Enhanced Network Intrusion Detection To Compute High Performance In Witty Environment.

Neha Suresh Sutar.
Department Of Information Technology
Siddhant College of Engineering, Sudumbare,
Pune, India.
neah141@gmail.com

Jyoti Pingalkar.
Department Of Information Technology
Siddhant College of Engineering, Sudumbare,
Pune, India.
jyoti.pingalkar@gmail.com

Abstract

With the huge use of internet on all fields, the network security has been rehabilitated into the key concept for lot of fiscal and dealing web applications. Intrusion Detection is very important for security in network. Failure of intrusion detection systems (IDS) has granted an occasion for data mining to make much important charity to the area of intrusion detection. In early stage, many researchers are using data mining approaches for creating IDS. Now, we propose a new concept by using data mining such as neural network-fuzzy clustering and radial basis support vector machine for serve IDS to reach superior detection rate. The proposed idea has four steps such as k-means clustering, Threat normalization, Neuro-Fuzzy logic and at last classification using radial Support Vector Machine. K-means clustering is used to create linear clusters which are produced from training dataset. After clustering we have to do Threat Normalization. Then, based on the produced clusters, different Neuro-fuzzy models are created. Classification using radial SVM is executed to detect intrusion has occurred or not. At the end, evaluating the performance using TP_Rate, FP_Rate, Precision, Recall and Accuracy for demonstration of the innovative approach, the result of proposed idea on KDD CUP 1999 dataset is established. The results shows that our new idea do enhanced detection than Conditional Random Fields (CRF), Naïve Bayes and K-Means Technique...

Index Terms—IDS, SVM, Neuro-Fuzzy, TP_Rate, FP_Rate, Precision, Recall, Accuracy

I. INTRODUCTION

Nodes that cannot communicate directly depend on their neighbors in order to forward their messages to the appropriate destination. Applications of mobile ad hoc networks have increased requirements in order to ensure high quality of service for the provided services. Security in such infrastructure-less networks has been proven to be a challenging task. Many security threats arise against mobile ad hoc networks, as they are inherently vulnerable due to the way the build and preserve connectivity characteristics. The open medium presents the network with the first and most serious vulnerability.

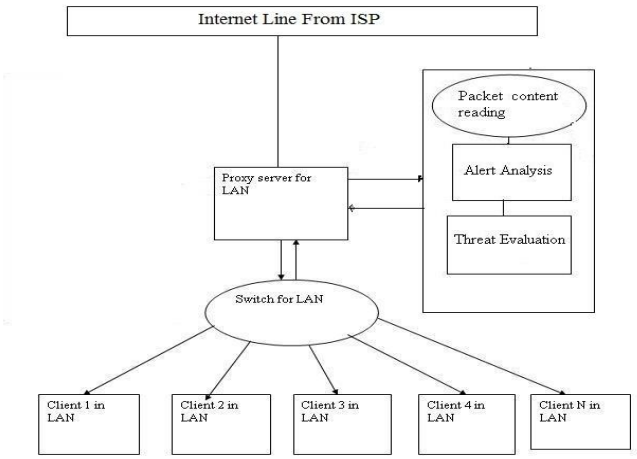

Fig. 1 Network Diagram of IDS

Unlike wired networks where an aggressor in order to launch an attack has to gain access to a wired infrastructure, firewalls and gateways, in ad hoc networks there is no clear line of defense. Every node is vulnerable and the good performance of the network depends on every node or at least on every node participating in a path from the source to a given destination.

Figure 1 shows that the network diagram for general threat evaluation, here the number of clients is connected to LAN which is connected by using switch. The switch is connected to the proxy server and proxy server will control the number of clients. The data coming from outside network like from internet service provider is connected to proxy server. Means whatever data coming or going in the network will go through proxy server [11]. Now Security issue arises that when and where we provide security to the data.

On sever side we will analyze the transceiver data, firstly read the contents of packet for analysis purpose. The packet contains the information about data, source, destination, type of service etc. Now we have to analyze that data whether normal or not. For this experiment we use KDD Cup 99 dataset which contains 41 attributes. We have to evaluate type of attack whether DOS, Probe, R2L, U2R and other data is normal data. Means we are grouping incoming data into five categories.
The core quandary of Neural Network-based IDS [5] exists in two concepts:

1. Weaker level detection firmness.

2. Low-grade level detection accuracy, especially for low-frequent attacks, e.g., Remote to Local (R2L), User to Root (U2R)

In good turn of the above two aspects, the main purpose is that the allotment of different types of attacks is avoidable. For low-frequent attacks, the receptiveness sample size is too small as compared to high-frequent attacks. It makes Neural Network not simple to study the lettering of these attacks and therefore detection accuracy is much worse.

To resolve the above two dilemma, we offer a description approach for Neural Network and Fuzzy clustering using radial Support Vector Machine (SVM) to the advance detection of accuracy for low-frequent attacks and detection loyalty.

[1]To get better result in the detection accuracy and detection constancy, many researches have been accomplished. Now in current stage, the research concentrate lies in using two methods such as rule-based specialist systems and statistical approaches. But when larger datasets comes, the results of rule-based proficient systems and statistical approaches become not as good as their need. Thus many of data mining techniques have been used to resolve the problem of intrusions. Out of these techniques, Artificial Neural Network (ANN)[6]is one of the largely used techniques and has been dominant in solving much amalgamated reasonable harm.

The common process of Neuro-Fuzzy using radial Support Machine (SVM) approach has the following four phases. In the first stage, K-means clustering which generate number of clusters for training phase. Then Threat normalization is performed on produced clusters. Then apply Neuro-Fuzzy logic on normalized data. At last stage use the classification using the radial support vector machine. Also comparing our result with conditional random fields (CRF) which shows our proposed idea gives better result than existing system.

II. Related work on Intrusion detection system

Intrusion Detection System (IDS) is fragment into two categories: misuse detection and anomaly detection systems. Misuse detection is used to recognize intrusions that compare known attack only. On other hand, anomaly detection system is an approach to investigate for malicious or abnormal behavior that pervert from already existing normal patterns. In this paper our interest is in to increase detection rate [2] and to reduce false alerts. To detect the intrusion form various concepts have been introduced and proposed over the current network analysis. Near the beginning stage, rule-based expert systems and statistical approaches are two distinctive ways to detect intrusion. A rule-based expert system Intrusion Detection System can detect known attack with higher detection rate, but it is very problematical to detect new coming attacks, and its signature database needs to[12] update physically. Statistical-based Intrusion Detection System hire various statistical concepts containing primary module analysis cluster and multivariate analysis, [9] Bayesian analysis and regularity and simple significance tests. However these type of Intrusion Detection System needs to gather sufficient data to produce a complex mathematical model, which is not practical in the case of complicated network data traffic.

To overcome these limitations, a number of data mining methods have been introduced. From these methods, Neural Network is one of the most used methods and has been successfully applied to intrusion detection. According to various types of Neural Network, these techniques can be classified into the following three categories: supervised Neural Network-based intrusion detection, unsupervised Neural Network based intrusion detection, and hybrid Neural Network based intrusion detection. Supervised Neural Network is applied to Intrusion Detection System generally contains multi-layer feed-forward (MLFF) neural networks and recurrent neural networks used MLFF neural networks for anomaly detection based on user activity behaviors. However in practically the number of training sets are very huge and the allocation of training set is unwarranted, the MLFF neural networks is simple to attain the local minimum and thus permanence is lower. Particularly, for low-frequent attacks, the detection precision is very low. Some of the researchers have judge against the efficiency of supervised Neural Network with other approaches such as support vector machine (SVM) and multivariate adaptive regression spines. Supervised Neural Network had been shown to have lower detection performance than SVM and multivariate adaptive regression spines. The second group uses unsupervised Neural Network to categorize input data and separate normal behaviors from abnormal or intrusive ones. Using unsupervised Neural Network in intrusion detection has many advantages. The key advantage is that unsupervised Neural Network can develop their analysis of new data without retraining. Fox was the first to submit an application a self-organizing map (SOM) to study the characteristics of normal system movement and classify statistical variations from the normal trends. Using supervised learning Neural Network, the performance of unsupervised Neural Network is also lower. Particularly for low-frequent attacks, unsupervised Neural Network also gets lower detection precision. The third group is hybrid Neural Network which integrates supervised Neural Network and unsupervised Neural Network, or to add Neural Network with other data mining methods to detect intrusion. The objective for using the hybrid Neural Network is to conquer the disadvantages of individual Neural Network.

Proposed employing a Neural Network and Fuzzy clustering using radial support vector machine for classifying the pattern of the attack or intrusion on the network. Neural Network for both visualizes intrusions using Kohenen's SOM and classify attacks using a elastic proliferation neural networks. The system gives usually better results than IDS based on RBF networks only. Proposed an intrusion detection technique based on evolutionary neural networks in order to find out the makeup and weights of the call sequences. The hybrid flexible neural-tree-based Intrusion Detection System [8] based on flexible neural tree, evolutionary algorithm and particle swarm optimization. For Neural Network based intrusion detection, hybrid ANN has been the leaning. But different ways to produce hybrid Neural Network [10] will highly power the performance of intrusion detection. Different hybrid Neural Network models should be properly created in order to give out different aims. The hybrid Neural Network, called neural network and fuzzy clustering, to resolve the two limitations of current Neural

Network-based Intrusion Detection System. i.e., lower detection precision for low-frequent attacks and weaker detection stability. Neural network and Fuzzy clustering approach introduces fuzzy clustering technique into ordinary Neural Network. Using fuzzy clustering technique, the whole training set can be divided into number of clusters which have limited size and lower complexity. Based on these clusters, the stability of individual Neural Network can be enhanced, the detection precision, especially for low-frequent attacks, can also be enhanced at last classify these data with radial SVM for detection purpose whether attack occurred or not.

III. EXPERIMENTAL WORK

In proposed idea we are presenting the whole framework of the new approach. After that we talk about the four main modules of neuro-fuzzy clustering using radial support vector machine (SVM), as k-means clustering module for building the different clusters for training phase, then we have to do Threat Normalization on produced clusters, apply neuro-fuzzzy logic,At last the classification using radial SVM to detect whether the given data is intruder or normal. Also we are mapping our proposed result with Conditional Random Fields (CRF) which will show the dtetction rate of neuro-Fuzzy using radial SVM is more and reduce false alert rate and accuracy of detection is increased. [13] The SVM vector contains attribute values obtained by giving input to each of the data through all of the neuro-fuzzy logic, and an supplementary attribute which has connection value of each of the data. At last step, classification is executed by using radial SVM[2] to detect intrusion has happened or not.

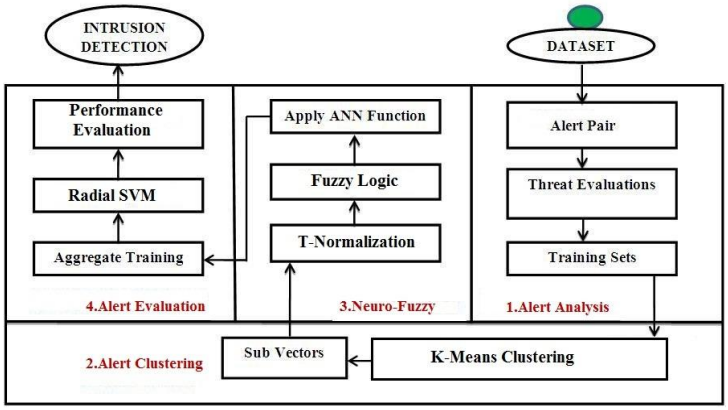

Fig.2 Working of Neuro-Fuzzy using Radial SVM to Calculate Performance

Experimental work contains four modules to find out intrusion is happened or not using neuro-fuzzy radial SVM as follows:

Module 1: Alert Analysis

The module 1 shows that we are giving dataset as input to module 1 which will generate alert pair on given dataset. Here we are using random sampling technique for taking data as training data from dataset. [7]When we take data for processing firstly we will read the contents of coming packet and generate alert pairs using minimum distance. After that we are evaluating the incoming data into five categories named by DOS, Probe, R2L, U2R and lastly remaining data as normal data by using following attributes

DOS: Duration, protocol_type, flag, src_bytes, count, dst_host_same_srv_rate ,dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate.
Probe: duration, protocol_type, service, flag, src_bytes.
R2L: duration, protocol_type, service, flag, src_bytes, hot, num_failed_logins, logged_in, num_compromised, num_file_creations, num_shells, num_access_files, is_host_login, is guest_login.
U2R: hot, num_compromised, root_shell, num_root,num_file_creations, num_shells, num_access_files, is_host_login.

By using these attributes we are finding the attack type of incoming data and other than this data as normal data.
Once we get the data we will pass to next module which is alert clustering.

Module 2: Alert Clustering

In this module we have to apply k-means algorithm on produced data from module 1 which convert these data into linear form using Euclidian Distance. Here we are finding the centroids of given data which will divide in to two homogeneous parts from non-homogeneous part. Repeat this step again and again till we don't have data. At last we will get the linear structured data for following processing. Through K-means clustering the training data is grouped into five groups wherein four groups will be a type intrusion or attack and one with normal data. Due to the size and complexity of every cluster is reduced by using k-means clustering.

Module 3: Neuro Fuzzy Module

The most important aim of our proposed idea is neuro-fuzzy logic module. The clusters which are obtained from alert cluster to normalize the threats from dataset. Now we have to apply fuzzy logic on normalized data to remove unwanted data from the clusters to reach out outcome. Fuzzy logic will mathematically do the computations on the given data. Whatever data produced from fuzzy logic will give input to neural network. Neural network is responsible for finding the pattern of attack. It will check the behavior of data and predict the pattern of attack or it is normal data. But the performance of neural network is not good as compare to support vector machine. For that reason we are using the neuro-fuzzy radial SVM which enhance the detection rate also and false alarms reduces.

Module 4: Alert Evaluation

Here on alert evaluation model we evaluate the TP, TN, FP, and FN from number of clusters which are coming from neuro fuzzy module. When the previous model data comes for evaluation we classify that data using radial support vector machine. Here we are finding the true positive, true negative, false positive, false negative from various clusters using the neuro-fuzzy radial SVM.
The evaluation of our proposed techniques is True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive shows the number of attacks

records that are correctly found as attack. Means true positive detection is properly detection of attack in IDS. True negative shows number of valid records correctly classified. Means IDS does not made mistake on detection. False positive shows incorrectly classified as attack in normal activity. And lastly false negative shows detect as normal activity whereas it is an attack. We have to enhance the accuracy level of detection.

IV. MATHEMATICAL MODEL

4.1 Set Theory

1. Let S = { }be as a System for Alert Identification and classification

2. Identify input as N
Where N = Input Document of KDD cup data set and Live data captured at router side
S={ N}

3. Identify A as Alert Identification and classification
S = { N,A}

4. Identify process P

S = {N, A, P}
P={Fc, At,Ts,Fa}
Where
Fc=Fuzzy Clustering
At=ANN Training
Ts=Training Set
Fa=Fuzzy Aggregation

5. S = { N,A, Fc, At,Ts,Fa]

4.2 Fuzzy clustering technique

Set F:
F0=Initializing Data Sets.
F1=calculating centers vectors
F2= Updating Vectors
F3=Creating Subset Vectors

4.3 ANN Training

Set A:
A0=Detecting pattern of every subset
A1=Feed-forward neural networks trained with the back- propagation algorithm to predict intrusion.
A2=Applying ANN Functions

4.4 Fuzzy Aggregation

Set G:
G0= Let the whole training set TR as data to input the every trained ANN_i and get the outputs:
G1=Form the input for new ANN:
G2=Train the new ANN. We can use Y_{input} as input and use the whole training set TR's class label as output to train the new ANN.

V. EXPERIMENTS AND RESULTS

To evaluate the performance of neuro- fuzzy using radial support vector machine approach, a series of experiments on KDD CUP 1990 dataset were carried out. In our proposed idea we implemented and evaluated the proposed method is in java on windows PC with core dual 1.83 GHz CPU and 512 MB.

5.1 Data preparation

The data set afforded for the 1999 KDD Cup was originally organized by MIT Lincoln labs for the 1998 Defense Advanced Research Projects Agency (DARPA) Intrusion Detection Evaluation Program, with the goal of evaluating research in intrusion detection, and it has become a target data set for the evaluation of Intrusion Detection Systems. It have approximately 49, 00,000 data connections information. Attacks fell in one of the four categories of following:
DOS-Denial [14] of Service denial of particular host service or continuously busy in other task (e.g. a mail bomb), R2L- try to get access to local machine from remote machine or unauthorized access from a remote machine (e.g. send mail), U2R- try to get access on remote machine means unauthorized access to super user or root functions (e.g. a buffer overflow attack), Probing inspection and other snooping for vulnerabilities (e.g. port scanning) [12]. In this paper we will use the subset of the original dataset which consist the separate records which are useful for our proposed idea. Data Preprocessing is the key task for reducing the attribute of KDD cup 1999 dataset. This process is approved out in two steps. The first step contains mapping symbolic-valued attributes to numeric valued attributes. In second step attributes are condensed by using Information expand. The KDD dataset has 41 attributes and after applying information expands approx. 21 attributes remain.

5.2 Evaluation strategy and Result

For our proposed experiments we are using KDD CUP 99 dataset. KDD CUP 99 contains 41 fields as an attributes and 42nd field as a label for attack type. In our idea we have taken selected features. The 42nd field can be comprehensive as Normal, DOS, Probing, U2R, and R2L. The performances of each approach are measured according to the Accuracy, Detection Rate and False Positive Rate using the following terms:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{True Positive Rate} = \frac{TP}{TP+FP}$$

$$\text{False Positive Rate} = \frac{FP}{FP+TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where,
FN is False Negative, TN is True Negative, TP is True Positive, and FP is False Positive
The detection rate is the number of intrusions detected by the IDS divided by the number of intrusions in the data set. The false positive rate is the number of normal activity that is misclassified as intrusions divided by the number of normal activities in the data set.

Following is the expected results showing high precision and recall in Fig 3 & Fig 4

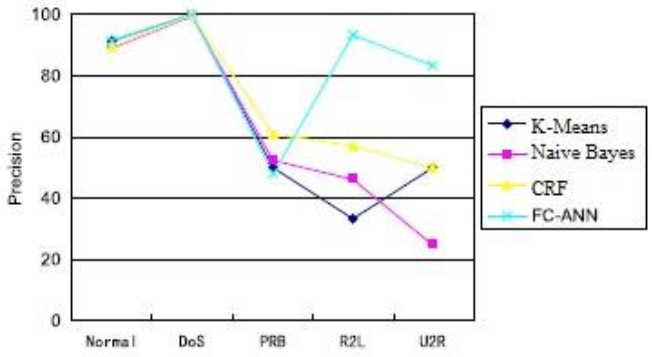

Fig 3 Precision result

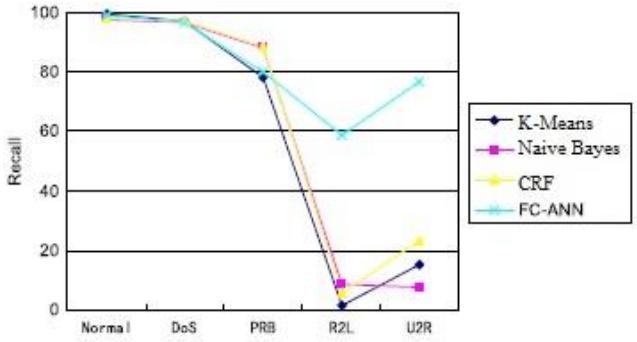

Fig 4 Recall result

VI. Conclusion & Future Work

Now days, the research on neural network approach and machine learning methods to enhance the network security by examining the activities of the network as well as that of threats is found in the quick force. The huge quantity of database is increasing quickly resulting in measured rise in the security attacks. The existing IDS is unsuccessful to modernize the audit data rapidly it occupy human intervention thus decrease the performances. The paper establishes the architecture of the Intrusion Detection System along with facial appearance of an ideal intrusion detection system. The study also describe the grouping and challenges if the IDS. In this paper we examine the neural network approach and the radial support vector machine learning technique in rise above the challenges of the IDS. Further there is need to plan the system which will limitations the current issues of IDS and also the system must offer a higher performance in detecting the threats and security intrusions.

VII. Acknowledgement

My sincere thanks go to Siddhant College Of Engineering for providing a strong platform to develop our skill and capabilities. I would like to thanks to my guide & respected teachers for their constant support and motivation. Last but not least, I would like to thanks all those who directly or indirectly help me in presenting the given paper.

VIII. Reference

[1] Yan Zhang, Shuguang Huang, Yongyi Wang "IDS Alert Classification Model Construction Using Decision Support Techniques, 2012 International Conference on Computer Science and Electronics Engineering.

[2] Annie George, "Anomaly Detection based on Machine Learning: Dimensionality Reduction using PCA and Classification using SVM', International Journal of Computer Applications" (0975 – 8887) Volume 47– No.21, June 2012.

[3] V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad, "A Review of Anomaly based Intrusion Detection Systems' International Journal of Computer Applications "(0975 – 8887) Volume 28– No.7, August 2011.

[4] Neethu B, "Classification of Intrusion Detection Dataset using machine learning Approaches' International Journal of Electronics and Computer Science Engineering "1044 ISSN-2277-1956. Available Online at www.ijecse.org.

[5] CHEN Bo, Ma Wu, "Research of Intrusion Detection based on Principal Components Analysis‖, Information Engineering Institute", Dalian University, China, Second International Conference on Information and Computing Science, 2009.

[6] T. J.Hastie, R. J.Tibshirani, and J. H.Friedman."The elements of statistical learning: Data mining, inference, and prediction", Springer-Verlag, 2001.

[7] R.Rifkin, A.Klautau "In defense of one-vs-all classification‖, Journal of Machine Learning Research", 5, pp.143-151, 2004.

[8] T. G.Dietterich, G.Bakiri. "Solving multiclass learning problems via error-correcting output codes‖, Journal of Artificial Intelligence Research" , 2, pp. 263-286, 1995.

[9] B. Pfahringer. "Winning the KDD99 Classification Cup: Bagged Boosting‖, SIGKDD Explorations", 1(2), pp.65-66, 2000.

[10] Xin Xu*, "Adaptive Intrusion Detection Based on Machine Learning: Feature Extraction, Classifier Construction and

Sequential Pattern Prediction,‘ International Journal of Web Services Practices”, Vol.2, No.1-2 (2006), pp. 49-58.

[11] A. Gardner, A. Krieger, G. Vachtsevanos, and B. Litt, “One-class novelty detection for seizure analysis from intracranial EEG,‖ J. Machine Learning Research (JMLR)”, vol. 7, pp. 1025–1044, Jun. 2006.

[12] Dayu Yang, Alexander Usynin, and J. Wesley Hines, “Anomaly-Based Intrusion Detection for SCADA Systems‖ IAEA Technical Meeting on Cyber Security of NPP I&C and Information systems”, Idaho Fall, ID, Oct.2006.

[13] J. Ma and S. Perkins, “Online novelty detection on temporal sequences‖ ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)”, Washington, DC, Aug. 2003.

[14] Dayu Yang, Alexander Usynin, and J. Wesley Hines, “Anomaly-Based Intrusion Detection for SCADA Systems‖ IAEA Technical Meeting on Cyber Security of NPP I&C and Information systems” , Idaho Fall, ID, Oct.2006.