

Research on Computer Network Virus Defense Technology in Cloud Technology Environment

Zhao Sheng, Han HuiShan, Shi XueKui

Xingtai polytechnic college
593826346@qq.com

Abstract

With the rapid development of the Internet, the antivirus software of the network is always emerging and constantly changing. Traditional detection methods can't effectively kill the new viruses and malicious software, the complexity of which also makes itself easy to be attacked by malicious software. The emergence of cloud computing has changed the status quo. Therefore, the architecture model of virus malware detection based on cloud computing is proposed in this paper. Based on the combination of the method for detecting malicious software virus based on cloud computing and the algorithm analysis theory in machine learning, a new type of distributed CFO algorithm is proposed, and the closed environment of cloud computing virtual machine nodes is used to realize dynamic behavior monitoring to the virus malware, then the distributed fluctuations PIF algorithm is used to describe the process of dynamic analysis and analysis reporting, besides, the wave algorithm is carried out corresponding improvement based on the analysis of the environment. Experimental results show that the model can detect the conditional trigger behavior of virus malware, so as to find the conditions for triggering malicious behavior and the input data that satisfy these conditions and the performance of this monitoring system is greatly improved compared with the common single machine system.

Keywords: *cloud computing; virus malware analysis; behavior based detection; central gravity optimization algorithm*

1. Introduction

With the development and popularization of Internet application technology, network security has been paid more and more attentions. Much antivirus software has been developed. Antivirus software is mainly to perform real-time monitoring and scanning disk. However, the effectiveness of traditional detection methods has been widely questioned. Due to the traditional detection methods can't effectively kill the new viruses and malicious software, and the complexity of which also makes itself easy to be attacked by malicious software. The emergence of cloud computing has changed the status quo. Cloud computing is a product of the integration of distributed computing, grid computing, utility computing, virtualization technology and other computer technology and network technology. It gathers a large number of computer resources, and provides users with a variety of IT services through the Internet, and then the users pay a fee in accordance with the amount of the use. Cloud computing can provide security services to end users. Cloud security services, namely, a large number of clients are used to monitor the behavior of the software in the network to get the latest information of the malicious software, such as Trojans, worms and others, and these information is sent to the cloud server for automatic analysis and processing, finally, the solutions of these virus malware are sent to each client. Based on this, this paper briefly introduces the related technologies, uses the PIF algorithm to describe the process of the detection and analysis of suspicious files, focuses on the distributed virus detection

mechanism in the overall architecture, and describes the virus malware dynamic behavior analysis system in detail. In addition, a static behavior detection method based on cloud computing is proposed on the basis of overall architecture of cloud computing - client, and an optimized distributed CFO algorithm is used in the cloud computing distributed environment, and through the integration of neural network to realize the classification of suspicious files. Experiment in cloud platform Eucalyptus shows that the model can detect the conditional trigger behavior of virus malware, so as to find the conditions for triggering malicious behavior and the input data that satisfy these conditions, and the performance of this monitoring system is greatly improved compared with the common single machine system.

2. Research Statuses

At present, the antivirus software has been popularized, however, the virus software has not been effectively curbed, and on the contrary, the virus infection rate has been rising. In view of the limitation of traditional virus detection methods, many scholars have put forward the methods of virus detection based on cloud computing. The development of foreign related technology is more mature. A new idea was put forward firstly, the strong distributed parallel processing capability of cloud computing was used to transplant the work of virus detection and analysis into the cloud computing to carry on, the analysis and testing of the executable files were carried out in the cloud (Jon Oberheide Et al. 2008) [1]. Intel company further expanded the method, a complete model of cloud virus detection was proposed, it added an archive feature to save the virus malware related features in the document (Carlos Rozas et al. 2009)[2]. Xin Wang used the cloud computing technology to make up for the shortcomings of the traditional virus detection methods, and extended the technology to the military network, and got good results(Xin Wang et al. 2010)[3]. Salah proposed a reliable model, the model not only could detect malicious virus software, but also could provide effective intercept service for distributed spam DDOS, the performance of the system was improved(Salah et al. 2013)[4]. A new type of MD5 lookup method is proposed, which improved the efficiency of virus malware detection (Nen-Fu Huang et al. 2011) [5]. The antivirus malware detection was extended to the mobile devices, and an Android application sandbox system was proposed, it could carry on the dynamic and static detection of the suspicious files (Batyuk L et al. 2010) [6]. Compared with the relatively mature research abroad, domestic related research started late. At present, virus malware often uses the code confusion technology, antivirus software often can't find the contents of the document. In order to effectively curb the development of such viruses and malicious software, many scholars have made a quite effective attempt. The malware signature automatic detection system AMSDS was proposed, it was smaller than the traditional signature database, and in virus detection model, users only needed to install a lightweight cloud signature collection, when the AMSDS couldn't detect the file, it would be reported to the cloud server(Wei Yan et al. 2009)[7]. A model CloudSEC of cooperative security system was proposed, it could resist a large number of distributed intrusion, and could be applied to cooperative security services in the cloud(Jia Xu et al. 2010)[8]. The remainder of this paper is organized as follows. Section 3 describes the related theories and key technologies: such as cloud technology, PIF algorithm, CFO algorithm and so on. Section 4 gives the design and construction process of the architecture model of virus malware detection based on cloud computing. Section 5 presents a real experiment to evaluate the model. Conclusion is summarized in Section 6.

3.Key Technologies

3.1 Cloud Technology

Cloud technology refers to a hosting technology that the hardware, software, network and other resources are unified to achieve the calculation, storage, processing and sharing of the data in the wide area network (WAN) or local area network (LAN). It is a product of the development of grid computing, distributed computing, utility computing, virtualization technology and service oriented technology [9]. Cloud computing makes full use of Internet to gather a large number of software and hardware resources to form a huge pool of resources, the ordinary users can enjoy the IT service provided by Internet. Cloud computing services are divided into IaaS, PaaS, and SaaS three types. This paper colligates the solutions of different manufacturers, and constructs a cloud computing architecture. Cloud computing system is a four layers structure, including: SOA component layer, the middle layer of management, resource pool layer, physical resource layer. Data center is the core of cloud computing technology, and its reliability has a great impact on the upper layer services, Google and other companies attach great importance to the construction of data centers.



Figure 1. Cloud Technology

3.2 Eucalyptus

Eucalyptus is the open source software research framework, it uses the modular design, and the components of which can be upgraded and replaced, so as to provide a good research platform for related researchers. Eucalyptus relies on Xen for virtualization. Eucalyptus provides access to computing resources and data through a variety of interfaces. The biggest innovation of Eucalyptus is that the IaaS can be achieved in the research environment for installation and maintenance, so as to facilitate the experimental modification and expansion. Completely based on IaaS, the calculation and storage facilities of which can be used by academic organizations, it provides a modular experimental platform, and allows researchers to test and research the scalability, security, and resource scheduling of the cloud computing. Eucalyptus has a very simple modular structure, it can be easily expanded. Eucalyptus uses open source web service, its internal structure is completely open, each component is composed of a number of services, which has a well defined document description interface. In general, Eucalyptus is particularly suitable for the study of cloud computing [10].

3.3 PIF Algorithm

Distributed PIF algorithm system is undirected graph $S=(V, E)$, V is a collection of nodes, it refers to the process of the collection in this paper, E is the channel of nodes, it refers to the process of communication among the channels of communication messages in this paper [11]. Process state information is a collection of all relevant variables state in the process. System configuration is a collection of related state of all processes in the system. In this paper, C represents all possible configurations in the distributed algorithm system. The protocol P of the distributed system is a collection of two elements that defines in C , namely " \rightarrow ". A calculation process e of the protocol p is a maximal sequence, the sequence satisfies: $e=\gamma_0, \gamma_1 \dots, \gamma_i, \gamma_{i+1}, i \geq 0, \forall \gamma_i \in C : \gamma_i \rightarrow \gamma_{i+1}$. Assuming that the γ_{i+1} is present, or γ_i is the termination configuration. The maximum sequence indicates that the sequence is either an infinite sequence or a finite sequence. In this system, all possible computation sets are \mathcal{E} .

Assuming that $x \vdash P$ represents : $X \in X$ satisfies the assertion P that defines in C . Definition 1: Assuming that T represents a task, SPT represents an examination rule of T , thus: $\forall e \in \mathcal{E} : e \vdash SPT$.

Assuming that PIF starts the operation from a process called the root node, the process is represented by r . Rule 1: The finite calculation process: $e=\gamma_0, \gamma_1 \dots \gamma_i, \gamma_{i+1}, \gamma_t \in \mathcal{E}$ is a PIF cycle process, when the following conditions are true: if the process r calculates $\gamma_0 \rightarrow \gamma_1$ broadcast a message m , then:

[PIF1]Any process $p \neq r$ has only $i \in [1, t-1]$, which meets p in the broadcast phase.

[PIF2]In the γ_t , r accepts the confirmation message that sent by any non root process participates in the broadcast phase.

3.4 CFO Algorithm

CFO algorithm refer to search for the optimal value in the objective function $f(x_1, x_2 \dots x_{N_d})$, region $\Omega : \min(x_k) \leq x_k \leq \max(x_k) \quad 1 \leq k \leq N_d$. X_k is a variable of N_d dimension search space, and N_d is the dimension of the decision space of the algorithm, $\min(X_k)$ and $\max(X_k)$ represent the minimum and maximum values of the algorithm in the K dimension [12]. The basic CFO algorithm includes the motion equation of two proton, the formulas are shown as follows. The equations determine the trajectory of the proton group in the iteration.

$$a_{j-1}^{-p} = G \sum_{\substack{k=1 \\ k \neq p}}^{N_d} U(M_{j-1}^k - M_{j-1}^p) \bullet (M_{j-1}^k - M_{j-1}^p)^\alpha \times \frac{(R_{j-1}^{-k} - R_{j-1}^{-p})}{\|R_{j-1}^{-k} - R_{j-1}^{-p}\|^\beta} \quad (1)$$

$$R_j^{-p} = R_{j-1}^{-p} + a_{j-1}^{-p}, j \geq 1 \quad (2)$$

a_{j-1} is the acceleration of proton p in the iterative step $j-1$. R_{jp} is expressed as follows:

$$R_j^{-p} = \sum_{k=1}^{N_d} x_k^{p,j} \overline{e_k} \quad (3)$$

Where R_{jp} is the position vector of the proton p in the iterative step j , where x_{kp} is the k dimensional coordinates, e_k is the unit vector in the direction of the x_k axis.

$1 \leq p \leq N_p$ and $0 \leq j \leq N_t$ indicate the number of protons and the corresponding iterative steps. N_p and N_t are the total number of protons and the total number of iterations.

$$M_{j-1}^p = f(x_1^{p,j-1}, x_2^{p,j-1}, \dots, x_{N_d}^{p,j-1}) \quad (4)$$

It represents the current position of the proton P, and the value of objective function in J-1 iteration. The remainder of the proton has an adaptive value at each iteration step:

$$M_{j-1}^p, k = 1, \dots, p-1, p, p+1, \dots, N_p \quad (5)$$

G represents gravity constant. U represents a single step function:

$$U(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Where α and β is 1 and 3 respectively. Generally in order to facilitate the calculation, β can be valued as 1. In the specific implementation of the algorithm, the quality is:

$$MASS_{CFO} = U(M_{j-1}^k - M_{j-1}^p) \cdot (M_{j-1}^k - M_{j-1}^p)^\alpha \quad (7)$$

4. Virus Detection System based on Cloud Computing

4.1 Overall Framework

This paper presents the architecture of the cloud detection, including the following components: First is the light host agent software. It can be run on the terminal system, such as the desktop system and mobile devices, this program can identify the new suspicious files, and send these files to the cloud for analysis. Followed is the network service component, the component can accept the suspicious files from the proxy host program, the different commercial antivirus engines are installed to parallel analyze the suspicious files, so as to find virus malware. At the same time, the behavior analysis engine is used to analyze the reported suspicious files, and then the results are reported to the host agent, so as to determine whether these suspicious files are safe. The last is the archive service component, it stores information about the results of the file analysis, and provides an interface for querying and operating management [13].

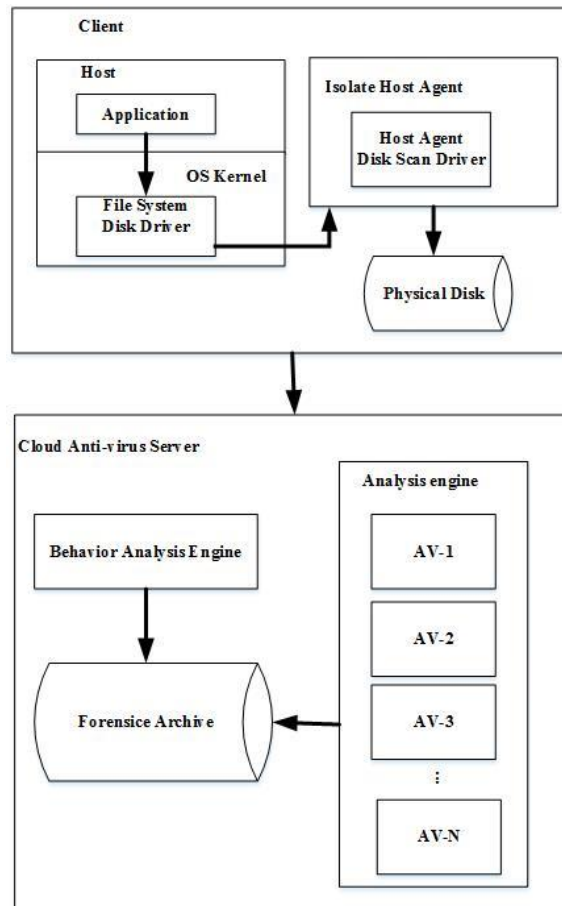


Figure 2. Overall Framework of Virus Detection System based on Cloud Computing

As shown in Figure 2, there are two important components in the network service section: One is the light detection engine - the collection of the heterogeneous detection analysis engine, the other is the heavy detection analysis engine - behavior detection engine. The two detection engines can save the detection result information through the archive service component, and provide the users with management and search service.

4.2 Distributed Light Detection Engine

Host Agent Software

The core part of the host agent software is the file UID generator. File UID generator provides the concise description of the suspicious files. File UID can be the only description of the file. The simplest way to get file UID is to use the file password hash function method. The file password hash function can provide a fairly good defense against attacks.

Network Service Component

In the architecture, the most important component is the network service part, which takes on the task of analyzing suspicious files. The core of the network service is to determine whether the submitted suspicious files are virus malware or normal files. Different from the current existing analysis methods, the cloud computing distributed parallel environment is used in the architecture, each

submitted file is detected and analyzed by a set of detection engine, so as to determine whether the file is malicious files [14].

Detection Engine Settings

Due to the scalability, the system can arbitrarily add additional detection engine. In the cloud, the light and heavy detection techniques are both used to analyze. The parallel detection and analysis engine needs cooperation and interactive information to improve the detection efficiency. This paper uses Eucalyptus open source cloud computing platform to achieve specific architecture. Based on the characteristics of open source cloud computing cluster manager, a distributed algorithm based on maximum independent set is proposed. The algorithm uses greedy algorithm to elect the node as the independent point, so as to construct minimum dominating set, that is, maximum independent set. Then the idea of divide and conquer is used to add non independent connectivity to connect independent set.

4.3 Dynamic Behavior Analysis system of Virus Malware

Based on the distributed environment of cloud computing, the distributed fluctuation algorithm is improved, virus state is added on the basis of the original basic PIF wave algorithm. When the behavior of virus malicious software is found, the corresponding node immediately sets the state of the virus, terminates the whole analysis process, and forms the analysis report, then reports to the general users that the virus malicious software is found. The architecture of this paper is based on the analysis and exploration of the multi branch path of the virus malware based on cloud computing. Cloud network architecture can be used to search for more than one branches of virus malware, so as to improve the efficiency [15].

Algorithm Description

PIF algorithm consists of four main stages: broadcast stage, feedback stage, feedback stage traces, and the virus stage. In order to improve the efficiency, the feedback phase and the clearance phase can be implemented simultaneously. In this paper, the PIF algorithm can ensure that the two phases don't affect each other. In the cleaning phase, the removal method of the leaf node is equivalent to the internal node. As shown in Figure 3, the cleanup phase is to remove the cycle traces of the previous PIF algorithm, and make preparations for the next PIF algorithm cycle.

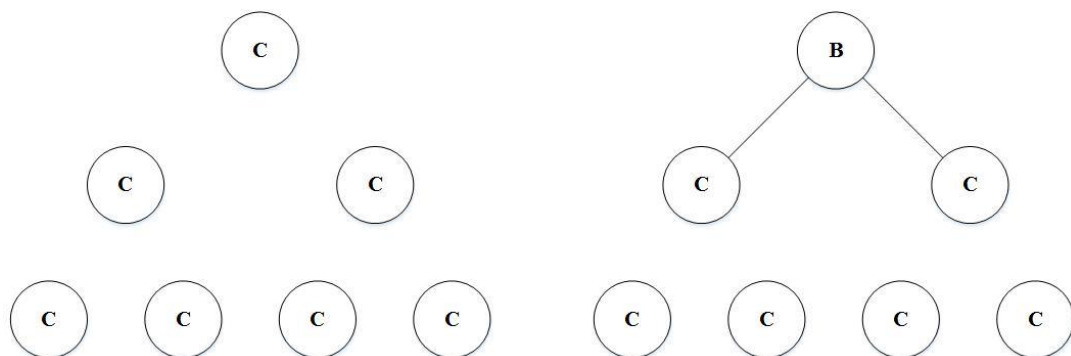


Figure 3. Initial Configuration of the System and the Configuration of the Second Round

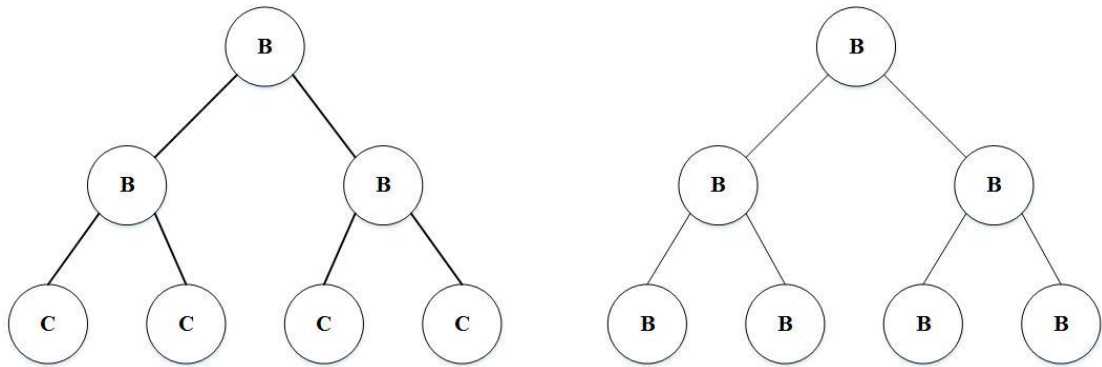


Figure 4. Configuration of the Third Round and the Configuration of the Fourth Round

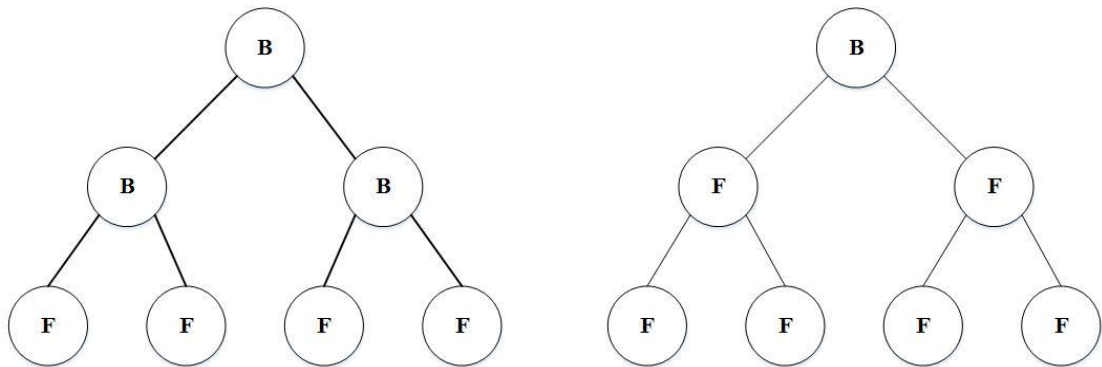


Figure 5. Configuration of the Fifth Round and the Configuration of the Sixth Round

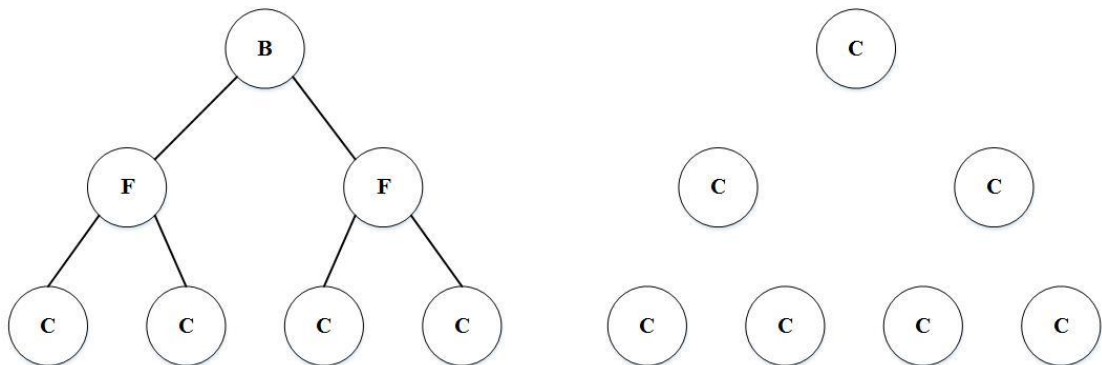


Figure 6. Configuration of the Sixth Round and the Configuration of the Seventh Round

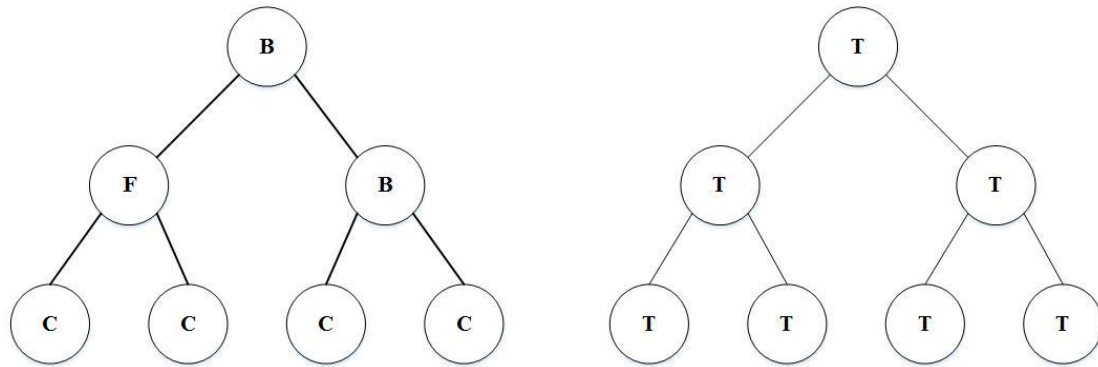


Figure 7. Virus Position T and the Entire System Location T

4.4 Static Behavior Analysis System of Virus Malware

The method based on the static behavior analysis is the way that the suspicious file system call sequence is statically analyzed to determine the behavior of the virus malicious software. The database update frequency of the detection method based on the static behavior is far below the detection method based on the characteristic value, the feature codes of many viruses malware are different, but they have the same behavioral characteristics. So once the new malware attacks appear, it is unnecessary to update the feature database.

CFO Algorithm Training Neural Network Classifier

The artificial neural network is used as a pattern classifier to classify suspicious files. A suitable algorithm for the training of artificial neural networks is required. This section uses the CFO algorithm to train the artificial neural network. CFO algorithm is a completely deterministic heuristic optimization algorithm, the algorithm can carry on the global multidimensional search. Due to the characteristics of the global and complete certainty of the CFO algorithm, the algorithm is especially suitable for training the artificial neural network to find the optimal solution in the multidimensional space.

Distributed Multi Objective CFO Algorithm Training Integrated Network

Although the CFO algorithm is robust and effective in solving the single objective problem, the algorithm can't solve the multi-objective optimization problem, so this paper propose a distributed multi-objective optimization CFO algorithm, which can solve the problem of multi objective function. Assuming that there are three groups of protons S_1, S_2, S_3 , the size is N , the three different objective function value Ψ_1, Ψ_2, Ψ_3 is respectively optimized. Each group is always aimed at one of the objective functions. Assuming that $a_{t-1,s}^p, R_{t-1,s}^p, M_{t-1,s}^p$ respectively represents the velocity, position, and objective function value of the p proton in group s in the $t-1$ iteration step. The improved algorithm proton groups constantly update according to the following two formulas:

$$a_{t-1,j}^p = G \sum_{k=1}^{N_p} E(M_{t-1,s}^k - M_{t-1,j}^p) \bullet (M_{t-1,s}^k - M_{t-1,j}^p)^\alpha \times \frac{(R_{t-1,s}^k - R_{t-1,j}^p)}{\|R_{t-1,s}^k - R_{t-1,j}^p\|^\beta} \quad (8)$$

$$R_{t,j}^p = R_{t-1,j}^p + a_{t-1,j}^p, t \geq 1 \quad (9)$$

Where j represents the number of protons in the group, the subscript s represents the topological structure of the migration. Specific definition is in formula 10. Algorithm always assumes that the search behavior of a proton group is always affected by the neighbor's behavior.

$$s = \begin{cases} 3 & \text{if } j = 1 \\ j-1 & \text{if } j = 2, 3 \end{cases} \quad (10)$$

In the multi-objective optimization problems of the three objective functions Ψ_1, Ψ_2, Ψ_3 , three proton groups correspond to three different objective functions, which finally determine the optimal solutions. The final result is not necessarily the optimal solution of the three objective functions, but it is a compromise solution among the three objective functions. Distributed multi objective CFO algorithm always assumes that the three proton groups carry on the optimization process in three different PC. The three PC machines are connected to a local area network, services are allowed to migrate from one PC to another PC. The sub network in the integrated network is a forward type neural network, the network has two layers of neurons: the input layer, hidden layer and output layer. The number of nodes in the input layer is a , the number of nodes in the hidden layer is: b_i ($i=1, 2, n$), the output node number is 1. In each study, the hidden layer nodes of the i sub network are calculated by using the following formula:

$$f(s_j) = \frac{1}{1 + \exp(-s_j)} s_j = \sum_{i=1}^a w_{ij} x_i - \theta_j \quad (j = 1, 2, \dots, b_i) \quad (11)$$

Where a is the input layer nodes, w_{ij} is the connection weights of input layer, and θ_j is the deviation of j hidden layer nodes, x_i is the input of i .

$$0 = \sum_{j=1}^{b_i} w_j f(s_j) - \theta \quad (12)$$

θ is the deviation of i , the next step is to choose a coding scheme to represent the weight and bias of forward network. This paper provides a concrete example of the use of the matrix encoding scheme, as shown in following formula:

$$probe(i) = [W_{ab}, b_b, W_{bo}, b_o] \quad (13)$$

$$W_{ab} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}, b_b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, W_{bo} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, b_o = b_o \quad (14)$$

Where W_{ab} is connection weight matrix from the input layer to the hidden layer, the B_b is the deviation matrix of the hidden layer, and the W_{bo} is the connection weight matrix from the hidden layer to the output layer, B_o is the deviation matrix of the output layer.

5. Experimental Analyses

This paper compares the virus detection and analysis system based on cloud with other virus detection and analysis methods, including MyDoom, NetSky, Tribe Flood Network (TFN), and Perfect Keylogger. Experimental environment: host Pentium 2.8 GHz, dual core CPU, 4GB RAM. Operating system: Ubuntu 8.10, version: Linux2.6.30. The test results as shown in Figure 8.

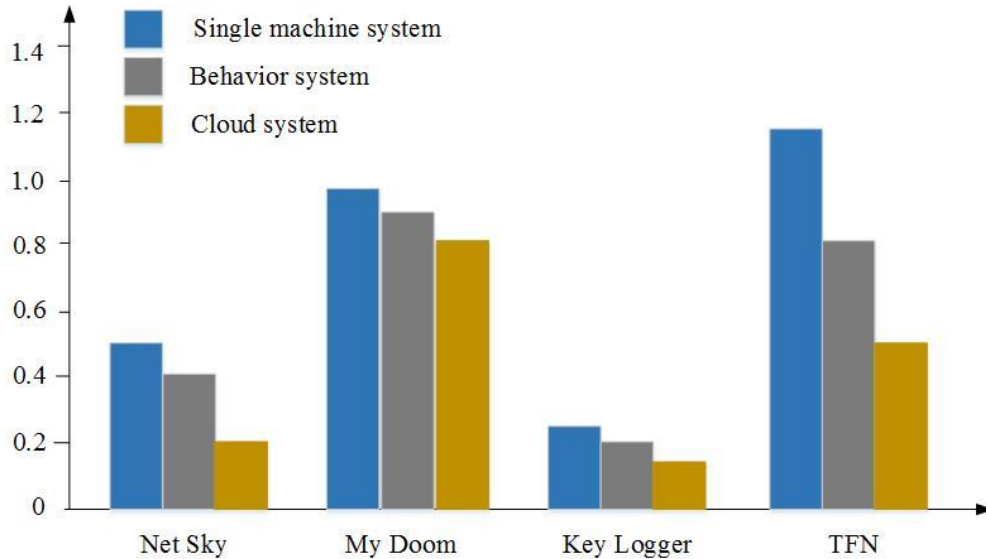


Figure 8. Experimental Result

It can be seen that the advantages of virus detection and analysis system based on cloud in time. The analysis time of Key Logger, TFN and Net Sky these three kinds of virus malware respectively increases by about 50%, the analysis time of My Doom is also improved.

6. Conclusion

With the rapid development of Internet technology, the virus malicious software spreads through the developed network. Traditional detection methods can't effectively kill the new viruses and malicious software. However, the powerful distributed computing capabilities of cloud computing are used to carry on the effective monitoring of virus malware and its variants has become a trend. In view of this situation, this paper proposes a model of virus detection architecture based on cloud computing. This model mainly consists of three parts: the distributed parallel detection mechanism, the dynamic behavior analysis and detection mechanism, the static behavior analysis and detection mechanism of virus malware. In the cloud computing environment, the CFO algorithm is applied to neural network training to classify suspicious files, and monitor the static behavior of malicious software. The PIF algorithm is applied to the dynamic behavior analysis of the virus malware, which further improves the efficiency of the analysis. In addition, the maximum independent set algorithm is used to select the virtual machine node to optimize network structure. These are the main innovation points of this paper. The experimental results show that the distributed parallel detection based on cloud computing greatly improves the detection accuracy compared with the single computer system. The static behavior analysis method based on cloud computing has strong generalization ability and high precision. Compared with the traditional method, detection rate of this method has greatly improved. But since the detection is synchronously completed in each virtual machine node, so the synchronization problem of each virtual machine node is a key problem. The in-depth study of the problem has become a hot research topic.

References

- [1] J. Oberheide, "Cloud AV: N-Version Anti-virus in the Network Cloud", Proceedings of the 17th Usenix Security Symposium, (2008), pp. 91-206.
- [2] C Rozas,H Khosravi,D Kolar Sunder,Y Bulygin.Enhanced Detection of Malware.Intel Technology Journal, vol. 13, no. 2, (2009).
- [3] X Wang, "Research on the anti-virus system of military network based on cloud security", 2010 International Conference on Intelligent Computing and Integrated Systems, (2010), pp. 656 - 659
- [4] K Salah, A Calero.S. Zeadally,.; Al-Mulla, "Using Cloud Computing to Implement a Security Overlay",IEEE Network Security & Privacy, vol. 11, no. 1, , (2013), pp. 44–53.
- [5] N-Fu Huang,C-N Kao,Rong-Tai Liu, "A novel software-based MD5 checksum lookup scheme for anti-virus systems. International Wireless Communications and Mobile Computing Conference (IWCMC), (2011), pp. 207 - 212
- [6] L Batyuk A.D ,Schmidt, S A Camtepe , S Albayrak, "An Android Application Sandbox system for suspicious software detection".International Conference on Malicious and Unwanted Software (MALWARE), (2010), pp. 55 – 62.
- [7] W Yan, E Wu, "Toward Automatic Discovery of Malware Signature for Autivirus Cloud Computing". Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering , (2009), vol. 4, pp. 724-728
- [8] J Xu,J Yan ,L He ,P Su CloudSEC: A Cloud Architecture for Composing Collaborative Security Services. IEEE Second International Conference on Cloud Computing Technology and Science,(2010), pp. 703–711.
- [9] W Yi, "Research on computer network security defense technology [J]", network security technology and application, (2015), no. 5, pp. 59-59.
- [10] P Deng, "Research on computer virus and its defense technology in network environment [J]". Silicon Valley, (2014), vol. 7, no. 4, pp. 83-84.
- [11] L Weijie, "Study on the implementation path of network security technology in the background of cloud computing". network security technology and application, (2015) no. 5, pp. 48-48.
- [12] Y Nenghai, H Zhuo, X Jiajia, "Progress in research on cloud security", Journal of electronics, (2013), vol. 41, no. 2, 371r381.
- [13] X Ying, "Cloud security under the framework of virus prevention research on the key technology of [D]", Beijing University of Posts and Telecommunications, (2013).
- [14] W Xiaodi, Z Yunyong, LDi, "Cloud computing virtualization security technology of", Telecom Science, (2015), vol. 31, no. 6, 2015154
- [15] Q Wei, Q Pan, Z Deqing, "Multi engine detection mechanism in the cloud defense system", Journal of Wuhan University, (2014), no. 5, p. 3.

Authors



Zhao Sheng, He is now a lecturer in Xingtai polytechnic college. His main research direction is the field of computer security.