A STUDY ON INTRUSION DETECTION BASED ON KDDCUP'99 BENCHMARK DATASET

Pavan Kaur, Dr.Dinesh Kumar

***Abstract*— A hybrid model for feature selection and intrusion detection is important issue in intrusion detection. The selection of feature in attack attribute and normal traffic attribute is challenging task. The selection of known and unknown attack is also faced a problem of classification. There is multiclass problem during the classification of data. Intrusion detection is a problem of transportation infrastructure protection owing to the fact that computer networks are at the core of the operational control of much of the nation's transportation. The objective is to detect the Intrusion from network from different dataset using KNN, SVM and GA in Weka tool .A comparative analysis of different feature selection methods based on KDDCUP'99 benchmark dataset. The performances are evaluated in terms of detection rate, root mean square error and computational time.**

***Index Terms*— KNN, SVM, Attacks etc.**

I. INTRODUCTION

During the last few years there is a dramatic increase in growth of computer networks. There are various private as well as government organizations that store valuable data over the network. This tremendous growth has posed challenging issues in network and information security, and detection of security threats, commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. The security attacks can cause severe disruption to data and networks. Therefore, Intrusion Detection System (IDS) becomes an important part of every computer or network system. Intrusion detection (ID) is a mechanism that provides security for both computers and networks. Feature selection and feature reduction is important area of research in intrusion direction system. The size and attribute of intrusion file are very large. Due to large size of attribute the detection and classification mechanism of intrusion detection technique are compromised in terms of detection rate and alarm generation. For the improvement of intrusion detection process various authors and researchers work together for feature reduction and feature selection for intrusion detection system. In current scenario the feature reduction and selection process focus on entropy based technique[6]. Some authors used neural network model such SOM and RBF neural network model for classification of intrusion data during attacking mode and normal mode of network traffic. On the mechanism of detection intrusion detection divide into two section host based intrusion detection system and network based intrusion detection system. Host based intrusion detection system in generally know as signature based intrusion detection system. Instead signature based intrusion detection system come along with another variant is called anomaly based intrusion detection. In anomaly based intrusion detection various technique are used such as supervised learning and unsupervised learning. In network intrusion Detection, independent and redundancy attributes leads to low detecting rate and speed of classification algorithms. Therefore, how to reduce network attributes to raise performance of classification algorithms by applying optimal algorithm has become a research branch of intrusion Detection[8,9]. A new approach for network intrusion detection feature selection based on PCNN-SVM attribute selection and reduction is presented in the paper. The available approaches for intrusion detection focus on improving detection accuracy and restraining false alarms, and given enough time, most of them can achieve satisfactory results in terms of these criteria. However, in practice, intrusion detection is a real-time critical mission, that is, intrusions should be detected as soon as possible or at least before the attack eventually succeeds. In addition, there is usually an initial training period for an intrusion detector to characterize the observable object's behavior, and most existing methods are based on the assumption that high quality labeled training data are readily available. Present a new approach; based on pulse Coupled Neural Networks (PCNN) to identify important input features for intrusion detection. Through identifying the important inputs and redundant inputs, a classifier can achieve the reduced problem size, faster training and more accurate results[1,3]. Then, applied modified Gaussian Support Vector Machines (GSVMs) based on training algorithm to, anomaly detection over noisy data. GSVMs effectively address the over-fitting problem introduced by the noise in the training data set. With GSVMs, the incorporation of an averaging technique in the standard support vector machines makes the decision surface smoother and controls the amount of regularization automatically. Moreover, the training algorithm can significantly reduce training time with better generalization performance and fewer support vectors while maintaining high detection accuracy. They thus require less computational overhead and running time and so are more desirable for real time intrusion detection.

1.1 Intrusion Detection System

An intrusion is an attempt to compromise the integrity, confidentiality, availability of a resource, or to bypass the security mechanisms of a computer system or network. James Anderson introduced the concept of intrusion detection in 1980 [1].It monitors computer or network traffic and identify malicious activities that alerts the system or network

Manuscript received April 25, 2015

Pavan Kaur, M.Tech IT-Research scholar, Department of CSE, Guru Kashi University, Bathinda(pb)

Dr.Dinesh Kumar, Associate Professor, Department of CSE, Guru Kashi University, Bathinda(pb)

administrator against malicious attacks. Dorothy Denning proposed several models for IDS in 1987 [2]. Approaches of IDS based on detection are anomaly based and misuse based intrusion detection. In anomaly based intrusion detection approach [3], the system first learns the normal behavior or activity of the system or network to detect the intrusion. If the system deviates from its normal behavior then an alarm is produced. In misuse based intrusion detection approach [4], IDS monitors packets in the network and compares with stored attack patterns known as signatures. The main drawback is that there will be difference between the new threat discovered and signature being used in IDS for detecting the threat. Approaches of IDS based on location of monitoring are Network based intrusion detection system (NIDS) [5] and Host-based intrusion detection system (HIDS) [6]. NIDS detects intrusion by monitoring network traffic in terms of IP packet. HIDS are installed locally on host machines and detects intrusions by examining system calls, application logs, file system modification and other host activities made by each user on a particular machine.

1.2 Feature Selection

Due to the large amount of data flowing over the network real time intrusion detection is almost impossible. Feature selection can reduce the computation time and model complexity. Research on feature selection started in early 60s [9]. Basically feature selection is a technique of selecting a subset of relevant/important features by removing most irrelevant and redundant features [10] from the data for building an effective and efficient learning model [11].

II. LITERATURE SURVEY

Aditya Shrivastava1 et.al [2013] have proposed a hybrid model for feature selection and intrusion detection. Feature selection is important issue in intrusion detection. The selection of feature in attack attribute and normal traffic attribute is challenging task. The selection of known and unknown attack is also faced a problem of classification. PCNN is dynamic network used for the process of feature selection in classification. The dynamic nature of PCNN select attribute on selection of entropy. The attribute entropy is high the feature value of PCNN network is selected and the attribute value is low the PCNN feature selector reduces the value of feature selection. After selection of feature the Gaussian kernel of support vector machine is integrated for classification. Our detection rate is very high in compression of other neural network model such as RBF neural network and SOM network. For the empirical evaluation used KDDCUP99 dataset and measure detection rate precision and recall of proposed model.[1]

JAYSHRI R. PATEL et. al [2013] proposed a Decision Trees are considered to be one of the most popular approaches for representing classifier for various disciplines such as statistics, machine learning and data mining. Classification of Intrusion detection, according to their features into either intrusive or non intrusive class is a widely studied
problem. Decision trees are useful to detect intrusion from connection records. In this paper, we evaluate the performance of various decision tree classifiers for classifying intrusion detection data. The aim of this paper is to investigate the performance of various decision tree classifiers for ranked intrusion detection data. Information Gain is used to provide ranking to intrusion detection data. Decision tree classifiers evaluated areC4.5, CART, Random Forest and REP Tree. [2].

Megha Aggarwal et.al [2013], presented there is a dramatic increase in growth of computer networks. There are various private as well as government organizations that store valuable data over the network. This tremendous growth has posed challenging issues in network and information security, and detection of security threats, commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. The security attacks can cause severe disruption to data and networks. Therefore, Intrusion Detection System (IDS) becomes an important part of every computer or network system.Intrusion detection (ID) is a mechanism that provides security for both computers and networks. [3]

Venkata Suneetha Takkellapati et. al [2012] proposed As the cost of the data processing and Internet accessibility increases, more and more organizations are becoming vulnerable to a wide range of cyber threats. Most current offline intrusion detection systems are focused on unsupervised and supervised machine learning approaches. Existing model has high error rate during the attack classification using support vector machine learning algorithm. Besides, with the study of existing work, feature selection techniques are also essential to improve high efficiency and effectiveness. Performance of different types of attacks detection should also be improved and evaluated using the proposed approach. In this proposed system, Information Gain (IG) and Triangle Area based KNN are used for selecting more discriminative features by combining Greedy k-means clustering algorithm and SVM classifier to detect Network attacks. This system achieves high accuracy detection rate and less error rate of KDD CUP 1999 training data set. [4]

III. PROBLEM FORMULATION

A hybrid model for feature selection and intrusion detection is important issue in intrusion detection. The selection of feature in attack attribute and normal traffic attribute is challenging task. The selection of known and unknown attack is also faced a problem of classification. There is multiclass problem during the classification of data. Intrusion detection is a problem of transportation infrastructure protection owing to the fact that computer networks are at the core of the operational control of much of the nation's transportation. The feature ranking and selection problem for intrusion detection is similar in nature to various engineering problems that are characterized by: *f*

- Having a large number of input variables $x = (x_1, x_2, \ldots, x_n)$ of varying degrees of importance to the output y; i.e., some elements of x are essential, some are less important, some of them may not be mutually independent, and some may be useless or irrelevant (in determining the value of y) .*f*
- Lacking an analytical model that provides the basis for a mathematical formula that precisely describes the input output relationship, y = F (x).

- Having available a finite set of experimental data, based on which a model (e.g. neural networks) can be built for simulation and prediction purposes.

The major problem is given below:

1. Security
2. Authentication
3. Attackers

IV. OBJECTIVE

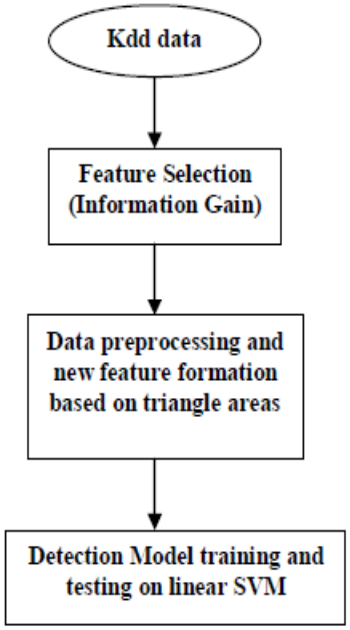
In today's era detection of security threats that are commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. Highly confidential data of various organizations are present over the network so in order to preserve that data from unauthorized users or attackers a strong security framework is required. Intrusion detection system plays a major role in providing security to computer networks.

Our Objectives are as follows:

- The objective is to detect the Intrusion from network from different dataset using KNN , SVM and GA in Weka tool
- A comparative analysis of different feature selection methods based on KDDCUP'99 benchmark dataset.
- To evaluate the performance are evaluated in terms of detection rate, root mean square error and computational time.
- The six feature selection algorithms are used to evaluate the performance.

V. METHODOLOGY

This work is to detect the intrusion from network. It is based upon weka tool. Their are the programmable files containing the information about the dataset. The Intrusion detection system deals with large amount of data which contains various irrelevant and redundant features resulting in increased processing time and low detection rate. Therefore feature selection plays an important role in intrusion detection. There are various feature selection methods proposed in literature by different authors. In this a comparative analysis of different feature selection methods are presented on KDDCUP'99 benchmark dataset and their performance are evaluated in terms of detection rate, root mean square error and computational time.



flow of the work

As the network environment has grown rapidly, so has the problem of intrusions. MIT kdd99 dataset is currently available approaches to dealing with intrusions can be categorized.

CONCLUSION

An intrusion is an attempt to compromise the integrity, confidentiality, availability of a resource, or to bypass the security mechanisms of a computer system or network. Some authors used neural network model such SOM and RBF neural network model for classification of intrusion data during attacking mode and normal mode of network traffic. On the mechanism of detection intrusion detection divide into two section host based intrusion detection system and network based intrusion detection system. Highly confidential data of various organizations are present over the network so in order to preserve that data from unauthorized users or attackers a strong security framework is required. Intrusion detection system plays a major role in providing security to computer networks. In this paper I reviwed different datasets and techniques and this work is further implem,ented to detect the Intrusion from network from different dataset using KNN , SVM and GA in Weka tool .A comparative analysis of different feature selection methods based on KDDCUP'99 benchmark dataset.

REFERENCES

[1]. Megha Aggarwal et.al " Performance Analysis Of Different Feature Selection Methods In Intrusion Detection" International Journal Of Scientific & Technology Research Volume 2, Issue 6, June 2013.

[2]. Aditya Shrivastava et.al " A Novel Hybrid Feature Selection and Intrusion Detection Based On PCNN and Support Vector Machine" Aditya Shrivastava et al, Int.J.Computer Technology & Applications,Vol 4 (6),922-927, IJCTA | Nov-Dec 2013.

[3]. Jayshri R. Patel et.al "Performance Evaluation Of Decision Tree Classifiers For Ranked Features Of Intrusion Detection " Journal Of Information, Knowledge And Research In Information Technology, ISSN: 0975 – 6698| NOV 12 TO OCT 13.

[4]. Venkata Suneetha Takkellapati et.al "Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine " International Journal of Engineering Trends and Technology- Volume3Issue4-2012.

[5]. Gong, S. (2011),‖ Feature Selection Method for Network Intrusion Based on GQPSO Attribute Reduction‖, International Conference on Multimedia Technology (ICMT), 6365 – 6368.

[6]. Nyguen, H. and Franke, K. et al.(2010)‖Improving effectiveness of intrusion detection by correlation feature selection‖, International conference on availability, reliability and security, 17-24.

[7]. Sridevi,R. and Chattemvelli ,R.(2012) ―Genetic algorithm and Artificial immune systems: A combinational approach for network intrusion detection International conference on advances in engineering, science and management (ICAESM-2012),494-498.

[8]. Wu, S.X. &Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. Applied Soft Computing Journal, 10, 1–35.