ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 1, January 2013

Network Intrusion Detection System (NIDS) based on Data Mining

S.A.Joshi, Varsha S.Pimprale
Sinhagd College of Engineering, Pune University

Abstract*— With the tremendous growth in information technology, network security is one of the challenging issue and so as Intrusion Detection system (IDS). IDS are an essential component of the network to be secured. The traditional IDS are unable to manage various newly arising attacks. To deal with these new problems of networks, data mining based IDS are opening new research avenues. Data mining is used to identify new patterns which were not known previously from large volume of network dataset. New Intrusion Detection Systems are based on sophisticated algorithms in spite of signature based detection. Data mining method uses binary classifiers and multiboosting simultaneously. Features are selected using binary classifiers for more accuracy in each type of attack. Multiboosting is used to reduce both the variance and bias. With data mining, it is easy to identify valid, useful and understandable pattern in large volume of data. Thus the efficiency and accuracy of Intrusion Detection system are increased and security of network so is also enhanced.

Index Terms— **NIDS, Data Mining, Feature Selection, Multiboosting.**

I. INTRODUCTION

With the rapid development in network technology during the past decade and the unprecedented growth of the Internet, people have become increasingly aware of the threats to personal privacy through computer crime. Attackers have become more sophisticated in the methodologies they use to intrude into the corporate networks [7].The importance of security of the computer networks will continue to increase as more business is conducted over the Internet [1]. Intrusion Detection is an area growing in relevance as more and more confidential data are stored and processed in networked systems. Intrusion Detection encompasses a range of security techniques designed to detect malicious system and network activity or to record evidence of Intrusion. Intrusion can be defined as “The act of thrusting in or of entering into a place or state without investigation, right or welcome” or it can be any unauthorized system or network activity on one or more of computers or networks [2]. Data mining can improve a network intrusion detection system by adding a new level of observation to detection of network data indifferences. Data mining provides an extra level of intrusion detection by identifying the boundaries for usual network activity so it can distinguish common activities from uncommon activities [10]. This paper is organized as follows: Section-II gives the idea behind the existing system and current scenario of computer networks. Section-III gives the details of the existing system. Finally the last section presents comparison of the scenario for IDS.

II. MOTIVATION

Almost all networks are protected by firewalls. However these firewalls are not always effective against the emerging intrusion attempts. Various methods based on Knowledge Development and Data mining which can help to improve Intrusion Detection Systems (IDSs) is Classification, Sequential Analysis, Time series Analysis, Prediction, Clustering and Association rules generation. Although data mining in intrusion detection is a fairly new method of maintaining network security, the data mining technique has been around for a long time and serves a variety of different purposes that involve both legitimate uses and malicious intentions by hackers who are trying to breach network security. KDD’99 (Knowledge Discovery in Databases) cup set is generally used for this purpose. KDD Cup is the leading Data Mining and Knowledge Discovery competition in the world, organized by ACM SIGKDD - Special Interest Group on Knowledge Discovery and Data Mining, the leading professional organization of data miners [13].

A. Detection Techniques

1**. *Signature-based detection techniques:*** In this method, IDS inspects the monitored packets on the basis of proof of attacks according to a predefined and existing model for specific known attacks. This method is capable to detect the known attacks and the disadvantage is it will work for predefined known attacks only. So if any new attack is invented it will give false or negative result.

2. *Anomaly-based detection techniques:* It does not require the prior knowledge of attacks and thus it can also detect new attacks. In this techniques, the IDS inspects the system activities on the basis of detection any deviation from existing model of normal and expected behaviour through the system.

According to the monitored system, the source of input information can be on a host or network or host and network. Thus IDS is further classified into three categories as follows [9]:

i. Network-based intrusion detection system (NIDS)

It is an independent platform that identifies intrusions by examining network traffic and monitors multiple hosts. Network intrusion detection systems gain access to network traffic by connecting to a network hub, network switch configured for port mirroring, or network tap.

ii. Host-based intrusion detection system (HIDS)

It consists of an agent on a host that identifies intrusions by analyzing system calls, application logs, file-system modifications (binaries, password files, capability databases, Access control lists, etc.) and other host activities and state. In a HIDS, sensors usually consist of a software agent.

iii. Hybrid Intrusion detection system (Hybrid IDS)

It complements HIDS system by the ability of monitoring the network traffic for a specific host; it is different from the NIDS that monitors all network traffic [1].

In computer security, a Network Intrusion Detection System (NIDS) is an intrusion detection system that attempts to discover unauthorized access to a computer network by analysing traffic on the network for signs of malicious activity.

B. Features Selection

Feature selection is an important asset in building classification model. Classification algorithm focuses on feature selection for improving the detection of attacks that occur infrequently in the training data and multiboosting for reducing both variance and bias. Elimination of useless features enhances the accuracy of detection which speeding up the computation. Features are to be extracted one at a time and testing the performance of a classification algorithm against the extracted features [3]. Feature selection contributes to improve the overall accuracy, reduces the number of false alarms and improves the detection of instances in the training data.

C. Multiboosting

It uses a decision committee technique that combines AdaBoost with the wagging. The effect of combining different classifiers can be explained with the theory of bias-variance decomposition [3]. Bias refers to an error due to a learning algorithm while variance refers to an error due to the learned model. Total expected error of a classifier is the sum of the bias and the variance. Multiboosting performs better than bagging [3].

III. EXISTING SYSTEM

A. *Data Set*

The data set that used for the existed system is KDD'99 cup version of the 1998 'Defence Advanced Research Projects Agency' (DARPA) intrusion evaluation dataset by MIT Lincoln Laboratory. Generally KDD'99 cup set is used for IDS.

B. *Types of attacks*

The training data set will contain 24 attack types that could be classified into four major categories as follows:

- Probing or Probe attack
- Denial-of-Service (DoS) attack
- User-to-Root (U2R) and
- Remote to Local (R2L) attack

For each record of KDD cup dataset different features are to be extracted.

C. *Existing Model*

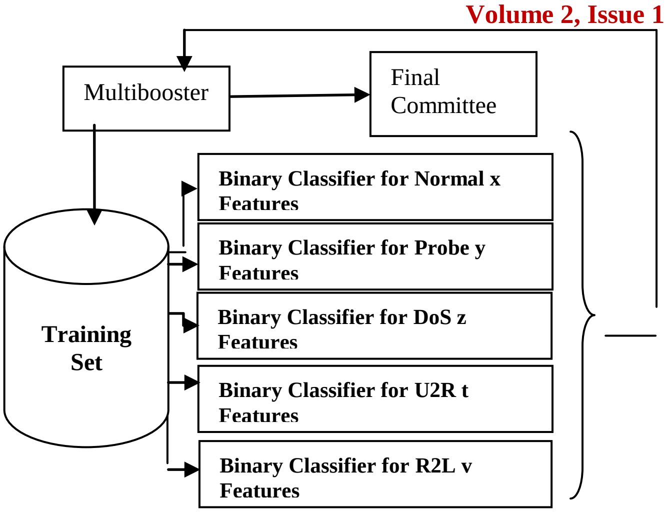

Multibooster
Final Committee
Training Set
Binary Classifier for Normal x Features
Binary Classifier for Probe y Features
Binary Classifier for DoS z Features
Binary Classifier for U2R t Features
Binary Classifier for R2L v Features

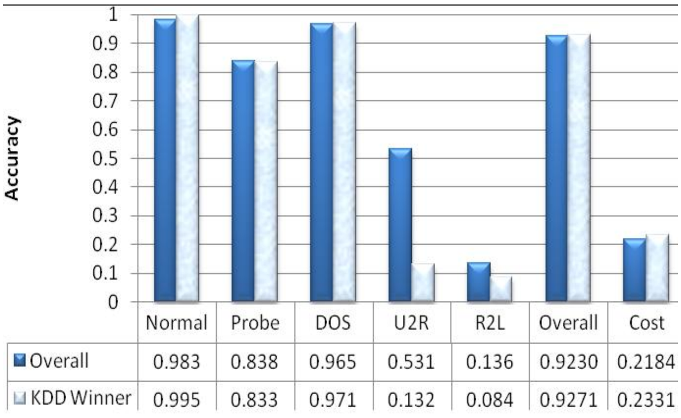
Fig 1: Existing System Model

The existed system is described as follows:

1. A sample training set is generated by multiboosting.
2. Binary classifiers use the relevant features for each class of the attack. Information gain or gain ratio is then calculated by selecting different features for each classifier. This is required to identify the normal and other class. This phase is the ensemble of the binary classifiers whose output will be then decided by the arbitrary function based on confidence level of each classifier.
3. Ensemble classifier is then used by multibooster to calculate if any error and generate the next training set.
4. Finally the committee is formed to be used with the existing system of IDS.

IV. CONCLUSION

With the rapid development in the field of computer security, data mining techniques are also one of the way by which security can be provided to the network. The existing system works with 94% accuracy using gain ratio and high detection rates for U2R and R2L attacks as shown in the figure-2. Also it works well than the winning entry of the KDD’99 cup in terms of cost and accuracy. To overcome the anomaly detection IDS are developed to detect unknown and known attacks, false positives and false negatives. Thus Data mining can help improve intrusion detection by addressing every mentioned problem. Data mining (DM), automatically searches large volumes of data for patterns using association rules efficiently. More work can be done using different algorithms of data mining like Naive Bayes, C4.5, ID3 on different datasets. The Future work will also extend this analysis to the PREDICT (Protected Repository for the Defense of Infrastructure against Cyber Threats) dataset or other network environments.



	Normal	Probe	DOS	U2R	R2L	Overall	Cost
■ Overall	0.983	0.838	0.965	0.531	0.136	0.9230	0.2184
□ KDD Winner	0.995	0.833	0.971	0.132	0.084	0.9271	0.2331

Fig 2: Performance of multiboosting + ensemble of binary classifiers with feature selection using information gain.

(Courtesy: D.Christine, Z.Wenjun et.al. "A New Data-Mining based Approach for Network Intrusion Detection", International Research Conference on Communication Networks and Services, Page no. 372-377, 2009)

ACKNOWLEDGMENT

I take this opportunity to express my gratitude to my Guide Prof.S.A.Joshi for her continuous support and valuable guidance. These inputs have helped me a lot.

REFERENCES

[1]. "Intrusion Alert – An Ethical Hacking guide to intrusion detection", by Ankit Fadia, Manu Zacharia, first edition, 2007.

[2]. Chetan R & Ashoka D.V., "Data Mining Based Network Intrusion Detection System: A Database Centric Approach ", International Conference on Computer Communication and Informatics, 2012.

[3]. D.Christine, Z.Wenjun et.al. "A New Data-Mining based Approach for Network Intrusion Detection", International Research Conference on Communication Networks and Services, Page no. 372-377, 2009.

[4]. Deepthy K Denatious, Anita John, "Survey on Data Mining Techniques to Enhance Intrusion Detection", International Conference on Computer Communication and Informatics, 2012.

[5]. M.Chunyu, C.Wei, "A study of Intrusion Detection System based on Data Mining", IEEE International Conference on Future Computer and Communication, Page no.186-189, 2010.

[6]. P. Prasenna, A.V.T RaghavRamana, "Network Programming and Mining Classifier For Intrusion Detection Using Probability Classification", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012.

[7]. Robert F. Erbacher, Michael J. Shevenell, "Design and Implementation of an Open Network and Host-Based Intrusion Detection Tested with an Emphasis on Accuracy and Repeatability", Ninth International Conference on Information Technology- New Generations, page no. 409-416,2012.

[8]. Sanjay Kumar Sharmai, Pankaj Pande/, Susheel Kumar Tiwari2 et.al., "An Improved Network Intrusion Detection Technique based on k-Means Clustering via NaIve Bayes Classification", IEEE-International Conference On Advances In Engineering, Science And Management, page no. 417-422, 2012.

[9]. X.Ming, Z.Changjun "Applied Research on Data Mining Algorithm in Network Intrusion Detection", International Joint Conference on Artificial Intelligence, Page no. 275-277, 2009.

[10]. Y.Qing, W.Xiaoping et.al. "An Intrusion Detection Approach based on Data Mining", IEEE International Conference on Future Computer and Communication, Volume-1, Page No.695-698, 2010.

[11]. Reema Patel, Amit Thakkar, et.al."A survey and comparative analysis of data mining for NIDS", IJSCE, ISSN: 2231-2307, volume-2, issue-1, and march-2012.

[12]. Z.Ali Othman, E.Eljadi, "Network Anomaly Detection Tools based on Association Rules", IEEE International Conference on Electrical and Informatics, 2011.

[13]. http://maths.anu.edu.au/~steve/pdcn.pdf.

[14]. http://en.wikipedia.org/wiki/Association_rule_learning.

[15]. www.sigkdd.org/kddcup/.

[16].http://www.sans.org/securityresources/idfaq/data_mining.php.