Integration Misuse and Anomaly Detection Techniques on Distributed Sensors

Shih-Yi Tu† Chung-Huang Yang‡ Kouichi Sakurai††

†Graduate Institute of Information and Computer Education,
National Kaohsiung Normal University, 116, Ho Ping First Road, Kaohsiung 802, Taiwan
tshihyi@gmail.com
‡Graduate Institute of Information and Computer Education,
National Kaohsiung Normal University, 116, Ho Ping First Road, Kaohsiung 802, Taiwan
chyang@computer.org
††Faculty of Engineering, Department of Materials Science and Engineering,
Kyushu University, 744 Motooka, Nishi-ku, Fukuoka-city, Fukuoka 819-0395
sakurai@csce.kyushu-u.ac.jp

Abstract Network Instruction Detection System (NIDS) tools serve as the first line of defense for network hacking, but most of the open source Network Instruction Detection System (NIDS) tools only provide misuse detection for well-known attacks. Most IDS systems use "Rule" or "Signature" to look for the known malicious that is misuse detection method. Nowadays popular software is Snort (Network type) or OSSEC (Host type). However, only use misuse detection technology that will not look up the unknown malicious and prevent the system. Existing IDS technology of anomaly detection include statistical based, knowledge based and machine learning based detection techniques. Machine learning based detection technology is applied to detect unknown malicious. The methods of machine learning include Bayesian Networks, Markov Models, Fuzzy Logic techniques, Clustering and Outlier detection, and Genetic Algorithm.

This research develops a method that allows both misuse and anomaly instruction detection on the distributed sensors. We apply Bayesian network to analyze each sensor's network packets and integrity information for misuse detection. We first establish a distributed system as Prelude IDS to integrate Snort and OSSEC for detection known malicious attacks. In addition, we use NTOP and CACTI collecting network and each sensor's host resources anomaly information. Next, we analyze sensor's network packets and anomaly behavior using Bayesian network with WEKA to identify the interdependencies between the attributes and then calculate their probabilities. These calculated probabilities are then used to distinguish unknown attack from the new attacks and to improve the instruction detection performance.

1 Introduction

In recent years, malware has become a serious threat to information security. Malware includes Bots, Virus, Worm, Trojan Horses, Spyware, and Adware. Most of the malware could destroy the computer operations and steal secret data to profit illegally. Although installing antivirus software could protect users to some degree, but it is not possible to guarantee users from being attacked. A built-in IDS will help systems against external attacks.

The scheme proposed in this paper combines OSSEC (host type) [11] and Snort (network type) [14] with the hybrid Prelude IDS to become a hybrid IDS) [10, 18]. Through this hybrid IDS data can be collected with increased quantity and detail, and these data can be used to increase the accuracy of attack detection and to reduce faulty detection rate.

In order to improve the IDS defection performance, many researchers applies machine learning in detecting unknown malicious attacks [1, 2, 7, 9, 16]. Machine learning has the capability of detecting "Anomaly" behaviors when facing unknown attacks.

2 Current Detection Technologies

There are currently several lines of research in misuse and anomaly detection, and they will be briefly discussed below.

2.1 Intrusion Detection System

Intrusion Detection System (IDS) is a hardware- or software-based detection tool [19] and it depends primarily on monitoring network traffic or system logs for detecting suspicious activities. According to the detection range, IDS can be classified into NIDS and HIDS.

NIDS is responsibly mainly on real-time scanning all original network packet headers and the command syntax they use [14]. The network packets are analyzed through comparing the contents of the packets with the characteristics of known malicious attackers, and from such analyses NIDS determines whether it may be under attack and an alert should be sent. NIDS may initiate a counter measure while detecting malicious attacks, and it may be deployed on large-scale detection at a low cost.

HIDS, on the other hand, install agents inside of the operating systems (OS) to be protected [11]; and these agents integrate closely with the OS services to monitor system events. HIDS can detect intrusions that are missed by NIDS, for example, hacker behaviors on the victim host or planting "backdoor" on the systems. Without additional hardware, HIDS can remedy the events missed by NIDS. However, because the host-based IDS must constantly monitor the system, it consumes quite a bit system resource and may affect the performance of the console itself.

A hybrid IDS combines the advantages of NIDS and HIDS, and this paper proposes a hybrid IDS to detect known malicious activities. A hybrid IDS (e.g., Prelude IDS) operates both network-based (e.g., Snort) and host-based (e.g., OSSEC) IDS to offer comprehensive detection of all types of security threat events.

Prelude IDS uses a distributed architecture [5, 12, 20] and it collects and integrates data from distributed sensors. For reporting events Prelude IDS accepts Intrusion Detection Message Exchange Format (IDMEF) as the common languages [20], and it allows users to mix-and-match the deployment of these systems according to their strengths and weaknesses to obtain the optimal implementation.

2.2 Bayesian Network

Bayesian Network is often used in IDS for anomaly detection [1, 7, 9, 16], and it can be used to show the probabilistic relationships among variables and their conditional independencies. Currently there are many researchers applying Bayesian Network to anomaly detections on malware.

Using Bayesian Network a mechanism can be built to predict anomaly behaviors from common normal behaviors; in this paper, we collected the necessary information from network and hosts as variables which are evidences of the attack. Through built Bayesian network shows the probabilistic relationships among those variables. It means that Bayesian network can based on previous learning, estimate the probability of occurrences for all conditions. When a new type of attack does occur, it will be analyzed and "learned"; and the new probability structure will be calculated for predicting the unknown (new) attack.

3 The proposed misuse and anomaly detection approach

Bayesian Network is one of the methods used in machine learning. The relationships of various variables will display as a graphical model. Many researches used Bayesian Network to detect [1, 2, 9, 16] for new release or unknown malicious attacks. Therefore, we chose Bayesian networks to learn with Prelude IDS alerts events as knowledge model.

In the proposed approach of Bayesian networks that have to lean from various variables (evidences). The implemented system stores the hosts various resources information and network packet flow as the evidences when the attack happened.

According to [1], [2], [8] and [9] when an attack occurs the resources of the host resources and network flow will produce an obvious change in the moment. The most special cases are worm attack. In addition, according to collected the malicious attack information and malware by our system. The statistic of historical database (2009/03/15~2009/7/31), we found the most malicious threat is warm (45.63%) as shown in Fig. 1. Therefore we collect network packets flow and host resources information as the related variables for building Bayesian Networks. The network information is collected by NTOP [10]. NTOP is a kind of network flow monitoring tool. The host resources information is collected by CACTI [4]. The built Bayesian Networks shall be tested with warm attacks that collected in this study.

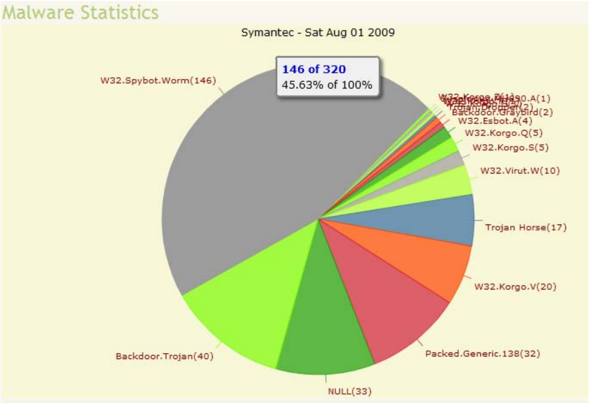

Fig.1 Malware statistics in this study

3.1 System Design

For the purpose of reducing cost we apply open source tools to implement our system. The proposed system architecture is shown in Fig. 2. If Prelude IDS detect the suspicious network or host activities, the system will get the related information from database and calculate the probabilistic with WEKA [3]. WEKA will produce an XML file and deliver it to JavaBayes [6]. The calculated probabilistic will be as the threshold value. The threshold will be translated into *security level* with Prelude IDS. Security Level is divided into four layers including *Info*, *Low*, *Medium* and *High*. If security level is *medium* the system will send a security level alert to inform the system administrator.

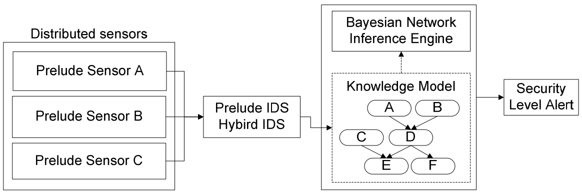

Fig. 2 System architecture

In this study, three mainly components of the system describe as following:

3.1.1 *Preldue IDS*

Prelude IDS is hybrid IDS including Snort and OSEEC. Snort is used to monitor the network activity and OSSEC is used to monitor the distributed host resources information. Therefore, we can get the network and host information at the same time.

3.1.2 *Knowledge Model*

This knowledge model is based on Bayesian Network which is learned with WEKA. At the first we collected the related data are network packet information (such as *IP/TCP/UDP/ICMP*) with NTOP and host resources information (such as *CPU usage*, *memory usage*, *network traffic*, *user login*) with CACTI. Bayesian Network had to learn with those variables.

Fig. 3 is a graphic of a part of knowledge model. It shows us the relationships between the various parameters of host information in normal situation. Each node means a parameter. The host related information includes *user usage time, users, average loading* and *memory usage rate* (*memory free* and *memory swap*). According to the training result influence the *RunningProcess* factors are *TimeInterval, users, MemoryFree, MemorySWAP,* and *AeverageLoading.* Therefore we can get information if the attack occurs and then the host will exert an influence on *runningprocess*.

Bayesian Network learned with standard version of WEKA. In order to evaluate the structure of Bayesian Network, we chose K2 search algorithm for evaluation on training dataset [3]. Search algorithm usually used to evaluate the structure of Bayesian Network. The data sources of knowledge model are from historical database. Using WEKA could help us to train the better knowledge model.

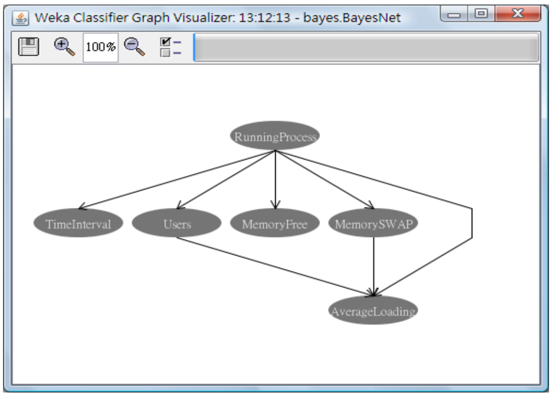

Fig. 3 A part of knowledge model

3.1.3 *Bayesian Network Inference Engine*

In this study proposed an inference mechanism to detect unknown or new attack with JavaBayes. JavaBayes is an API could be combined with WEKA. JavaBayes provides inference function for prediction the unknown attack. When JavaBayes receives the learned result of Bayesian Network, it will to produce an inference result. This result could be translated into security level. If security level is *medium*, the system will send an alert to administrator.

4 Conclusion and future work

Integrate misuse and anomaly detection technologies is the mainly future tendency [5, 13]. Many researches used Bayesian Network for detecting unknown attacks with NIDS [1] or HIDS [9]. Those researches are valid for detecting the malicious activities. The detection rates are almost 100% accuracy for unknown attacks [1, 9]. Therefore we proposed an approach to combine NIDS and HIDS with Bayesian Network.

In order to reduce the building costs we applied various open source tools to implement a hybrid IDS as misuse detection technology and Bayesian Network as anomaly technology. Though the proposed system to detect the known and unknown suspicious malicious attacks.

Table 1 is a properties comparison table with other researches. We combined with those researches advantages and proposed an integrated approach to improve the performance of hybrid IDS.

Table 1 Properties comparison

	[1]	[8]	[14]	[15]	This study
Host detected	No	Yes	No	No	Yes
Network detected	Yes	No	Yes	Yes	Yes
Anomaly detection	Yes	Yes	Yes	No	Yes
Misuse detection	Yes	Yes	Yes	Yes	Yes
Distributed	Yes	No	No	Yes	Yes
Open Source tool	Yes	Yes	Yes	Yes	Yes

Current work is focus on the test phase for implemented Bayesian Networks as knowledge model for predicting unknown attack. We will also test the detection performance with the malware are collected by our system (special the warms) to improve the IDS detection rate for known and unknown malicious attacks. In the future, we can integrate *Honeypot* to Prelude IDS for collecting more related information of attack. That will help us to realize the attack behaviors for further prevention system.

Reference

[1] P. G. Bringas and Y. K. Penya, “Next-generation Misuse and Anomaly Prevention System,” *Lecture Notes in Business Information Processing*, in press.

[2] J. Bigham, X. Jin, D. Gamez and C. Phillips, “Hybird workflow and Bayesian Networks to Correlate Information in the Protection of Large Scale Critical

Infrastructures," *Electronic Notes in Theoretical Computer Science*, 2005.

[3] R. R. Bouckaert, Bayesian Network Classifiers in Weka, September, 2004.

[4] CACTI, http://www.cacti.net/

[5] P. Gracia-Teodora, J. Diaz-Verdejo, G. Macia-Fernandez and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers and Security*, August 2008.

[6] JavaBayes, http://www.cs.cmu.edu/~javabayes/

[7] C. Kruegel, D. Mutz, W. Robertson and F. Valeur, Bayesian Event Classification For Intrusion Detection, Computer Security Applications Conference, pp. 14-23, 2003.

[8] A. Lakhina, M. Crovell and C. Diot, Characterization of Network-Wide Anomalies in Traffic Flows, Proceedings of the ACM SIGCOMM Internet Measurement Conference, Oct 2004.

[9] R. Moskovitch,S. Pluderman, I. Gus, D. Stopel, C. Feher, Y. Parmet, Y. Shahar and Y. Elovici, Host Based Intrusion Detection using Machine Learning, Intelligence and Security Informatics 2007 IEEE, May 2007.

[10] NTOP, http://www.ntop.org/

[11] OSSEC, http://www.ossec.net/

[12] Prelude IDS, http://www.prelude-ids.com/

[13] A. Patcha and J. M. Park, "An Overview of Anomaly Detection Techniques: Existing Solution and Latest Technological Trends," *Computer Networks*, 2007.

[14] Snort, http://www.snort.org/

[15] J. Timofte, "Intrusion Detection using Open Source Tools," *Revista Informatica Economică*, 2008.

[16] W. Tylman, Misuse-Based Intrusion Detection Using Bayesian Networks, International Conference on Dependability of Computer Systems, 2008.

[17] B. Weiland, Intrusion Detection with Heterogenous Sensors, July 2007.

[18] Wikipedia, Bayesian Network, http://en.wikipedia.org/wiki/Bayesian_network

[19] Wikipedia, Intrusion detection system, http://en.wikipedia.org/wiki/Intrusion_detection_system

[20] C. Yasm, "Prelude as a Hybrid IDS Framework," *SANS Institute*, March 2009.