Acta Technica *61*, No. 4A/2016, 39–48

Network anomaly traffic monitoring based on improved immune genetic algorithm

Binwu Ji[1], Rui Huang[1]

Abstract. A novel network anomaly traffic monitoring method is proposed. The network anomaly traffic monitoring system consists of three steps, that is, 1) Data preprocessing, 2) SVM training, and 3) Anomaly detection. In step 1, a feature set is constructed from training dataset. In step 2, parameters of SVM are optimized by an improved Immune genetic algorithm. In step 3, network anomaly traffic is detected by classifying testing samples to "Normal" and "Attacks". Particularly, the main innovation of this paper is to integrate the immune concept to genetic algorithm together to optimize parameters of SVM classifier. The improved immune genetic algorithm is able to restrain the degenerative phenomena arising in the process of evolution, and then makes the fitness of population increase slowly. Finally, experiments are conducted based on 1998 DARPA dataset and KDD-CUP'99 dataset. Experimental results prove that the proposed method can detect network anomaly traffic more accurately after optimizing SVM parameters by the improved immune genetic algorithm.

Key words. Network anomaly traffic monitoring, Immune genetic algorithm, crossover operation, mutation operation, memory set.

1. Introduction

In recent years, with the rapid development of Internet technology, the traditional network cannot meet the growing requirements of Internet users [1]. How to effectively manage Internet has been attracted more and more attention, and network traffic monitoring has been a crucial issue in computer network [2], [3]. The data of network traffic provides significant information for computer network, and is also very crucial to distribute network resources and analyze of service quality and computer network security [4].

Due the diversification of IP business and the complexity of the network environment, it is difficult to detect abnormal in a small amount from massive network data flow [5]. The abnormality in network traffic denotes the irregular and transpar-

[1]Guilin University of Aerospace Technology, Guilin, 541004, China

ent changes in it. There are two major types of network abnormality, such as local incident and overall incident [6], [7]. Local incident includes temporary congestion, attack of refusing serving, on the other hand, overall incident contains network route abnormality [8].

Network traffic is able to represent the activities of the network and user behaviors, therefore, network traffic analysis can help us to understand and manage the network [9], [10]. Particularly, the anomaly detection and the application type identification of the network traffic are two types of the basic and important issues. In this paper, we aim to propose an efficient algorithm to detect and discover traffic anomalies. The major innovations of this paper are to introduce immune genetic algorithm in network anomaly traffic monitoring. Immune genetic algorithm has been widely used in many fields, such as weapon system portfolio optimization [11], articulated industrial robotic manipulator [12], touch panel cover glass design [13], fault diagnosis of rolling-element bearings [14], dynamic clustering approach [15], the bi-level linear programming problem [16], Emulating human society education and experiential inheritance mechanism [17].

The rest of the paper is organized as follows. We express the framework of the network anomaly traffic monitoring system in section. In Section 3, we illustrate how to optimize SVM parameters by an improved immune genetic algorithm. Section 4 designs and implements a series of experiments to show the effectiveness of the proposed algorithm. Section 5 concludes the whole paper.

2. Overview of the network anomaly traffic monitoring system

The framework of network anomaly traffic monitoring system is shown in Fig. 1, which is made up of three steps: 1) Data preprocessing, 2) SVM training, and 3) Anomaly detection. In the first step, feature set is constructed from training dataset, which is history data of Network anomaly traffics. In the second step, SVM classifier is trained after the data normalization process, afterwards, parameters of SVM are optimized by an improved Immune genetic algorithm. In the third step, anomaly detection is done by classifying testing samples to two classes (Normal and Attacks). Particularly, in this work, the attacks are considered as anomalies.

It can be observed from Fig. 1 that the key component of the proposed system is SVM classifier, and then the network anomaly traffic monitoring problem is converted to a classification problem.

Suppose that the training samples are represented as $x_i, y_i,\ i = 1, 2, \cdots, l$, conditions $x_i \in R^n$, $y_i \in \langle 1, -1\rangle$ are satisfied, $y_i(x) = wx + b$ is the discriminant function, and n is the dimension of training sample space. In particular, SVM aims to solve the following optimization problem:

$$\min \|w\|^2 /2\,, \tag{1}$$

$$\text{s.t.}\ \ y_i[wx_i + b] - 1 \geq 0\,,\quad i \in \langle 1,\, 2,\, 3,\, \cdots,\, l\rangle\ . \tag{2}$$

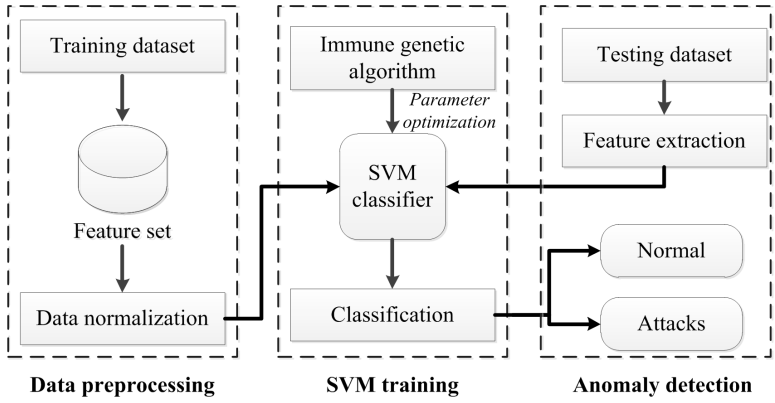

Fig. 1. Framework of network anomaly traffic monitoring system

Expressions (1) and (2) are convex functions, so that the required optimization problem can be transformed into solving a quadratic convex optimization problem. Among them, $\|w\|/2$ is the classified interval, and b is the constant of discriminant function.

Here vector x_i is mapped to a high-dimensional space by the function $\varphi()$, and C denotes the penalty parameter.

Thus, the classification issue is tackled by the following equation.

$$\operatorname{sgn}\{(w \bullet x) + b\} = \operatorname{sgn}\{\sum_{i=1}^{l} \alpha^* y_i (x \bullet x_i + b^*)\} \tag{3}$$

where $(x_i \bullet x)$ is the vector product of two vectors, while α^*, b^* are parameters of the classification hyperplane.

To promote the SVM's performance, Gaussian kernel is utilized in our work, that is defined as follows.

$$K(x_i, x) = \mathrm{e}^{-\frac{|x - x_i|^2}{\sigma^2}}, \tag{4}$$

where x_i is the kernel function center and σ is the width parameter of the function, which controls the radial range of the function.

3. SVM parameters estimation by immune genetic algorithm

To optimize parameters of SVM, we introduce the immune concept to genetic algorithm, by using local characteristic information to obtain optimal solutions. That is to say, the immune genetic algorithm (denoted as IGA) refrains the degenerative phenomena arising in the process of evolution, and then lets the fitness of population increase slowly. Next, we proposed an improved immune genetic algorithm, which is used to optimize parameters of SVM. The flowchart of the immune genetic algorithm is illustrated in Fig. 2.

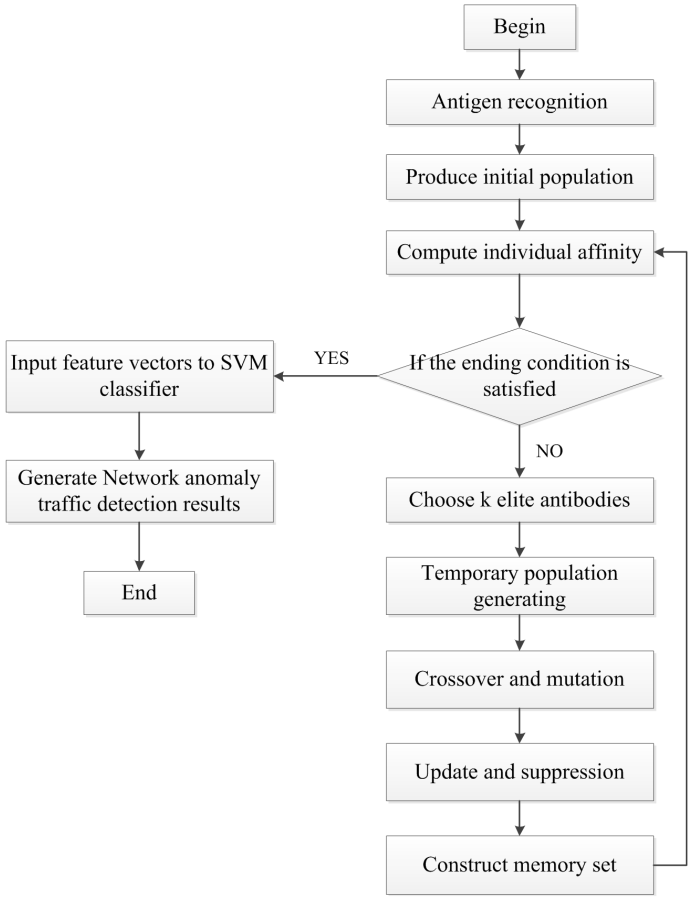

Fig. 2. Flowchart of immune genetic algorithm

As shown in Fig. 2, the immune genetic algorithm is made up of the following steps:

1. Randomly produce an initial population of antibodies.

2. Estimate the value of affinity for each antibody by the following equation.

$$c(x,y) = \frac{1}{1+d(x,y)}\,, \tag{5}$$

where

$$d(x,y) = \frac{1}{n}\sum_{i=1}^{n} |x(i)-y(i)|\,, \tag{6}$$

$d(x,y)$ denoting the proximity degree between the antibody x and y.

3. Choose k antibodies which have the largest affinity value.
4. Clone the antibodies of step 3.
5. Execute crossover and mutation operations for the cloning sets.
6. Construct memory set and then update the antibodies in it.
7. If the ending condition is satisfied, input feature vectors to SVM classifier and then generate network anomaly traffic detection results.
8. Otherwise, go to step 2.

In order to promote the performance of immune genetic algorithm, some modifications are made in this work.

The probability of choosing antibody K is calculated as follows

$$P(k) = \frac{f_k \mathrm{e}^{-\chi C_k}}{\sum_{i=1}^{N} f_k \mathrm{e}^{-\chi C_k}}\,, \tag{7}$$

where parameter χ denotes the regulation, which refers to the weight of affinity f_k and concentration c_k, and N refers to the number of antibody in a given population. We randomly choose antibody $x(k)$ and $x(l)$ in a population, and crossover the process of $x(k)$ and $x(l)$ in the qth bit is defined as follows

$$\begin{aligned} x_{kq} &= (1-\eta)x_{kq} + \eta x_{lq}\,, \\ x_{lq} &= (1-\eta)x_{lq} + \eta x_{kq}\,, \end{aligned} \tag{8}$$

where η is a random number which is valued in the range [0,1].

Afterwards, the mutation operation is executed to avoid premature convergence of the immune genetic algorithm. For a given antibody $x(k)$, the mutation of qth base is defined by the following equation.

$$\begin{aligned} x_{kq} &= (b_q - x_{kq}) \cdot f(t) \quad \text{for } \eta > 0.5\,, \\ x_{kq} &= (x_{kq} - a_q) \cdot f(t) \quad \text{for } \eta \leq 0.5\,. \end{aligned} \tag{9}$$

Here, parameters a_q and b_q denote the low and upper bounds, $f(t)$ means the probability of mutation, and t means the number of iterations.

4. Experiment

In order to test the capability of the proposed method, we collect five weeks of the 1998 DARPA dataset [18], which has been successfully exploited in network intrusion detection evaluation. The DARPA dataset is made up of nearly 1.5 million traffic instances with half of them is tagged attacks. Particularly, to illustrate the input data, six types of network traffic instance are chosen in this dataset, including: 1) connection time, 2) protocol type, 3) source port, 4) destination port, 5) source IP

address and 6) destination IP address. Afterwards, a 14-dimensional feature vector is input to SVM, which is illustrated in Table 1.

Table 1. Features utilized in network anomaly traffic monitoring

Feature description	Number of features
Connection time	3
Packet type	1
Source port number	1
Destination port number	1
Address of source IP	4
Address of destination IP	4

Table 1 shows that fourteen features are used to describe feature vector of network anomaly traffic monitoring. Moreover, to enhance the accuracy of network anomaly traffic detection, we normalize all features to the range [0,1]. The performance criteria used in this work are 1) Attack Detection rate (denoted as DR) and 2) False Alarm rate (denoted as FA). DR is defined as the ratio between the number of truly detected attacks and the whole attacks; on the other hand, FA is defined as the ratio between the number of normal connections which are falsely classified as attacks and the whole normal connections.

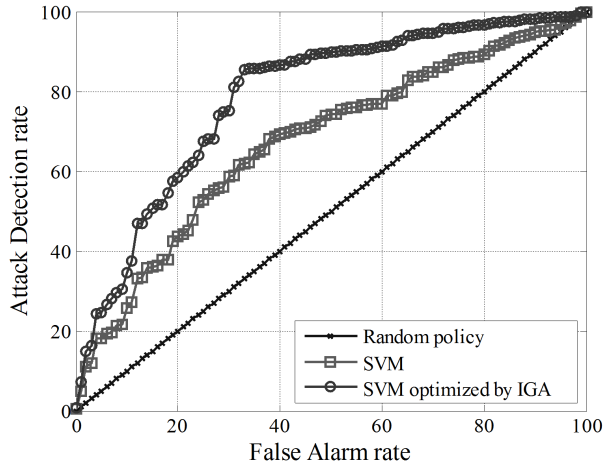

Fig. 3. ROC Curve under 20 % attack distribution

Integrating experimental results from Fig. 3 to Fig. 6, we can see that the proposed SVM classifier optimized by IGA performs better than SVM, that is to say the improved genetic algorithm is able to estimate parameters of SVM with high accuracy. Moreover, with the degree of attack distribution increasing, the areas of ROV curve significantly decrease.

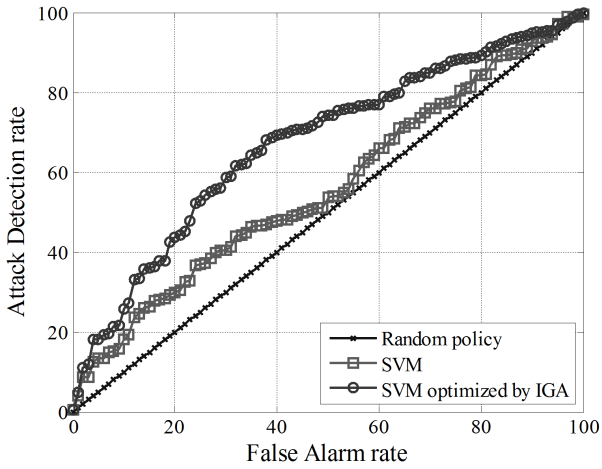

Fig. 4. ROC Curve under 40 % attack distribution

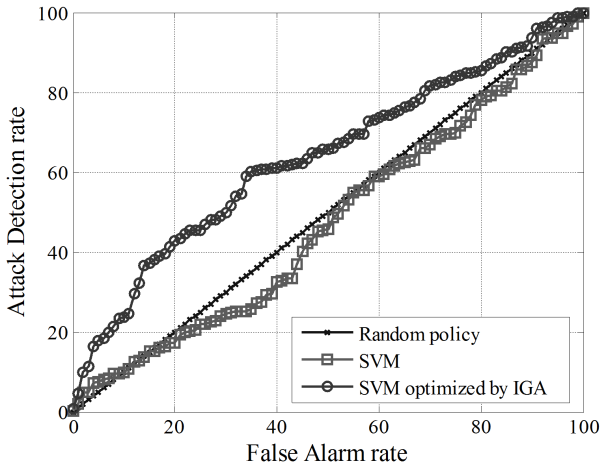

Fig. 5. ROC Curve under 60 % attack distribution

Apart from the above experiments, we also utilize KDD-CUP'99 anomaly detection dataset to further test the effectiveness of our method. In KDD-CUP'99 dataset, the simulated attacks are classified into four classes, which are 1) DoS: Denial of Service, 2) R2L: unauthorized access from a remote computer, 3) U2R: unauthorized access to super users, and 4) Probing: surveillance and probing for vulnerabilities. The accuracy of network anomaly detection for all the above four types of attacks is given in Fig. 7.

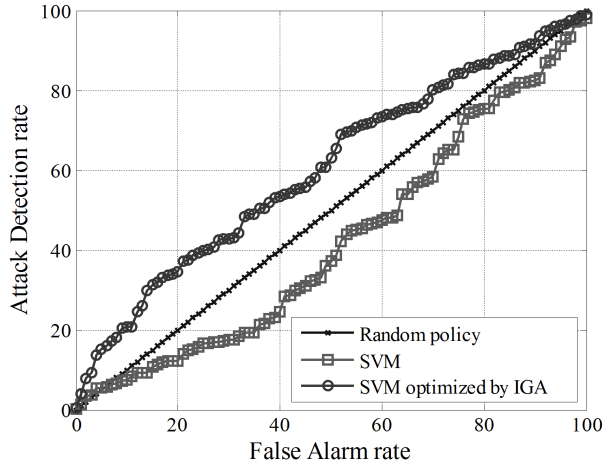

Fig. 6. ROC Curve under 80 % attack distribution

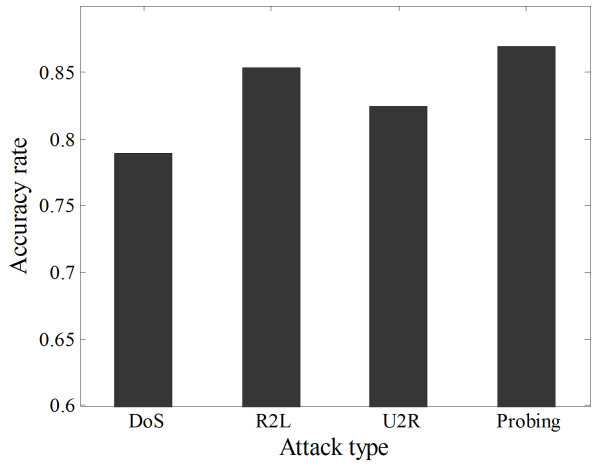

Fig. 7. Network anomaly detection accuracy for different attack types

Figure 7 demonstrates that the proposed algorithm can effectively detect four different types of attack in the KDD-CUP'99 dataset.

5. Conclusion

This paper presents a new network anomaly traffic monitoring method based on an improved immune genetic algorithm. The network anomaly traffic monitoring system is made up of three steps, including a) Data preprocessing, b) SVM train-

ing, and c) Anomaly detection. The main idea of this paper lies in that we convert the network anomaly traffic detection to a classification problem. Experiments are conducted using both 1998 DARPA dataset and KDD-CUP'99 dataset, and experimental results reveal that the proposed method can detect network anomaly traffic with high accuracy.

References

[1] Y. Lou, P. Li, X. Hong: *A distributed framework for network-wide traffic monitoring and platoon information aggregation using V2V communications.* Transportation Research Part C: Emerging Technologies *69* (2016), 356–374.

[2] Y. Liu, J. Ling, Q. Wu, B. Qin: *Scalable privacy-enhanced traffic monitoring in vehicular ad hoc networks.* Soft Computing *20* (2016), No. 8, 3335–3346.

[3] T. D. Truong, G. Cheng, A. Jakalan, X. J. Guo, A. P. Zhou: *Detecting DGA-based botnet with DNS traffic analysis in monitored network.* J Internet Technology *17* (2016), No. 2, 217–230.

[4] J. J. Fernández-Lozano, M. M. Guzmán, J. M. Ávila, A. G. Cerezo: *A wireless sensor network for urban traffic characterization and trend monitoring.* Sensors *15* (2015), No. 10, 26143–26169.

[5] A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato, H. Hlavacs: *The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring.* IEEE Trans. Intelligent Transportation Systems *16* (2015), No. 5, 2551–2572.

[6] A. Baiocchi, F. Cuomo, M. De Felice, G. Fusco: *Vehicular ad-hoc networks sampling protocols for traffic monitoring and incident detection in intelligent transportation systems.* Transportation Research Part C: Emerging Technologies *56* (2015), 177–194.

[7] W. Xue, L. Wang, D. Wang: *A prototype integrated monitoring system for pavement and traffic based on an embedded sensing network.* IEEE Trans. Intelligent Transportation Systems *16* (2015), No. 3, 1380–1390.

[8] J. Zhang, L. Jia, S. Niu, F. Zhang, L. Tong, X. Zhou: *A space-time network-based modeling framework for dynamic unmanned aerial vehicle routing in traffic incident monitoring applications.* Sensors *15* (2015), No. 6, 13874–13898.

[9] C. W. Chang, G. Huang, B. Lin, C. N. Chuah: *LEISURE: load-balanced network-wide traffic measurement and monitor placement.* IEEE Trans. Parallel & Distributed Systems *26* (2015), No. 4, 1059–1070.

[10] M. A. Hannan, C. T. Gee, M. S. Javadi: *Automatic vehicle classification using fast neural network and classical neural network for traffic monitoring.* Turkish Journal of Electrical Engineering and Computer Sciences *23*, (2015), No. Sup. 1, 2031–2042.

[11] S. Yang, M. Yang, S. Wang, K. Huang: *Adaptive immune genetic algorithm for weapon system portfolio optimization in military big data environment.* Cluster Computing *19* (2016), No. 3, 1359–1372.

[12] H. C. Huang, S. S. D. Xu, C. H. Wu: *A hybrid swarm intelligence of artificial immune system tuned with Taguchi-genetic algorithm and its field-programmable gate array realization to optimal inverse kinematics for an articulated industrial robotic manipulator.* Advances in Mechanical Engineering *8* (2016), No. 1, paper 1687814015626380.

[13] T. S. Wang: *Integrating grey sequencing with the genetic algorithm-immune algorithm to optimize touch panel cover glass polishing process parameter design.* IJ Production Research *54* (2016), No. 16, 4882–4893.

[14] W. A. Yang, M. Xiao, W. Zhou, Y. Guo, W. Liao, G. Shen: *Trace ratio criterion-based kernel discriminant analysis for fault diagnosis of rolling element bearings using binary immune genetic algorithm.* Shock and Vibration *(*2016), No. 15, paper 8631639.

[15] R. J. Kuo, S. H. Lin, Z. Y. Chen: *Integration of particle swarm optimization and immune genetic algorithm-based dynamic clustering for customer clustering.* IJ Artificial Intelligence Tools *24* (2015), No. 5, paper 1550019.

[16] R. J. Kuo, Y. H. Lee, F. E. Zulvia, F. C. Tien: *Solving bi-level linear programming problem through hybrid of immune genetic algorithm and particle swarm optimization algorithm.* Appl. Math. Compu. *266* (2015), 1013–1026.

[17] D. Song, X. P. Fan, Z. L. Liu: *An immune memory optimization algorithm based on the non-genetic information.* Acta Physica Sinica *64* (2015), No. 14, paper 140203.

[18] M. M. Mohammadi, A. Akbari, B.Raahemi, B. Nassersharif, H. Asgharian: *A fast anomaly detection system using probabilistic artificial immune algorithm capable of learning new attacks.* Evolutionary Intelligence *6* (2014), No. 3, 136–156.

Received November 16, 2016