International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Semantic Host Based Intrusion Detection

Dipali Suhalal Patil†* and Atul Dusane†

†Computer Engineering, SSBT's COET, Bambhori, North Maharashtra University, Jalgaon, Maharashtra, India

Accepted 20 Jan 2015, Available online 01 Feb 2015, **Vol.5, No.1 (Feb 2015)**

Abstract

Today it is very important to provide a high level security to protect highly sensitive and private information. Intrusion detection system is an essential technology in network security. Host-based misuse intrusion detection system design is very challenging due to the high false alarm rate. This system introduces a new host-based anomaly intrusion detection methodology using discontiguous system call patterns, in an attempt to increase detection rates whilst reducing false alarm rates by combining misuse and anomaly based detection techniques. The key concept is to apply a semantic structure to kernel level system calls in order to reflect intrinsic activities hidden in high-level programming languages, which can help understand program anomaly behavior. The semantic method possesses an inherent resilience to mimicry attacks, and posses a high level of portability between different operating system versions.

Keywords: *Misuse detection, anomaly detection, system call pattern, host based system.*

1. Introduction

Information security is one of the major issues concerned by computer professions in recent years. While human daily life is more and more dependent on computers, the number of cyber crimes, as well as the impact caused by the cyber crime is growing in incredible rates.

This shows the importance to protect our information assets from attacks and damages. To solve security issues along with traditional security majors Intrusion Detection Systems are invented.

Intrusion detection is the process of identifying and responding to suspicious activities targeted at computing and communication resources. An intrusion detection system (IDS) monitors and collects data from a target system that should be protected, processes and correlates the gathered information, and initiates responses when evidence of an intrusion is detected (Creech, *et al*,2014).

Source of input, decides the IDS type which can be network based systems or host-based systems. Network-based intrusion detection systems (NIDSs) gathers input data by observing network traffic (e.g. captured packets). Whereas host based intrusion detection systems (HIDSs), depends on events collected and monitored by the hosts. The HIDS agent monitors system integrity, application activity, file changes, host network traffic, and system logs. Using common hashing tools, file time stamps, system logs and monitoring system calls and the local network interface gives the agent insight to the current state of the local host. If an unauthorized change or activity is detected, it will alert the user via a pop-up, alert the central management server, block the activity, or a combination of the three. The decision is based on the policy that is installed on the local system (E.Kesavulu Reddy, *et al*, 2011).

Intrusion detection approach can be further classified as anomaly based IDS and signature or misuse based IDS. The anomaly based IDS attempts to detect activities that differ from the normal expected system behavior. It is solution to the problem of signature based IDS. This IDS technique is able to detect attack without prior knowledge of attack.

The signature-based IDS uses pre-known attack scenarios and compare them with incoming packets traffic but if there is an unknown attack that is its signature is not stored in IDS database then IDS is unable to detect attack. For this database need to be updated periodically.

Network traffic analysis and Audit trail analysis are the source of information for the IDS. Auditing is a mechanism to collect information regarding the activity of users and applications. Different auditing levels can be specified, and, in addition to system calls, security-relevant higher-level events can be generated as well (e.g. login events).

System calls represent the rawest interaction between a program and the host system, and have virtually no abstraction of data. Other methods such as log file analysis introduce an unavoidable level of

*Corresponding author: **Dipali Suhalal Patil**

obfuscation as they rely on data which has already been interpreted and formatted to produce the logs (Creech, *et al*,2014).

2. Related Work

F. Bin Hamid Ali and Y. Y. Len developed a host based intrusion detection system for log files. Host-based IDS are usually implemented by choosing proper parameter present in the host and using this parameter as the input to a decision engine. Audit trail and log files are some of the parameters present in a computer to provide this feature which can be used in operating system–level intrusion detection systems and application-level intrusion detection systems. The use of log files has several drawbacks. One of the drawback is that log file can be manipulated(Bin Hamid Ali, *et al*, 2011) .

S. Forrest *et al*. suggested the solution to a log file approach using system calls. As system call provides raw interaction between kernel and program, from which decision feature can be formulated. The system call approach provides very small details of each and every field without botheration of management of log files. The work done in this paper introduces a completely new way of understanding the raw system call traces, using a true semantic interpretation to improve results (Forrest, *et al*, 1996).

Shruthi.K.R *et al*. presented a host based intrusion detection system using semantic approach to system call patterns by using discontiguous system call patterns. The system provided resistance to mimicry attacks (Shruti K.R, *et al*, 2014) .

Gideon Creech and Jiankun Hu presented a semantic approach to host based IDS using contiguous and discontiguous system call patterns. Author invented a new semantic based algorithm . It used Extreme Learning Machine (ELM) as a decision engine detecting the malicious activity (Creech, *et al*,2014).

Karthikeyan K.R. presented a survey on intrusion detection tools and techniques. Author provided comparison between various intrusion detection tools (Karthikeyan K.R, *et al*, 2010).

K.Ganesh presented a semantic based intrusion detection scheme where state transition analysis, pattern matching and data mining techniques are combined to increase the detection rate. Patterns and rules are formulated based on the events detected by WSN (Ganesh K, *et al*, 2011).

Sandeep Kumar proposed an approach for misuse detection by using the pattern matching on audit trails under audit trails under UNIX to detect system attacks using colored-petri nets.(Sandeep Kumar, *et al*, 1994).
Debar have proposed pattern recognition techniques that identifies and store signature patterns of known intrusions. The activities are matched with known patterns of intrusion signatures to identify the possible intrusion. If patterns matches then it is reported as intrusions. Using this only known intrusions can be detected, unknown attack pattern cannot be detected (H. Debar, *et al*, 1999).

Ji-Rong Wen,*et al*. proposed a new approach to query clustering using user logs and their contents(Ji-Rong Wen,*et al*, 2005).

3. Preliminaries

3.1 The System Call Approach

Forrest suggested the use of system call approach for intrusion detection. He collected the normal system call traces for each program and using this traces for classification of behavior. System call represents the rawest interaction between host system and the program. In this method only system calls without any argument are observed. Semantic patterns in the system call traces are considered in this approach (S.Forrest, *et al*, 1996).

2.2 Extreme Learning Machine

The extreme learning machine (ELM) methodology is an extremely powerful decision engine in the artificial neural network family. Many types of decision engine in neural networks suffer from a high training overhead, traditionally requiring extensive resources and many iteration to reach an effective level of training. ELM completes training much faster than other traditional decision engine as it requires only one pass training. While training this decision engine only the weights between the hidden layer and output nodes are adjusted. These things make ELM approach attractive for use as a part of host based intrusion detection system (G.B.Huang, *et al*, 2004).

2.3 Semantic Analysis

The use of semantic principles to a computer system concentrates on defining a scalable set of rules governing the combination of terminating units. These rules allow complicated concepts to be expressed in terms of simple components. The conversion of highly abstract user actions to low-level kernel events is an involved process, and the formal structure provided by semantically inspire rules greatly assists in producing an effective interface.

G.creech proposed that a context-free grammar (CFG) could be applied to the system call traces as well as to the language structure used in creating the multitude of high-level programs present in any operating system. Under this new approach, system calls can be viewed as letters, with a string of contiguous system calls thus forming a word. Applying a CFG, the resulting word lists can now be combined to form phrases. It should be noted that when compiling the multi-word lists, the eventual phrases may not in fact have occurred in the training data under consideration (G.creech, *et al*, 2014).

2.4 Misuse detection using pattern matching

Misuse detection attempts to record knowledge about attacks as well as patterns and monitors for the occurrence of these patterns.

A pattern is a subsequence of system calls that a process can generate. Intrusion detection systems that make use of this system call, need to build the table of characteristic patterns. The patterns are determined by letting the process invoke as many subcommands as possible, then extracting the patterns from the corresponding sequences of system calls. A pattern matching algorithm is applied to match on the fly the system calls generated by the process examined with entries of the pattern table. Based on how well the matching can be done, conclusion is made whether the sequence of system calls represents normal or anomalous behavior.

General steps involved in this process are: 1.Recording system calls 2. Producing training data 3. Building the process model 4. Comparing real data with the process model 5. Finding attacks.

4. Proposed System

The existing system uses the anomaly based detection approach for detecting the intrusion. Anomaly detection can detect unknown intrusions. Anomaly detection technique is able to detect novel or newly generated and unknown attacks as it attempts to detect intrusions that have a significant deviation from normal behavior of legitimate user. But drawback of anomaly detection technique is that the non-intrusive behavior falling outside the normal range may be identified as an intrusion which in turn results high false positive error, also a large amount of data and audit trial is to be analyzed to model normal behavior . Anomaly detection may cause a higher rate of false alarms. In practice, an Intrusion Detection System (IDS) integrate misuse detection and anomaly detection techniques to enhance the performance of the intrusion detection.

The objective of the IDS is to find out more and more attacks in order to provide a high level security to computer system. The proposed system will use the hybrid detection technique i.e. combines the misuse detection technique with already existing anomaly based detection scheme. System will use semantic analysis of system call patterns as input. Pattern recognition and pattern matching techniques will be used as a part of the misuse detection scheme.

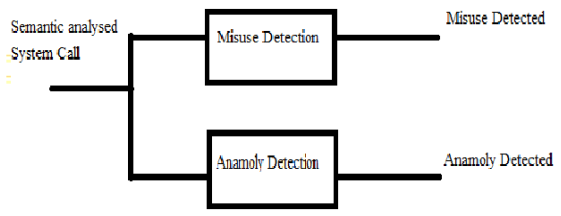

Fig.1 Architecture for Semantic Host based Intrusion Detection System

Conclusions

Internet and local networks have become everywhere and essential now a days. So organizations are increasingly employing various systems that monitor IT security breaches because intrusion events are growing day by day. The proposed system will try to improve the security of the single system i.e. host system. A semantic approach greatly degrades the ability of an attacker to bypass security systems. Any such attempt has a clear semantic pattern and will thus be caught by the guardian IDS. Anomaly detection scheme previously produces high false alarm rate. Combining previous approach with the misuse detection technique will reduce the false alarm rate resulting in improved performance of the system, ultimately increasing the security of the system. The proposed approach can increase core performance of system in terms of detection rate by improving the possibility of detection of newly introduced attacks.

References

Kumar Sandeep and Spafford Eugene H(1994), A pattern matching model for misuse intrusion detection.

S. Forrest, S. Hofmeyr, A. SoMayaji, and T. Longstaff,(1996), A Sense of Self for Unix Processes,*Proc. IEEE Symp. Security and Privacy*, pp. 120-128.

Debar H. and Dacier Marc and Wespi Andreas(1999), Towards a taxonomy of intrusion-detection systems, *Computer Networks*, vol 31,number 8, pages 805-822.

G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew(2004), Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, *Proc. IEEE Int'l Joint Conf. Neural Networks*, vol. 2,pp. 985-990.

Ji Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang(2001), Clustering User Queries of a Search Engine, *Proceeding WWW '01 Proceedings of the 10th international conference on World Wide Web* , Pages 162-168.

Karthikeyan K.R. and A. Indra(2010), Intrusion detection tools and techniques a survey, *International Journal of Computer Theory and Engineering*, vol. 2, no. 6, pp. 901-906.

E.Kesavulu Reddy, V.Naveen Reddy, P.Govinda Rajulu(2011), A Study of Intrusion Detection in Data Mining, *Proceedings of the World Congress on Engineering WCE 2011*, vol. 3.

K. Ganesh, M. Sekar, and V. Vaidehi(2011), Semantic intrusion detection system using pattern matching and state transition analysis," *Recent Trends in Information Technology (ICRTIT), 2011 International Conference* , pp. 607-612.

F. Bin Hamid Ali and Y. Y. Len(2011), Development of host based intrusion detection system for log files, *Business, Engineering and Industrial Applications (ISBEIA), 2011 IEEE Symposium* , pp. 281-285.

G. Creech and J. Hu(2014), A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns, *Computers, IEEE Transactions*, vol. 63, no. 4, pp. 807-819.

Shruthi.K.R, Shweta Hiremath, Tejaswini V Nalwade, Mangala C N(2014), A Host Based Intrusion Detection System Using Semantic Approach To Sytem Call Patterns, *IEEE Sponsored International Conference On Empowering Emerging Trends In Computer, Information Technology & Bioinformatics International Journal of Computer, Information Technology & Bioinformatics (IJCITB)*, Volume-2, Issue-2 .