www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 12 Dec. 2016, Page No. 19494-19497

A Comparative Study of Perceiving Intrusion Using Data Mining Techniques

M.Deepa[1], Dr.P. Sumitra[2],
[1]Ph.D Research Scholar, Department of Computer Science,
Legithasai2010@gmail.com
[2]Professor, Department of Computer Science,
sumithravaratharajan@gmail.com
[1,2]Vivekananda College of Arts and Sciences for Women (Autonomous),Elayampalayam.

ABSTRACT: By the rapid development of the computer network during the past few years, the security of information issue comes to be more and more difficult. The Intrusion Detection Systems (IDS) can be used widely for protecting network. Data mining techniques are extensively used, due to some attributes like the scalability, adaptability and validity. This paper focuses on review of the existing intrusion detection system by using data mining techniques and discussing on various disputes in the existing system based on certain classification parameters such as accuracy, detection rate, false alarm etc.

Key Terms: Data Mining, Intrusion Detection System, Classification, Clustering.

I. INTRODUCTION

In recent years, with the terrific growth in networked computer resources, a variety of network-based applications have been developed to provide services in different areas such as ecommerce services, social media services, banking services, government services, etc. These Internet applications need a satisfactory level of security and privacy. On the other hand, the intruder create many vulnerable programs that attacks the various information on the networks .There is an increasing availability of tools and tricks for attacking and intruding networks. Compared with previous protection system, the Intrusion detection System (IDS) has come to be a key factor for the security of the network in the current online world. The data mining approach used in the field of IDS yields an improvement of detection rate, managing the false alarm rate and reduce false positive rate.

II. INTRUSION DETECTION SYSTEM

The intrusion detection system is an approach that presets the intrusion that are occurred on the network. The intrusion has many types namely viruses, worms, Trojan horse, etc. The foregoing defense system like firewall, virtual private network haven't a sufficient ability for recognizing critical intrusions from the network. The role of IDS is to trap the hacker's presence on the network and inform to the network administrator or user of the system and also raises alarms or signals when the security violations are occurred. The figure 1 describes the overall architecture of IDS. Initially the information can be retrieved from the database, which is checked by the firewall. It can be protected by the IDS and sends the information to the corresponding network.IDS plays an important role to secure the network and its main goal is to view the network activities automatically to identify the malicious attacks. Over the years, the researchers and designers have used many techniques to design the IDS. But, there have been limitations exist in present intrusion detection systems.

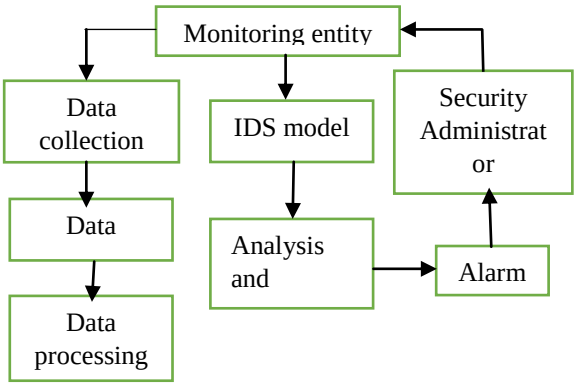

Figure 1: Overall architecture of Intrusion detection system

The IDS notices many attacks on the networks like (i) scanning attacks, (ii) denial of service (DOS) attacks, and (iii) penetration attacks. Each of these three categories of computer attacks has distinct signatures and behaviors - to which IDS is designed to analyze, detect and triggers an alarm when encountered. Once an alarm is set, the network administrators has to analyze the logs to decide whether these unexpected activity is indeed anomalous.IDS use either (i) Signature based detection or (ii) anomaly based detection

A. Signature based detection

A Signature based detection system monitors the network traffic for matches to the signature that is preconfigured and predetermined by the domain experts. Once frequent instances tie with the preconfigured domain then the IDS take the appropriate actions. This type of intrusion detection system can only detect the known threats but the unknown attacks are not identified by this method, and whenever a new software is arrived then this model needs an updating.

B.Anomaly based detection

The new or unknown threats are identified by an anomaly based detection approach. This method builds a

models of normal network behavior (called profiles) that are then used to detect the significantly deviate from the profiles. Thus given the promising capabilities of anomaly-based network intrusion detection systems (A-NIDS), this approach is currently a principal focus of research and development in the field of intrusion detection.

III. DATA MINING

Data mining techniques are used widely in various fields. The Network Intrusion Detection System is designed by using the data mining concept based on some attribute such as detect correct attacks within less time and raise the appropriate alarm. The intrusion detection system have been formatted by using various data mining techniques in an efficient manner. Data mining techniques can detect both known and unknown threats. In data mining a lot of techniques are available for analyzing and detecting the intrusion on the network. Some of the techniques are association rules, clustering, and classification, etc.

Association rule

Association rules mining identifies association among database attributes and their values. It is a pattern discovery technique which does not serve to solve classification problems nor predict problems.

Classification

Classification is the process of learning a function that maps data objects to a subset of a given class set. There are two goals of classification, first finding a good general mapping that can predict the class of so far unknown data objects with high accuracy. Second to find a compact and Understandable class model for each other classes

Clustering techniques

Clustering groups the data elements into different groups based on the similarity between within a single group Cluster partitions the data set into clusters or equivalence classes. Cluster methods divided into two categories based on the cluster structure namely Non Hierarchical and Hierarchical –connection oriented. This paper focuses on the study of existing intrusion detection task by using data mining techniques and discussing on various issues in existing IDS based on data mining techniques.

IV. LITERATURE REVIEW

Chong eik loo etc (2016) has designed to sense an abnormal traffic patterns by collecting an information of normal traffic pattern. In this approach obtain maximum energy because there is no way for sharing information from neighboring nodes. Every node on the network has an individual IDS. The anomaly based approach is used to model the distribution of training points by consuming clustering algorithm which is based on a fixed width. In this model the detection rate have been increased and the false positive rate will be In this method each node has independent IDS so it leads to high configuration rate. [1]

Jaina Patel, Krunal Panchal (2015). In this paper the author integrates both type of detection techniques. The irregularity data was identified by SNORT. Anomaly based IDS use both "k-means and CART" CART (Classification and Regression Trees) for classifying normal and abnormal activities in the network. The author evaluates the data from KDD Cup Dataset. The proposed assemblage is familiarized to maximize the effectiveness in identifying attacks and achieve high accuracy rate as well as low false alarm rate. The approach CART proceeds not as much of time to form decision tree than the earlier one. [2]

G. V. Nadianmai and M. Hemalatha (2014) in this paper the author point out four concerns which are found in an existing one like Classification of Data, High Level of Human Interaction, Lack of Labeled Data, and Effectiveness of Distributed Denial of Service (DoS) attack. The author suggested the EDADT algorithm, Hybrid IDS model, Semi-Supervised Approach and Varying HOPE RAA algorithm for solving the above stated problems. In this an enhanced data adapted decision tree algorithm is used to effectively classify the data into normal and attack without any classification. The algorithm SNORT and anomaly based approaches are being used to reduce the workload of network administrator. The frequently occurred data are classified by hybrid IDS pre-defined rules. The issue related to belling the unlabeled data is solved using Semi-Supervised approach where with the small amount of labeled data, the large amount of unlabeled data can be labeled. The author practice varying clock drift for explaining the Distributed Denial of Service Attack. This varying clock drift in network based applications makes it difficult for the intruder to access the port that has been used by the legitimate client [3].

Yogita B. Bhavasar, Kalyani C. Waghmare(2013) recommended intrusion detection system using Support Vector Machine (SVM). The NSL-KDD Cup'99 data set is engaged by their verification. The NSL-KDD Cup'99 data set took more time for construct the SVM model. This model yields high accuracy, detection rate and low false positive rate. In this proposed work author has used Gaussian Radial Basis Function but it needs extensive memory requirements for classification in many cases.[4]

Sahilpreet Singh, Meenakshi Bansa(2013), have proposed a paper suggested the Multilayer Perceptron feed forward neural network and use back propagation algorithm for detecting the intrusion on the network. This approach classify the attacks in an efficient way and deliver high accuracy with low error rate. The author use NSL KDD dataset and WEKA machine learning tool. [5]

Chitrakar, Chuanhe (2012) discussed the SVM classification and k-medoids clustering. By using k-medoids clustering similar data instances are grouped and the resulting clusters are classified by using SVM classifiers. This approach safeguard from all the attacks like Dos, probe, U2R, R2L. This approach yields high accuracy rate but it takes more time when the dataset is very large. [6]

N. S. Chandolikar, V. D. Nandavadekar (2012), use J48 decision tree classifier. The author suggested many approaches for their evaluation like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error and Kappa statistics measures. The KDD Cup 99 data set is used for their verification. In this approach the search space is divided into rectangle region. [7]

AhmedYoussef and Ahmed Emam(2011) recommended Data mining and Network Behavior Analysis concepts. This approaches detect intrusions on networks more excellently. The NBA architecture is formed for monitoring the inbound and outbound traffics and to ensure the overall security of the network at all levels. NBA based intrusion detection system tracks signature based security attacks faster. [8]

D. Md. Farid, N. Harbi and M. Z. Rahman(2010), concerns Naive Bayesian classifier and ID3 algorithm. Author also talks some problems that are present in the existing one, such as handling continuous attribute, missing attribute values and treducing noise in training data. This approach solve above problems and achieves good detections rate and low false positives and also eliminates redundant attributes from training data set. This model used Knowledge Discovery Data Mining (KDD) CUP 99 dataset for experiment [9].

Tao Peng, Wanli Zuo considered two things namely novel FP-tree structure and FP-growth mining methods. The FP-growth mining integrates divide-and-conquer strategy that compress the frequent item from database into a frequent-pattern tree, and takings mining of the FP-tree using Apriori candidate generation algorithm. The author used DARPA 2000 data set for detecting Intrusion. The proposed structure is a distributed architecture that consists of sensor, data preprocessor, extractors of features and detectors [10].

TABLE 1 COMPARITIVE STUDY OF PAPERS

Author(s) Name	Paper Name	Methodology used	Limitations
Chong Eik Loo (2016)	Intrusion Detection for Routing Attacks in Sensor Networks	Collect normal traffic pattern and then identify abnormal patterns	95% detection rate for a 5% false positive rate but high expensive
Jaina Patel,Krunal Panchal (2015)	Effective Intrusion Detection System using Data Mining Technique	CART identifies abnormal activities and SNORT generates alerts	High accuracy and high false positive
G. V. Nadianmai, M. Hemalatha (2014)	Effective approach toward Intrusion Detection System using data mining techniques	Classifies the data , label the unlabeled data then Dos is solved by clock drift	Combining algorithms produce better result
Yogita B. Bhavasar, Kalyani C. Waghmare (2013)	Intrusion Detection System Using Data Mining Technique: Support Vector Machine	Classifies the data by using SVM	Detection rate is increased & False positive rate is decreased but it requires high memory requirements for classification
Sahilpreet Singh, Meenakshi Bansa (2013)	Improvement of Intrusion Detection System in Data Mining using Neural Network	All the nodes are connected so there is no cycle.	This approach delivers accuracy in high manner and decrease the false error rate. Doesn't scale well and trained MLP not updated
Chitrakar, Roshan, Chuanhe (2012)	Clustering Anomaly based Intrusion Detection using Hybrid Learning Approach Of combining k-Medoids and SVM Classification"	Similar data instances are grouped by k- medoids and resulting clusters are classified by using SVM classifiers	Higher accuracy. Time complexity is more when the dataset is very large.
N. S. Chandolikar, V. D. Nandavadekar (2012)	Efficient algorithm for intrusion attack classification by analyzing KDD Cup 99	Similar data instances are classified by J48 classifier	Doesn't handle nonnumeric data, pruning is necessary
Ahmed YoussefAhmed Emam (2011)	Network intrusion detection using data mining and network behavior analysis.	Monitor inbound and outbound traffic	High accuracy and low false positive but needs improvement for false positives in remote to user (R2L) attack.
D. Md. Farid, N. Harbi , M. Z. Rahman(2010)	Combining naive bayes and decision tree for adaptive intrusion detection"	It performs balance detections and keeps false positives at acceptable level for different types of network attacks.	It produces accuracy result and high detection rate but requires large searching time and generate long rules.
Tao Peng, Wanli (2006)	Data Mining for Network Intrusion Detection System in Real Time	The frequently occurred instances are grouped into patterns	It takes more time for analyzing patterns. It have need of many scans for identifying patterns

V.CONCLUSION

There is an essential need for developing a new approaches for detecting the attacks that are rapidly found on the internet. This paper has evaluated different classification and clustering data mining techniques on the root of detection rate, accuracy, execution time and false alarm rate, for detecting the intrusion on the network. The results also suggest that no individual techniques can achieve better performance than the hybrid one. In order to achieve the better accuracy, high detection rate and to reduce false alarm rate the existing algorithms can be modified.

VI. REFERENCES

[1] Chong Eik Loo, Mun Yong Ng, Christopher Leckie, Marimuthu Palaniswami,Intrusion Detection for Routing Attacks in Sensor Networks., International Journal of Distributed Sensor Networks, 2006, pp.313-332.

[2]Jaina Patel, Krunal Panchal "Effective Intrusion Detection System using Data Mining Technique", JETIRJune 2015, Volume 2, Issue 6

[3] Nadiammai G. V, Hemalathain M,―Effective approach toward Intrusion Detection System using data mining techniques,Cairo University, Elsevier,Egyptian Informatics Journal, 2014, pp. 37-50

[4] Ogita B. Bhavasar, Kalyani C. Waghmare "Intrusion Detection System Using Data Mining Technique: Support Vector Machine" 2013 International Journal of Emerging Technology and Advance Engineering ,March 2013,volume 3, Issue 3.

[5]Sahilpreet Singh Meenakshi Bansa ,"Improvement of Intrusion Detection Systemin Data Mining using Neural Network",IJARCSSE, September 2013,Volume 3, Issue 9

[6] Chitrakar R, Chuanhe H, Clustering Anomaly based Intrusion Detection using Hybrid Learning Approach of combining k-Medoids and Naïve Bayes Classification, In Proceedings of 8th IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), 2012, p1-5.

[7] N.S.Chandolikar and V.D. Nandavadekar, "Efficient algorithm for intrusion attack classification by analyzing KDD Cup 99", Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on ISSN :2151-7681, (2012) September 20-22, pp. 1 - 5.

[8]Ahmed Youssef and Ahmed Emam, "Network Intrusion detection using data mining and network behavior analysis", International Journal of Computer Science & Information Technology (IJCSIT) ,Dec 2011,Vol 3, No 6.

[9]D. Md. Farid, N. Harbi and M. Z. Rahman,"Combining naive bayes and decision tree for adaptive intrusion detection", International Journal of Network Security & Its Applications (IJNSA), April 2010**,**vol. 2, no. 2.

[10] Tao Peng, Wanli Zuo "Data Mining forNetwork Intrusion Detection System in RealTime" IJCSNS International Journal of Computer Science and Network Security, February 2006,vol6