ISSN: 2057-5688

A SURVEY - AN APPROACH ON THE STUDY OF INTRUSION DETECTION MECHANISMS

[1]BATHULA SREEJA, [2]Dr.D. RAMESH

[1]M. Tech Scholar, [2]Professor, Dept. Of CSE,

JNTUHU COLLEGE OF ENGINEERING, JAGTIAL, T.S., INDIA

Abstract: Intrusion Detection System (IDS) is an application program that acts as a security system that acts as a security layer for the infrastructure. It analyzes the network to determine the possibility of malicious activities. Increased use of the internet raises questions of how to safeguard digital data with a security approach. In the past, IDS technology has grown exponentially to keep pace with the rapid advancements in cybercrime. Today, hackers employ a variety of methods of attack to get the personal information of our computers. There are a variety of intrusion detection techniques strategies, algorithms and methods will provide a defense against these types of attacks. This primary goal of this article is to provide a thorough investigation for intrusion detection. the history and life-cycle, techniques for detection of intrusion, the kinds of attacks, various methods and tools, as well as the challenges and solutions.

I. INTRODUCTION:

Today, with the spread internet and the growing popularity of Internet and the online processes that require security It has become an essential requirement to offer the security for networks. There are many sources of threats that can be found in software, particularly because operating systems and software grows more effective and larger in size. The intruders that do not have the right to access this data could steal private and valuable data belonging to users of networks [1]. Firewalls are either software or hardware devices that are placed the middle of two or more networks to prevent attacks by isolating the networks with the rules and guidelines set for them. It is evident that firewalls don't suffice to protect a network fully since attacks that originate from outside the network are stopped while inside attacks aren't. This is why intrusion detection systems (IDSs) are responsible [2]. IDSs are utilized to prevent attacks or recover from them, at a minimum cost, or analyze security issues to prevent them from being repeated. IDSs gather information on a computer or computer network to identify misuses and attacks within the network. Some IDSs simply analyze attacks, while some attempt to stop the attack during the time of the attack. Three kinds of data are utilized by IDSs. These include networks traffic information, system-level test data, and system status files [2,3].

II. LITERATURE SURVEY

This paper outlines the application of supervised machine algorithms to aid in detection of anomalies. The algorithms employed include SVM, J decision tree and decision table as well as the naive Bayes. The goal is to figure out the source of the irregularity in normal flow of traffic. The model is based on information and using machine learning algorithms, this data is then trained into a model of prediction [4]. Then, after training, data is collected to ensure whether the model that was predicted is accurate. The data that is used by the algorithm must be comprehended to the algorithm. This is why we choose the input that contains instances, records, or examples. They have specific attributes that distinguish them. They are referred to as feature vectors [5]. Certain instances also have class types. The algorithm that requires each input to be a feature class type or vector it is referred to as the supervised algorithm. Some of which require semi-supervised. which does not require any and is trained using a patterns or similarity in the input are referred to as unsupervised. The dataset that was used by the author is KDD dataset. The 492021 instances used are for the training program and 311029 are to test the data [6]. In the model presented in this paper, FPR for DOS attacks is very low, J48 has the highest precision, and SVM is a very accurate model with zero FPR for attacks using u2r. 2. In this paper the authors have created an algorithm that uses the supervised machine algorithm for the detection of anomalies. The algorithms utilized in this study include C4.5 decision tree as well as Naive Bayes classification. It is proven that c4.5 is accurate when taking decisions as in the paper. The proposed model in this paper consists of three components that are processing, pre-processing and post processing. the aid of C4.5 decisions, the model that is trained in accordance with the inputs given [7]. Learn the chosen algorithms for machine learning. Based on the training data, classification of the recorded data to distinguish between normal and abnormal activities is carried out. Two parameters are extracted: both X and Y. In the case of x, it is the protocol that is used: TCP, IP, UDP and Y is the name given to attack: probe, U2R, etc. [8].

Processing Naive Bayes classification method calls for K (k-1)/2 2-class Naive Bayes classification techniques, where each is trained with information taken from 2 classes. The binary Tree Naive Bayes is currently being utilized to detect IDS. Based on the features of various intrusion detection types they have designed four Naive Bayes classifiers that can identify the five states that are regular condition (NS) along with the other four states: intrusion state, or attack states: Denial of Service (DoS) Remote to user (R2L) Users to Root (U2R) and Probing. Post processing: minimize false positive rates [9]. The N voting algorithm is employed to group the attacks by IP addresses, both from the origin and destination. After grouping, a check is made to determine if the attack is normal or not. If 3 out of 5 are considered to be attacks then the IP addresses are classified as malicious. This model has an accuracy of as high as 99.48 percent. The paper is a discussion of SDN (software development networks) built on NIDS. The paper provides a brief overview of how various machine learning algorithms are employed in NIDS.

It segregates the algorithms in 3 types:

Supervised: where data is labeled algorithms like random forests and SVM are used to classify of data that is unsupervised in which the data is created as an element that is structural and the unidentified data is predicted. algorithms like self-organizing maps or clustering are employed. However, clustering is recommended to be avoided as it could produce inaccurate results when the center isn't appropriate [10].

Semi-supervised: If there is a large amount of data not available, these algorithms are utilized: Spectral Graph Transducer and Gaussian Fields approach, used to identify malicious behavior, and a semi-regulated method of clustering MPCK means to increase the effectiveness that the detector system can achieve. The methods used for this study are the following: 1. Decision tree algorithm for feature selection and random forests to determine classification 2.PCA to select features and SVM as a classifier. 3.Softmax regression as a classifier and an additional encoder to further reduce and accuracy of around 92 percent. According to this model, the decision tree has high accuracy and has low false alarm [11].

In this paper What is SDN and how it can be employed is discussed. SDN is an approach to improve security. It is utilized with several tools. Most popular tool being open flow. SDN is used in different simulation tools, such as the ns2, ns3, or omnetand ++. The issues discussed in this paper are that SDN could be vulnerable to attacks like DOS, DDOS and so on. A majority of the papers are focused on KDD dataset that is old and, despite having a significant amount of information, but the data is out of date and a new dataset's must be utilized. The determination of the optimal quantity of parameter parameters to model was an obstacle [12].

1) Tchakoucht TA, Ezziyyani M on "Building an efficient intrusion detection system for high-speed networks probes and DoS attack detect." Procedia Comput Sci. 2018. The solution proposed was light and was suitable for detecting attacks on high-speed networks. Of the 41 features utilized in KDD'99, the top features were utilized to improve effectiveness and accuracy in the design. Six methods of selecting features were utilized. A resampled version KDD'99 utilized to test the system. The results showed an excellent detection accuracy of approximately 99.6 percent, as well as false-positive rates of about 0.3 percent for DoS attacks that use C4.5. A precision of 99.8 percent and a false positive rate of approximately 2.7 percent were found for Probe attacks made using Naive Bayes [13]. It was discovered that processing time could be reduced when compared with the best-chosen feature subset. The feature subset suggested is therefore recommended for use in high-speed networks. It has 19 features to detect probes and 9 features to aid in DoS detection.

2) Sahasrabuddhe A, et al. Research on "Intrusion detection technique employing techniques for data mining" In Int Rej Eng Tech. 2017. This paper discusses the various types of SQL injection attacks as well as the various data mining techniques used for intrusion detection. The paper then discusses an algorithm that uses it's Naive Bayes classifier for detecting SQL Injection attacks. It explains the different types of SQL injection attacks that may be executed on a computer system, such as Tautologies, Blind Injection, Piggy-Backed Queries Et al. It examines the necessity of an intrusion-detection system in order to identify SQL Injection attacks since it is among the oldest cyber-attacks. The various methods of data mining used are reviewed in a brief manner and then converge on the Naive Bayes classifier, which is utilized in the model proposed.

3) Ferhat K. Sevcan An on "Controlling fraud through Spark's machine-learning libraries" in the journal Int J. Appl. Math. Electron Comput. 2018. This paper suggests the use of K-Means ' clustering algorithm to spot abnormal behavior of networks. A commonly utilized algorithms used in the field of data mining is K-Means clustering technique. These are algorithms that split data by itself into sub-clusters or smaller clusters. It puts statistically similar data within the same group. Apache Spark is an open-source unified analysis engine designed for big data processing, is being employed. They found ten anomalies out of four hundred thousand records in the KDD'99 10 percent data set with the proposed method.

4) Lee Y on "Toward scalable internet traffic measurement and analysis using Hadoop" In ACM SIGCOMM. Commun Rev 2013. A solution that is based in Apache Hadoop for packet analysis (NetFlow trace analysis) is suggested. A performance in the range of 14 Gbps was obtained using the 200-node Hadoop testbed that could handle 5TB of files data. The proposed solution uses the binary format used to obtain trace data concurrently, MapReduce algorithm to aid in Netflow, TCP, IP and HTTP analysis, as well as a Hive-based system that can be used to simplify queries. The solution proposed includes numerous tools that build on this, including standard five-tuple flows statistics TCP Re-transmission stats along with DDoS analysis. A variety of experiments were conducted to test accuracy and Scalability, as well as a

thorough examination of CoralReef and the RIPE's Pcap. The paper concludes with the suggestion that this approach could provide the ability to scale out Hadoop to handle the ever-growing amount of traffic data.

5) Peng K. Et al. On "Intrusion detection system built on decision tree and Big Data in fog environment" 2018. This paper examines Intrusion Detection System in the viewpoint of Fog Computing. Fog computing Fog nodes, being near the users' devices are not able to perform computing tasks and therefore could face certain security issues. This system Fog nodes could be destroyed through traditional security threats to networks and therefore an intrusion detection System (IDS) could provide a security option that can be utilized within Fog [14]. Fog environment.

6) Manzoor MA, Morgan Y on "Real-time support vector machine-based network intrusion detection system employing Apache Storm" in IEEE 7th Annual IEMCON in 2016. IDS is a crucial component to manage the security of a network. It acts as a security device for networks. This paper proposes the high-speed Intrusion Detection System which can operate in real-time. The system presented within the research paper a Support Vector Machine based Intrusion detection system for networks which employs the KDD99 Data-Set. The proposed system is designed to manage huge quantities of streaming data i.e., Big Data. Results of an study conducted with the proposed system have shown that it's suitable for processing stream-processing information from networks for intrusion detection that is highly accurate [15].

7) Dahiya P, Srivastava DK on "Network intrusion detection on a massive dataset's by using Spark" In Procedia Comput Sci. 2018. The model that is proposed in this paper is effective and quick while being efficient in detecting intrusions. The UNSW data set NB-15, which is both large and small datasets was used for the evaluation of the performance of the model presented in this paper. Different algorithms for feature extraction and classification were used used to build the model. Utilizing CNN has led to the realization that accuracy in the small datasets was affected, but it also reduced the time required to train the model was also decreased. In contrast, LDA improves the accuracy, but with a cost: an increasing the amount of time needed

for the training of the model with both large and small datasets. The most effective method was identified to be the Decision Tree algorithm and LDA was identified as the best method for feature reduction. The Intrusion detection technique proved to be quicker and more efficient with an algorithm called random forest as well as LDA. Accuracy wise, the Random Tree algorithm was found to be superior to other Algorithms. It was able to categorize the data in a precise manner as normal traffic, and also into different attacks. Methods for feature reduction when employed have been found to improve efficiency of model. After conducting the tests above with Apache Spark, it can be conclusively concluded that Apache Spark method is better than the other methods, and is faster and more efficient [16].

8) Wang H, Xiao Y, Long Y on "Research of intrusion detection algorithms that is based on the parallel SVM and Spark." In the the 7th Annual IEEE (ICEIEC) in 2017. SVM is a tried and tested tool that is powerful in the field of the classification of data and regression [17]. Some of the SVM algorithms that were developed to date include: SMO (Sequential Minimal Optimization) Light SVM, libSVM, and others. While they've been shown that they are feasible to some extent, none can be employed to handle large amounts of data at a larger scale. When the volume of the sample slowly increases the dramatic increase in the amount of time and memory needed to train models employing an SVM algorithm. Also, it is a well-known fact that the single SVM Algorithm cannot effectively deal with data sets that are at the higher end in regards to size. To overcome this issue of optimizing SVM to manage large quantities of data it was decided to make the algorithm more parallel. This technique reduces the size of data by using the strategy divided and conquer. This paper suggests in its the implementation a combination of the parallelized implementation for SVM and PCA reduction in dimensionality, using Bagging on Apache spark. Optimization was achieved by removing any bottleneck i.e., difficulties in handling massive data sets, while also enhancing the efficiency of network information with high dimensions data.

III. INTRUSION DETECTION AND MACHINE LEARNING

The concept behind applying methods of machine learning for the detection of intrusions is to create the model using the data used to train. The data set is an assortment of data instances each one of which is defined using a list of features (features) and their labels. The attributes may be of various types, such as continuous or categorical. The attributes' nature determines the effectiveness of anomaly detection methods. For instance, distance-based techniques were initially designed to work with continuous attributes and typically do not deliver satisfying results with categorical attributes. The labels that are associated with data instances usually take forms of binary numbers i.e., normal and abnormal. However certain researchers have used different kinds of attacks, including DoS U2R, R2L, and Probe instead of the label of anomalous. In this way, learning techniques are capable of providing more information about the different types of anomalies. However, the results of experiments show that the current methods of learning aren't precise enough to detect the nature of anomalies. Because labeling is typically done by humans and obtaining a precise dataset that is representative of all kinds of behavior can be quite costly. In turn, based on the quality of the labels there are three operating modes described for anomaly detection methods such as Supervised Learning, Unsupervised Learning, and Semi supervised Learning

Machine learning methods are based on the creation of an implicit or explicit model. One distinctive feature of these methods is the need to label data to build the model's behavior this process puts huge requirements on the resources. In many instances the applicability of machine learning techniques is comparable with the one of the statistical methods, but the latter is focused on developing an improved model performance based on previous data. Thus, machine learning in IDS is able to modify its strategy of execution as it receives new information. This could make it beneficial to employ the same strategies for every situation.

Pros: -1) Flexibility and adaptability capture of interdependencies.

Cons: -1) High depended on the assumption about the behavior accepted into the system.

IV. INTRUSION DETECTION ATTACKS

A. Denial-of-Service (DOS) Attacks

There are two major kinds of denial-of-service (DoS) attacks that are flooding and flaw exploits. Flooding attacks are often simple to execute. For instance, you can start a DoS attack with the command ping. It will send the victim a massive amount of packets of ping. If the attacker is able to access more speed than that of the target, it will quickly and efficiently overburden the victim. For instance, an SYN flood attack will send TCP/SYN packets with a fake source address to the victim. The victim will be prompted to open up half open TCP connections. The attacker will then send a TCPSYN/ACK message and then wait for an ACK to be returned. Because the ACK does not arrive and the victim will eventually run out of resources and wait for ACKs from a host that is not there.

B. Eavesdropping Attacks

It's the plan to interfere with communications by the attacker. The attack could be carried out via phone lines or via email.

C. Spoofing Attacks

The attacker is disguised as a user to alter the data and gain access to unlawful events happening within the network. IP Spoofing is a common scenario where the system connects with a trusted person and allows permission to an attack.

D. Intrusion attacks or User to Root Attack (U2R)

An intruder attempts to gain access to the system or traverse the network. Buffer overflow is an attack that typically happens when a web server receives more information than it is able to handle. This results in the loss of information.

E. Logon Abuse Attacks

An attack that abuses logons could bypass security and authentication mechanism and provide a user with additional advantages.

F. Application-Level Attacks

The attacker targets weaknesses of the application layer.

V. COMPARISION OF DIFFERENT TOOLS AND TECHNOLOGY

The research has been aided with grants provided by the Natural Sciences and Engineering Research Council of Canada as well as Dalhousie University Electronic Commerce Executive

Committee found that Dragon did better than, or scored as high than the other three open-source solutions in all five categories that deal with insider traffic. Except for one of the categories, it spotted at the least 50 percent or more of the attacks within each. Furthermore, several instances of attacks detected with confidence levels of two. The Denial-of-Service category snared only one of eight attacks. 73 percent of the attack-related information contained in log entries was at the level 2. Snort did well with both the DOS as well as Probe categories, catching nearly 50% of attacks in each, but did not perform as well in the three other categories. The majority of the data contained in its attacks related data was at confidence levels 2. This study showed that when certain traditional attacks are employed, no system detected them. It also proved the fact that Intrusion Detection Systems could not be responsible for attacks that originate from the misuse of features that are perfectly legitimate. There are still flaws in the manner Operating Systems are designed and constructed. More research is needed to determine the effectiveness of these devices in the face of attacks, and is specifically tailored to destroy those intrusion detection mechanisms. School of Future Studies & Planning found that there are a variety of tools available today that can help businesses fight the inevitable system and network attack. The use of IPS as well as IDS technologies are just two of the numerous resources which can be utilized to improve control and visibility in a computing environment for corporate use. IDS and IPS are designed to serve as an infrastructure of technology that can meet the requirements of tracking and monitoring network attacks, and detect them through the logs of IDS systems and preventing the action of IPS systems. If the host is a victim of sensitive systems, confidential information and strict compliance requirements It is a good idea option to utilize IDS and IPS or both in networks [18].

International Journal of Computer Science and Information Technologies demonstrated that both firewalls as well as intrusion detection system need to be upgraded to ensure a reliable security of a network. They're not as reliable (especially with regard the false negatives as well as false positives) and aren't easy to administer. In order to ensure an effective computer-based security system it is highly recommended to utilize a mixture of different types of intrusion detection systems. These technologies will have to be improved in the near future due to the ever-growing security requirements of organizations and advances in technology that allow more effective operations detection and prevention systems. This paper offered a fresh method of looking at network research that includes different types of firewalls and kinds of intrusion detection, and which are essential comprehensive, comprehensive, and mutually exclusive to facilitate the honest comparison of firewalls, intrusion detection systems and assist in focusing on the research area that is devoted to new developments like Intrusion Prevention System.

VI. NEEDS AND CHALLENGES:

IDS technology is going through many improvements. Since the IDS implementation, it's clear that it is essential to an organization. IDS technology does not require humans to intervene. In the present, IDS technology provides some sort of automation such as notifying the administrator in the event of detection of an activity that is malicious and removing the dangerous connection for a pre-determined time period, and altering the access control list of a router to prevent the malicious connection. Each time an event happens, the IDS logs must be monitored. The logs should be monitored every day is necessary to analyze the actions that are identified through IDS. IDS over a long period of time. The sensor manager's ratio is something to be celebrated. It is crucial to establish the base policy prior to starting the IDS implementation to avoid false positives. IDS sensors can transmit a number of false positives results to the sensor, and the proportion could be insufficient.

VII. FUTURE SCOPE

IDS implementation is contingent on implementation's success. Planning is essential for the development and implementation phase. In the majority of cases, it is beneficial to use a hybrid approach of host and network IDS. The choice of which one to choose is dependent on companies. Network-based IDS is a good option for many companies due to its capability for monitoring of multiple platforms, and in addition, it does not require application to run onto a production system like hosts-based IDS. Certain organizations offer hybrid solutions. Therefore, the resources available are necessary to build a system prior to installing a sensor based on host [20]. This IDS technology is active rather than proactive. this technology relies with attack signatures. Signatures are described as a pattern of attacks, which was defined earlier. The signature database has to be updated each time a different type of attack is discovered and then they are corrected in the database. The frequency of updating the signatures differs between vendors [19].

VIII. CONCLUSION

The principal goal of this article is to give a general overview of the need and benefits of the intrusion detection systems. The paper offers a thorough analysis about the various types of IDS the life cycle, different areas, types of attacks and tools. IDS are increasingly essential to security today in the business and network users. IPS describes the preventive security measures. The phases of the lifecycle are created and the stages depicted. There are still issues to face. The methods of the detection of misuse and anomalies are clearly illustrated, and further techniques are available. The research will focus in a comparative analysis of popular data mining algorithms that are applied to IDS and improving IDS based on classification. IDS with the use of selective feedback methods

IX REFERENCES:

[1] E. M. Tchakoucht TA, "Building a fast intrusion detection system for high-speed-networks: probe and DoS detection," 2018.

[2] S. A, "Survey on intrusion detection system using data mining techniques," 2017.

[3] S. A. Ferhat K, " Big Data: controlling fraud by using machine learning libraries on Spark," 2018.

[4] L. Y, "Toward scalable internet traffic measurement and analysis with hadoop," 2013.

[5] P. K, "Intrusion detection system based on decision tree over Big Data in fog environment," 2018.

[6] M. Y. Manzoor MA, "Real-time support vector machine based network

intrusion detection system using Apache Storm," 2016.

[7] S. D. Dahiya P, "Network intrusion detection in big dataset using Spark," 2018.

[8] X. Y. L. Y. Wang H, "Research of intrusion detection algorithm based on parallel SVM on Spark," 2017.

[9] K. M. Gupta GP, "A framework for fast and efficient cyber security network intrusion detection using Apache Spark".

[10] A. A. M. A. H. J. Ahmed M, "A survey of network anomaly detection techniques," 2016.

[11] A. A. A. P. M. Mazhar Rathore, "Real time intrusion detection system for ultra-high-speed big data environments.".

[12] N.-U. K. T.-M. C. Sung-Hwan Ahn, "Big Data Analysis System Concept for Detecting Unknown Attacks".

[13] S. C. Y. N. a. M. Bakhtiarib, "Advanced Persistent Threat Detection Based On Network Traffic Noise Pattern and Analysis.".

[14] a. Q. Z. ,. G. H. ,. J. X. H. Wenying Feng, "Mining Network Data for Intrusion Detection through Combining SVM with Ant Colony".

[15] Z.-H. A. H. M. Kayacik HG, " Selecting features for intrusion detection: a feature relevance analysis on kdd99 intrusion detection datasets.," 2005.

[16] C. Xiang and S. M. Lim, “Design of multiple-level hybrid classifier for intrusion detection system,” in Workshop on Machine Learning for Signal Processing, 2005, pp. 117–122.

[17] B. Daniel, C. Julia, J. Sushil, P. Leonard, N. N. Wu,“ADAM: Detecting intrusions by data mining”, Proceedings of the 2001 IEEE, workshop on Information Assurance and Security, West Point, NY, 2001.

[18] Murali A, Rao M, “A Survey on Intrusion Detection Approaches,” Information and Communication Technologies, 2005. ICICT 2005. First International Conference on DOI: 10.1109/ICICT.2005.1598592, Year: 2005, pp: 233 – 240

[19] Mrutyunjaya Panda, and Manas Ranjan Patra “ NETWORK INTRUSION DETECTION USING NAÏVE BAYES ”, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007

[20] Li Xiangmei Qin Zhi “The Application of Hybrid Neural Network

Algorithms in Intrusion Detection System “978-1-4244-8694-6/11 ©2011 IEEE

[21] Xiangmei Li ,”Optimization of the Neural-NetworkBased Multiple Classifiers Intrusion Detection System “,978-1-4244-5143-2/10 ©2010 IEEE

[22] Naeem Seliya Taghi M. Khoshgoftaar, ”Active Learning with Neural Networks for Intrusion Detection”, IEEE IRI 2010, August 4-6, 2010, Las Vegas, Nevada, USA 978-1-4244-8099-9/10

[23] H.H. Hosmer, Security is fuzzy!: applying the fuzzy logic paradigm to the multipolicy paradigm, Proceedings of the 1992-1993 workshop on New security