Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
3(2), 2015, 93-97

ISSN 2347 – 7393

Short Communication

SURVEY ON HOST BASED INTRUSION DETECTION SYSTEMS

Kulkarni Sagar S. and Prof. Kahate Sandip A.

Department of Computer, SP COE, Otur, Pune, India

Abstract: Recently, use of security tools is increased because of increasing security threats emerging day by day. Intrusion detection system is one of such security tool that helps users to detect security threats. Intrusion detection system uses various approaches such as knowledge, behaviour to detect security vulnerabilities. This paper reviews different host based intrusion detection system and finally concludes with future requirements of host based intrusion detection system.

***Keywords*—** Anomaly Detection; Intrusion Detection; Statistical based.

Introduction

During the last decade network communication technology brings many digital devices close to each other. Current business often runs totally on data that, so any disaster to computer may bring huge loss to business. The branch of computer security focuses not only focuses on data security, but also whole computer security. There are many tools that helps to keep computer system secure from different threats such as firewalls, access control lists, etc. The limitation of these tools triggers the development of intrusion detection system. Now days nearly every big organization makes use of intrusion detection system [1][2].

The intrusion detection system is a specially designed security tool to detect security vulnerabilities [1][2]. The detection process can be real time or offline detection. There are two types of IDS, HIDS and NIDS. This categorization is based on the installation location of IDS. The HIDS can use various types of information such as log files, CPU usage, system call patterns, command sequence executed by user to detect intrusions. The NIDS uses network packets, router table for detecting intrusions. For building an IDS two approaches are used either knowledge based or anomaly based. In knowledge based system knowledge about specific attacks and system vulnerabilities is used to detect intrusions. The example of knowledge based system includes Expert system, signature based system. These type of

For Correspondence:
kulsagar325@gmail.com
Received on: June 2015
Accepted after revision: June 2015
Downloaded from: www.johronline.com

systems are simple, but such systems can not detect zero-day attacks or even variations of attack whose knowledge is given to system. The behaviour based system is uses system behaviour to detect anomalies. It is very powerful method and has the ability to detect unknown attacks [2][3].

There are number of researches that make use of log file in host based intrusion detection systems. But, using log file for intrusion detection creates problem such as, log file can grow very large and therefore management of log files becomes problematic. Also attacker can erase footprint from log files making it impossible to detect intrusion occurrence. Therefore, many researches uses another source of information called system call. The system call is called whenever a program needs some kind of service from operating system. It is strong source of information than log file, as attacker cannot hide or erase its footprints [1][2][3].

The usefulness of anomaly based IDS attracted attention of many researchers, this is because this type of system does not requires signature for each attack [1][2]. The effectiveness of anomaly based system is depends on source of information, how the information is used and threshold chosen. This paper takes journey that analyzes different HIDS and their limitations, so that useful concepts from these researchers can be taken to build new HIDS that can provide high DR with small FPR.

The rest of paper is organized as follows: Section 2 covers literature review. and section 2 contains concluding remarks.

Literature

Haystack [4] is earlier IDS that uses anomaly based approach for intrusion detection. This system applied statistical approach over feature such as CPU uses, I/O activity, etc. The monitored features range values are used as normal profile and any large deviation with respect to the normal range values is considered as anomaly. The limitation of this system is that all features are considered as independent so that there was high FPR.

MIDAS [5] is an expert system. This type of system uses rules created by human expert to detect anomalies. These rules are created by analyzing various log files. Later, the analyzing and creation of rules is done using program. This system suffers from same limitations as that of other misuse based systems, because it uses rules detect anomalies. The major problem faced by this type of system is that how to construct rules that detect anomalies and did not matches any normal activity.

In early 1980s researchers were impressed by the functionalities of human immune system and motivated towards building system that mimics functionality of human immune system. During development major problem faced by researchers was that, how computer can identify self from non-self. Forrest et al [6] model given way to researchers to identify self from nonself in computer. Their model [6] generates detectors that can identify nonself. The detectors are generated using negative selection process.

After initial work on identifying self from nonself in computer [6], Forrest et al [7] uses system call patterns to detect anomalous behaviour. The theory was that, program code is static and it uses fixed set of system calls. Therefore whenever program runs it will generate set of system call patterns. If during normal runs all the fixed length system call patterns are extracted and used to represent normal behaviour of program, then all the system call patterns not present in normal profile database is an anomaly. All the mismatches occurred during the execution of program are used to detect anomaly. The limitation of this model inherently lies in fact that, it is practically impossible to collect all types of normal runs of a program. Also as the length of system call trace can be different, setting threshold over mismatch is difficult task.

Forrest et al [8] proposed extension to their earlier model [7], in this model they used hamming distance measure to compare program behaviour, also decision of anomaly is based on mismatches occurred in local region. Therefore, this model could work even if system is trained using incomplete normal data, also the system call trace length have no effect on threshold.

Wareender et al [9] proposed sequence time delay model in which, fixed length system call patterns are used to construct normal profile of a program. The threshold was given on LFC, to detect anomaly. If during detection, any sequence does not found in database it was considered as mismatch. All the mismatches occurred in LFC (Locality Frame Count) are summed up. Warrender et al [8] extended STIDE model by considering rarity of system call patterns. According to them, rare system call patterns are suspicious. Therefore, they measure frequencies of system call patterns and eliminated rare system call patterns. The detection phase was similar as that of STIDE [8].

Somayaji et al [9] proposed model was inspired by working of human immune system. This model uses similar concept as that of Forrest et al [7], but with a difference that process is delayed based on the strength of anomaly signal. According to author, attack can only be successful if the required system calls executed in time. Therefore, if anomalous system call is delayed for certain period then it is not possible to compromise the system. This model [9] is included in many Linux kernels as a first step for preventing intrusions.

Vardi et al [10] revolutionalized Warrender et al [9] t-STIDE model, in which the rare system call patterns are identified using vocabulary concept. According to Vardi et al [10], using frequency based measure for measuring rarity is problematic. Therefore, in their model they calculated rarity of system call patterns as the number of system call traces in which the given pattern is seen. This rarity index measure does not create bias as that occurred in frequency based measure.

Lee et al [11] proposed model uses data mining technique to create generalized rules for detecting intrusions. For effectiveness, it is important that training data supplied must contain greater number of abnormal samples than that of normal samples. Rules are created randomly and support is calculated. If support is greater than threshold then rule is included in database to detect anomaly.

Xuan et al [12] uses HMM to decide whether the mismatch occurred due to incompleteness or because of intrusion. The initial normal profile generation is similar to Forrest et al [8] work, during detection the mismatch sequences are given for finding the probability required to produce the given sequence. If the probability required to produce the given sequence is lower than given threshold then it is considered as anomaly. Finally, threshold over LFC is given for detecting intrusions.

Wespi et al [13] observed problems associated with fixed length patterns and proposed the use of variable length patterns for intrusion detection. The variable length patterns are constructed using Teiresias pattern matching algorithm. During detection, number of system call that are mismatched in given trace are used for decision purpose.

Liao et al [14] uses text categorization technique for classifying normal system call trace from abnormal one. There model maps system call as letters, system call trace as a document. Finally these documents are categorized using kNN classifer. For doing so, process system call trace is converted into vector and cosine similarity was calculated to find similarity between different processes. During detection, similarity of test trace is calculated, if it turns to be 1, then it is categorized as normal. Otherwise, average similarity is calculated by aggregating k nearest neibour.

Ye et al [15] proposed model uses variable length patterns for detecting intrusions. Initial variable length patterns are constructed using Teiresias pattern matching algorithm. Then, during detection the average value of hamming distance between patterns encountered is used for anomaly detection.

Syed et al [16] proposed model uses kernel events to detect intrusions. According to author, process calls number of system call during its lifetime, therefore it is complex to gain important information about their activity. Their model considers fixed number of kernel states such as, File System, memory management, interprocess communication, networking, etc. A process can be either any one state at a time. During detection is carried out by calculating probabilities of occurrence of states in normal and abnormal traces.

Creech et al [17] proposed model uses semantic theory for intrusion detection. According to them, system call patterns are not random combination of each other, therefore it is possible to generate any normal trace if all normal patterns are known. The use of semantic theory is motivated from the fact that, if a program follows some kind of grammar, then sequences of system calls generated after execution of these programs will also follow similar grammar. Their model extract semantic information about system call patterns and this semantic information is used to detect valid and invalid system call patterns. To use this theory, they map system calls into individual letters, then all the fixed length patterns of 2-n are extracted as words, finally different phrases with their occurrence count is built. This phrase-count dictionary is used as normal profile in their model. During detection, the normal phrases seen in traces are greater than anomalous one then trace is considered as intrusive one. The problem with there model is that, it requires high amount of training time particularly for dictionary construction, also best DR reaches up to 85% with FPR of 10%.

Conclusion

It has been found that variable length patterns are useful for reducing the dictionary size required to profile normal activity of a program. But, it is also found while experimenting that, pattern extraction using method proposed by Wespi et al [13] does not cover all system call patterns, this is due to selection of longest pattern while pattern matching. It is also found that, the use of single threshold for intrusion detection generates FPR, therefore in future systems there should be use of more than one threshold. Finally, the intrusion detection system must make use of approximate pattern matching to reduce FPR, encountered due to incomplete training.

References

[1] John McHugh, Alan Christie, and Julia Allen, ”The Role of Intrusion Detection Systems, IEEE SOFTWARE,SEP 2000.

[2] Mehdi Bahrami and Mohammad Bahrami, “An overview to Software Architecture in Intrusion Detection System”, Soft Computing And Software Engineering (JSCSE), 2011.

[3] Herve Debar, “An Introduction to Intrusion-Detection Systems”, IBM Research, 2011.

[4] Debra Anderson, Teresa F. Lunt, Harold Javitz, Ann Tamaru, Alfonso Valdes, "Haystack: an intrusion detection system”, Aerospace Computer Security Applications Conference, Oct. 7481, Dec 1988.

[5] M. M. Sebring, E. Shellhouse, M. E. Hanna, and R. A. Whitehurst, "Expert systems in intrusion detection: A case study", Proceedings of the 11th National Computer Security Conference, Oct. 7481, 1988.

[6] Stephanie Forrest, and Alan Peterson, "Self - Nonself Discrimination in Computer”, Proceeding of 1994 IEEE Symposium on Research in Security and Privacy, 1994.

[7] S. Forrest,S. A. Hofmeyr and A. SoMayaji, ”A sense of self for Unix Processes”, IEEE Symposium, May 1996.

[8] S. Forrest,S.A. Hofmeyr and A. SoMayaji, ”Intrusion Detection Using Sequences of System Calls”, IEEE Symposium, May 1996.

[9] C. Warrender, S. Forrest, and B. Pearlmutter, ”Detecting intrusions using system calls: alternative data models”, Proceedings of the 1999 IEEE Symposium,1999.

[10] A. Somayaji and S. Forrest, "Automated Response Using System-Call Delays.", Proceedings of the 9th USENIX Security Symposium, The USENIX Association, Berkeley, 2000

[11] Wen-Hu Ju and Yehuda Vardi, “Profiling Unix Users And Processes Based On Rarity of Occurrences Statistics with Applications to Computer Intrusion Detection”, Fourth Aerospace Computer Security Applications Conference, October 1988

[12] John Andreas Wespi and Herv Debar, “An intrusion detection system based on the teiresias pattern discovery algorithm”, Proceedings of EICAR, 1998.

[13] Wenke Lee and Salvatore J. Stolfo , "Data Mining Approaches for Intrusion Detection",7th USENIX Security Symposium, Jan 1998.

[14] Xuan Dau Hoang, Jiankun Hu, Peter Bertok, "A Multi-layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls" ,IEEE, October 1988.

[15] Ye Du, Ruhui Zhang, and Youyan Guo, "A Useful Anomaly Intrusion Detection Method Using Variable-length Patterns and Average Hamming Distance", Journal of Computers, Aug 2010.

[16] Syed Shariyar Murtaza, Wael Khreich, Abdelwahab Hamou-Lhadj, Mario Couture, "A Host-based Anomaly Detection Approach by Representing System Calls as States of Kernel Modules", IEEE 24th International Symposium on Software Reliability Engineering (ISSRE), 2013.

[17] G. Creech and J. Hu.,"A Semantic Approach to Host-based Intrusion Detection Systems Using Contiguous and Dis-contiguous System Call Patterns", IEEE Transactions on Computers, 2014.