Research Journal of Applied Sciences, Engineering and Technology 5(3): 937-943, 2013
DOI:10.19026/rjaset.5.5043
ISSN: 2040-7459; E-ISSN: 2040-7467

Submitted: June 20, 2012 Accepted: July 23, 2012 Published: January 21, 2013

Research Article

New Intelligent Computer Intrusion Detection Method Using Hessian Local Linear Embedding and Multi-Kernel Support Vector Machine

[1]Fei Hu, [2]Guoxiang Zhong, [3]Qiong Bo and [4]Yang Lei
[1]Network Management Center,
[2]Department of Research,
[3]Library, Chongqing University of Education, Chongqing, 400065 China
[4]Modern Educational Technology Center, Chongqing Nankai Secondary School,
Chongqing, 400030 China

Abstract: Computer networks frequently collapse under the destructive intrusions. It is crucial to detection hidden intrusions to protect the computer networks. However, a computer intrusion often distributes high dimensional characteristic signals, which increases the difficulty of intrusion detection. Literature review indicates that limited work has been done to address the nonlinear dimension reduction problem in computer intrusion detection. Hence, this study has proposed a new intrusion detection method based on the Hessian Local Linear Embedding (HLLE) and multi-kernel Support Vector Machine (SVM). The HLLE was firstly used to reduce the dimension of the original intrusion date in a nonlinear manner. Then the SVM with multiply kernels was employed to detect the intrusions. A real computer network experimental system has been established to evaluate the proposed method. Four typical intrusions have been tested. The test results show high effectiveness of the new detection method. In addition, the new method has been compared with the single-kernel SVM with Local Linear Embedding (LLE) or Principal Component Analysis (PCA). The comparison results demonstrate that the proposed HLLE plus multi-kernel SVM can provide the best computer intrusion detection rate of 97.1%.

Keywords: Computer networks, HLLE, intrusion detection, multi-kernel SVM

INTRODUCTION

Along with the rapid development of internet and the associated application networks, network security has become a prominent and tough problem, in particular, intrusions and attacks on computer network systems becomes more complex and diverse. Huge economic losses have been caused by the computer and network intrusions and attacks every year. Therefore, it is essential to detect the intrusions and attacks in time to prevent damages of computers and networks.

The diversity and the evolution of the intrusion viruses make it very difficult in detecting and identifying the undergoing network intrusion. From the early worm virus to the recent shock, shock waves and the panda incense viruses, the attacking objects almost include all computer system accessed to the internet. The attacking viruses will cost the system resources, manipulate data and steal the confidential information, leading to massive economic losses. American business magazine "Information Weekly" has published a survey on the network intrusion and indicates that a network attack happens every second in the global scope. In such a situation, network security has become an urgent and practical problem and received worldwide attentions. How to develop and use the existing security technology to protect all kinds of resources from damage is the hot spot in the research field of network security. Effective intrusion detection technology is the key issue to solve this problem.

The Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R) and Probe or Scan (PoS) are the most representative computer intrusion types. These four computer intrusion types cover almost all known computer intrusion types. The biggest thing in common of these four intrusion types is that the high dimensions of the attack feature data. Usually, there are more than 41 features need to check when one aims to detect these computer intrusions. Hence, it is very difficult in identifying an intrusion. To overcoming this problem, some advanced data processing methodologies have been proposed. Among the machine learning is a very useful technology in the field of computer and network security. However, the computer intrusion feature data is always contaminated by background noise. In addition, the high dimensionality of the intrusion data increases the difficulties in the intrusion detection (Abadeh *et al*., 2011). Hence, it is crucial to remove the

Corresponding Author: Fei Hu, Network Management Center, Chongqing University of Education, Chongqing, 400065 China

noise and reduce the intrusion features into a low dimensional space. As a result, the intrusion recognition can be enhanced.

Up to date, the Artificial Intelligence (AI) technology has been extensively used in the computer intrusion detection. Representative methods include the Artificial Neural Network (ANN) (Nabil *et al.*, 2012), Fuzzy logic (Abadeh *et al*., 2011) and Support Vector Machine (SVM) (Vapnik, 1995), etc. Elshoush and Osman (2011) gave a comprehensive review on the intrusion detection using AI methods. Nabil *et al*. (2012) adopted the artificial neural networks to detect the network intrusions. Abadeh *et al*. (2011) used the fuzzy inference to detect intrusion in computer networks. Yi *et al*. (2011) employed the SVM for network intrusion detection. Hu *et al*. (2003) proposed an anomaly detection algorithm, based on robust Support Vector Machines (SVM), which can effectively detect the intrusions even if noise existed. The success of these intelligent methods relies on proper and sufficient pre-processing of the original image data. However, the ANN or the Fuzzy detection model often requires a large amount of training samples. It is not realistic in practice because sufficient training samples are usually difficult to be ensured. As a result, if not enough training is provided the detection ability of the ANN or Fuzzy model should decay. In contrast, the SVM does not have this shortcoming and performs well even with a very small amount of samples. Hence, SVM is competent for parametric and/or non-parametric statistics and is found to be powerful in computer network intrusion detection (Mukkamala and Sung, 2003). Nevertheless, SVM is original to be suitable for binary classification tasks, which means SVM is usually used in one-class problem. The computer network intrusions are always a multi-class problem. How to use the SVM to deal with different types of intrusions is very important to ensure the security of computer networks. Since the multi-kernel SVM is proven to be effective in the computer intrusion detection (Li *et al*., 2012a), it is worthy testing its outcomes in the multi-class recognition problem in the computer intrusion detection.

On the other hand, many problems in network security begin with the preprocessing of raw high dimensional signals. This kind of preprocessing aims to attain more useful representations of the high dimensional signals in the form of low dimensional information. As a result, the subsequent operations such as classification, de-noising and visualization, etc., can be handed in an easy way (Hinton and Sejnowski, 1999). The high dimension of the intrusion feature data not only increases the computation cost but the detection rate. So, before the intrusion detection using the SVM, it is crucial to reduce the dimension of the intrusion feature data. One of the most useful methods in the dimension reduction is the Principal Component Analysis (PCA) (Li *et al.*, 2011a). PCA can project the high dimensional features into a low dimensional space and thus eliminate useless ones. However, PCA is a kind of linear reduction approach and hence performs not well for the nonlinear cases (Li *et al*., 2012b). In contrast to PCA, the Hessian Local Linear Embedding (HLLE) (David and Carrie, 2003) has strong ability of dealing with nonlinear dimension reduction (Goldberg *et al.*, 2008). HLLE is a kind of manifold learning algorithm, specifically, it is an extension of the Local Linear Embedding (LLE) algorithm (Roweis and Saul, 2000). The HLLE can reduce a high dimensional feature space into a much lower dimension space by keeping the topological structure of the original feature space. As a result, the output of HLLE preserves the nonlinear properties of the original space. Although HLLE has been found to be very powerful in the application of image processing and medical pattern recognition, etc., little work has been done to address the computer intrusion detection. Hence, the outcomes of the HLLE should be evaluated in the application of computer intrusion detection.

To enhance the performance of computer network intrusion detection, the present work has proposed a new method based on the integration of HLLE and multi-kernel SVM. The HLLE was firstly used to reduce the dimension of the original feature space of the intrusion data in a nonlinear manner. Then the multi-kernel SVM acted as a multi-class classifier to recognize the types of the intrusions. Compared with Li's work (Li *et al*., 2012a), the proposed method in this study has conducted the nonlinear dimension reduction using HLLE. Thus, the presented work has better performance in the computer network intrusion detection. Experimental tests have showed high performance of the proposed method.

THE PROPOSED DETECTION APPROACH

Since there is an inner correlation between intrusion and its future data, which is often impossible to be described via analytical model, the SVM (Vapnik, 1995) is employed to learn their relationships. The concept of the kernel trick allows SVM to be able to find the decision function from small datasets. In this study the multi-kernel SVM is used for the intrusion detection. Moreover, to improve the generalization of the SVM model, the HLLE algorithm is introduced to reduce the dimension of the input feature data. The theories about the multi-kernel SVM and HLEE as well as the proposed intrusion detection processes are briefly described as follows:

The Hessian Local Linear Embedding (HLLE): HLLE (David and Carrie, 2003) extends the continuum theory for LLE to recover efficient parameterization without requiring convexity. Given an intrusion feature space $M = [m_1\ m_2\ ...\ m_i] \in R^p$, the goal of HLLE is to project M into a low dimensional space $N = [n_1\ n_2\ ...\ n_i] \in R^q$ ($q<<p$). It takes six steps for HLLE to achieve the dimension reduction (David and Carrie, 2003):

- Compute k neighbors of each sample m_i, i.e., $M_i = [m_{i1}\ m_{i2}\ ...\ m_{ik}]$
- Compute the tangent coordinates V_i of M_i
- Estimate the Hessian matrix H_i for each sample
- Construct quadratic term H
- Calculate the null space U of H
- Calculate the low dimensional embedding N of M

It can be seen that HLLE has the same framework of LLE. The only difference is that HLLE adopts the Hessian transform to replace the local reconstruction weight matrix in LLE. For more details of HLLE, it can refer to David and Carrie's work (David and Carrie, 2003).

The Support Vector Machine (SVM): According to Vapnik's definition of SVM (Vapnik, 1995), the basic SVM can be described as:
if sample set:

$$S = \{(x_{j,}\ y_j) \mid j = 1, 2, ..., n\} \in R^p \times \{-1, 1\}$$

is linear separable, then exist a hyperplane $\omega^T \mathbf{x} + b = 0$ that makes arbitrary sample $(x_{j,}\ y_j)$ satisfy the following condition:

$$\begin{cases} \omega^T x_j + b \geq +1\ (when\ y_j = +1) \\ \omega^T x_j + b \leq -1\ (when\ y_j = -1) \end{cases} \quad (1)$$

where,
ω : The weight vector
b : A constant

Then the goal of SVM is to find the optimal hyper plane $\omega^T \mathbf{x} + b = 0$ and make (ω, b) subject to the following convex quadratic optimization problem:

$$\begin{cases} \min(\dfrac{\|\omega\|^2}{2}) \\ s.t.\ y_j(< \omega, x_j > + b) \geq 1 \end{cases} \quad (2)$$

when the sample set S is not linear separable, a slack variable ξ is needed such that Eq. (2) can be rewritten as:

$$\begin{cases} \min(\dfrac{\|\omega\|^2}{2} + C\sum_{j=1}^{n} \xi_j) \\ s.t.\ y_j(< \omega, x_j > + b) \geq 1 \end{cases} \quad (3)$$

where, C is a constant.

To solve the convex quadratic optimization problem, the Lagrange multiplier method has been introduced to transform it into a dual problem. Then, the original convex quadratic optimization can be solved by the maximum of the following dual problem:

$$\begin{cases} L(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j < x_i, x_j > \\ s.t. \sum_{j=1}^{n} \alpha_j y_j = 0, 0 \leq \alpha_j \leq C \end{cases} \quad (4)$$

where, α is the Lagrange multiplier. Thus, the optimal discriminant function can be expressed as:

$$d(x) = \sin\{\sum_{j=1}^{n} \alpha^*_j y_j < x_j, x > + b^*\} \quad (5)$$

when the sample set S is nonlinear separable, the kernel function $K(x_j, x)$ (Vapnik, 1995) is used to replace the dot product in Eq. (5). Common kernel functions include polynomial kernel, Radial Basis Function (RBF) kernel and sigmoid kernel.
The polynomial kernel is defined as:

$$K(a, b) = (a^T b + 1)^d, d = 1, 2, ... \quad (6)$$

The RBF kernel is defined as:

$$K(a, b) = \exp(\frac{\|a - b\|^2}{2\sigma^2}) \quad (7)$$

where, σ is a constant.

The sigmoid kernel is defined as:

$$K(a, b) = \tanh(\beta_1 a^T b + \beta_2) \quad (8)$$

where, β is a constant.

In practical classification problem, the choice of the kernel function is very important. Proper kernel relates directly to the classified accuracy. However, the above kernels are efficient for their own scopes of application. To overcome this problem, the multiply

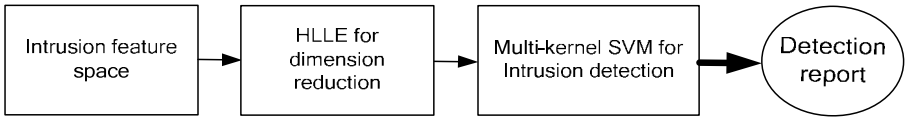

Fig. 1: The diagram of the computer intrusion detection

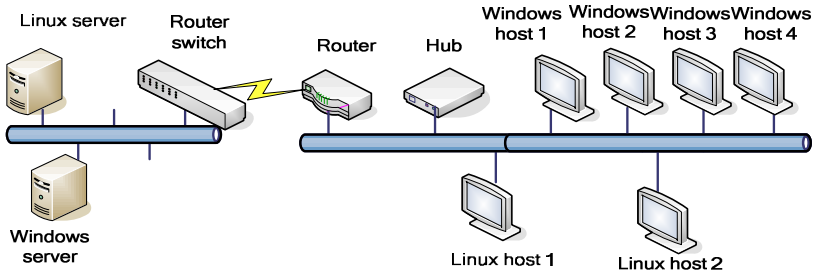

Fig. 2: The principle of the experiment tests

kernels have been used for SVM. Then, the optimal discriminant function can be rewritten as:

$$d(x) = \sin\{\sum_{j=1}^{n}[\alpha_j^1 y_j K_1(x_j, x) + \alpha_j^2 K_2(x_j, x) + \ldots + \alpha_j^k K_k(x_j, x)] + b^*\} \quad (9)$$

Equation (9) adopts k kernels to optimize the discriminate function. Hence, it has more powerful generalization ability than single kernel.

The computer intrusion detection: The proposed processes of the computer intrusion detection are given as follows:

- Preprocess the computer network intrusion feature data into required format.
- Reduce the dimension of the intrusion feature data using HLLE and obtain the new low-dimensional nonlinear intrusion features.
- Train the multi-kernel SVM using the reduced intrusion features.
- Test the performance of the SVM model for the computer intrusion detection.

A diagram of the proposed intrusion detection method is illustrated in Fig. 1.

EXPERIMENT TESTS AND RESULTS

In order to evaluate the performance of the proposed computer intrusion method, experiment tests have been implemented in this study. Figure 2 shows the experiment principle. A mini computer network has been established to conduct the experiments. The computer network is composed of a linux server, a windows server, the web link, two linux host and four windows hosts. Some manual attacks have been simulated and tested using this experiment system.

In this study, the Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R) and Probe or Scan PoS) are introduced into the experiment system to validate the new detection method. Thirty two attribute features are monitored and recorded for every intrusion. These features include the bytes issued from source to destination, the bytes from destination to source, (duration, teardrop, Neptune etc. There are 2,000 samples for each intrusion type and hence the total samples are 8,000.

After the data acquisition, the HLLE is used to reduce the 32 intrusion features into 2 manifold features. In the HLLE processing, the neighbors' k selects 8. The dimension reduction results are shown in Fig. 3. To highlight the efficiency of HLLE, HLLE is compared with LLE and PCA herein and Fig. 4 and 5 show the dimension reduction results of LLE and PCA. It can be seen in Fig. 3 that the intrusion feature space has been grouped into 4 clusters with evident demarcations although there are some overlaps between R2L and U2R. However, in the dimension reduction of LLE in Fig. 4, cluster R2L mixes up with U2R. Thus, LLE only separates three groups from the original feature space. It is also noticeable in Fig. 5 that the inter-class is confused in R2L, U2R and Pos when the PCA algorithm is applied. Hence, only two groups have been clustered from the original feature space by PCA. In addition, it needs emphasize that the inner-class distance and inter-class distance of Fig. 3 are much bigger than that in Fig. 4 and 5. This means that the performance of the dimension reduction of HLLE is superior to the LLE and PCA.

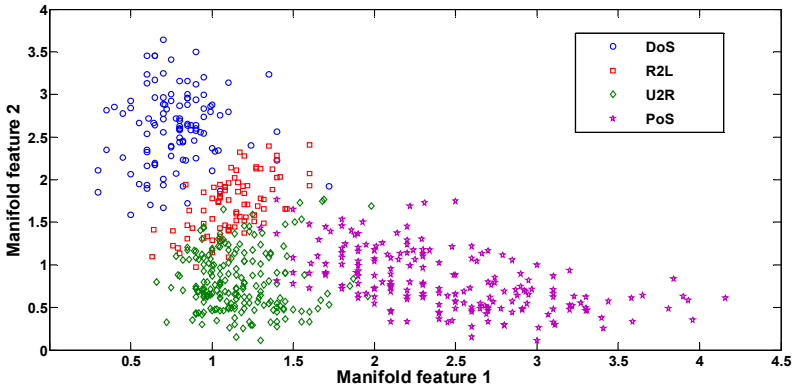

Fig. 3: The dimension reduction via HLLE (embedding dimension q = 2)

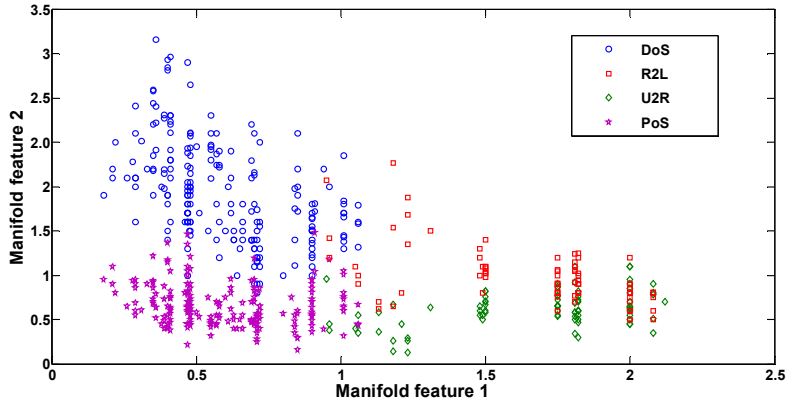

Fig. 4: The dimension reduction via LLE (embedding dimension q = 2)

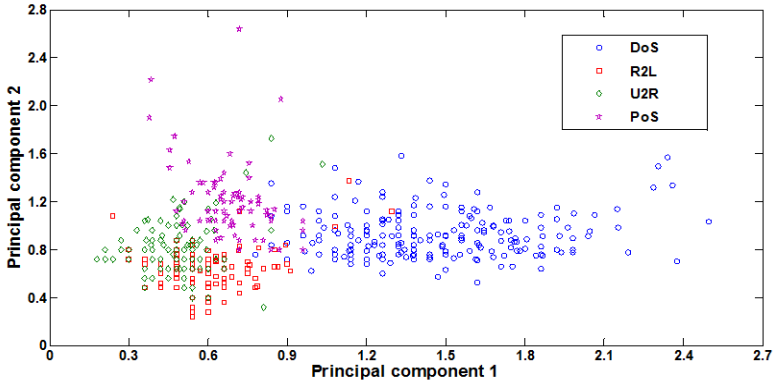

Fig. 5: The dimension reduction via PCA (principal component number is 2)

Table 1: The intrusion detection results using single-kernel SVM

Dimension reduction method	Training (%)			Testing (%)		
	Polynomial kernel	RBF kernel	Sigmoid kernel	Polynomial kernel	RBF kernel	Sigmoid kernel
Non	83.3	84.3	83.6	83.1	83.1	82.9
PCA	90.1	89.5	90.3	89.6	90.2	89.8
LLE	92.6	91.9	92.3	91.4	91.6	91.3
HLLE	94.1	94.3	93.7	93.8	93.5	93.2

Table 2: The intrusion detection results using multi-kernel SVM

Dimension reduction method	Intrusion detection rate (%)					
	All polynomial kernels	Polynomial + RBF kernels	Polynomial + sigmoid kernels	All RBF kernels	RBF + sigmoid kernels	All sigmoid kernel
Non	87.6	87.7	87.7	86.1	85.4	86.9
PCA	93.2	93.5	93.5	92.9	91.8	93.3
LLE	94.7	95.2	94.6	94.1	94.1	94.3
HLLE	96.3	97.1	96.5	95.5	95.7	95.8

After the dimension reduction, a new feature space $F_{2000\times2}$ is obtained, where 2000 denotes the samples and 2 denotes the dimension. Then, the SVM is applied to learn the feature space and recognize the intrusion patterns. In this study, half of the feature space $F_{2000\times2}$ is used to train the SVM and the rest is for testing.

As mentioned above, the multi-kernel SVM is adopted in the intrusion detection. Three kernels have been used for the multi-kernel SVM. Moreover, comparison of multi-kernel SVM and single-kernel SVM has been implemented. Table 1 shows the intrusion detection results using single-kernel SVM while Table 2 shows the intrusion detection results using multi-kernel SVM.

It can be seen in Table 1 that the performance of different kernels in the single-kernel SVM is almost the same. Owing to the dimension reduction processing, the intrusion detection rate has been enhanced significantly. This is because the original intrusion feature space contains lots of redundant information that disturb the SVM detection. However, these redundant features and noise information have been eliminated by the dimension reduction processing and hence the intrusion detection rate has increased. What is more important, the dimension reduction efficiency of HLLE is better than that of LLE while LLE outperforms to PCA. This is because HLLE and LLE can preserve the nonlinear information of the original data and HLLE is an improved version of LLE. As a result, the HLLE plus SVM provide the best intrusion detection results. Proof of this statement can also be found in Table 2. In Table 2, the multi-kernel SVM has been tested. It can be seen in the table that the polynomial plus RBF kernels can provide the best intrusion detection rate of 97.1%. It also can be noticed that by the use of multiply kernels, the intrusion detection rate has been enhanced against to single-kernel SVM.

CONCLUSION

Computer networks make the modern society prosperous. However, they often suffer the intrusions. To protect the computer networks, a new intrusion detection method base on the integration of Hessian Local Linear Embedding (HLLE) and multi-kernel Support Vector Machine (SVM) is proposed in this study. The innovation of this new detection method lies on that it adopts the nonlinear data dimension reduction and flexible kernels to improve the generalization ability of the SVM. A series of experiments have been conducted in a real computer network system. The test results demonstrate that the HLLE can capture distinct features from the original intrusion data and the multi-kernel SVM can detect the intrusions precisely. Future study will evaluate the new method in practice application.

REFERENCES

Abadeh, M., H. Mohamadi and J. Habibi, 2011. Design and analysis of genetic fuzzy systems for intrusion detection in computer networks. Expert Syst. Appl., 38: 7067-7075.

David, L. and G. Carrie, 2003. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. PNAS, 100: 5591-5596.

Elshoush, H. and I. Osman, 2011. Alert correlation in collaborative intelligent intrusion detection systems: A survey. Appl. Soft Comput. J., 11: 4349-4365.

Goldberg, Y., A. Zakai, D. Kushnir and Y. Ritov, 2008. Manifold learning: The price of normalization. J. Mach. Learn. Res., 9: 1909-1939.

Hinton, G. and T. Sejnowski, 1999. Unsupervised Learning and Map Formation: Foundations of Neural Computation. MIT Press, Cambridge, MA.

Hu, W., Y. Liao and V. Vemuri, 2003. Robust support vector machines for anomaly detection in computer security. Proceedings of the 2003 International Conference on Machine Learning and Applications, pp: 23-24.

Li, Z., J. Zhang and S. Hu, 2011a. Incremental support vector machine algorithm based on multi-kernel learning. J. Syst. Eng. Electron., 22: 702-706.

Li, Z., X. Yan, C. Yuan, Z. Peng and L. Li, 2011b. Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. Mech. Syst. Signal Pr., 25: 2589-2607.

Li, Z., X. Yan, Y. Jiang, L. Qin and J. Wu, 2012a. A new data mining approach for gear crack level identification based on manifold learning. Mechanika, 18: 29-34.

Li, Y., W. Li and G. Wu, 2012b. An intrusion detection approach using SVM and multiple kernel method. Int. J. Advancements in Computing Technol., 4: 463-469.

Mukkamala, S. and A. Sung, 2003. Feature selection for intrusion detection with neural networks and support vector machines. Transp. Res. Record, 1822: 33-39.

Nabil, E., K. Hadjar and E. Nahla, 2012. A mobile agents and artificial neural networks for intrusion detection. J. Softw., 7: 156-160.

Roweis, S. and L. Saul, 2000. Nonlinear dimensionality reduction by locally linear embedding. Sci., 290: 2323-2326.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. 1st Edn., Springer-Verlag, Berlin.

Yi, Y., J. Wu and W. Xu, 2011. Incremental SVM based on reserved set for network intrusion detection. Expert Syst. Appl., 38: 7698-7707.