Design and Development of Suginer Filter for Intrusion Detection Using Real Time Network Data

Revathi Sujendran and Malathi Arunachalam
Department of Computer Science, Government Arts College, India

Abstract: *By rapid use of the Internet and computer network all over the world makes security a major issues, so using the intrusion-detection system has become more important. All the same, the primary issues of Intrusion-Detection System (IDS) are generating high false alarm rate and fails to detect attacks, which make system security more vulnerable. This paper proposed a new concept of using Suginer Filter to identify IDS. The Takagi-Sugeno fuzzy model is structured based on Neuro-fuzzy method to generate fuzzy rules and wiener filter is used to filter out attack as a noise signal using fuzzy rule generation. These two methods are combined to detect intrusive behavior of the system. The proposed suginer filter (Sugeno+Wiener) uses completely a different research structure to identify attacks and the experiment was evaluated on live network data collected, which shows that the proposed system achieves approximately 98.46% of accuracy and reduce false alarm rate to 0.08% in detecting different real time attacks. From the obtained result it's clear that the proposed system performs better when compared with other existing machine learning techniques.*

Keywords: *Intrusion detection, wiener filter, artificial neural network, knowledge discovery dataset, network socket layer, defense advanced research projects agency, support vector machine.*

Received October 10, 2014; accepted September 7, 2015

1. Introduction

Intrusion-Detection Systems (IDSs) is an evolving technology for protecting computer networks. For instance, in earlier day's Denial-of-Service (DoS) attack cannot cause serious disasters, but today, successful DoS attacks can cause great financial loss to organizations. The goal of intrusion-detection systems is to detect anomalous or misuse behavior of system and notify to network administrators about the activities. Many intrusions-detection tools have security weaknesses such as failing to encrypt the log files, ignoring access control, and failing to perform integrity checks, etc., An IDS is more secure than other security tools, such as firewalls [2].

Earlier research system based on two major concepts known as anomaly detection and signature detection based on abnormal behavior of the system. Initially IDS consists of collection of audit data from the observed system. Then this data is either preprocessed or directly applied to the detector to generate an alarm. The main aim of IDS is to increase detection rate and to reduce false alarm rate in detecting attacks [5]. Recently, the researcher mainly focused on anomaly detection based on proposed methodologies such as data mining, neural network, fuzzy logic and so on, but now a day's Neuro-fuzzy rule generation plays a vital role in detecting intrusive behavior of the system.

In this paper, a new concept of Suginer filter is used to detect intrusion. It consists of Takagi-sugeno fuzzy.inference system to generate linguistic rule based on the Neuro fuzzy concept [7]. The Adaptive Neuro Fuzzy Inference System (ANFIS) has the ability to construct a model based on input data and map to corresponding output, So ANFIS is used as a fuzzy classifier to classify intrusive activity based on its fuzzy rule generation [9]. The Wiener filter is an optimum linear filter used to filter out noise signal from the desired signal as data. It mainly used to reduce minimum mean square error signal and reduce the amount of noise present in data signal [1].

The rest of the paper is organized as follows; some of the related works are described in section 2. Section 3 explains Takagi-sugeno model and wiener filter. Section 4 details the proposed suginer system architecture. Section 5 describes the experimental result analysis and section 6 draws some conclusion and future works of the proposed system.

2. Related Work

To detect intrusion various research area has been focused such as data mining, Artificial Neural Network, etc., But now-a-days researcher focus on several soft computing and computational intelligence methods to detect anomalous behavior of the system. Hoang *et al*. [8] has proposed anomaly detection system using multiple detection engines and fuzzy inference system. Experimental results show that the proposed scheme reduced false alarm rate effectively. A novel intrusion detection was proposed by Elhag *et al*. [6] by combining genetic fuzzy system within the pairwise learning framework for development of the system which improves the precision rate. Luo and Xia

[11] proposed a four-angle-star based visualized feature generation approach to generate numerical features and to detect intrusion, Knowledge Discovery Dataset (KDD) cup99 data are used to identify the efficiency of the proposed system. A new approach of Counting Bloom Filter (CBF) is used by Brindha and Senthilkumar [3] to minimize error and to improve the false rate of intrusion detection, which is achieved efficiently by using Virtex 4 (XC4VSX25) Field Programmable Gate Array (FPGA) hardware. Similarly, various data mining techniques are used to detect Structure Query Language (SQL) injection attacks which were proposed by Kim and Lee [10] based on Support Vector Machine (SVM) classifier along with several kernel functions.

Toosi and Kahani [20] proposed ANFIS based Neuro fuzzy classifier to detect intrusion. The experimentation is analyzed based on the KDD cup data set to obtain better result genetic algorithm is used for optimization.

3. Fuzzy Inference System and Wiener Filter

3.1. Takagi-Sugeno Model

In this paper a new method of Takagi-Sugeno model based on Multiple Input Single Output (MISO) is used as a fuzzy rule generation engine. The Sugeno FIS based on linguistic variable uses if-then part. The premise of T-S model is a linguistic variable, but the consequent part is non fuzzy variable equated as function part. In this paper the rule generation is based on seven linguistic terms which are used as membership function to handle both continuous and discrete attributes. Figure 1 draws the Membership functions of the seven linguistic values as: Very Small (VS) Small (S), Medium Small (MS), Medium (M), Medium Large (ML) and Large (L), Very Large (VL)

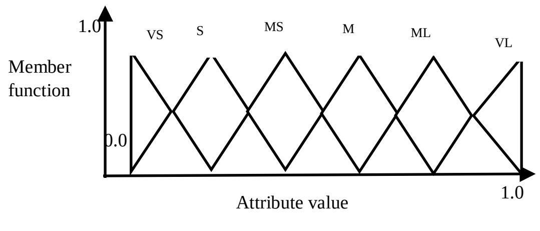

Figure 1. Membership functions of seven linguistic values (VS: Very Small, S: small, MS: medium small, M: medium, ML: medium large, L: large, VL: Very Large).

The first two parts of the fuzzy inference process as, falsifying the inputs and applying the fuzzy operator, are same as Mamdani type. The main difference between mamdani and Sugeno model is its output member function, be either linear or constant. A typical rule in a Sugeno fuzzy model has the form [19]. If Input 1 = x and Input 2 = y, then Output is z = ax + by + c for a zero-order Sugeno model, the output level z is a constant (a=b =0). The output level Z_i of each rule is weighted by the firing strength w_i. For example, for an AND firing rule strength be

$$w_i = \text{AndMethod } (F1(x), F2(y)) \tag{1}$$

Where *F1*(*x*) and *F2*(*x*) are the membership functions for Inputs 1 and 2. The final output of the system is computed based on weighted average as

$$Final\ output = \frac{\sum_{i-1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

The main advantage of Sugeno model is [19]

- It is computationally efficient.
- It works well with linear optimization and adaptive techniques.
- It's well suited for mathematical analysis.
- No defuzzification is needed; output is calculated based on weighted averages.

In this paper ANFIS structure is used to identify the parameter of Sugeno model using hybrid learning where the number of rules are not limited, a zero-order Sugeno model has unlimited approximation power for matching any nonlinear function on a compact set. This can be proved using the Stone-Weierstrass theorem. This ANFIS is named as universal approximate.

3.2. Wiener Filter

Wiener filter is used for estimates linear desired signal form other related noisy signal. It plays a vital role in linear estimation, signal restoration and system identification. The main idea of using wiener filter is to calculate minimum mean square error based on the average square distance between the normal filter output and desired output [14]. The signal component $X_i(n)$ be a N-dimensional dataset based on linear combination of normal data $S_i(n)$ and anomaly data $A_i(n)$, expressed mathematically by

$$X_i(n) = S_i(n) + A_i(n) \tag{3}$$

The wiener filter is mainly used to filter out normal and abnormal data and used to calculate minimum mean square error $e_{min}(n)$ based on normal output data y_i (*n*) and desired output *r* (*n*). The error value is calculated using

$$e_{\min}(n) = r\ (n) - y_i(n) \tag{4}$$

Rearranging Equation (4) we get

$$r(n) = y_i(n) + e_{\min}(n) \tag{5}$$

Let E_{MSE} denotes the Minimum Mean Square Error, defined by:

$$E_{MSE} = E[|e_{\min}(n)|^2] \qquad (6)$$

Hence, by evaluating the MSE on both sides of Equation (5), and applying it to the principle of orthogonality we get:

$$\sigma_r^2 = \sigma_{y_i}^2 + E_{MSE} \qquad (7)$$

Where σ_r^2 , $\sigma_{y_i}^2$ is the variance of the desired response and estimated output; on assuming the random variable to zero, the Equation (7) is:

$$E_{MSE} = \sigma_r^2 - \sigma_{y_i}^2 \qquad (8)$$

The mean square value will lies between zero and one, on dividing Equation (8) be divide by σ_r^2 obtaining

$$\frac{E_{MSE}}{\sigma_r^2} = 1 - \frac{\sigma_{y_i}^2}{\sigma_r^2} \qquad (9)$$

Clearly, this is possible because σ_r^2 is never zero, except when *r*(*n*) is zero for all *n*. Let

$$\kappa = \frac{E_{MSE}}{\sigma_r^2} \qquad (10)$$

Where *k* be normalized mean-squared error, in terms of rewriting Equation (10) in the form:

$$\kappa = 1 - \frac{\sigma_{y_i}^2}{\sigma_r^2} \qquad (11)$$

Where *k* be positive between the $0 \le \kappa \le 1$ To find the normal and abnormal states of the network, the autocorrelation matrix R admits the eigen-decomposition For instance, the eigenvector places a boundary between normal data traffic and abnormal traffic based on minimum and maximum values. The values beyond maximum may leads to abnormality of data flow. The design of a Wiener filter needs *a priori* knowledge about the statistics of the data to be processed. The filter is optimum only when the statistical features of the input data relate to *a priori* information about the filter [4]. When this information is not known completely, it may not be possible to design the Wiener filter to be optimum. In this research wiener is used to filter out the abnormality of the network data based on MMSE. The fuzzy rule generation is used to train wiener filter for error detection and postulate the abnormality of the network data.

4. Proposed Suginer System Architecture

4.1. Dataset Description and Extraction

So far the intrusive activities of the system are examined based on dataset available. The existing benchmark dataset used by researchers are Defense Advanced Research Projects Agency (DARPA) [13], KDDcup99 [12] and Network Socket Layer (NSL)-KDD [15] to detect intrusion. There are lots of statistical degrading issues in these existing dataset [12], so the proposed system uses live network server log data collected from every 2 second of data transmission in the network, such dataset has collected as live network data from Linkware Solution Pvt Ltd [18] which consist of 4 major attack category as DOS, Probe, U2R and R2L. Table 1 show number of records in dataset.

Table1. No of records in real time dataset.

Attack Name	No of Records
Normal	5917854
DoS	7683641
U2R	165920
R2L	3785161
Probe	3418944
Total	20971520

The dataset consist of 41 attribute out of which only 11 attributes are selected for detection phase. The attribute is selected based on new hybrid feature selection method of incorporating simplified swarm optimization with the random forest algorithm [17]. These 11 attributes consist of 7 discrete and 4 continuous attributes, data types are listed in Table 2.

Table 2. Reduced attribute by SSO-RF technique.

1. Protocol Type- Discrete
2. Service Name- Discrete
3. Flag status- Discrete
4. Size of Source in Bytes- Continuous
5. Size of Destination in Bytes- Continuous
6. Emergency Flag- Discrete
7. Connection Status- Discrete
8. Number of Outbound commands- Continuous
9. Accessed Files Count- Continuous
10. Host Login Status- Discrete
11. Guest Login Status- Discrete

4.2. Proposed Suginer Filter

In the proposed system, initially Sugeno based fuzzy rules are generated simultaneously using Lagrange's Interpolation along with successive approximation method. It's mainly based on a sequence of rule approximation that converges to the solution and is constructed recursively (Iteratively) that is, each new rule is calculated on the basis of the preceding rule generation. Let f(x) be a continuous function on the interval [a, b], the rule generation equation for *x=f(x) be* $x_1, ..., x_n,...$ such that $x_1 = f(x_0)$, $x_2 = f(x_1)$, $x_3 = f(x_2)$ ... Iterate the same process until $(x_n – x_{n-1})$ smaller than some specified tolerance (max IDR reached). The result of the successive approximation is then processed using Lagrange's interpolation to classify the rule as 7 segmented data which are then filtered using wiener to generate the data as normal or attack category. The procedure for suginer filter is as follows.

4.2.1. Procedure for Fuzzy Rule Generation using Wiener Filter to Detect Attacks

- *Step 1*. The input dataset D is divided parallel into five subsets as D= {D_i; where i= 1≤i≤5} are sent as input signal $x_i(n)$ to wiener filter
- *Step 2*.The wiener filter is modeled as a linear combination of normal dataflow $s_i(n)$ and abnormal data flow $A_i(n)$ where $x_i(n)= s_i(n)+ A_i(n)$
- *Step 3*. Wiener filter is used to filter abnormal data as noise signal and extract normal data as s_i.
- *Step 4*. To achieve balanced estimator Sugeno fuzzy predictor used 7 linguistic variable such as (VS, S, MS, M, ML, L and VL) to generate rules to achieve better results.
- *Step 5*. Based on rule generation wiener filter is used as the error detector to classify attack data in the network using parallel computation. The filter mainly used to reduce minimum mean square error ($E_{MSE)}$.
- *Step 6*. The correlation matrix generated by wiener filter is used to calculate chi square test which produces results as either 0 or 1.
- *Step 7*. Result 0 indicates significant indifference known as normal data and result as 1 indicates significant difference known as attack or abnormal data.

The pictorial representation of the suginer filter is illustrated in Figure 2, shows that the five input are given parallel to Suginer filter to filter out attacks and extract normal data as output.

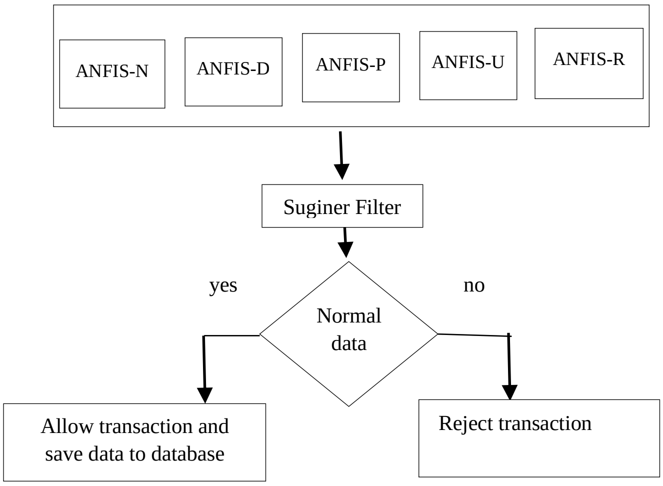

Figure 2. Flow of proposed suginer filter.

5. Experimental Result Analysis

All the correctly classified live dataset are collected as on Table 2 are used as training and testing data to evaluate the classifier. The experimentation has been done on 30% of training data and use full dataset for testing phase to analyze the performance of the proposed system. The training dataset based on Sugeno fuzzy rule generation train the classifier in parallel and the Wiener filter is used as the error detector to filter out anomaly data based on rule generation. The below Table 3 shows various factors as DR, FAR, training time and testing time.

Table 3. Performance of proposed system.

Models	Class Name	Accuracy (DR)	FAR	Train Time (Sec) 30%	Test Time (Sec) Full Dataset
Proposed Suginer Filter	Normal	98.42%	0.09%	66.2	202.8
	DOS	98.62%	0.09%	61.1	201.7
	Probe	97.62%	0.13%	66.2	202.8
	U2R	98.82%	0.07%	62.1	201.9
	R2L	98.84%	0.04%	62.4	202.4

The performance of the proposed suginer filter for intrusion detection detect attacks more accurately for live dataset. The proposed work has also been compared with several machine learning techniques such as fuzzy logic to generate various linguistic rules along with other hybrid method such as using an SVM classifier to classify the data as normal or attack. The genetic algorithm is used to optimize the various rules generated by the fuzzy system in detecting intrusion, In addition, this paper also compared our proposed method with an adaptive Kalman filtering technique based on fuzzy rule generation to detect various attacks, the comparative results of these techniques are listed in Table 4. The major difference between the existing and the proposed system is the use of fuzzy logic, which does not have learning and training capability, such drawback leads to a proposed system of combining Neuro-computing with fuzzy logic as Neuro-fuzzy system to generate more complex rules easily. Based on below table, it’s clear that the suginer filter shows high performance in detecting various attacks category in computer networks. Also, this method is more flexible and convenient for every situation due to its automatic rule generation in real time scenario.

Table 4. DR and FPR for other various machine learning techniques.

Techniques	DR	FAR
Fuzzy Logic	74.5%	2.51%
SVM + Fuzzy logic	80.7%	2.30%
Kalman Filter+Fuzzy Logic	81.4%	1.97%
GA+Fuzzy logic	83.5%	1.80%
Fuzzy Logic +Wiener filter	88.76%	1.34%
Proposed Suginer Filter	**98.46%**	**0.08%**

The pictorial representation of the comparison strategy for various machine learning techniques are illustrated in Figure 3, which shows high DR and very low FAR for proposed suginer filter.

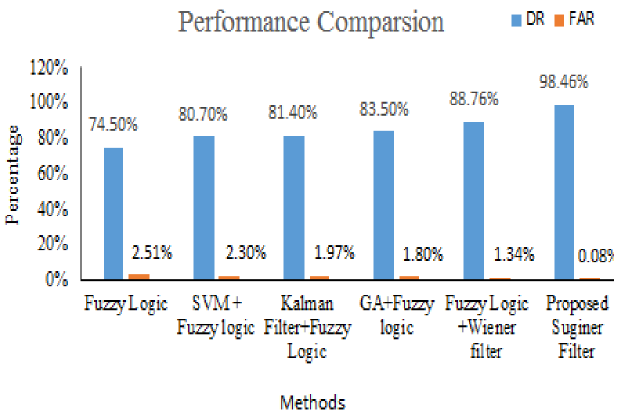

Figure 3. Performance comparison.

The detection rate and false positive rate are calculated based on

$$(DR) = \frac{Total\ number\ of\ correctly\ classified\ attacks}{Total\ number\ of\ ins\tan ce} * 100 \quad (10)$$

$$(FAR) = \frac{Total\ number\ of\ misclassified\ ins\tan ce}{Total\ number\ of\ ins\tan ce} * 100 \quad (11)$$

The signal generation graph is calculated based on a collection of real time data on various measures for 100 seconds has been shown below, that number of data transmission on x axis and total number of bytes transferred in the y axis. On using wiener filter the yellow line indicates the normal flow of data which shown in Figure 4.

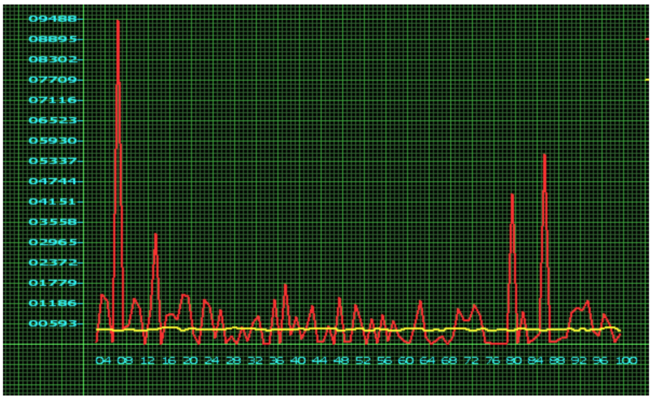

Figure 4. Data flow based on a suginer filter.

The Minimum mean square error for live network data be MMSE= 0.0187 has been calculated using Equation (9) based on the number of source byte transmission features.

All the research work based on intrusion detection has shown best detection rate only in DOS and Probe attack rather than other two, where in real time scenario U2R and R2l plays a major role. This research work shows significant outperform for U2R and R2l attack in both accuracy and false alarm rate [16].

6. Conclusions

In this paper Suginer filter has been proposed, which are used to detect intrusive activity for live network data and efficiently extract many rules for classification. As an application, intrusion detection system has been developed for the real time scenario to detect attacks and the result shows the proposed system can effectively perform the detection. Detection Rate and False Alarm Rate have been considered as two important measures for security system, which has been focused on this research work along with training and testing time.

Initially in the proposed system for feature selection a new concept of hybrid SSO-RF technique is used to reduce attribute more effectively. The proposed Suginer method shows high performance, efficiency than other existing methods. The important function of the proposed system is its parallel computation and extract fuzzy rule automatically, in addition wiener filter uses statistical calculations based on a chi square test which significantly different between normal and attack data. The advantage of automatic rule generation is, it may generate various rules based on real time attacks and it automatically calculates deviation for normal connection. The future work may be focused on using some optimization technique to improve detection rate.

References

[1] Al-Kasassbeh M., "Network Intrusion Detection with Wiener Filter-Based Agent," *World Applied Sciences*, vol. 13, no. 11, pp. 2372-2384, 2011.

[2] Axelsson S., "Research in intrusion-detection systems: A survey," Technical Report, 1998.

[3] Brindha P. and Senthilkumar A., "Network Intrusion Detection System: An Improved Architecture to Reduce False Positive Rate," *Journal of Theoretical and Applied Information Technology*, vol. 66, no. 1, pp. 618-626, 2014.

[4] Celenk M., Conley T., Graham J., and Willis J., "Anomaly Prediction in Network Traffic using Adaptive Wiener Filtering and ARMA Modeling," *in Proceedings of IEEE International Conference Systems, Man and Cybernetics*, Singapore, pp. 3548-3553, 2008.

[5] Depren O., Topallar M., Anarim E., and Ciliz M., "An Intelligent Intrusion Detection System for Anomaly Misuse detection in Computer Networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713-722, 2005.

[6] Elhag S., Fernández A., Bawakid A., Alshomrani S., and Herrera F., "On the Combination of Genetic Fuzzy Systems and Pairwise Learning for Improving Detection Rates on Intrusion Detection Systems," *Expert Systems with Applications*, vol. 42, no. 1, pp. 193-202, 2015.

[7] Full´er R., *Introduction to Neuro-Fuzzy Systems*, Springer, 1999.

[8] Hoang X., Hu J., and Bertok P., "A Program-Based Anomaly Intrusion Detection Scheme using Multiple Detection Engines and Fuzzy

Inference," *Journal of Network and Computer Applications*, vol. 32, no. 6, pp. 1219-1228, 2009.

[9] Jang J., "ANFIS Adaptive-Network-Based Fuzzy Inference System," *IEEE Systems, Man, and Cybernetics Society*, vol. 23, no. 3, pp. 665-685, 1993.

[10] Kim M. and Lee D., "Data-Mining Based SQL Injection Attack Detection using Internal Query Trees," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5416-5430, 2014.

[11] Luo B. and Xia J., "A Novel Intrusion Detection System Based on Feature Generation with Visualization Strategy," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4139-4147, 2014.

[12] Lu W., Tavallaee M., Bagheri E., and Ghorbani A., "A Detailed Analysis of the KDD CUP 99 Data Set," *in the Proceeding Of the IEEE Symposium on Computational Intelligence in Security and Defense Applications*, Ottawa, pp. 1-6, 2009.

[13] MIT Lincoln Labs, 1998 DARPA Intrusion Detection Evaluation, http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/ index.html, last Visited, 2008.

[14] Mulgrew B., Grant P., and Thompson J., *Digital Signal Processing: Concepts and Applications*, Macmillan, 1999.

[15] Nsl-kdd Data Set for Network-based Intrusion Detection Systems, Available on: http://nsl.cs.unb.ca/NSL-KDD/, Last Visited, 2009.

[16] Qassim Q., Patel A., and Mohd-Zin A., "Strategy to Reduce False Alarms in Intrusion Detection and Prevention Systems," *The International Arab Journal of Information Technology*, vol. 11, no. 5, pp. 500-506, 2014.

[17] Revathi S. and Malathi A., "Feature Extraction Using Sim-Swadorest Optimization Algorithm for Intrusion Detection," *in Proceedings of The International Conference on Recent Innovations in Computer Science and Information Technology*, Singapore, pp. 75-79, 2014.

[18] Revathi S., Linkware Technologies Private Limited, Network Simulator Capture (NSC) Dataset. Accessed, https://www.linkware.in/, Last Visited, 2013.

[19] Takagi T. and Sugeno M., "Fuzzy Identification of Systems and its Applications to Modeling and Control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 116-132, 1985.

[20] Toosi A. and Kahani M., "A new Approach to Intrusion Detection based on an Evolutionary Soft Computing Model using Neuro-fuzzy Classifiers," *Computing Communication*, vol. 30, no. 10, pp. 2201-2212, 2007.

Revathi Sujendran received her MSc degree in computer science from St. Joseph College of Arts and Science, Cuddalore, Tamilnadu, India, in 2008 and her MPhil degree in computer science from Bharathidasan University, Trichy, Tamilnadu, India, in 2009. She is now currently pursuing her PhD degree at PG and Research, Department of Computer Science, Government Arts College, affiliated to Bharathiar University, Coimbatore, Tamilnadu, India. She has published 24 Research paper which includes national, International and conference proceedings publications. She has visited Singapore for international conference and got excellent best paper award. Her current research interests include network security, data mining, and computational intelligence.

Malathi Arunachalam Graduated from Bharathidasan University in 1989 and completed M.Sc (Computer Science) in 1991 under the same University. She has also received qualified degree of M.phil and Ph.D respectively in Computer Science in the year 2002 and 2012 from Bharathiar University, Coimbatore, India. She has more than two decades of teaching experience and 14 years of research experience. She has completed a funding project by UGC. She is guiding 8 Ph.D scholars and 3 M.Phil Scholars. She has guided and produced 13 M.Phil Scholars. Currently she is working as an Assistant Professor, PG and Research Department of Computer Science, Government Arts College, Coimbatore. She has published 70 Research paper which includes national, International and conference proceedings publications. She has authored three books.