Intrusion Detection System Using Hybrid Approach (Clustering and Classification)

Jay Kareliya[1] Mr. Krunal Panchal[2]
[1]M.E Student
[1,2]Department of Computer Engineering
[1,2]L.J. Institute of Engineering & Technology

Abstract— Now a day's security is the primary concerned in the field of computer science. Intrusion detection system provides stronger security services with the help of rules. Intrusion Detection System (IDS) has recently emerged as an important component for enhancing information system security. However, constructing and maintaining a misuse intrusion detection system for a network is labor– intensive, since attack scenarios and patterns need to be analyzed and categorized. Moreover, the rules corresponding to the scenarios and patterns need to be carefully hand-coded. In such situations, data mining can be used to ease this inconvenience. All most all-existing intrusion detection systems focus on attacks at low-level, and only produced isolated alerts. It is known that existing IDS can't find any type of logical relations among alerts. This research proposes an intrusion detection system that uses a combination of classification and clustering algorithms to detect intrusions. Basically this model work on misuse and anomaly detection mode, it will use an approach to extract features from arriving data packets and will apply the algorithm to get the rule for match normal and abnormal behavior. The main advantage of this approach is that the system can be trained with unlabeled data and is capable of detecting previously "unseen" attacks.

Key words: Classification, Clustering, K-means, Decision Table, Intrusion Detection System

I. INTRODUCTION

With online business more important now than in yesteryears, importance of securing data present on the systems accessible from the Internet is also increasing. If a system is compromised for even a small time, it could lead to huge losses to the organization [2]. Nowadays, computer networks are so complex, nearly everyone with a computer has connected it to the Internet to access information and transmit messages and as complexity increases the question of security becomes more and more familiar as well as the depth knowledge of computer network protocols namely. Intrusion detection is one of the aspects to resolve the problem of network security. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents. IDSs primarily focus on identifying possible incidents and detecting when an attacker has successfully compromised a system by exploiting vulnerability in the system Imperfectness of intrusion detection system has given an opportunity for data mining to make several important contributions to the field of intrusion detection. For controlling intrusion, intrusion detection systems are employed. The three important characteristics of intrusion detection systems are accuracy, extensibility and adaptability [3].Our aims to suggest a mechanism for detecting, unknown intrusions by identifying packets that are normal and to flag any packets that significantly deviate from the behaviour of these normal packets, these deviations are called anomaly or outlier [1]. The proposed Hybrid Intrusion detection system affects the execution performance and analysis of security [2].Clustering analysis is a common unsupervised anomaly detection method, and often used in Intrusion Detection System (IDS), which is an important component in the network security. The proposed approach applies clustering on all data into the corresponding group and after that applies a classifier for classification purpose. The proposed work will explore C4.5 and K-Means methods for intrusion detection and how it is useful for IDS. Traditional instance-based learning methods can only be used to detect known intrusions, since these methods classify instances based on what they have learned. They rarely detect new intrusions since these intrusion classes has not been able to detect new intrusions as well as known intrusions. Intrusion Detection System (IDS) using hybrid approach is getting popularity due to its adaptability to the changes in the behaviour of network traffic as it has the ability to detect the new attacks.

IDS Categories:

- Misuse detection systems
- Anomaly detection systems

A. *Misuse-Based Detection:*

Misuse detection systems model attacks as a specific pattern and are more useful in detecting known attack patterns. [6]

B. *Anomaly-Based Detection:*

In anomaly detection, the system administrator defines the baseline, or normal, state of the network, traffic load, breakdown, protocol, and typical packet size. [6].

Intrusion detection systems (IDSs) play a key role in detecting such malicious activities and enable administrators in securing network systems [9]. Two key criteria should be met by IDS for it to be effective: (i) ability to detect unknown attack types, (ii) having very less miss classification rate [9].

II. INTRUSION DETECTION SYSTEM IN DATA MINING

This section is going to be present general idea on a new proposed concept for intrusion detection system, which will enhance efficiency as compare existing intrusion detection system. The proposed concept is using data mining techniques [2]. Data mining techniques have been successfully applied in many different fields including

marketing, manufacturing, process control, fraud detection, and network management. We have already known that IDS most efficient technique against network attacks since they allow network administrator to detect policy violations [1].Hybrid intrusion detection systems comprise of misuse detection and anomaly detection systems that can detect both known and unknown intrusions [3].A new hybrid learning approach that combines clustering and classification algorithm. The best possible accuracy and detection rate can be achieved by using Hybrid learning approaches [4]. Hybrid IDS is still fresh topic that researcher work on it, Hence the aim of this research is to combine both algorithms that are signature based and anomaly based in order to enhance system security and help IDS user to make decision against detected attacks by the system [4]. Different classifiers can be used to form a hybrid learning approaches. But also combination of clustering and Classification technique [1] can be used in the process of hybridization, to achieve better results.[4]. View fact is that there is no perfect approach to avoid or protect intrusions from various events, it is very important to detect or identify them at the initial level of occurrence and take necessary or required actions for reducing or decreasing the likely damage. Most of the data mining techniques like clustering and classification have been functional on intrusion detection. In the proposed model we are using data mining concept. We will apply to data mining for anomaly detection field of intrusion detection. Anomaly detection approaches are capable to detect attacks with good accuracy and to achieve good detection rates.

III. LITERATURE SURVEY

Here proposed concept is going to be present general idea as showing in figure 1 for intrusion detection system, which will enhance efficiency as compare existing intrusion detection system [1]. The proposed concept is using data mining techniques. We will apply to data mining for anomaly detection field of intrusion detection. Anomaly detection approaches are capable to detect attacks with good accuracy and to achieve good detection rates. However, the rate of false alarm using anomaly approach is equally high. In order to maintain the good accuracy and detection rate while at the same time to lower down the false alarm rate [2],

We propose a concept which is the combination of 2 different techniques K-Mean clustering and c4.5 classification. For the first stage in the proposed hybrid IDS model, similar data instances will be grouped based on their behaviors by using a K-Means clustering as a pre-classification component[1][2][6]. For the second stage, C4.5 decision tree algorithm will be used which will create the nodes. At last Decision Table Majority rule based approach applied. Following is the proposed IDS, which divided into following module [2]:

A. Database Creation (Suggested Technique):

- Selecting and generating the data source (KDD 99')
- Data scope transformation and pre-processing

B. Data Mining Techniques:

- K-Means Cluster Technique
- C4.5 Classification
- Decision Table Majority Rule Base Approach

C. Proposed System:

- K-Mean with C4.5 and Decision Table Majority Rule Based Approach

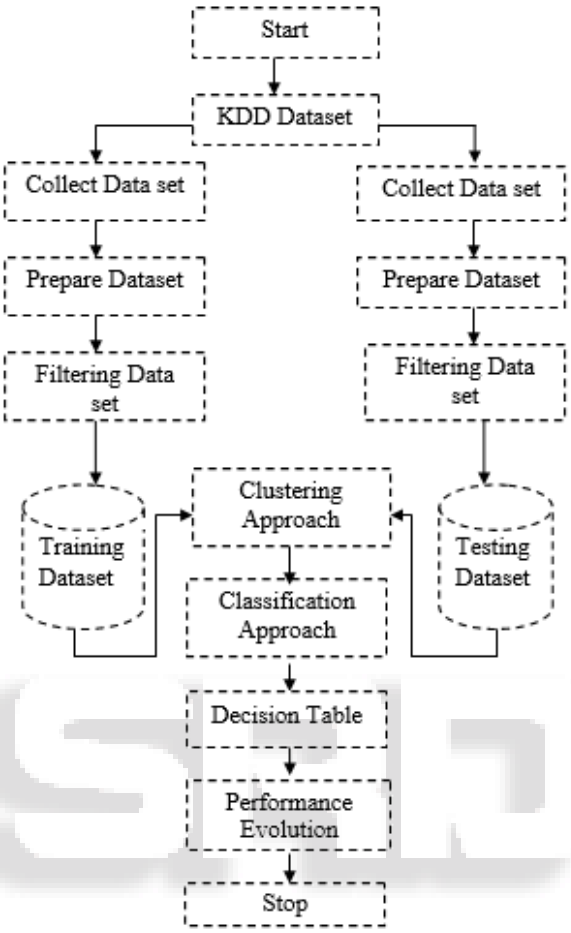

Fig. 1: Block diagram for propose concepts using combination of clustering and classification

1) Clustering:

Clustering is a division of data into groups of similar kind of objects. Each group or cluster contains objects that are similar among themselves but dissimilar with the others [2]. Clustering is an unsupervised learning because the class labels are not known [3]. A group of measurements and observations are done for the existence of the data in a cluster. Some clustering algorithms are: k-Means [3], Agglomerative Hierarchical clustering and classification and DBSCAN [6]. I use k-means clustering in this work.

2) Classification:

Classification is a data mining technique that maps data into predefined groups or classes [8]. It is a supervised learning method which requires labelled training data to generate rules for classifying test data into predetermined groups or classes.[3]. It is a two-phase process. The first phase is the learning phase, where the training data is analysed and classification rules are generated. The next phase is the classification, where test data is classified into classes according to the generated rules[8].It builds a classification model from the search domain and decides the class domain for each given object using one of the methods - k-nearest

neighbour [9],Naïve Bayes [6][9], Decision tree [1][3], and Support Vector Machine[5].

Among the cluster mining algorithms , K-Means is one of the most popular and well-known methods mainly used due to its simple concept , easy implementation and comprehensible mining result[18]

3) Decision Table:

Decision Table is one of the possible simplest hypothesis spaces, and usually they are easy to understand. A decision table is a managerial or encoding tool or technique for the demonstration of separate functions. This can be viewed as a matrix where the higher rows identify sets of circumstances and the lesser ones sets of events to be in use while the matching circumstances are fulfilled; thus each column, called a rule, describes a procedure of the type "if conditions, then actions". Given an unlabeled instance, decision table classifier searches for exact matches in the decision table using only the features in the schema (it is to be noted that there may be many matching instances in the table) [1]. If no instances are found, the majority class of the decision table is returned; otherwise, the majority class of all matching instances is returned. Decision Table is the technique of classifier, which is responsible for correct match of every attribute standards all to meet and thus remove the well-built independence conjecture.

IV. Evaluation Mesurement

During experiments, we will choose KDD Cup'99 benchmark dataset [1] which will suitable for us to evaluation and comparison.The data set contains 24 attack types. All these attacks fall into four main categories: DoS, U2R and R2L, Probe as follows [2][3].

A. Denial Of Service (Dos):

The attacker makes some computing resources too busy or memory resources too full to handle legitimate requests, or denies legitimate users access to a machine[2][8][9].
Examples are Apache2, Back, Land, Mailbomb, SYN Flood,Pingof death, Process table, Smurf, Teardropv[8][9].

B. Remote To Local (R2L):

The attacker who does not have an account on a remote machine sends packets to that machine over a network and exploits some vulnerability to gain local access as a user of that machine. Examples are Dictionary, Ftp_write, Guest, Imap, Named, Phf, Sendmail, Xlock [8][9].

C. User To Root (U2R):

In this type of attacks an attacker starts out with access to a normal user account on the system and is able to exploit system vulnerabilities to gain root access to the system. Examples are Eject, Loadmodule, Ps, Xterm, Perl, Fdformat [8][9].

D. Probing:

In this type of attacks an attacker scans a network of computers to gather information or find known vulnerabilities [3]. An attacker with a map of machines and services that are available on a network can use this information to look for exploits. Examples are Ipsweep, Mscan, Saint, Satan, Nmap[8][9].

To evolution we have selected some parameters to compare results between existing system and proposed system [2]. An Intrusion Detection System (IDS) requires high accuracy and detection rate as well as low false alarm rate [1]. In general, the performance of IDS is evaluated in term of accuracy, detection rate, and false alarm rate as in the following formula [1]:

- True positive (TP): number of malicious records that are correctly classified as intrusion.
- True negative (TN): number of legitimate records that are not classified as intrusion.
- False positive (FP): number of records that are incorrectly classified as attacks.
- False negative (FN): number of records that are incorrectly classified as legitimate activities.

- Accuracy = (TP+TN) / (TP+TN+FP+FN)
- Detection Rate = (TP) / (TP+FP)
- False Alarm = (FP) / (FP+TN)
- Classification Rate = $(TP+TN)$ / $(TP+TN+FP+FN)$

V. Conclusion

A presented approach is a hybrid approach which is a combination of clustering and classification. Presented Approach is a hybrid approach, which is the combination of K-mean, clustering, K-nearest and Decision Table Majority rule based approach. The proposed approach was compared and evaluated on KDD'99 dataset. Combining more than one data mining algorithms may be used to eliminate disadvantages of one another. C4.5 algorithm performs well in constructing decision trees and extracting rules from the dataset and K-Means is a well-known data-mining algorithm that has been used in attempt to detect anomalous user behavior as well as unusual behavior in network traffic.

So, it will give good result by giving high detection rate and low false alarm rate. The algorithm used in this work has the ability to detect unknown intrusion, in addition to known type intrusion.

Future research work should pay closer concentration or attention to the data mining process. Either more works should address the (semi-automatic) generation of high quality labeled training data, or the existence of such data should no longer be assumed. It involves improving the detection rates for unknown attacks.

Better anomaly methods may be used to improve this model as anomaly detection stands better chance of detecting unknown attacks. Intelligent algorithms and techniques are necessary to improve the performance of IDS.

References

[1] Aditi purohit, Hitesh Gupta " Hybrid Intrusion Detection System Model using Clustering, Classification and Decision Table" IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 9, Issue 4 (Mar. - Apr. 2013).

[2] Vasim Iqbal Memon, Gajendra Singh Chandel "A Design and Implementation of New Hybrid System for Anomaly Intrusion Detection System to Improve Efficiency" IJERA ISSN : 2248-9622, Vol. 4, Issue 5(Version1), May 2014.

[3] Om, H. and Kundu, A. "A hybrid system for reducing the false alarm rate of anomaly intrusion detection system" Recent Advances in Information Technology (RAIT), 1st IEEE International Conference on 15-17 March 2012.

[4] Virendra Barot and Durga Toshniwal "A New Data Mining Based Hybrid NetworkIntrusion Detection Model" IEEE 2012.

[5] Wang Pu and Wang Jun-qing "Intrusion Detection System with the Data Mining Technologies" IEEE 2011.

[6] Z. Muda, W. Yassin, M.N. Sulaiman and N.I. Udzir "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification" 7th IEEE International Conference on IT in Asia (CITA) 2011.

[7] Dewan M.D. Ferid, Nouria Harbi, "Combining Naïve Bayes and Decision Tree for Adaptive Intrusion detection", Intl.Journal of Network Security and Application (IJNSA), Vol-2, pp. 189-196, April 2010.

[8] Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul HonraoDe, "Predicting students' performance using ID3 and c4.5 classification algorithm", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.5, September 2013.

[9] R.Karthick,VipulP.Hattiwale,B.Ravindran,"Adapti ve Network Intrusion Detection System using a Hybrid Approach", IEEE 2012.

[10] M.Panda and M.Patra"NETWORK INTRUSION DETECTION USING NAÏVE BAYES" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[11] Jiawei Han, "Data Mining: Concepts and Techniques", Second Edition, China Machine Press, 2007.

[12] Mohammadreza Ektefa, Sara Memar , and Fatimah Sidi," Intrusion Detection Using Data Mining Techniques" , 978-1-4244-5651-2/10/ -2010 IEEE.

[13] YU-FANG ZHANG, ZHONG-YANG XIONG, XIU-QIONG WANG," Distributed Intrusion Detection Based On Clustering", 2005 IEEE.

[14] Mohammad Khubeb Siddiqui and Shams Naahid," Analysis of KDD CUP 99 Dataset using Clustering based Data Mining", International Journal of Database Theory and Application (IJDTA)Vol.6, No.5 -2013.

[15] SnehalA.Mulay,P.R.Devale, G.v. Garje," Intrusion Detection System using Support Vector Machine and Decision Tree", International Journal of Computer Applications (0975 - 8887)Volume 3 - No.3, June 2010.

[16] Dewan Md. Farid, Nouria Harbi, Suman Ahmmed, Md. Zahidur Rahman, and Chowdhury Mofizur Rahman," Mining Network Data for Intrusion Detection through Naïve Bayesian with Clustering International Scholarly and Scientific Research & Innovation Index Vol:4,No: 6 2010-06-20.

