International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 08, August 2016
www.ijates.com
ijates
ISSN 2348 - 7550

SURVEY OF VARIOUS APPROACHES OF TYPES OF INTRUSION DETECTION TECHNIQUES BASED ON DATA MINING

Y.S.Kalai Vani[1], Dr. M. Krishnamurthy[2]

[1]Asst. Professor, Dept. of Computer Science, Sindhi College of Commerce, Bangalore

[2]Professor & Head ME(CSE), Dept. of Computer Science & Engineering,KCG Institute of technology Chennai

ABSTRACT

Due to the widespread proliferation of computer networks, attacks on computer systems are increasing day by day. Preventive measures can stop these attacks to some extent, but they are not very effective due to various reasons. This lead to the development of intrusion detection as a second line of defense. Security of information system has become an important issue these days because of increasing number of public and private services used through the internet. Due to the increasing risk, Intrusion Detection Systems (IDS) have become a critical component to secure the systems and network. Data mining is a technique that provides higher automated capabilities to detect intrusion. Data mining techniques such as classification, clustering and association rule are used for intrusion detection. This review paper aims to present various data mining classification techniques and types of intrusion detection.

Keywords: Intrusion Detection (IDS), Networks, Public, Secure.

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 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. LITETATURE SURVEY

Interruption recognition framework assumes a vital part in distinguishing vindictive exercises in PC frameworks. The accompanying talks about the different terms identified with interruption discovery. Interruption is a kind of malevolent movement that tries to prevent the security viewpoints from claiming a PC framework.

It is characterized as any arrangement of activities that endeavors to trade off the trustworthiness, secrecy or accessibility of any asset. i) Data uprightness: It guarantees that the information being transmitted by the sender is not adjusted amid its transmission until it achieves the expected recipient. It keeps up and guarantees the precision and consistency of the information from its transmission to gathering. ii) Data privacy: It guarantees that the information being transmitted through the system is open to just those recipients why approved get the individual information should. It guarantees that the information has not been perused by unapproved clients. iii) Data accessibility: The system or a framework asset guarantees that the required information is open and usable by the approved framework clients upon interest or at whatever point they require it. Interruption identification is the procedure of observing and investigating the occasions happening in a PC framework keeping in mind the end goal to recognize noxious exercises occurring through the system. ID is a region developing in hugeness as more touchy information are put away and prepared in organized frameworks.

Interruption Detection framework is a mix of equipment and programming that recognizes interruptions in the system. IDS screen every one of the occasions gathering so as to occur in the system and breaking down data from different zones inside of the system. It distinguishes conceivable security breaks, which incorporate assaults from inside and outside the association and henceforth can recognize the indications of interruptions.

The primary goal of IDS is to alert the framework director at whatever point any suspicious action is recognized in the system. By and large, IDS makes two presumptions about the information set utilized as data for interruption recognition as tails: i) The measure of ordinary information surpasses the anomalous or assault information quantitatively. ii) The assault information varies from the typical information qualit1atively.

Mitchell and Deepali*[2]* presented a survey on clustering algorithms namely K-means, Y-means and Fuzzy C-means. Among these Fuzzy C-means can be considered as an efficient algorithm since it allows an item to belong to more than one cluster and measures the quality of partitioning. Fuzzy C-clustering also offers high detection rate and lower false positive rate for intrusion detection. Tripti and Khomlal *[3]* have presented the basic concepts and requirements for IDS.

Functional requirements and classification of IDS have also been presented. V. Jaiganesh, S. Mangayarkarasi and Dr. P. Sumathi *[4]* have proposed different types of intrusion detection system and highlights techniques of intrusion detection. In this paper various data mining algorithms for IDS implementation have been presented such as Support Vector Machine, kernelized support vector machine, extreme learning machine and kernelized extreme learning machine.

Manoranjan, Sateesh and Sudhir *[5]*have shown that neural network can be used as a method for training and learning IDS with classification rate of 100% and false positive rate of 0%. To classify traffic correctly neural network does not need huge amount of training data and detect unknown attacks. Amanpreet, Gaurav and Gulshan *[6]*have presented a survey on various types of intrusion, intrusion detection methodologies, data mining classification and clustering techniques. Sneh

2.1. Types of Intrusion detection systems.

A. 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.

B. 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.

2.2. Detection Techniques

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

A. 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.

B. 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.

III. RESEARCH BACKGROUND

3.1. Techniques for Intrusion Detection

Each malevolent action or assault has a particular example. The examples of just a percentage of the assaults are known while alternate assaults just demonstrate some deviation from the typical examples. In this way, the methods utilized for identifying interruptions depend on whether the examples of the assaults are known or obscure. The two principle strategies utilized are: A. Peculiarity Detection: It depends on the supposition that interruptions dependably mirror a few deviations from ordinary

examples. The ordinary condition of the system, activity load, breakdown, convention and bundle size are characterized by the framework chairman ahead of time. In this way, irregularity identifier thinks about the present condition of the system to the typical conduct and searches for vindictive conduct. It can recognize both known and obscure assaults. B. Abuse Detection: It depends on the information of known examples of past assaults and framework vulnerabilities.

3.2. Advantages and Disadvantages of Anomaly Detection and Misuse Detection

The fundamental weakness of abuse discovery methodologies is that they will distinguish just the assaults for which they are prepared to recognize. Novel assaults or obscure assaults or even variations of normal assaults regularly go undetected. The fundamental favorable position of irregularity identification methodologies is the capacity to recognize novel assaults or obscure assaults against programming frameworks, variations of known assaults, and deviations of typical utilization of projects paying little respect to whether the source is an advantaged inside client or an unapproved outer client. The burden of the oddity identification methodology is that understood assaults may not be distinguished, especially on the off chance that they fit the set up profile of the client. Once recognized, it is frequently hard to describe the way of the assault for criminological purposes. At last a high false positive rate may come about for a barely prepared identification calculation, or then again, a high false negative rate may come about for a comprehensively prepared abnormality discovery approach.[4]

3.3 Data Mining Approaches for Intrusion detection

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. Prediction makes 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)

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.

3.4 Need of Data Mining In Intrusion Detection:

Information Mining alludes to the procedure of removing covered up, beforehand obscure and valuable data from extensive databases. It is an advantageous method for separating examples and spotlights on issues identifying with their attainability, utility, effectiveness and versatility. Hence information mining procedures distinguish designs in the information set and utilize these examples to identify future interruptions in comparative information. The accompanying are a couple of particular things that make the utilization of information mining critical in an interruption discovery framework:

1. Manage firewall rules for irregularity recognition.
2. Analyze extensive volumes of system information.
3. Same information mining instrument can be connected to diverse information sources
4. Performs information synopsis and perception.v) Differentiates information that can be utilized for deviation examination.
5. Clusters the information into gatherings such that it have high intra-class likeness and low between class similitude.

IV. TYPES OF INTRUSION DETECTION SYSTEM

1. The MINDS System [36]

The Minnesota Intrusion Detection System (MINDS), uses data mining techniques to automatically detect attacks against computer networks and systems. While the long-term objective of MINDS is to address all aspects of intrusion detection, the system currently focuses on two specific issues

2. EMERALD (SRI) [35]

EMERALD is a software-based solution that utilizes lightweight sensors distributed over a network or series of networks for real-time detection ofanomalous or suspicious activity. EMERALD [12] sensors monitor activity both on host servers and network traffic streams. By using highly distributed surveillance and response monitors, EMERALD provides a wide range of information security coverage, real-time monitoring and response, protection of informational assets.

3. IDSs in the Open Market

Various systems that employ data mining techniques [15]have already been released as parts of commercial security packages.– Dshield,, RealSecure SiteProtectort.

V. 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 . This paper surveyed the different types of Intrusion detection system to find solution to the new emerging attacks.

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. Research Volume 3, Issue 4, April-2012
5. Mrs. SnehaKumari, Dr. ManeeshShrivastava "A Study Paper on IDS Attack Classification Using Various Data Mining Techniques" International
6. Journal of Advanced Computer Research Volume-2 Number-3 Issue-5 September-2012.
7. Mitchell D'silva, DeepaliVora Comparative Study of Data Mining Techniques to Enhance Intrusion Detection International Journal of Engineering
8. Research and Applications (IJERA) Vol. 3, Issue 1, January -February 2013. [3] S.A.Joshi, VarshaS.Pimprale Network Intrusion Detection System
9. (NIDS) based on Data Mining" International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013.
10. Reema Patel, AmitThakkar, AmitGanatra "A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems'
11. International Journal of Soft Computing and Engineering (IJSCE) Volume-2, Issue-1, March 2012.
12. AnkitaAgarwal Multi Agent Based Approach For Network Intrusion Detection Using Data Mining Concept Journal of Global Research in Computer Science, 3 (3), March 2012.
13. Miss. Prajkta P. Chapke& Prof. A.B. Raut Intrusion Detection System using Fuzzy logic and Data Mining Technique International Journal