Acta Technica *62* No. 2A/2017, 245–254

Mobile Internet anomaly traffic detection technology research based on improved wavelet neural network

Qingshan Li[1, 2]

Abstract. Mobile Internet anomaly traffic detection has a very important significance for ensuring the effective operation of the network and raising the robustness of service providing ability. Quantum particle swarm optimization algorithm is combined with the wavelet neural network, and the parameters of the neural network are optimized by quantum particle swarm optimization. We build a wavelet neural network model based on quantum particle swarm algorithm. Because quantum particle swarm optimization algorithm is easy to fall into local optimum, and it reduces the diversity of population and the global search ability at the same time, we put forward improved quantum particle swarm optimization algorithm based on adaptive local search scheme. Then the improved scheme is used to optimize parameters of wavelet neural network. The experiment results show that the proposed scheme has higher detection rate of abnormal state and lower misjudgment rate of normal state than tradition quantum particle swarm optimization algorithm.

Key words. Anomaly traffic, wavelet neural network, detection rate.

1. Introduction

Wireless communication, mobile broadband and embedded technology progress, makes the calculation function of smart phones and other mobile devices enhanced. Smart phones become the unity equipment combining telecommunication network and Internet communication, and gradually become information center of personal. Along with the rapid growth in the number of users, its malicious software is more and more, which is not only harmful to mobile device security, but also does bad to user privacy property security. How to protect the safety of mobile devices and network is concerned by academia and industry.

In the smartphone security protection technology, operating system security reinforcement and digital signature technology is not enough to prevent malicious

[1]School of EECS, Peking University, Beijing, 100871, China
[2]MOE Key Lab of Network and Software Security Assurance, Peking University,Beijing, 100871, China

software. Most of the mainstream smartphone security software based on feature matching can detect public malicious software, but it cannot detect the unknown malware, and due to the particularity of mobile network, the characteristic library of the software is difficult to guarantee the timely update. For mobile network security protection, the traditional intrusion detection system is difficult to cope with the new mobile network DoS attack. The anomaly detection technology can detect unknown malicious programs, as well as new DoS attack, which is one of the current research focuses in the field of mobile Internet security protection.

For more complex data sets, different anomaly detection techniques have different difficulties. The anomaly detection technology based on clustering and the adjacent method has low detection performance for high dimensional data set. When the dimension is high, the distance between the normal instance and abnormal instance is difficult to draw. These two kinds of technology need to use suitable distance measurement standard to determine the exception. Anomaly detection technology based on spectrum can reduce the dimensions, but performance depends on the hypothesis that normal and abnormal instance has difference after they are projected onto a low dimension. The anomaly detection technology based on classification needs to train normal and abnormal instance. Anomaly detection based on statistical method depends on the assumption of distribution function. Anomaly detection based on information theory needs sensitivity standard to determine the abnormal event. Statistical and signal-based network traffic recognition for anomaly detection was investigated by Michal [8]. Network anomaly detection by cascading k-Means clustering and C4.5 decision tree algorithm was proposed by Muniyandi [9]. Discriminating DDOS attack from flash crowds by means of flow correlation coefficient was put forward by Yu Shui [10]. A design of history based traffic filtering with probabilistic packet marking Against Dos attacks was put forward by Tadashi Kiuchi [11]. Discriminating DDoS attack traffic from flash crowd through packet arrival patterns was presented by Therasak Thapngam [12]. A kind of real time DDoS detection method using fuzzy estimators was proposed by S. N. Shiaeles. Distributed collaborative DDoS detection method based on traffic classification features was investigated by Z. Xiong. D. Stevanovic investigated detection method of malicious and non-malicious website visitors by means of unsupervised neural network learning. Statistical and signal-based network traffic recognition for anomaly detection was proposed by M. Choras. Here, we investigate mobile Internet traffic anomaly detection based on wavelet neural network aiming at the high-dimensional nonlinear behavior of network traffic on small-time scale and propose improved scheme in view of drawbacks of wavelet neural network. It is organized as follows. In the next section, a kind of mobile Internet anomaly traffic detection method based on improved wavelet neural network is put forward. In section 3, in order to test the performance of mobile Internet anomaly traffic detection method, experiments are done. In the end, some conclusions are given.

2. Mobile Internet anomaly traffic detection based on improved wavelet neural network

Wavelet neural network (WNN) and quantum particle swarm optimization are used in traffic anomaly detection of mobile Internet. Wavelet neural network is trained by quantum particle swarm optimization. The parameters combination of wavelet neural network is taken as a particle in the quantum particle swarm optimization algorithm. The parameter vector with the optimal fitness value is searched, then wavelet neural network trained by quantum particle swarm is used in anomaly traffic detection. The wavelet neural network structure is shown in Fig. 1

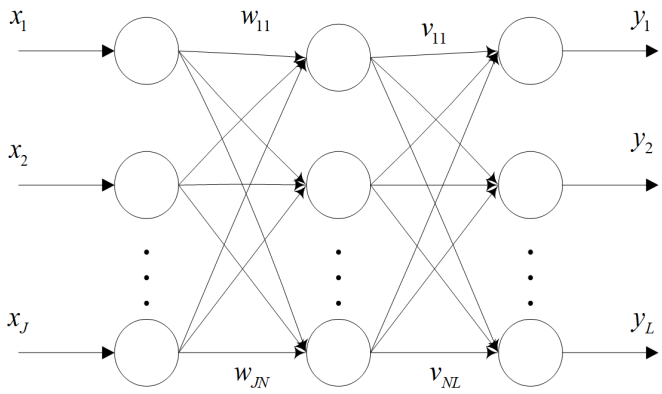

Fig. 1. Adopted wavelet neural network structure

The structure of continuous wavelet neural network is similar to BP neural network, which includes the input layer, hidden layer and output layer. Neuron activation function of hidden layer is wavelet function $\psi(x)$. Training process of wavelet neural network is based on the idea of error back propagation, weight and wavelet parameters are adjusted according to gradient descent direction.

Function $\psi(x) \in L^2(R)$ is the activation function of hidden layer node. There holds

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\psi(w)|^2}{|w|} dw < \infty \,. \tag{1}$$

The output result of node i in the output layer is

$$y_i(t) = \sigma(x_n) = \sigma\left(\sum_{j=1}^{N} v_{ji}\psi_{a_j b_j}(\sum_{k=1}^{J} w_{kj}x_k(t))\right) \,, \quad i = 1, 2, \ldots, L \,, \tag{2}$$

where $\sigma(x_n)$ is the Sigmoid function

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \,. \tag{3}$$

Wavelet neural network uses gradient descent method to adjust the network connection weights, expansion coefficient and shift coefficient based on the error function.

$$E = \frac{1}{2}\sum_{i=1}^{L}\left(y_i(t) - d_i(t)\right)^2 . \tag{4}$$

Here, t represents the current time, x_k represents the input vector of input node k, y_i represents output vector of output node i, w_{kj} is the weight of input node k and hidden layer node j. Quantity v_{ji} represents the weight of hidden layer node and output layer node i, a_j and b_j are expansion coefficient and shift coefficient of the hidden layer node j, E is the error function, and d_i represents the expected output of output node i.

Further

$$\text{net}_j = \sum_{k=1}^{J} w_{kj}x_k(t)\,, \tag{5}$$

$$\psi_{a_jb_j}(\text{net}_j) = \psi\left(\frac{\text{net}_j - b_j}{a_j}\right), \tag{6}$$

and

$$y_i(t) = f\left(\sum_{j=1}^{N} v_{ji}\psi_{a_jb_j}(\text{net}_j)\right). \tag{7}$$

The gradient descent method is used to optimize the wavelet neural network parameters, which is easy to make the algorithm trapped into local optimum and cause oscillation effect. PSO algorithm and QPSO algorithm can overcome the above shortcomings. But quantum particle swarm optimization algorithm and standard PSO is easy to fall into local optimum. Therefore, wavelet neural network parameters are optimized by improved quantum particle swarm optimization algorithm. In the Dth dimension of space, there are m particles. The position of the ith particle is $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$, $i = 1, 2, \ldots, m$. The local optimum of the ith particle is $p_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$ and the optimal position of the whole particle swarm is $p_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$, $g \in \{1, 2, \ldots, M\}$. The evolution equation of quantum particle swarm optimization is

$$\text{mbest}(t) = \frac{1}{m}\sum_{i=1}^{m} p_i(t) = \left[\frac{1}{m}\sum_{i=1}^{m} p_{i1}(t), \frac{1}{m}\sum_{i=1}^{m} p_{i2}(t), \ldots, \frac{1}{m}\sum_{i=1}^{m} p_{iD}(t)\right], \tag{8}$$

$$p = (r_1p_{id} + r_2p_{gd})/(r_1 + r_2)\,, \tag{9}$$

$$X_{id}(t+1) = p_{id}(t) \pm \beta\,|\text{mbest}(t) - X_{id}(t)|\ln\frac{1}{u}\,. \tag{10}$$

In formulas (8)–(10), $i = 1, 2, \ldots, m$, $d = 1, 2, \ldots, D$, u, r_1 and r_2 are random number belonging to [0,1]. Symbol t represents the current iteration times, D represents the dimension of particle, M represents the swarm scale, $p_i(t)$ represents the

current best position of the particle, which means the ith particle in the tth iteration. Quantity $p_g(t)$ represents the global optimal position and mbest(t) represents the average best position in the swarm, which means average value of particle in the tth iteration. Symbol $p_{id}(t)$ represents the random point between $p_i(t)$ and $p_g(t)$. Quantity β represents the shrinkage and expansion coefficient. When quantum particle swarm optimization algorithm is applied to the practical problem, there are many kinds of control methods for the parameter β. One simple way is to set β as a fixed value.

Another effective method is based on linear decreasing. We put forward a kind of adaptive mechanism based on error function

$$z = \frac{f_i - f_{\text{gbest}}}{\min(\text{abs}(f_i), \text{abs}(f_{\text{gbest}}))}\,, \tag{11}$$

where f_i represents the best fitness value of the ith particle, $f_i = f(\text{pbest}(i))$. Symbol f_{gbest} represents fitness value of gbest. The error function is used to represent proximity degree between the particle and the global optimal position gbest. For a certain particle, if value of the error function is the smaller, it means that particle is closer to the global optimal point. As a result, the search area of the particle is narrowed. For the point that is away from the global optimal value, the value of β should be smaller, and, the value of β should be bigger for the point near to the global optimal value. This is because it is almost impossible to search the point far away from their current position, so the value of β should be larger. Otherwise, it is impossible to search a new optimal position. On the contrary, for particles away from the global optimal value, the value of the parameter β is set to be small to ensure the convergence of population. In this way, we construct one adaptive function and can work out value of β according to error function of fixed particle.

In order to improve the algorithm global search and local search ability, and prevent algorithm trapped in local optimal solution, we put forward the improvement strategy, namely adaptive local search strategy. This strategy adjusts the size of local search neighborhood area adaptively according to the search state of swarm algorithm. Changing the size of the neighborhood area is implemented by neighborhood function. Each component of current solution is added by a random variable

$$p_{id}^{'} = p_{id} + \varepsilon\,, \tag{12}$$

where $p_{id}^{'}$ is the solution in the neighborhood area of p_{id}, and ε is a random real number belonging to $(-\delta, \delta)$. Symbol δ is determined by iteration times and fitness value of current optimal solution as

$$\delta = \frac{a}{\text{iteration}} \left| \overline{f(p)} - f(p_g) \right|\,, \tag{13}$$

where $f(p_g)$ is the fitness value of current optimal solution p_g, $\overline{f(p)}$ is the average fitness value of all particles in current swarm. iteration is current iteration times and a is a given real number. The process of improved quantum particle swarm optimization algorithm is as follows:

Step 1. Initialize the population size, the number of iteration, the number of swarm, dimension, individual best position and position of current global optimal solution.

Step 2. Update all particles according to evolution equation.

Step 3. Initialize the position of p_i and p_g with the smallest fitness value.

Step 4. Adaptive local search algorithm is used to search particle swarm is carried out and fitness value of each particle is worked out.

Step 5. Update p_i and p_g.

Step 6. If it meets the constraint condition, the optimal solution is outputted. Otherwise, it turns to step 2 to search until the condition is satisfied.

Then the improved quantum particle swarm optimization algorithm is used to optimize parameters of wavelet neural network. Firstly, we determine the maximum iteration times $T_{\max}$, population size M, particle search space dimension D, shrink and expansion coefficient a_j, shift coefficient b_j, node connection weight value w_{kj} and v_{ji}. We define a vector $x_i = \{a_i, b_i, w_i, v_i\}$. Then x_i is trained and optimized as one particle of improved quantum particle swarm optimization algorithm. The particle with optimal value is mapped to wavelet neural network parameters. The training sample is input to train wavelet neural network and square error of each network on the training set is worked out:

$$E = \frac{1}{2U} \sum_{s=1}^{U} \sum_{p=1}^{L} |c_s^p - f_s^p|^2 \ , \tag{14}$$

where c_s^p and f_s^p represent ideal and actual outputs of training sample s on the output node p. Symbol U represents the total number of sample and L represents the number of output node. Thirdly, we determine whether the algorithm meets termination condition. If it does not meet the condition, new particle individuals are generated according to quantum particle swarm optimization algorithm and these new individuals are optimized again. If it meets termination condition, the parameters with the optimal fitness value is taken as the final results.

3. Experiment and analysis

Ns-2 is used as simulation environment, and the data set is KDD CUP 99 data set, which contains 5000000 connection records, including four types of attacks, namely DOS, U2R and R2L and PROBE. Ten percent of the data is taken as wavelet neural network training samples. The training sample contains 22 kind of different attack way with 494021 connection records, of which only 97278 connection records is normal, and each connection record has 41 different attributes. Network node receives data from the mobile Internet, then the first data is standardized. After standardization, these network data is input into wavelet neural network for testing. KDD CUP 99 data set has 41 different properties in each connection record, so the structure of wavelet neural network is 41-60-5. Five output layer nodes respectively correspond to the four types of attacks and a normal type. Particle swarm

optimization, quantum particle swarm optimization and improved quantum particle swarm optimization is used to optimize parameters of wavelet neural network. Through the simulation experiment results, we compare advantages and disadvantages of three kinds of method. The population size is $M = 100$, the maximum iteration times is $T_{\max} = 1000$, the values of w_{kj} and v_{ji} belong to (-1,1). The value ranges of a_j and b_j belong to (1,100). In the quantum particle swarm optimization, $\beta = 0.5 \cdot (T_{\max} - T)/T_{\max} + 0.5$. In the particle swarm optimization algorithm, inertia weight is determined by $w = (w_{\text{init}} - w_{\text{end}})(T_{\max} - t)/T_{\max} + w_{\text{end}}$ and $T_{\max} = 1000$ means the maximum iteration times. Symbol t represents the current iteration times, w_{init} represents the initial inertia weight, w_{end} represents termination inertia weight and the value of w is from 0.4 to 0.9, $c_1 = c_2 = 2$. The average detection rate and misjudgment rate is shown in Table 1. Detection rate of four kinds of network anomaly is shown in Table 2. In the simulation experiment, the network parameters of wavelet neural network are very different respectively after optimized by above three kinds of training algorithm. From Table 1, we can see that detection rate and misjudgment rate of improved quantum particle swarm optimization is better than the other two kinds of algorithm. Conhat improved quantum particle swarm optimization has faster convergence speed and it has better convergence effect under the same number of iteration times. In mobile internet anomaly detection, the improved quantum particle swarm optimization has several advantages, which not only improves detection rate of network traffic abnormal state, but also reduces the misjudgment rate of normal state.

Table 1. Average detection rate and misjudgment rate

algorithm	average detection rate	average misjudgment rate
PSO	90.87	8.79
QPSO	93.17	5.67
Improved QPSO	95.01	4.65

Table 2. Detection rate of four kinds of network anomaly

algorithm	DOS	PROBE	U2R	R2L
PSO	88.77	88.57	94.27	91.87
QPSO	91.43	91.20	96.48	93.57
Improved QPSO	92.63	93.17	97.51	95.09

4. Conclusion

Network anomaly detection, which establishes the normal network traffic behavior model to detect the abnormal behavior of the network, is an important means of intrusion detection. In recent years, with the continued growth of the number of mobile Internet users and the rapid deployment of new network application, threat

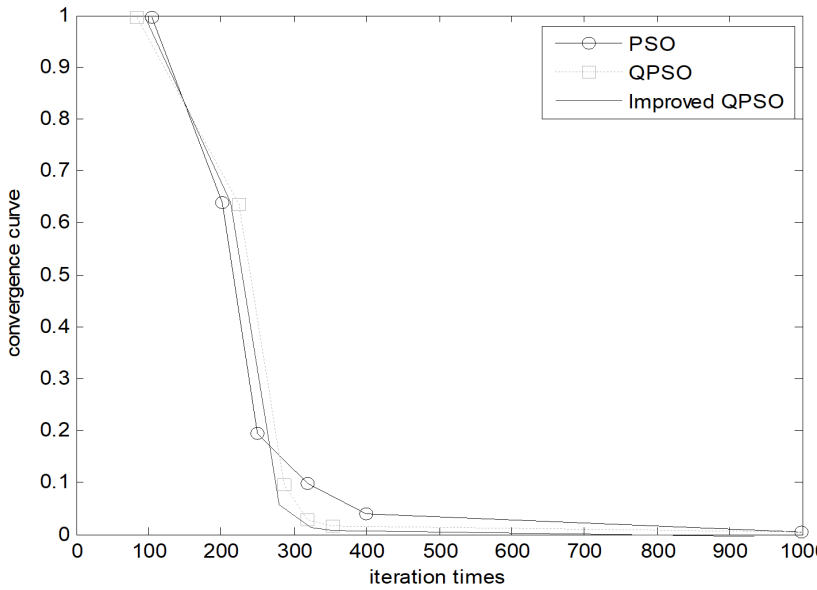

Fig. 2. Convergence curve of three kinds of algorithm

of attack against the network traffic has become increasingly serious. Aiming at the high-dimensional nonlinear behavior of network traffic on small-time scale, a novel wavelet neural network model optimized by improved quantum wavelet neural network is presented. The experiment results show that the proposed scheme is effective. In the condition that mobile node has limited computing power and battery capacity, data dimension reduction methods should be investigated. We should retain the network characteristics as much as possible, minimize the data dimension, decrease the complexity of the algorithm and reduce battery consumption of mobile node.

References

[1] X. Ma, Y. Chen: *DDoS detection method based on chaos analysis of network traffic entropy*. IEEE Communications Letters *18* (2014), No. 1, 114–117.

[2] J. Yang, D. Woolbright: *Correlating TCP/IP packet contexts to detect stepping-stone intrusion*. Computers & Security *30* (2011), Nos. 6–7, 538–546.

[3] X. Wu, Y. Chen: *Validation of chaos hypothesis in NADA and improved DDoS detection algorithm*. IEEE Communications Letters *17* (2013), No. 12, 2396–2399.

[4] Y. Gao, C. Chen, J. Bu, W. Dong, D. He: *ICAD: Indirect correlation based anomaly detection in dynamic WSNs*. IEEE Wireless Communications and Networking Conference, 28–31 March 2011, Cancun, Quintana Roo, Mexico, IEEE Conference Publications (2011), 647–652.

[5] G. Thatte, U. Mitra, J. Heidemann: *Parametric methods for anomaly detection in aggregate traffic*. IEEE/ACM Transactions on Networking *19* (2011), No. 2, 512–525.

[6] J. Yu, H. Kang, D. H. Park: *An in-depth analysis on traffic flooding attacks detection and system using data mining techniques.* Journal of Systems Architecture *59* (2013), No. 10, Part B, 1005–1012.

[7] Y. Ye, T. Li, Q. Jiang, Y. Wang: *CIMDS: Adapting postprocessing techniques of associative classification for malware detection.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) *40* (2010), No. 3, 298–307.

[8] M. Choraś, Ł. Saganowski, R. Renk, W. Holubowicz: *Statistical and signal-based network traffic recognition for anomaly detection.* Expert Systems *29* (2012), No. 3, 232–245.

[9] A. P. Muniyandi, R. Rajeswari, R. Rajaram: *Network anomaly detection by cascading K-Means clustering and C4.5 decision tree algorithm.* Procedia Engineering *30* (2012), 174–182.

[10] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, F. Tang: *Discriminating DDoS attacks from flash crowds using flow correlation coefficient.* IEEE Transactions on Parallel and Distributed Systems *23*, (2012), No. 6, 1073–1080.

[11] T. Kiuchi, Y. Hori, K. Sakurai: *A design of history based traffic filtering with probabilistic packet marking against DoS attacks.* IEEE/IPSJ International Symposium on Applications and the Internet, 19–23 July 2010, Seoul, South Korea, IEEE Conference Publications, (2010), 261–264.

[12] T. Thapngam, S. Yu, W. Zhou, G. Beliakov: *Discriminating DDoS attack traffic from flash crowd through packet arrival patterns.* IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 10–15 April 2011, Shanghai, China, IEEE Conference Publications, (2011), 952–957.

Received July 12, 2017