Fast Hierarchical Relevance Vector Machine towards Network Intrusion Detection System

V. Jaiganesh, PhD
Professor
PG & Research Department of Computer Science
Dr. N.G.P. Arts and Science College
Coimbatore, Tamil Nadu, India.

ABSTRACT
Internet becomes very essential needs in today's life because internet has become a public network world wide. The art of detecting inappropriate, incorrect, or anomalous activity is Intrusion Detection (ID). It is a security service which monitors and analyzes system events for finding, and creating a real-time or near real-time scenario, trials to access system resources in an unauthorized manner. In the proposed Fast Hierarchical Relevance Vector Machine (FHRVM), Analytical Hierarchy Process Method (AHP) is used to select the input weights and hidden biases. The algorithm is used to analytically determine the output weights and the Iterative Learning Mechanism (ILM) algorithm is employed in order to learn the network through Relevance Vector Machine (RVM). It is established by developing a probabilistic Bayesian learning structure which is capable enough to derive accurate prediction models. Such prediction models will exploit considerably fewer basis functions. These incorporate the benefits of valid predictions; elimination of non-impact attributes along with it will facilitate usage of arbitrary functions. Simulation has been carried out using Math works MATLAB R2012a. KDD Cup 1999 dataset is taken for testing the performance of the proposed work and the results indicate that FHRVM has achieved higher detection rate and low false alarm rate than that of existing RVM algorithm.

Keywords
Intrusion Detection System (IDS), Relevance Vector Machine (RVM), Levenberg Marquardt and Analytical Hierarchical Process.

1. INTRODUCTION
Network instruction discovers the 'burglar alarms' or 'intrusion alarms' of the computer and network security field system. The chief idea of this to shield a system by using a combination of an alarm that sounds at any time the site's security is negotiated, and an entity – most often a spot security officer can respond to the alarm and take the proper action. An ID tries to identify attempts to hack or break into a computer system or cruelty it. IDSs may monitor packets passing over the network and monitor system files, log files, otherwise set up ruse systems that attempt to trap hackers.

Incursion detection is needed in today's computing atmosphere because it is impossible to keep rapidity with the current and potential threats and vulnerabilities in the computing system techniques. This type of surroundings is constantly developing and changing fueled by new technologies and internet systems. To make matters inferior, bullying and vulnerabilities in these surroundings are also constantly developing. Intrusions discover products are tools to assist in managing threats and vulnerabilities in this varying environment. Intrusion detection assaults are segmented into two groups they are

Host-based attack (2-4): When the data reaches frn the records of various behaviors of host that includes audit record of operating system, system log files, information from application programs etc. Considering Windows NT 4.0 operating system, contains records of almost three events such as operating system events, safety events and application program events.

Network-based attack (5-7): When the data is transmitted over the network several records will be generated contains network segment. The network segments contains information, eg: internet packets.

For the past decade several machine learning algorithms and data mining techniques have been used for identifying network intrusion detection. This paper is organized as depicted in the Fig.1.

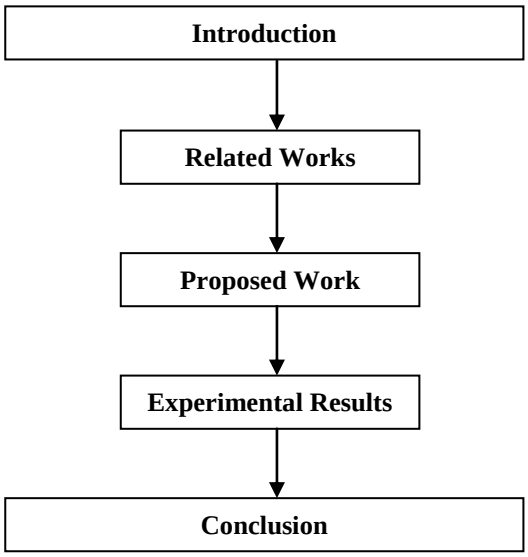

Fig.1 Flow of Research

1.1 Categories of IDS
Abuse detection vs. anomaly detection: In abuse detection, these evaluate the information it collects and compares it to large databases of attack signatures. Basically, it looks for a precise attack that has previously been documented. Similar to a virus detection system, this type of detection software is only as good as the database of attack signatures that it uses to compare packets against. In difference detection in this the system administrator defines the baseline normal and the state of the network traffic load their breakdown, important

protocol, about their typical packet size. These types of detector monitor network segments to compare their state to the normal baseline and look for anomalies.

Network-based vs. host-based systems: In a network-based system, it analyzes the individual packet curving through the network. It can detect malevolent packets that are designed to be overlooked by a firewall simplistic filtering rules. In host-based system the activity on each individual computer or host is examined by the IDS.

Relevance Vector Machine (RVM) is a Bayesian learning model for deterioration and categrization of identical functional form to the Support Vector Machine (SVM). RVM can be generalized well and provide detections the cost. This technique employs RVM classification.

Arrangement of the Paper can be as follows : Section 2 provides the related works involved in intrsion systems and the techniques used here. Proposed methodology is described in the Section 3 and 4 that gives the experimental results to the proposed work. Section 5 cocludes the paper with future works followed by the referenced manuscripts.

2. RELATED WORKS

Instruction detection is the process of monitoring and examining events occurred in a computer or network and presenting the results to the administrator. The related research is started in early 1980's stage. Then the instruction detection continued through numerous major DARPA projects and other Government Programs. Later in the beginning of the 1990s it becomes a hot research topic and commercial IDRS started to emerge when the internet epoch arrived.

Unlike a traditional network, which only submissively transforms data package, an active network allows the network lump to accomplish mobile code passed in packets. The proposed IDRS merges distributed monitoring techniques and data mining methods to analyze the threats of the incoming data and respond to the interruption effectively.

The KDDCup99 datasets comprises very high training examples. These datasets could be a common benchmark for analysis of intrusion detection techniques. This dataset contains number of connection records where each connection is a series of packets including values of 41 features. Four main types of attack are present in this datasets: Denial of Service (DoS), probe, user to root (U2R), and remote to local (R2L). Separate, Continuous, and Symbolic are the feature values in this data set. Range of values for some of these features is very huge and assorts.

An increasing trend that speaks to this issue is that the training of intrusion detection system. This system area unit aims towards police work threatening things that occur in malice of alternative security measures and follow two main paradigms: Anomaly detection and Misuse detection systems. These are supposed to be utilized in conjunction with further security measures imposing the protection policy. Like motion detectors during a building, they symbolize a second line of protection session behind the locks on the doors and windows.

The next section explains about the proposed work which is the combination of RVM, LM and AHP method.

3. PROPOSED WORK

The proposed methodology named as Fast Hierarchical Relevance Vector Machine is used for Intrusion detection system is dealt in this section. The overall methodology of the proposed work is shown in the Fig.2.

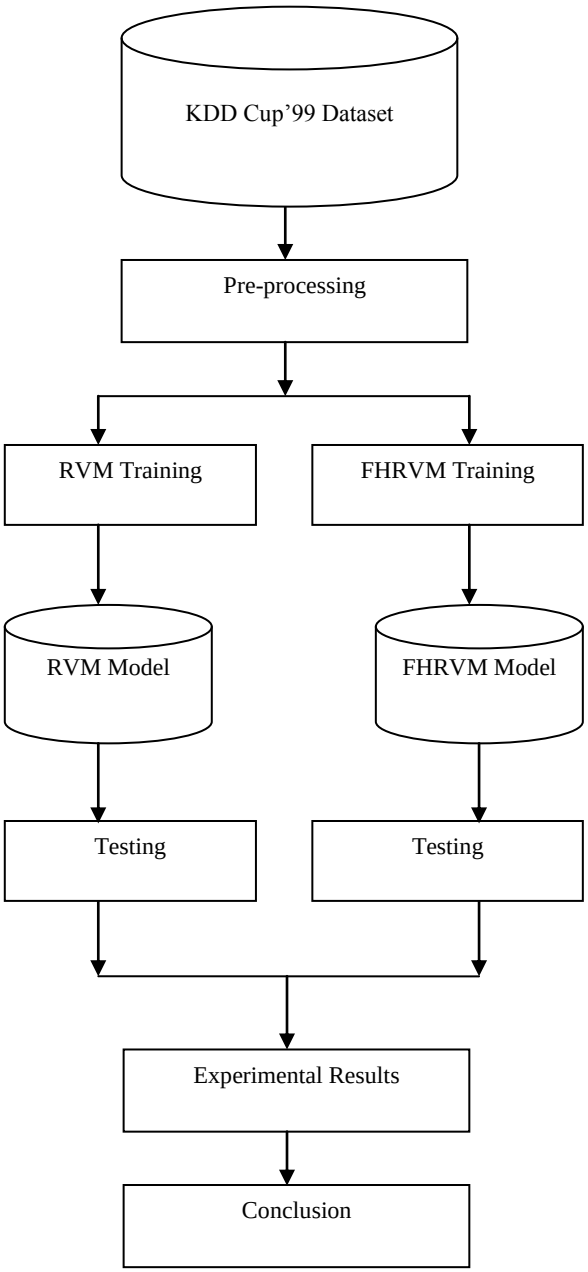

Fig.2 Methodology

3.1 Data collection and Preprocessing

The benchmark dataset KDD Cup'99 has been taken for testing the performance of the proposed FHRVM. KDD Cup'99 is the dataset which is being used by many researchers as the benchmark dataset for their testing. Major attacks are contained in the dataset that includes back dos buffer overflow u2r, ftp_write, guess_passwd, imap, ipsweep probe, land dos, load module u2r, multihop, neptune dos, nmap probe, perl u2r, phf, pod dos, portsweep probe, rootkit u2r, satan probe, smurf dos, spy, teardrop dos, warezclient r2l, warezmaster.

41 attributes (features) contained in KDD Cup'99 dataset. The above said attacks could be broadly comes under the categories such as DOS, Probe, U2R and R2L attacks.

The important among all the protocols such as TCP, UDP and ICMP are used to train the proposed mechanism and the training dataset has not included all the types of attacks. The

testing has been carried with 10% of KDD Cup'99 dataset. Replication of data that is present in the dataset is eliminated which belongs to labeled data as "normal". The proposed FHRVM is involved in classifying the given dataset into normal and malicious attacks.

3.2 Sparse Bayesian Learning for Regression

The standard probabilistic formulation is followed with the given a data set consisting of input-target pairs $:\{Xn, tn\}^N_n$. The scalar-valued target function only is considered. It is assumed that the targets are samples from the model with additive noise:

$$t_n = y(X_n;W)+E_n \quad (1)$$

Where E n are independent samples from some noise process then it is assumed to be mean-zero Gaussian with variance.

The said link is involved to a non-linear function in order to extend the logistic model that has non-linear transformations of the features that are taken or selected as input. The classifiers are of two types such as

- Linear classifiers – The linear classifiers are capable enough to classify high dimensional data with reduced time complexity.
- Kernel classifiers – Kernel classifiers are non-linear in nature and have the added functionality of learning. The time consumption is greater when comared with the linear classifier but that can be compromised with ability of learning.

The latent or hidden variables are interpreted and allowed to exploit the probability link.

3.3 Relevance vector machine

Relevance vector machine is simply a specialization of a spares Bayesian model which uses the same kernel based data dependency. The important features of RVM are the predictors that are inferred. Such features of RVM are exceedingly sparse in nature which contains relatively few ''relevance vectors". The next section discusses on the working mechanism of RVM.

The mapping relationship with this work falls under the category of Multiple Input Single Output shortly called as MISO. The input vectors may be n and the output is coined as Independent Identically Distributed nature. It is noteworthy to gain from the observations that it contains mean-zero Gaussian noise with variance observations.

The complexity of the model along with over-fitting avoidance zero mean Gaussian prior probability mechanism is used. *The* kernel is designed with Gaussian function along with sparse Bayesian learning framework. *C*ross validation on the validation set is used to get good unified kernel width.

3.4 Analytical Hierarchy Process Method

The analytical hierarchy process (AHP) method is one of the extensively used multi-criteria decision-making (MCDM) methods. One of the main advantages of this method is: it can effectively handle both qualitative and quantitative data, uses of AHP do not involve cumbersome mathematics, and it is very easy to handles multiple criteria and easier to understand. Decomposition, pair-wise comparisons and priority vector generation and synthesis are the principles of AHP.

The AHP is an approach based on pair-wise comparisons, which can be used to determine relative weights of individual criteria or alternatives. Pair-wise comparisons are usually quantified by the linear scale or the nine-point intensity scale. By the linear scale, each linguistic phrase is mapped to one value in a set of available values. A matrix that has a pair-wise comparison is generated.

3.5 The Levenberg–Marquardt training algorithm

A mathematical description of the LM neural network training algorithmic rule has been given by Hagan and Menhaj [12]. The LM algorithmic rule was originally designed to function an intermediate optimization algorithmic rule between the Gauss–Newton (GN) methodology and gradient descent algorithmic rule, and address the restrictions of every of these techniques.

The LM algorithmic rule is thought of a trust-region modification to GN methodology or a bridge between GN method and gradient descent algorithmic rule. The GN method has glorious quadratic convergence properties. However, these properties are extremely dependent on the initial values. If the initial values are not designated properly, this algorithmic rule might simply diverge. Additionally, the accuracy of the GN methodology becomes marginal once it reaches the proximity of a minimum.

Gradient descent algorithmic program, on the opposite hand, though is restricted by a slow speed of convergence, exhibits wonderful behavior within the locality of a minimum purpose. The performance of gradient descent algorithmic program is additionally not adversely affected by the selection of the initial values. The update rule for the weights of the neural network is carried out using the equation (3) as stated below,

$$W= - [V^2E(W)]^{-1}.VE(w) \quad (3)$$

Where $V^2E(W)$ represents the Laplacian of the energy function, and is also referred to as the Hessian matrix. The Hessian term can be written as:

$$V^2E(W)=J^T(W).J(W)+S(W) \quad (4)$$

$$S(W)=\Sigma^N_{t-1}e_t(W).V^2e_1(W) \quad (5)$$

Where the term $e_1(W)$ denotes the error vector of the neural network for pattern I and J (W) symbolizes the Jacobin Matrix.

The term second derivative is incredibly costly to calculate because the quantity of computations will augment exponentially with the dimensions of the network. The update rule for the GN method may be written as (6):

$$W= - [J^T(W).J(W)]^{-1}.J^T(W)e(W) \quad (6)$$

Given the above introduction, the LM modification to the GN method is as follows:

Note that when parameter μ is large, the above expression approximates gradient descent *(with learning rate 1/ μ)* even as a minuscule μ, the algorithm approximates the GN method. By adaptively adjusting the parameter μ, the LM algorithm can maneuver between its two extremes – the gradient descent and the GN algorithm. By doing so, the LM method can combine the advantages of gradient descent and the GN algorithms, while bypassing their limitations. Change in the μ parameter is performed similarly to the modification of the adaptive eradiating rate in back-propagation algorithm. When a step would result in an increased energy function, μ is multiplied by a constant μ_{Inc}>1 to drive the algorithm towards the incline descent algorithm and thus obtain more stability.

On the other hand, when a step would result in a decreased energy function, μ is multiplied by $\mu_{Dec}=1/\mu_{Inc}$ to drive the algorithm towards the GN algorithm, and thus gain more celerity. The stopping criteria for the LM method are similar to that of the BP method [16][17].

It is assumed that the algorithm have converged when the sum of squares has been reduced to approximate error goal.

FHRVM is a mechanism consists of the features of RVM, LM and AHP method. The next section discusses in the experimental results along with the dataset chosen.

4. EXPERIMENTAL RESULTS

KDD Cup'99 is initially used for The Third International Knowledge Discovery and Data Mining Tools Competition. There are round about 4,940,000 kinds of training data and 3,110,291 testing data in the KDD Cup'99 dataset. KDD Cup99 is an audited set of standard dataset which includes training and testing set. Data has the following four types of attacks as found below,

- DoS
- Probe
- U2R and
- R2L.

Data has the following four types of attacks as found below,

- TCP
- UDP and
- ICMP.

Totally 600 database are composed from the KDD cup99, these database are used for categorization. Here three type of organization are done, those are RVM and FHRVM.

4.1 Performance Measures

Assessment of the proposed FHRVM against RVM is is compared using the below mentioned performance metrics.

- Detection Rate
- False-alarm rate

Detection Rate: Accuracy can be mentioned as the proportion of attack detected among all attack data, such as true positive (TP). Thus detection rate is

$$\text{Detection Rate} = \frac{TP}{TP+FN}*100\%$$

False Alarm Rate: False alarm rate can be mentioned to the proportion that a normal data is falsely detected as attack behavior, such as false positive (FP). Thus false alarm rate is

$$\textbf{False Alarm Rate} = \frac{FP}{FP+TN}*100\%$$

Detection and identification of attack and non-attack behavior can be done using TP, TN, FP, and FN. The description is given below,

True Positive (TP): The number of attack detected when it is actually attack.

True Negative (TN): The number of normal detected when it is actually normal.

False Positive (FP): The number of attack detected when it is actually normal.

False Negative (FN): The number of normal detected when it is actually attack.

4.1.1 Correctly Detected Attack

The acceptably detected assault is based on the precision of the four attacks. The results of correctly detection rate for different types of attacks and protocols are shown in Fig.3 and Fig.4 respectively. From the results it is observed that the proposed FHRVM has high detection accuracy than RVM.

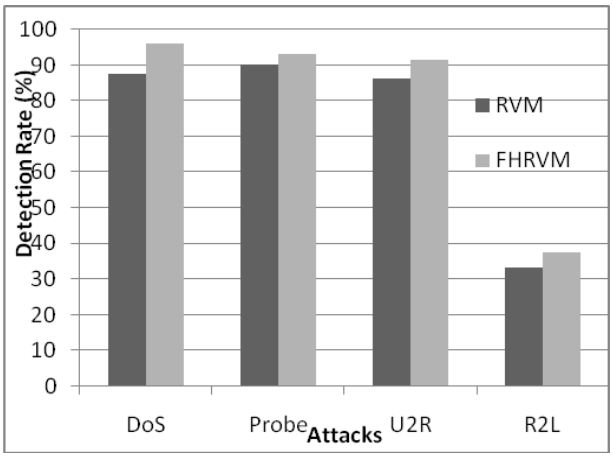

Fig.3. Comparison of Detection Rate against Attacks

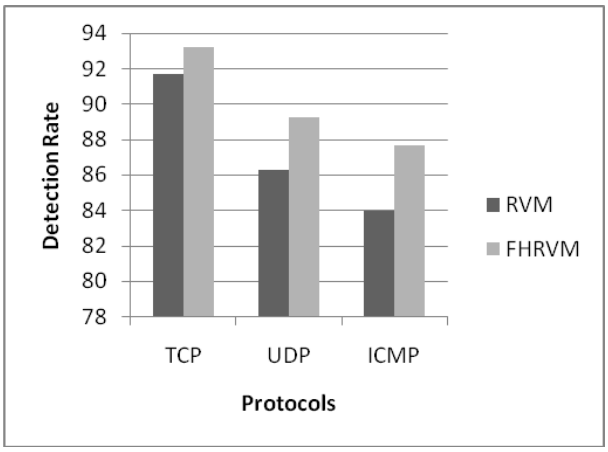

Fig.4. Comparison of Detection Rate against Protocsols

4.1.2 False Alarm rate

False alarm rate indicates the percentage of normal data which is erroneously recognized as attack. The false alarm rate comparison against attacks and protocols is shown in Fig.5 and Fig.6 respectively. From the results it is experimented that the proposed FHRVM has comparatively less false alarm rate than conventional RVM.

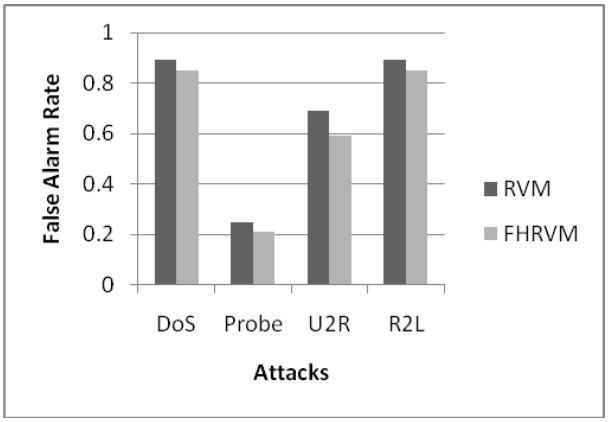

Fig.5 Comparison of False Alarm Rate against Attacks

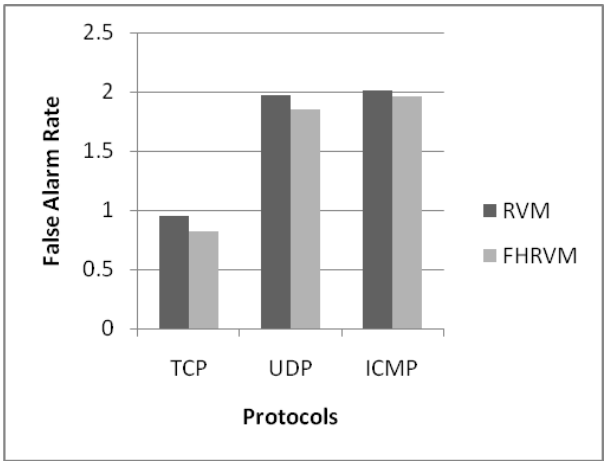

Fig.6 Comparison of False Alarm Rate against Protocsols

The next section concludes the research work followed with references.

5. CONCLUSION

At present, sanctuary within the network communication is of a momentous concern. Being the authentic fact that information are thought of because the valuable quality of an organization, giving the security against the interlopers is incredibly essential. Intrusion detection system tries to identify security attacks of intruders by investigating many information records determined in processes on the network. This research work proposes a Fast Hierarchical Relevance Vector Machine for detection of attacks that are happens more in networked environment. The performance of the proposed FHRVM is compared with the conventional RVM and the simulation results have proven that the proposed mechanism FHRVM is better in terms of detection rate and false alarm rate. As a future work, numerous training algorithms are utilized to enhance its performance.

6. REFERENCES

[1] K.Vivek . Kshirsagar, M.Sonali . Tidke & Swati Vishnu, "Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012

[2] D. Anderson, T. Frivold and A. Valdes, "Next-generation intrusion detection expert system (NIDES): a summary", Technical Report SRI-CSL-95-07. Computer Science Laboratory, SRI International, Menlo Park, CA, 1995.

[3] S. Axelsson, "Research in intrusion detection systems: a survey", Technical Report TR 98-17 (revised in 1999). Chalmers University of Technology, Goteborg, Sweden, 1999.

[4] S. Freeman, A. Bivens, J. Branch and B. Szymanski, "Host-based intrusion detection using user signatures", Proceedings of the Research Conference. RPI, Troy, NY, 2002.

[5] D. Marchette, "A statistical method for profiling network traffic", Proceedings of the First USENIX Workshop on Intrusion Detection and Network Monitoring, Santa Clara, CA, Pp. 119–128, 1999.

[6] R.G. Bace, "Intrusion Detection", Macmillan Technical Publishing, 2000.

[7] F. Chuan, P. Jianfeng, Q. Haiyan, and W. R. Jerzy, "Alert fusion for a computer host based intrusion detection system," in Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, 2007, pp. 433-440.

[8] W. Rensheng and V. N. Jeffrey, "Search strategy optimization for intruder detection," IEEE Sensors Journal, Vol. 7, 2007, pp. 315-316.

[9] J. G. Tront and R. C. Marchany, "Internet security: Intrusion detection and prevention in mobile systems," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007, pp. 162-162.

[10] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi and Lilly Suriani Affendey, "Intrusion Detection Using Data Mining Techniques", IEEE, 2010. ISBN: 978-1-4244-5651-2/10

[11] Avolio, M.Frederick. "A multidimensional approach to internet security." netWorkervol. 2.2, April 1998, pp. 15-22.N.S.Chandolikar1 & V.D.Nandavadekar2, "Comparative Analysis Of Two Algorithms For Intrusion Attack Classification Using Kdd Cup Dataset", International Journal of Computer Science and Engineering (IJCSE) Vol.1, Issue 1 Aug 2012 81-88 © IASE.

[12] R.J. Kuo, S.C. Chi, S.S. Kao, "A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network," Computers in Industry, vol. 47, pp 199-214, 2002.

[13] T.L. Saaty, "Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process," vol. VI, RWS Publication, Pittsburgh, 1994.

[14] Y.-C. Hu, J.-F. Tsai, "Backpropagation multilayer perception for incomplete pair wise comparison matrices in analytic hierarchy process," Applied Mathematics and Computation, vol. 180, pp 53–62, 2006.

[15] B.G. Kermani, S.S. Schiffman, H.T. Nagle, "Performance of the Levenberg–Marquardt neural network training method in electronic nose applications," Sensors and Actuators B, vol. 110, pp 13-22, 2005.

[16] M. Hgan, H. Demuth, and M. Beale, "Neural Network Design," Boston: PWS, 1996.

[17] V.Jaiganesh , Dr.P.Sumathi "An Efficient Intrusion Detection using Fast Hierarchical Relevance Vector Machine",JATIT, ISSN: 1992 - 8645, E- ISSN: 1817-3195,, Volume No: 62, Issue No: 1, April 2014