A Review on Data Mining Techniques for Network Based Intrusion Detection System

Deepak Kumar Rathore

Assistant professor, Department of CSE, LNCT, Bhopal (India)

rathore.rath@gmail.com

ABSTRACT
Data mining techniques have been successfully applied in the fields including marketing, manufacturing process control, fraud detection, and network management. Over the past two decade years, a growing number of research projects have applied data mining to intrusion detection. In this paper we presents a survey for the intrusion detection using various clustering and classification techniques, in addition also focus on some evolutionary and optimization methods to improve the detection rate for network.

Keywords: - Intrusion Detection System, KDD-CUP, Data Mining, Neural network, Support Vector machines.

INTRODUCTION

The balance between detection rate and false positive rate become more challenging when normal activity and anomalous activity are not static. The activity on the network can change and the IDS must be aware of this change and adapt accordingly. If not, the ability of the IDS to provide accurate and reliable results is greatly diminished. Therefore, an IDS must adapt to different environments, which potentially bring different activity and behavior unseen by the IDS [4].

The intrusion detection model based on the attribute-weighted clustering, as shown in below figure .The model first pre-treats collection of data, chooses training samples, reduces attributes in decision tables, produces reduced output rules to construct rule base of safe system and intrusion detection detector [10].

The initial intrusion model needs gradually perfect and improvement in subsequent studies to reach the best detection effect.

The functions for intrusion detection system such as Monitoring and analyzing both consumer and structure activities, Analyzing structure configurations and vulnerabilities, Assessing structure and file reliability, Ability to be familiar with patterns typical of attacks, Analysis of uncharacteristic activity patterns and Tracking user strategy violations.

Artificial neural network is an information processing model that is inspired by the biological nervous systems, such as brain, process information. It tries to represent the physical brain and thinking process through electronic circuit or software. Artificial neural network is the network of individual neurons. Each neuron is a neural network acts as an independent processing element. Each processing element (neuron) is fundamentally a summing element followed by an activation function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer.

Data mining works with both unsupervised and supervised techniques apart from these we can also used some classification techniques to improve the detection ratio and enhanced the performance of the system also increasing the accuracy. Support vector machine is another classification technique which classify the data and features into the plane i.e. hyper plane by reduce the margin of data in both mode such as linear mode and non-linear mode. Support vector machine is used as a super set of classifier techniques which generate the more accurate results than other classification techniques.

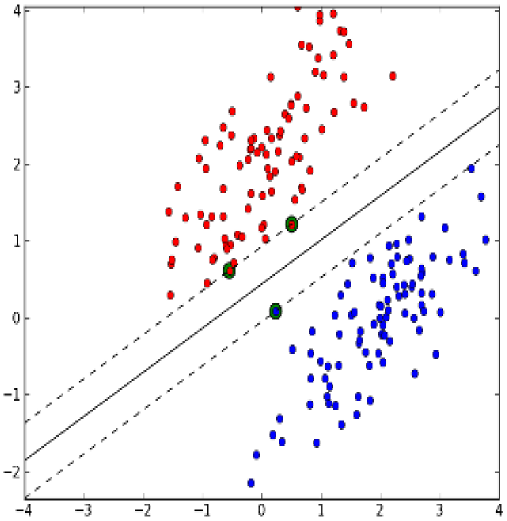

Fig 1: Support vector machine classification data.

Here Below diagram figure shows the percentage wise distribution of the research paper under various methodologies that are applied in the creations of IDS. The most commonly and widely applied approach is the hybrid approach.

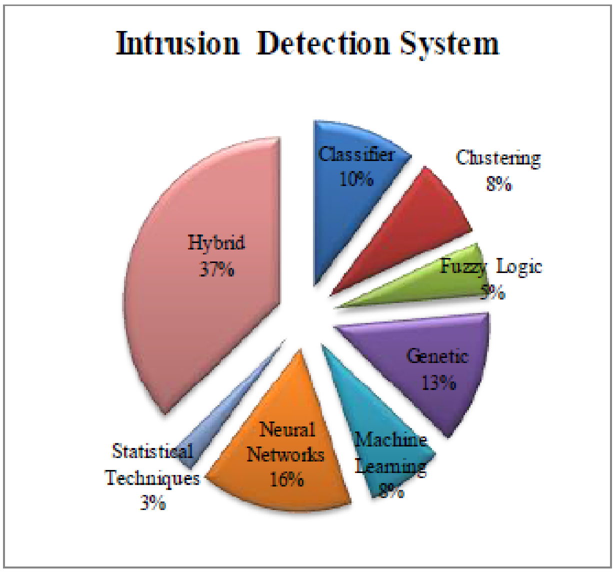

Fig 2: the percentage distribution of the number of papers under various IDS approaches.

The rest of this paper is organized as follows in the first section we describe a introduction of about Intrusion detection system and their techniques. in section II we discuss about the rich literature survey for the about Intrusion detection system and various researchers techniques for the same. In section III we discuss about the dataset and show a table for KDDCUP Features Dataset, finally in section IV we conclude the about our paper which is based on the literature survey and specify the future scope.

II RELATED WORK

In this section we discuss about the previous work done in the field of intrusion detection system using various techniques such as some classification techniques, evolutionary techniques and optimization methods, these sections further describe various author research work for the security of system and after that we formulate a problem for the solution in future work.

[1] In this paper, they aim to address this issue by proposing a simple Artificial Neural Network (ANN) based IDS model. The proposed IDS model uses the feed forward and the back propagation algorithms along with various other optimization techniques to minimize the overall computational overhead, while at the same time maintain a high performance level. Experimental results on the benchmark NSL-KDD dataset shows that the performance (accuracy and detection rate) of the proposed ANN based IDS model is at par and in some cases even better than other IDS models.

[3] This paper proposes an efficient intrusion detection architecture which named NIDERC (Network Intrusion Detection based on Ensemble Rough Classifiers). The NIDERC contains a new algorithm of attribute reduction which combined Rough Set Theory with Quantum Genetic Algorithm, a method of establishing multiple rough classifications and a process of identifying intrusion data. The experimental results illustrate the effectiveness of proposed architecture.

[4] In this paper, they implemented an Evolutionary General Regression Neural Network (E-GRNN) as a two-class classifier for intrusion

detection based on features of application layer protocols (e.g.,http, ftp, smtp, etc.) used in simulated network traffic activities. The E-GRNN is an evolutionary search-inspired General Regression Neural Network, which extracts the most salient features to reduce computational complexity and increase accuracy.
[5] This study proposed an SVM-based intrusion detection system, which combines a hierarchical clustering algorithm, a simple feature selection procedure, and the SVM technique. The hierarchical clustering algorithm provided the SVM with fewer, abstracted, and higher-qualified training instances that are derived from the KDD Cup 1999 training set. It was able to greatly shorten the training time, but also improve the performance of resultant SVM. The simple feature selection procedure was applied to eliminate unimportant features from the training set so the obtained SVM model could classify the network traffic data more accurately.

[8] In this work they adopted a two-stage alarm correlation technique to improve the accuracy of an IDS. The aim of the first stage is the reduction of the large volumes of detected alerts. They used two alternatives: SOM with k-means and Neural gas with FCM, in order to cluster low level alerts into meaningful partitions or meta-alerts that contain all alarms triggered by the same event in a certain time. Experimental results show that neural gas with FCM provides better clustering results.

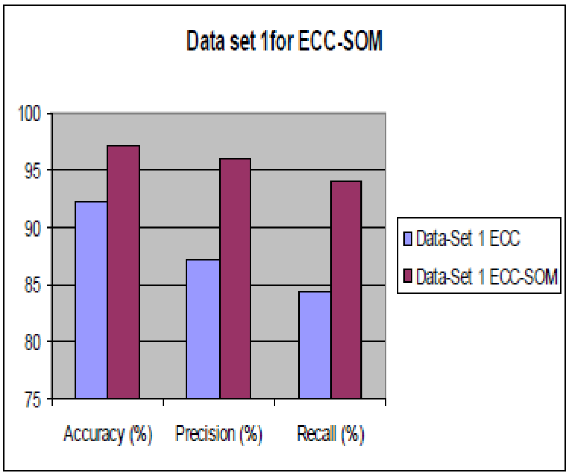

Fig 3: Intrusion detection system classification ratio using Ensemble cluster and classification techniques with self organizing network [13].

III KDDCUP
This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 the Fifth International Conference on Knowledge Discovery and Data Mining [4].

Four main class of attack	**categories of attack**
Denial of Service (DoS)	back, land, neptune, pod, smurt, teardrop
Remote to User (R2L)	ftp_write, guess_passwd, imap, multihop, phf,spy, warezclient, warezmaster
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Probing(Information Gathering)	ipsweep, nmap, portsweep, satan

Table 1: Classification of attack types [4].

IV CONCLUSIONS AND FUTURE WORK
Intrusion detection based upon computational intelligence is currently attracting considerable interest from the research community. In this paper we study various research paper for the intrusion detection system using various techniques and algorithms, in future we will develop a secure intrusion detection system mechanism whose create a difference between the normal and abnormal system activity and generate a more efficiently alert signal for the our system.

REFERENCES:-
[1] Basant Subba , Santosh Biswas, Sushanta Karmakar, “A Neural Network Based System for Intrusion Detection and Attack Classification,” IEEE 2016. pp. 1-6.

[2] Gaby Abou Haidar and Charbel Boustany, “High Perception Intrusion Detection Systems Using Neural Networks,” IEEE, 2015, pp. 497-501.

[3] SHEN Li, FENG Lin, "An Efficient Architecture for Network Intrusion Detection Based on Ensemble Rough Classifiers," The 8th International Conference on Computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka. pp. 1411-1415.

[4] James Brown, Mohd Anwar, Gerry Dozier, "An Evolutionary General Regression Neural Network Classifier for Intrusion Detection," 978-1-5090-2279-3/16, IEEE 2016. pp. 1-5.

[5] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai and Citra Dwi Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines," Elsevier, 2011, pp. 306-313.

[6] Asaf Shabtai, Uri Kanonov and Yuval Elovici, "Intrusion detection for mobile devices using the knowledge-based, temporal abstraction method," The Journal of Systems and Software, 2010, pp. 1524–1537.

[7] Gaby Abou Haidar and Charbel Boustany, "High Perception Intrusion Detection Systems Using Neural Networks," IEEE, 2015, pp. 497-501.

[8] Hachmi Fatma, Limam Mohamed, "A two-stage technique to improve intrusion detection systems based on data mining algorithms," IEEE 2013. pp. 1-6.

[9] Adil Fahad, Zahir Tari, Ibrahim Khalil, Ibrahim Habib, Hussein Alnuweiri, "Toward an efficient and scalable feature selection approach for internet traffic classification," Computer Network, Elsevier ltd. 2013. pp. 1-18.

[10] D.P.Gaikwad and Ravindra C. Thool, "Intrusion Detection System Using Bagging Ensemble Method of Machine Learning," IEEE, 2015, pp. 291-295.

[11] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Proceedings of the 2009 IEEE Symposium on computational intelligence in security and defense applications. pp. 1-6.

[12] Feng Du, "An Effective Pattern Matching Algorithm for Intrusion Detection," 2012 International Conference on Computer Science and Electronics Engineering, IEEE. pp. 34-38.

[13] Deepak Rathore, Prof. Anurag Jain, "Design Hybrid method for intrusion detection using Ensemble cluster classification and SOM network", International Journal of Advanced Computer Research , 2012.