Review Paper on Intrusion Detection Using Data Mining Technique

Snehal S. Somwanshi¹, Soniya N. Madavi²

¹Post Graduate Student, ²Assistant Professor, Maharashtra Institute of Technology, Aurangabad, Maharashtra, India.

Abstract—Nowadays Intrusion Detection System has various issues regarding set of rules and attack classification. Intrusion Detection Systems that practices solitary algorithm, the accuracy and detection rate were not accurate. Increase in the false alarm rate was also come across. Grouping of algorithm is performed to solve this problem. This paper represents two fusion algorithms for developing the intrusion detection system. C4.5 and Support Vector Machine (SVM) algorithms are combined to maximize the accuracy, which is the advantage of C4.5 and reduce the wrong alarm rate which is the advantage of SVM.

Keywords—C4.5, Data Mining, Hybrid Algorithm, Internet attacks, Intrusion Detection System, Support Vector Machine.

I. INTRODUCTION

In today’s situation network is the heart of computer. Internet gives many advantages but has some disadvantages too. It is used as a apparatus for crime. For instance, on February 7th, 2000 the first DoS attacks of great volume where launched and they targeting the computer systems of large companies like Yahoo!, eBay, Amazon, CNN, ZDnet and Dadet[6]. Hence there is an immediate need to recognized and detect the attacks. Recent intrusion detection solicitations facing complex problems. These applications has to be oblige, consistent, extensible, easy to manage, and have low maintenance cost. In current situation, intrusion detection systems (IDSs) using data mining technique have attracted to many researchers because of it gives high accuracy, good simplification to novel types of intrusion, and stout performance in a changing environment. Data mining is indispensable for pattern discovery, Data miners are proficient at using specialized software to find regularities and irregularities in the large data sets.

In the case of detecting attacks, Intrusion Detection System can be classified as host-based and network-based [1]:

HOST-BASED IDS: In this data come from the records of various host activities, including audit record of operation system, system logs, application programs information, and so on.

NETWORK-BASED IDS: In this data is mainly collected from network generic stream going through the network segments, such as: Internet packets.

Intrusion Detection Techniques divides into two categories [1]:

ANOMALY DETECTION: It verifies the normal system behaviour like the network bandwidth, ports, device connections and protocols. If there is any anomalous activity in system or network it informs to the administrator.

MISUSE DETECTION: It is the ability to identify malevolent intrusions based on a known pattern for the activity.

There are four different types of attacks [5]. Attacks in IDS-

Denial of service attack (Dos):
It is a type of attack in which the attacker makes the memory too busy or too full to handle the requests.

User to Root Attack (U2R):
It is a type of attack in which attacker tries to access the normal user account.

Remote to Local Attack (R2L):
In this type attacker sends packets to a machine over a network but it does not have an account on that machine.

Probing Attack:
It is an activity to gather information about the network of computers.

II. RELATED WORK

In 1980, the idea of intrusion detection began with Anderson's [9] He presented a intimidation classification model that develops a security monitoring surveillance system based on detecting abnormalities in user behavior. There are various techniques that have been used in the development of Intrusion Detection System which are categorized as single and hybrid techniques.

M. Jiang et al (2011) [10] In this paper he focused on a collective model for misuse and anomaly intrusion detection. Common behaviour rule set are developed by using the clustering analysis algorithm to detect the new anonymous attacks and association rule mining algorithm was useful to detect the known attack rapidly.

M. Panda et al (2011) [11] In this paper they have proposed a hybrid intelligent intrusion detection system by merging the two classification algorithm for making the decision more accurate and rapid. First the classification or clustering was applied in the whole dataset and the output is applied to another classification algorithm. They applied 10-fold cross validation method, and the result achieved is in the form of normal or intrusion.

Yogita B.Bhavsar et al (2013) [2] In this paper they have proposed Intrusion Detection System using data mining technique SVM(Support Vector Machine) and in their proposed system SVM is used for classification and verification regarding the effectiveness of the proposed system is done by conducting some experiments using NSL-KDD Cup’99 dataset which is improved version of KDD Cup’99 data set. The SVM is one of the most noticeable classification algorithms in the data mining area, but its disadvantage is its large training time. For this drawback they proposed system that carried out some experiments using NSL-KDD Cup’99 data set. The experimental results show that it can reduce extensive time required to build SVM model by performing proper data set pre-processing. Also when they do proper selection of SVM kernel function such as Gaussian Radial Basis Function because of this it produce results that shows attack detection rate of SVM is increased and False Positive Rate (FPR) is decrease.

Jashan Koshal, et al(2012) [3] This paper represented two hybrid algorithms for developing the intrusion detection System.C4.5 decision tree algorithm and Support Vector Machine (SVM) algorithm are combined to maximize the accuracy and because of the combination of algorithm result shows that increase in the accuracy and detection rate. And they have conclude that the cascading of algorithm gives better results than single algorithm in detection of Intrusions.

Cascading the Algorithm improves the accuracy and reliability. Therefore Researchers are concentrating on cascading techniques for developing the Intrusion Detection System.


M. Jiang et al (2011) [10] In this paper he focused on a collective model for misuse and anomaly intrusion detection. Common behaviour rule set are developed by using the clustering analysis algorithm to detect the new anonymous attacks and association rule mining algorithm was useful to detect the known attack rapidly.

M. Panda et al (2011) [11] In this paper they have proposed a hybrid intelligent intrusion detection system by merging the two classification algorithm for making the decision more accurate and rapid. First the classification or clustering was applied in the whole dataset and the output is applied to another classification algorithm. They applied 10-fold cross validation method, and the result achieved is in the form of normal or intrusion.

Yogita B.Bhavsar et al (2013) [2] In this paper they have proposed Intrusion Detection System using data mining technique SVM(Support Vector Machine) and in their proposed system SVM is used for classification and verification regarding the effectiveness of the proposed system is done by conducting some experiments using NSL-KDD Cup’99 dataset which is improved version of KDD Cup’99 data set. The SVM is one of the most noticeable classification algorithms in the data mining area, but its disadvantage is its large training time. For this drawback they proposed system that carried out some experiments using NSL-KDD Cup’99 data set. The experimental results show that it can reduce extensive time required to build SVM model by performing proper data set pre-processing. Also when they do proper selection of SVM kernel function such as Gaussian Radial Basis Function because of this it produce results that shows attack detection rate of SVM is increased and False Positive Rate (FPR) is decrease.

Jashan Koshal, et al(2012) [3] This paper represented two hybrid algorithms for developing the intrusion detection System.C4.5 decision tree algorithm and Support Vector Machine (SVM) algorithm are combined to maximize the accuracy and because of the combination of algorithm result shows that increase in the accuracy and detection rate. And they have conclude that the cascading of algorithm gives better results than single algorithm in detection of Intrusions.

Cascading the Algorithm improves the accuracy and reliability. Therefore Researchers are concentrating on cascading techniques for developing the Intrusion Detection System.


M. Jiang et al (2011) [10] In this paper he focused on a collective model for misuse and anomaly intrusion detection. Common behaviour rule set are developed by using the clustering analysis algorithm to detect the new anonymous attacks and association rule mining algorithm was useful to detect the known attack rapidly.

M. Panda et al (2011) [11] In this paper they have proposed a hybrid intelligent intrusion detection system by merging the two classification algorithm for making the decision more accurate and rapid. First the classification or clustering was applied in the whole dataset and the output is applied to another classification algorithm. They applied 10-fold cross validation method, and the result achieved is in the form of normal or intrusion.

Yogita B.Bhavsar et al (2013) [2] In this paper they have proposed Intrusion Detection System using data mining technique SVM(Support Vector Machine) and in their proposed system SVM is used for classification and verification regarding the effectiveness of the proposed system is done by conducting some experiments using NSL-KDD Cup’99 dataset which is improved version of KDD Cup’99 data set. The SVM is one of the most noticeable classification algorithms in the data mining area, but its disadvantage is its large training time. For this drawback they proposed system that carried out some experiments using NSL-KDD Cup’99 data set. The experimental results show that it can reduce extensive time required to build SVM model by performing proper data set pre-processing. Also when they do proper selection of SVM kernel function such as Gaussian Radial Basis Function because of this it produce results that shows attack detection rate of SVM is increased and False Positive Rate (FPR) is decrease.

Jashan Koshal, et al(2012) [3] This paper represented two hybrid algorithms for developing the intrusion detection System.C4.5 decision tree algorithm and Support Vector Machine (SVM) algorithm are combined to maximize the accuracy and because of the combination of algorithm result shows that increase in the accuracy and detection rate. And they have conclude that the cascading of algorithm gives better results than single algorithm in detection of Intrusions.

Cascading the Algorithm improves the accuracy and reliability. Therefore Researchers are concentrating on cascading techniques for developing the Intrusion Detection System.
As it is made in java, it permits user to put on data mining plus machine learning algorithm to their data regardless of the platform and policy of the computer. It is substantially accessible on internet and comes under GNU license.[12].

A. C4.5 Decision Tree Algorithm

It is an algorithm used for classification in data mining. It is a stimulation algorithm that creates a model using the data set and deduce out some assumption. Based on the assumptions deduced, it tries to classify the new data set. It is also called as Classification Tree Algorithm.

At each node of the tree C4.5 selects an attribute that mostly splitting the samples into subsets. The splitting criteria use information gain. The attribute with the highest information gain is chosen to make decision. For building decision tree [3]

1. Check for base classes
2. For each attribute a find the information gain from splitting A.
3. Let A is a best attribute with the highest information gain.
4. Create a decision node that splits the A.
5. Recurse on the sub lists obtained by splitting A and add those nodes as children’s of nodes.

It can handle both continuous and discrete data. It can handle the missing attributes values. After finishing it goes back for pruning. C5.0 is the new advanced version of C4.5 is now available in the market.

Following figure2[15] shows basic working of decision tree. It follows the divide and conquers scheme. As it is decision tree it permits drawing decisions by trialing different potential considerations. The node of tree portraying the testing of attributes values. It allows testing of condition with several relational and comparison operators. Decision tree can be used for both description and prediction.

B. Support Vector Machine

The Support Vector Machine(SVM) is one of the greatest prosperous classification algorithms in the data mining area. SVM uses a high dimension space to find a hyper-plane to perform binary classification. SVM approach is a classification technique based on Statistical Learning Theory (SLT)[2]. It is based on the idea of hyper plane classifier. The main goal of support vector machine is to build a hyperplane or set of hyperplanes to separate tuples which belongs to two separate classes.[2]/[3]

Following figure3 shows Hyper plane of Support Vector Machine separating the two different classes.
In the above fig3/3 The distance between the data and hyper plane is shown. In the left image, the distance among the data and the hyper plane is small and in the right image the space is larger, which makes classification easy.

IV. CONCLUSION

This paper shows a survey of the various data mining techniques that have been applied to detect intrusions. As Overall Investigation shows that the intrusion detection system uses single algorithm not gives accurate results. Therefor to increase accuracy and reliability of Intrusion detection process we can use combinations of algorithms that helps us to detect intrusions more accurately.

REFERENCES

[1] Intrusion Detection Using Data Mining Techniques, Mohammadreza Ektefa,Department of IS UPM Serdang Malaysia mrektefa@gmail.com., Sara Memar Department of CS UPM Serdang, Malaysia sr.memar@gmail.com

[2] Intrusion Detection System Using Data Mining Technique: Support Vector Machine Yogita B. Bhavsar1, Kalyani C.Waghmare2 Post Graduate Student, Assistant Professor, Pune Institute of Computer Technology, Pune, Maharashtra, India

[3] Cascading of C4.5 Decision Tree and Support Vector Machine for Rule Based Intrusion Detection System Jashan Koshal, Monark Bag Indian Institute of Information Technology Allahabad, Uttar Pradesh211012,Indiajashankoshal@gmail.com,monarkbag@gmail.com

[4] A Survey on Building Intrusion Detection System Using DataMining Framework V.Jaiganesh, Assistant ProfessorDepartment of Computer ScienceDr.N.G.P. Arts andScience College Coimbatore-mail:jaiganeshree@gmail.comM.ThemozhiAssistantProfessorDepar tment of InformationTechnologyAvinashilingam University forWomen, Coimbatore-mail:themozhi@gmail.com


[13] Comparison betweenID3 and C4.5 in Contrast to IDS Surbhi Hardikar*,Ankur Shrivastava and Vijay Choudhary


[16] Choosing an Intrusion Detection System that Best Suits yourOrganization Dennis Mathew GSEC Practical v1.4b Option A


Acknowledgements

I would like to take this opportunity to express my thoughtful appreciation and deep regard to my Project Guide Prof. Miss S.N.Madavi, for her exemplary guidance, valuable feedback and constant encouragement for above survey.

As well thankful to Prof.Mrs.K.V.Bhosale, Prof.Mrs V.Kala for their valuable guidance.