Network Intrusion Detection using Neural Network Based Classifiers

[1]Ashalata Panigrahi and [2] Manas Ranjan Patra
Department of Computer Science, Berhampur University, Berhampur, India

Abstract: *Rapid expansion of computer networks throughout the world has made data security a major concern. In the recent past, there have been incidences of cyber-attacks which have put data at risk. Therefore, developing effective techniques to secure valuable data from such attacks is the need of the hour. Several intrusion detection techniques have been developed to deal with network attacks and raise alerts in a timely manner in order to mitigate the impact of such attacks. Among others, ANN methods can provide multilevel, multivariable security system to meet organizational needs. In this work, we have applied four prominent neural network based classification techniques, viz., Self-Organizing Map, Projective Adaptive Resonance Theory, Radial Basis Function Network, and Sequential Minimal Optimization to predict possible intrusive behavior of network users. The performance of these techniques have been evaluated in terms of accuracy, precision, recall / detection rate, F-Measure, and false alarm rate on the standard NSL-KDD intrusion dataset.*

Keywords: *Intrusion detection, ANN, Classification, SOM, PART, RBFN, SMO, Ant Search, Random Search*

I. INTRODUCTION

In the present world scenario, a variety of network based applications have been developed to provide services in different application areas such as banking, government, ecommerce, military, education, etc. With this there has been growing number of security threats, and thus there is need for securing data against different types of network attacks. Traditional intrusion prevention techniques such as firewalls, data encryption, access control have failed to fully protect networks and systems from increasingly sophisticated attacks and malwares. As a result, intrusion detection systems(IDS) have become indispensable component of any security infrastructure to minimize the impact of network attacks. The goal of Intrusion Detection System is to identify unusual access behavior and try to secure the system. Essentially, intrusion detection systems can be classified as misuse detection system and anomaly detection system. In a misuse-based IDS, attacks are represented as a pattern or a signature whereas anomaly based IDS identifies activities that deviate from the normal behavior of the monitored system and has the potential to detect unknown attacks.

Soft computing is an innovative approach to construct computationally intelligent systems consisting of artificial neural networks, fuzzy inference system, approximate reasoning and evolutionary computation. Gary Stein et al. [1] used Decision Tree classifier for Intrusion detection with GA based feature selection to improve the classification abilities of the decision tree classifier. GA based feature selection helps in selecting the best features for the decision tree classifiers. Mohammadreza Ektefa et al. [2] have shown the efficacy of C4.5 algorithm over SVM in detecting intrusions and false alarm rate. A hybrid feature selection approach called Genetic Quantum Particle Swarm Optimization (GQPSO) has been proposed in [3] which is capable of reducing redundant and irrelevant features while detecting intrusive behavior. This enhances the performance of intrusion detection as compared to PSO and QPSO algorithms. Yet another hybrid technique has been proposed in [4] that uses Intelligent Dynamic Swarm based Rough-Set (IDS-RS) for feature selection and Simplified Swarm Optimization (SSO) for attack classification. First, the most important features are extracted using IDS-RS, and then a novel Weighted Local Search (WLS) scheme is used to enhance the performance of SSO classifier. WLS is responsible for mining intrusion patterns to determine the appropriate solution based on the neighborhood of the present solution generated by SSO. This approach offers reasonably accuracy rate but it could be further improved. Koc, Mazzuchi & Sarkani [5] have integrated HNB with various discretization and feature selection methods to increase the accuracy rate and decrease the error rate. In HNB, the entire features are considered as independent or unbiased to each other. Each attribute in the HNB model has a hidden parent, in which the parameter on the training dataset estimates to unite the attributes that have relation with sequential minimal optimization. Fanping Zeng et al [6] have proposed a new anomaly detection method based on rough set reduction and HMM. In this work Hidden Markov model has been applied for detecting intrusions and it is proved to be a better tool. The proposed approach classifies and simplifies long observation sequence with the help of rough set and decision condition obtained by rough set reduction which could be applied for further detection. This is suitable for anomaly detection with high accuracy and low false alarms.

Neural network algorithms have emerged as artificial intelligence technique that can be applied to real-life problems. ANN can handle large number of weights and obtain the desire result. Here, the objective is to combine different neural network techniques to build intrusion detection models which can exhibit low false alarm rate and high detection rate.

II. METHODOLOGY

A. Self-Organizing Map (SOM)

The goal of Self-Organizing Map (SOM) [7] is to transform an input data set of arbitrary dimension to a one/two dimensional topological map. SOM is capable of discovering the feature map of an input data set by developing a topology preserving map that describes neighborhood relations of the data points [4]. The SOM array is essentially a fixed size grid of nodes. Here, the training is based on competitive learning, such that neuron with weight vector that is most similar to the input vector is adjusted towards the input vector and the neuron is said to be the "winning neuron" or the Best Matching Unit (BMU). Next, the weights of the neurons close to the winning neuron are also adjusted where the magnitude of change depends on the physical distance from the winning neuron.

B. Projective Adaptive Resonance Theory (PART)

Projective Adaptive Resonance Theory (PART) [8] is a new neural network architecture proposed to provide a solution to high-dimensional clustering problems. The architecture of PART is based on adaptive resonance theory (ART) which is very effective for self-organized clustering in full dimensional space. ART focuses on similarity of patterns in the full dimensional space but may fail to find patterns in subspaces of higher dimensional space.

C. Radial Basis Function Network (RBFN)

Radial basis function neural (RBFN) network [9] is a nonlinear hybrid network which contains an input each other throughout conditional mutual information. HNB significantly increased the accuracy rate of attacks compared to other leading methods, such as decision tree, neural network, and layer, a single hidden layer with a non-linear RBF activation function and an output layer. It is a simple network structure with better approximation and faster learning capabilities. The hidden layer uses Gaussian transfer function. The width and centers of the Gaussians are adjusted by unsupervised learning rules and supervised learning is applied to the output layer of the RBF network. The Gaussian functions respond only to a small region of the input space. The key to a successful implementation of these networks is to find suitable center for the Gaussian functions. Here, the simulation starts with the training of an unsupervised layer with an objective of deriving the Gaussian centers and the widths from the input data. These centers are encoded within the weights of the unsupervised layer using competitive learning. The advantage of the radial basis function network is that it finds the input to output map using local approximations. Usually the supervised segment is simply a linear combination of the approximations. Since linear combiners have few weights, these networks train extremely fast and require fewer training samples.

D. Sequential Minimal Optimization (SMO)

The SMO algorithm [10] is an optimization approach for the SVM quadratic program. It derives benefits from the sparse nature of the support vector problem and the simple nature of the constraints in the SVM Quadratic Programming (SVMQP) with a view to reduce each optimization step to its minimum form. Here, it is possible to break large QP problems into smallest possible QP problems which can be solved analytically. The memory requirement is linear in the training data set size, thus enabling SMO to handle very large training data sets.

III. THE PROPOSED MODEL

The prime objective of the proposed model is to develop an intrusion detection system with high detection and low false alarm rate. In order to achieve this four ANN based techniques such as SOM, PART, RBFN, and SMO have been employed. Two effective feature selection methods such as Ant search and Random search are applied on NSL-KDD dataset to select the most relevant features.

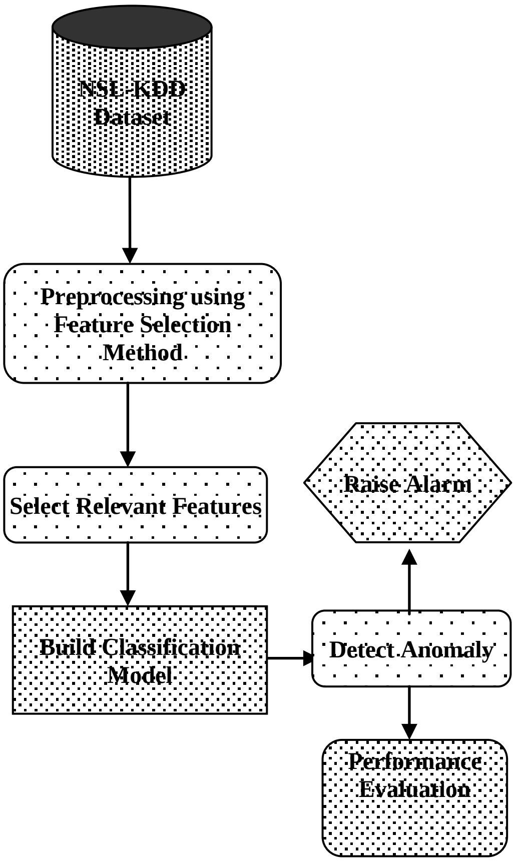

Fig. 1 Proposed Model

IV. EXPERIMENTAL SETUP

A. NSL-KDD Dataset

Experiments have been conducted using the well-known NSL- KDD dataset [11] which is a reduced version of the original KDD’99 dataset. The data set comprises of 41 attributes and a total of 125973 records, of which 67343 are normal and 58630 represent four different types of attacks as indicated in figure 2.

1) Denial of Service (DOS): Here, a computing resource is made too busy by an intruder so that it fails to process user requests.

2) User to Root (U2R): Here, an attacker enters into a system as a normal but eventually gain root access to the system by exploiting system vulnerabilities.

3) Remote to Local (R2L): In this type of attack an attacker who does not have an account on a remote machine sends packets to that machine over a network and exploits some vulnerability to illegally gain local access as a user of that machine

4) Probing: In this type of attack, an attacker scans a network of computers to gather information about system vulnerabilities and tries to exploit those to gain illegal access.

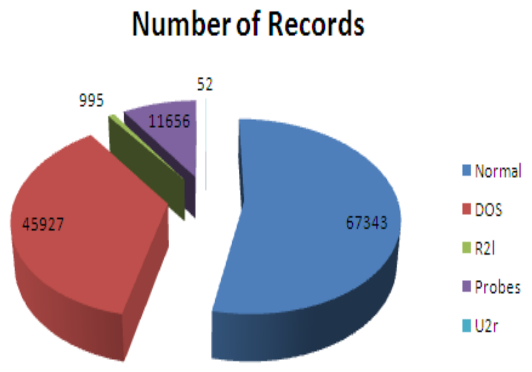

Fig. 2 Distribution of Records

B. Feature Selection

Feature selection identifies the optimal features of a dataset that could improve the performance of classification algorithms. For problems involving high dimensional feature space, there may be redundant or irrelevant features that would deteriorate the performance of classifiers. Thus, removing such redundant/irrelevant features has immense benefit. In our work, the two search based techniques namely, Ant search and Random search have been applied for selection of important features as described in Table I.

Table I: Selected features using feature selection methods

Feature Selection Method	No. of Features Selected	Feature Name
Ant Search	10	Flag, Src_bytes,Logg_in, R_shell, Se_se_rt, Sa_srv_rt, Di_srv_rt, Ds_Rate, Ds_d_h_rt, Ds_h_r
Random Search	12	Service, Flag, Src_bytes,Dst_bytes, Logg_in, R_shell, Count, Se_se_rt, Sa_srv_rt, Di_srv_rt, Ds_d_h_rt, Ds_h_r

C. Cross-Validation

Cross validation calculates the accuracy of the model by separating the data into two different populations, a training set and a testing set. We have adopted a 10-fold cross-validation process wherein the dataset is randomly partitioned into 10 mutually exclusive approximately equal size folds; T_1, T_2, …..,T_{10} out of which 9 folds are used to training and the remaining for testing. The process is repeated 10 times in a turn-taking manner by changing the folds. The 10 sets of results thus obtained are averaged to produce a single model estimation.

D. Confusion Matrix

The confusion matrix is a table with two rows and two columns that reports the number of False Positive, False Negative, True Positive, and True Negative. A confusion matrix can be used to evaluate the ability of an IDS to accurately predict whether an attempt to access a network is attack or not. The four possible scenarios are depicted in Table II.

We measure the performance of the model by computing the accuracy, detection rate, precision, F-value, false alarm rate, and fitness values as follows:

Accuracy measures the probability that one can correctly predict positive and negative scenarios:

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

Precision is a measure of the accuracy provided that a specific class has been predicted:

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

Detection rate measures the probability that one can correctly predict positive scenarios:

$$\text{Detection Rate or Recall} = \frac{TP}{TP + FN}$$

F- value is the harmonic mean of Precision and Recall which measures the quality of classification:

$$\text{F - Value} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Table II: Confusion matrix for IDS

		Predicted Class	
		Negative Class (Normal)	***Positive Class (Attack)***
Actual Class	***Negative Class(Normal)***	True Negative (TN)	False Positive (FP)
	Positive Class (Attack)	False Negative (FN)	True Positive (TP)

False Alarm Rate is the ratio of the number of normal instances incorrectly labelled as intrusion and the total number of normal instances:

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

V. RESULTS AND DISCUSSION

The aim of this paper was to develop intrusion detection models so as to protect computer networks from illegal access. Towards this end, we have employed four ANN based classification techniques to build an effective intrusion detection model. While doing so the performance of each of the techniques was measured in terms accuracy, precision, detection rate, F-value, and false alarm rate. The values are represented in Table III.

It is clearly observed that PART classification technique with random search gives the highest accuracy of 99.8381%, highest detection rate of 99.8107%, and low false alarm rate of 0.1381. These results suggest that PART classification technique outperforms other three techniques, thus qualify to be a potential candidate for building and effective IDS.

Table III: Performance Comparison of Neural Network based Classifiers

Feature Selection Method used	**Classification Techniques used**	**Evaluation criteria used**				
		Accuracy in %	Precision in %	Detection Rate / Recall in %	F-Value in %	False Alarm Rate in %
Ant Search	SOM	85.8446	78.2802	96.3141	86.3658	23.267
	PART	**99.5475**	**99.8681**	**99.4696**	**99.5188**	**0.3757**
	RBFN	91.0314	93.9921	86.2425	89.9507	4.7993
	SMO	91.8832	96.6266	85.5466	90.7497	2.6001
Random Search	SOM	84.7166	77.9295	93.7029	85.0914	23.1048
	PART	**99.8381**	**99.8413**	**99.8107**	**99.826**	**0.1381**
	RBFN	94.094	95.4941	91.634	93.5242	3.7643
	SMO	96.6453	96.9388	95.8178	96.3751	2.6343

REFERENCES

[1] G. Stein, B. Chen, A. S. Wu, K. A. Hua, "Decision tree classifier for network intrusion detection with GA based feature selection" ACM-SE 43: Proceedings of the 43rd annual Southeast regional conference- Vol. 2, p.136-141, 2005..

[2] M.Ektefa, S. Memar, F. Sidi, L. S. Affende. "Intrusion detection using Data Mining Techniques", Proceedings of IEEE International Conference on Information Retrieval & Knowledge Management, Exploring Invisible World, CAMP' 10, p. 200-203, 2010..

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

[4] Y.Y. Chung, N.Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)", Applied Soft Computing, p. 3014-3022, 2012..

[5] L. Koc, T.A.Mazzuchi, and S. Sarkani, "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier", Expert Systems with Applications, 39(18), p.13492- 13500, 2012..

[6] F. Zeng, K. Yin, M. Chen. "A new anomaly detection method based on rough set reduction and HMM", Eight IEEE/ACIS International Conference On Computer and Information science, p. 285-289, 2009..

[7] T.Kohonen, "The Self-Organizing Map". In the Proceedings of the IEEE, Vol.78, Issue: 9, p. 1464-1480, 1990.

[8] Y. Cao, J. Wu, "Projective ART for clustering data sets in high dimensional spaces" Neural Networks, vol.15, p. 105-120, 2002.

[9] S.V. Chakravarthy and J.Ghosh, "Scale based clustering using radial basis function networks", Proceeding of IEEE International Conference on Neural Networks, p.897-902, 1994.

[10] J. Platt, "Fast training of support vector machines using sequential minimal optimization", in Advances in Kernel Methods – Support Vector Learning, MIT Press, 1998.

[11] M.Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set", Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications, p.1-6, 2009