International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062165

A Survey on the Network Intrusion Detection System Using Data Mining Techniques

Saumya Saraswat, Rahul Yadav, Phalguni Pathak
Department of Computer Science Application, ITM University, Gwalior, Madhya Pradesh, India

ABSTRACT

The idea of making everything available easily and universally has led to a revolution in the field of networking. Despite the tremendous growth of technologies in the field of networks and information technology, we still cannot avoid the theft / attack of our resources. This may not apply to small organizations, but it is a serious problem regarding industry / business or national security. Organizations face an increasing number of threats every day in the form of viruses, intrusions, etc. Since organizations have opted for many different mechanisms in the form of intrusion detection and prevention systems to protect themselves from this type of attack, there are many breach security systems that go undetected. To understand safety hazards and intrusion detection and prevention (IDPS) systems, we will first analyze common security breaches and then discuss what the different opportunities and challenges are in this particular field. In this document, we conducted a survey on the overall progress of intrusion detection systems. We analyze the existing types, techniques and architectures of intrusion detection systems in the literature. Finally, the future scope is mentioned.

Keywords : Architecture, Attack, Detection, IDS, Prevention, Detection

I. INTRODUCTION

The intrusion detection field has grown considerably in recent years and a large number of intrusion detection systems have been developed to meet different needs. Many past events have shown that intrusion prevention procedures alone, such as encryption and authentication, which often represent a first line of defense, are not enough. As the system becomes more complex, there are also more weaknesses, which lead to more security problems. Intrusion detection can be used as a second defense wall to protect the network from such problems. If intrusion is detected, a response can be initiated to prevent or minimize damage to the system. After perimeter checks, firewalls and authentication and access checks block certain actions, some users can use a computer system. Most of these checks are preventive: they prevent known bad things from happening. Intrusion Detection Systems (IDS) complement these preventive checks as the next line of defense.

The definition of an intrusion detection system does not include the prevention of the occurrence of intrusions, but only the detection and information of an operator. There are some Intrusion Detection Systems (IDPS) that attempt to react when they detect an unauthorized action. This reaction generally includes trying to contain or stop the damage, for example by terminating a network connection. When an IDS detects an intrusion, it logs the event, stores the relevant data / traffic, notifies an administrator and in some cases tries to intervene. In addition to the obvious benefits of an IDS, the stored data and records provide valuable forensic information and can be used as evidence in a legal case against the attacker. Most intrusion detection systems try to do their job in real time, but there are also intrusion detection systems that don't work in real time, due to the nature of the analysis they perform or because they are analysis oriented. forensic. An IDS is much like an unease system, some being more forward-looking and intellectual than others. When To create an IDS, many problems need to be considered, such as data collection, data preprocessing, intrusion detection, reporting and response. Among them, the recognition of intruders is at the center. The audit data are examined and compared with detection models, which describe models of intrusive or benign behavior, in order to identify successful and unsuccessful intrusion attempts. Many intrusion detection systems have been proposed in traditional wired networks, where all traffic must pass through switches, routers or gateways. Therefore, IDS can be easily added and implemented on these devices [1, 2]. On the other hand, MANETs do not have these devices. In addition, the support is open, so legitimate and malicious users can access it. Furthermore, there is no clear parting between normal and unusual activities in a mobile surroundings. Since nodes can move arbitrarily, false routing information can come from a compromised node or from a node that has outdated information. Therefore, current IDS techniques in wired networks cannot be applied directly to MANET.

Cloud security is an evolving subdomain of computer security, network security and, more generally, information security. There are numerous security problems / concerns associated with cloud computing, but these problems fall into two broad categories: security problems faced by cloud providers (organizations that provide software, platform or infrastructure as a service through (cloud computing) and security problems addressed by its customers (companies or organizations that host applications or store data in the cloud) [3, 4, 5]. In a public cloud that allows a shared environment of more tenants, with increasing number of users, the risks to security become more intense and diverse. Attack surfaces subject to attacks and security mechanisms that ensure effective protection on the client and server side [3, 4, 5] must be identified. Private clouds are considered more secure than public clouds; They still have multiple problems that, if left unattended, can lead to significant security gaps rity.

This paper does not provide a method to deal with new attacks but tries to explain the current techniques and their potent approaches to deal with attacks. This paper also explains the future scope for IDS's face.

II. GENERAL ARCHITECTURE OF AN IDS

A panoramic view of an IDS reveals that it is a security system that continuously monitors a target system and produces audit evidence. These audit tests contain processed data generated from information from the target system [6]. Audit tests can be automatically inspected by some tools, online or offline and / or used by a manual authority (an administrator) who analyzes these records very carefully [7]. The general architecture of an IDS is shown in Figure 1. The location of an IDS is a big problem and depends on several factors, such as level of security, budget and environment. In general, an IDS is placed at the entry / exit points of the network or with the hosts themselves, or sometimes a combination of both. The job of an IDP is simply to monitor data, analyze it and prevent intrusion. An anomalous behavior detected by an IDP system can be treated automatically or by causing alarms in the manual station. An IDP distinguishes between normal and abnormal behavior based on knowledge or policies already defined in its database.

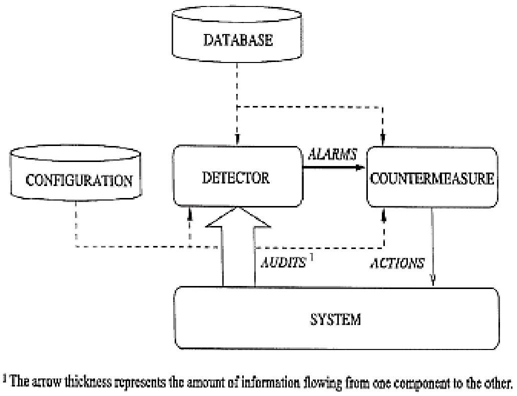
DATABASE
CONFIGURATION
DETECTOR
ALARMS
COUNTERMEASURE
AUDITS 1
ACTIONS
SYSTEM
1 The arrow thickness represents the amount of information flowing from one component to the other.

Fig. 1: General Architecture of an IDS [1]

III. TYPES OF IDS

There are many different ways to classify different types of IDS in a production network. These classifications are not mutually exclusive; for example, a network-based IDS might use the signature-based detection approach. The following diagram describes the most common IDS classification methodologies, although the list is certainly not exhaustive.

3.1. Host-based IDs

A host-based IDS (HIDS) is an IDS that generally activates within a computer, node or device. Its main function is internal observing, although many alternatives of HIDS have been developed that can be used to monitor networks [8]. Mainly, it monitors and analyzes the internal components of a computer, node or device. An HIDS determines if a system has been compromised and warns administrators accordingly [9]. For example, it is possible to detect a fraudulent program that suspiciously accesses the resources of a system or to discover that a program has corrupted the registry. HIDS were the first types of intrusion detection software designed [10]. Unlike network-based IDS, an HIDS can inspect the entire flow of communications. NIDS evasion practices, such as shattering attacks or session interweaving, are not applied because HIDS can inspect the fully recombined session as presented to the operating system [11]. Encrypted communications can be monitored because an HIDS inspection can observe traffic before it is encrypted. This suggest that HIDS monograms will still be able to compete with common attacks and not be blinded by encryption. An HIDS is also able to perform additional system-level checks that can only be performed by the IDS software installed on a host computer, such as checking file integrity, monitoring the registry, analyzing the registry, detecting rootkit and active response [11].

3.2. Network-Based Intrusion Systems

NIDS works at the network level by analyzing packets traveling here and there on a network. Its existence is clearly isolated from network firewalls and can be considered as a second level of security systems. They work invisibly and therefore can be more effective. NIDS can detect large amounts of data and, if it encounters abnormal activity, it can block all traffic related to that particular service.

3.3. Signature-based detection

In the signature-based approach, an IDS searches for packages and compares them with predefined rules or schemes known as signatures defined in the database. These attack signatures convey specific traffic or activities based on known intrusive activities [10]. The main benefit of this practice is the modest and well-organized processing of audit data. Signature-based approaches have a much lower false positive rate. On the other hand, the very natural surroundings of signature-based detection means that such an methodology is fruitless against zero-day attacks for which an established set of attack rules or methods have not yet been discovered. With the frequency of new attacks and malicious activity every hour, a signature-based IDS is as good as the last one in its signature and rule set database.

3.4. Anomaly-based detection

Fault-based IDS work by identifying user models or user groups that have already been defined. This approach seeks variations and deviations from established reference behavior which could indicate malice. It implies a greater amount of processing that the anomaly detector uses to study the behavior of the system since its audits [12]. The baseline must first be created for system, network or program activity. This baseline is the outline of what a scenario, usage, bandwidth or normal behavior would look like in a specific network environment. Subsequently, any activity that deviates from the baseline is treated as a possible intrusion [13] and a warning would be generated. The foremost enhancement of the anomaly-based approach is its ability to detect zero-day attacks, as it does not depend on an established signature database, but simply on deviations from an established baseline. The deeds of each target system or network is exclusive, therefore anomaly-based approaches use custom profiles which in turn make it difficult for an attacker to know for sure what activity he can perform without triggering an alarm. On the other hand, anomaly-based IDS have a high false positive rate. It also takes time to establish a reference behavior when it is first placed in a new network environment or host device. These systems are also more complex and the difficulty of associating an alarm with the specific event that triggered this alarm [10].

IV. DETECTION TECHNIQUES

From different sources, systems such as expert rule-based systems, state transition analysis, and genetic algorithms are direct and efficient ways to implement signature detection. In the detection of anomalies, inductive sequential models, artificial neural networks, statistical analyzes and data extraction methods were used. There are several types of frameworks used for anomaly-based detection. This section presents an extensive study of the various intrusion detection classification techniques and hybrid detection techniques. Some proposed methods could be described as follows.

4.1 Bayesian networks

Bayesian networks are probabilistic graphical models representing sets of variables and their probabilistic dependencies. Bayesian theory is named after Thomas Bayes. Its theory can be explained as follows: if the events A1, A2, ... and An constitute a partition of the sample space S such that P (Ak) ≠ 0 for k = 1,2, ..., n, then for any event B such that P (B) ≠ 0:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)} = \frac{P(A_i)P(B|A_i)}{P(B)}$$

4.2. Genetic algorithm (GA)

GA is a search technique used to find a suitable solution to search problems. Genetic algorithms have been applied to anomaly detection in many ways, as they are a flexible and powerful research method. Some network intrusion detection approaches have used genetic algorithms for instance classification, while others, such as the fuzzy data mining approach, have applied this technique for selecting features. To list a GA benefit, select the best functionality and it has better efficiency, but its method is complex.

4.3. Abnormal detection

The anomalous detection approach is based on the idea of semi-supervised learning in which the system would learn the basic data and would consider any instance that does not fit the normal data profile as an anomaly. Most anomaly detection algorithms require a reference dataset to train the model and assume that anomalies can be treated as patterns never before observed. Since an anomalous value is defined as a data point very different from the rest of the data, therefore, based on a certain measure, we use various anomalous detection schemes to see how effectively these schemes can handle anomaly detection problems. In statistical based outlier detection techniques, the data points are modeled using a stochastic distribution and these points are determined to be anomalous according to their relationship with this model.

4.4. Clustering

This technique is based on two important assumptions [14]. First, most network connections represent normal traffic and only a very small percentage of that traffic is harmful. Secondly, malicious traffic is statistically different from normal traffic. Anomalies will be detected based on cluster size, meaning large clusters must be basic data, while the rest are malicious attacks. Clustering is unsupervised learning. There is no need to label the data and natural patterns are extracted from the data. It does not require the use of a tagged dataset for training

4.5. Neural networks

Neural networks are networks of computational elements that cooperatively appliance complex mapping functions. First, networks are trained with a tagged dataset. Test instances are placed on the network to be classified as normal or abnormal. An example of the neural network technique widely used in anomaly detection is Support Vector Machines (SVM) [15]. This technique would be operative if the exact characteristics of the attack are already known. However, these intrusions change constantly due to the individual approaches adopted by the attackers and the periodic changes made to the software and hardware of the target systems. Due to the wide variety of attacks and attackers, despite your dedicated commitment to constantly updating the rules base of an expert system, you can never hope to accurately identify the variety of intrusions. For these changing natures of these network attacks, we need a flexible defensive system capable of analyzing these huge amounts of network traffic in a less structured way than rule-based systems. For example, a neural network-based signature detection system could solve many of the problems encountered in rule-based systems. The intrinsic speed of neural networks is another advantage of this approach, as it requires timely identification of attacks and the speed of processing neural networks could allow intrusion responses to take place before irreversible damage can be caused. to the system. It has a high signal-to-noise ratio and requires more time and a more sample training phase.

V. FUTURE SCOPE IN IDS TECHNIQUES

The successful growth of artificial intelligence has posed a major challenge in incorporating this new field into intrusion detection systems. Currently limited by its new implementation [16], it will constitute an important contribution to the IDS methodology.

The use of fuzzy logic can also be effective in IDS. His ability to process big data and derive meaning and patterns can be applied to find attacks. Gradually, he continues to learn to track past penetrations and to analyze data from more recent ones.

VI. CONCLUSION

We have introduced an overview of the different types of intrusion detection systems, approaches, methodologies and techniques for IDS. Each IDS technique and class has its superiority and limitations, so we must keep this in mind when selecting the best approach. We focus our study on the most common intrusion detection models, such as NIDS and HIDS, and on the anomaly and signature detection approach. This survey study offers an overview of some intrusion detection approaches. Some approaches work best in one setting but then become weak in other settings. We need to develop a generic technique that can help us protect our networks in any environment. This requires a detailed understanding of existing techniques and their shortcomings, so that researchers can find ideas for overcoming weaknesses and developing a much stronger approach to dealing with intrusions.

VII. REFERENCES

[1]. Y. F. Jou, F. Gong, C. Sargor, X. Wu, S. Wu, H. Chang, and F. Wang, “Design and Implementation of a Scalable Intrusion Detection System for the Protection of Networks Infrastructure,” Proceedings of DARPA Information Survivability Conference and Exposition, vol. 2, pp. 69-83, January 2000.

[2]. E. Y. K. Chan et al., “IDR: An Intrusion Detection Router for Defending against Distributed Denial-of-Service (DDoS) Attacks,” 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04), pp. 581-586, May 2004.

[3]. Shyam Nandan Kumar, “DecenCrypto Cloud: Decentralized Cryptography Technique for Secure Communication over the Clouds.” Journal of Computer Sciences and Applications, vol. 3, no. 3 (2015): 73-78.

[4]. Shyam Nandan Kumar, and Amit Vajpayee, “A Survey on Secure Cloud: Security and Privacy in Cloud Computing”, American Journal of Systems and Software, vol. 4, no. 1, pp. 14-26, 2016

[5]. Shyam Nandan Kumar, and Amit Vajpayee, “ASP: Advanced Security Protocol for Security and Privacy in Cloud Computing.” American Journal of Information Systems, vol. 4, no. 2, pp. 17-31. 2016.

[6]. Debar, Hervé, Marc Dacier, and Andreas Wespi. "Towards a taxonomy of intrusion-detection systems." Computer Networks 31.8 (1999): 805-822.

[7]. Karlzén, Henrik. "An Analysis of Security Information and Event Management Systems- The Use or SIEMs for Log Collection,Management and Analysis." (2009).

[8]. Singh, A., & Singh, M. (2014). Analysis of Host-Based and Network-Based Intrusion Detection System. International Journal of Computer Network and Information Security IJCNIS, 41-47.

[9]. Pieter de Boer, Martin Pels, "Host-based Intrusion Detection Systems", Revision 1.10 – 2005, p: 5-7.

[10]. Gupta, M. (2015). Hybrid Intrusion Detection System: Technology and Development. International Journal of Computer Applications IJCA, 5-8.

[11]. Hay, A., & Cid, D. (2008). OSSEC host-based intrusion detection guide. Burlington, Mass.: Syngress Pub.

[12]. James Cannady, Jay Harrell, 1996. "A Comparative Analysis of Current Intrusion Detection Technologies", p: 6.

[13]. Gangwar, A., Sahu, S., Int. Journal of Engineering Research and Applications; ISSN : 2248-9622, Vol. 4, Issue 4(Version 1), April 2014, pp.67-72.

[14]. L. Portnoy, E. Eskin, and S.J. Stolfo. Intrusion detection with unlabeled data using clustering. In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA2001), pages 76–105. Philadelphia, PA, 2001.

[15]. S. Mukkamala, G. Janoski, and A. Sung. Intrusion detection using neural networks and support vector machines. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), volume 2, 2002

[16]. Frank, Jeremy. "Artificial intelligence and intrusion detection: Current and future directions." Proceedings of the 17th National Computer Security Conference. Vol. 10. 1994.

Cite this article as : Saumya Saraswat, Rahul Yadav, Phalguni Pathak, "A Survey on the Network Intrusion Detection System Using Data Mining Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 3, pp. 31-36, May-June 2020. Available at doi : https://doi.org/10.32628/CSEIT2062165
Journal URL : http://ijsrcseit.com/CSEIT2062165