ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

A Survey on Efficient Data Mining Techniques for Network Intrusion Detection System (IDS)

P.Kalarani[1], Dr.S. Selva Brunda[2]

Assistant Professor, Department of CT and IT, Kongu Arts and Science College, Erode, Tamil Nadu, India[1]

Professor and Head, Department of CSE, Sasurie College of Engineering, Erode, Tamil Nadu, India[2]

Abstract: Network security technology has become crucial in protecting government and industry computing infrastructure. Modern intrusion detection applications facing complex problems. These applications has to be require reliable, extensible, easy to manage, and have low maintenance cost. In recent years, data mining-based intrusion detection systems (IDSs) have demonstrated high accuracy, good generalization to novel types of intrusion, and robust behavior in a changing environment. Still, significant challenges exist in the design and implementation of production quality IDSs. Instrumenting components such as data transformations, model deployment, cooperative distributed detection and complex engineering endeavor. The IDS used data mining techniques for the network security, because to protect the network from various attacks and malicious traffic. This survey paper describes the Data mining approaches which are used to the detect intrusion in a network.

Key words: Anomaly detection, Data mining, Intrusion detection system, Misuse detection.

I. INTRODUCTION

Data mining is the process of discovering interesting patterns (or knowledge) from large amounts of data. The data sources can include databases, data warehouses, the Web, any other information repositories or data that are streamed into the system . Data mining is also called KDD (Knowledge Discovery in Databases). The goal of data mining is process is used to extract information from the dataset and it is changed into an understandable structure.

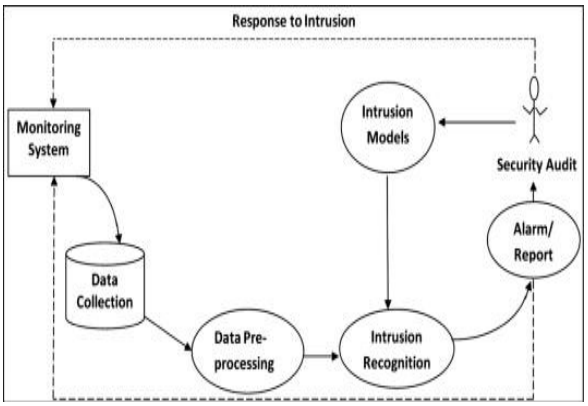

Fig.1 Overall structure of Intrusion Detection System.

The *intrusion detection* is software that automates the intrusion detection process. The intrusion has many types namely viruses, worms, Trojan horse, etc. The normal detection system like firewall, virtual private network are failed to detect some critical intrusions from the network.

The IDS finding the threats give response to the network administrator or user of the system and also raises alarms or signals when the security violations are occurred. In figure 1 describes how the intrusion detection can be processed. The information can be retrieved from the database, then it is checked by the firewall and finally it protected by the IDS after that it send the information to the corresponding network.

II. DETECTION TECHNIQUES

The components of intrusion detection system are: data source, analysis engine and response manager. The data source contains two categories namely host based and network based. The analysis engine gets the information from the data source and analyzes the attacks.

The response manager detects the attacks and gives response to the corresponding users. The Intrusion detection has two categories for detecting attacks in the network or host. They are:

a) Anomaly or Statistical detection
b) Misuse or Signature detection

Most of the researchers use these two techniques for detection rate and false alarm rate.

Anomaly Detection

This detection technique is comparing user's current behavior with usual behavior which is already stored in the database.

It used to detect the unknown attacks. By using statistical techniques to find patterns of activity that appears to be abnormal.

It is failed in high detection rate. But the goal of anomaly detection system is to find the intrusion in the system timely.

Misuse Detection

This type of detection technique system use patterns of well-known attacks of the system to match and identify known intrusions .The misuse detection system is ability to detect only the known attacks which already stored in the database and generate the fewer false alarm rates. The disadvantage of this system is unable to detect the newly invented attacks.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

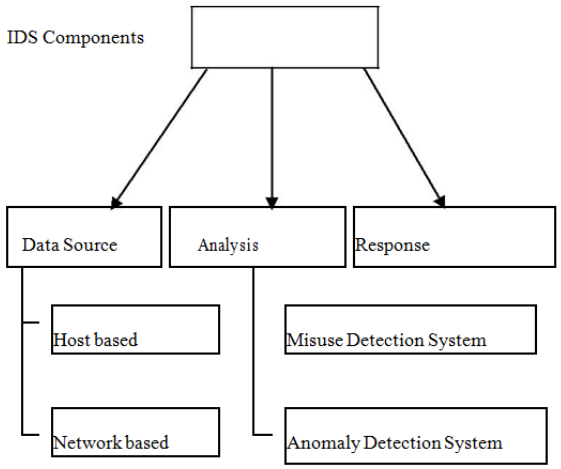

Fig. 2 Components of Intrusion Detection System

2.1 Data Mining meets Intrusion Detection

Anomaly Detection or Profile Matching, this technique is based on the normal behavior of a subject (e.g., a user or a system) any action that significantly deviates from the normal behavior is considered as an intrusive action. Misuse detection catches intrusions in terms of the characteristics of known attacks or system vulnerabilities any action that conforms to the pattern of a known attack or vulnerability is considered intrusive. The anomaly approach is focused on normal behaviors patterns. When a new kind of activity becomes acceptable (does not contradict to security policy), the normal behavior pattern database must be updated; otherwise the activity will be treated as an intrusion and will result in false positives. Attacks and deviations from normal activity are anomaly by definition and deserve the IDS user's attention.

Although anomaly detection can find out unknown patterns of attacks, it also suffers from several drawbacks. A general problem of all anomaly detection approaches, with the exception of the specification-based technique, is that the subject's normal behavior is modeled on the basis of the (audit) data collected over a period of normal operation. If undiscovered intrusive activities occur during this period, they will be taken as normal activities. In addition, because a subject's normal behavior usually changes over time (for example, a user's behavior may change when he moves from one project to another), the IDSs that use the above approach usually allow the subject's profile to gradually change. So, this gives an intruder the chance to gradually train the IDS and trick it into accepting intrusive activities as normal.

Also, because these approaches are all based on summarized information, they are insensitive to stealthy attacks. Because of some technical reasons, the current anomaly detection approaches usually suffer from a high false-alarm rate. Another difficult problem in building such models is how to decide the features to be used as the input of the models (e.g., the statistical models). In the existing models, the input parameters are generally decided by domain experts (e.g., network security experts) in ad hoc ways. So, it is not guaranteed that all the features related to intrusion detection will be selected as input parameters. Missing important intrusion-related features makes it difficult to distinguish attacks from normal activities, having non-intrusion-related features could introduce "noise" into the models and thus affect the detection performance. Misuse Detection or Signature Matching: Misuse detection is said to be complementary to anomaly detection. In misuse detection approach, firstly abnormal system behavior is defined, and then define any other behavior, as normal behavior.

Its main advantage is simplicity of adding known attacks to the model. Therefore, this systems look for well-defined patterns of known attacks or vulnerabilities. They can catch an intrusive activity even if it is so negligible that the anomaly detection approaches tend to ignore it. Attacks and deviations from normal behavior are taken as anomalies. The disadvantage of misuse detection is that it cannot detect novel or unknown attacks. As a result, the computer systems protected solely by misuse detection systems face the risk of being comprised without detecting the attacks. In addition, due to the requirement of explicit representation of attacks, the detection system requires the nature of the attacks to be well understood. It implies that human experts must work on the analysis and representation of attacks.

So, it is time consuming and error prone. Additionally, intrusion detection systems (IDSs) are categorized according to the kind of input information they analyze. This leads to the distinction between host-based and network-based IDSs. Host-based IDSs analyze host-bound audit sources such as operating system audit trails, system logs, or application logs. Network-based IDSs analyze network packets that are captured on a network.

III. APPLIED DATA MINING BASED INTRUSION DETECTION TECHNIQUES

Data mining look for hidden patterns and trends in data warehouse that is not immediately apparent from summarizing the data, and there is no query involved but use the concept interestingness criteria i.e specification of data such as Frequency, Rarity, Correlation, Length of occurrence, Consistency, Repeating/ periodicity, abnormal behavior, and other patters of interestingness. The algorithms which are used for intrusion detection based on data mining techniques are listed as follows

3.1 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. Association rule mining requires two thresholds i.e Minimum support and Minimum Confidence. Example: Apriori for mining Association rules Algorithm. [Agrawal and Srikant, 1994] [1] and [2]

3.2 Classification

Classification is the process of learning a function that maps data objects to a subset of a given class set. There

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

are two goals of classification, First finding a good general mapping that can predict the class of so far unkown data objectswith high accuracy. Second to find a compact and understandable class model for each other classes [1] and [2]

3.3 Clustering techniques

Clustering group's 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. [1], [2], [3]

3.4 Decision Tree

Decision tree initially builds a tree with classification. Each node represents a binary predicate on one attribute, one branch represents the positive instances of the predicate and the other branch represents the negative instances. Construction of Decision Tree does not require any domain knowledge and can handle high dimensional data. [3][4]

3.5 Genetic Algorithms

Method: learning examples are stored in relational database that are represented as relational tuples. It solves the problems with multiple solutions and easily transferred to existing models [3][4]

3.6 K -Nearest Neighbour

An object classification process is achieved by the majority vote of its neighbours. The object is being assigned to the class most common amongst its k nearest neighbours. If k=1, then the object is simply assigned to the class of its nearby neighbour. Its Implementation task is simple and Easy for parallel implementations.[2]

3.7 Support vector Machine

Method: A support vector machine is a classification and regression technique it constructs a hyper plane or set of hyper planes in a high or infinite dimensional space. It is able to model complex and nonlinear decision boundaries. [4]

3.8 Neural Network

Method: A Neural Network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Implicitly detect the complex nonlinear relationships between dependent and independent variables.

Highly tolerant the noisy data. Availability of multiple training algorithms.[3] [4]

3.9 Bayesian Method

Bayesian classifier based on the rules. It uses the joint probabilities of sample classes and observations. The algorithm tries to estimate the conditional probabilities of classes given an observation. Naïve Bayesian Classifier simplifies the computations It exhibit high accuracy and speed when applied to large databases.[4]

3.10 Fuzzy Logic

The Fuzzy logic has been used for both anomaly and misuse intrusion detection. It uses linguistic variables and allows imprecise inputs, permits fuzzy thresholds. Rule base or Fuzzy sets easily modified.[3][4]

4. A Review of Literature

This section discusses about various detection algorithms for network security.

[1] ***Network Intrusion Detection System based on Data Mining – S.A. Joshi, et. al.,***

In this paper the author discuss about the data mining algorithms 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

[2]Anomaly Detection in Network using Data mining Techniques – Sushil Kumar Chaturvedi, et.al

Given dataset is pre-processing and then the data can be partition into training and testing. The third stage the dataset is applied in C4.5 and SVM algorithm. The author of this paper compares these two algorithms and find out the detection rate comparison and false alarm rate comparison. By using these two data mining techniques they justify the C4.5 algorithm is better than the SVM.

[3]Application of Genetic Algorithm in Intrusion Detection System – Omprakash Chandrakar, et. al.,

This paper describes about basic concepts of network intrusion detection system, components and types of attacks. The IDS contains the three types of components namely data source, analysis engine, response manager. This paper gives the overview of genetic algorithm.

The genetic algorithm randomly selected the input (chromosome) and calculates the fitness value for each generated initial chromosome. The iteration has performed some specific operations namely sorting, selection, crossover, mutation and finally calculates the fitness value for chromosome.

[4] ***A Review of Intrusion Detection System in Computer Networks - Abhilasha A Sayar, et.al.,***

In this paper the author discuss about the classification of Intrusion detection system, advantageous and disadvantageous 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.

[6] ***A Survey on Intrusion Detection using Data Mining Techniques - R. Venkatesan, et al.,***

This paper describes the overview of the intrusion detection system and its each technique. The authors discuss pros and cons of anomaly detection and misuse detection. By combining these two categories and data mining approaches, then include the Apriori association rule algorithm for calculating the confidence levels. Apriori algorithm employs an iterative approach known as a level wise search, where k-item sets are used to explore (k + l)-item sets [5].

V. CONCLUSION

This survey paper study on various techniques which are used to detect the attacks from unknown users. The intrusion detection system components are useful to know about the process of detection. The IDS is combined with the data mining techniques and algorithms detect the threats and give immediate response to the user, and also find the percentage of detection rate, false alarm rate, and confidence level. There is a much research scope involved for the research community in this field to find the right kind of generalization of the IDS model .i.e it should not be neither too general nor too specific which is not Reliable. The challenges to find solution to the new emerging attacks with using current data mining based intrusion detection techniques in different fields is a new research domain.

ACKNOWLEDGMENT

I Give thanks to Almighty God to give an opportunity for doing research. And wish to acknowledge Dr.S.Selva Brunda., my Advisor and other known & unknown research scholars for their support and sharing ideas, views in their respective papers which really gave me an inspiration to do A Survey on Efficient Data Mining Techniques for Network Intrusion Detection System (IDS).

REFERENCES

[1] S.A.Joshi, Varsha S.Pimprale, "Network Intrusion Detection System (NIDS) based on Data Mining", International Journal of Engineering Science and Innovative Technology, Vol. 2, No. 1, January 2013, ISSN. 2319-5967.

[2] Sushil Kumar Chaturvedi, Prof. Vineet Richariya. Prof. Nirupama Tiwari, "Anomaly Detection in Network using Data mining Techniques", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 5, May 2012, ISSN. 2250-2459.

[3] Omprakash Chandrakar, Rekha Singh, Dr. Lal Bihari Barik, "Application of Genetic Algorithm in Intrusion Detection System", International Institute for Science, Technology and Education, Vol. 4, No. 1, 2014, ISSN. 2224-5774.

[4] A.R. Jakhale, G.A. Patil, "Anomaly Detection System by Mining Frequent Pattern using Data Mining Algorithm from Network Flow", International Journal of Engineering Research and Technology, Vol. 3, No.1, January 2014, ISSN. 2278-0181.

[5] R. Venkatesan, Dr. R. Ganesan, Dr. A. Arul Lawrence Selvakumar., "A Survey on Intrusion Detection using Data Mining Techniques", International Journal of Computers and Distributed Systems, Vol. 2, No. 1, December 2012, ISSN. 2278-5183.

[6] Abhilasha A Sayar, Sunil. N. Pawar, Vrushali Mane., "A Review of Intrusion Detection System in Computer Network", International Journal of Computer Science and Mobile Computing, Vol. 3, No. 2, February 2014, pp. 700 - 703.

[7] Monowar H. Bhuyan, Bhattacharyya DK, Kalita JK. An effective unsupervised network anomaly detection method. In: International conference on advances in computing, communications and informatics, no. 1; 2012. p. 533–9.

[8] R.Venkatesan, Dr.R.Ganeshan, Dr. A.Arul Lawrence Selvakumar "A Survey on Intrusion Detecion using Data Mining Techniques" in International journel of Computers and Distributed System, December 2012.

[9] Kamini Maheshwar and Divakar Singh "A Review of Data Mining based Intrusion Detection Techniques" in International Journal or Innovation in Engineering & Management (IJAIEM) Feb 2013.

[10] Harshna and NavneetKaur "Survey paper on Data Mining techniques of Intrusion Detecion " in International Journal of Science,Engineering and Technology Research (IJSETR), April 2013.

BIOGRAPHIES

P.KALARANI is Working as an Assistant professor in the Department of CT & IT in Kongu Arts and Science College, Erode, Tamilnade, India. Pursuing her research in Bharathiar University, Coimbatore.

DR.S. SELVA BRUNDA Working as a Professor and Head in the Department of Computer Science and Engineering in Sasurie College of Engineering, Erode, Tamilndu, India.