Raj Kumar Pandey et al. 2019, 7:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

A Survey on Intrusion Detection System Techniques and Features for Identification

Ph.D Scholar Raj Kumar Pandey
Dept. of Computer science & Engineering, AISECT University
Bhopal , M.P. , India
sharma01.scope@gmail.com

Dr. Shiv Shakti Shrivastava
Dept. of Computer Science & Engineering , AISECT University
Bhopal, M.P. , India
shivshakti18@gmail.com

Abstract-The intrusion detection systems (IDSs) are essential elements when it comes to the protection of an ICT infrastructure. Intrusion detection systems (IDSs) are widespread systems able to passively or actively control intrusive activities in a defined host and network perimeter. Recently, different IDSs have been proposed by integrating various detection techniques, generic or adapted to a specific domain and to the nature of attacks operating on. In this paper survey was done on the various techniques of intrusion detection system where some of supervised and unsupervised intrusion detection techniques are discussed. Here methodology of various researchers is explained with their steps of working. Different types of attacks done by the intruders were also discussed.

Keywords- Anomaly Detection, ANN, Clustering, Genetic Algorithm, Intrusion Detection.

I. INTRODUCTION

Intrusion detection systems (IDSs) are one of the most important entities when it comes to Information and communications technology (ICT) infrastructure protection against cyber attacks. IDSs weaponries defenders with fundamental means to detect offensive events and consequently trigger optimal counteraction plans against them [1], [2]. In fact, the everlasting battle between defenders and attackers has taken the form of an ''arm race'', where both sides constantly upgrade their arsenals in order to prevail against each other.

The emergence of new attacks spurs the academia and industry to investigate for novel methodologies which are able to closely monitor this race and adapt rapidly to the changes in the field. In principle, IDSs fall into two major categories, namely Anomaly Detection Systems and Misuse Detection Systems. The former regulate their detection engine to identify as intrusive incidents those that exhibit deviations from a predefined normal behavioral profile. This kind of IDSs are able to identify previously unseen attacks, but are known to produce high false alarm rates, rendering them a questionable solution especially for complex infrastructures, where the standardization of the normal profile is challenging. On the other hand, misuse IDSs relies on known signatures trying to designate traffic instances to legitimate or attack traffic classes. This kind of IDS lacks the ability of identifying new attack patterns or deviations from known ones, and their performance depends on the freshness of the signatures database. Hence, the IDSs administrator needs to put significant effort to keep the misuse detection model up to date.

If we additionally consider the fact that the protected environment is a dynamic ecosystem where new devices and/or services may appear or leave the network at any moment (e.g., the Internet of Things), it becomes clear that the adaptability issue becomes a burden on administrators' shoulders. This burden becomes even heavier as the growth of communication networks pushes IDSs into the big data era, where the increased volume of the transmitted data surpasses the limits of human processing capabilities.

So paper main motivation on the intrusion detection techniques as per the types of attacks or network. Here main motive of this paper to brief various techniques used by the authors / researchers in previous years for increasing the detection rate while

reducing the false alarm rate. Paper highlights the limitations and problems still present in the previous approaches of researcher. So new work will focus to overcome those problems. Whole paper is organized into few sections where second section gives summary of the work done by different author in this field of IDS, here comparison table of researcher approach was shown with their limitations. Third section explained various techniques for intrusion detection include supervised and unsupervised both. Finally types of attacks were brief in fourth section.

II. TECHNIQUES OF IDS

1. Misuse identification [9]

Utilizes examples of the definitely known attacks or the framework's delicate spots to coordinate and distinguish intrusions. For example, in the event that somebody tries to figure a secret key, a mark manage for this sort of conduct could be that 'excessively numerous fizzled login attempted in indicated time' and occasion of his compose may bring about rising an alarm. Misuse recognition observed to be not proficient against the not known attacks that have no coordinated guidelines or examples yet.

2. Anomaly location [9]

banners watched exercises that withdraw impressively from the customary use profiles as inconsistencies, that is, conceivable intrusions. For example a profile of a client may contain the found the middle value of frequencies of some framework summons in his or her logging sessions. Also, for a logging session that is being observed on the off chance that it has altogether lower or higher frequencies an abnormality ready will be raised. Anomaly identification is a successful strategy for discovering perfect or not referred to attacks as the learning is never required with respect to the intrusion attacks. Be that as it may, in the meantime it tends to raise a larger number of cautions than misuse identification since whatever occasion occurs in a session, ordinary or unusual conduct, if their frequencies are significantly separate among the threshold found the middle value of frequencies of the client it will raise an alarm.

3. Supervised learning

In conceptual terms the supervised learning can be seen as a teacher having knowledge of the environment derived from input-output examples. The teacher provide consultancy to the neural network telling it what is normal and abnormal traffic pattern, in the sense of what is classified as malicious and non-malicious.Basically the supervised learning operates as a portion of network connection is to be analyzed and labeled with the help of the teacher [2, 11]. Afterwards the labeled training data is used by the learning algorithm to generalize the rules. Finally the classifier uses the generated rules to classify new network connections and gives alert if a connection is classified to be malicious.

4. Unsupervised learning

Unlike the supervised learning, unsupervised learning does not have a teacher to tell what is a 'good' or 'bad' connection. It has the ability to learn from unlabeled data and create new classes automatically. In with the use of a clustering algorithm it is illustrated how unsupervised learning operates [8]. First, the training data is clustered using the clustering algorithm. Second, the clustered weight vectors can be labeled by a given labeling process, for example by selecting a sample group of the data from a cluster and label that cluster center with the major type of the sample. Finally, the labeled weight vectors can be used to classify the network connections.

5. Compare Supervised and unsupervised

Monitoring network traffic shows a lot of activities in the sense of different data packets being sent forth and back constantly. Of course the magnitude of this activity depends on the network monitored. If a network of a home computer, which is only used for e-mail checking and internet browsing, is monitored, it will show little traffic activity, but if a busy server on the Internet is monitored, it will show a great deal of activity. Intrusion detection systems should be able to monitor and categorize (or label) traffic at the same time regardless of the size of the traffic activity. But in networks with large traffic rate, labeling data becomes a tough task. It is time-consuming and normally only the small part of provided data may be labeled [7]. At packet level it may be impossible to unambiguously assign label to data. On the other hand in real application one can never be sure that a set of labeled data examples are enough to cover all possible attacks [10]. These considerations are important and should be taken into account when choosing network paradigm.

6. Genetic Algorithm

Genetic algorithms are unsupervised search procedures often used for optimization problems. Genetic algorithm is based on the principles of evolution and natural selection of chromosomes. An initial population of chromosomes is generated

randomly where each chromosome represents a possible solution to the problem (a set of parameters).The evaluation function is used to calculate the "goodness" of each chromosome. In evaluation, two operators, crossover and mutation, are used to generate the new population or rules. Then, the best individual or chromosome is selected as the final result once the optimization criteria in met

III. RELATED WORK

Moukhafi et al. [12] combined a hybrid genetic algorithm (GA) and an SVM with PSO for feature subset selection in their proposed intrusion detection system. This system was successful in differentiating DoS attacks from other types of attack with an accuracy of almost 100%; however, it could not discriminate normal class signals from other types of attacks with reasonable accuracy.

Vajayanand et al. [13] tried to improve classification accuracy by proposing a hybrid feature-selection technique based on a GA and mutual information (MI) for an SVM-based classifier. They also proved (by illustrating their experimental results) that an SVM-based classifier is successful in achieving better performance than an artificial neural network (ANN). The highest accuracy achieved in their experiment was 96% when the classifier was trained with 400 samples. The results showed that utilizing both the GA and MI could need as few as three informative features to obtain these results. However, considering the battery and computation-cost limitations of IoT devices, this scheme does not seem like a promising solution.

Kabir et al. [14] proposed an optimum allocation-based least square SVM (OA-LS-SVM) for intrusion detection systems. This technique first combines the training and testing datasets. Then, an optimal allocation (OA) scheme determines the volume of training and testing sets. Later, it selects representative samples directly from training and testing datasets for the classifier. Although this paper presented some interestingly satisfactory results, it can miss some important information or features in the dataset owing to their limiting the training dataset to samples having a specific relation with a representative sample. Furthermore, obtaining all the samples from training and testing datasets is a challenge difficult to overcome.

Yu et al. [15] designed an IDS model by stacking dilated convolution auto encoders (DCAEs) for learning features representations from unlabeled data. In their experiments, the authors tested the generalization ability of their detection model by testing it with previously unknown attacks. Even though the authors pose their trained model against new attacks, they do not proceed to any automated retraining method. Rather they aim to exclusively test the generalization ability of the learned features employed to statically train the IDS. Additionally, the authors approach the problem as a binary one (normal/attack), while in our classification case we deliver a multi-classification method.

Chuan long Yin [16] "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks" this paper, investigates how to model an intrusion recognition framework dependent on profound learning, and propose a profound learning approach for intrusion detection utilizing intermittent neural systems (RNN-IDS). In addition, this paper has considered the execution of the model in twofold arrangement and multiclass characterization, and the quantity of neurons and distinctive learning rate impacts on the execution of the proposed model.

In [20] introduces a deep learning solution for NIDSs. The authors utilize the self-taught learning methodology exclusively as an unsupervised feature learning method for supporting statically trained IDS. Instead, our solution provides a more powerful setup of the STL in conjunction with MAPE-K methodology to deliver a deep learning methodology for adaptive IDSs. In fact, contrary to [20], author pass the IDS through environmental changes to prove that our approach is able to generate knowledge out of unknown environments.

Ma et al. [21] propose an IDS algorithm based on Spectral Clustering (SC) [19. Through SC the proposed method is able to identify cluster centers that divide a raw dataset into data clusters with similar features. Those data clusters are fed as training data into DNN's of multiple layers. The algorithm trains as many DNN's as the clusters identified by the SC and aggregates the final result in an ensemble way. However, the proposed deep learning approach does not provide any kind of addictiveness to the system.

IV. NETWORK ATTACKS

A more profound comprehension of PC attacks is required to distinguish intrusion and security dangers. An ordinary PC attack can be summed up into a five stage approach.

1. **Reconnaissance:** The assailant gathers abnormal state data of the framework.
2. **Scanning:** Using the data gathered in the past advance, the assailant distinguishes potential vulnerabilities in the framework and gathers point by point data about the system, for example, arrange topology, ports utilized and firewall rules.
3. **Gaining Access:** There are two approaches to access the framework relying on the authenticity of the client. An approved client misuses the provisos in the working framework or different applications running in the framework. An ill-conceived client makes utilization of the system to join the framework. DoS (Denial of Service) are one such case in which the web server is shelled with different demands all the while that it in the long run crashes.
4. **Maintaining Access:** The invader approaches the framework and tries to extricate data from the framework and hold control.
5. **Covering Tracks:** with a specific end goal to practice nonstop control over the framework, the invader alters framework logs and other pertinent data to guarantee that there is no hint of contradiction in the security framework.
6. The easy and common criterion for describing all computer network attacks and intrusions in the respective literature is to the attack types [1]. In this chapter, this work categorizes all computer attacks into the following classes:
7. **Denial of Service (DoS) attacks:** Denial of Service (DoS) attacks mainly attempt to "shutdown a whole network, computer system, any process or restrict the services to authorized users" [2]. There are mainly two types of Denial of Service (DoS) networking attacks
8. Operating system attacks
9. Network attacks
10. In denial of service attack, operating system attacks targets bugs in specific operating system and then can be fixed with patch by patch, on the other hand networking attacks exploits internal limitation of particular networking protocols and specific infrastructure.
11. **SSH:** Secure Shell is a protocol that provides authentication, encryption and data integrity to secure network communications. Implementations of Secure Shell offer the following capabilities: a secure command- shell, secure file transfer, and remote access to a variety of TCP/IP applications via a secure tunnel. Secure Shell client and server applications are widely available for most popular operating systems. The secure shell protocol allows users to log in remote terminals in a secure fashion. It does this by performing authentication using a passphrase and a public keying, and subsequently encrypts all information transmitted or received, guaranteeing its confidentiality and integrity.
12. **Probing: (surveillance, scanning): Probing (surveillance, scanning)** attacks scan the networks to identify valid IP addresses and to collect information about them (e.g. what services they offer, operating system used). Very often, this information provides a tacker with the list of potential vulnerabilities that can later be used to perform an attack against selected machines and services. These attacks use known vulnerabilities such as buffer overflows [8] and weak security points for breaking into the system and gaining privileged access to hosts. Depending upon the source of the attack (outside attack vs. inside attack), the compromises can be further split into the following two categories:
13. **R2L (Remote to Local):** Attacks, where an attacker who has the ability to send packets to a machine over a network (but does not have an account on that machine), gains access (either as a user or as the root) to the machine. In most R2L attacks, the attacker breaks into the computer system via the Internet. Typical examples of R2L attacks include guessing passwords (e.g. guest and dictionary attacks) and gaining access to computers by exploiting software vulnerability (e.g. phf attack, which exploits the vulnerability of the phf program that allows remote users to run arbitrary commands on the server).
14. **U2R:** (User to Root): Attacks, where an attacker who has an account on a computer system is able to misuse/elevate her or his privileges by exploiting vulnerability in computer mechanisms, a bug in the operating system or in a program that is installed on the system. Unlike R2L attacks, where the hacker breaks into the system from the outside, in U2R compromise, the local user/attacker is already in the system and typically becomes a root or a user with higher privileges.

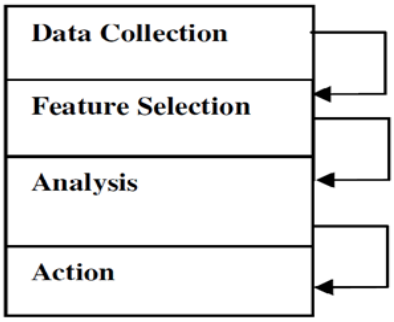

Fig.1 U2R.

The most common U2R attack is buffer overflow, in which the attacker exploits the programming error and attempts to store more data into a buffer that is located on an execution stack.

V. FUNCTION STEPS OF IDS

Figure 2 Key functionalities of an Intrusion Detection System [14].

1. **Data Collection:** This module collects the real world data in system and provided it as the input to the IDS. In case of network based IDS, packets of data in the transmission are get collected and within the host dependent IDS, details such as disk usage, system process, call stack etc are logged.
2. **Feature Selection:** Large amount of data is available in the network and a subset of it is usually analyzed for the intrusion. As an example, Internet-Protocol (IP) address of source and the target system, protocol type, header length and size could be studied for possible intrusion.
3. **Analysis:** This module defines the method that is used to analyze data. One approach which is the use of the rule dependent IDS in which the incoming-traffic is checked against pre-defined signature or pattern. Another method is the use of anomaly dependent IDS in which the behavior of the system is analyzed and mathematical models are employed.
4. **Action:** This scheme explains how a system should act to the possible attacks within a system. It can either inform the system administrator with entire data which is needed via icons of alarm/email or it can play an active part in the system by dropping packets so that it does not enter the system or closing ports [1]. Distributed IDs It gathers audit data from multiple hosts and possibly the network that connects the host. It detects attacks involving multiple hosts. Network based IDs It uses network traffic as the audit data source, relieving the burden on the hosts that usually provide normal computing services. It detects attack from network. NIDS uses a passive interface to capture network packets for analyzing. NIDS sensors placed around the globe can be configured to report back to a central site, enabling a small team of security experts to support a large enterprise. NIDS systems scale well for network protection because the number of actual workstations, servers, or user systems on the network is not critical – the amount of traffic is what matters .Provide better security against DOS attacks.

VI. CONCLUSION

Different techniques are discussed in this paper to support the security of an organization against threats or attacks. On the other side attackers are discovering new techniques and ways to break these security policies. IDS provides the facility to detect and prevent from attacks by inheriting multiple approaches like secure mobile agent, virtual machine; high throughput string matching, multilayer and distributed approach provide greater and strongest security against multiple attacks. This paper gives a deep understanding of various methods adopt by the researcher for classifying the intrusion into their class of attacks. Here discussion of different attacks was also mention. Techniques for learning of patterns of intrusion detection were detailed and it was obtained that use of neural network is good for multiclass partition. As research is continuous process so it is desired to develop a algorithm which can dynamically identify intruder behavior.

REFERENCES

[1]. Shaohua Teng, Naiqi Wu, Senior, Haibin Zhu, Senior, Luyao Teng, and Wei Zhang. "SVM-DT-Based Adaptive and Collaborative Intrusion Detection". IEEE/CAA JOURNAL OF AUTOMATICA SINICA, VOL. 5, NO. 1, JANUARY 2018.

[2]. Kai Peng, Victor C.M. Leung, Qingjia Huang. "Clustering Approach Based on Mini Batch Kmeans for Intrusion Detection System over Big Data". IEEE Transaction 2169-3536 © 2017. .

[3]. Aljurayban, N.S Emam, A. (21-23 March 2015). Framework for Cloud Intrusion Detection System Service. Web Applications and Networking (WSWAN), 2015 2nd World Symposium on, p1-5

[4]. Barolli Leonard, Elmazi, Donald; Ishitaki, Oda, Tetsuya; Taro; Yi Liu, Uchida, Kazunori. (24-27 March 2015). Application of Neural Networks for Intrusion Detection in Tor Networks. Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on, p67-72.

[5]. Koushal Kumar, Jaspreet Singh Batth "Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, September 2016

[6]. R. Karthik, Dr.S.Veni, Dr.B.L.Shivakumar "Improved Extreme Learning Machine (IELM) Classifier For Intrusion Detection System" International Journal of Engineering Trends and Technology (IJETT) – Volume-41 Number-2 - November 2016

[7]. Premansu sekhara rath, 2manisha mohanty, 3silva acharya, 4monica aich "optimization of ids algorithms using data mining technique" International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-4, Issue-3, Mar.-2016

[8]. Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey "Intrusion Detection Using Data Mining Techniques", 978-1-4244-5651-2/10/$26.00 ©2010 IEEE

[9]. YU-XIN MENG," The Practice on Using Machine Learning For Network Anomaly Intrusion Detection" Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, 978-1-4577-0308-9/11/$26.00 ©2011 IEEE

[10]. Liu Hui, CAO Yonghui "Research Intrusion Detection Techniques from the Perspective of Machine Learning" 2010 Second International Conference on Multimedia and Information Technology 978-0-7695-4008-5/10 $26.00 © 2010 IEEE

[11]. Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach For Intrusion Detection Using Recurrent Neural Networks" current version November 7, 2017. Digital Object Identifier 10.1109/ACCESS.2017.2762418.

[12]. [26] M. Moukhafi, K. El Yassini, and S. Bri, ''A novel hybrid GA and SVM with PSO feature selection for intrusion detection system,'' Int. J. Adv. Sci. Res. Eng., vol. 4, pp. 129–134, May 2018.

[13]. [27] R. Vijayanand, D. Devaraj, and B. Kannapiran, ''A novel intrusion detection system for wireless mesh network with hybrid feature selection technique based on GA and MI,'' J. Intell. Fuzzy Syst., vol. 34, no. 3, pp. 1243–1250, 2018.

[14]. [28] E. Kabir, J. Hu, H. Wang, and G. Zhuo, ''A novel statistical technique for intrusion detection systems,'' Future Gener. Comput. Syst., vol. 79, pp. 303–318, Feb. 2018.

[15]. [26] Y. Yu, J. Long, and Z. Cai, ''Network intrusion detection through stacking dilated convolutional autoencoders,'' Secur. Commun. Netw., vol. 2017, Nov. 2017, Art. no. 4184196.

[16]. Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks" current version November 7, 2017.

[17]. [20] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, ''A deep learning approach for network intrusion detection system,'' in Proc. 9th EAI Int. Conf. BioInspired Inf. Commun. Technol. (Formerly BIONETICS), 2016, pp. 21–26.

[18]. [21] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen, ''A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks,'' Sensors, vol. 16, no. 10, p. 1701, 2016.

[19]. U. von Luxburg, ''A tutorial on spectral clustering,'' Statist. Comput., vol. 17, no. 4, pp. 395–416, Dec. 2007.