ISSN 2047-3338

Analysis of ANN-based Echo State Network Intrusion Detection in Computer Networks

Dr. S. Saravanakumar[1], T.A. Mohanaprakash[2], Ms. R. Dharani[3] and Dr. C. Jaya Kumar[4]

[1,2]Department of CSE, Panimalar Institute of Technology, Chennai, India
[3]Department of IT, Panimalar Institute of Technology, Chennai, India
[4]Department of CSE, RMK Engineering College, Chennai, India

***Abstract*– The complexity of attacks on computer systems is increasing rapidly. The current network is complicated due to the high throughput and the multi- uniformity of actions. Intrusion detection is a process of monitoring the various computer networks and systems for violations of security and this can be automatically done with the help of an intrusion detection system. An Intrusion Detection System (IDS) is a critical component for secure information management. IDS plays a major role in detecting and disrupting various attacks before cooperating with the software. This paper presents the investigations carried out on different neural network structures using a number of algorithms for intrusion detection. Also this paper proposed an Echo State Network (ESN) structures for intrusion detection. The proposed algorithm has faster convergence and better performance in IDS. The objective of this paper is to implement the ESN algorithm and compare with other neural network algorithms in a networked environment. The performances of different methods have been implemented and compared using the Knowledge Discovery and Data mining (KDD) dataset to experiment the performance of ESN in classifying the Local Area Network (LAN) intrusion packets.**

***Index Terms*– ANN, Denial of Service, Echo State Network, Intrusion Detection, Malicious Attacks and Neural Networks**

I. INTRODUCTION

A computer system should provide confidentiality, integrity and assurance against Denial of Service (DoS). 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 [14]. Every organization of even small size is connected to the 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.

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents which are violations of computer security policies. Intrusion detection can be used to guard a host computer or network against being a source or a victim of an attack. Intrusion detection system is software that automates the intrusion detection process. IDS has become increasingly vital over the last decade as network information systems have grown into the daily life of most businesses, government agencies and private citizens [13]. Because of the increasing dependence in which companies and government agencies have their own computer networks and the importance of protecting these systems from attack is critical. A single intrusion of a computer network can result in the loss of unauthorized utilization or modification of large amounts of data and cause users to question the reliability of all of the information on the network.

IDSs can be categorized into three types namely, network-based intrusion detection, router-based intrusion detection, and host-based intrusion detection [16]. Network-based intrusion detection, operate at the gateway of a network and examines all the incoming packets. Router-based intrusion detection is installed on the routers to prevent intruders from entering into the network. Finally, the host-based intrusion detection receives the necessary audit data from the hosts operating system and analyzes the generated events to keep the local system secure. A centralized scheme is proposed to schedule authentication and intrusion detection, which needed a centralized controller [12]. This is more opted for a single system rather than a network with distributed systems with random mobility. There are numerous methods of responding to a network intrusion [6], but they all require the accurate and timely identification of the attack. IDS detect DoS attacks either by using a priori knowledge of the types of known attacks or by recognizing deviations from normal system behaviors. DoS attacks aim at denying or degrading legitimate users access to a service or network resource, or at bringing down the servers offering

such services.

An intruder can get into the system through primary intrusion, system intrusion and remote intrusion [4]. The IDS responses to set of actions when detects intrusions. Some of the responses are mentioned in [5], 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 [2] and the intruders are still able to enter computer networks. IDS expand the security provided by the access control systems by using system administrators with a warning of the intrusion [8].

Various algorithms are used to model the attack signatures and normal behavior response patterns of the systems. There are three algorithms used to model the various attacks, [9] and are named as, naive Bayes, Artificial Neural Network (ANN), and Decision Tree (DT). The naive Bayes classifier is based on a probabilistic model. This model will assign the most likely class for a given instance. ANN model is a pattern recognition technique. This technique has the capacity to adaptively model the user or the system behavior. DT model is a machine learning technique. This model is used to organize the attack signatures into a tree structure. IDS will create two kinds of errors. One is False Positive (FP) and another is False Negative (FN). FNs mainly results in security breaches since intrusions are not detected. Therefore, no alert is raised. The False Negative Rate (FNR) is used to measure the secure characteristics of the IDS. A low FNR means a low possibility that intrusion can occur without detection [16].

The importance of the present work is to analyze the potential benefits of ANN algorithms as intrusion detection software in a computer network connected with internet facility. When an ANN is properly explored for its complete implementation in 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, etc.

The rest of the paper is organized as follows. Section 2 describes about the artificial neural networks. Section 3 illustrates the implementation of the various algorithms used for the analysis. Section 4 details the experimental setup and analysis, and finally conclusions are given in section 5.

II. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) has computing elements that are based on the structure and function of the biological neurons. The new algorithms 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.

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 perceptions. Some of the dynamical networks (recurrent networks) have output feedback, state feedback and feed forward 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 [3] self organizing features maps [11]. 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.

III. ALGORITHMS

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

A. Back Propagation Algorithm (BPA)

BPA is one of the most studied and used algorithm for neural networks learning [10]. 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 Fig. 1. The classification performance of BPA is shown in Table 1. In Table 2, false acceptance rate and false rejection rates are shown.

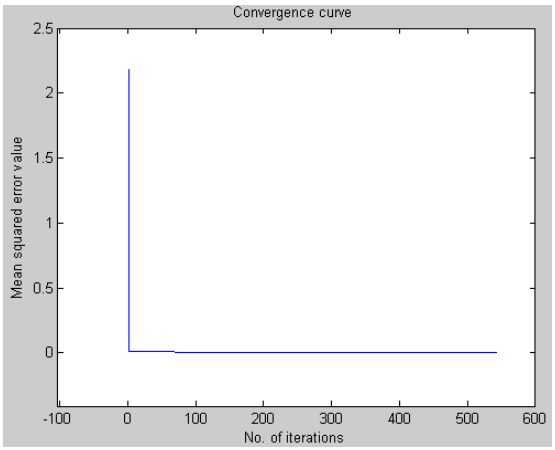

Fig. 1: Mean squared error curve

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)

B. Echo-state Neural Network (ESNN)

ESNN [6], [7] 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 Fig. 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.

Training of ESNN

Training an ESN is a simple linear regression model task. In this, only the output activation function, f is calculated, while Win, W, and W back never change after the initialization. The training is divided into the following three steps.

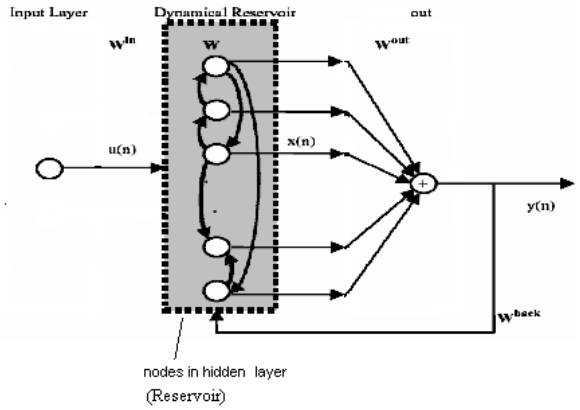

Fig. 2: An echo state neural network (ESNN)

1). Network Initialization

- W^{in} and W^{back} are generated randomly
- Random sparse matrix W is generated and scaled to have a spectral radius of α, where $\alpha < 1$ to ensure the presence of echo states in the network

2). Sampling Network Training Dynamics

- Network inner units are initialized arbitrarily. For example, x(0) = 0.
- The inner units states are updated for n = 0,1, 2, ... , T, using the equation,

$$x(n+1) = f(W^{in}\, u(n+1) + Wx(n) + W^{back}\, d(n)) \qquad (1)$$

where d(t) is the teaching signal and d(0) = 0.

- Network states before a washout time T_o are ignored due to their dependency on the initial state.
- Network states (u(n + 1), x(n + 1), d(n − 1) after To are collected in a state collecting matrix M of size $(T - T_o + 1) \times (K + N + L)$.
- $f^{out\ -\ 1}(d(n))$ values after T_o are collected in a teacher collecting matrix T of size $(T - T_o + 1) \times L$.

3). Computing Output Weights

Output weights are computed by evaluating the pseudo-inverse matrix of M, multiplying it by T, and then transposing it.

$$W^{out} = (M^{+}T)^{t}$$

C. Radial Basis Function (RBF)

RBF have been found to be widely used for the interpolation of scattered data [15]. 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 [1]. 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

IV. EXPERIMENTAL ANALYSIS

It is mandatory to use huge amount of patterns to be presented for training Echo State Neural Network (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. Fig. 3 shows the performance of the ESNN.

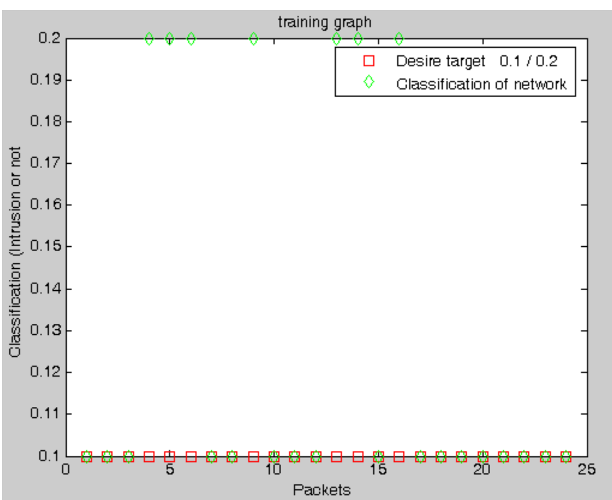

Fig. 3: Packet classification

The simulation results were obtained from the standard KDD data set. It is a well defined as normal and with different types of attack for TCP, UDP, ICMP, etc. A set of sample data set is shown in Table 4. Each row is a pattern. The fields in each pattern describe the properties of respective packet.

Table 4: Sample KDD dataset

Sl. No.	Packet Details
1	0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.00,0.00,0.00,0.00,0.00,0.00,normal.
2	0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.00,0.00,0.00,0.00,0.00,0.00,normal.
3	0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.00,0.00,0.00,0.00,0.00,0.00,normal.
4	0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.00,0.00,0.00,0.00,0.00,0.00,snmpget attack.
5	0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.01,0.00,0.00,0.00,0.00,0.00,snmpget attack.

Table 5: Sample dataset used for training

Sl. No.	Patterns used for training Input to ESNN after uncorrelating the features of Patterns	Target outputs
1	0 .2 .01 .1 105 146 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0.00 0.00 0.00 0.00 1.00 0.00 0.00 255 254 1.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00	.1
2	0 .2 .01 .1 105 146 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0.00 0.00 0.00 0.00 1.00 0.00 0.00 255 254 1.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00	.1
3	0 .2 .01 .1 105 146 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0.00 0.00 0.00 0.00 1.00 0.00 0.00 255 254 1.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00	.1
4	0 .2 .01 .1 105 146 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 0.00 0.00 0.00 0.00 1.00 0.00 0.00 255 254 1.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00	.2
5	0 .2 .01 .1 105 146 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 0.00 0.00 0.00 0.00 1.00 0.00 0.00 255 254 1.00 0.01 0.01 0.00 0.00 0.00 0.00 0.00	.2

Instead of KDD data set, free sniffer software's like network sniffer, packet sniffer and more software's can be used to extract the values of a packet, which can be further labeled as normal or an attack to be used for training. The contents of the packet should be suitably modified into meaningful numerical values. A sample dataset used for training all the three algorithms, back propagation algorithm, echo-state neural network algorithm and radial basic function are shown in Table 5.

The topology of ESN used is 41 x 20 x 1; no. of nodes in the input layer is 41, no. of nodes in the hidden layer is 20 and no. of nodes in the output layer is 1. The labeling is set as 0.1 (normal) or 0.2 (attack). It is mandatory to use huge amount of patterns to be presented for training ESN. However, it would take enormous amount of time for the ESN to learn the patterns. Hence, only 24 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. Table 6 gives number of patterns used for training and testing the performance of ESNN in classifying the intrusion packet. Table 7 gives number of patterns classified and misclassified.

Table 6: Distribution of patterns chosen for training

Packet Type	Total number used for training
Normal	17
Intrusion	7

Table 7: Classification performance

Packet Type	Total number tested	No. Classified	No. Misclassified
Normal	17	15	2
Intrusion	7	2	5

V. CONCLUSIONS

The paper has been carried out to achieve faster and better performance from a set of already available information in the database. With the existing training and testing data, the classification performance is 100%. In this paper, KDD dataset has been considered to experiment the performance of ESN in classifying the LAN intrusion packets. A topology of 41 x 20 x 1 had been chosen. The future work will involve in implementing an echo state neural network for classification of intrusion packet and suggested to implement other combinations of supervised and unsupervised ANN by incorporating additional intrusion data.

REFERENCES

[1] Arvind Rapaka, Alexander Novokhodko, Donald Wunsch, "Intrusion Detection Using Radial Basis Function Network on Sequences of System Calls", IEEE, 2003.

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

[3] 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.

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

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

[6] Jaeger H., "Short term memory in echo state networks", Tech. Rep. No. 152, Bremen: German National Research Center for Information Technology, 2002.

[7] Jaeger, H.; Tutorial on training recurrent neural networks, covering BPPT, RTRL,EKF and the "echo state network" approach (Tech. Rep. No.159).; Bremen: German National Research Center for Information Technology, 2002.

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

[9] Katar C., "Combining multiple techniques for intrusion detection", International Journal on Computer Science and Network Security (IJCSNS), Vol. 6, No. 2B, pp. 208-218, Feb. 2006.

[10] Khalil Shihab, "A Back Propagation Neural Network for Computer Network Security", Journal of Computer Science, Vol. 2, Issue 9, pp. 710-715, Science Publications, 2006.

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

[12] Liu J., Yu F., Lung C. H., and Tang H., "Optimal combined intrusion detection and biometric-based continuous authentication in high security mobile ad hoc network", IEEE Transactions on Wireless Communication, Vol. 8, No. 2, pp. 806-815, February 2009.

[13] Qing-Hua L, Sheng-Yi Jiang, Xin Li, "A Supervised Intrusion Detection Method", Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August, 2004.

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

[15] Sarra S. A., "Integrated Multiquadric Radial Basis Function Approximation Methods", PERGAMON Computers and Mathematics with Applications 0, Elsevier, 2006.

[16] Shengrong Bu, Richard Yu F., Xiaoping P. Liu, Peter Mason, and Helen Tang, "Distributed Combined Authentication and Intrusion Detection with Data Fusion in High-Security Mobile Ad hoc Networks", IEEE Transactions on Vehicular Technology, Vol. 60, No. 3, pp. 1025-1036, March 2011.

Dr. S. SARAVANAKUMAR have more than 9 years of teaching and research experience. He did his Postgraduate in ME in Computer Science and Engineering at Bharath engineering college, Anna university, chennai, and Ph.D in Computer Science and Engineering at Bharath University, Chennai. He occupied various positions as Lecturer, Senior Lecturer, Assistant Professor and Associate Professor & HOD. He has published more than 15 research papers in High Impact factor International Journal, National and International conferences and visited many countries like Taiwan, Bangok and Singapore. He has guiding a number of research scholars in the area Adhoc Network.

T. A. MOHANAPRAKASH has more than 6years and 4 months of teaching and research experience. He did his Postgraduate in M.Tech in Information Technology at Sathyabama University Chennai .He occupied various position as Lecturer, Senior Lecturer and Assistant Professor. He has published 8 research papers in International journal, International conferences and National conferences.

Ms. R. DHARANI has more than 8 years of teaching and research experience. She did her Under Graduate in B. Tech, Information Technology at Anna University and Postgraduate in M. Tech, Computer Science and Engineering at Dr. M. G. R University, Chennai. She occupied various positions as Lecturer, Senior Lecturer and Assistant Professor. She had published more than 10 research papers in International Journal, National and International conferences. She is guiding number of research scholars in the area of Ad-hoc Network.

Dr. C. JAYAKUMAR has more than 14 years of teaching and research experience. He did his Postgraduate in ME in Computer Science and Engineering at Anna University, chennai, and Ph.D in Computer Science and Engineering at Anna University, Chennai. He occupied various positions as Lecturer, Senior Lecturer, Assistant Professor and Associate Professor, Professor & HOD. He has published more than 35 research papers in High Impact factor International Journal, National and International conferences and visited many countries like Taiwan, Bangok and Singapore. He has guiding a number of research scholars in the area Ad-hoc Network.