IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 9, Issue 4 (Mar. - Apr. 2013), PP 103-107
www.iosrjournals.org

Hybrid Intrusion Detection System Model using Clustering, Classification and Decision Table

Aditi purohit[1], Hitesh Gupta[2]
[1](*M.tech Scholar Cse Department, PCST/ RGPV, India*)
[2](*Hod Cse Department, PCST/ RGPV, India*)

Abstract : *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; Transmission Control Protocol (TCP), Internet Protocol (IP) being the most common ones amongst others like User Datagram Protocol (UDP) etc...One of the two most publicized to security is intruder (other is virus) generally referred to as hacker or cracker. Intrusion detection systems are an important component of defensive measures protecting computer systems and networks from abuse. In this paper we proposes, a new hybrid learning approach that combines K-Mean clustering, Naive Bayes (statistical) also known as KMNB with Decision Table Majority (rule based) approaches. An experiment will be Carrie out to evaluate the performance of the proposed approach using KDD Cup '99 dataset. The experimental results will shows that new type of attack can be detected effectively in the system, so that the efficiency and accuracy of intrusion detection system will improve terms of detection rate as well false positive rate with reasonable prediction time.*

Keywords - Data mining, Intrusion detection System, Security, Protocol, Data Base.

I. INTRODUCTION

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; Transmission Control Protocol (TCP), Internet Protocol (IP) being the most common ones amongst others like User Datagram Protocol (UDP) etc...One of the two most publicized to security is intruder (other is virus) generally referred to as hacker or cracker. 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 behavior of these normal packets, these deviations are called anomaly or outlier. This mechanism can be visualized on a 2D topological map formulated by a Data Mining. This so called method of detection is named anomaly detection.

Rest of the paper are as follows: section II presents a brief survey where we show the previous research study on Intrusion Detection System and Problem Formulation, section III presents a proposed work methodology and proposed model. Finally, section IV presents concluding remarks The introduction of the paper should explain the nature of the problem, previous work, purpose, and the contribution of the paper. The contents of each section may be provided to understand easily about the paper.

II. LITERATURE SURVEY AND ANALYSIS

Intrusion Detection System (IDS) have become an important building block of any sound defense network infrastructure. Malicious attack have brought more adverse impact on the network than before increasing the need for effective approach to detect and identify such effects more effectively. Naive Bayes is one of the classification models that predicts very fast due to the less complexity functioning of it. Fast prediction is also the reason for a lot work done in recent years using Bayesian approach. In [1] a new hybrid model has suggested that ensembles Naive Bayes (statistical) and Decision Table Majority (rule based) approaches. In [2] authors have discussed on network security through Intrusion Detection Systems (IDSs). We have already known that IDS most efficient technique against network attacks since they allow network administrator to detect policy violations. However, traditional IDs are vulnerable to original and novel malicious attacks. Also, it is very inefficient to analyze from a large amount volume data such as possibility logs. In addition, there are high false positives and false negatives for the common lOSs. Furthermore in this paper authors have discussed also on data mining technique and how its help full in IDS system. Thus, how to integrate the data mining techniques into the intrusion detection systems has become a hot topic recently. Herr, authors presented the whole techniques of the IDS with data mining approaches in details. In [3] author discussed on Intrusion Detection System (IDS) where IDS is the most important technique to achieve higher security in detecting unknown\malicious\abnormal activities for a couple of years. Anomaly detection is one of intrusion detection system. Current anomaly detection is often associated with high false alarm with moderate accuracy and detection rates when it's unable to detect all types of attacks correctly. To overcome this problem, authors have suggested a hybrid learning approach. In this approach they have combine two different technique one is K-Means clustering and second is Naïve Bayes classification. In this authors have used clustering technique of all data into the corresponding group before applying a classifier for classification purpose. Authors have performed experiment using KDD Cup '99 dataset. Result show that the presented approach performed better in term of accuracy, detection rate with reasonable false alarm rate. In[4] an algorithm for adaptive network intrusion detection using naive Bayesian classifier and decision tree is presented, which performs balance detections and keeps false positives at acceptable level for different types of network attacks, and eliminates redundant attributes as well as contradictory examples from training data that make the detection model complex. The presented algorithm also addresses some difficulties of data mining such as handling continuous attribute, dealing with missing attribute values, and reducing noise in training data. Due to the large volumes of security audit data as well as the complex and dynamic properties of intrusion behaviors, several data mining based intrusion detection techniques have been applied to network-based traffic data and host-based data in the last decades. In[5] authors evaluated the performance of various rule based classifiers like: JRip, RIDOR, NNge and Decision Table using ensemble approach in order to build an efficient network intrusion detection system. they use KDDCup'99, intrusion detection benchmark dataset (which is a part of DARPA evaluation program) for experimentation. In[7] describe an adaptive network intrusion detection system, that uses a two stage architecture. In the first stage a probabilistic classifier is used to detect potential anomalies in the traffic. In the second stage a HMM based traffic model is used to narrow down the potential attack IP addresses. Various design choices that were made to make this system practical and difficulties faced in integrating with existing models are also described. In [8] presented a hybrid IDS by integrated signature based (Snort) with anomaly based (Naive Bayes) to enhance system security to detect attacks. This research used Knowledge Discovery Data Mining (KDD) CUP 99 dataset and Waikato Environment for Knowledge Analysis (WEKA) program for testing the proposed hybrid IDS. Accuracy, detection rate, time to build model and false alarm rate were used as parameters to evaluate performance between hybrid Snort with Naïve Bayes, Snort with J48graft and Snort with Bayes Network. In **[10]** author presented a hybrid intrusion detection system for wireless local area networks, based on Fuzzy logic. In this Hybrid Intrusion Detection system, anomaly detection is performed using the Bayesian network technique and misuse detection is performed using the Support Vector Machine (SVM) technique. The overall decision of system is performed by the fuzzy logic. For anomaly detection using Bayesian network, each node has a monitoring agent and a classifier within it for its detection and a mobile agent for information collection. The anomaly is measured based on the naïve Bayesian technique. For misuse detection using SVM, all the data that lie within the hyper plane are considered to be normal whereas the data that lie outside the hyper plane are considered to be intrusive. The outputs of both anomaly detection and misuse detection modules are applied by the fuzzy decision rules to perform the final decision making.

Problem Analysis: Traditional Intrusion Detection System (IDS) focus low –level attacks and only generate solated attacks to achieve higher security in detecting malicious activities for a couple of years. Anomaly detection is one of intrusion detection system. Current anomaly detection is often associated with high false alarm with moderate accuracy and detection rates when it's unable to detect all types of attacks correctly. Naive bayes is very fast in prediction as it processes training set only once to store statistics and use it to predict the unforeseen record, But suffers in performance due to independence attribute assumption. Decision Table Majority (DTM) is the classifier that is doing exact match of each attribute values all to gather and thus removes the strong independence assumption but very slow in processing, as it is comparing each test record against every test record. By use of universal hash table, the complexity proportionate to number of attributes. So with good subset of features, DTM can perform really well. Another thing we have observed that authors have used an association algorithm which is produced inefficient result for large data set in term of efficiency of executed algorithm.

III. PROPOSED WORK

This section we are going to be present proposed Intrusion Detection System model using efficient data mining approach. The proposed model will enhance efficiency for proposed intrusion detection system. The proposed model is based on combination of three different techniques K-Mean clustering, Naive Bayes (statistical) and Decision Table Majority (rule based) approaches for intrusion detection. There is different type of attack incorporated in the dataset which is fall into various categories. Table 1 is showing the details of attacks.

Table 1: Types of Attack

Type of Attacks	Attacks Category	Description	TCP/IP Layer Category
back	DoS	denial-of-service (fack Address generate)	Application Layer
land	DoS	denial-of-service (fack Address generate)	Transport Layer
buffer_overflow	U2R	unauthorized access to local superuser (root) privileges	Application Layer
ftp_write	R2L	unauthorized access from a remote machine	Application Layer
multihop	R2L	unauthorized access from a remote machine	Transport Layer
nmap	Probe	surveillance and other probing	Application Layer
portsweep	Probe	surveillance and other probing	Transport Layer
TCP SYN FLOOD	DoS	denial-of-service (fack Address generate)	Transport Layer

The objective of this paper is to provide comparative study of intrusion detection system using various techniques where we will show that our suggested technique will be produced batter result. The performance and strength of suggested technique will expected to be better than conventional technique of intrusion detection system and highly effective against attack.

Proposed Technique:

Here we are presenting general idea on a new hybrid model for intrusion detection system which will enhance efficiency as compare existing intrusion detection system. In the proposed model we are using data mining concept. Data mining techniques have been successfully applied in many different fields including marketing, manufacturing, process control, fraud detection, and network management. Over the past five years, a growing number of research techniques have applied data mining to various problems in intrusion detection. In this 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, We propose a model which is the combination of three different techniques K-Mean clustering, Naive Bayes (statistical) and Decision Table Majority (rule based) approaches shown in figure 1 For the first stage in the proposed hybrid IDS model, We will group similar data instances based on their behaviors by using a K-Means clustering as a pre-classification component. For the second stage, we will use Naïve Bayes classifier. In this we will classify the resulting clusters into classes like normal and abnormal. We will found that data that has been misclassified during the earlier stage may be correctly classified in the subsequent classification stage. Next Stage resultant of the classification process will pass on Decision Table Majority (DTM) for final evolution.

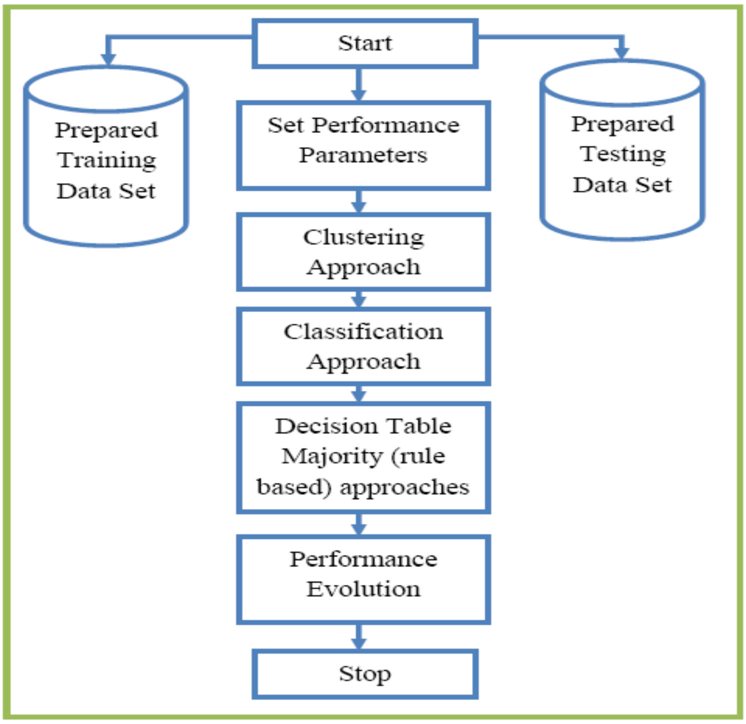
Start
Prepared Training Data Set
Set Performance Parameters
Prepared Testing Data Set
Clustering Approach
Classification Approach
Decision Table Majority (rule based) approaches
Performance Evolution
Stop

Figure 1: Proposed Hybrid IDS

Evaluation Measurement:

During experiments, we will choose KDD Cup'99 benchmark dataset [6] which will suitable for us to evaluation and comparison between the proposed approaches and the previous approaches. The entire data set will contain approximately 500,000 instances with 41 features because it's sufficient to calculation. The training dataset will contain 12-24 types of attack, while the testing data contains more than 10 types of additional attack. Our dataset will cover four major categories of attacks which is Probe, DoS, R2L and U2R. To evolution we have selected some parameters to compare results between existing system and proposed system. An Intrusion Detection System (IDS) requires high accuracy and detection rate as well as low false alarm rate. In general, the performance of IDS is evaluated in term of accuracy, detection rate, and false alarm rate as in the following formula:

- Accuracy = (TP+TN) / (TP+TN+FP+FN)
- Detection Rate = (TP) / (TP+FP)
- False Alarm = (FP) / (FP+TN)
- True positive (TP) when attack data detected as attack.
- True negative (TN) when normal data detected as normal.
- False positive (FP) when normal data detected as attack.
- False negative (FN) when attack data detected as normal.

On the basic of this table 2 is concluding result

Table 2: Packet Behaviors

Actual	**Predicted Normal**	**Predicted Attack**
Normal	TN	FP
Intrusions (attacks)	FN	TP

IV. CONCLUSION

This paper will be improve detecting speed and accuracy as a goal, and proposing more efficient associate and cluster rules mining method as comparing algorithm to abnormal detecting experiment based on network, and will improve the support and credit. In this paper, an hybrid model through combination of K-Means clustering, Naïve Bayes classifier and Decision Table Majority (rule based) approach is proposed. As we know that a naïve Bayesian network is a restricted network that has only two layers and assumes complete independence between the information nodes. This poses a limitation to this research work. In order to alleviate this problem so as to reduce the false positives, active platform or event based classification may be thought of using Bayesian network. We continue our work in this direction in order to build an efficient intrusion detection model. In Future we will implement our proposed model and will use KDD Cup '99 benchmark dataset in comparison between proposed and existing IDS.

REFERENCES

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

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

[3] 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.

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

[5] Joseph Derrick,Richard W. Tibbs, Larry Lee Reynolds "Investigating new approaches to data collection,management and analysis for network intrusion detection". In Proceeding of the 45th annual southesast regional conference, 2007. DOI = http://dl.acm.org/citation.cfm ?doid=1233341.1233392

[6] M.Panda, M. Patra, "Ensemble rule based classifiers for detecting network intrusion detection", in Int. Conference on Advances in Recent Technology in Communication and Computing, pp 19- 22,2009.

[7] R Rangadurai Karthick, Vipul P. Hattiwale and Balaraman Ravindran " Adaptive Network Intrusion Detection System using a Hybrid Approach" 2012 IEEE

[8] Safwan Mawlood Hussein, Fakariah Hani Mohd Ali and Zolidah Kasiran "Evaluation Effectiveness of Hybrid IDS Using Snort with Naïve Bayes to Detect Attacks" IEEE 2012

[9] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir "A K-Mean and Naïve Bayes Learning Approach for Bateer Intrusion detection system ", Information technology journal 2011 ISSN 1812-5638

[10] M.Moorthy and S.Sathiyabama "A Hybrid Data Mining based Intrusion Detection System for Wireless Local Area Networks" International Journal of Computer Applications (0975 – 8887)Volume 49– No.10, July 2012

[11] Vijay Katkar S. G. Bhirud "Novel DoS/DDoS Attack Detection and Signature Generation" International Journal of Computer Applications (0975 – 888)Volume 47– No.10, June 2012

[12] Mrutyunjaya Panda1 and Manas Ranjan Patra "NETWORK INTRUSION DETECTION USING NAÏVE BAYES" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007

[13] Skorupka, C., J. Tivel, L. Talbot, D. Debarr, W. Hill, E. Bloedorn, and A. Christiansen 2001. "Surf the Flood: Reducing High-Volume Intrusion Detection Data by Automated Record Aggregation," Proceedings of the SANS 2001 Technical Conference, Baltimore, MD.

[14] KDD. (1999). Available at http://kdd.ics.uci.edu/databases/ - kddcup99/kddcup99.html

[15] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, Classification and regression tres. Monterey, CA: Wadsworth & Books/Cole Advanced Boks & Software, 1984.

[16] http://www.webopedia.com/TERM/I/intrusion_detection _system.html

[17] LI Min "Application of Data Mining Techniques in Intrusion Detection" 2005