Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (22): 5099-5105, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.5099.5105

The Study on Network Intrusion Detection Technology Based on Improved Principal Component Analysis

Yu-Ting Lu
ChangzhouTextileGarmentInstitute, 213164, Changzhou, China

Abstract: Targeting at various defects of the current intrusion detection technologies, such as the high omission ratio, poor self-learning ability and the relatively high error alarming rate, the paper proposed a intrusion detection algorithm based on improved neural network. According to the experimental results, the intrusion detection result makes full use of the advantages of KPCA and ICA and it boasts ideal intrusion detection performance and the fine learning ability.

Key words: Intrusion detection, principal component analysis, independent component analysis, neural network

INTRODUCTION

With the development of openness, sharing and interconnection of information age computer network, especially the emergence of the Internet, the importance of network and its impact on society has also been increasing. While enjoying the various benefits brought about by the network, people also face along with new insecurities generated by new technologies. Therefore, the Internet and network information security has become a hot issue of concern for governments, major industries and enterprises and leaders. The research on intrusion detection system began in the 1980s. The earlier intrusion detection method only detects when the intrusion mode already exists in the invasion pattern and you can not identify the intrusion that does not exist in the database. Manual analysis and writing new rules of invasion should be done for new intrusions. This intrusion detection method is not only with poor ability to detect unknown network attack, narrow detection range and has poor defensive ability to the widely used scripting attacks, slow detection speed, poor real time and expansion potential as well as the high cost.

Most current intrusion detection products are based on a single detection technology to detect the intrusion. However, intrusion or attack behavior in an ever-changing network environment are becoming integrated and complicated. Intruders usually while implementing the invasion or attack, take a variety of invasive means, in order to guarantee the success of the invasion. Multiple intrusion means could conceal the true purpose of attack or invasion in the early stage (Hu *et al*., 2010). On the other hand, the current intrusion detection technology itself has many defects and there are a lot of technical problems still awaiting a breakthrough. Although a large number of commercial products appeared, intrusion detection systems still face considerable problems.

Described in this paper is intrusion detection algorithm based on neural network. First it introduced feature selection and feature extraction algorithm that with Relief algorithm it achieves intrusion feature selection and through PCA, KPCA and ICA, the three algorithms, realizes intrusion feature extraction. In this process, through the tests on KDDCUP99 datasets it compared the performance of PCA and KPCA algorithm and also compared the merits and demerits of PCA, KPCA and ICA three algorithms in terms of intrusion feature extraction (Sheluhin *et al*., 2011). It elaborated the theory of neural networks. It gave detailed led description of the process of this integrated neural network construction and used the advantages and disadvantages of the two methods of KPCA, ICA and combined weighted ensemble classifier intrusion detection of integrated neural network based on genetic algorithm. In addition it used genetic algorithms to achieve the right value adjustments referred to in this article as intrusion detection algorithm based on neural network (Bahrololum *et al*., 2009). Finally, the simulation results proved the detection performance of the algorithm.

BASIC THEORY OF INTRUSION DETECTION

Intrusion detection is through the collection and analysis of key information in the computer network or computer system to check whether there exist actions violating the security strategy and the attacked objects in the network or system. Intrusion detection activities are capable of real-time monitoring system, real-timely discovering aggressive behaviors and taking appropriate measures to avoid the attack or trying to reduce the harm caused by the IDS typically uses two basic analytical

methods to analyze events, detect intrusions, misuse detection and anomaly detection (Hong, 2012). Misuse detection target is to find the known intrusion mode. It is the analytical methods used by the majority of commercial IDS products. Misuse detection analysis method attempts to detect abnormal patterns of behavior of the system which is seldom used in the actual IDS.

Misuse detection: Misuse detection is also known as feature detection. Assume the intruder's activities could be indicated by mode. The task of the system is to detect whether the main activities comply with these patterns. Misuse detection depends on the mode library (feature database). If you did not construct a good pattern library, you will fail to detect intruders. This method can only detect the invasion mode in the model base and it could not detect that not in the mode library. Misuse Detection Model is shown by Fig. 1.

Misuse detection can effectively detect known attacks, produce fewer false alarms but it needs to constantly update attack feature database to be able to detect new attacks. Therefore, the system's flexibility and adaptability are relatively poor (Richard and Tan, 2012). It has lots of troubles and systematic underreporting.

As for realization methods of specific modeling, now there are expert systems, pattern matching, the state transition, neural networks, artificial immunology and other technologies applied to misuse intrusion detection system.

Anomaly detection: The anomaly detection's assumption is that the activities of the intruder are abnormal activities in the body. According to the idea it establishes activity contour for authorized users of normal activity and it compares the current activity status with behavior contour of the authorized users. When it violates the authorized user's behavior contour it is regarded as intrusion. Unfortunately there is intersection of intrusion and contour of authorized the user behavior. Its advantage is that it does not depend on the attack mode, yet able to discover new attacks.

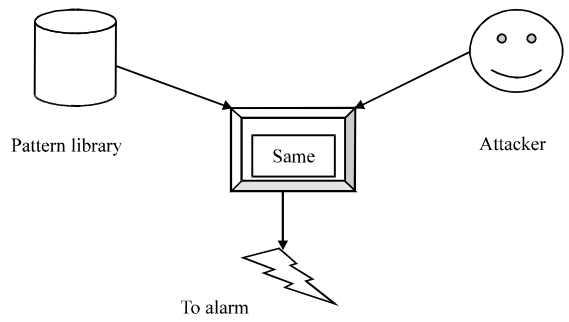

Fig. 1: Misuse detection model

The key of anomaly detection method lies in how to describe and establish the outline of the normal behavior of the system and make the system update and automatically generate new behavior contour according to the actual use. Anomaly detection methods do not need to define for each intrusion, so it could effectively detect the unknown new attack patterns. But due to the sharp shift of normal users and the network's behaviors, the system has high error alarm rate. In order to describe the normal behavior patterns, anomaly detection methods often require a large number of system event log "trainingset." Commonly used methods of anomaly detection: anomaly detection method based on statistical analysis, anomaly detection method based on model predictions, case learning method based on the similarty, method based on neural network.

Therefore, in lot of that intrusion attacks are becoming more comprehensive; attacker indirect; attacked object complex, the paper applies the artificial intelligence technology to the research and development of intrusion detection systems, in order to achieve a project-oriented application with perfect algorithm model, secure and effective intelligent intrusion detection system. Now it has become an important development direction of intrusion detection systems.

Intrusion detection methods have a variety of strategies and methods for the detection of abnormal intrusion are often not fixed. Intelligent computing technologies' application in intrusion detection will greatly improve the efficiency and accuracy of the detection. Intelligent detection methods currently adopted include: Neural network technology, genetic algorithms, immune systems, fuzzy theory, expert systems, machine self-learning and other related technologies.

INTRUSION DETECTION ALGORITHM BASED ON PCA NEURAL NETWORK

In order to obtain ideal intrusion detection, efforts should be made in two ways: First, to build a good classifier; second to look for a good representation of the problem that the selected input features could provide the most useful information for the classifier. For the former, people try to use a variety of methods based on different principles to detect intrusions. For the latter, there are generally two ways to get a better expression of the problem: feature selection and feature extraction. It is necessary to reduce the dimensions while carrying out intrusion detection to detection data. Existing dimension reduction feature mainly adopts feature selection and feature extraction (also feature structure).

Feature selection and feature extraction in the intrusion detection system are shown in Fig. 2-4. Here the

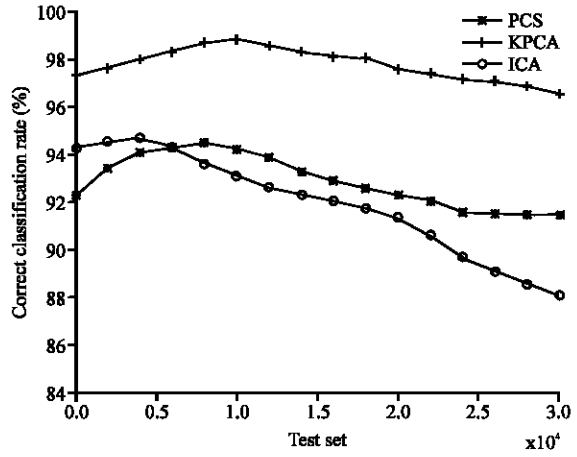

Fig. 2: Comparison of correct classification rates

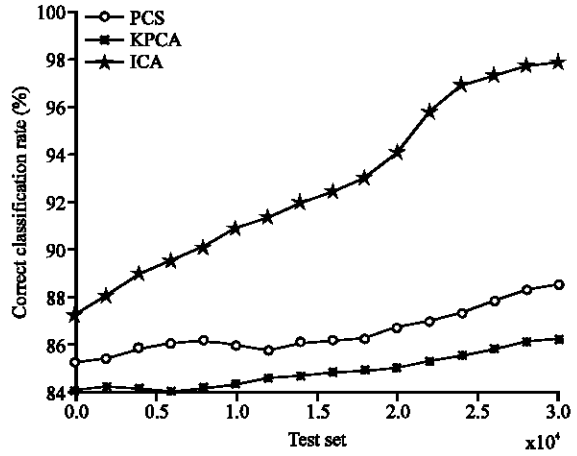

Fig. 3: Comparison of false alarm rates

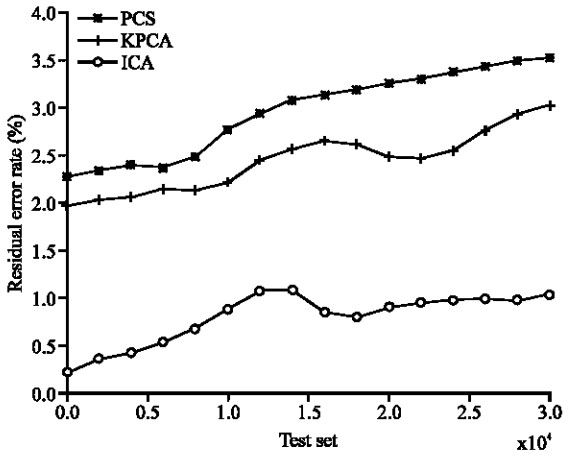

Fig. 4: Comparison of mission rates

irrelevant feature refers to those features that have no relationship with the classification process. Removing irrelevant features, this paper applies the Relief algorithm. To realize feature extraction, this study adopts Kernel Principal Component Analysis (KPCA) method and independent principal component analysis (ICA) method.

Feature selection: The method borrows the idea of nearest neighbor learning algorithm and its theoretical basis is: A good feature should make a similar sample of the nearest neighbor have same or similar feature values, so that the feature value of nearest different samples will be sharply different, thus giving weight to the corresponding features and carrying out feature sorting. The greater weight means that the stronger the classification ability of the feature will be. On the contrary it means the weak classification ability. Through setting the feature subset number and the threshold value, one can carry out corresponding feature selection. Relief algorithm is described as follows:

- **Input:** The feature vector set of training examples and category mark×N-feature dimension, m-itinerate times
- **Output:** Weight vector W corresponding to the various sub-vectors of the feature vector

```
Initialize the weight vector as 0, namely
W[A]: = 0. 0;
    for i: = 1 to m
    Randomly select example Ri;
    Calculate nearest hit H and nearest miss M;
    for A: =1 to N
    W[A]:=W[A]-diff(A,Ri,H) /m
+diff(A,Ri,M) /m;
    end;
```

Parameters need explanation: diff (A, Ri, H) represents the difference of sample R and H on the feature A. If the feature is discrete, then:

$$\text{diff}(A,R_i,H) = \begin{cases} 0 & \text{value}(A,R_i) = \text{value}(A,H) \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

If the feature is continuous, then:

$$\text{diff}(A,R_i,H) = \frac{|\text{value}(A,R) - \text{value}(A,H)|}{\max(A) - \min(A)} \quad (2)$$

where, H and M represent the nearest hit and nearest miss of training set to the real example Ri. Relief applies only to the two types of Kononenko (Wang *et al*., 2011). In view of this situation, extending Relief algorithms will get Relief algorithm. Relief can be applied to the multi-class sample situation. Differently, Relief algorithms in processing multi-class problem, does not select the near sample from different classes but rather select from a sample set of each of the different classes. And it does not select a nearest neighbor samples but k nearest neighbor samples instead.

Relief algorithm description is as follows:

- **Input:** The feature vector set of the training real example and category mark×N-feature dimension, m-itinerate times, k-nearest neighbor samples
- **Output:** Weight vector W corresponding to the various sub-vectors of the feature vector

Initialize the weight vector as 0, namely
W[A]: =0. 0;
for i: =1 to m
Randomly select example Ri
Calculate k nearest hits Hj
As for each category of C C,C?class (Ri)
Obtained from category C, k nearest misses Mj(C)
for A: =1 to N

$$W[A]:= W[A]-\sum_{j=1}^{k} diff(A,R_i,H_j)/(m.k)+$$

$$\sum_{C\#class(R_i)}\left[\frac{P(C)}{1-P(Class(R_i))}\sum_{j=1}^{k} diff(A,R_i,M_j)()C)/(m.k)\right]$$

end;

Relief randomly extracts a sample Ri at a time from the training samples and from the sample set with the same class of Ri to find out k nearest hits of sample Ri. From the samples with different class of Ri, find out k nearest misses. Then update each feature weight value. Repeat the above process by m times. Kononenko, Marko and others based on statistical theories analyzed the Relief algorithm (Zhou *et al*., 2011). They concluded that when sample set is adequately large, Relief evaluation could meet the following approximation:

$$W[\text{feature I}] = P$$

(feature i is valued differently| nearest sample of different class)-P (feature i is valued differently |nearest sample of the same class).

Here W [feature i] indicates the weighted value of feature i. P (feature i is valued differently| nearest sample of different class) indicates the probability that nearest samples of different classes have different i values; P (feature i is valued differently |nearest sample of the same class) indicates that probability that nearest samples of the same class have different i values.

Feature extraction: From the perspective of optimization, feature re-construction methods' advantage lies in that number of features of the new structure is less than the original and the ability to carry maximum useful original feature information. In addition, they remain irrelevant to new features. The feature extraction is a singular way to improve the performance of the classifier. It obtains new low dimension features through transforming the high-dimensional input feature in order to achieve the effect of data dimensionality reduction.

The Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Independent Component Analysis (ICA) are common and effective feature extraction methods. In this section, first it introduced PCA and KPCA algorithm and carried out analysis and comparison of the two methods. And then it introduced the ICA algorithm and compared the three algorithms. It concluded that KPCA and ICA based classifiers are mutually complimentary in terms of accuracy rate and omission ratio.

Principal component analysis: PCA principal component analysis refers to a method that extracts the linear features of data. As a feature extraction technology it has been applied in many model recognition areas. The emphasis of the section will be placed on researching the value of PCA in intrusion detection feature extraction.

Given that a group of the centralized the input vector is:

$$x_t, t=1,...,l\sum_{t=1}^{1} x_t = 0$$

each m-dimensional, $xt = (x_t(1), x_t(2),..., x_t(m))^T$, usually, m<l, each x_t is transformed by PCA linear transformation into a new vector.

Here U is an orthogonal matrix of m×m, u_i of columni is the eigenvector of sample variance matrix:

$$C=\frac{1}{l}\sum_{t=1}^{1} x_t, x_t^T$$

In other words, we should first obtain the following Eigen value with PCA:

$$\lambda_i\mu_i = CU_i, i=1,...,m \quad (3)$$

In the above, λ_i is an eigenvalue of C and u_i is the corresponding eigenvector. Conduct following orthogonal transformation on x_t to get component s_t:

$$S_t(i) = U_i^T x_t, i=1,...m \quad (4)$$

The new component is referred to as a main component. Eigenvector u_i and the corresponding eigenvalue u_i are arranged in descending order. When we only use the first few eigenvectors, the number of principal component λ_i at the same time will reduce, thus realizing the effect of dimension reduction.

Algorithm of nuclear principal component analysis : The proposal of nuclear principal component analysis is considered as the nonlinear extension of PCA. It introduces nuclear method into PCA. It first maps the original input vector into a high dimensional feature space and then calculates with PCA method. The linear PCA method in space $\phi(x_t)$ maps with nonlinear PCA methods in space x_t. As the dimension assumption of space $\phi(x_t)$ is larger than that of training sample l. Therefore, KPCA solves such eigenvalue problem as follows:

$$\lambda_i \mu_i = \tilde{C} u_i, i = 1, ..., l \quad (5)$$

Here:

$$\tilde{C} \frac{1}{l} \sum_{t=1}^{l} \phi(x_t) \phi(x_t)$$

is the sample variance matrix of $\phi(x_t)$ λ_i a nonzero eigenvalue of C; u_i, corresponding eigen vector.

Equation 5 can be transformed into the following problem of eigenvalue:

$$\tilde{\lambda_i} \tilde{\alpha_i} = K\alpha_i, i = 1, ..., l \quad (6)$$

In the above equation, K is the nuclear matrix of l×l. The value of each element of K is equal to the inner product of vector x_i and x_j in high dimensional feature space $\phi(x_t)$ and $\phi(x_j)$, namely:

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \quad (7)$$

In light of the mapping of $x_t \rightarrow \phi(x)$, the nuclear function $K(x_i, x_j)$ can totally replace $\phi(x_{tt}) \bullet \phi(x_j)$ and weneedn't to precisely calculate arbitrary dimension $\phi(x_j)$ which has already been solved by the deployment of nuclear function K. In addition, functions satisfying the conditions of Mercer's can be applied with nuclear method K. Postulate that λ_i is a eigenvalue of K, then $\lambda_i = l \lambda_i$, a_i is the corresponding eigenvector of K, meeting the condition that:

$$u_i = \sum_{j=i}^{l} \alpha_i(j) \phi(x_j) \ (\alpha_i(j), \ j = 1, ..., l \text{is component of } \alpha_i)$$

Besides, in order to ensure that the eigenvector $\phi(x_t)$can satisfy the condition that $u_i \bullet u_{i=1}$, each a_i should be standardized by corresponding eigenvalue:

$$\tilde{\alpha_i} = \frac{\alpha_i}{\sqrt{\tilde{\lambda}}}$$

Hence, based onestimation, we can calculate the principle component x_t with the following equation:

$$S_t(i) = u_i^T \phi(x_t) = \sum_{j=i}^{l} \tilde{\alpha_i}(j) K(x_j, x_t), i = 1, ..., l \quad (8)$$

In the above wquation, the sample input vector $\phi(x_t)$ is made up of some centre data, namely:

$$\sum_{j=i}^{l} \phi(x_t) = 0$$

The nuclear matrix on training set K and test set K_t can be modified with the following equation:

$$\tilde{K} = (I - \frac{1}{l} l_l l_l^T) K (I - \frac{1}{l} l_l l_l^T)) \quad (9)$$

$$\tilde{K} = (K_t - \frac{1}{l} l_{lt} l_l^T K)(I - \frac{1}{l} l_l l_l^T) \quad (10)$$

Here, I is the n-D unit matrix l_t; the quantity of test data; l_t and l_{lt}, the full l vector of length of l and l_t.

From equation 10, we found that, compared to the PCA, KPCA can extract more main components in that the maximum amount of the main component in the KPCA is l instead of m. Similarly, the dimension of S_t reduces in that it only considers relatively greater feature vectors.

Independent component analysis algorithm: The second-order statistics, just as the previously described PCA, can not solve the problem of higher order statistical characteristics. However, the higher-order statistics can better describe the probabilistic statistical characteristics of the signal and can contain the Gaussian noise. The purpose of the independent component analysis is to decompose mixed signal into independent components rather than related components of the PCA. ICA, compared with PCA, can make better use of the statistics between the signals.

As a method of data analysis, ICA was proposed by Jutten and his colleges to solve the problem of separating the blind source (Gogoi *et al*., 2010). Nowadays it is also widely used in such fields as intrusion detection and feature extraction. The general model of ICA is as follows:

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_m(t) \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & a_{mm} \end{bmatrix} \bullet \begin{bmatrix} s_1(t) \\ \vdots \\ s_n(t) \end{bmatrix} \quad (11)$$

If:

$$\begin{cases} x_1(t) = x_1(t), \cdots, x_m(t)]^T \\ A = (a_{ij})_{m \times n} \\ s(t) = [s_1(t), \cdots, s_n(t)]^T \end{cases} \quad (12)$$

Then the above model can be simplify as follows:

$$X(t) = As(t) \quad (13)$$

In the above: x_t is the known observation data, S_t is the n-dimensional vector composed of n unknown independent source signals. $A = (a_{ij})_{m \times n}$ is an unknown m×n dimensional mix matrix, waiting for estimation. If we obtain the mix matrix pass the appropriate algorithm estimated is obtained, the generalized inverse matrix A through appropriate algorithm, we can estimate the independent component with the known observational data vectors x_t after calculating the inverse matrix of A in broad sense.

$$S(t) = Wx(t) \quad (14)$$

SIMULATION EXPERIMENT AND RESULT ANALYSIS

The data set used in the study is the current relatively authorities test data set KDDCUP99 in the field of intrusion detection. In the experiment, 1 stands for DOS attack; 2, Probing attack; 3, U2R attack; 4, R2L attack. In the KDDCUP99 training set there are altogether 2984154 data, among which 1062847 are normal data, 1879191 are data about DOS attacks, 41023 are data about U2R attacks, 1002 new data about U2R attack, 91 are data about R2L attack. 31539 data are selected as training data for this experiment, the detailed distribution of which are as Table 1 follows:

The test data is part of the 10% test data set of KDDCUP99. Among the selected 32768 data, 19646 are normal data, 12668 are about DOS attack, 415 are about Probing attack, 5 are about U2R attack, 34 are about R2L attack. The test data set and the training data are different in probability distribution and the test data set includes some attack types never appeared in the training data set which makes the intrusion detection more practical.

During the study, we first chose 31539 data from KDDCUP99 training set in accordance with the proportion of each data type. Then, conduct dimension reduction with methods like PCA, KPCA and ICA and after that input those data into BP neural network for training. In the end, use the trained network to classify the data of 10% test set of KDDCUP99. BP neural network consists of one hidden layer and four neurons on the hidden layer. The learning rate is 0.05 and the expected error is 0.005.

Table 1: Distribution of various data types in training

Types of attack	Data amount
Normal	14500
DOS	14500
R2L	91
U2R	1002
Probing	1446

From Fig. 2-4 are, respectively comparison of correct classification rates, comparison of false alarm rates and comparison of omission rates. The above definitions are used to describe the performance of the algorithm and system for intrusion detection.

From the comparison of classification accuracy rate in Fig. 2, adopting PCA method, the system's classification accuracy rate could peak 94% with an average standing at 93% through BP neural network training; adopting KPCA method, the number could peak 99% with an average of 98% after BP neural network training; adopting ICA method, the classification accuracy rate is lower than the former two which peaks 94.5% with an average of 93%. Therefore, from the classification accuracy rate, KPCA is superior to PCA and ICA.

From the false alarming rate comparison of Fig. 3, adopting PCA method, the rate is minimized to 5% with an average of 6% after BP neural network training; adopting KPCA method, the rate is minimized to 4% with an average of 5% after BP neural network training; adopting ICA method, the rate is lower than the former two which is minified to 7% with an average of 12%. Therefore, from the false alarming rate, KPCA is superior to PCA and ICA. From the omission rate comparison of figure 4, adopting PCA method, the rate is minimized to 2.4% with an average of 3% after BP neural network training; adopting KPCA method, the rate is minimized to 2% with an average of 2.5% after BP neural network training; adopting ICA method, the rate is lower than the former two which is minified to 0.2% with an average of 0.8%. Therefore, from the omission rate, KPCA is superior to PCA and ICA.

CONCLUSIONS

Adopting improved PCA algorithm to establish intrusion detection system will obtain high classification accuracy rate, low false alarming rate but high omission rate; however, adopting ICA algorithm to construct the detection system will get a very low omission but relatively low accuracy rate. In other words, improved PCA and ICA are complementary in terms of performances on classification accuracy rate and omission rate.

REFERENCE

Bahrololum, M., E. Salahi and M. Khaleghi, 2009. An improved intrusion detection technique based on two strategies using decision tree and neural network. J. Convergence Inform. Technol., 4: 96-101.

Gogoi, P., B. Borah and D.K. Bhattacharyya, 2010. Anomaly detection analysis of intrusion data using supervised and unsupervised approach. J. Convergence Inform. Technol., 5: 95-110.

Hong, L.X., 2012. The research on anti-interference technology for network of optimized support vector machine based on genetic algorithm. Adv. Inform. Sci. Ser. Sci., 4: 119-125.

Hu, L., K. Tang, Y. Ku and K. Zhao, 2010. Improvement on intrusion detection technology based on protocol analysis and pattern matching. J. Convergence Inform. Technol., 5: 86-94.

Richard, M.R. and G.Z. Tan, 2012. Innate-inspired automated intrusion response mechanism for a network intrusion detection system. J. Convergence Inform. Technol., 7: 194-201.

Sheluhin, O.I., A.A. Atayero and A.B. Garmashev, 2011. Detection of teletraffic anomalies using multifractal analysis. Int. J. Adv. Comput. Technol., 3: 174-182.

Wang, P.F., Y. Hu and L. Li, 2011. An efficient automaton based matching algorithm and its application in intrusion detection system. Int. J. Adv. Comput. Technol., 3: 278-285.

Zhou, R., H. Wang, G. Feng, F. Guo, B. Li and F. Gao, 2011. A behavioral model of ephemeral ports and its applications. Adv. Inform. Sci. Serv. Sci., 3: 9-16.