Research on Algorithm Based on Secure Computer Network Defense

Tian Jifeng[1, a]

[1]School of information engineering, Shandong Yingcai University, Jinan, 250104, China

[a]i_drizzle@126.com

Keywords: Network security, Adaptive virtual framework, Particle swarm, Immune

Abstract. In order to solve the problems that the network security defense measures are independent, passive and lagging, and anomaly detection needs an effective training set, a scalable dynamic compound virtual network framework and strategy for active defense is designed and realized in this paper, a classification method based on real network data is also proposed. While PSO-FCM clustering algorithm is used to analyze the data in real network, immune evolutionary algorithm is used to dynamically adjust the number of clusters. Experimental results show that this algorithm has correct cluster and standard of network data, and can extract relatively pure training data from the network data.

Introduction

With the rapid development of computer network technology and the improvement of social informatization, computer network plays an increasingly important role in the fields of politics, economy, military, life and so on [1] . However, there are a lot of hacker attacks, Trojans, viruses and other security threats in the internet. Besides, with the continuous innovation of network attack technology and the emergence of new types of attack tools, the vulnerability of the network becomes more and more serious, and the network security issues are increasingly severe. If we don't understand our network security situation, which leads to information theft, destruction or hostile attacks, at best, such situations result in loss of entity economy; at worst, they cause paralysis in the whole country's politics, economy and military, even social orders out of control. How to fully and effectively aware of the status of network security has become an urgent problem to be solved[2, 3]. The existing network security defense technology mainly includes firewall, intrusion detection system, user authentication, data encryption and decryption, vulnerability scanning, anti-virus software and so on, but any single security technology can not ensure the security of network and system, and that most of the security defense technology is passive and lagging behind.

A dynamic compound virtual framework that can be used in all levels of network is designed and realized in this paper, which provided an initiative, pro-active and real-time intrusion prevention. An intrusion model based on improved immune optimization algorithm is proposed, which solved the problems that false alarms cause system load increasing and redundant information affects efficiency.

Overall framework and strategy of dynamic composite virtual network

Faced with complex and diverse network attacks and passive defense of the information network, we proposed to combine honeypot technology, network topology visualization technology, attack signature automatic extraction technology, intrusion detection technology and firewall technology together to design and realize a dynamic compound virtual framework that can be used in all levels of network and to provide an initiative, proactive and real-time intrusion prevention. A dynamic adaptive compound virtual network framework is shown in figure1. It is the compound virtual network within the dotted line that composed by front-end low-interaction honeypot network and back-end high-interaction honeypot network, which is based on honeypot technology. In order to realize initiative defense function of the compound virtual network, we designed and explored 5 major functional modules: Scanning module, policy module, characteristics generation module, data analysis module and linkage module.

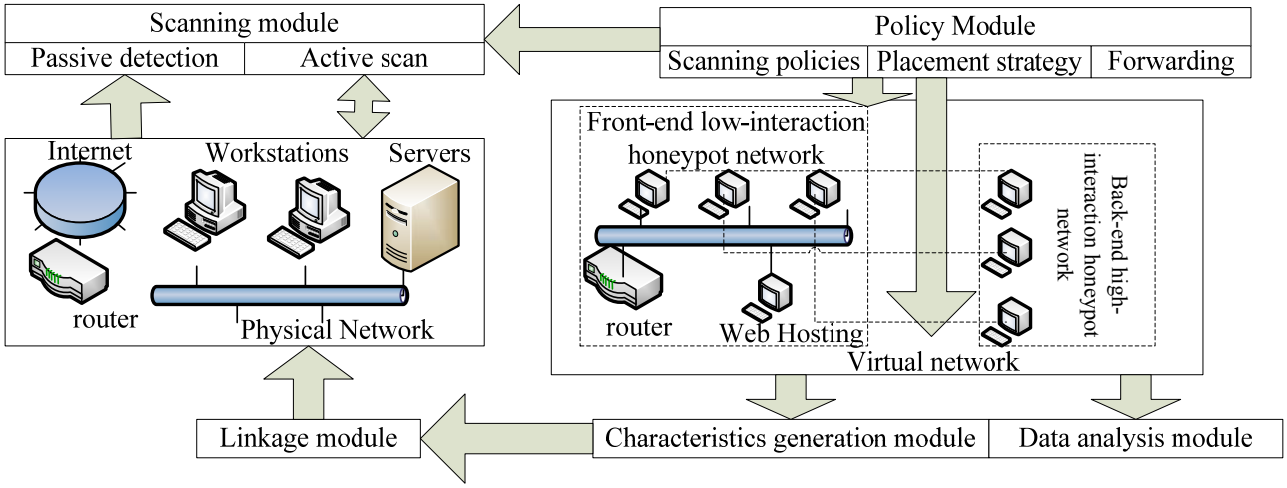

Figure1 dynamic compound virtual network framework

It is based on active scan and supplemented by passive detection, which has complementary advantages. Using Nmap as the active scanning tool, as the active vertical

Using nmap as active scanning tools do variable time interval of the active vertical and horizontal scanning to achieve physical network host discovery, operating system identification as well as port and service discovery, discovery technology uninterrupted passive service discovery with the help of a chapter proposed NetFlow network service and comprehensive real-time identification, discovery, and physical network tracking information and change tracking, also as far as possible to reduce the active recognition and tracking the impact of physical network, consumption takes up the least amount of system resources.

The basic idea of deploying front Honeyd virtual honeypot is using free IP addresses to arrange Honeyd virtual honeypot as much as possible, extend address space, attract attacks to protect the physical network and host, at the same time to assure the glory of the virtual Honeypot and fraudulent concealment, enable virtual M network close to the characteristics of physical network, and create the virtual honeypot configuration of the operating system to maintain the original physical network operating system distribution ratio.

If we use NP to represent the total number of physical hosts, NPoi to represent number of physical hosts that run certain operating systems, NFIP to represent total number of idle IP addresses that can be used to deploy virtual hosts, NHDoi to represent number of virtual hosts(honeypots) that can be used to simulate i^{th} operating system, so:

$$NHDo_i = (\frac{NFIP}{NP} \times NPo_i \times (1 - NUIP)) \qquad (1)$$

NUIP represents IP address proportion that is reserved in the idle IP. $NCIO_i$ represent the number of connections expected to accept the invasion, so the number of operating system/high-interaction honeypot that need to be arranged is:

$$NPHo_i = NCIo_i / NCo_i \qquad (2)$$

The average number of connections of low-interaction honeypot network in given operating system i is shown as $NHCo_i$, and the number of low-interaction honeypot of real connections is:

$$NHDRo_i = \min(NCIo_i / NHCo_i, NHDo_i) \qquad (3)$$

So, when an intruder invaded virtual network, probability of interaction with virtual host is:

$$P_{HF} = \sum_i HNHDR / (\sum_i NPo_i + \sum_i NHDRO_i) \qquad (4)$$

The number of Honeyd virtual honeypot, specified operating system, open ports and services are specified by a configuration file, the specific configuration of the system that according to the scan results and allocation strategies is dynamically updated, Honeyd server adopt the daemon to read configuration files to create and update the virtual honeypot, and manage the entire low interaction virtual networks. In front end, the low-interaction honeypot network is based on network feature

generation technology of honeycomb, and back-end high-interaction honeypot is based on host feature generation technology of Argos, both sides generated attack characteristics at the same time in order to achieve complementary feature extraction. And a new feature purification method to remove redundant features is put forward, in order to reduce the number of feature generation, and further reduce characteristics of redundant information, decrease each feature of the size (in bytes), and improve the generation characteristics of effectiveness, availability.

Design of defense algorithm based on improved fuzzy particle swarm

Design of intrusion data classification algorithm

From the strategic points, intrusion detection is divided into misuse detection and anomaly detection. Misuse detection is developed earlier, also known as knowledge based detection. In the detection, if the intrusion behavior characteristics match with the known intrusion signature database, the detection is considered to have the intrusion behavior. [5,6] This detection strategy can be targeted to establish a detection system, and has high detection efficiency, but it'll be powerless in the face of unknown intrusion behaviors. Then the anomaly detection strategy is appeared, which is assumed to be different from the normal behavior. This strategy will establish a normal behavior model base, during the detection, it'll compare the current behavior with the model base, when there is a big difference, and it is considered that the current behavior may be an invasion; therefore, this detection strategy has the potential to detect unknown intrusions.

Based on clustering algorithm of particle swarm K-means (PSO-KM), this algorithm improved the overall searching abilities of K-means. PSO-FCM is introduced into intrusion detection, and the ideal detection results are obtained, but this algorithm need to make sure the cluster number c, and the certain of c value has played a decisive role in detection results, which cannot be clustered facing unknown network data. At the same time, this algorithm is also easy to fall into the local optimal situation.

(1) Particle coding

In order to construct a particle swarm with different numbers of clusters, the position of each particle is composed of a vector matrix and a control vector. The vector matrix is a $C_{max} \times l$ matrix, C_{max} is the biggest cluster number, 1 is the dimensions of the data, and each row of the matrix represents a cluster center. The control vector is a column vector of the C_{max}, the numerical value is composed of "0" and "1", "1" represents the dimension of the corresponding matrix is valid, while the "0" is not valid. The total number of "1" represents the numbers of clusters of particles owned by the particle swarm.

(2) Fitness function and diversity function

According to the structure of the fitness function, it is defined as the formula (5).

$$f = \frac{1}{|B_1^{-1} B_2|} \qquad (5)$$

In the formula, B_1 is the matrix within samples; B_2 is between-class scatter matrix. The smaller the value of the fitness function is, the better the clustering results. The more the particles disperse in the space, the higher of the diversity is; the less they disperse, the lower of the diversity is. Take the average distance of the group as the evaluation function of diversity; it is defined as formula (6):

$$D = \frac{1}{N \cdot R} \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{l} (x_{ij} - \overline{x_j})^2} \qquad (6)$$

In the formula, N represents the size of the particle swarm, R represents the maximum radius of the space, x_{ij} is the j dimensional number of No.i particle, $\overline{x_j}$ is the average value of the j dimensional number of total particle.

When the diversity function value is lower than a reference value, the particle swarm optimization is carried out to enhance the diversity of the particle swarm, reference value defined as:

$$D_0(t)=\begin{cases}\alpha(1-\dfrac{t}{\beta\cdot t_{\max}}) & t<\beta\cdot t_{\max}\\ 0 & \beta\cdot t_{\max}<t<t_{\max}\end{cases} \qquad (7)$$

In the formular, αand β are control coefficients, t_{max} is maximum iterations. $D_0(t)$ can be reduced from α to 0, which meet the requirements of the algorithm for diversity.

(3) Immune evolution and adjustment

According to the fitness value of particles, the particles of high fitness are selected to enter the next iteration, after the probability of selection in each of the immune evolution, save the highest part of the adaptation of particles as memory particles as the variation and replacement of the guidance data. It is defined as:

$$P_1=\frac{f(x_i)}{\sum_{i=1}^{c} f(x_i)} \qquad (8)$$

$$P_2=\frac{f_{\max}-f}{f_{\max}} \qquad (9)$$

In the formula, f_{max} represents fitness value of the best particles.

(4) Type algorithm

In the final clustering results marking $G=\{C_1,C_2,\Lambda\ ,C_k\}$, according to the various types of abnormal factors to sort, and mark them as "normal" and "attack" according to certain proportion. Define abnormal factors C_i as $Z(C_i,y)$ as followed(10):

$$Z(C_i,y)=\left(\frac{\sum_{i\neq j} d(C_i,C_j)^y}{k-1}\right)^{\frac{1}{y}} \qquad (10)$$

In the formula, k represents numbers of clusters. When y < 1, the smaller the distances, the greater the impact of abnormal factors. When y > 1, the bigger the distances, the greater the impact of abnormal factors. When marking types, mainly affected by such recent distance, so y ≤ 1. According to abnormal factors $Z(C_i,y)$ decreasing order, $C_i(1\le i\le k)$ meet the smallest d in the formula(11), mark $C_1,C_2,\Lambda\ ,C_d$ as "attack", and $C_{d+1},C_{d+2},\Lambda\ ,C_k\}$ as "normal".

$$\frac{\sum_{i=1}^{d}|C_i|}{|G|}\ge\lambda \qquad (11)$$

Intrusion data classification process

Fusion control particle swarm optimization and immune evolution of intrusion data classification process are shown as figure 2.

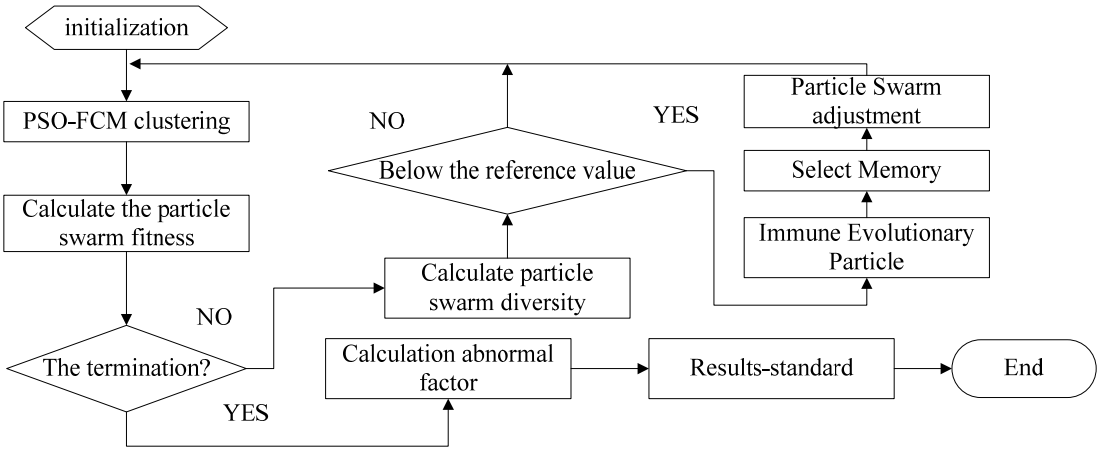

Figure 2 Intrusion data classification process

Basic PSO-FCM clustering algorithm process

PSO-FCM algorithm is as followed:

Step 1：Initialization of the particle swarm and parameters, set the fuzzy coefficient as m and the number of clusters as c, and c sample points were randomly selected as cluster centers；

Step2：To calculate new cluster centers according to (:7) in chapter 4.3.2, and calculate the corresponding membership matrix to form a new particle by formula(8)；

Step 3：The reciprocal of the formula (6) is used to calculate the fitness of the particles, to record the individual and global extremes, and to update the position and velocity of the particles;

Step4：To decide should terminate or not, if it is in accordance with the termination conditions, then end the algorithm, if not, return to (b).

Example verification of computer network defense algorithm

Choose a subset including 40000 records from 10% of KDD CUP 99 database, and to test fusion control particle swarm and classification method of immune evolutionary intrusion data. In this subset, normal data covers 96%. Before classification, a data preprocessing method is used to normalize the 41 attributes of the data set.

Using PSO-FCM clustering algorithm, fusion control particle swarm, and immune evolution of intrusion data algorithm respectively, and we can classify the experimental data. To set certain cluster numbers for PSO-FCM, c = 10、 c = 15 and c = 20, and set biggest cluster number Cmax = 30 for improved cluster setting. The convergence curve is shown in Figure 3.

From figure 3 we can see that because cluster numbers of PSO-FCM algorithm is certain, all particles are classified into different clusters according to iterations, therefore all curves are steep and the convergence speed is relatively fast. After 5 times of iteration of PSO-FCM when c = 10, we can see a minimum particle fitness, and when c = 15 or c = 20, after 7 and 10 iterations, the clustering algorithm is completed. But in the number of different clusters, the final fitness of the particle has large fluctuation. When c = 10, c = 15 and c = 20, the numbers are respectively 0.18, 0.06 and 0.08, because the right clustering number is 16, when c = 15, we got better clustering results, the clustering result is directly related to the accurate determination of the number of clusters. Because of several kinds of particles' coexistence, at the original state, improved clustering algorithm resulted in big fitness value, during the iteration, the curve has shown several steep, while the algorithm adjust particles scales and immune and evolutes particle swarm, optimization speed of fitness value has been improved, finally, it got the best particle fitness number 0.03 after 15 iteration, which is better than PSO-FCM clustering algorithm.

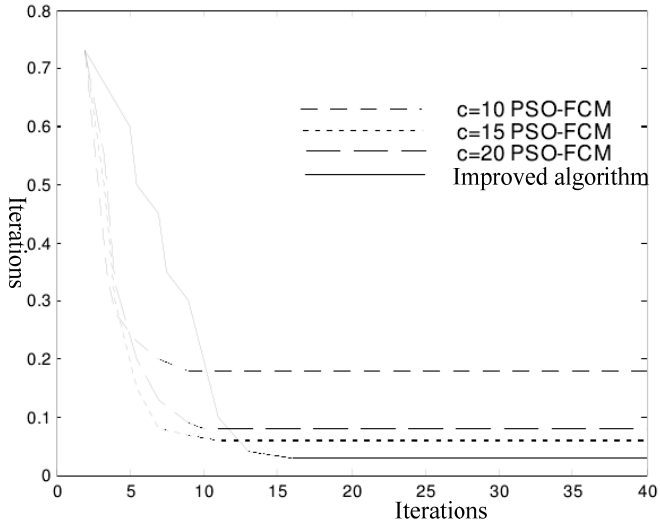

Figure 3 the convergence curve

The clustering results involved 16 clusters, the results of marking accroding to abnormal factors (λ = 0.05) and the sizes of clusters are shown in table 1.

Table 1 the results of marking accroding to abnormal factors (λ = 0.05) and the sizes of clusters

number	size	Normal number	Attack number	Marking as abnormal factors	Marking as size
1	19570	19553	17	normal	normal
……	……	……	……	……	……
6	1582	1580	2	normal	normal
7	1149	127	1022	attack	normal
8	225	0	225	attack	attack
……	……	……	……	……	……
16	10	9	1	normal	attack

We can see from table 1 that size has great numbers and normal numbers covers a lot, so we canhave right marking type such as type 1, the size is 19670, normal number is 19553, which covers 99.91%. Right now, both two types of clusterings mark it as normal ones; while about large numbers of clusters and attack numbers have wrong clustering, for example 7, the size is 1149, and abnormal number is 1022, which covers 88.95%, and the other algorithm also mark it as normal; and about small numbers of clusters, it has most normal clusters, which is also wrong like 16. As you see, this algorithm that adopts clustering based on abnormal numbers are scientific, and has right clustings.

Conclusion

Network attacks are diverse and complex, and the network security defense measures adopted by the network are independent, passive and lagging. In this paper based on honeypot technology, and the integration of network topology visualization technology, attack signature automatic extraction technology, intrusion detection technology and firewall technology, the design and implementation of the oriented active defense of the scalable dynamic composite virtual network framework and the five basic functional modules, to provide active, proactive, real-time intrusion prevention system.

The data clustering algorithm of PSO-FCM was improved, various controllable particle swarm have same time iterative evolution, and finally make the data polymerization in an optimal cluster number, which overcomes the defects of the cluster number is determined in advance; while immune principle is introduced into the particle evolution, which has kept the diversity of particles and avoided the precocity of clustering algorithm; in the final clustering results, the standard algorithm based on abnormal factor reduces the standard errors. Finally, the experimental results show that the

proposed algorithm has the correct clustering and standard of the network data, and can extract the relatively pure training data from the network.

References

[1]Liu He-an．Immune based novel intrusion prevention model．Application Research of Computer, Vol 29（7），2012,pp 2712-2714.

[2] Li Pengwei, Ge Wenying．Research on network virus intrusion prevention technology．Coal Technology, Vol 31（09），2012,pp219-221

[3] Liu Yanhua, Zhou Liuhong, Chen Guolong．Feature selection of the intrusion detection logs based on particle swarm optimization．Journal of Fuzhou University(Natural Science Edition), Vol 39（6），2011,pp 811-818

[4] Hu Daoyuan, Min Jinghua．Network security technology．Beijing: Tsinghua University Press, 2004，pp120

[5] CNCERT．CNCERT Internet security threat report(37)．Beijing: CNCERT, 2014

[6] Macfarlane R., Buchanan W., Ekonomouet E., et al. Formal security policy implementations in network firewalls. Computers & Security,2012,V 31(2): 253-270

[7] Yao Yunzhi, Tian Yuling. An improved artificial immune-based intrusion detection model. Computer applications and software, Vol （1），2014, pp308-310.

[8] Xia Qin, Wang Zhiwen, Lu Ke. Improved artificial immune intrusion detection model. Journal of Xi’an Jiaotong University, 2013, (2).