SURVEY OF NETWORK ANOMALY DETECTION USING MARKOV CHAIN

Brindasri S[1] and Saravanan k[2]

[1]M.E(CSE), Department of Computer Engineering, Erode Sengunthar Engineering college, Anna University(Chennai)
[2]Assistant Professor, Department of Computer Engineering, Erode Sengunthar Engineering college, Anna University(Chennai)

ABSTRACT

Recently an internet threat has been increased. Our motive is detect the intrusion in the network in concise. The real time issue such as DoS attack in banking, companies, industries and organization have been increased significantly IDS has been used in both server and host side. The major challenge is to effectively predict the periods of threats and protect the server from the unauthorized user. In this study, a novel probabilistic approach is proposed effectively to detect the network intrusions. It uses a Markov chain for probabilistic modelling of abnormal events in network systems. The degree of abnormality of the incoming data is performed on the basis of the network states.

KEYWORDS

Anomaly detection, K-Mean Algorithm, Data Clustering, ID3 Decision tree, Markovian chain.

1. INTRODUCTION

[1]Intrusion detection is referred as scanning, it used to scan and detect any intrusion occur in the system. It is a technology developed to assess the security of a computer system. Intrusions is the activities that violate the security policy in the system. The main security goals are confidentiality, integrity, availability, Possession of a computer or network[2]. IDS purpose is used to identify and report unauthorized user or unapproved network activities in the system There are 2approches are Anomaly and misuse detection. [2] Anomaly it attacks from outside the organization and misuse it attacks from within the organization. SBIDS also known as misuse detection that is signatures of known attacks is already stored and according to that it match the attacks which is already present in the stored data. It will signal an intrusion if a match is found. [3] The main drawback in this detection is that it cannot detect any new attacks whose signatures are unknown. This means that an IDS using misuse detection will only detect known attacks or attacks that are similar enough to a known attack to match its signature.[3] ABIDS is the Novelty detection is the identification of new or unknown attack in the testing process.

The current process in intrusion detection is to combine both host based and network based information to develop hybrid systems. Intrusion Detection System it combine both HIDS and NIDS to identify the attacks effectively. So In hybrid system both kinds of IDS is used to detect the attack simultaneously in the network.

DOI : 10.5121/ijcseit.2014.4105

Host based Intrusion Detection System (HIDS) is used to analyse the intrusions by analyzing system calls, application logs, and data modifications. It is used on personal computer to provide secure to the system host based Intrusion detection system used.

Network Intrusion Detection System (NIDS) are placed at strategic points within a network to monitor traffic from all devices on the network.

2. RELATED WORK

2.1 TAXONOMY OF ANOMALY BASED INTRUSION DETECTION SYSTEM

IDS aim at detecting attacks against information system. It is difficult to provide a secure information system and to prevent it for lifetime[2]. Maintenance of such system is difficult and costly. IDS have been used for this purpose. Two types of IDS are there. They are Signature Based IDS(SBIDS) and Anomaly Based IDS(ABIDS). ABIDS have been used to detect intrusion in a network efficiently.

2.2 A CLUSTERING-BASED METHOD FOR UNSUPERVISED INTRUSION DETECTIONS

Detection of intrusion attacks is an important issue in network security. A novel method [4] is proposed to compute the cluster radius threshold. The data classification is performed by an improved nearest neighbor (INN) method. A powerful clustering-based method is presented for the unsupervised intrusion detection (CBUID). The clustering based methods for the intrusion detection and prevention in the network by classify the states Low accuracy detection compare with existing system.

The parameter such as detection rate (DR), false alarm rate (FR) and detection rate for unknown attack types to measure performance of the intrusion detection methods. False Alarm Rate defined as the number of 'normal' patterns classified as attacks (False Positive) divided by the total number of 'normal' patterns. The detection rate is the ratio of the detected attack records based on the attacks in the original record.

The outlier factor of cluster is performed on the basis of the local deviation of a given data point. It used to compute the cluster radius threshold. The data classification has been performed by an improved nearest neighbour method. A powerful clustering based method has been used for the unsupervised intrusion detection.

2.3 AN EFFICIENT INTRUSIONS DETECTION BASED ON DECISION TREE CLASSIFIER USING FEATURE REDUCTION

Large computational value has always been a restraint in processing huge network intrusion data [5]. Dataset was available (KDD data set) for the purpose of intrusion detection. It contains 42 attributes and the classes are categorized into five main classes (one normal class and four main intrusion classes. Here the performance is based on correctly classified instances, incorrectly classified instances, kappa statistic, mean absolute error, root mean squared error, relative absolute error, root relative squared error and time .the comparison performed for 41 and 11 attributes. The four classifier models on the dataset were built and tested by means of 10-fold cross-validation, due to the method of information gain feature reduction it enable better accuracy of 99% and true positive rate is high with a value of one and false positive rate is low value is zero.j48 is better than other three classifier.

2.4 Intrusion Detection Systems Using Decision Trees and Support Vector Machines

Investigation and evaluation of intrusion detection mechanism by the tree data mining techniques [6] with the Support Vector Machines (SVM) was found. The Intrusion detection with Decision trees and SVM were tested with benchmark 1998 DARPA Intrusion Detection data set. The work enable the accuracy of the attack detection is low.

The Classification algorithm is inductively learned to construct a model from the pre-classified data set. Each data item is defined by values of the attributes. Classification may be viewed as mapping from a set of attributes to a particular class. The Decision tree classifies the given data item using the values of its attributes. The decision tree is initially constructed from a set of pre-classified data. The main approach is to select the attributes, which best divides the data items into their classes. The main problem here is deciding the attribute, which will best partition the data into various classes. The ID3 algorithm uses the information theoretic approach to solve this problem. Information theory uses the concept of entropy, which measures the impurity of a data items.

SVMs are able to handle only binary class classification problems. They divided the data into the two classes of “Normal” and “Attack” patterns, where the Attack is collection of four classes of attacks (Probe, DOS, U2R, and R2L). The classifier is learned from the training data and it is used on the test data to classify the data into normal or attack patterns. This process is repeated for all classes. The training time and testing times are also less for decision tree compared to the SVM.

2.5 An Implementation of Intrusion Detection System Using Genetic Algorithm

Intrusion Detection Systems has become a powerful component in day today life in network security. Intrusion Detection System (IDS), by applying genetic algorithm (GA)[7] to efficiently detect various types of network intrusions. GA uses evolution theory to information evolution in order to filter the traffic data and thus reduce the complexity. KDD benchmark dataset is used to measure the performance of the system and obtained reasonable detection rate.

Here both Genetic Algorithms (GAs) and Genetic Programming (GP) is used for detecting intrusion detection in the system. Some uses GA which is derived from the classification rules. GA is used to select required features and to determine the optimal and minimal parameters perform the operation in which different AI methods were used to derive acquisition of rules.

Problems are fidelity problem, resource usage problem, reliability problem. GA is used for solving various problems such as it include three impact factors that vitally impact on the effectiveness of the algorithm and application process. They are: i) the fitness function ii) the representation of individuals iii) the GA parameters- Crossover and mutation occurs in rest of the population which becomes the population of new generation. The process runs until the generation size comes down to 1 (one). And it obtain 0.9500 as the detection rate.

2.6 Evaluation Of Fuzzy K-Means And K-Means Clustering Algorithms

Clustering used to detect possible attacks in one of the branches of unsupervised learning[8]. Fuzzy sets reduces spurious alarms and intrusion detection. It uses KDDCUP99 data that are

collected based on DARPA innovation in 1998 for IDS designers that are used in several investigation to find attacks and intrusion.
These data are simulated in seven weeks for intrusion detection, KDDCUP99 data have 41 properties which are divided in to four parts Fundamental Properties, Content, Traffic property based on time, Traffic property based on the host**:** K-means algorithm acts better than fuzzy k-means in 66% of DOS attacks.

By comparing the average distance between the sum of data points belong to one cluster from center of that cluster ,Fuzzy K-Means and K-Means algorithm, K-Means clustering algorithm give better result but when choosing different mode of study Fuzzy K-Means gives better result.

2.7 A MARKOV CHAIN MODEL OF TEMPORAL BEHAVIOR FOR ANOMALY DETECTION

A Markov chain model is a discrete time series stochastic process specifies the random variable changes at discrete points in time [11]. Let y(t) be a random variable representing the state of a system at time t, where t=0,1,2.. This process satisfying the following two assumptions:

- The probability distribution of the previous state at time t+1 depends on the current state t, and it include dependent of time.
- A state transition from t to t+1 is independent of time

The Markov chain model is learned from historic data of the system's normal behaviour. The behaviour is analyzed to infer the probability of the process. A low probability indicates the abnormal activities and it clearly separates normal from intrusive activities. Capturing activities in network by network traffic data and audit data. Here it uses audit data from a UNIX Based Host Machine. The audit records contains a variety of information including the event type, user ID, IP address, and attacks. Here it extracted the event type characteristics of an audit event. The learning and inference algorithm of the Markov model for intrusion detection were implemented using c++. By using any threshold value it detects the attacks , it distinguishes normal activities from the attack activities with the 0% false alarm rate and the 100% detection rate.

2.9 COMPARISON TABLE FOR VARIOUS ALGORITHM AND TECHNIQUE

Table 1 shows comparison of various algorithm and technique used to find out intrusion detection.

Table 1 Comparison for various algorithm and technique

Algorithm (or) Techniques	Advantage	Disadvantage
K-Means	Simplest algorithm to solve clustering problem	It generates high false positive and negative
Fuzzy K-Means	Easy to find out the type of DoS attack which have occur	It creates bottle neck problem
Data Mining	It helps to prevent hacked IP address It identify false alarm generators	It creates bottle neck problem
HOP-COUNT	It used to detect whether hacking attempt is made	It does not identify type of attack
Genetic Algorithm (GA)	It used to filter the data Reduced the complexity Measure the fitness of the process False positive rate is low	It need to improve more statistical analysis Complex equation are needed

3. K-MEANS CLUSTERING ALGORITHM

K-Means is an unsupervised algorithm which is used for clustering. K-Means follows simple and easy way to classify given dataset. Here k centers take one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point it needs to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop it produces the k centers. The location of the k center is based on the changing of location.

3.1 K-MEANS ALGORITHM

The k-means algorithm is used for partitioning; each cluster is represented on the basis of mean value.

Method

Step 1:Choose k objects from D as the initial cluster centers
Step 2:Repeat Step 3
Step 3: Assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.

Step 4: Update the cluster means, i.e., calculate the mean value of the objects for each cluster.
Step 5: Until no change

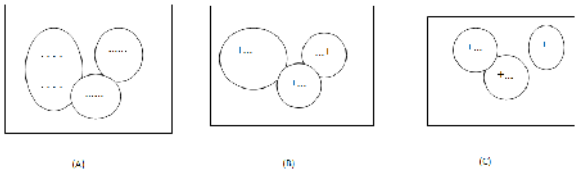

Fig. 1 Clustering a set of objects based on the k-means method

4. MARKOV CHAIN MODEL

A Markov chain is a continous process that predicts a change in the future using analysis on transitional characteristics from one state to another state. It describes a system with state transition probability and provides a powerful method for analyzing the operation of a system composed of finite states.

A Markov chain is a sequence of random variables X_1, X_2, X_3, ... with the Markov property, namely that, given the present state, the future and past states are independent. Formally,

$$\Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots \ldots, X_n = x_n) = \Pr(X_{n+1} = x | X_n = x_n)$$

The possible values of X_i form a countable S called the state space of the chain. Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states.

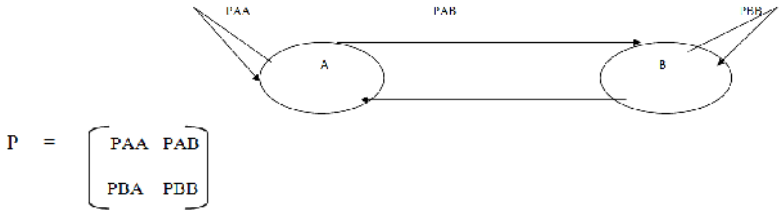

Fig. 2 Flow diagram of markov chain model

5. CONCLUSION

The existing system is implemented using k-means algorithm with Markov chain model. The system detects the attacks such as NIDS and HIDS. The rate of detection and rate of prevention measure is low. In future to improve the detection performance the work might be implemented by Eigen value and threshold values to detect the intrusion in network.

REFERENCES

[1] Poonam Dabas, Rashmi Chaudhary(2013), ”Survey Of Network Intrusion Detection Using K-Mean Algorithm”, International Journal Of Advanced Research In Computer Science And Software Engineering, Volume:3, Issue: 3,ISSN: 2277,pp. 30-35.

[2] Manasi Gyanchandani, Rana.J.L, Yadav.R.N(2012), ”Taxonomy of Anomaly Based Intrusion Detection System: A Review”, International Journal of Scientific and Research Publications, Volume:2, Issue:12,1 ISSN 2250-3153.

[3] Jyothsna.V, Rama Prasad.V.V, Munivara Prasad.K(2011), ”A Review of Anomaly based Intrusion Detection Systems”, International Journal of Computer Applications (0975 – 8887)Volume:28 No:7, pp: 283–304.

[4] Jiang, S., Song, X., Wang, H., Han, J., & Li, Q. (2006), ” A clustering-based method for unsupervised intrusion detections”, Pattern Recognitions Letter, 27(7), pp.802–810.

[5] Yogendra Kumar Jain , Upendra, ” An Efficient Intrusions Detection Based On Decision Tree Classifier Using Feature Reductions”, International Journal Of Scientific And Research Publications,Volume 2,Issue 1, January 2012 ISSN 2250-3153.

[6] Peddabachigari.S, Abraham.A & Thomas(2004), ”Intrusion detection systems using decision trees and support vector machines”, International Journal of Applied Science and Computations, pp. 118-134.

[7] Mohammad Sazzadul Hoque, Md. Abdul Mukit, Md. Abu Naser Bikas(2012), ”An Implementation Of Intrusion Detection System Using Genetic Algorithm”, International Journal Of Network Security & Its Applications (IJNSA), Volume:4, pp.109-120.

[8] Farhad Soleimanian Gharehchopogh, Neda Jabbari, Zeinab Ghaffari Azar(2012), ”Evaluation of Fuzzy K-Means And K-Means Clustering Algorithms In Intrusion Detection Systems”, International Journal Of Scientific & Technology Research Volume:1, Issue 11, ISSN 2277-8616 66, pp. 283–304.

[9] Doob, John Wiley & Sons, Dreger, Kreibich.C, Paxson.V, & Sommer.R (2005), ”Enhancing the accuracy of network-based intrusion detection with host-based context”, Lecture Notes in Computer Science, 3548,pp. 206–221.

[10] Mukkamala.S, Janoski.G & Sung.A.H(2002), ” Intrusion detection using neural networks and support vector machines”, In Proceedings of IEEE International Joint Conference On Neural Networks , pp. 1702–1707.

[11] Nong Ye(2000), ”A Markov Chain Model of Temporal Behavior for Anomaly Detection”, Proceedings of the 2000 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point,pp. 171-174.