Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT INTRUSION DETECTION USING FAST HIERARCHICAL RELEVANCE VECTOR MACHINE

[1] V. JAIGANESH, [2] Dr. P. SUMATHI

[1] Doctoral Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, TN, India.

[2] Doctoral Research Supervisor, Assistant Professor, PG & Research Department of Computer Science, Government Arts College, Coimbatore, Tamilnadu, India.

E-mail: [1]jaiganeshree@gmail.com, [2]sumi_rajes@yahoo.com

ABSTRACT

Internet is a useful source of information in everyone's daily activity. Internet becomes a globally used public network. Significance of Intrusion detection system (IDS) in computer network security is well proven. In order to protect the organization data, Intrusion Detection System (IDS) offers protection from external users and internal attackers. Intrusion detection is the processes of examining the events which happens in a computer system or network and evaluate them for signs of possible events, which are imminent threats of violation of computer security policies, standard security practices and acceptable use policies. In this paper a new algorithm is introduced to find out the protocol and their attacks are Fast Hierarchical Relevance Vector Machine (FHRVM). These algorithms are formed by the combination of RVM with LM-AHP. The main goal of the paper, that successful identification of attacks in reduced false alarm rate. To demonstrate this by exploiting a probabilistic Bayesian learning framework, in this paper derive an accurate prediction models which typically utilize dramatically fewer basis functions than a comparable KSVM and FHELM. These include the benefits of probabilistic predictions, automatic estimation of `nuisance' parameters and the facility to utilize arbitrary basis functions. The experiment is carried out with the help of MATLAB by using KDD Cup 1999 dataset and the results indicate that the proposed technique can achieve higher detection rate and very low false 7alarm rate than the regular KSVM, FHELM algorithms.

Keywords: *Intrusion Detection System(IDS), Support vector machine (SVM), Extreme Learning Machine (ELM),Relevance Vector Machine (RVM), Levenberg-Marquardt (LM)*

1. INTRODUCTION

Information security could be a matter of significant worldwide concern because the incredible development in connectivity and accessibility to the web has generated an incredible security threat to information systems worldwide [1]. Security is turning out to be a significant issue because the web applications are developing continuously. The major security systems mostly focusing on encryption, firewall and access control. However all these approaches cannot promise perfect security. Hence, the security of a system can be improved with the introduction of intrusion detection system.

The capability of IDS in categorizing a wide range of intrusions in real time with correct results is more significant. The patterns of user behavior and inspection records are observed and the intrusions are traced [2].Intrusion detection attacks are segmented into two groups namely,

- Host-based attacks [3-5] and
- Network-based attacks [6, 7].

In case of a host-based attacks, the intruders aim at a particular machine and attempts to get access to privileged services or resources on that particular machine. Detection of these kinds of attacks typically uses routines that acquire system call data from an audit-process which monitors all system calls made with the support of each user. In case of network-based attacks, it is extremely complicated for legitimate users to use various network services by purposely occupying or disrupting network resources and services. Intruders attack these systems by transmitting huge amounts of network traffic, utilizing familiar faults in networking services, overloading network hosts, etc.

Detection of these kinds of attacks uses network traffic data (i.e., tcp dump) to look at traffic addressed to the machines being monitored. Various types of intrusion detection have developed and it does not recover the needful for the internet atmosphere [8, 9]. Nowadays modern computer systems are mostly used by IDS and it is necessary also. This type of Intrusion detection technologies are classified into two major groups:

- Misuse detection and
- Anomaly detection.

A misuse detection system traces intrusion activities that follow recognized patterns. These patterns justify a suspect collection of sequence of activities or operations that probably is dangerous. The most important disadvantage of this detection is that it doesn't have the potential to trace or detect new reasonable intrusions, i.e., certain events that have not occurred the past. An anomaly detection system examines event knowledge and identifies pattern of activities that seem to be standard. If an event lies outside of the patterns, it's thought-about as a potential intrusion [18].

But this type of method is not sufficient to find out the new kind of attacks. So to overcome this Support Vector Machine (SVM) are handled in this paper are used in IDS. It is one of the type of machine learning method and it plot the in high dimensional feature space, labeling each vector by its class. SVMs categorize data by finding out a collection of support vectors, which are members of the set of training inputs that outline a hyper plane in the featurespace.SVM are the kind of classifiers, in this paper SVM is reinitialize developed for binary classification [10] and it can be exploited to classify the attacks.

Relevance Vector Machine (RVM) and Extreme Learning Machine (ELM) has been used in modeling, control and several other applications [21],[22],[23] and [24]. The significance of this research work is to provide effective machine learning using several mechanisms. Since RVM, ELM have contributed high in various research dimesions, this research work provides network intrusion detection system to learn intrusions and other perspectives. Analytical Hierarchical Process (AHP) method [25] also used for decision support.

Speed and scalability is one of the major advantages of SVM, so SVM is used IDS. More over the classification complexity of SVM does not based on the dimensionality of the feature space and hence they can potentially learn a huge collection of patterns. Also SVM provides a standard mechanism to fit the surface of the hyper plane to the data by utilizing the kernel function. In this paper, radial basis kernel function of SVM is tuned using Levenberg-Marquardt (LM) learning and tested using KDD Cup 1999 dataset based on the detection rate and very low false alarm rate.

2. LITERATURE REVIEW

A model is obtainable by Zhang Hongmei [11] to produce a solution for the crisis of low accurateness and high false alarm rate in network model of SVM with the assistance of rough set feature that is employed to reduce. Rough set approach will be used to take care of the reduction of original datasets, the initial dataset is computed and is employed to train individual SVM classifier for assortment and increase the variability between individual classifiers, and to extend the possibility of discovering accuracy increasing as a result.

SVM has been confirmed to be an efficient classifier in several applications, however its practice is still imperfect as the data allocation information is underutilized in determining the result hyper plane. The majority of existing kernels are working in nonlinear SVMs establish the likeness between a pair of pattern images depending on the Euclidean inner product or the Euclidean detachment of equivalent input patterns, that omits data sharing tendency and makes the SVM fundamentally a classifier. Toward a paradigm of kernels, Defeng Wang et al., [12] gives some idea for exact data knowledge into existing kernels. By using Agglomerative Hierarchical Clustering (AHC) to locate out the data structure primary for each class adaptively in the input space, after this process by using data distribution information built up the Weighted Mahalanobis Distance (WMD) kernels. With the help of cluster size also find out the Similarity between two images not only by Mahalanobis Distance (MD).However WMD kernels are not assured to be Positive Definite (PD) or Conditionally Positive Definite (CPD), acceptable categorization outcome can still be achieved of regularizes in SVMs with WMD kernels are empirically optimistic in pseudo-Euclidean (pE) spaces.

The individuality of security attacks has to be one of the aims to develop the intention towards novel intrusion detection approaches, different from those exists in conventional networks. Josephet al., [13] developed an autonomous host-dependent IDS for identifying malicious sinking behavior. With the help of cross-layer features, the detection accuracy is increase by the system and also it examines the routing behavior.

For learning and adjustment to new kind of attack circumstances and network surroundings, two machine learning approaches are exploited. SVMs and Fisher Discriminate Analysis (FDA) are utilized collectively to develop better accuracy of SVM and quicker speed of FDA. Rather than using all cross-layer features, features from MAC layer are connected/interrelated with features from additional layers, in that way reducing the feature set without reducing the information content.

3. METHODOLOGY

For information sharing networks are used by various multi-level organizations. On the other hand, the usage of networks has paved the way for intruders to attack the communication path and to steal the valuable asset (data) of any organization. Therefore necessity of Intrusion Detection System developed in to an essential and significant field of research.

More number of techniques is present today and gives more security to the network, but most of these techniques are static. To overcome this one of the dynamic state is the intrusion detection it preserves the network security by examines the attack. Three methods are carried on this section to provide a potential solution for IDS problem.

3.1 Kernelized SVM (KSVM)

A kernel function and its parameter have to be chosen to build a SVM classifier. Three main kernel functions have been used to build SVM classifiers. They are

- Linear Kernel Function, $K(x,z) = \langle x,z\rangle$
- Polynomial Kernel Function, $K(x,z) = (\langle x,z\rangle + 1)^d$, d is the degree of polynomial.
- Radial basis function $K(x,z) = \exp\{\frac{-|x-z|^2}{2\sigma^2}$, σ is the width of the function.

This section introduces a novel Levenberg-Marquardt learning for tuning the sigma (σ) in a RBF kernel of SVM. In this case, each attribute has its individual sigma constraint related with it. The values of the tuned σ are then utilized as a measure for variable selection [15].

3.1.1 Sigma Tuning Algorithm

In this section, the sigma tuning algorithm is explained. MetricQ^2 is selected as an error metric, represented as E(σ), which is based on the vector σ, Leave-One-Out (LOO) Kernel Partial Least Squares (K-PLS) is used to obtain an initial Q_0^2, value depends on an initial starting guess for the sigma-vector rep[resented as σ_0, a second-order gradient descent technique is exploited to reduce the objective function E(σ), discover the optimal choice for σ. The search process begins from the initial point $E(\sigma_0) = Q_0^2$, The value of σ, is transformed based on the minimization of the leave-one-out tuning error, instead of directly minimizing the training error (Figure 1.1).

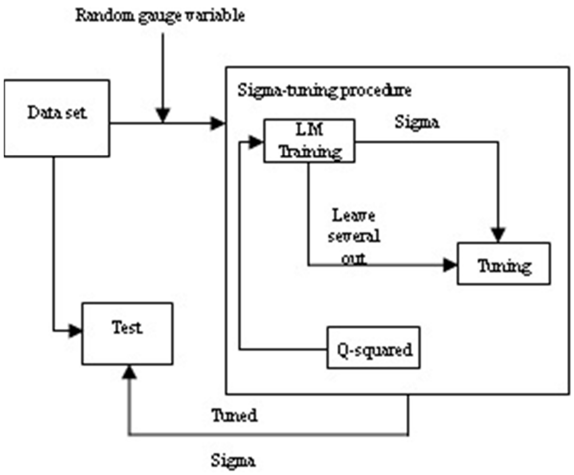

Figure 1: Process flow for Sigma Tuning in RBF Kernel of SVM using LM Training

Based on the Newton's rule for finding a least amount in a multi-dimensional space, the relation between e(σ) and σ at the minimum can be given as,

$$\sigma = \sigma_0 - H^{-1}\forall E(\sigma_0) \qquad (1)$$

Where h is the hessian matrix. $\nabla e(\sigma_0)$ represents a vertical vector, as represented as:

$$\forall E(\sigma_0) = \forall E(\sigma) | \sigma = \sigma_0 = \begin{pmatrix} \frac{\partial E}{\partial \sigma_1} | \sigma = \sigma_0 \\ \vdots \\ \frac{\partial E}{\partial \sigma_m} | \sigma = \sigma_0 \end{pmatrix} \qquad (2)$$

After rearranging, the equation can be transformed as

$$H\Delta\sigma = -\nabla E(\sigma_0) \qquad (3)$$

Where $\Delta(\sigma = \sigma - \sigma_0)$. With the intention of efficiently proceeding towards a converged solution, Levenberg-Marquradt learning is utilized.

3.1.2 Levenberg-Marquradt Learning

Levenberg-Marquardt (LM) learning developed from expansion of Error Back Propagation (EBP) algorithm dependent techniques. It provides a significant exchange between the speed of the Newton algorithm and the strength of the steepest descent technique. There are the two fundamental theorems of LM algorithm. In the back-propagation algorithm, the performance index F(w) to be reduced is defined as the sum of squared errors between the target outputs and the network's simulated outputs,

$$F(w)=e^{T}e \qquad (4)$$

where w = [w_1,w_2…,w_n] includes all weights of the network, e indicates the error vector comprising the error for all the training samples.

The increment of weights Δw, when training with the lm method is calculated as follows:

$$\Delta w=[J^{T}J]+\mu I]-1J^{T}e \qquad (5)$$

where J denotes the jacobian matrix, μ represents the learning rate which is to be updated using the β depending on the resultant. Especially, μ is multiplied by decay rate β(0<β<1) whenever F(w) diminishes, while μ is divided by β whenever F(w) increases in a new step.

This is achieved by adding a small scalar λ to the diagonal elements in the HessianH, as expressed by:

$$(\mathrm{H}+\lambda I)\Delta\sigma=-\nabla E(\sigma_{0}) \qquad (6)$$

In this method, the algorithm begins with a first-order approach and regularly proceeds towards the second-order approach outlined below. The equation (6) is solved for Δσ. It is to be noted that each element $\frac{\partial E}{\partial \sigma}|_{\sigma=\sigma 0}$ in the right side of equation (2) will be computed by numerical perturbation as given below:

$$\frac{\partial E}{\partial \sigma}|_{\sigma=\sigma 0}\approx\frac{\Delta E}{\varepsilon}|_{\sigma=\sigma 0}=\frac{E(\sigma_{i}+\varepsilon)}{\varepsilon}|_{\sigma=\sigma 0} \qquad (7)$$

Where ε represents a small perturbation value acting on the ith component in σ. E(σ_i) is the performance metric Q^2obtained from the change in the ith component of σ only.

A second approximation will be introduced before solving the above equations. Because the elements of the Hessian are expensive to evaluate, a fast and efficient approximation for the Hessian matrix is introduced.

In general, the second partial derivatives can be numerically computed. On the other hand, in order to speed up the calculation process, the second-order partial derivatives are approximated.

Δσ is then solved numerically from equation (6) with a fast conjugate gradient based equation solver in order to avoid calculating the inverse of the Hessian matrix, H. As a result of the fairly accurate evaluation of the Hessian, a heuristic coefficient α will be introduced in the iterative updating procedure for the elements.

3.2 Proposed Fast Hierarchical Extreme Learning Machine (FHELM)

In this section, a FHELM approach using ELM and LM algorithm. The Proposed FHELM which makes use of both the advantage of ELM and LM. First calculate the output weight by using the ELM algorithm then update the parameter with the help of LM algorithm.

3.2.1 Extreme Learning Machine

With in the case of learning an arbitrary function with zero training error, Baum had offered several constructions of SLFNs with adequate hidden neurons [14]. Though, in practice, the number of hidden neurons required to accomplish an appropriate generalization performance on novel patterns is much less. The resulting training error may not approach to zero have been minimized.

ELM at random allocates and fixes the input weights w_i and biases b_i supported on some constant probability distribution function in the case of learning a structured function. The exceeding problem is finely established and acknowledged as a linear system optimization problem. The overview of performance of a SLFN leans to be enhanced with smaller magnitude of output weights.

ELM Algorithm known to be a training of

(i) At random assign input weights w_i and biases b_i according to some constant probability density function.

(ii) Evaluate the hidden layer output matrix H.

(iii) Evaluate the output weights

In addition to the probability density function is a standardized distribution function in the range from −1 to 1. Then LM is applied here to compare LM for each input from ELM classifier.

3.3 Proposed Fast Hierarchical Relevance Vector Machine (FHRVM)

The Relevance Vector Machine (RVM) was introduced by as a Bayesian counterpart. In this section, a FHRVM approach using RVM, LM and

AHP algorithm. The proposed FHRVM makes use of both the advantage of RVM with LM and AHP. The hierarchical model is formulated using AHP which is most helpful for decision support. The Relevance Vector framework provides a means for solving regression and classification problems, but in this paper look for models which are highly sparse by selecting a subset from a larger pool of candidate kernel functions (one for each example in the training set). A key concept is the use of continuous hyper parameters to govern model complexity and thereby avoid the in- tractable problem of searching over an exponentially large discrete space of model structures.

In order to reduce the dimensionality of the hyper parameter space, specify a prior structure which reflects the possibility of correlation between the hyper parameters of the coefficients distribution and hence it is possible to segregate a unique solution.

RVM has been used for classification in the proposed method. Relevance vector machine (RVM) is a special case of a sparse linear model in which the basic functions are formed by a kernel function. AHP is used for hierarchical model that aid in decision support.

The sparseness property of the RVM enables choosing proper kernel automatically at each location by pruning all irrelevant kernels; hence it is possible that two different kernels remain on the same location.

RVM has several advantages which includes the number of relevance vectors can be much smaller than that of support vectors , RVM does not need the tuning of a regularization parameter (C) as in SVM during the training phase. Thus the proposed dataset can be classified using RVM classifier.

At first the input weights are generated and followed by hidden biases. This method uses the analytical hierarchy process method to select the input weights and hidden biases. Then, the corresponding outputs weights are analytical computed by using the RVM algorithm once and randomly generate the output hidden biases. Then, update these parameters (all weights and biases) by using LM algorithm.

3.3.1 Significance of Analytical Hierarchical Process (AHP) in FHRVM

AHP is a multi-criteria decision-making approach has attracted the interest of many researchers mainly due to the nice mathematical properties of the method and the fact that the required input data are rather easy to obtain.

The AHP is a decision support tool which can be used to solve complex decision problems. It uses a multi-level hierarchical structure of objectives, criteria, sub criteria, and alternatives. The pertinent data are derived by using a set of pair wise comparisons. These comparisons are used to obtain the weights of importance of the decision criteria, and the relative performance measures of the alternatives in terms of each individual decision criterion. If the comparisons are not perfectly consistent, then it provides a mechanism for improving consistency.

The basic procedure to carry out the AHP consists of the following steps:

- Structuring a decision problem and selection of criteria
- Priority setting of the criteria by pair wise comparison (weighing)
- Pair wise comparison of options on each criterion (scoring)
- Obtaining an overall relative score for each option

3.3.2 Significance of Levenberg-Marquradt Learning in FHRVM

The LM algorithm combines the significant properties of steepest descent method and Gauss-Newton algorithm by inheriting the speed advantage of Gauss-Newton algorithm and utilizes the stability nature of steepest descent method. The LM algorithm is a combined training process that switches the steepest descent algorithm towards quadratic approximation along with speed convergence.

4. RESULTS AND DISCUSSIONS

The proposed IDS are experimented using MATLAB and the dataset used is KDD Cup99 dataset. MATLAB is used to evaluate the intrusion detection. KDD Cup99 dataset comes from DARPA 98 Intrusion Detection Evaluation handled by Lincoln laboratory at MIT [17].

It shows how the results that have found in this proposed FRVM. This gives the way to highlight research reflects, differs from and extends current knowledge of the area intrusion detection. Results provided by the performance analysis explain the findings.

The performance measure used to evaluate the proposed KSVM, FHELM and FHRVM against RVM, SVM and ELM. KDD Cup99 is an audited set of standard dataset which includes training and testing set.

The data has been tested in the following three major protocols such as TCP, UDP and ICMP.

Data has the following four major groups of attacks such as DoS, R2L, U2R and Probing. The metrics that are taken into account are as below,

- False Positive (FP): Corresponds to the number of detected attacks but it is actually normal.
- False Negative (FN): Corresponds to the number of detected normal instances but it is really an attack. These attacks are the major target of intrusion detection systems.
- True Positive (TP): Corresponds to the number of detected attacks and it is in fact attack.
- True Negative (TN): Corresponds to the number of detected normal instances and it is actually normal.

The performance measure used to evaluate the proposed KSVM, FHELM and FHRVM is

- Detection rate and
- False-alarm rate

The accuracy of an intrusion detection system is computed based on the detection rate and false alarm rate.

Detection rate indicates the percentage of detected attack among all attack data, and is given as,

Table 1: Comparison Table for Detection Rate of Protocols and Attacks

METHODS	DETECTION RATE							FALSE ALARM RATE						
	TCP (%)	UDP (%)	ICMP (%)	DoS (%)	Probe (%)	U2R (%)	R2L (%)	TCP (%)	UDP (%)	ICMP (%)	DoS (%)	Probe (%)	U2R (%)	R2L (%)
SVM	89.59	81.24	80.89	85.84	89.09	85.84	15.90	1.98	2.89	2.67	1.92	0.68	1.32	1.92
KSVM	91.56	83.24	83.19	90.12	90.26	90.12	26.95	1.62	2.64	2.45	1.26	0.31	0.84	1.26
ELM	90.19	85.27	83.32	85.84	85.84	66.67	15.9	1.74	2.47	2.48	1.34	0.45	0.79	1.34
FHELM	92.46	87.51	86.34	93.39	89.39	73.65	26.95	1.54	2.16	2.13	1.24	0.30	0.75	1.24
RVM	91.67	86.24	83.95	87.65	90.04	86.24	33.12	0.95	1.97	2.01	0.89	0.25	0.69	0.89
FHRVM	93.22	89.25	87.64	95.89	93.24	91.53	37.42	0.82	1.85	1.96	0.85	0.21	0.59	0.85

$$Detection \quad Rate = \frac{TP}{TP + FN} \times 100$$

$$False \ Alarm \ Rate = \frac{FP}{FP + TN} \times 100$$

Table 1 gives the comparison table for the detection rate and false alarm rate of protocols. Protocols are TCP, UDP, and ICMP and the attacks taken are DoS, Probe, U2R and R2L. These protocols are detected by proposed system of KSVM, FHELM and FHRVM. From the table it is clearly noticed that the proposed FHRVM proves the better results against proposed KSVM and FHELM.

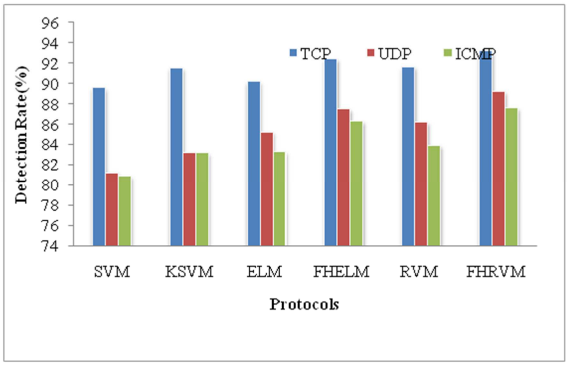

Figure 2: Comparison Graph for Detection Rate

Figure 2 and 3 shows the detection rate for protocols and attacks. It can be concluded that the detection rate of proposed FHRVM shows maximum percentage when comparing with proposed FHELM and KSVM and also with SVM, ELM and RVM.

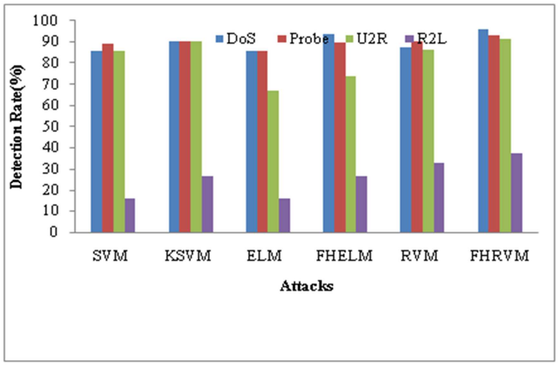

Figure 3: Comparison Graph for Detection Rate

Figure 4 and Figure 5 shows the false alarm rate for protocols and attacks. It can be concluded that the false alarm rate of the proposed FHRVM shows lesser percentage when comparing with proposed KSVM and FHELM and also with SVM, ELM and RVM.

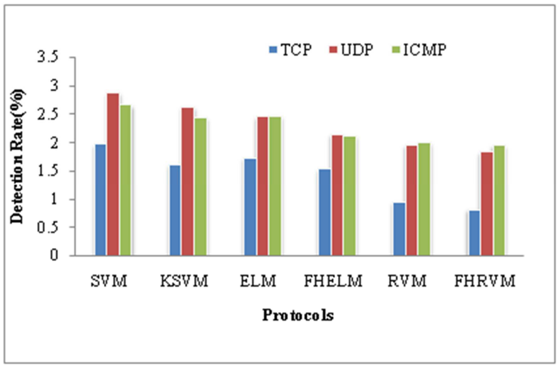

Figure 4: Comparison Graph for false alaram rate Protocols

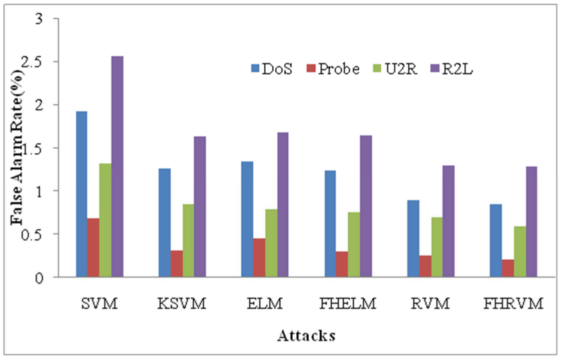

Figure 5: Comparison Graph for False Alarm Rate of Attacks

5. CONCLUSIONS AND FUTURE WORKS

At present, security inside the network communication is of a major concern. Being the fact that data are considered as the valuable asset of an organization, providing security against the intruders is very essential. Intrusion detection system utmost identifies security attacks of intruders by investigating several data records observed in processes on the network. But since an efficient mechanism is still in need. This research proposed an Efficient Intrusion detection system that is formed by using Modified Classification process of SVM, ELM and RVM. RVM Classification is for good generalization performance. LM algorithm is used to obtain a converged solution. AHP is the method which is capable enough for qualitative and quantitative data classification with decision support.

Hence the proposed FHRVM method proves better result against proposed KSVM and FHELM. The experiment is carried out in MATLAB by using KDD Cup 1999 dataset and the results indicate that the proposed system of FHRVM can provide better detection rate and low false alarm rate than the proposed KSVM and FHELM. As a future work, the proposed mechanism can be implemented in real time scenarios.

REFRENCES:

[1] S. Mukkamala, G. Janoski and A. Sung, “Intrusion detection using neural networks and support vector machines”, Proceedings of International Joint Conference on Neural Networks (IJCNN '02), Vol. 2, Pp. 1702–1707, 2002.

[2] Snehal A. Mulay, P.R. Devale and G.V. Garje, “Intrusion Detection System using Support Vector Machine and Decision Tree”, International Journal of Computer Applications, Vol. 3, No.3, Pp. 40-43, 2010.

[3] 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.

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

[5] 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.

[6] K. Ilgun, R.A. Kemmerer and P.A. Porras, “State transition analysis:A rule-based intrusion detection approach”, IEEE Trans. SoftwareEng,Vol. 21, No. 3, Pp. 181–199, 1995.

[7] 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.

[8] R.G. Bace, “Intrusion Detection”, Macmillan Technical Publishing, 2000.

[9] B.V. Dasarathy, “Intrusion detection, Information Fusion”, Vol. 4, No. 4, Pp. 243-245, 2003.

[10] M. Cramer, James Cannady and Jay Harrell,“New Methods of Intrusion Detection using Control-Loop Measurement”,Proceedings of the Technology in Information Security Conference, Pp 1-10, 1995.

[11] Zhang Hongmei, “SVM ensemble intrusion detection model based on Rough Set feature reduct”,Chinese Control and Decision Conference (CCDC '09), Pp. 5604–5608, 2009.

[12] Defeng Wang,D.S. Yeung and E.C. Tsang, “Weighted Mahalanobis Distance Kernels for Support Vector Machines”, IEEE Transactions on Neural Networks, Vol. 18, No. 5, Pp. 1453-1462, 2007.

[13] J.F.C.Joseph, Bu-Sung Lee,A. Das and Boon-Chong Seet, “Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FDA”, IEEE Transactions on Dependable and Secure Computing, Vol. 8, No. 2, Pp. 233-245, 2011.

[14] Wun-Hwa Chen, Sheng-Hsun Hsu and Hwang-Pin Shen, “Application of ELM and ANN for intrusion detection”, Computers & Operations Research, Vol. 32, Pp. 2617-2634, 2003.

[15] Long Han, Mark J. Embrechts, Boleslaw K. Szymanski, Karsten Sternickel and Alexander Ross, “Sigma Tuning of Gaussian Kernels Detection of Ischemia from Magnetocardiograms”, Computational Modeling and Simulation of Intellect: Current State and Future Perspectives, Pp. 1-14, 2011.

[16] Witten, I. H., and Frank E. (1999) DataMining: Practical Machine Learning Tools and Techniques with Java Implementations,Morgan Kaufmann, San Francisco.

[17] KDD Cup network intrusion dataset,http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[18] Kyaw Thet Khaing, “Enhanced Features Ranking and Selection using Recursive Feature Elimination(RFE) and k-Nearest Neighbor Algorithms in Support Vector Machine for Intrusion Detection System”, International Journal of Network and Mobile Technologies, Vol. 1,No. 1, Pp. 8-14, 2010.

[19] Dimitris G. Tzikas , Liyang Wei, Aristidis Likas, Yongyi Yang and Nikolas P. Galatsanos, “A Tutorial On Relevance Vector Machines For Regression And Classification With Applications”.

[20] Pijush Samui, Gautam Bhattacharya, Sarat Kumar Das, ., “Support Vector Machine and Relevance Vector Machine Classifier in Analysis of slopes”, International Association for Computer Methods and Advances in Geomechanics (IACMAG) 1-6 October, 2008.

[21] Pak-Kin Wong, Qingsong Xu, Chi-Man Vong, Hang-Cheong Wong, " Rate-Dependent Hysteresis Modeling and Control of a Piezostage Using Online Support Vector Machine and Relevance Vector Machine ," IEEE Trans. on Industrial Electronics, vol. 59, no. 4, pp. 1988-2001, Apr 2012.

[22] Z. Bai, G.-B. Huang, D. Wang, H. Wang and M. B. Westover, "Sparse Extreme Learning Machine for Classification," (in press) IEEE Transactions on Cybernetics, 2014.

[23] G.-B. Huang, X. Ding, and H. Zhou, “Optimization Method Based Extreme Learning Machine for Classification”, Neurocomputing, vol. 74, pp. 155-163, 2010.

[24] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, “Semi-supervised and Unsupervised Extreme Learning Machines,” (in press) IEEE Transactions on Cybernetics, 2014.

[25] Yucheng Dong, Zhi-Ping Fan, and Shui Yu, " Consensus Building in a Local Context for the AHP-GDM with the Individual Numerical Scale and Prioritization Method," IEEE Transactions on Fuzzy Systems (accepted on February 11, 2014).