Data Mining for Malicious Code Detection and Security Applications

Dr. Bhavani Thuraisingham

Louis A. Beecherl, Jr. Distinguished Professor
Director of the Cyber Security Research Center
Department of Computer Science
Eric Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
Richardson, Texas
http://www.utdallas.edu/~bxt043000/

ABSTRACT

Data mining is the process of posing queries and extracting patterns, often previously unknown from large quantities of data using pattern matching or other reasoning techniques. Data mining has many applications in security including for national security as well as for cyber security. The threats to national security include attacking buildings, destroying critical infrastructures such as power grids and telecommunication systems. Data mining techniques are being investigated to find out who the suspicious people are and who is capable of carrying out terrorist activities. Cyber security is involved with protecting the computer and network systems against corruption due to Trojan horses, worms and viruses. Data mining is also being applied to provide solutions such as intrusion detection and auditing.

The first part of the presentation will discuss my joint research with Prof. Latifur Khan and our students at the University of Texas at Dallas on data mining for cyber security applications. For example, anomaly detection techniques could be used to detect unusual patterns and behaviors. Link analysis may be used to trace the viruses to the perpetrators. Classification may be used to group various cyber attacks and then use the profiles to detect an attack when it occurs. Prediction may be used to determine potential future attacks depending in a way on information learned about terrorists through email and phone conversations. Data mining is also being applied for intrusion detection and auditing. Other applications include data mining for malicious code detection such as worm detection and managing firewall policies.

This second part of the presentation will discuss the various types of threats to national security and describe data mining techniques for handling such threats. Threats include non real-time threats and real-time threats. We need to understand the types of threats and also gather good data to carry out mining and obtain useful results. The challenge is to reduce false positives and false negatives.

The third part of the presentation will discuss some of the research challenges. We need some form of real-time data mining, that is, the results have to be generated in real-time, we also need to build models in real-time for real-time intrusion detection. Data mining is also being applied for credit card fraud detection and biometrics-related applications. While some progress has been made on topics such as stream data mining, there is still a lot of work to be done here. Another challenge is to mine multimedia data including surveillance video. Finally, we need to maintain the privacy of individuals. Much research has been carried out on privacy-preserving data mining.

In summary, the presentation will provide an overview of data mining, the various types of threats and then discuss the applications of data mining for malicious code detection and cyber security. Then we will discuss the consequences to privacy.

Dr. Bhavani Thuraisingham is the Louis A. Beecherl, Jr. I Distinguished Professor in the Erik Jonsson School of Engineering and Computer Science at the University of Texas at Dallas (UTD) effective September 2010. She joined UTD in October 2004 as a Professor of Computer Science and Director of the Cyber Security Research Center which conducts research in data security and privacy, secure networks,

secure languages, secure social media, data mining and semantic web. She is an elected Fellow of three prestigious organizations: the IEEE (Institute for Electrical and Electronics Engineers), the AAAS (American Association for the Advancement of Science) and the BCS (British Computer Society). She is the recipient of numerous awards including the IEEE Computer Society's 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management" and the 2010 Research Leadership Award for Outstanding and Sustained Leadership Contributions to the field of Intelligence and Security Informatics" presented jointly by the IEEE Intelligent and Transportation Systems Society Technical Committee on Intelligence and Security Informatics in Transportation Systems and the IEEE Systems, Man and Cybernetics Society Technical Committee on Homeland Security. She served as served as an IEEE Distinguished Lecturer between 2002 and 2005. She was also quoted by *Silicon India* magazine as one of the seven leading technology innovators of South Asian origin in the USA in 2002.

Prior to joining UTD, Dr. Thuraisingham was an IPA (Intergovernmental Personnel Act) at the National Science Foundation (NSF) in Arlington, VA, from the MITRE Corporation for three years. At NSF she established the Data and Applications Security Program and co-founded the Cyber Trust theme and was involved in interagency activities in data mining for counter-terrorism. She worked at MITRE in Bedford, MA between January 1989 and September 2001, first in the Information Security Center and later as a department head in Data and Information Management as well as Chief Scientist in Data Management in the Intelligence and Air Force centers. At MITRE she led team research and development efforts on secure data management and real-time data management for NSA, AFRL, SPAWAR, CECOM and CIA. She also served as a technical consultant in information security and data management to the Department of Defense, the Department of Treasury and the Intelligence Community for over 10 years and served as an expert consultant to the Department of Justice in 2001. Thuraisingham's industry experience includes six years of research and product development as well as technology transfer at Control Data Corp. and Honeywell Inc. in Minneapolis, MN. While in industry and at MITRE, she was an adjunct professor of computer science and member of the graduate faculty first at the University of Minnesota and later at Boston University between 1984 and 2001. She also worked as visiting professor soon after her PhD, first at the New Mexico Institute of Technology and later at the University of Minnesota between 1980 and 1983.

During her six years at UTD, Dr. Thuraisingham has established and leads a strong research program in Intelligence and Security Informatics which now includes 5 core professors and the team has generated over $12 million in research funding from agencies such as NSF, AFOSR, IARPA, NGA, NASA, ONR, ARO and NIH as well as corporations such as Raytheon Inc. The research projects include an NSF Career Grant, an AFOSR Young Investigator Program Award and a DoD MURI Award on Assured Information Sharing. Her current focus includes three activities: (i) studying how terrorists and hackers function so that effective and improved solutions can be provided (ii) initiating interdisciplinary programs integrating social sciences and information sciences and (iii) transferring the technologies developed at the university to commercial development efforts. She is also instrumental in establishing UTD's MS Track in Information Assurance and is a Co-PI of the $1.7 million NSF Scholarship for Service Award in Cyber Security. She teaches courses in data and applications security, trustworthy semantic services and digital forensics and collaborates with the DFW corporations as well as North Texas Regional Computer Forensics Laboratory for student projects. She also writes motivational articles including one on CS Careers in the Global Economy.

Dr. Thuraisingham's research interests are in data security and data mining for counter-terrorism. Her work in information security and information management has resulted in over 100 journal articles, over 200 refereed conference papers and workshops, three US patents and several IP disclosures. She is the author of ten books in data management, data mining and data security including one on data mining for counter-terrorism and another on Database and Applications Security and is completing her eleventh book on Data Mining Tools for Malware Detection, and is the editor of twelve books She has given over 70

keynote presentations at various technical conferences and has also given invited talks at the White House Office of Science and Technology Policy and at the United Nations on Data Mining for counter-terrorism. She serves (or has served) on editorial boards of leading research and industry journals including several IEEE and ACM Transactions, the VLDB Journal, and also served as the Editor in Chief of Computer Standards and Interfaces Journal. She has contributed to multiple standards activities including Navy's Next Generation Interface efforts, Object Management Group's Real-time computing and C4I efforts, and more recently the Open Geospatial Consortium's semantic web efforts. In addition, she has been an instructor at AFCEA's (Armed Forces Communications and Electronics Association) Professional Development Center since 1998 and has served on panels for the Air Force Scientific Advisory Board and the National Academy of Sciences including one on protecting children from inappropriate content on the Internet chaired by Hon. Dick Thornburgh in 2000. She is continuing with these efforts and recently participated in EastWest Institute's 1st Worldwide Security Summit panel on protecting our children in cyberspace. She is a member of several professional organizations including the Association for Computing Machinery, IFIP 11.3 Working Group in Data and Applications Security and AFCEA. She has chaired over ten conferences and has served in over 100 conferences program committees.

Dr. Thuraisingham received her BS degree in Mathematics and Physics with first class at the University of Ceylon, her M.Sc degree in Mathematical Logic at the University of Bristol, UK and her PhD degree in Theory of Computation at the University of Wales, UK. She strongly believes in continuing education and has also received a number of professional qualifications to enhance her 30 year career since 1980 including an MS in Computer Science focusing in computer systems and networks at the University of Minnesota, Java Development Certification from Learning Tree International, the Certificate in Terrorism Studies at St. Andrews University, Scotland and CISSP (Certified Information Systems Security Professional) certification with ISC2.

Dr. Thuraisingham promotes Math and Science to high school students as well as to women and underrepresented minorities, and is a member of the Society of Women Engineers (SWE). She has given featured addresses at conferences sponsored by WITI (Women in Technology International) and SWE and received the 2001 Woman of Color Research Leadership Award from Career Communications Inc. Articles on her efforts, her vision as well as her team's research have appeared in multiple magazines including the *Dallas Morning News*, the Boston Globe, ABC News, *D Magazine, MITRE Matters* the *DFW Metroplex Technology* magazine. She has also appeared in DFW Television giving her opinions on cyber security.

Dr. Thuraisingham is the founding president of "Bhavani Security Consulting, LLC" a company providing services in consulting and training in Cyber Security and Information Technology. She is also the founder and a member of the board of directors of "Infosec Analytics, LLC", a spin-off company from UTD developing tools in malware detection and information sharing.