Presented at the Simposium Internacional de Tecnologias de Informacion 2004, Campus Queretaro, 26-28 August 2004.

Machine Learning Methods for Intrusion Detection

Prof. Rao Vemuri
rvemuri@ucdavis.edu

in collaboration with

Yihua Liao, Wenjie Hu and Alexandro Pasos
Universityof California, Davis
USA

Abstract

Intrusion detection has always played an important role in computer security research. Two general approaches to intrusion detection are currently popular: misuse detection and anomaly detection. In misuse detection – basically a pattern matching method – one gathers signature patterns of known intrusive attacks by, say, inspecting the audit records on a victim machine. Signature patterns generated by user activities are then compared with the stored patterns of, say, authorized users. Those matched are then labeled as intrusive activities. That is, misuse detection is essentially a model-reference procedure. While misuse detection can be effective in recognizing known intrusion types, it tends to give less than satisfactory results in detecting novel signature patterns.

Anomaly detection, on the other hand, looks for patterns in signatures that deviate from the normal. Although a number of methods have been used with varying degrees of success, this method suffers from the basic difficulty in defining what is "normal." Methods based on anomaly detection tend to produce many false alarms because they are not capable of discriminating between abnormal patterns triggered by an otherwise authorized user and those triggered by an intruder.

Regardless of the approach used, almost all intrusion detection methods rely on some sort of signature data – tracks of activity left behind by users. People trying to outsmart an intrusion detection system can deliberately cover their tracks by consciously changing their behavior patterns. Some examples of obvious features that a user can manipulate are the identity of the host machine from where an attack originated, time of log-in and the command set used. This, coupled with factors emanating from privacy issues, makes the monitoring of user activities a less attractive option.

An alternative to learning *user* behavior and building *user profiles* is to model results that are one or more steps removed from user activity. Looking for statistical deviations in the self-similar nature of network traffic is one possibility. Although no one seems to have attempted this method so far, a reference to this possibility can be found in the literature.

Learning *program* behavior and building *program profiles* is another possibility. Indeed building program profiles, especially those of privileged programs, has become a popular alternative to building user profiles. Capturing the system call history associated with the execution of a program is one way of creating the execution profile of a program. Program profiles appear to have the potential to provide concise and stable descriptions of intrusion activity. Furthermore, they are less prone to subjectivity than user behavior profiles because it would be difficult for attackers to cover their signature tracks left in system call history. To date, almost all the research in this area was focused on using short sequences of system calls generated by individual programs. The local ordering of these system call sequences was then examined and classified as normal and intrusive. There is one theoretical and one practical problem with this approach. Theoretically, no justification was provided for this definition of "normal" behavior. Notwithstanding this theoretical gap, this procedure is tedious and costly because it is difficult and time consuming to build and maintain profiles to all the programs (i. e. system programs and application programs). Although the system programs are not generally updated as often as the application programs, the execution traces of system programs are likely to be dynamic also, thus making it difficult to characterize "normality."

This paper treats the system calls differently. Instead of looking at the local ordering of the system calls, this method uses the frequencies of system calls to characterize program behavior. This stratagem allows the treatment of long stretches of system calls as one unit thus allowing one to bypass the need to build and maintain separate databases for each program within a process. Using the text processing metaphor, each system call is then treated as a "word" in a long document and the set of system calls generated by a process is treated as the "document." This analogy makes it possible to bring the full spectrum of well-developed text processing methods to bear on the intrusion detection problem.

We at the University of California, Davis have been trying a variety of machine learning methods to address these issues, including the k-nearest neighbor classification method, support vector machines and fuzzy connectionist methods. Highlights of results from these efforts will be presented at the conference.

Acknowledgement

Work reported here is support in part by a grant from AFOSR, USA. The work of Alexandro Pasos was supported by CONASYT, Mexico.

References.

Copies of many of these papers are available at http://www.cs.ucdavis.edu/~vemuri/publications

Khaled Labib and V. Rao Vemuri, "Detecting and Visualizing Denial of Service And Network Probe Attacks Using Principal Component Analysis," SAR'04 the *3rd Conference on Security and Network Architectures*, La Londe, Cote d'Azur (France), June 21-25, 2004

Rawat, Sanjay, Arun K. Pujari, V. P. Gulati, V. Rao Vemuri, "Intrusion Detection using Text Processing Techniques with a Binary-Weighted Cosine Metric," *International Journal of Information Security*, Springer-Verlag, Submitted 2004.

Liao, Yihua and V. Rao Vemuri and Alejandro Pasos, "A General Framework for Adaptive Anomaly Detection with Evolving Connectionist Systems", *SIAM International Conference on Data Mining*, Lake Buena Vista, FL, April 22-24, 2004.

Wenjie Hu, Yihua Liao, and V. Rao Vemuri, Robust Anamoly Detection using Support Vector Machines, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Submitted**.**

Wenjie Hu, Yihua Liao, and V. Rao Vemuri, Robust Support Vector Machines for Anamoly Detection in Computer Security, *International Conference on Machine Learning*, Los Angeles, CA, July 2003**.**

Khaled Labib and V. Rao Vemuri, “NSOM: A Real-time Network-Based Intrusion Detection System Using Self-Organizing Maps," *Networks and Security,* Submitted*,* 2002.

Yihua Liao and V. Rao Vemuri, “Using Text Categorization Techniques for Intrusion Detection, *Usenix: Security 2002,* San Francisco, Aug. 2002.

Yihua Liao and V. Rao Vemuri, “Use of K-nearest Neighbor Classifier for Intrusion Detection,, *Networks and Security,* Vol. 21, no. 5, pp 438-448, 2002.

Dao, Vu and V. Rao Vemuri, “Computer Network Intrusion Detection: A Comparison of Neural Networks Methods, *Differential Equations and Dynamical Systems* (Special Issue on Neural Networks), Published, 2002.

Dao, Vu and V. Vemuri, “Profiling Users in the UNIX OS Environment, *International ICSC Conference on Intelligent Systems and Applications*, University of Wollongong, Australia, Dec. 11-15, 2000.