Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 2063 – 2069

Procedia Engineering

www.elsevier.com/locate/procedia

International conference on modeling optimisation and computing (ICMOC-2012)

A COMPREHENSIVE SURVEY ON APPROACHES TO INTRUSION DETECTION SYSTEM

Deepa A J[1], Dr.V.Kavitha[2]

1. Research Scholar, Anna university of Technology Tirunelveli, Tirunelveli.

2 Associate Professor, Dept. of CSE, University College of Engineering, Nagercoil.

Abstract

As there is a wide spread of Internet services all over the world, many kinds and large number of security threats are increasing. Since it is not technically feasible to build a system with no vulnerabilities, Intrusion detection System, which can effectively detect intrusion accesses have attracted attention. Various architectures and different soft computing based approaches have been proposed to detect computer network attacks. In this paper we present a survey of intrusion detection systems. The survey was about the existing types, techniques and approaches of intrusion detection systems in the literature.

Keywords: *Intrusion Detection System, Security Threats, Network.*

* Corresponding author. Tel.: +91-9994589876;

E-mail address: ajdeepajames@gmail.com.

1. INTRODUCTION

Due to the open society of the internet, the security of our computer systems and data is always at risk. The most important requirements for handling data are availability, integrity and confidentiality. Computer break-ins and misuse have become common features. Security mechanisms of a system are designed to prevent unauthorized access to system resources and data. The field of trying to detecting these intrusions called Intrusion Detection. An intrusion detection system monitors the activities of a given environment and decides whether these activities are malicious (intrusive) or legitimate (normal)

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.06.248

based on system integrity, confidentialty and the availability of the information resources. The intrusion Detection System collects information about the system being observed. This collected audit data is processed by the detector and eliminates unnecessary information from the audit data and then makes a decision by monitoring the activities of the system, analyzing the activities to determine that any of the activity is violating the security rules. Once the intrusion detection system determines that an unusual activity or an activity that is known to be an attack occurs, it then generates an alarm to alert the security administrator. Intrusion Detection System can also initiate a proper response to the malicious activity.

Intrusion Detection is traditionally divided into two categories, i.e, Misuse detection and anomaly detection. Misuse detection searches for specific patterns or sequence of programs and user behaviors that match well known intrusion scenarios. Anomaly detection develops models of normal network behaviors, and new intrusions are detected by evaluating significant deviations from the normal behaviour.The advantage of anomaly detection is that have not been observed yet. There are many reasons that make use of Intrusion detection System for more secures and protect data on computers. Such as, detecting the attacks and other security violations that are not prevented by other security measures, detecting and dealing with the preambles to attacks and it act as quality control for security design and administration, especially of large and complex enterprises.

2. EVOLUTION OF INTRUSION DETECTION SYSTEM

The goal of intrusion detection is to monitor network assets to detect anomalous behavior and misuse. This concept has been around for nearly twenty years but only recently has it seen a dramatic rise in popularity and incorporation into the overall information security infrastructure.

In 1980, James Anderson's seminal paper, written for a government organization, introduced the notion that audit trails contained vital information that could be valuable in tracking misuse and understanding user behavior. With the release of this paper, the concept of "detecting" misuse and specific user events emerged. His insight into audit data and its importance led to tremendous improvements in the auditing subsystems of virtually every operating system. Anderson's conjecture also provided the foundation for future intrusion detection system design and development. His work was the start of host-based intrusion detection and IDS in general.

In 1983, Dr. Dorothy Denning, [1] began working on a government project that launched a new effort into intrusion detection development. Their goal was to analyze audit trails from government mainframe computers and create profiles of users based upon their activities. One year later, Dr. Denning

helped to develop the first model for intrusion detection, the Intrusion Detection Expert System (IDES), which provided the foundation for the IDS technology development that was soon to follow.

In 1984, SRI developed a means of tracking and analyzing audit data containing authentication information of users on ARPANET, the original Internet. Soon after, SRI completed a Navy SPAWAR contract with the realization of the first functional intrusion detection system, IDES.

In 1988, the Haystack project at Lawrence Livermore Labs released another version of intrusion detection for the US Air Force. This project produced an IDS that analyzed audit data by comparing it with defined patterns.

The subsequent iteration of this tool was called the Distributed Intrusion Detection System (DIDS). DIDS augmented the existing solution by tracking client machines as well as the servers it originally monitored. Finally in 1989, the developers from the Haystack project formed the commercial company, Haystack Labs, and released the last generation of the technology, Stalker. Crosby Marks says that "Stalker was a host-based, pattern matching system that included robust search capabilities to manually and automatically query the audit data." The Haystack advances, coupled with the work of SRI and Denning, greatly advanced the development of host-based intrusion detection technologies.

In 1990, Heberlein introduced the idea of network intrusion detection Heberlein was the primary author and developer of Network Security Monitor (NSM), the first network intrusion detection system NSM was deployed at major government installations where network traffic analysis provided massive amounts of information. This new awareness generated more interest in the field of intrusion detection and investments in that market increased significantly. Heberlein's contributions also extended to the DIDS project where, along with the Haystack team, he introduced the first idea of hybrid intrusion detection. The work of the Haystack project and the introduction of the Network Security Monitor revolutionized the IDS field and brought it into the commercial world.

Commercial development of intrusion detection technologies began in the early 1990s. Haystack Labs was the first commercial vendor of IDS tools, with its Stalker line of host-based products. SAIC was also developing a form of host-based intrusion detection, called Computer Misuse Detection System (CMDS).

The Lawrence Berkeley National Laboratory announced Bro in 1998, which used its own rule language for packet analysis from libpcap data. Network Flight Recorder (NFR) in 1999 also used libpcap. APE was developed as a packet sniffer, also using libpcap, in November, 1998, and was renamed Snort one month later, and has since become the world's largest used IDS/IPS system with over 300,000 active users.

The Audit Data Analysis and Mining (ADAM) IDS in 2001 used tcpdump to build profiles of rules for classifications.

In 2003, Dr. Yongguang Zhang and Dr. Wenke Lee argue for the importance of IDS in networks with mobile nodes.

Currently, market statistics show that IDS is amidst the top selling security vendor technologies and should continue to rise. Furthermore, government initiatives, such as the Federal Intrusion Detection Network, (FIDNet) are also adding impetus to the evolution of IDS. Advancements in IDS will ultimately push security technology into a whole new arena of automated security intelligence.

3.Types of intrusion detection system

Network-Based IDS: Network-based IDS monitors network traffic using techniques like packet sniffing to collect network traffic data and tries to detect malicious activity such as denial of service attacks; port scans or even attempts to crack into computers. [2].

Host-Based IDS: Host-based IDS monitors and analyzes system calls, application logs, file-system modifications and other host activities to identify the intrusion such as unauthorized remote login attempt, attempt to access unprivileged data. It normally works with Network-based IDS.

***Stack-based* IDS:** Stack-based IDS examines the packets as they go through the TCP/IP stack and, therefore, it is not necessary for them to work with the network interface in promiscuous mode. This fact makes its implementation to be dependent on the Operating System that is being used.

Protocol-Based IDS: Typically protocol-based IDS are installed on a web server, and they are used for monitoring and analysis of the protocol in use of the computing system. If there is a deviation from intended behaviour of protocol then it can be detected as intrusion [3].

Graph-Based IDS: Graph-based IDS concerned with detecting intrusions that involve connections between many hosts or nodes. A graph consists of nodes representing the domains and edges representing the network traffic between them.

4. RECENT APPROACHES IN IDS

Association Rules: In [4],[5],[6] Data mining approaches for intrusion detection include association rules and frequent episodes, which are based on building classifiers by discovering relevant patterns of program and user behavior. Association rules and frequent episodes are used to learn the record patterns that describe user behavior. These methods can deal with symbolic data, and the features can be defined in the form of packet and connection details. However, mining of features is limited to entry level of the packet and requires the number of records to be large and sparsely populated; otherwise, they tend to produce a large number of rules that increase the complexity of the system.

Data Clustering Methods: In [6],[7] Methods such as the k-means and the fuzzy c-means have also been applied extensively for intrusion detection. One of the main drawbacks of the clustering technique is that it is based on calculating numeric distance between the observations, and hence, the observations must be numeric. Observations with symbolic features cannot be easily used for clustering, resulting in inaccuracy. In addition, the clustering methods consider the features independently and are unable to capture the relationship between different features of a single record, which further degrades attack detection accuracy.

Bayesian network: In [8],[9],[10] Bayesian Network is a graphical representation of the joint probability distribution function over a set of variables. The network structure can be represented in Bayesian Network as a Directed Acyclic Graph where each node represents a random variable and each edge between nodes shows the relation between nodes. Individual invents which occurs during attack are represented as nodes in the graph and relationship between those events are represented as edges of the graph and this graph is then used to detect the intrusion. Bayesian network can also be used for intrusion detection . However, they tend to be attack specific and build a decision network based on special characteristics of individual attacks. Thus, the size of a Bayesian network increases rapidly as the number of features and the type of attacks modeled by a Bayesian network increases.

Hidden Markov models (HMMs): In [11],[12],[13] To detect anomalous traces of system calls in privileged processes Hidden Markov Models are applied.However,modeling the system calls alone may not always provide accurate classification as in such cases various connection level features are ignored. Further, HMMs are generative systems and fail to model long-range dependencies between the observations .

Decision tree: The decision trees select the best features for each decision node during the construction of the tree based on some well-defined criteria. One such criterion is to use the information gain ratio. Decision trees generally have very high speed of operation and high- attack detection accuracy even if dealing with a large amount of data.

Support Vector Machine (SVMs): Though the neural networks can work effectively with noisy data, they require large amount of data for training and it is often hard to select the best possible architecture for a neural network. Support vector machines have also been used for detecting intrusions. Support vector machines map real-valued input feature vector to a higher dimensional feature space through nonlinear mapping and can provide real-time detection capability, deal with large dimensionality of data, and can be used for binary-class as well as multiclass classification.

Honey pot: The honey pot is used as bait in the form of a vulnerable system to trap hackers and keep them away from accessing the critical information in the main system. In this technique alarming adversaries, initially detected by the IDS, will be rerouted to a honeypot network for a more close investigation. If as a result of this investigation, it is found that the alarm decision made by the IDS of the agent is wrong, the connection will be guided to the original destination in order to continue the previous interaction. This action is hidden to the user. Such a scheme significantly decreases the alarm rate and provides a higher performance of IDS.

Genetic algorithms (GAs): Genetic algorithms mimic the natural reproduction system in nature where only the fittest individuals in a generation will be reproduced in subsequent generations, after undergoing recombination and random change.

Fuzzy logic: In [14] A set of rules can be created to describe a relationship between the input variables and the output variables, which may indicate whether an intrusion has occurred

5. CONCLUSION AND FUTURE SCOPE

This paper reviews and tries to summarize different types, methods and approaches for intrusion detection system and also provides a strong platform to detect anomalies. This survey focuses that there are limited researchers in detecting application based intruders. The future step for this proposal is the development of application based intrusion detection system.

6. REFERENCES

1. D. Denning, “An Intrusion-Detection Model,” Proc. 1986 EEE Symp. Sec. Privacy, Apr. 7–9, 1986, pp. 118–31.
2. D.S. Kim and J.S. Park, “Network-Based Intrusion Detection with Support Vector Machines,” Proc. Information etworking, Network-ing Technologies for Enhanced Internet Services Int’l Conf. (ICOIN ’03),pp. 747-756, 20
3. L. Portnoy, E. Eskin, and S. Stolfo, “Intrusion Detection with Unlabeled Data Using Clustering,” Proc. ACM Workshop Data Mining Applied to Security (DMSA), 2001.
4. T. Abraham, IDDM: Intrusion Detection Using Data Mining Techniques, http://www.dsto.defence./gov.au/publications/ 2345/DSTO-GD-0286.pdf, 2008.
5. R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rules between Sets of Items in Large Databases,” Proc. ACM SIGMOD, vol. 22, no. 2, pp. 207-216, 1993.
6. Khosravifar B and Bentahar J 2008) An experience improving intrusion detection systems false alarm ratio by using honeypot. IEEE, 22nd Intl. Conf. on Advanced Information Networking and Applications.pp: 997-1004.

7. W. Lee, S. Stolfo, and K. Mok, “Mining Audit Data to Build Intrusion Detection Models,” Proc. Fourth Int’l Conf. Knowledge Discovery and Data Mining (KDD ’98), pp. 66-72, 1998.
8. N.B. Amor, S. Benferhat, and Z. Elouedi, “Naive Bayes vs. Decision Trees in Intrusion Detection Systems,” Proc. ACM Symp. Applied Computing (SAC ’04), pp. 420-424, 2004.
9. C. Kruegel, D. Mutz, W. Robertson, and F. Valeur, “Bayesian Event Classification for Intrusion Detection,” Proc. 19th Ann. Computer Security Applications Conf. (ACSAC ’03), pp. 14-23, 2003.
10. Lu Huijuan, Chen Jianguo and d Wei Wei (2008) Two stratum Bayesian network based anomaly detection model for intrusion detection system. IEEE, Intl. Symp. on Electronic Commerce & Security.pp:482-487.
11. Y. Du, H. Wang, and Y. Pang, “A Hidden Markov Models-Based Anomaly Intrusion Detection Method,” Proc. Fifth World Congress on Intelligent Control and Automation (WCICA ’04), vol. 5, pp. 4348-4351, 2004.
12. S. Forrest, S.A. Hofmeyr, A. Somayaji, and T.A. Longstaff, “A Sense of Self for Unix Processes,” Proc. IEEE Symp. Research in Security and Privacy (RSP ’96), pp. 120-128, 1996.
13. C. Warrender, S. Forrest, and B. Pearlmutter, “Detecting Intrusions Using System Calls: Alternative Data Models,” Proc. IEEE Symp. Security and Privacy (SP ’99), pp. 133-145, 1999.
14. H. Shah, J. Undercoffer, and A. Joshi, “Fuzzy Clustering for Intrusion Detection,” Proc. 12th IEEE Int’l Conf. Fuzzy Systems (FUZZ-IEEE ’03), vol. 2, pp. 1274-1278, 2003.