http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 7 (3): 510-515, 2008
ISSN 1812-5638

PCA-ICA Ensembled Intrusion Detection System by Pareto-Optimal Optimization

[1]Yu Gu, [2]Bo Zhou and [3]Jiashu Zhao
[1]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]Yunnan Institute of Technology and Information, Kunming 650000, China
[3]School of Science, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: A novel Ensemble Intrusion Detection System is proposed in this study. In this system, Principle Component Analysis (PCA) and Independent Component Analysis (ICA) feature extraction approaches are used to construct two Support Vector Machine (SVM) classifiers. Then the results are combined to pursue higher performance. Because the costs of false positive error and false negative error are asymmetric in IDS, we introduce Pareto-Optimal Approach to obtain the optimal weight for the ensemble system. Experiments on the data set KDD Cup 1999 Data show that the proposed system outperforms standard SVM, PCA SVM and ICA SVM.

Key words: Ensemble, intrusion detection, feature extraction, pareto-optimal

INTRODUCTION

Intrusion Detection Systems (IDS) is to recognize and notify the users' various security events, e.g., incidents and anomalies, which can be observed from users' behaviors. The challenge for an IDS is how to discover a users or programs behavior from large audit data. Machine learning approaches were successfully applied to intrusion detection due to their learning abilities. These approaches including but not limited to Artificial Neural Network (ANN) (Verwoerd and Hunt 2002; Joo *et al*., 2003; Zhang *et al*., 2003), Artificial Immune (Aickelin *et al*., 2004; Harmer *et al*., 2002), Markov Model (Ye *et al*., 2001; Du *et al*., 2004), Support Vector Machines (Mukkamala and Sung, 2003; Chen *et al*., 2005), are being used to build IDS.

In recent years, Support Vector Machine (SVM) is gaining much popularity as one of those effective methods for machine learning (Scholkopf *et al*., 1997). It has been found that SVMs perform better than neural networks for intrusion classifications (Sung and Mukkamala, 2004). Moreover, SVM is more suitable for intrusion detection, for it is faster while training (Mukkamala and Sung, 2003), independent of data dimension and can learn incrementally (Ralaivola and D'Alche-Buc, 2001).

Whereas, in the former studies (Gu *et al*., 2005), we found that false negative error is very high in intrusion detection by standard SVM algorithm. False negative error is incurred when the IDS does not function properly or mistakenly ignores an attack (Joo *et al*., 2003; Iheagwara *et al*., 2004). The cost of false negative error is higher than that of false positive error, which means IDS misinterprets normal packets or activities as attacks, because it reflects the possibility that the system is threatened (Joo *et al*., 2003; Iheagwara *et al*., 2004).

An intrusion detection ensemble system is presented in this study, which investigated two feature extract techniques, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) and SVM for classification. We discovered that PCA feature extraction had a low false positive error and a high false negative error and ICA feature extraction was opposite. Whereas, to combine these two costs into a single scalar objective function, we need incorporating a priori information into the aggregation method, which is a difficult task. Therefore, the multi-object optimization technology is adopted to obtain Pareto-front solutions. The final solution set is presented to user for estimation.

PCA-ICA ENSEMBLED IDS

For an intrusion system, when false negative error occurs, this means that the attack will succeed and the target resource will be damaged. In present experiment, standard SVM has a high false negative error, which is a big threat to the security of the system. The reason might be the complicated relativity among the features. Feature extraction is an unsupervised approach that aims to search the proper features in data. So the features extracted are likely to better reflect the essential characteristics of intrusion.

In this study, we have investigated the intrusion detection algorithm based on PCA and ICA for feature

Corresponding Author: Yu Gu, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

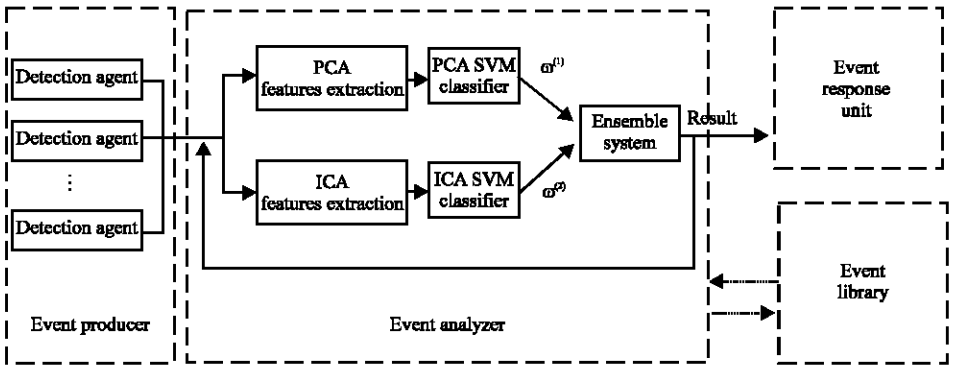

Fig. 1: Structure model of intrusion detection ensemble system

extraction, respectively. PCA is a multivariate statistics method and its basic idea is to seek a projection that best represents the data in a least-squares sense. Independent Component Analysis is also a linear transform technology. While Principal Component Analysis seeks directions in feature space that represent the data in a sum-squared error sense, Independent Component Analysis instead seeks directions that are most independent from each other.

The framework of the system is shown in Fig. 1. This system is composed of four parts: (1) Event producer; (2) Event analyzer; (3) Event response unit; (4) Event library. The mechanism of this ensemble classification system is as follow: each detecting Agent of event producer collects and unifies audit data and then submits the data to event analyzer. Principal component analysis and independent component analysis extract the data separately. SVM is used to classify the data extracted, viz. (Here, it means constructing an optimal hyperplane in two different spaces.) The optimal hyperplane thus can be constructed in the corresponding principle component space and independent component space, by using SVM. Then, event analyzer combined the outputs from two individual classifiers (PCA SVM and ICA SVM) to judge whether the event is an intrusion. If event analyzer verdicts that the event is suspect, the event response will adopt corresponding step or simply record it in log file. Event library is used to preserve support vector sets of PCA SVM and ICA SVM, the weight of the ensemble system and the rules learned from intrusion events, etc.

The learning method of combined classifiers in the system is called Ensemble. It was shown in many domains that an ensemble is often more accurate than any of the single classifiers in the ensemble (Brown *et al*., 2005).

ENSEMBLE WEIGHT OPTIMIZATION

In an Intrusion Detection System, the cost of false negative error is higher than of false positive error. However, with high false positive error, large amount of normal networks being regarded as intrusion by IDS will bring much trouble to the administrators. That is to say, the key point of IDS is how to make both false negative error and false positive error as low as possible. After investigating FP, FN of PCA SVM and ICA SVM, we can see: When PCA SVM classifier works, the system has low false positive error; while ICA SVM classifier works and the system has low false negative error. It showed that the errors of PCA SVM and ICA SVM distributed in the different position of error space. If we combine them, the ensemble might be an effective way to detect intrusion.

The optimization problem is:

$$\min \begin{cases} \text{FN\_error} \\ \text{FP\_error} \end{cases} \tag{1}$$

In present intrusion detection ensemble system, the output of event analyzer is obtained by weighted average of the output of PCA SVM and ICA SVM in order to optimize the upper formula.

$$y = \text{sgn}(\omega^{(1)}c^{(1)} + \omega^{(2)}c^{(2)}) \tag{2}$$

$\omega^{(k)}$ (k = 1, 2) are, respectively the weights of the two individual classifiers:

$$\begin{aligned} \omega^{(k)} &\geq 0 \\ \sum \omega^{(k)} &= 1 \end{aligned} \tag{3}$$

$c^{(k)}$ (k = 1, 2) are, respectively the output of PCA SVM classifier and ICA SVM classifier:

$$c^{(k)} = \sum \alpha_i^{(k)} y_i^{(k)} K(x_i^{(k)}, x_j^{(k)}) + b^{(k)}$$

where, K( ) is kernel function.

Traditional methods to solve (1) attempt to combine these two costs (false negative error and false positive

error) into a single scalar objective function (Pietraszek, 2004; Joo *et al*., 2003) and then find the optimal solution. A drawback to this approach is that we need incorporating a prior information into the aggregation method (Abbass, 2003). To choose a optimal tradeoff is difficult, for instance, in Pietraszek (2004), the authors selected an individual classifier that gives a good tradeoff between false positives and false negatives for cost ratio ICR = 50, while the ratio 5 is considered to be proper (Joo *et al*., 2003).

In this study, the two objective functions (false negative error and false positive error) are optimized simultaneously. Usually, the minimization of a certain objective (e.g., FP_error) implies degradation in another objective (FN_error). Since our aim is at minimizing two criteria during the search, it will produce a set of classifiers which is the tradeoff between the criteria. Therefore, we need to find the Pareto-optimal set in our problem. A solution $z\in\Omega$ dominates a solution z' and we write $z<z'$ if and only if $\exists k \in \{1,..., K\}$: $f_k(z)<f_k(z')$ and $\nexists k \in \{1,..., K\}$: $f_k(z)>f_k(z')$. The elements of the set $\{z|\nexists z' \in\Omega: z'<z\}$ are called Pareto-optimal.

In our experiments, the niched Pareto multiobjective genetic algorithm (NP-GA) (Kupinski and Anastasio, 1999) is used to maintain stable subpopulations of good solution. With NP-GA's, a niche method is used to maintain stable subpopulations of good solution. An objective vector is optimized instead of a scalar fitness function in GA. Before the selection is performed, the population is ranked based on the concept of dominance, incorporates the multiobjective nature of the problem into the selection mechanism. When two or more solutions in a tournament have the same rank, the winner of a tied tournament is the solution that has the smallest niche count. The niche count estimates the density of solutions in a localized region (niche) around an individual.

The complete algorithm of the evolution of the intrusion detection ensemble system can be described below:

Algorithm:

Input: The training set (x_i, y_i), i = 1, ..., *l* collected and packed by detecting Agent:

- **Step 1:** Project the training data to the corresponding feature space $(x_i^{(k)}, y_i^{(k)})$ by PCA and ICA and the process is
- **Step 1.1:** Principle component analysis
  Calculate the ith principal component adopting the following formula

$$e_1 = \arg\max_{\|e\|=1} E\{(e^T x)^2\}$$

$$e_k = \arg\max_{\|e\|=1} E\{[e^T (x - \sum_{i=1}^{k-1} e_i e_i^T x)]^2\},\ \ k = 2,\ldots,d$$

- **Step 1.2:** Independent component analysis
  ICA can be regarded as maximizing this function:

$$J_G(w_i) = [E\{G(w_i^T x)\} - E\{G(v)\}]^2,\ i = 1,2,\cdots,p$$

where, function G is usually

$$G_1(u) = \frac{1}{a_1}\log(\cosh a_1 u),\ 1 \le a_1 \le 2,\ \text{or}\ G_2(u) = -\exp(-u^2/2),$$

is a standard Gauss random variable.

- **Step 2:** Using SVM, separately search optimized classification hyperplane in the feature spaces of PCA and ICA.

$$\begin{aligned} &\min_{w,b,\xi}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{1}\xi_i\right),\\ &\text{s.t.}\ \ y_i^{(k)}\left(\left\langle w, \Phi(x_i^{(k)})\right\rangle - b\right) \ge 1-\xi_i,\\ &\qquad \xi_i \ge 0, \qquad i = 1,\cdots,l \end{aligned}$$

It turns to be a quadratic optimization problem as follows.

$$\begin{aligned} &\max_{a}\ \ w(a) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i^{(k)} y_j^{(k)} K(x_i^{(k)}, x_j^{(k)})\\ &\text{s.t.}\ \ \sum y_i^{(k)}\alpha_i^{(k)} = 0,\\ &\qquad 0 \le \alpha_i^{(k)} \le C, \qquad i = 1,\cdots,l. \end{aligned}$$

- **Step 3:** Find a set of Pareto optimal weights of ensemble system via NP-GA.

NP-GA divides each generation into several classes and selects some individuals with high fitness from each class to represent this class. Then generates a new group by crossover and mutation and selection by using sharing function.

The niche count m_i for the ith solution is given by

$$m_i = \sum_{j\in Pop} s(d_{ij})$$

where, d_{ij} is the distance between solutions i and j. $s(d_{ij})$ is the so-called sharing function given by

$$s(d_{ij}) = \begin{cases} 1 - \dfrac{d_{ij}}{\sigma_{share}}, & d_{ij} \le \sigma_{share} \\ 0, & d_{ij} > \sigma_{share} \end{cases}$$

Here, σ_{share} is called the niche radius, which represents the maximum distance between solutions that will result in an increase in their niche counts.

- **Step 4:** Output. Output the Pareto sets of the optimized weights in the ensemble system and store the information of support vector sets of PCA SVM and ICA SVM in Event library.

RESULTS AND DISCUSSION

The data set used in the experiments is KDD Cup 1999 Data (Blake and Merz, 1998), which is used as the experimental data set with many kinds of network intrusions simulated. The raw data includes a wide variety of intrusions simulated in the network. A connection is a sequence of TCP packets starting and ending at some well-defined times, between which data row to and from a source IP address to a target IP address under some well defined protocol. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type.

Because SVMs are only suitable for binary classifications, we partition the data into two classes of Normal and Attack (Probe, DoS, U2R, R2L) patterns, where the attack is the collection of the four classes of attacks instances in the data set. The (training and testing) data set contains 25 000 randomly generated points, 10 000 of which are training samples and 15,000 are testing samples. Because these data sets are randomly selected from KDD data set, the testing set contains 4 intrusion attacks that have not appeared in training set. We hope the system also performs well on these new intrusion behaviors.

Training is done using the Radial Bias Function (RBF) kernel option; an important point of the kernel function is that it defines the feature space in which the training set examples will be classified. In our former studies (Gu *et al*., 2005), we have also found that Gauss kernel is better than polynomial kernel, based on the comparison between the performances of Gauss function and polynomial function in intrusion detection.

Because of the former researching (Gu *et al*., 2005), we extract 9 dimensions in PCA and ICA feature extraction. The parameter of NP-GA is set as following: population size = 20, number of generations = 200, probability of crossover = 0.8, probability of mutation = 0.1 and niche distance (σ_{share}) = [0.25, 0.45].

We firstly discuss how the system performance is affected by the weights of PCA SVM and ICA SVM.

Figure 2 presents the variety of false negative error and false positive error with different weights. From the Fig. 2 it can seen that if $\omega^{(2)} = 0$, viz. $\omega^{(1)} = 1$, PCA SVM

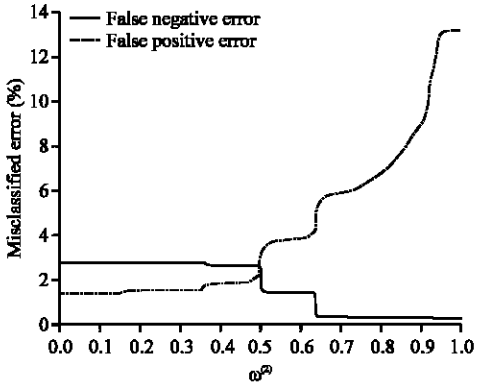

Fig. 2: Effect of weight on system performance

works. The system has a lower false positive error rate (= 1.44%), false negative error rate is 2.71%. With the increasing of $\omega^{(2)}$, false negative error rate of the system is decreased and has decreased dramatically near 0.503 and 0.641. In most of the cases, ($0 \le \omega^{(2)} \le 0.65$), false negative error and false positive error change oppositely, that is, when one of them decreases, another one increases, which shows that it is very difficult to choose one single optimal weight, so we must make a tradeoff between this two criterion. While $\omega^{(2)} > 0.65$, false negative error basically keeps constant, while false positive error ascends. When $\omega^{(2)} = 1$(ICA SVM operates), false positive error reaches its tiptop, 13.14%.

Then, we have investigated the performance of ensemble system and SVM. Table 1 shows the classification performance of the classifier based on standard SVM, PCA feature extraction (PCA SVM), ICA feature extraction (ICA SVM) and the individual classifiers with the highest accuracy and with the lowest false negative error in Pareto-optimal set which is gained by using the niched Pareto multiobjective genetic algorithm (NP-GA).

We can see from Table 1 the accuracy of individual a in Pareto-optimal set is the highest, which is close to PCA SVM and the accuracy of standard SVM is less than it. Individual b in Pareto-optimal set has the false negative error which is the same as in ICA SVM, moreover, its accuracy is much higher than ICA SVM, reaching 94.39%. Although SVM has a high accuracy, its false negative error is the highest, reaching 4.04%, 15 times higher than the individual classifier with the lowest false negative error in Pareto-optimal set.

To express the system's performance more intuitively, we propose a popular measurement, an ROC graph, which plots Accuracy against false negative rate. Compared to other measurements, ROC curves provide a visual tool for examining the tradeoff between the ability

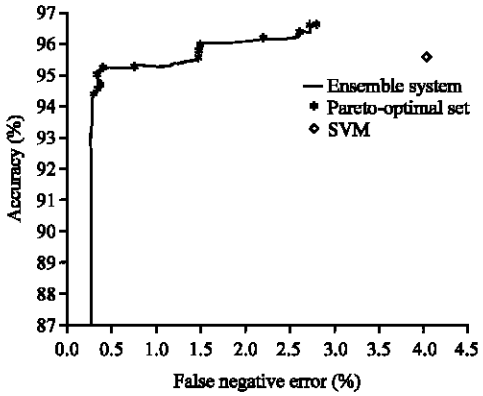

Fig. 3: ROC curves and points

Table 1: Detail results of SVM and ensemble system

Classifier	Accuracy (%)	False negative error (%)	False positive error (%)
SVM	95.56	4.04	1.34
PCA SVM	96.54	2.71	1.44
ICA SVM	87.14	0.25	13.14
Pareto-optimal[a]	96.56	2.71	1.41
Pareto-optimal[b]	94.39	0.27	6.04

[a]The individual classifier with the highest accuracy in Pareto-optimal set, [b]The individual classifier with the lowest false negative error in Pareto-optimal set

of a classifier to correctly identify positive cases and the number of negative cases that are incorrectly classified. Figure 3 presents the ROC curve of the ensemble system, Pareto-optimal set and ROC point of SVM. Point (0, 1) in the figure corresponds to perfect classification.

We can see from the figure that the accuracy of most individuals in Pareto-optimal set selected by NP-GA is higher than or similar to the accuracy of SVM and its false negative error is lower than SVM' s. At the left side of ROC figure, the individuals of Pareto-optimal set have lower accuracy than SVM and the false negative errors are 10 times lower than SVM. Therefore, our approach is more effective for intrusion detection.

The Pareto optimal solutions gained will be submitted to users, who will choose a security strategy based on the current performance of the system. For example, high accuracy strategy can be adopted in peacetime, whereas strategy with a low false negative error is adopted when there is threaten. Therefore, the system can have better protection.

CONCLUSIONS

In this study, after having investigated feature extracting of PCA and ICA, we found that classifier based on PCA feature extraction has a higher accuracy, while it also has a higher false negative error ; on the other side, the classifier based on ICA feature extraction has a very low false negative error and low accuracy. The reason might be that ICA pays more attention on the independence of features while PCA cares the features that can mostly represent the original data.

We combined two feature extracting methods utilizing the idea of ensemble learning, to construct an IDS ensemble system. In the first place, IDS ensemble system adopts support vector machine as classification algorithm, which utilizes quadratic optimization technique to construct an optimal classification hyperplane between two classes and ensures high accuracy of IDS. In the second place, PCA and ICA separately reduce false negative error and false positive error, demonstrating the diversity between individual classifiers.

Because the costs of false positive error and false negative error is asymmetric in IDS, we introduced multi-object optimization when seeking the optimal weight, not to find a single optimal solution but a Pareto-optimal set that is sent to users for making final decision. The experiment indicates that this system is very effective, especially for the lower false negative error compared to SVM, which is very important in an intrusion system.

For the data sets with asymmetric losses, such as intrusion detection, cancer detection, etc, it is a significant problem to find the optimal decision with the lowest cost. For intrusion detection, we utilized two feature detection approaches and multi-objective optimization to solve this problem. In the future work, we intend to bring this technique into cancer detection.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their careful reading of this research and for their helpful and constructive comments. This study was supported by National Natural Science Foundation of China under grant No. 60575045, Yunnan Provincial Natural Science Fund under grant No. 2005F0028Q, the open research fundation from the key lab of wireless sensor networks on Yunnan and Scientific Research Fund Projects of Yunnan Education Department under grant No. 6Y0006D.

REFERENCES

Abbass, H.A., 2003. Pareto neuro-ensembles. Adv. Artificial Intel., 2903: 554-566.

Aickelin, U., J. Greensmith and J. Twycross, 2004. Immune system approaches to intrusion detection. A review. Artificial Immune Systems Proceedings, 3239: 316-329.

Blake and Merz, 1998. UCI repository of machine learning databases. http://www.ics.uci.edu/mlearn/MLRepository.html: Technical Report of Department of Information and Computer Science, University of California, Irvine.

Brown, G., J. Wyatt, R. Harris and X. Yao, 2005. Diversity creation methods: A survey and categorisation. Inform. Fusion, 6 (1): 5-20.

Chen, R.C., J. Chen, T.S. Chen, C. Hsieh, T.Y. Chen and K.Y. Wu, 2005. Building an intrusion detection system based on support vector machine and genetic algorithm. Advances in Neural Networks -ISNN 2005, Pt 3, Proceedings, 3498: 409-414.

Du, Y., H.Q. Wang and Y.G. Pang, 2004. HMMs for anomaly intrusion detection. Computational and Information Science Proceedings, 3314: 692-697.

Gu, Y., J. Zheng, J. Sun and Z. Xu, 2005. Intrusion detection method based on independent component analysis and support vector machine. J. Xi'an Jiaotong Univ., 39 (8): 876-869.

Harmer, P.K., P.D. Williams, G.H. Gunsch and G.B. Lamont, 2002. An artificial immune system architecture for computer security applications. IEEE Trans. Evol. Comput., 6 (3): 252-280.

Iheagwara, C., A. Blyth, T. Kevin and D. Kinn, 2004. Cost effective management frameworks: The impact of IDS deployment technique on threat mitigation. Inform. Software Technol., 46 (10): 651-664.

Joo, D., T. Hong and I. Han, 2003. The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors. Exp. Syst. Appl., 25 (1): 69-75.

Kupinski, M.A. and M.A. Anastasio, 1999. Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves. IEEE Trans. Med. Imag., 18 (8): 675-685.

Mukkamala, S. and A.H. Sung, 2003. Feature selection for intrusion detection with neural networks and support vector machines. Transportation Security Infrastructure Prot., (1822): 33-39.

Pietraszek, T., 2004. Using adaptive alert classification to reduce false positives in intrusion detection. Recent Adv. Intrusion Detection, Proceedings, 3224: 102-124.

Ralaivola, L. and F. D'Alche-Buc, 2001. Incremental support vector machine learning: A local approach. Artificial Neural Networks-Icann. Proceedings, 2130: 322-330.

Scholkopf, B., K.K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik, 1997. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans. Signal Processing, 45 (11): 2758-2765.

Sung, A.H. and S. Mukkamala, 2004. The feature selection and intrusion detection problems. Advances Computer Science-Asian. Proceedings, 3321: 468-482.

Verwoerd, T. and R. Hunt, 2002. Intrusion detection techniques and approaches. Comput. Commun., 25 (15): 1356-1365.

Ye, N., X.Y. Li, Q. Chen, S.M. Emran and M.M. Xu, 2001. Probabilistic techniques for intrusion detection based on computer audit data. IEEE Trans. Syst. Man Cybernetics Part A-Systems Humans, 31 (4): 266-274.

Zhang, C.L., J. Jiang and M. Kamel, 2003. Comparison of BPL and RBF network in intrusion detection system. Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, 2639: 466-470.