International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 07 | July 2019 www.irjet.net p-ISSN: 2395-0072

A Review on Application of Data Mining Techniques for Intrusion Detection

E. Vijay Kumar[1], Dr. B. Indira Reddy[2]

[1]Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology, Telangana, India
[2]Professor, Dept. of Information Technology, Sreenidhi Institute of Science and Technology, Telangana, India

---

Abstract - *The fundamental idea of applying the data mining techniques for the intrusion detection is to determine the security infringements in information system. Enormous quantity of data is processed in the data mining and it finds covered up data as well as disregarded data. In order to find out the intrusion in a network, data mining has various methods and algorithms such as Neural networks, Classification tree, Clustering and Genetic Algorithms etc. Detecting attacks is a fundamental need in networks. This review paper is all about providing the information about different algorithms that were used to detect the intrusion in a network.*

Key Words: Intusion Detection, Data Mining Algorithms, k-means algorithm, naïve bayes classifiers, SVM.

1. INTRODUCTION

Static security components, for example, firewalls can give a sensible dimension of security, however unique security components like Intrusion Detection System ought to likewise be utilized.

The intrusion Detection Systems were classified as Misuse Detection and Anomaly detection.

1.1 Anomaly Detection

It alludes to recognize unusual demeanour of host or network. It really alludes to putting away highlights of client's standard practices snared on database, at that point its contrast client's present conduct and database. On the off chance that there happens a deviation gigantic enough, thereby it is acknowledged that the tested data is anomalous. The designs identified are known as anomalies. They are named as outliers.

1.2 Misuse Detection

Initially, it characterizes uncharacteristic system demeanour, and after that it characterizes any other demeanour, as regular demeanour. It assumes that uncharacteristic demeanour and bustle has an easy to characterize the model. It progresses within the low level of false alarm and rapid of detection. In any case, it is unable to find the non-pre-elected attacks within the highlight library, so the plenteous new assaults were not recognized. The intrusion detection systems (IDS) utilizing the conventional strategies are constrained in recognizing obscure interruption demeanour and upgrading the profile in genuine time and the upkeep will be moderate and overwhelming with these systems. In order to resolve these issues a few research had been carried out by the means of the data mining technology into the intrusion detection system, which encourages them to create accurate type of model for detection without human intervention from a enormous data by audit method.

1.3 Types of IDS

There are various types of Intrusion Detection Systems were there. They are

- Host Based IDS
- Network Based IDS
- Hybrid Intrusion Detection Systems

1.3.1 Host Based IDS

It is used to recognize the intrusion that usually happens on a solitary host system. It collects the audit information from audit trails that had been performed regularly on the host system and screens exercises such as astuteness of the system, host based network traffics, record changes and system logs. In the event that there's any illegal alter or activity is recognized, it makes cautious to the client through alarm or by a pop-up menu and notifies and provide suggestions to the central administration server. The development is obstructed by the Central administration server. The judgment ought to be founded based on the methodology provided by the local system.

1.3.2 Network Based IDS

It always monitors and examines network activity so that it can provide security to the network of a system from being attacked. It recognizes the malevolent activities such as network traffic attacks and denial-of-service attacks. Inorder to monitors packet traffic, Network based intrusion detection system employs the enormous sensors.

1.3.3 Hybrid Intrusion Detection Systems

The later improvement in intrusion detection is to aggregate the both sorts of host-based intrusion detection system and network-based intrusion detection to develop and implement the hybrid intrusion detection systems. Hybrid intrusion detection system can adapt to any situation and it provides the higher level of security. In order to analyze the whole network or the specific part of the network, It uses the combination of both IDS sensor areas and reports attacks.

1.4 Drawbacks of conventional IDS

Intrusion Detection Systems have numerous disadvantages because of the quick advancement of technology. The disadvantages of the conventional and existing intrusion detection systems are:

- Threshold detection
- False positives
- False negatives
- Updates lag

1.4.1 Threshold Detection

Several traits of system and client demeanour are communicated as far as checks, with some dimension set up as reasonable. Such demeanour traits can incorporate the quantity of records gotten to by a client in a given timeframe, the quantity of fizzled endeavours to login to the framework, the measure of CPU used by a procedure. Utilize this procedure in Anomaly Based Intrusion Detection System produce an abnormal state of false positives alerts.

1.4.2 False positives

If the ordinary attacks are erroneously delegated pernicious and treated appropriately then it is termed as false positives. In order to anticipate the false positive from happening repeatedly, the arrangements were made to examine and survey the IDS design.

1.4.3 False negatives

If the intrusion detection system unable to recognize the occurrence of the event or attack then it is known as false negatives.

1.4.4 Updates lag

The update lag is the fundamental problem that jumps out at Signature-Based Intrusion Detection System. At the end of the day, will be dependably a lag between the IDS's upgrades and existence of new thread.

2. LITERATURE SURVEY

“Sheng Yi Jang” has developed a intrusion detection system based on the clustering technique. Unlabeled datasets are used to obtain the clusters. As the datasets were increased the system will have the better performance.

“Sang Hyun Oh” projected an anomaly intrusion identification strategy by clustering typical client conduct. A discovery strategy which uses a clustering algorithm for demonstrating the customary demeanour of a client's exercises has been utilized. clustering disposes of the mistake brought about by statistical analysis. Thus the continuous exercises of the client are demonstrated more precisely than the statistical analysis.

“Inho Kang” had developed a strategy for intrusion detection system that utilizes a classification algorithm. Wide range of attacks were detected by the many systems, but this this system has one-class classification technique that upgrades the recognition execution for dangerous attacks.

“Mrudula Gudadhe”, have presented the idea of an inventive entirety enhanced decision tree to be utilized in intrusion detection system. The essential objective of this methodology is to join obvious guidelines to summarize a structure with the end goal that exhibition of the sole is improved.

“Juan Wang”, introduced an intrusion detection algorithm dependent on decision tree innovation. The study of his experiment states that the C4.5 decision tree is a successful method for the usage of decision tree and it gives very good precision. But his experiment is failed to modify the error rate.

“Xiang M.Y and Chang et.al”, have presented the idea of multiple level tree classifier for Intrusion detection system and upgraded the detection rate. The observation of this experiment demonstrates that the classifier can recognize known attacks yet it shows more false alarm rate for new assaults.

“Peddabachigiri S”, developed a intrusion detection system by combining decision tree and support vector machine (DTSVM) which are generally utilized as classification techniques and produces mammoth discovery rate.

“Mrutyunjaya panda”, had carried out the experiments by the means of various data mining techniques for intrusion detection system and they have observed that the accuracy and the performance of Naive bayes classifier is better when contrasted with the exactness accomplished in case of different decision tree algorithm however decision tree is unequivocally utilized algorithm for

recognizing obscure sorts of attacks when contrasted with Naive Bayes.

Eric, developed a intrusion detection system that will identify the outliers in the association of log for anomalies in the network traffic through k-means algorithm.

Eskin, chan et al., presented a intrusion detection system to identify the abnormalities in the network traffic by implementing fixed width and k-closest clustering strategy.

Panda and patra, contrasted various data mining algorithms with recognize the intrusion in network and inferred that their methodologies expanded the detection rates and decreased false alarm rate.

Tsai et al, presented the intrusion detection system to decrease the false alarm rates by implementing k-means algorithm and SVM.

JingTao Yao, developed the intrusion detection system by utilizing the enhanced SVM. The experimental observations states that the performance of the system is improved in precision when compared to conventional SVM methods.

In order to analyze the KDD dataset, an intrusion detection system was designed by the Mohjeran. This system uses the fuzzy logic along with the neural networks to improve the precision.

Manisha kansra presented a intrusion detection system by utilizing the tools like snort, WEKA. In this experiment J48graft and naive bayes algorithm was applied in each tool to analyze the KDD dataset and compare the results.

Balaraman Ravindran had developed the hybrid intrusion detection system that will work in two phases. In the first phase it detects the anomalies by using the probabilistic classifier and in the second phase, the system reduces the attack on IP addresses by utilizing the Hidden Markov Model.

"Sivaranjani S" had proposed the intrusion detection system by applying the Correlation Feature Selection for the purpose of feature extraction and Density Based Spatial Clustering. The experimental results state that the system is good at recognizing the anomalies.

"Ali Yazdian Varjani" had brought the new idea to develop the intrusion detection system that will solve the sensitivity shortage at initial centers in the K-means algorithm by utilizing the MinMax K-means algorithm. By using this system the clustering quality is increased.

"Pradeep singh" had developed the two hybrid intrusion detection systems, the first system uses the Random Forest classifier and Gaussian Mixture Clustering, the second system uses the Random Forest classifier and K-means clustering. He compared the two systems with respective of performance in identifying the false alarm rate and observed that the system with Random Forest classifier and K-means clustering is good.

"Manoj Kumar" had proposed a intrusion detection system for the cloud computing. This system utilizes the outlier detection technique to recognize the false detection. The advantage of this system is it can detect the attaks without the prior knowledge.

"Anand sukumar" J V had developed the intrusion detection system where improved genetic k-means algorithm is utilized to recognize the intrusion. It is observed that, when there are less datasets the system has the less accuracy rate.

"Neha G. Relan" had developed the intrusion detection system by the means of C4.5 decision tree algorithm and with the pruning of C4.5 decision tree. The observations of the experiment states that C4.5 decision tree algorithm is detecting with less accuracy.

3. CONCLUSION

In this review paper, we have provided the different types of models that will detect the intrusion in the network. In future the data mining algorithms were integrated and used to achieve the accuracy for prediction and classification of attacks in KDD dataset.

REFERENCES

[1] S. Varuna, P.natesan. "An integration of k-means clustering and naïve bayes bayes classifier for Intrusion Detection", 2015 3rd Intetrnational Conference on Signal Processing, Communication and Networking (ICSCN), 2015.

[2] A.S Subaira, P.Anitha. "Efficient classification mechanism for intrusion detection system based on data mining techniques:A survey",2014 IEEE 8th International Conference on Intelligent Systems and Control(ISCO), 2014.

[3] Nadya EI Moussaid, Ahmed Toumanari. "Overview of intrusion detection using data mining and features selection", International Conference on Multimedia Computing and Systems (ICMCS) 2014

[4] Relan, Neha G., and Dharmaraj R. Patil. "Implementation of network intrusion detection system using variant of decision tree algorithm", International Conference on Nascent Technologies in the Engineering Field (ICNTE), 2015.

[5] Dr.S.Vijayarani and Ms. Maria Sylviaa.S, INTRUSION DETECTION SYSTEM-A STUDY.",International Journal of Security, Privacy

and Trust Management (IJSPTM) Vol4, No 1, February 2015.

[6] Madhulika Deshmukh, Prof. S.K. Shinde. "Intrusion Detection System Using Clustering ", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6656-6658

[7] Mohsen Eslamnezhad, Ali Yazdian Varjani. "Intrusion Detection Based on MinMax K-means Clustering", International Symposium on Telecommunications (IST) 2014

[8] Jiang, ShengYi, Xiaoyu Song, Hui Wang, Jian-Jun Han, and Qing-Hua Li. "A clustering-based method for unsupervised intrusion detections." Pattern Recognition Letters 27, no. 7 (2006): 802-810.

[9] Oh, Sang Hyun, and Won Suk Lee. "An anomaly intrusion detection method by clustering normal user behavior." Computers & Security 22, no. 7 (2003): 596-612.

[10] Pradeep Singh, M.Venkatesan, "Hybrid Approach for Intrusion Detection System ", International Conference on Current Trends toward Converging Technologies 2018

[11] M.Kabir-Gambo, Azman-Yasin, "Hybrid Approach for Intrusion- Detection Using the Combination of K -Means Clustering Algorithm and RandomForest Classification", The International Journal Of Engineering And Science,vol. 6, pp. 93-97, Jan. 2017.

[12] Gunupudi-Rjesh K, Mangathayru N, G Narsimha, "Similarity measure of intrusion-detection with gaussian function", Rev. Tec. Ing. Univ. Zulia. , vol. 39,no.2, pp. 173-183, 2016

[13] E. E. Papalexakis, A. Beutel, and P. Steenkiste, "Network anomaly detection using co-clustering," in Proc. 2012 Int. Conf. Advances in Social Networks Anal. and Mining (ASONAM 2012), 2012, pp. 403-410.

[14] K. Wankhade, S. Patka, and R. Thool, "An Overview of Intrusion Detection Based on Data Mining Techniques," in Proc. 2013 Int. Conf. Commun. Syst. and Network Technologies, 2013, pp. 626-629.

[15] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proc. 2nd IEEE Symp. Computational Intell. for Security and Defence Applicat., 2009.

[16] Cuixiao Zhang; Guobing Zhang; Shanshan Sun. "A Mixed Unsupervised Clustering-based Intrusion Detection Model",Third International Conference on Genetic and Evolutionary Computing 2009

[17] Patel A.,Sammarvar,S., and Naik, A. DataMining Vs.Statistical Techniques for Classification of NSL-KDD Intrusion Data. International Journal of Computer Science and Information Technologies, Vol 5(4), 2014.ISSN:075-9646

[18] Siddiqui, M.K., and Naahid, S.Analysis of KDD CUP 99 Dataset using Clustering based Data Mining. International Journal of Database Theory and Application Vol.6, No. 5. pp.23-24. 2013

[19] Sabhani, M., and Serpen, G. Why Machine Learning Algorithms Fail in Misuse Detection on KDD Intrusion Detection Dataset. Intelligent Data Analysis, vol 6. (Jne 2004).

[20] Srilatha Chebrolu, Ajith Abrahama,,*, Johnson P. Thomas, Feature deduction and ensemble design of intrusion detection systems, Elsevier Ltd. doi:10.1016/j.cose.2004.09.008

[21] Salvatore Pontarelli, Giuseppe Bianchi, Simone Teofili. Traffic-aware Design of a High Speed FPGA Network Intrusion Detection System. Digital Object Indentifier 10.1109/TC.2012.105, IEEE TRANSACTIONS ON COMPUTERS.

[22] Sailesh Kumar, "Survey of Current Network Intrusion Detection Techniques", available at http://www.cse.wustl.edu/~jain/cse571-07/ftp/ids.pdf.

[23] S. A. Khayam, Recent Advances in Intrusion Detection, Proceedings of the 26th Annual Computer Security Applications Conference, Saint-Malo, France, pp. 224-243, 42, 2009

[24] J. Xu, J. Wang, S. Xie, W. Chen and J. Kim, "Study on Intrusion Detection Policy for Wireless Sensor Networks", International Journal of Security and Its Applications, vol. 7, no. 1, (2013) January, pp. 1-6.