologies, policies, and processes), which protect the information infrastructure from cyber threats. Some of its key elements are,

- techniques to insure the integrity, availability, reliability, security and authenticity of information in various databases,
- techniques to detect, prevent, thwart, and recover from network-based attacks from outsiders,
- techniques to monitor, detect, prevent and thwart insider attacks, and
- techniques to maintain operations in the face of partial failures.

In this talk we will discuss various areas of information assurance, and present some ideas on the issues that need to be addressed going forward.

Bio of the presenter: Jaideep Srivastava received his B.Tech. from the Indian Institute of Technology, Kanpur, India, in 1983, and M.S. and Ph.D. from the University of California - Berkeley in 1985 and 1988, respectively. Since 1988 he has been on the faculty of the University of Minnesota, where is a Professor. For over 15 years he has been active as a researcher, educator, and consultant in the areas of databases, data mining, and multimedia. Throughout his career Dr. Srivastava has had an active collaboration with the industry, both for collaborative research and technology transfer. Between 1999 and 2001 Dr. Srivastava was on leave from the University of Minnesota, during which period he has spent time at Amazon.com (www.amazon.com) as the Chief Data Mining Architect, and at Yodlee Inc. (www.yodlee.com) as Director of Data Analytics. Dr. Srivastava is an often-invited participant in technical as well as technology strategy forums. He has given more than a hundred talks in various industry, academic, and government forums. He is on the editorial boards of the IEEE Transactions on

Knowledge & Data Engineering, and the WWW Journal and has been a guest editor for the Data Mining & Knowledge Discovery Journal. He is the program co-chair for PAKDD 2003 and the conference co-chair for the M2003 data mining conferences. The federal government has solicited his opinion on computer science research as an expert witness. He has served in an advisory role to the governments of India and Chile on various software technologies.

Presenter: Jau-Hwang Wang
Department of Information Management, Central Police University
56 Shu-Ren Road, Ta-Kang, Kwei-Shan
Tao-Yuan, Taiwan, ROC 333
Phone: 886-3-3274323
Fax: 886-3-3285189
E-mail: jwang@sun4.cpu.edu.tw

Title: **Cyber Forensics – An Introduction**

Abstract:
Forensics is defined as the application of science to laws enforced by police agencies in a criminal justice system. Similarly, *cyber forensics* can be defined as the application of computer science to laws. Although forensic science has been around for more than a century and comparatively cyber forensics has a brief history, the basic methodologies in determining the evidential value of crime scene and related evidence remain consistent. While traditional forensics professionals use fingerprints, DNA typing, and ballistics analysis to make their case, cyber forensics professionals have to develop new tools for collecting, preservation, examining, extracting, and evaluating digital data in an effort to establish intent, culpability, motive, means, methods, and loss resulting from cyber crime. Traditional evidences, such as fingerprints, are more tangible than digital data. Digital data are relatively *soft* and highly *volatile*. Thus, cyber forensics professionals need to make the digital data harder or at least as hard as the traditional evidences. Furthermore, due to pervasive internet connectivity, the scope of cyber crime incidents are often across national boundaries and not isolated to a single system. Cyber forensics professionals often have to track offenders across the digital matrix. While many of software tools that the system and network administrators use to monitor and test network connectivity can be used to track cyber offenders across the Internet, the techniques of network forensics largely remain to be developed. For example, traditional crime lead discovery are often depended on informants via social networks. However, the dehumanization and depersonalization due to computerization will decrease if not eliminate the functionalities of social informants. Furthermore, as the society has become more and more dependent on computers and computer networks and as the cyber activity become a significant portion of the daily life for general publics, cyber security and cyber forensics will become very important in this information age. New techniques, such as data mining, must be developed to help law enforcement agencies discovering criminal relationships and social networks in the cyber space. This talk will outline the main works in the field of cyber forensics and discuss possible research issues, especially in the area of data mining application.

Bio of the presenter: Jau-Hwang Wang received his bachelor degree from Central Police University, Taiwan, ROC, in 1981, and his MS degree in 1986 from University of Alabama, Birmingham, and his Ph. D. degree in 1992 from University of Minnesota. Since 1992 he has been on the faculty of Central Police University, where he is a professor in the department of information management. He has served as the department head from 1995 to 1998. He is the founder of the master program in the department and is one of the key initiators to establish an information police program for the university. For over 10 years he has been active as a researcher, educator, and consultant in the areas of information systems, data mining,

information retrieval, operating system security, computer crime investigation and computer forensics. He has established and led a data mining and information retrieval research laboratory, where 7 people have received their masters. Most graduates have joined the police task force in the area of police information system and computer crime investigation in Taiwan. Dr. Wang is also the founder of the annual conference on information management and its application in law enforcement and has organized and served on the program committee of a number of workshops and conferences. His current research interests include data mining, information retrieval, computer forensics, computer security and virus detection.

Presenter: Vipin Kumar
CS Department, University of Minnesota
200 Union Street SE, 4-192, EE/CSci Building
Minneapolis, MN 55455
Phone: 612 - 624 - 8023
Fax: 612 - 625 - 0572
E-mail: kumar@cs.umn.edu

Title: **Detecting and Characterizing Novel Network Intrusions using Anomaly Detection**

Abstract: This talk introduces the Minnesota Intrusion Detection System (MINDS), which uses a suite of data mining techniques to automatically detect attacks against computer networks and systems. While the long-term objective of MINDS is to address all aspects of intrusion detection, in this talk we present two specific contributions. First, we show how the behavior-based anomaly detection approach of MINDS is suitable for detecting new and previously unknown types of intrusions, which often indicate emerging threats. Specifically, we present an anomaly detection algorithm that assigns a score to each connection based on its probability of being an intrusion. Experimental results on live network traffic at the University of Minnesota show that our anomaly detection techniques are very promising. In particular, during the past few months our techniques have been successful in automatically detecting several novel intrusions that could not be identified using state-of-the-art signature-based tools such as SNORT. Many of these have been reported on the CERT/CC list of recent advisories and incident notes. Second, we show how using an association pattern analysis, which creates a rule-based model for the novel intrusions detected, can summarize the knowledge in the scored connections. Given the very high volume of connections observed per unit time, such characterization of novel attacks is essential in enabling a security analyst to understand emerging threats. Experimental evaluation shows that the MINDS approach is very useful in creating accurate summaries of novel attacks.

Bio of the presenter: Vipin Kumar received the B.E. degree in electronics & communication engineering from University of Roorkee, India, in 1977; the M.E. degree in electronics engineering from Philips International Institute, Eindhoven, Netherlands, in 1979; and the Ph.D. degree in computer science from University of Maryland, College Park, in 1982. He is currently Director of Army High Performance Computing Research Center and Professor of Computer Science at the University of Minnesota. Kumar's current research interests include parallel computing, parallel algorithms for scientific computing problems, and data mining. His research has resulted in the development of the concept of isoefficiency metric for evaluating the scalability of parallel algorithms, as well as highly efficient parallel algorithms and software for sparse matrix factorization (PSPACES), graph partitioning (METIS, ParMetis, hMetis) and dense hierarchical solvers. He has authored over 100 research articles, and coedited or coauthored 5 books including the widely used text book "Introduction to Parallel Computing" (Publ. Benjamin Cummings/Addison Wesley, 1994) and coedited book “Data Mining for Scientific and Engineering Applications” (Publ. Kluwer Academic Publishers, October 2001). Kumar serves on the editorial boards of IEEE Concurrency, Parallel Computing, the Journal of Parallel and Distributed Computing, and served on the editorial board of IEEE Transactions of Data and Knowledge Engineering during 93-97. He is a Fellow of IEEE, a member of SIAM, and ACM.

Presenter: Jeff Undercoffer
University of Maryland, Baltimore County
Department of Computer Science and Electrical Engineering
1000 Hilltop Circle
Baltimore, MD 21250
Phone: 410-455-3500.
E-mail: undercoffer@umbc.edu

Title: **Data Mining for Cyber Threat Analysis: What are we digging for and where are we digging?**

Abstract:
Current data mining technology provides a broad range of tools that enable knowledge discovery, attribute relationships, pattern evaluation and model creation in large collections of data. These technologies have been applied within the domain of intrusion detection, and most often have used network flows, characterized by TCPdump data, as a data source. Although network flows are an important source of data, they are incomplete by themselves. Reliance upon them as a sole source of data limits the intrusion detection system (IDS) to the syntactic level -- a level where they are operating at the border of ignorance. The overarching consequence of continued operation at this level is that false positive rates will never diminish and true positive rates will fail to improve.

This talk presents a method for an IDS to operate at a semantic level, so that the IDS has an understanding of the elements of an intrusion and what they represent. We define a target centric ontology, where the ontology is representative of the semantics necessary to move beyond the border of ignorance. We detail how the ontology may be completed by data mining techniques in order to identify the relationships that hold between the low level constructs of the ontology. We also suggest how data mining techniques may be used to augment the reasoning process where intrusive events are identified and separated from benign events. In short, we identify where to dig and what to dig for.

Bio of the presenter: Jeffrey L. Undercoffer is a non-traditional PhD student at the University of Maryland, Baltimore County. He is a retired Supervisory Special Agent of the United States Secret Service. Mr. Undercoffer was last assigned to the elite Presidential Protective Division where he was responsible for the integrity, security, and continual operation of the computer systems employed to protect the President of the United States and the White House Complex. Mr. Undercoffer, who holds a M.Sc. in Software Engineering and a B.Sc. in Computer Science, held the position of Solutions Director, World Wide Computer Security Practice, at Unisys Corporation where he was responsible for the design and implementation of security controls for the protection of critical infrastructures (policy, hardware, and software), computer incident response capability, and computer forensics. Mr. Undercoffer has published articles defining security architectures and protocols.

Presenter: Dave DeBarr
MITRE Corporation
7515 Colshire Drive MS W640
McLean, VA 22102-7508
Phone: 703 - 883 - 6231
E-mail: debarr@mitre.org

Title: **Using Association Discovery and Cluster Analysis to Summarize Network Security Data**

Abstract:
Network security analysts often review prioritized event logs from Network-based Intrusion Detection System (NIDS) sensors in an attempt to find interesting events; i.e. events that require further action on their part, such as checking for possible compromise, updating server software, or modifying firewall rules. Unfortunately, the important events can be buried within a much larger group of alerts. For example, one site generates about 1.7 million event records per day. Our discussion will focus on a general approach to event correlation and drill-down capabilities using aggregation, association discovery, and cluster analysis.

Our goal is to help the analyst sift through a large volume of data quickly to find the interesting events. Alerts and other event data can be grouped by address to summarize contiguous periods of activity (using heuristic-based anomaly detection to decide when to aggregate by destination address instead of source address). Frequent event sets can be summarized to provide an overview of common scans, worm activity, and false alarms. Cluster analysis of infrequent event sets can be used to summarize unusual behaviors.

These higher-level summaries allow an analyst to quickly get an overview of what's happening on the network, then drill down on activities of interest. Using five weeks of data from a real-world environment, we were able to provide a high-level overview of 58 million alerts/events with only 126 summary records. Details of the methods used to summarize this data will be provided, including the use of Jaccard coefficients to cluster event labels.

Bio of the presenter: Dave is currently employed as a data mining engineer for the MITRE Corporation, a non-profit organization operating federally-funded research and development centers for the DoD, FAA, and IRS. He is involved in supporting anomaly detection and graph-based data mining efforts for network security and law enforcement application domains. His previous employers include both digiMine and SAIC. His research interests include large-scale cluster analysis and adaptive modeling.

Presenter: Parthasarathy Srinivasan
Dept. of Computer and Information Science, Ohio State University
395 Dreese Lab
2015 Neil Ave
Columbus, OH-43210, USA
Office: 693 Dreese Lab
Phone: 614 - 292 - 2568
Fax: 614 - 292 - 2911
Email: srini@cis.ohio-state.edu

Title: **NIC-based Intrusion Detection: A Feasibility Study**

Abstract:
We present and evaluate a NIC-based network intrusion detection system. Functions such as signature-based and anomaly-based packet classification are performed on the NIC, which has its own processor and memory. This makes the system virtually impossible to bypass or tamper with as can be the case with software-based systems that rely on the host operating system to function. We empirically evaluate such systems from the perspective of quality and performance (bandwidth of acceptable messages) under varying conditions of host load. The preliminary results we obtain are very encouraging and lead us to believe that such NIC-based security schemes could very well be a crucial part of next generation network security systems.

Bio of the presenter: Srinivasan Parthasarathy is currently an Assistant Professor in Computer and Information Sciences at the Ohio State University. He was born on March 22 1970 in Cleveland, Ohio. He holds a B.E. in Electrical Engineering (1992) from the University of Roorkee, India, an M.S. in Electrical and Computer Engineering (1994) from the University of Cincinnati, and an M.S. (1996) and PhD (2000) in Computer Science from the University of Rochester. His research interests are in data mining and parallel and distributed systems and he has published widely in these areas. He has consulted for Intel Corporation and held visiting positions at the Rochester Institute of Technology and at the University of Rochester. He has served on the program committees of several international conferences and workshops and is the co-chair of the High Performance Data Mining Workshop the past two years. He is a recipient of an Ameritech Faculty Fellowship for the year 2001-2002.

Presenter: Kenji Yamanishi
Internet Systems Research Laboratories,
NEC Corporation
4-1-1 Miyazaki, MIyamae-ku, Kawasaki Kanagawa, 216-8555,JAPAN
E-mail: k-yamanishi@cw.jp.nec.com

Title: Detecting Anomalies and Change-points for Cyber Threat Analysis

Abstract:

In this talk we introduce new methods for detecting statistical anomalies and change points with their applications to intrusion detection. We are concerned with the following three issues:

1) how to detect intrusions of unknown patterns, which any signature-based method doesn't detect,
2) how to detect intrusions adaptively even when their attack pattern may change over time,
3) how to detect the emergence of concentrated attacks in a data stream.

We show a statistical/machine-learning based approach to these issues and its empirical demonstration.

Bio of the presenter: Kenji Yamanishi is a research fellow at Internet Systems Research Laboratories, NEC Corporation. He has been working in the field of computational learning theory, machine learning, and information theory. His recent interests include data mining(anomaly detection, event detection) with applications to cyber threat analysis and text mining with applications to CRM. He received BS, MS, and PhD from University of Tokyo. He received the best paper award from IEICEB!J The Institute of Electronics, Information and Communication Engineers B!K in 1990.He is a chair of Information-Based Induction Sciences Working Group in IEICE and an editorial member of IEICE journal, etc.