IEEE ISI 2008 Invited Talk (I)

Data Mining for Security Applications: Mining Concept-Drifting Data Streams to Detect Peer to Peer Botnet Traffic

Dr. Bhavani Thuraisingham

The University of Texas at Dallas

Abstract

There has been much interest on using data mining for counter-terrorism and cyber security applications. For example, data mining can be used to detect unusual patterns, terrorist activities and fraudulent behavior. In addition data mining can also be sued for intrusion detection and malicious code detection. Our current research is focusing extensively on data mining for security applications in general and data mining for botnet detection in particular. Our presentation will addressed both aspects of data mining applications.

The term bot comes from the word robot. A bot is usually referred to automated software capable of performing certain functions. A botnet is a network of bots that are used by a human operator or botmaster to carry out malicious actions. Botnet is one of the most extreme tools used in cyber-crime these days including DDoS attacks, phishing, spamming, and spying on remote computers. Often businesses, governments, and individuals are facing multi-million-dollar damages caused by hackers with the help of these botnets. It is a major challenge to the cyber-security research community to combat this threat.

Botnets have different topologies and protocols. The most prevalent botnets are Internet Relay Chat (IRC)-based, having a centralized architecture. There are many approaches available to detect and take down IRC botnets. On the other hand, Peer to Peer (P2P) is a relatively newer technology used in botnets. P2P botnets use P2P protocols to communicate among the bots and the botmaster. These botnets are distributed, having no central point of failure. Besides, they are relatively smaller than their IRC counterparts. As a result, these botnets are more difficult to detect and destroy than the IRC botnets. Moreover, most of the current research related to P2P botnets are in the analysis phase. The main goal of our project is to devise an efficient technique to detect P2P botnets. We approach this problem from a data mining perspective. We are developing techniques to mine network traffic for detecting P2P botnet traffic.

The presentation will first provide an overview for data mining for security applications and then discuss our research to the botnet problem which follows from an important observation that network traffic (as well as botnet traffic) is a continuous flow of data stream. Conventional data mining techniques are not directly applicable to stream data because of two vital problems associated with them: potentially infinite in length, and concept drift. We propose a technique that can efficiently handle both problems. Our main focus is to adapt three major data mining techniques: classification, clustering, and outlier detection to handle stream data. Our preliminary study on the development of new stream classification techniques for P2P bothnet detection has generated encouraging results. In addition to botnet detection, we will also discuss our research on data mining for malicious code detection and intrusion detection.

ACKNOWLEDGEMENT: This research is a collaborative effort between the University of Texas at Dallas and the University of Illinois at Urbana Champagne on data mining for botnet detection. I thank PhD student Mehedy Masud and Associate Professor Latifur Khan from the University of Texas at Dallas and Professor Jiawei Han from the University of Illinois Urbana Champagne for the collaboration.

Biography: Dr. Bhavani Thuraisingham joined The University of Texas at Dallas (UTD) in October 2004 as a Professor of Computer Science and Director of the Cyber Security Research Center in the Erik Jonsson School of Engineering and Computer Science She is an elected Fellow of three professional organizations: the IEEE (Institute for Electrical and Electronics Engineers), the AAAS (American Association for the Advancement of Science) and the BCS (British Computer Society) for her work in data security. She received the IEEE Computer Society's prestigious 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management." She was quoted by Silicon India Magazine as one of the top seven 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. At NSF she established the Data and Applications Security Program and co-founded the Cyber Trust theme and was involved in inter-agency 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 was later a department head in Data and Information Management as well as Chief Scientist in Data Management in the Intelligence and Air Force centers. She has served as an expert consultant in information security and data management to the Department of Defense, the Department of Treasury and the Intelligence Community for over 10 years. Thuraisingham's industry experience includes six years of research and development at Control Data Corp. and Honeywell Inc. in Mpls, MN. While she was in Industry and 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.

Dr. Thuraisingham's work in information security and information management has resulted in over 70 journal articles, over 200 refereed conference papers and workshops, and three US patents. She is the author of seven 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 eighth book on Trustworthy Semantic Web. She has given over 30 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 and currently serves as the Editor in Chief of Computer Standards and Interfaces Journal. She is also 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.

Dr. Thuraisingham promotes Math and Science to high school students as well as to women and underrepresented minorities and has given featured addresses at conferences sponsored by WITI and SWE. Articles on her efforts as well as her vision have appeared in multiple magazines including the Dallas Morning News, The D Magazine, The MITRE Matters and the DFW Metroplex Technology Magazine. She enjoys advising and motivating her several research students pursuing MS and PhD degrees in data mining and data security at UTD and mentors assistant and associate professors related to her field at the university.

Dr. Thuraisingham was educated in the United Kingdom both at the University of Bristol and at the University of Wales.