Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Quantum Particle swarm optimization based network Intrusion feature selection and Detection

Hong-mei Zhang *. Hai-Hua Gao **,Xing-Yu Wang***

** School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China (Tel:086-021-64251019; e-mail:hmzhang@ guet.edu.cn).*
*** School of Information and Communication engineering, Guilin University of Electronic Technology, Guilin 541004，China*

Abstract: Considering the relevance among features, which filter-based feature selection method fails to deal with, a kind of hybrid quantum particle swarm optimization and support vector machines based network intrusion feature selection wrapper algorithm is put forward. The subset of features is represented using quantum superposition characteristic and probability representation, among which superposition characteristic can make a single particle represent several states, thus potentially increases population diversity. Every particle in the quantum particle swarm stands for a selected subset of features. A probabilistic mutation is adopted to avoid local optimal and a taboo search table is used to enlarge particle swarm's search space and avoid repeated computation. The fitness of particle is defined as the correct classification percentage by SVM using a training set whose patterns are represented using only the selected subset of features. The results of experiments demonstrate that the proposed method can be an effective and efficient way for feature selection and detection via using the data sets of KDD cup 99.

1. INTRODUCTION

With the exponential outspread and development of computer network technology, network intrusion presents a comprehensive and integrated evolution trend, which makes the currently used rule based misuse detection system inefficient, thus cause high false alarm rate and undetected rate. So it is critical to develop novel methods for anomaly detection. In recent year, machine learning based anomaly detection has become the popular research issue, which views intrusion detection modelling as a data classification problem. Measurements of system activity such as TCP/IP connections, system logs, resource usage, command traces and audit trails are used to produce a classification of the state of the system as normal or abnormal, and great achievements have been reported.

In order to evaluate the performance of applying varied machine learning based methods to intrusion detection, the 1998 DARPA Intrusion Detection Evaluation project was conducted by MIT Lincoln Labs, which set up a LAN network and logged normal and attack network traffic. These records were reduced and processed by domain experts to yield KDD Cup 99 dataset for competition [1]. Based on KDD Cup 99, many researches have been conducted, which falls into two key parts: detection model generation and intrusion feature analysis.

For the former, numerous machine learning methods such as artificial immune theory (X.R. Yang et al. 2002), Bayesian parameter estimation (S. Cho et al., 2004), clustering (S.H. Oh et al., 2003), data fusion (Y. Wang et al., 2004) and neural networks (A.H. Sung et al., 2003) etc has been adopted to build good detection model. Support Vector Machines (SVM) (V.N. Vapnik,1995) is a newly proposed machine learning approach, owing to its remarkable characteristics such as good generalization performance, the absence of local minimal and the sparse representation of solution, SVM has become a popular research method in intrusion detection, and good results are reported (A.H. Sung et al.,2003; H. Li et al., 2003).

For the latter, generally, there are two ways to obtain better represented features from the original samples. One is feature selection; the other is feature extraction. Feature selection approach is based on the idea that, though all available indicators can be used as the inputs of classifier, irrelevant or redundant features may have a negative effect on its accuracy and deteriorate its generalization performance. In addition, reducing the number of features may help decrease the cost of acquiring data and might make the classification models easier to understand. However, the literature indicates that many machine learning techniques applied in IDSs are often used for classification, and very little scientific efforts aimed at modelling efficient IDS feature selection. Sung (A.H. Sung et al., 2003) have demonstrated that a large number of features are unimportant and may be eliminated, without significantly lowering the performance of the IDS. Chebrolu (S. Chebrolu et al., 2004) selected important features based on the Markov blanket model and train the Bayesian network and CART (Classification and Regression Trees) with the reduced data set. Their experiment results on the KDD Cup 99 show that the most important features are heavily overlapped, and using the most important features can yield results comparable to those using all available features. Results also demonstrate that feature selection can reduce training and detection time. However, most of those methods assume that all features are independent and without any mutual influence.

This paper presents a novel hybrid Quantum Binary Particle Swarm Optimization (QBPSO) and SVM based model for

intrusion feature selection and detection. Particle swarm optimization is a newly emerged evolutionary technique based on swarm intelligence theory and has been successfully applied to function optimization, neural network training, etc. BPSO is the binary version of PSO. The subset of features is represented using quantum superposition characteristic and probability representation, among which superposition characteristic can make a single particle represent several states, thus potentially increases population diversity. And probability representation makes particle's state is represented according to a certain probability. Every particle in QBPSO stands for a selected subset of features. A probabilistic mutation is adopted to avoid local optimal and a tabu search table is used to enlarge quantum particle swarm's search space and avoid repeated computation. The fitness of particle is defined as the correct classification percentage by SVM using a training set whose patterns are represented using only the selected subset of features. Thus through quantum particle swarm optimization, the network intrusion feature selection and classification is achieved.

The rest of this paper is organized as follows. In section 2, the detail of QBPSO algorithm for feature selection is presented and a Modified version of QBPSO is proposed. In section 3, we evaluated the proposed algorithm with KDD 99 dataset and compare it with feature ranking method of Fisher Discrimination Rating. In the last section, we summarized the experiment result and reach the conclusions.

2. QBPSO for Network Intrusion Feature Selection

2.1 Particle swarm optimization

Particle swarm optimization is an evolutionary computation techniques inspired on the choreography of a bird flock (J. Kennedy et al., 1995). Each particle stands for a possible optimal solution and has two feature parameters: n-dimension position vector and n-dimension velocity vector, for the i-*th* particle, they are denoted by $x_i = (x_{i1}, x_{i2}, \cdots x_{in})$ and $v_i = (v_{i1}, v_{i2}, \cdots v_{in})$ respectively. The fitness of each particle is its value of object function for current position, which determines the performance of the possible solution. PSO is initialized with a group of random particles at the beginning of the algorithm. In each iteration round, the fitness of each particle is first to be calculated, then each particle's own best previous position (pbest) $p_i = (p_{i1}, p_{i2}, \cdots p_{in})$ and best previous position of the whole or neighbour swarm (gbest) $g = (g_1, g_2, \cdots g_n)$ are updated. Then particle's current velocity and position are recalculated by using the following equations:

$$v_i = v_i + c_1 \times k_1 \times (p_i - x_i) + c_2 \times k_2 \times (g - x_i) \quad (1)$$

$$x_i = x_i + v_i \quad (2)$$

where c1 and c2 are two positive constant named as learning factors, k1 and k2 are random numbers in the range of (0, 1), v_i is the current velocity of the i-*th* particle. Such an adjustment of the particle's movement through the space changes each particle toward its pbest and the current gbest value. If the minimum error criterion is satisfied or iteration rounds reach a predefined threshold, the algorithm is terminated.

The PSO technique described above is the real valued PSO, whereby each dimension can take on any real valued number and is difficult to solve combinatory optimization problem. Therefore James Kennedy proposed a discrete binary particle swarm optimization. Each dimension of the particle can only take on the discrete values of 0 or 1. The velocity of each particle represents the probability of bit taking the value of 1 (J. Kennedy et al., 1997). In next section, we presented a kind of hybrid quantum theory and binary particle swarm optimization algorithm for feature selection.

2.2 Representation of QBPSO particle

In quantum binary particle swarm optimization, each bit of particle is represented by quabit, which has two basic states $|0>$ and $|1>$. Differed from classical bit representation, at any particular time, the state of quabit can be represented as the linear combination of basic states, which is called overlapped state as is shown in the following equation:

$$|\varphi>= \alpha |0> + \beta |1> \quad (3)$$

Among which, α and β are complex value, they are called probability breath which means the probability of $|0>$ is $|\alpha|^2$ and that of $|1>$ is $|\beta|^2$, they satisfied the following relations:

$$\alpha^2 + \beta^2 = 1 \quad (4)$$

In this paper, α and β are limited in real number region thus the Euclid representation of quabit is shown in the following equation:

$$|\varphi>= \cos\theta |0> + \sin\theta |1> \quad (5)$$

θ is the phase of quabit and its relationship with probability breath is shown as:

$$\theta = \arctan\frac{\beta}{\alpha} \quad (6)$$

Suppose the particle's dimension in QBPSO is denoted as m (which is equal to the dimension of training sample), the quantum representation of particle is the sum of probability breath or phase which are shown as the follows respectively:

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_m \end{bmatrix} \quad (7)$$

$$\text{Or } [\theta_1 \mid \theta_2 \mid \theta_3 \mid \cdots \mid \theta_m] \quad (8)$$

For instance, suppose the state of particle is like equation (9) composed of 3 dimensions:

$$\left[\begin{array}{c|c|c} \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{array}\right] \tag{9}$$

Thus, the state of this particle is:

$$\begin{aligned} &(\frac{1}{\sqrt{2}}|0>+\frac{1}{\sqrt{2}}|1>)\otimes(\frac{\sqrt{2}}{\sqrt{3}}|0>+\frac{1}{\sqrt{3}}|1>)\otimes(\frac{1}{\sqrt{2}}|0>+\frac{1}{\sqrt{2}}|1>) \\ &=(\frac{1}{\sqrt{3}}|00>+\frac{1}{\sqrt{6}}|01>+\frac{1}{\sqrt{3}}|10>+\frac{1}{\sqrt{6}}|11>)\otimes(\frac{1}{\sqrt{2}}|0>+\frac{1}{\sqrt{2}}|1>) \\ &=\frac{1}{\sqrt{6}}|000>+\frac{1}{\sqrt{6}}|001>+\frac{1}{2\sqrt{3}}|010>+\frac{1}{2\sqrt{3}}|011> \\ &\quad+\frac{1}{\sqrt{6}}|100>+\frac{1}{\sqrt{6}}|101>+\frac{1}{2\sqrt{3}}|110>+\frac{1}{2\sqrt{3}}|111> \end{aligned} \tag{10}$$

From above equation, the state of particle is the linear combination of eight basic bit, which means it represents the information of eight general particle. This representation method increases the variety of particle swarm thus expands the searching space of algorithm.

Based on the above description, suppose in the process of QBPSO based intrusion feature selection, the dimension of features is denoted as D, the scale of particle swarm is N, for the convenience of representation, the j-*th* dimension of the i-*th* particle is represented as:

$$\begin{aligned} &\left[\begin{array}{c} particle(1,i,j) \\ particle(2,i,j) \end{array}\right] = \left[\begin{array}{l} \alpha_{ij},\ i=1,\cdots,N;\ \ j=1,\cdots,D \\ \beta_{ij},\ i=1,\cdots,N;\ \ j=1,\cdots,D \end{array}\right] \\ &s.t.\ \ particle(1,i,j)^2+particle(2,i,j)^2=1\ \ \forall i,j \end{aligned} \tag{11}$$

In initial, the value of α and β can be set randomly in the limitation of equation (4), also it can be set equally to the unique value of $1/\sqrt{2}$.

2.3 Particle State Observation of QBPSO

During the observation process, the overlapped state of particle is collapsed as one of the basic state, which in the discrete QBPSO is limited to 0 or 1, then the observe value is utilized for the particle's fitness evaluation. For the j-*th* dimension of the i-*th* particle, suppose the selection probability of the dimension to be collapsed to 1, then the final value is obtained as the follows:

$$ParticleValue(i,j)=\begin{cases} 1 & if\ \ particle(2,i,j)>=p_{ij}\ \ [commonly\ to\ be\ set\ as\ rand(1)] \\ 0 & else \end{cases} \tag{12}$$

2.4 Particle Velocity Update Rule of QBPSO

In the algorithm of QBPSO, suppose the displacement of particle is the phase θ and its variety is the flying velocity $v=\Delta\theta$. For each dimension of particle, suppose $pbest\theta_i$ denotes the history optimal phase of the i-*th* particle, which also means the individual history optimal value: $pbest\theta_j=(pbest\theta_{j1},pbest\theta_{j2},\cdots,pbest\theta_{jD})$, $gbest\theta$ denotes the global history optimal phase for all particles: $gbest\theta=(gbest\theta_1,gbest\theta_2,\cdots,gbest\theta_D)$, then the update equation is described as follow:

$$\begin{aligned} &v_{jd}^{t+1}=\Delta\theta_{jd}^{t+1} \\ &=\omega\times\theta_{jd}^{t}+c_1\cdot rand()\cdot\left(pbest\theta_{jd}-\theta_{jd}\right)+c_2\cdot rand()\cdot(gbest_d-\theta_{jd}) \\ &=\omega\times\theta_{jd}^{t}+c_1\cdot rand()\cdot[\arctan(\frac{pbest\beta_{jd}}{pbest\alpha_{jd}})-\arctan(\frac{\beta_{jd}}{\alpha_{jd}})] \\ &+c_2\cdot rand()\cdot[\arctan(\frac{gbest\beta_d}{gbest\alpha_d})-\arctan(\frac{\beta_{jd}}{\alpha_{jd}})] \\ &j=1,2\cdots,N \qquad\qquad d=1,2\cdots,D \end{aligned} \tag{13}$$

among which, ω is the inertia weight, which can be set a constant, also can in the initial set as a bigger value, and gradually decreased using the following equation:

$$\omega=\omega_{\max}-\frac{\omega_{\max}-\omega_{\min}}{iteration_{\max}}\times iteration \tag{14}$$

c_1 and c_2 is the learning parameter, which commonly to be set as $c_1=c_2=2$. The velocity of each particle has the maxim value $v_{\max}$, if $v>v_{\max}$ then, $v=v_{\max}$. θ_{jd}^{t} is the phase value of the j-*th* particle in the t-*th* round, which is initialized as a rand number.

After the obtainment of phase, the quantum rotation gate is calculated to update the velocity, which is calculated as follows:

$$\left[\begin{array}{c} \alpha_{jd}^{t+1} \\ \beta_{jd}^{t+1} \end{array}\right]=\left[\begin{array}{cc} \cos\Delta\theta_{jd}^{t+1} & -\sin\Delta\theta_{jd}^{t+1} \\ \sin\Delta\theta_{jd}^{t+1} & \cos\Delta\theta_{jd}^{t+1} \end{array}\right]\left[\begin{array}{c} \alpha_{jd}^{t} \\ \beta_{jd}^{t} \end{array}\right] \tag{15}$$

Among which, $\Delta\theta_{jd}^{t+1}$ is the phase variety of the d-*th* dimension of the j-*th* particle in the (t+1)-*th* round, α_{jd}^{t+1} and β_{jd}^{t+1} is the probability representation of the d-*th* dimension of the j-*th* particle. From equation (15), we can get:

$$\begin{aligned} &(\alpha_{jd}^{t+1})^2+(\beta_{jd}^{t+1})^2 \\ &=(\cos\Delta\theta_{jd}^{t+1}\alpha_{jd}^{t}-\sin\Delta\theta_{jd}^{t+1}\beta_{jd}^{t})^2+(\sin\Delta\theta_{jd}^{t+1}\alpha_{jd}^{t}+\cos\Delta\theta_{jd}^{t+1}\beta_{jd}^{t})^2 \\ &=(\cos\Delta\theta_{jd}^{t+1}\alpha_{jd}^{t})^2-2\sin\Delta\theta_{jd}^{t+1}\cos\Delta\theta_{jd}^{t+1}\beta_{jd}^{t}\alpha_{jd}^{t}+(\sin\Delta\theta_{jd}^{t+1}\beta_{jd}^{t})^2 \\ &+(\sin\Delta\theta_{jd}^{t+1}\alpha_{jd}^{t})^2+2\sin\Delta\theta_{jd}^{t+1}\cos\Delta\theta_{jd}^{t+1}\alpha_{jd}^{t}\beta_{jd}^{t}+(\cos\Delta\theta_{jd}^{t+1}\beta_{jd}^{t})^2 \\ &=[(\sin\Delta\theta_{jd}^{t+1})^2+(\cos\Delta\theta_{jd}^{t+1})^2][(\alpha_{jd}^{t})^2+(\beta_{jd}^{t})^2]=1 \end{aligned}$$

The above mentioned basic velocity update rule that is directly derived from basic PSO algorithm has many parameters to be tuned. Furthermore, for this update rule, it is hard to control the particle's phase variety scope, which may cause the ruleless change. Besides this, this update rule may

cause the problem that the variety of velocity make the phase update scale very slim, thus let the search progress to stagnate. Here we proposed a new particle velocity update rule, which presented as follows:

FOR i = 1 to M (the particle number of QBPSO)
  Initially predefine static probability $a(0 < a < 1)$
  Decrease the value of Mu_thres according to the iteration rounds increase
  IF random（）<Mu_thres
    LOOP:
      generate P[i] randomly
      until P[i] is NOT IN taboo_table
    END LOOP
    UPDATE taboo_table=taboo_table+ P[i]
  ELSE
    FOR j=1 to N (N= dimension of the i-th particle)
      vij＝random（[0, 1]）
      (vij =velocity of the j-*th* dimension of the i-*th* particle)

$$\Delta\theta_{ij} = 0 \quad if \quad v_{ij} < a \tag{16}$$

$$\Delta\theta_{ij} = pbest\theta_{ij} - \theta_{ij} \quad if \quad a \le v_{ij} \le \frac{1}{2}(a+1) \tag{17}$$

$$\Delta\theta_{ij} = gbest\theta_{ij} - \theta_{ij} \quad if \quad \frac{1}{2}(a+1) < v_{ij} < 1 \tag{18}$$

    END FOR
  END IF
END FOR

In the above proposed update rule, equation (16) makes some dimension of particle has the opportunity to unchanged, thus not immediately trend to its history optimal value or global value, so as to expand the particle's local search ability, equation (17) (18) let particle conduct global search according itself and global history optimal information, thus guarantee algorithm can converge. Among which, the static probability is adopted to balance the particle's local search and global search. In initial, *a* is to be set a bigger value, which makes particle has the bigger chance to conduct local search, then gradually decrease its value to make algorithm converged to global optimal. In the update process, a whole mutation probability threshold is defined to let some particles in QBPSO generate mutation to escape local optimal.

3. Experiments setup and results

In this section, we adopted the MIT KDD Cup 99 dataset to validate the effectiveness of MQBPSO based network intrusion feature selection. The KDD 99 dataset contain 744 MB data with 4,940,000 records. The data set has 41 attributes for each connection record plus one class label. Although there were four categories of attack in the original input, we simplified the situation by merging four types into one category "Attack", and tried to differentiate them from those "Normal" connections. Upon this simplification, we transformed the original multi-class classification problem to a two-class problem, which helps us focusing on the effect evaluation of feature selection. We choose 23,000 unique records randomly from the KDD 99 dataset for our experiment, and 3,000 records are used to train the model and 20,000 records are used to test. In the process of sampling, we kick out the redundant samples for the purpose of more accurately evaluating our proposed method. In our sampling, the "Normal" samples and the "Attack" samples are of the same proportion in both the training and testing dataset. In the process of selecting attack samples, we reduced the portion of the easily detected "Probe" attack and included more hard-to-detect attack such as "U2R" and "R2L", which makes our dataset more challenging for classification. The anomaly detection simulation was carried out within a Matlab7.1 environment, which was running on a server powered by a Intel ™ 2.4G CPU, 1G RAM，Window XP.

Firstly, we adopt RBF SVM to KDD 99 dataset classification. The two SVM's parameters are both optimized using computation cost grid search, and using 5-CV's average accuracy as the optimal condition. The search scope of parameters grid is: kernel width $\ln\sigma^2 = \{-5:1:10\}$ and penal factor $\ln C = \{-5:1:10\}$. Then we got the optimal kernel parameters, which are C=1 and σ =0.3679. The parameter optimization result is show in Fig.1.

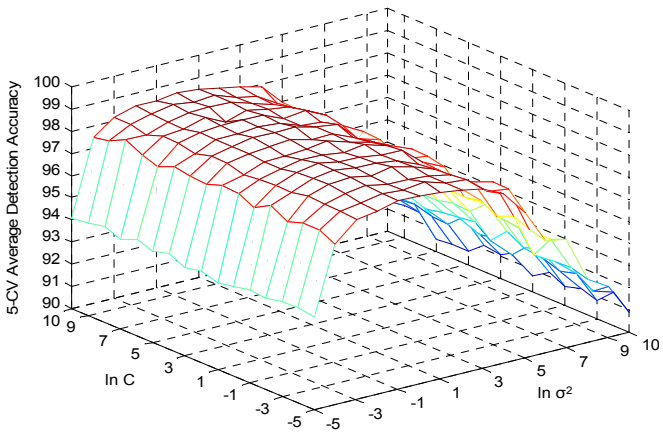

Fig. 1 Average 5-CV Validation Accuracy Mesh of kernel parameter Grid for Sampled KDD 99

Based on the above optimal kernel parameters, the MQBPSO based intrusion feature selection experiment is conducted. The scale of particle swarm is set to 20, and its dimension is 41, which is equal to the number of intrusion feature, the maxim iteration round is set to 100, the mutation probability for the first 70 round is set to 0.25, and 0.075 for the latter 30 round. The performance of MQBPSO_SVM intrusion detection is shown in Table 1.

To evaluate our proposed feature selection method, we compare it with filter based feature ranking using Pearson Correlation Coefficient (PCC) and Fisher Discrimination Rating (FDR) which are defined as follows.

PCC adopts the following feature ranking criterion:

$$CR_j = \frac{(\mathbf{x}_j - \mu_j)^T(\mathbf{y} - \mu_y)}{|\mathbf{x}_j||\mathbf{y}|}, j = 1,2,...N_{feat}$$

Where CR_j is the ranking of the j-*th* feature, $\mathbf{x}_j$ is the corresponding vector of feature j. $\mathbf{y}$ is the target label of

sample, μ_j and μ_y is the expectation value of category 1 and category 2 in feature j, N_{feat} is the dimension of feature space.

FDR adopts the following feature ranking criterion:

$$FDR_j = \frac{(\mu_{j,1} - \mu_{j,2})^2}{\sigma_{j,1}^2 + \sigma_{j,2}^2}, j = 1,2,...,N_{feat}$$

Where FDR_j is the ranking of the j-*th* feature, $\mu_{j,1}$ and $\mu_{j,2}$ is the average of category 1 and category 2 in feature j, $\sigma_{j,1}$ and $\sigma_{j,2}$ is the standard derivation of category 1 and category 2 in feature j, N_{feat} is the dimension of feature space.

Firstly, based on training samples, the ranking for each feature is calculated. The feature importance ranking based on FDR criterion is listed as: [33 29 4 34 12 23 38 39 25 26 3 32 27 30 40 41 28 35 31 10 22 37 36 8 11 24 14 19 13 7 6 9 16 1 5 17 18 2 15 21 20], and the feature importance ranking based on PCC criterion (using absolute value) is listed as: [33 29 4 34 12 23 38 39 25 26 3 32 27 30 40 41 28 35 31 10 22 37 36 8 11 24 19 14 13 7 6 9 16 1 5 17 18 2 15 21 20]. The result demonstrates that except feature 14 and 19, the feature importance ranking is same. Then SVM classification is conducted according to the sorted ranking with elimination one less important feature at each iteration round. In order to consider the performance synthetically, we define a synthetic index to represent the training accuracy (denoted as *TrainAccu*), test accuracy (denoted as *TestAccu*), undetected rate (denoted as *OmitDec*), false alarm rate (denoted as *FalseAlarm*) and the number of support vectors (denoted as *nSVs*). The defined synthetic index *compF* is shown as follows:

$$\min\ compF = w_1 \times (1 - TrainAccu) + w_2 \times (1 - TestAccu) + w_3 \times OmitDec + w_4 \times FalseAlarm + w_5 \times nSVs / TotalNo$$

Here we adopt $w_1 = w_2 = w_3 = w_4 = 1$ and $w_5 = 0.01$, then the *compF* value vary trend of SVM by elimination one less important feature each round according to the sorted ranking are shown in Fig.2 and Fig.3.

The Fig.2 and Fig.3 show that the variation trend of synthetic performance indexes based on FDR and PCC is very similar. Based on the minimal *compF* value, the optimal feature subset [33 29 4 34 12 23 38 39 25 26 3 32 27 30 40 41 28 35 31 10 22] is obtained. The figures also demonstrate that the obtained performance index variation trend is inconsistent with the feature's importance ranking, which means that there exists correlation relationship among network intrusion features.

Subsequently, a FDR-SVM based feature selection using the above calculation result (the optimal feature subset corresponds to minimal *compF*) is performed. The comparison summarized in Table 1.

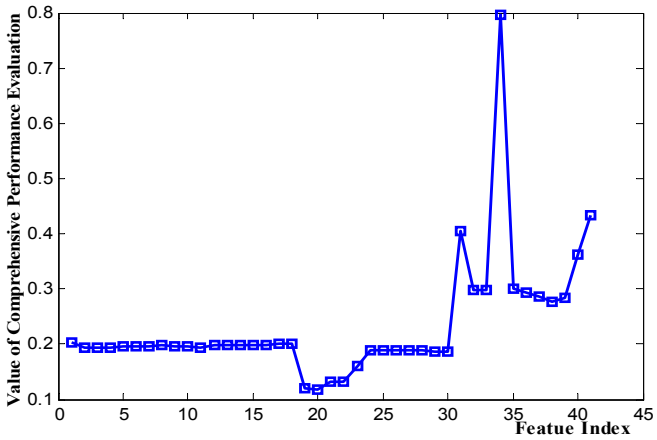

Fig.2 compF Value for Feature Iteration Selection Based on FDR Value for Sampled KDD 99 Dataset

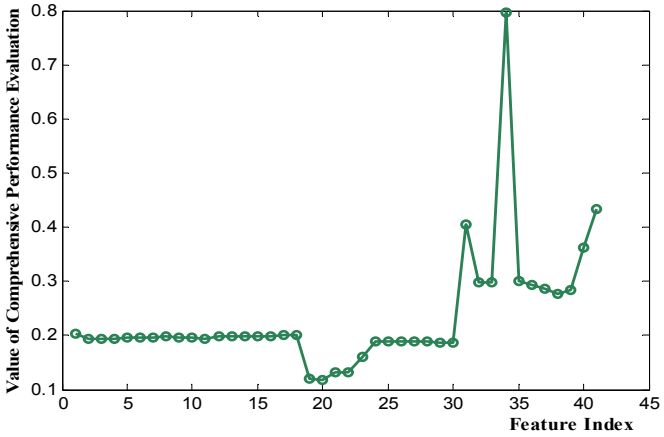

Fig.3 compF Value for Feature Iteration Selection Based on PCC Value for Sampled KDD 99 Dataset

Table 1 Performance Comparison between MQBPSO-SVM, C-SVM and FDR_SVM

Algorithm	Feature No.	No. of SVs	Training Accuracy (%)	Detection Accuracy (%)	False Alarm Rate(%)	Undetect Rate (%)	Training Time (sec)
MQBPSO _SVM	20	142	**99.99**	**97.77**	0.19	**4.2**	8340.00
C_SVM	41	163	99.98	93.55	0.24	12.7	**5.39**
FDR_SVM	21	163	99.98	93.55	0.24	12.7	248.97

Table 1 shows that MQBPSO based feature selection method can find a satisfactory feature subset that the corresponding SVM training accuracy and detection accuracy is superior to C-SVM without feature selection and filter-based feature selection using FDR and PCC criterion. However, it is time consuming, therefore can be applied on the phase of network intrusion feature design, to analyze the features that designed according to the domain expert, thus decrease the feature capture cost of network sensing. Since there is no theory that guarantees that MQBPSO based feature selection can obtain optimal feature subset each time, but can obtain a satisfactory subset with a certain probability. So in practical application, we can adopt classical feature selection algorithm as a reference, and in the permission of time cost limitation,

increase the iteration round and the number of particle, thus to obtain better solution.

4. CONCLUSIONS

This paper presents a novel quantum particle swarm optimization and support vector machines based network intrusion feature selection wrapper algorithm, considering the relevance among features, which filter-based feature selection method fails to deal with. The quantum superposition characteristic can make a single particle represent several states, thus potentially increases population diversity. The probability representation makes particle mutate according a certain probability to avoid local optimal. A taboo search table is used to enlarge particle swarm's search space and avoid repeated computation. Through comparison experiment with the classical intrusion feature selection, we find that there exist correlation relationship among network intrusion features, MQBPSO based wrapper feature selection is superior to those classical intrusion feature selection methods. To sum up, the proposed method is an effective and efficient way for feature selection and detection via using the data sets of KDD cup 99.

Acknowledgement: This research was supported by the Specialized Research Fund for the Doctoral Program of Higher Education (granted No.20040251010), the National "973" Key Basic Research and Development Program (No.2002CB312200), Shanghai Leading Academic Discipline Project (project No.B504), and the Guangxi Province youth Science Foundation (granted No.0728091).

REFERENCES

A.H. Sung, S.Mukkamala. (2003) Identify important features for intrusion detection using support vector machines and neural networks. *Proceedings of the 2003 Symposium on Application and the Internet*, pp. 209-216.

H.Li, X.H. Guan, X. Zan, et al. (2003) Network intrusion detection based on support vector machine. *Journal of Computer Research and Development*, **Vol.40, No.6,** pp.799-807.

J. Kennedy, R.C. Eberhart. (1995) Particle swarm optimization. *Proc. IEEE International Conference on neural networks*. Perth, WA, Australia. **Vol.23,** pp.1942-1948.

J. Kennedy, R. Eberhart. (1997) A discrete binary version of the particle swarm optimization algorithm. *Proc. of the l997 Conf. on Systems, Man, and Cybernetics* (SMC'97). **Vol.1**, pp.4104-4109.

R.P. Lippmann, J.W. Haines, D.J. Fried, J. Korba, et al. (2000) The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks*, **Vol.34, No.4,** pp.579-595

S.Cho, S. Cha. (2004) SAD: web session anomaly detection based on parameter estimation. *Computers & Security*, **Vol.23, No.4**, pp.265-351.

S.H. Oh and W.S. Lee. (2003) An anomaly intrusion detection method by clustering normal user behavior. *Computers & Security*, **Vol.22, No.7,** pp.596-612,.

S.Chebrolu, A.Abraham, et al. (2004) Feature deduction and ensemble design of intrusion detection systems. *Computer &Security*, **Vol.24, No.4**, pp. 295-307

V.N. Vapnik.(1995) *The Nature of Statistical Learning Theory*, Springer, New York.

X.R. Yang, J.Y. Shen, R.Wang. (2002) Artificial immune theory based network intrusion detection system and the algorithms design. *Proceedings of 2002 International Conference on Machine Learning and Cybernetics*, Beijing, pp. 73-77.

Y. Wang, H.H. Yang, X.Y. Wang, et al. (2004) Distributed Intrusion Detection System Based on Data Fusion Method, *The 5th World Congress on Intelligent Control and Automation*, Hangzhou, pp.4331-4334.