ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

A STUDY ON IDS USING CLASSIFICATION TECHNIQUES IN DATA MINING

K.Saranya#1,R.Bhavani*2,Dr.G.Padmavathi#3

#K.Saranya[1]
***R.Bhavani[2]**
Research Scholar,
Department of Computer Science,
Avinashilingam Institute for Home Science and Higher
Education for Women,
Coimbatore, India
saranyakumar.msc@gmail.com

Dr.G.Padmavathi,[3]
Professor and Head,
Department of Computer Science,
Avinashilingam Institute for Home Science and Higher
Education for Women,
Coimbatore, India

Abstract: Network security is one of the most vital one in this upgrading world, with the increasing number of attackers in network, Users also want more security from the intruders to their data; since the number of Intruder on the networks is increasing day-by-day . The data is transmitted from multiple sources and causes increased traffic intrusion and may lead intrusion problem. To avoid this problem an effective technique called IDS is raised, It is the act of detecting and responding to computer misuse to monitor and analyze the intrusion problem. By pertaining Data Mining techniques in network is a potential solution to develop better intrusion detection systems. Therefore this paper provides study of various classification techniques applied for Intrusion detection system
.

Keyword: Data mining, IDS, wireless network, classification and security.

I. INTRODUCTION

Even with today's sophisticated computer technologies (e. g., machine learning and data mining systems) extracting data from the large database is a critical task. However companies need to extract the information for large datasets, here the Data mining is a powerful technology with a great potential to help the companies to extract the important data from large datasets. Data Mining is the process of analyzing data from different perspectives and abridgment it into helpful information - information that can be used to enlarge revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for scrutinized data. It allows users to examine data from many different dimensions or angles, categorize it, and recapitulate the relationships identified. In some situation the extraction process of data may be done during classification and clustering methods [1]. In the endorsement field of Data Mining the intruder (attacker) also enhanced to hack the secretes data of others; it is the most significant one for many areas. Intrusion detection is a set of techniques and methods that are used to detect mistrustful activity in network with host level. The goal of intrusion detection is to determine intrusions into a computer or network, by scrutinize various network activities or attributes. Here intrusion refers to any set of actions that threatens the truthfulness, Intrusion detection systems fall into three basic categories: Signature-based intrusion detection systems, Anomaly detection systems and Specification-based Detection. Intruders have signatures, like computer viruses, that can be perceived with software [2]. The user can try to find data packets that include any known intrusion-related signatures or anomalies related to Internet protocols.

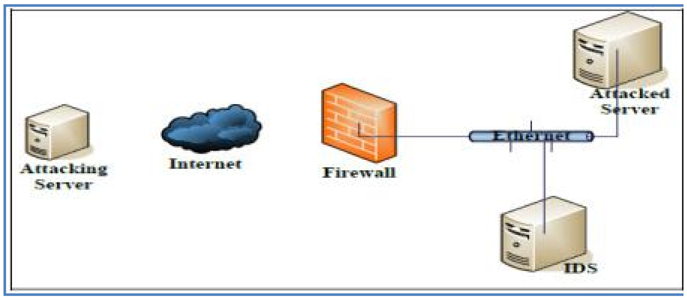

Figure 1: Architecture of IDS

Based upon a set of signatures and rules, the detection system is capable to find and log distrustful activity and produce alerts. Anomaly-based intrusion detection typically depends on packet anomalies present in protocol header parts. In some cases these methods produce recovered results compared to signature-based IDS.Specification based technique compare the behavior of substance with their associated security specifications that confine the correct behavior of the objects. Usually an intrusion detection system captures data from the network wireless network and can appropriate its rules to that data detect anomalies in it. Sniff is primarily a rule-based IDS, however input plug-ins are present to detect anomalies in protocol headers [3, 4]. This paper present the study of various intrusion techniques based on classification techniques.

II LITERATURE REVIEW

From the author Sahilpreet Singh and Meenakshi Bansal [5] present a survey of techniques of intrusion detection system using supervised and unsupervised learning. The techniques are classified based upon different approaches like Statistics, Data mining, Neural Network Based and Self Organizing Maps Based approaches. The detection type is borrowed from intrusion detection as either misuse detection or anomaly detection. It provides the reader with the major improvement in the malware research using these approaches and features the categories from the surveyed work based upon the above stated categories.

Huy Anh Nguyen and Deokjai Choi [6] assess that a network attacks have enlarged in number and severity over the past few years, Intrusion detection system (IDS) is gradually more becoming a critical component to protect the network. Due to huge volumes of security audit data as well as multifarious and dynamic properties of intrusion behaviors, optimizing performance of IDS becomes an important open problem that is getting more and more attention from the research community. The uncertainty to explore if certain algorithms achieve better for certain attack classes constitutes the enthusiasm for the reported herein. In this paper, they estimate concert of a comprehensive set of classifier algorithms using KDD99 dataset. Based on assessment results, best algorithms for each attack grouping is chosen and two classifier algorithm assortment models are proposed. The simulation result evaluation indicates that noticeable performance improvement and real-time intrusion detection can be achieved as we apply the proposed models to detect diverse kinds of network attacks.

S.A.Joshi et al.[7]in their paper acknowledged that with the incredible growth in information technology, network security is one of the challenging issue and so as Intrusion Detection system (IDS). The traditional IDS are unable to administer various newly arising attacks. To defeat this type of problem Data Mining techniques, Feature Selection, Multiboosting are applied. With data mining, it is easy to identify valid, useful and comprehensible pattern in large volume of data. Features are selected using binary classifiers for more exactness in each type of attack. Multiboosting is used to diminish both the variance and bias. Thus the efficiency and accuracy of Intrusion Detection system are enlarged and security of network so is also enhanced.

From Yogita B. Bhavsar [8] Security and privacy of a system is concession when an intrusion happens. Intrusion Detection System (IDS) plays vital role in network security as it detects different types of attacks in network. So here, they propose Intrusion Detection System using the data mining techniques: SVM

(Support Vector Machine). Here, Classification will be done by using SVM and verification concerning the usefulness of the proposed system will be done by conducting some experimentation using NSL-KDD Cup'99 dataset which is enhanced version of KDD Cup'99 data set. The SVM is one of the most important classification algorithms in the data mining area, but its drawback is its widespread training time. In this proposed system, the authors have carried out some experiments using NSL-KDD Cup'99 data set. The experimental results show that they can reduce extensive time necessary to build SVM model by performing proper data set pre-processing. Also when we do appropriate selection of SVM kernel function such as Gaussian Radial Basis Function is done, attack detection rate of SVM increases and False Positive Rate (FPR) decreases.

From Krishna Kant et al [9] analyze the increasing number of public and commercial services are used through the Internet, Here security of information becomes more of an imperative issue Intrusion Detection System (IDS) used against attacks On another way, some data mining techniques also contribute to intrusion detection. Some data mining techniques used for intrusion detection can be classified into two classes: misuse intrusion detection and anomaly intrusion detection. Misuse always refers to known attacks and destructive activities that exploit the known understanding of the system. Anomaly normally means a general activity that is able to indicate an intrusion. In this paper, assessment made between 23 related papers that use data mining techniques for intrusion detection. Data mining and soft computing techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Multivari-ate Adaptive Regression Spline (MARS), are measured . In this paper comparison shown between IDS data mining techniques and tuples used for intrusion detection. In those 23 related papers, 7 research papers use ANN and 4 ones use SVM, because of ANN and SVM are more reliable than other representation and structures. In addition, 8 researches use the DARPA1998 tuples and 13 researches use the KDDCup1999, because the standard tuples are much more convincing than others. There is no best intrusion detection model in the present time.

III CLASSIFICATION IN DATA MINING

Classification is a data mining (machine learning) technique used to foretell group membership for data instances. Classification is a data mining function that dispenses items in a collection to objective categories or classes. The goal of classification is to truthfully predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high recognition risks. Classifications models are tested by evaluating the predicted values to known target values in a set of test data [10]. The chronological data for a classification project is typically divided into two data sets: one for building the model; and the other for testing the model.

Classification algorithms in data mining and machine learning are given a set of inputs and they come up with a detailed class associated with those inputs. This study presents a an assortment of classification techniques of the data mining advance to solve the intrusion detection problems

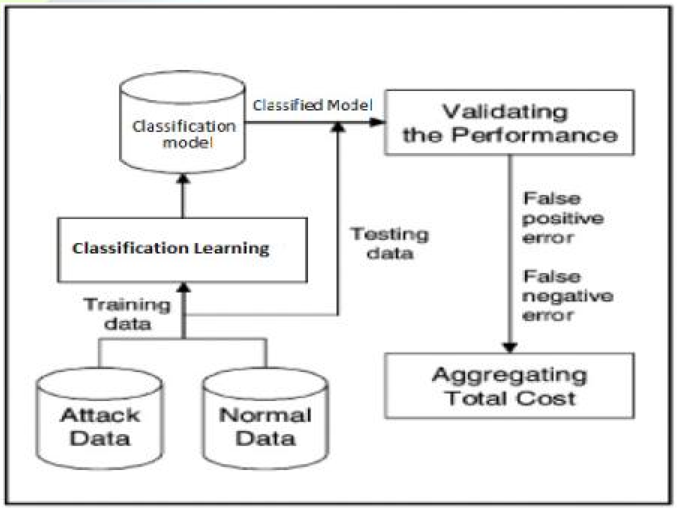

Figure 2: IDS using Classification model

IV IDS USING CLASSIFICATION TECHNIQUES:

The goal of classification is to assign objects (intrusions) to classes based on the values of the object's features. Classification technique can be used in misuse, anomaly and specification based detections. Classification is the task of oversimplify known structure to relate to new data. It is used to predict group relationship for data instances. To recognize the set of categories that new observation belongs to on the basis of a training data set that enclose observations (or instances) whose category membership is known. Classifier presentation depends greatly on the characteristics of the data to be classified. Intrusion detection systems of classification analysis fall into three basic categories:

- Misuse detection systems
- Anomaly detection systems
- Specification-based Detection

Intruders have signatures, like computer viruses, that can be distinguished using software. The user can try to find data packets that enclose any known intrusion-related signatures or anomalies related to Internet protocols. Based upon a set of signatures and rules, the uncovering system is able to find and log apprehensive activity and generate alerts. Anomaly-based intrusion detection usually depends on packet anomalies present in protocol header parts. In some cases these methods create better results compared to signature-based IDS

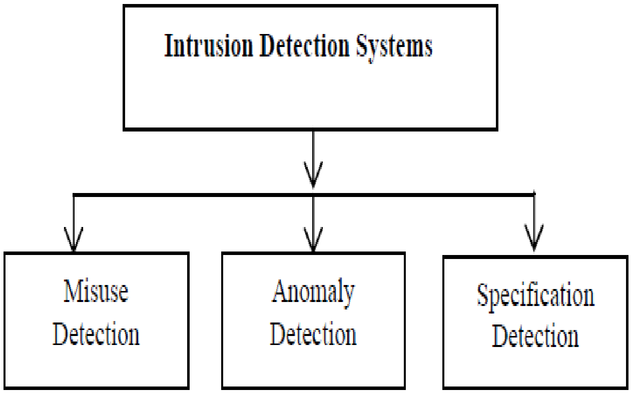

Figure 3: Classification analyses of IDS

Intrusion detection systems in wireless sensor networks are classified into three types on the basis of its detection techniques as,

a. Misuse detection: Misuse detection is also called as Signature-based intrusion detection. In this system the IDS is detected through behaviors of known attacks like antivirus software. Antivirus software compares the data with recognized code of virus. Similarly it is detected through signature of the behavior. In misuse detection, network traffic data are unruffled and labeled as "normal" or "intrusion". This labeled dataset is used as a training data to learn classifiers of different types. In Misuse detection, pattern of known malevolent activity is stored in the dataset and categorize distrustful data by comparing new instances with the stored prototype of attacks. Misuse detection finds intrusions by appear for activity corresponding to known techniques for intrusion. This generally involves the scrutinizing of network traffic in search of direct matches to known patterns of attack (called signatures). This is fundamentally a rule-based approach. A disadvantage of this approach is that it can only identify intrusions that follow pre-defined patterns. In misuse detection, there are state conversion analysis, pattern matching, model-based, keystroke monitoring and Expert system [12, 13].

b. Anomaly detection: It is different from Misuse detection. Here baseline of regular data in network data in network freight on network traffic, protocol and packet size etc are defined by system administrator and According to this baseline, Anomaly detector monitors new occurrence The new instances are contrast with the baseline, if there is any deviation from baseline, data is reported as intrusion. For this reason, it is also called performance based Intrusion detection system. In anomaly detection, the system describes the probable behavior of the network (or profile) in advance. Any significant divergences from this probable behavior are then descripted as possible attacks. Such deviations are not essentially actual attacks. They may simply be new network behavior that needs to be additional to the profile. The primary advantage of anomaly-based detection is the ability to detect novel attacks for which signatures have not been defined. Approaches to anomaly detection have neural network, Statistics, Predictive pattern generation, and sequence matching and supervising [12, 14].

c. Specification-based Detection: Specification based techniques compare the behavior of objects with their related security specifications that capture the correct performance of the objects. In the type of specification intrusion detection it detects the trespasser based on the behavior objects, If the object is harass on another object means this type intrusion detect based on that behavior only, so it will be workout on the time of executing progress. It will work out at the runtime only, so while in the offline or un-progress mode it does not work properly. And also this technique does not detect intrusions directly; it detects the effect of the intrusions as run-time violation of the specifications instead [12, 15].

Table 1: Comparison of different IDS techniques

Technique	**Advantage**	**Disadvantage**
Misuse detection	Higher detection rate, Low false alarm rate	High missing report, detect only known attack and failed in detecting new attackers
Anomaly detection	It can examine unknown data, minimizing alarm rate, detect new vulnerabilities and higher accuracy.	False positive result
Specification based Detection	Capture with associated object, detect at run-time and effective method	Run-time violation

This table shows the analyses and comparison of the IDS techniques with the help of Classification in Data mining.

V IDS USING CLASSIFICATION ALGORITHMS

a. K-Nearest Neighbor

It is one of the simplest classification techniques. It calculates the distance between different data points on the input vectors and assigns the unlabeled data point to its nearest neighbor class. K is an important parameter.

If k=1, then the object is assigned to the class of its nearest neighbor. When value of K is large, then it takes large time for prediction and influence the accuracy by reduces the effect of noise. The Knn is here used to find the attackers with the help of the nearest neighbor, with the help of the nearest parameter it assigned the class and estimate the attackers who will be the intruder and after that it will be intimate to the server.

b. Naive Bayes classifier

Naive Bayes classifier is probabilistic classifier. It predicts the class according to membership probability. To derive conditional probability, it analyzes the relation between independent and dependent variable. Where, X is the data record and H is hypothesis which represents data X and belongs to class C. Construction of Naive Bayes is easy without any complicated iterative parameter. It may be applied to large number of data points but time complexity increases. How it will help in IDS here is, the naïve Bayes probabilistic classifier, it classifies the attacker according to the membership probability, it have certain condition to analyze the intruder with the help of separating the normal and intruder data. It helps to find the intruder.

c. Support Vector Machine

The Support Vector Machine is one of the most successful 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).It is based on the idea of hyper plane classifier. The goal of SVM is to find a linear optimal hyper plane so that the margin of separation between the two classes is maximized. While intrusion behaviors happen, SVM will detect the intrusion. A classification task involves training set and testing set which consist of instances. Each instance in the training set contains one “target value" (class labels: Normal or Attack) and several “attributes" (features).The goal of SVM is to produce a model which predicts target value of data instance in the testing set which is given only attributes. From the verification of the above three algorithm SVM proves the best one to find the attackers in the network.

Table 2: Comparison of different classification algorithms in IDS

Algorithm	**Advantage**	**Disadvantage**
K-Nearest Neighbor	Find high detection rates.	Increasing in False result.
Naive Bayes classifier	Better accuracy than Knn, False alarm rate has been decreased	False positive result.
Support Vector Machine	Explore the Hit rate and False Rate on data set to detect no. of attacks and Shows maximum attacks and also increases the alert level	Fire alarms when nothing wrong in the network.

VI CONCLUSION

The security of user data plays a vital role in wireless network area. Detection of these attacks is the most significant issue in computer network security. For this, IDS play an important role to find out the attackers with the help of classifies: misuse detection, anomaly detection and Specification detection. There are several different methods for IDS detection; here the misuse and anomaly have some drawbacks like missing report, lack of missing attacks and high false rate but with the help of specification detection, it generate automatic detection examine the unknown data with associate one, providing minimizing alarm rate, can detect new vulnerabilities and higher accuracy.

VII REFERENCES

[1]. Jian Pei, Shambhu J. Upadhyaya Faisal Farooq, Venugopal Govindaraju, "Data Mining for Intrusion Detection– Techniques, Applications and Systems"

[2]. Ahmed Youssef and Ahmed Emam, "Network Intrusion Detection Using Data Mining And Network Behaviour Analysis".

[3]. Abhaya, Kaushal Kumar, Ranjeeta Jha, Sumaiya Afroz, "Data Mining Techniques for Intrusion Detection: A Review".

[4]. http://www.ll.mit.edu/IST/ideval/data/data_index.html

[5]. Sahilpreet Singh,Meenakshi Bansal, "A Survey on Intrusion Detection System in Data Mining"

[6]. Huy Anh Nguyen and Deokjai Choi, "Application of Data Mining to Network Intrusion Detection: Classifier Selection Model"

[7]. S.A.Joshi, Varsha S.Pimprale "Network Intrusion Detection System (NIDS) based on Data Mining" *International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013*

[8]. Yogita B. Bhavsar, Kalyani C.Waghmare Intrusion Detection System Using Data Mining Technique: Support Vector Machine

[9]. Krishna Kant Tiwari, Susheel Tiwari, Sriram Yadav, "Intrusion Detection Using Data Mining Techniques"

[10]. P. Kumar, P.R. Krishna, B. S Raju and T. M Padmaja, "Advances in Classification of Sequence Data", Data Mining and Knowledge Discovery Technologies. IGI Global, 2008, pp.143-174.

[11]. Steven Noel et al, " Modern intrusion detection, data mining, and degrees of attack guilt"

[12]. G.Keerthana and Dr.G.Padmavathi, "A Study on Sinkhole Attack Detection using Swarm Intelligence Techniques for Wireless Sensor Networks"

[13]. Wenke Lee, Salvatore J. Stolfo , Philip K. Chan, "Real Time Data Mining-based Intrusion Detection".

[14]. Jian Pei, Shambhu J. Upadhyaya, "Data Mining for Intrusion Detection – Techniques, Applications and Systems".

[15]. R. Sekar et al. Specification-based anomaly detection: a new approach for detecting network intrusions, in ACM CCS'02.

AUTHOR’S BIOGRAPHY

K.Saranya received her M.Sc Computer Science degree in 2011 from vysya college of Arts & Science, Salem. She is pursuing her M.Phil at Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. Her areas of interest are Data Mining, Security.

Dr.G.Padmavathi is the Professor and Head of computer science Department of Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. She has 27 years of teaching experience and one year of industrial experience. Her areas of interest include Real Time Communication, Wireless Communication, Network Security and Cryptography. She has significant number of publications in peer reviewed International and National Journals. Life member of CSI, ISTE, WSEAS, AACE and ACRS.