A SURVEY: INTRUSION DETECTION ON NETWORK USING DATA MINING TECHNIQUES

Y.S. Kalai Vani[1], T.Srinivasa Rao[2], G.Prabhu[3], Y.Keerthi[4], Geetha Nadepalli[5], Vanitha[6]

[1,2,3,4,5,6]*Department of MCA, Sambhram Institute of Technology, VTU, (India)*

ABSTRACT

Network security technology has become crucial in all fields such as business and government. Government and business infrastructure has large volume of data set to protect these data from intruders, the intrusion detection system using the data mining techniques are used.To avoid problems in modern intrusion detection system the efficient data mining techniques are used as it gives accuracy, good generalization, good quality. Intrusion detection System detects the various attacks in network traffic. This paper explains the different approaches of data mining technique to the intrusion in network.

Keywords: Intrusion Detection System, Data mining, Intruders

I. INTRODUCTION

An Intrusion Detection System (IDS) is a device or application that monitors network or system behavior from malicious activities and produces report to a management station. Intrusion detection is software that automates the intrusion detection process. This intrusion has many types namely viruses, worms, Trojans, Horse etc. The normal detection system like firewall, VPN (Virtual Private Network)are failed to detect critical Intrusions from the network.

Data mining is a sophisticated approach to search the data from the huge capacity of storage. It has different methodologies to mine the data [1]. Data mining has different set of applications in different areas such as Business, weather forecasting, financial and marketing, neural networks; intrusion detection such as credit card detection etc., Data mining has different strategies of analysis which is used to find the data in effective way. Data mining is popularly used to combat frauds because of its effectiveness. It is a well-defined procedure that takes data as input and produces models or patterns as output Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make a valid prediction

II. CLASSIFICATION OF INTRUSION DETECTION SYSTEM

2.1 Network Based Intrusion Detection System

Network Intrusion Detection scrutinize network traffic[1], network intrusion detection system on dedicated machines that observe the network flows and it combines with firewall.

2.2 Host Based Intrusion Detection System

Host based Intrusion detection system runs on machine it can theoretically observe and log any event occurring on the machine.

III. DETECTION TECHNIQUES

The intrusion detection system has two categories for detecting attacks in the network or host.

3.1 Anomaly/Statistical Detection System

Anomaly detection[2] assumes that intrusions will always reflect some deviations from normal patterns. Anomaly detection may be divided into static and dynamic anomaly detection. A static anomaly detector is based on the assumption that there is a portion of the system being monitored that does not change. The static portion of a system is the code for the system and the constant portion of data upon which the correct functioning of the system depends. For example, the operating systems‟ software and data to bootstrap a computer never change. If the static portion of the system ever deviates from its original form, an error has occurred or an intruder has altered the static portion of the system. Dynamic anomaly detection typically operates on audit records or on monitored networked traffic data. Audit records of operating systems do not record all events; they only record events of interest. Therefore only behavior that results in an event that is recorded in the audit will be observed and these events may occur in a sequence

3.2 Misuse/Signature Based Detection System

It is based on the knowledge of system vulnerabilities and known attack patterns. Misuse detection is concerned with finding intruders who are attempting to break into a system by exploiting some known vulnerability. Ideally, a system security administrator should be aware of all the known vulnerabilities and eliminate them. The term intrusion scenario is used as a description of a known kind of intrusion; it is a sequence of events that would result in an intrusion without some outside preventive intervention. An intrusion detection system continually compares recent activity to known intrusion scenarios to ensure that one or more attackers are not attempting to exploit known vulnerabilities. To perform this, each intrusion scenario must be described or modeled.

IV. DATA MINING APPROACHES FOR INTRUSION

Data mining is the search for the relationshipAnd global patterns that exist in large databases.Data Mining is frequently used to designate the process extracting useful information from large data bases.

Predictionmakes use of existing variables in thedatabase in order to predict unknown or future values ofinterest.

Description focuses on finding patterns describing the data and the subsequent presentation for user interpretation. The relative emphasis of both prediction and description differ with respect to the underlying application and the technique.

Classification: Classification builds up (from the training set) and utilizes a model (on the target set) to predict the categorical labels of unknown objects to distinguish between objects of different classes. These categorical labels are predefined, discrete and unordered. The research literature describes that classification[7] or prediction is the process of identifying a set of common features (patterns), and proposing models that describe and distinguish data classes or concepts. Common classification techniques include neural networks, the Naïve Bayes technique, decision trees and support vector machines. Such classification tasks are used in the detection of credit card, healthcare and automobile insurance, and corporate fraud, among other types of fraud and classification is one of the most common learning models in the application of data mining in fraud detection

Clustering: Clustering is used to partition objects into previously unknown conceptually meaningful groups (i.e. clusters), with the objects in a cluster being similar to one another but very dissimilar to the objects in other clusters. Clustering is also known as data segmentation or partitioning and is regarded as a variant of unsupervised classification The most common clustering techniques are the K-nearest neighbor, the Naïve Byes technique and self-organizing maps.

Prediction - Prediction estimates numeric and ordered future values based on the patterns of a data set.

Outlier detection - Outlier detection is employed to measure the distance between data objects to detect those objects that are grossly different from or inconsistent with the remaining data set Data that appear to have different characteristics than the rest of the population are called outliers The problem of outlier/anomaly detection is one of the most fundamental issues in data mining. A commonly used technique in outlier detection is the discounting learning

Regression - Regression is a statistical methodology used to reveal the relationship between one or more independent variables and a dependent variable (that is continuous-valued)

V. CLASSIFICATION OF DATA MINING TECHNIQUES FOR INTRUSION

5.1 Detection Models

To determine the main algorithms used for financial accounting fraud detection, we present a Review of data mining techniques identified in literature applied to the detection of financial fraud. The most frequently used techniques are logistic models, neural networks, the Bayesian belief network, anomaly detection and misuse detection.

Regression Models: The regression based models are mostly used in financial accounting fraud detection. The majority of them are based on logistic regression, stepwise-logistic regression, multi criteria decision making method and exponential generalized beta two. Logistic model is a generalized linear model that is used for binomial regression in which the predictor variables can be either numerical or categorical [5]. It is principally used to solve problems caused by insurance and corporate fraud.

Neural Networks – The neural networks are non-linear statistical data modeling tools that are inspired by the functionality of the human brain using a set of interconnected nodes [6]. Neural networks are widely applied in classification and clustering, and its advantages are as follows. First, it is adaptive; second, it can generate robust models; and third, the classification process can be modified if new training weights are set. Neural networks are chiefly applied to credit card, automobile insurance and corporate fraud.

Bayesian Belief Network - The Bayesian belief network (BBN) represents a set of random variables and their conditional independencies using a directed acyclic graph (DAG), in which nodes represent random variables and missing edges encode conditional independencies between the variables [7]. The Bayesian belief network is used in developing models for credit card, automobile insurance, and corporate fraud detection.

Decision Trees – A decision tree (DT) is a tree structured decision support tool, where each node represents a test on an attribute and each branch represents possible consequences. In this way, the predictive model attempts to divide observations into mutually exclusive subgroups and is used for data mining and machine learning tasks.

Naïve Bayes - Naïve Bayes is used as simple probabilistic classifier based on Bayes conditional probability rule. Naïve Bayes follows strong (naive) statistical independence assumptions for the predictor variables. It is an effective classification tool that is easy to interpret and particularly suited when the dimensionality of the inputs is high. In a study. Naïve Bayes methods are widely used in banking and financial fraud detection and claim fraud detection.

Fuzzy Logic : Fuzzy Logic is a mathematical technique that classifies subjective reasoning and assigns data to a particular group, or cluster, based on the degree of possibility the data has of being in that group. The expert fuzzy classification[8] techniques enable one to perform approximate reasoning that can improve performance in three ways. First, performance is improved through efficient numerical representation of vague terms, because the fuzzy technology can numerically show representation of a data item in a particular category. The second way performance is enhanced is through increased range of operation in ill-defined environments, which is the way that fuzzy methodology can show partial membership of data elements in one or more categories that may not be clearly defined in traditional analysis.

Anomaly detection: Anomaly Detection the process where to localize objects that are different from other objects (anomalies).It is a technique for improving the analysis of typical data Objects. These anomalous objects are exceptionalin some sense. Lie far away from other data points (outliers) have attribute values that deviate significantly from the expected or typical attribute values indicate errors in data.

This paper focuses on the anomaly detection which is used to find the fraud in credit card using online .Anomaly uses different strategies to find out the anomalous behavior in the credit card.

VI. A REVIEW OF LITERATURE

6.1 Network Intrusion Detection System based on Data

Mining – S.A. Joshi, et. al.,

In this paper the author discuss about the data miningalgorithms and Intrusion detection system to detect the unknown attacks from the dataset. There different kinds of
attacks but the authors of this paper discuss the few kinds
of attacks. They compares the four types of attacks are:

a) Probing attack

b) Denial of service

c) User to root

d) Remote to local

Intrusion Detection Using Data Mining Techniques . This paper focuses on the Intrusion detection using data mining technique and it compared the different techniques of data mining classification, clustering etc. It compares the IDS techniques under data mining Anomaly detection and misuse detection technique. It explains the data mining and soft computing technique for intrusion

6.2 Advances in Computing, Communications and Informatics (Icacci), 2013 International Conference on

`The main research method is clustering analysis with the aim to improve the detection rate and decrease the false alarm rate. Most of the previously proposed methods suffer from the drawback of k-means method with low detection rate and high false alarm rate. This paper presents a hybrid data mining approach encompassing feature selection, filtering, lustering, divide and merge and clustering ensemble

6.3 A Survey on Efficient Data Mining Techniquesfor Network Intrusion Detection System (IDS)P.Kalarani1, Dr.S. Selva Brunda2

This survey paper describes the Data mining approaches which are used to the detect intrusion in a network.

A Review of Intrusion Detection System in Computer Networks - AbhilashaASayar, et.al., In this paper the author discuss about the classification of Intrusion detection system, advantageous anddisadvantageous and its types.In this the IDS uses the artificial intelligence, fuzzy logic and neural network. The techniques are used to detect the intrusions in the images.

For example, in military the original information’s are changed into images and then send to another location.

By using the artificial intelligence with IDS the user can easily identify the unknown attacks. This paper is useful for beginners to study the basic concepts of Intrusion detection system and also detect all kind of images.

A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems Reema Patel, Amit Thakkar, Amit Ganatra This paper focuses on the techniques based on data mining technique which is used to detect the intrusion in the network. Different algorithms and comparisons

A Survey on Intrusion Detection using Data Mining Techniques - R. Venkatesan, et al., This paper describes the overview of the intrusiondetection system and its each technique. The authorsdiscuss pros and cons of anomaly detection and misusedetection. By combining these two categories and datamining approaches, then include the Apriori associationrule algorithm for calculating the confidence levels.Apriori algorithm employs an iterative approach known asa level wise search, where k-item sets are used to explore(k + l)-item sets [5].

VII. CONCLUSION

This survey paper study on various techniques which are used to detect the attacks from unknown users. Theintrusion detection system components are useful to know about the process of detection. The IDS is combined withthe data mining techniques and algorithms detect thethreats and give immediate response to the user, and alsofind the percentage of detection rate. There is much research scope involvedfor the research community in this field to find the rightkind of generalization of the IDS model . The challenges to find solution to the newemerging attacks with using current data mining basedintrusion detection techniques along with fuzzy logic data set.

REFERENCES

[1] S.A.Joshi, VarshaS.Pimprale, “Network Intrusion DetectionSystem (NIDS) based on Data Mining”, International Journal ofEngineering Science and Innovative Technology, Vol. 2, No. 1,January 2013, ISSN. 2319-5967.

[2] Sushil Kumar Chaturvedi, Prof. VineetRichariya. Prof. NirupamaTiwari, “Anomaly Detection in Network using Data miningTechniques”, International Journal of Emerging Technology andAdvanced Engineering, Vol. 2, No. 5, May 2012, ISSN. 2250-2459.

[3] International Journal of Advanced Computer Technology (IJACT) ISSN:2319-7900 21 Intrusion Detection Using Data Mining Techniques . Krishna Kant Tiwari 1 ,Susheel Tiwari 2 , Sriram Yadav 3

[4] A SURVEY ON DATA MININGTECHNIQUES International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014

[5] .V.Nadiammai, S.Krishaveni, M.Hemalatha – “A comprehensive Analysis and study in intrusion detection system using data mining Techniques”. IJCA, Volume 35 –No.8, December 2011.

[6] P.Kalarani1, Dr.S. SelvaBrunda- A Survey on Efficient Data Mining Techniquesfor Network Intrusion Detection System (IDS)

[7]. Krishna Kant Tiwari 1 ,Susheel Tiwari 2 , Sriram Yadav 3- International Journal of Advanced Computer Technology (IJACT) ISSN:2319-7900 21 Intrusion Detection Using Data Mining Techniques

[8] A.R. Jakhale, G.A. Patil, “Anomaly Detection System by MiningFrequent Pattern using Data Mining Algorithm from NetworkFlow”, International Journal of Engineering Research andTechnology, Vol. 3, No.1, January 2014, ISSN. 2278-0181

[9] Tanase, Matthew, “ One of These Things is not Like the Others: The State of Anomaly Detection”,2010, http://www.symantec.com/connect/articles/one-these-things-not-others-state-anomalydetection