Information Processing Letters 117 (2017) 19–24

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Research on dynamic heuristic scanning technique and the application of the malicious code detection model

Zhang Bo[a], Li Qianmu[a,*], Ma Yuanyuan[b]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Global energy interconnection research institute, Nanjing 210003, China

ARTICLE INFO

Article history:
Received 7 April 2016
Accepted 30 June 2016
Available online 7 July 2016
Communicated by X. Wu

Keywords:
Security indigital systems
Software engineering
Performance evaluation

ABSTRACT

With the rapid development of computer technology, people pay more attention to the security of computer data and the computer virus has become a chief threat to computer data security. By using an antivirus system that can identify randomly generated computer viruses and on the basis of the basic characteristics of the computer code, this paper investigates the heuristic scanning technique. This paper proposes the minimum distance classifier and detection model through the analysis of the malicious code. This model can identify unknown feature codes of illegal procedures and construct a healthy network environment by using a combination of model and experimental method, which can intercept the illegal virus program in the installation and operation stages.

1. Introduction

The rapid development of the network has brought the world within the scope of information sharing, which has significantly changed people's output and lifestyle. With the wide use of network in finance, defense, education, and other fields, there have also emerged several unsafe factors for network users. Network security has become a major issue in the process of the development of human social information. Therefore, the research on malicious code significantly contributes to improve network security.

There are several types of research on the malicious code detection technology such as the linkage of the firewall and intrusion detection technology, active defense technology, static signature detection technology, and behavior analysis technology [1]. Among them, the main technology is the behavior analysis technology, which can detect the signature of unknown illegal procedures. Furthermore, it is advantageous as it can minimize the behavior analysis. Johannes Kinder and coworkers described the malicious code by using the method of computer tree logic (CTL), and through the abstract generalization of CNF, this method has a good effect on proactive inspection, but the method can be transferred only through a level of assembly instructions. Zhangboyun used Naive Bayes and K-NN algorithms to detect unknown viruses. He also used a rough set to simplify the characteristic and avoid the loss of information. Relevant scholars from Germany placed the malicious code in the environment of the virtual machine software and analyzed the code by tracking program behavior. After repeated research of the scholars' work simultaneously at home and abroad, hackers, in order to increase the survivability of the malicious code, also adopted anti-debugging techniques to check whether the code is being debugged. Therefore, in the context of malicious behavior, we still need some security experts to study and analyze the resultant data, but the judgment process will consume much time [2].

On the basis of the results of predecessors' research, this paper has conducted further research on the malicious code detection technology. It mainly focuses on the analysis of the malicious code, discusses the description method

[*] Corresponding author.
E-mail address: lqmnust@sina.com (Q. Li).

http://dx.doi.org/10.1016/j.ipl.2016.06.014
0020-0190/© 2016 Elsevier B.V. All rights reserved.

of malicious behavior, and applies the behavior analysis technology in virus detection model. This paper provides guidance for future research in this field [3].

2. Research on dynamic heuristic scanning technique

2.1. Dynamic heuristic scanning technique

Dynamic heuristic scanning technique is a behavior-based technique to monitor the running of a dynamic computer program and restrict the dynamic behavior of the computer. During the running of a program, some malicious and illegal procedures are often generated that are in conflict with the general procedures; these are intercepted and stemmed by the dynamic heuristic scanning technique.

2.1.1. Characteristics of illegal viruses

According to the analysis and identification of dynamic heuristic scanning technique, the illegal virus usually has the following characteristics:

(1) Illegal procedures invade the memory and modify the total system memory to remain concealed from the disk operating system (DOS).

(2) While switching between programs and viruses, the antivirus system executes commands before the host program.

(3) Boot viruses are carried by the guidance virus; the start-up and executive commands are obtained by the main Boot sector and the Boot sector. Simultaneously, the virus can occupy INT13 interrupt and set up the virus code before system setting is completed.

(4) The viruses can also invade certain files such as EXE and BAT files to tamper them. Because illegal procedures exhibit behavior characteristics, the dynamic heuristic scanning technique will test them [4].

2.1.2. Principle of dynamic heuristic scanning technique

Because the dynamic heuristic scanning technique can identify the malicious code, its study has become the hot spot in the research of computer antivirus software. This technique can closely monitor the operating system (OS) and preserve the normal operation of the system. When the system32 file of the OS creates abnormal problems, the network port traffic will increase and an unknown program will run in the computer. The dynamic heuristic scanning technique will find it by analyzing the software, and hence, it is widely used in the field of antivirus software.

2.1.3. Role of dynamic heuristic scanning technique

(1) The technique mainly scans to determine the internal behavior of unauthorized network and sends a warning to network users.

(2) It detects the user's network connection; if there is a wide range of broadband usage, it will result in the depredations of network resources. The dynamic heuristic scanning technique ensures the safety of users whose Internet usage is within a certain range by analyzing and identifying the unauthorized file sharing software.

(3) This technique can detect illegal procedures through unauthorized P2P applications, which can gain access to the external network news secretly and detect the network-cheating behavior of fly standard channel data transmission.

(4) The dynamic heuristic scanning technique can determine the potential danger program, for example, it can detect the attackers' use of illegal procedures or malformed packets. Furthermore, the dynamic heuristic scanning technique can make accurate judgment [5].

Table 1
ROC test index of the statistic.

No.	Display	Type
(A)	True	The virus detection model identifies the virus
(B)	False	The virus detection model identifies virus to be a legitimate program
(C)	True	The detection model identifies program to be legitimate legal procedures
(D)	False	The detection model identifies program to be a legitimate virus

3. Research on the detection model based on dynamic heuristic scanning technique

3.1. Establishment of the model

Detection index is the basis for determining the merits of the model test results. In this paper, test results are determined mainly through false negatives and false positives. The false negatives view legal program as malicious code. The false positives view malicious code as normal legal procedures [6].

Let N be the number of procedures needed for the detection, m be the malicious codes, and n be the legitimate programs, then the condition $N = m + n$ should be satisfied, provided all the three have positive values. If n normal programs consist of p false alarm malicious codes, the following conclusions can be obtained:

$$\text{False alarm rate} = (p/n) \times 100\% \tag{1}$$

If q denotes the normal legal process among m malicious code programs:

$$\text{Rate of missing report} = (q/m) \times 100\% \tag{2}$$

3.2. Detection using dynamic heuristic scanning technique

In this paper, we chose four ROC indices of medical statistics (Table 1), whose interrelationship is shown in Table 2. As shown in the table, the accuracy of detection of malicious code to become a true positive rate is given as

$$TPF = a/(a + c) = 100\% - \text{False negative rate} \tag{3}$$

Detection-qualified rate of false positives becomes the false positive rate: