Development and Implementation of Artificial Neural Networks for Intrusion Detection in Computer Network

S. Saravanakumar[1], Umamaheshwari[2] , D.Jayalakshmi[3] and R.Sugumar[4]
1. Assistant Professor Senior, Department of SITE, VIT University, Vellore,
2. Professor and Head, Department of CSE, Bharath University, Chennai,
3. Lecturer, Department of EEE, Easwari Engineering College, Chennai,
4. Senior Lecturer, Department of IT, R.M.D. Engineering college, Chennai.

Abstract
The current network is complicated due to the high throughput and the multi-uniformity of actions. An intrusion detection system is a critical component for secure information management. Network Intrusion Detection Systems, which are part of the layered defense scheme, must be able to meet these organizational objectives in order to be effective. This work presents the investigations carried out on different neural network structures using a number of algorithms for intrusion detection. New algorithms have been presented which have faster convergence and better performance in Intrusion Detection System (IDS). The objective of this paper is to implement a new combination of artificial neural network algorithm which would be efficient in detecting intrusion in a networked environment. The performances of different methods have been compared.
Keywords:
pattern, network, intrusion detection, ANN, Malicious, Denial of Service.

I. Introduction

A computer system should provide confidentiality, integrity and assurance against denial of service. Due to increased connectivity, and the vast spectrum of financial possibilities that are opening up, more and more systems are subject to attack by intruders. Any system connected to internet cannot provide security without additional provision of intrusion detection elimination softwares [1]. Every organization of even small size is connected to internet. Due to functional requirements and cost factors, employees work from their home by connecting their systems with the main office. Employees exchange data in the form of revision, completion of the work assigned to them. Financial organization, automatic teller machines, landline telephones, cellular phones, wireless networks provide internet facilities. The equipment which rely upon main database stored in severs should not be damaged due to software threats in the form of intrusion. Military bases, nuclear research centers, organization with top level information should not be damaged in the form of alteration, corruption of information by any unknown activities entertained through the internet facilities by any one.

The primary ways an intruder can get into the system is through primary intrusion, system intrusion and remote intrusion [2]. The IDS responses to set of actions when detects intrusions. Some of the responses are mentioned in [3], which involves reporting results and findings to a pre-specified location, while others are more active automated responses. IDS can be viewed as the second layer of protection against unauthorized access to networked information systems because despite the best access control systems [4] and the intruders are still able to enter computer networks. IDS expand the security provided by the access control systems by providing system administrators with a warning of the intrusion [5]. The importance of the present research work is to explore the potential benefits of Artificial Neural Network (ANN) algorithms as intrusion detection software in a computer network connected to internet facility. When an ANN is properly explored for its complete implementation as intrusion detection software, most of the attacks can be detected. Some of the attacks are: Attempted break-ins, Masquerade attacks, Penetration of the security control system, Leakage, Denial of service, Malicious use.

Artificial neural networks (ANN) are computing elements, which are based on the structure and function of the biological neurons. There has been considerable interest shown in the development of new algorithms, which are faster and give better performance. ANN consists of interconnected processing units. The general model of processing unit consists of summing part followed by an output part. The summing part receives ‘n’ input values and weight values, and performs a weighted sum. The weighted sum is called the activation value. The sign of the weight for each input determines whether the input is excitatory (positive weight) or inhibitory (negative weight). The input and output could be the digital or analog data values. Several processing units are interconnected according to a selected topology of the network to achieve a pattern recognition task.

Manuscript received July 5, 2010
Manuscript revised July 20, 2010

The input of a processing unit may come from outputs of other processing units, and or from an external source. The output of each unit may be given to several units including it. A network can be static or dynamic; some of the static networks use the back propagation algorithm and radial basis function with multilayer perceptrons. Some of the dynamical networks (recurrent networks) have output feedback, state feedback and feedforward dynamics. The learning of the network can be supervised or unsupervised. In supervised learning, both inputs and outputs are presented to the network. In the unsupervised learning (self-recognizing networks), the inputs alone are presented to the network. Some of the algorithms for unsupervised learning are adaptive resonance theory [6] self organizing features maps [7]. One of the important applications of ANN is in pattern recognition analysis. A pattern is a set of inputs and outputs. Either supervised or unsupervised training method can be used to train an ANN, depending upon the network topology. In the supervised training, the difference between the calculated output of the network and the desired output of the pattern is minimized. To achieve the minimum difference, synaptic weights are updated. This procedure is adopted for all the patterns.

Lippmann [8] in his pioneering work has given the state of the ANN. Subsequently, in [9] the work has presented the latest developments in supervised learning. The desire to develop ANN started in the late forties was stated in [10]. The contribution by Webros in his doctoral dissertation in the area of feed forward neural networks was the first and the foremost. Conventionally a supervised learning employs the well known BPA with a linear function of weights and uses the steepest-desent method (SDM) for weight updation as found in [11]. Many others [12] have improved the BPA with faster convergence employing other updates. Hirose [13] have used an algorithm based upon the mean squared error to analyze the number of hidden nodes.

In addition to the above, optimal discriminant plane technique [14] used to map n-dimensional pattern into a 2-dimensional pattern to train the ANN. Echostate Neural Network [15] possesses a highly interconnected and recurrent topology of nonlinear Processing Elements (PEs) that constitutes a “reservoir of rich dynamics” and contains information about the history of input and output patterns. This has been attempted in intrusion detection. In the work, the different algorithms developed have been used on data obtained during Knowledge Discovery and Datamining (KDD) for intrusion detection. In addition, the well known XOR problem has been used as benchmark data.

II. Updation of the Synaptic Weights

Although a number of algorithms are investigated, a sample number of algorithms are presented here and their performances are discussed.

Back Propagation Algorithm (BPA)

The BPA uses the Steepest Descent Method (SDM) to reach a global minimum. The SDM uses the error in the output layer of the network to update the weights of the network so as to reach the minimum of the objective function, which is defined to the summation of squared error between the desired outputs and the network outputs. The algorithm uses a learning parameter called η. The algorithm works on supervised learning. The number of iterations required for different values of η for different range of synaptic weights for SDM, the number of iterations required for constant weights for SDM and the number of iterations required for different hidden nodes with one hidden layer for SDM were found. A comparison is made between the iterations required for one hidden layer and two hidden layers in SDM, the iterations required for the nodes in the hidden layer for different value of η for SDM and the iterations required by the nodes in the hidden layer with and without θ in SDM were found. However, it would take enormous amount of time for the ANN to learn the patterns. Hence only 1000 patterns have been considered for training purpose. The dataset has been separated as training and testing (intrusion detection). Training indicates the formation of final weights which indicate a thorough learning of intrusion and normal packets along with corresponding labeling. The convergence rate of BPA is shown in figure 1. The classification performance of BPA is shown in Table 1. In Table 2, false acceptance rate and false rejection rates are shown.

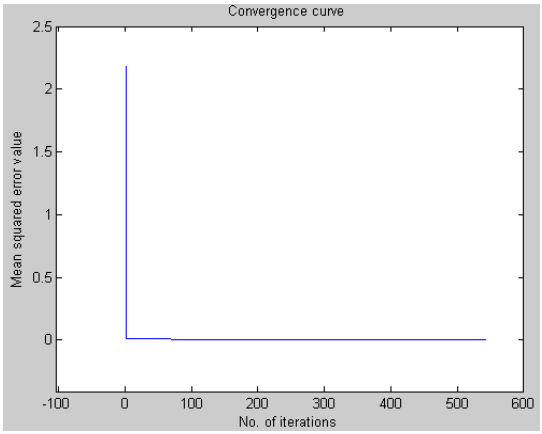

Figure 1 Mean squared error curve

Table 1 Classification performance

Packet type	Total No. tested	No. classified	No. misclassified
Normal	363	360	3
Intrusion	637	600	37

Table 2 False acceptance / Rejection rate

Packet type	False Acceptance Rate (FAR)	False Rejection Rate (FRR)
Normal	5.8% (37/637)	0.8% (3/360)
Intrusion	10.1%(37/363)	0.4%(3/637)

Echostate Neural Network (ESNN)

ESNN [16], [17] possesses a highly interconnected and recurrent topology of nonlinear PEs that constitutes a "reservoir of rich dynamics" and contains information about the history of input and output patterns. The outputs of internal PEs (echo states) are fed to a memory less but adaptive readout network (generally linear) that produces the network output. The interesting property of ESNN shown in figure 2 is that only the memory less readout is trained, whereas the recurrent topology has fixed connection weights. This reduces the complexity of RNN training to simple linear regression while preserving a recurrent topology, but obviously places important constraints in the overall architecture that have not yet been fully studied.

To train the ESNN, reservoirs and state matrix have to be used. The number of the iterations required for ESNN is lesser than the number of iterations required for SDM.

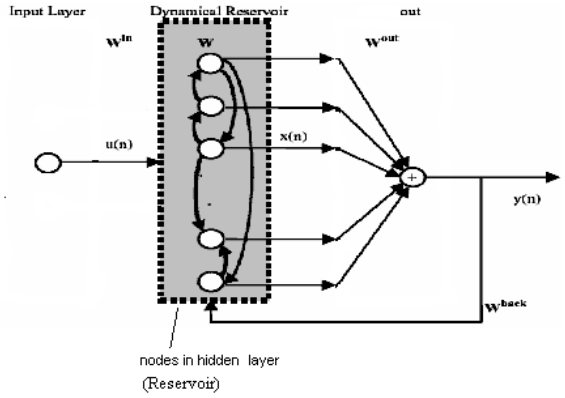

Figure 2 An echo state neural network (ESNN)

III. Experimental Analysis

It is mandatory to use huge amount of patterns to be presented for training ESNN. However, it would take enormous amount of time for the ESNN to learn the patterns. Only 24 patterns have been considered for training purpose. Training indicates the formation of final weights which indicate a thorough learning of intrusion and normal packets along with corresponding labeling. Figure 3 shows the performance of the ESNN.

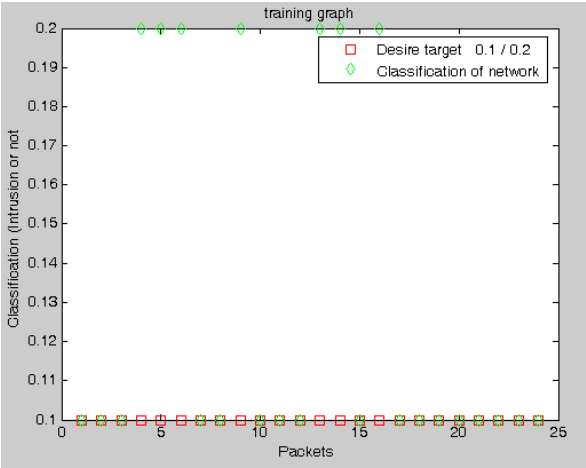

Figure 3 Packet classification

Radial Basis Function (RBF)

A Gaussian RBF monotonically decreases with distance from the centre. In contrast, a multiquadric RBF which, in the case of scalar input monotonically increases with distance from the centre. Gaussian-like RBFs are local and are more commonly used than multiquadric-type RBFs which have a global response. Radial functions are simply a class of functions. In principal, they could be employed in any sort of model (linear or nonlinear) and any sort of network (single-layer or multi-layer). RBF networks have traditionally been associated with radial functions in a single-layer network [18]. The simulation of intrusion detection has been implemented. Table 3 gives the distribution of patterns chosen for training and testing. This data set has been separated using variance analysis into training (183 patterns) and testing (2973 patterns).

Table 3 Distribution of patterns chosen for training and testing

Class	Training Pattern	Testing Pattern
1 (Normal	148	1286
2 (snmpgetattack)	7	735
3 (smurf)	28	932
Total	183	2983

Structural variants of ANN - Hierarchical structures (HS)

It is easy to implement direct classification method for a two-classification problem and requires no assumption about a suitable partition of the whole problem in to subtasks. With hierarchical classification one can design parts of whole system independently of each other. One

goal is to breakdown the complexity of the problem and defines the several sub problems each of which is easier to solve. There are two advantages to this approach. In order to classify and detect many attacks, hierarchical classification is adopted. Firstly, if we consider two classes with similar representations, a network for pair wise classification has more discriminative power than a network which has to separate the correct class against all others at the same time. This results from the fact that discriminating regions for patterns of these two classes are in conflict with those of other classes. Secondly, with this scheme we can construct more complex decision regions than with direct classification.

Algorithms with non-linear perception blocks

When non-linear summation is used the number of iterations required is lesser compared with that of the iterations required for linear summation. The non-linear summation is developed as follows:

Let $X = \{x_i\}$; i =1,…41. Where X is a pattern and x_i is the feature of the pattern. i.e, $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. The outer product of X is given in the matrix

$$X = \begin{bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_1x_4 & x_1x_5 & x_1x_6 \\ x_2x_1 & x_2x_2 & x_2x_3 & x_2x_4 & x_2x_5 & x_2x_6 \\ x_3x_1 & x_3x_2 & x_3x_3 & x_3x_4 & x_3x_5 & x_3x_6 \\ x_4x_1 & x_4x_2 & x_4x_3 & x_4x_4 & x_4x_5 & x_4x_6 \\ x_5x_1 & x_5x_2 & x_5x_3 & x_5x_4 & x_5x_5 & x_5x_6 \\ x_6x_1 & x_6x_2 & x_6x_3 & x_6x_4 & x_6x_5 & x_6x_6 \end{bmatrix}$$

The non-linear summation to the node in the next layer is given by

1) II $w_{ij}x_i = w_{ij}x_1 + w_{ij}x_2 + w_{ij}x_3 + w_{ij}x_4 + w_{ij}x_5 + w_{ij}x_6$(NL1),
2) $\sum w_{ij}x_i (i \neq j)$, = Off-diagonal elements of the above matrix (NL2)
3) $\sum w_{ij}x_i^2$ = Diagonal elements of the above matrix (NL3)
4) $\sum w_{ij}x_i (i \neq j)$, + $\sum w_{ij}x_i^2$ =Diagonal element and Off-diagonal elements of the above matrix (NL4)
5) linear plus NL1 (NL5)
6) linear plus NL2 (NL6)
7) linear plus NL3 (NL7)
8) linear plus NL4 (NL8)

Optimal Discriminant Plane (ODP)

The process of changing the dimensions of a vector is called transformation. The transformation of a set of n-dimensional real vectors onto a plane is called a mapping operation. The result of this operation is a planar display. The main advantage of the planar display is that the distribution of the original patterns of higher dimensions (more than two dimensions) can be seen on a two dimensional graph. The mapping operation can be linear or non-linear. In this work, the generalized declustering optimal discriminant plane is used. The mapping of the original pattern 'X' onto a new vector 'Y' on a plane is done by a matrix transformation, which is given by and φ1 and φ2 are the discriminant vectors (also called projection vectors).

$$Y = AX$$

where

$$A = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix}$$

The dimension of a pattern depends on the number of features. A n-dimension pattern has n-features and lies in a higher dimensional space. It is difficult to visualize the distribution of a pattern in a higher dimensional space. In order to visualize the distribution of the patterns, the pattern is mapped on to a two-dimensional space by using a transformation. A transformation of a set of n-dimensional pattern onto a two-dimensional plane is called a mapping operation. The result of this mapping is a planar display. Mapping denotes both transformation and its result. Fisher's optimal discriminant plane technique has been extended to pattern classification when the numbers of patterns are less. This technique has been successfully applied to pattern classification of KDD in intrusion detection.

Combination of ODP & SDM:

In this method, ϕ_1 & ϕ_2 vectors obtained in the ODM are multiplied with the input vectors into 2-dimensional vectors. After obtaining the 2-dimentional vectors training of ANN was done with SDM. Because of the conversion of n-dimensional input vectors into a 2-dimensional vector, the computation complexity is reduced. Due to this, the number of iterations required for reaching the required MSE is less and the classification performance is same as that of the classification performance of SDM.

Presentation of Normalized Patterns

The features of the patterns are normalized by taking the maximum value of each feature and dividing the values of the same feature for all the patterns. This is, because the outputs of the sigmoid function will never reach 0.0 or 1.0.When functional update method is used, the patterns are binary coded. Selection of patterns for training the neural network is important as they should be representative of all the patterns collected during machining. So, statistical techniques have been used to

select the training and testing patterns. For each class patterns with maximum variance VE_i^2 are selected. The maximum VE_i^2 of a pattern is found from the equation

$$VE_i^2 = \frac{\sum_{j=1}^{nf}(x_{ij} - \bar{x}_j)^2}{\sigma_i^2} \quad ; \quad \sigma_i^2 = \frac{1}{L}\sum_{i=1}^{L}(x_{ij} - \bar{x}_j)^2$$

where

nf is the number of features
L is the number of patterns
p_i is the pattern number
x_{ij} features of pattern 'X'
$\bar{x}_j$ is the mean for each feature
VE_i^2 is the variance of patterns
j is the feature
i is the pattern

V. Conclusions

The research work has involved in development of artificial neural networks for implementation in intrusion detection system. The back propagation algorithm, Radial basis function and Echostate neural network along with Fisher's linear discriminate function have been used. The input patterns are preprocessed with non linear vectors. The performances of different methods have been compared. The research work has been focusing in design and development of artificial neural network algorithms for implementation in intrusion detection. New attacks are evolved by intruders. Neural networks can be explored in combinations with genetic algorithms and fuzzy logic for the possibility of detecting new attacks.

References

[1] Sampada Chavan, Khusbu Shah, Neha Dave and Sanghamitra Mukherjee, 2004, "Adaptive Neuro-Fuzzy Intrusion Detection Systems", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04).

[2] Gang Kou, Yi Peng, Yong Shi, And Zhengxin Chen, (2006), "Network Intrusion Detection by Multi-group Mathematical Programming based Classifier", Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06).

[3] Helman, P. and Liepins, G., (1993), "Statistical foundations of audit trail analysis for the detection of computer misuse", IEEE Transaction on Software Engineering, 19(9):886-901.

[4] Baojun Zhang, Xuezeng Pan and Jiebing Wang, 2007, "Hybrid Intrusion Detection System for Complicated Network", Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007).

[5] Jingg-Sheng Xue, Ji-Zhou Sun, Xu Zhang, "Recurrent Network in Network in Network Intrusion Detection System", Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August, 2006.

[6] Carpenter, G.A., and Grossberg, 1987, "Self-organization of stable category recognition codes for analog input patterns", Applied optics, Vol. 26, No. 23, pp. 4919-4930.

[7] Kohonen T., 1990, "The self-organizing map", Proceedings of the IEEE, Vol. 78, No. 9, pp. 1464-1480.

[8] Lippmann R. P., 1987, "An introduction to computing with neural nets", IEE Transactions On Acoustics, Speech and Signal Processing Magazine, Vol. 35, No. 4, pp. 4-22.

[9] Hush D. R., and Horne B. G., 1993, "Progress in supervised Neural networks", IEEE Signal Processing Magazine, Vol. 10, No. 1, pp. 8-39.

[10] FNo. 148). Bremen: German National Research Center for Information Technology, 2001.

S. Saravana kumar an Assistant Professor of School Of Information Technology and Engineering at the University fVIT, Vellore, India. Currently, the pursuing research interests include the Application of Intrusion Detection In Computer Network

D.Jayalakshmi, Lecturer in EEE Dept. ,Easwari Engineering College, Chennai. Completed M.E.(Power Electronics And Drives)

R.Sugumar Senior Lecturer in IT Dept., RMD Engineering College, Chennai