International Journal of Security and Its Applications
Vol. 10, No. 5 (2016) pp.315-326
http://dx.doi.org/10.14257/ijsia.2016.10.5.30

Research on Intrusion Detection Systems and Unknown Malcode Detection based on Network Behavior

Xiaoyong YU*

School of Information Engineering, Suzhou University, Suzhou city, Anhui province, 234000, China

** yxy@ahszu.edu.cn*

Abstract

In all kinds of Internet security incidents, the most serious is malicious code. The increasingly serious problem caused by malicious code, not only make the enterprises and users suffered huge economic losses, but also makes network security facing serious threat. In this paper, based on the analysis of malicious code detection technology and detection system, the author designs and implements an unknown malicious code detection system based on network behavior analysis. Test results show that the detection system can distinguish three kinds of ARP attack; it can produce normal alarm information and achieve the desired results. At the same time, the network behavior analysis method needs to be further improved in order to achieve better analysis results, and provide more reliable results for the detection system.

Keywords*: Network behavior, malicious code, intrusion detection system, network security*

1. Introduction

Malicious code detection is security defense system relay firewall, data encryption and other traditional security measures to protect an important security technology, can attack of malicious code in the whole process of the system of real-time detecting and monitoring. Malicious code detection system in the process of malicious code attacks, timely warning, and minimize the damage caused by the loss, but also with the firewall and other defense system linkage, completely stop the attack behavior[1]. Constantly updated with the network scale continues to expand and the means of attack, malicious code detection technology also faces many challenges, how to effectively detect the unknown malicious code; how to improve the efficiency of detection of malicious code detection system, to adapt to network traffic increasing demand, how to reduce the malicious code detection system of false negative and false positive to improve the accuracy. Taking Internet as the representative of the information network technology application is more and more popular and widely, and the Internet is open, international and freedom in terms of increasing the degrees of freedom of the application at the same time, the security proposed higher requirements[2]. Open the network, resulting in network technology is open and any individuals, groups may derive the desired things, so network face the failure and attack may in many aspects. For example, attacks from the physical transmission lines can attack the network communication protocol and implementation attacks. A network of international also means that network attack not only from the local network of the user, it can from any machine on the Internet, that is to say, the network security faces is an international challenges. Freedom means that the initial use of the network did not provide any technical constraints; users can freely access the network, free to use and publish a variety of types of information[3]. Users are only responsible for their actions, without any legal restrictions. With the popularity of the

Internet, more and more organizations began to use Internet processing and transmission of sensitive data, at the same time on the Internet also spread everywhere and the spread of attacks and malicious code, making even into the Internet any system in the risk of attack will be. In various countries of the world, the security threat is increasing year by year, and the harm is also more and more big.

In all kinds of Internet security incidents, the most serious harm caused by malicious code, malicious code is becoming the major means of information warfare and network warfare, the increasingly serious threats of malicious code, not only the enterprises and users suffered huge economic losses, and the national security faced a serious threat. Malicious code is becoming more and more serious. These malicious code to the system install backdoors install Key logger, steal or damage the system of important data, tampering with government agencies and important information systems department, perform other malicious damage, in the user's knowledge, your computer is likely to have become some of the massive attack. Malicious code is so devastating, however, current of malicious code detection ability is still very limited, mainstream of malicious code detection system is often according to the malicious code has been the emergence of signature to detect the malicious code, such new malicious code if not being found in time and get the feature code may on the Internet is difficult to estimate the damage. Therefore, how to the unknown malicious code for effective detection is the focus of research at home and abroad, the goal of this study is through on network behavior analysis of network transmission of data, and reasonable classification of all kinds of network behavior, summarizes the network behavior of malicious code, through the network behavior of the unknown malicious code detection, and design and implementation of the unknown malicious code detection system based on network behavior analysis.

2. Literature Eeview

2.1. Malicious Code

Code Malicious, which is a program code against the security policy of the target system, causes the information leakage of the target system, the abuse of resources, the integrity and availability of the system. It can be transmitted through a storage medium or network, from a computer system to another computer system, without authorization to access or destroy the integrity of the computer system[4]. Usually, many people think that the "virus" represents all the programs that infect the computer and cause damage. In fact, the malicious code is more common, the virus is just a type of malicious code. Malicious code, including the computer virus, network worm, Trojan, etc. Non delegation and destruction are the two main features of malicious code. Computer viruses have the following characteristics:

1) ***Parasitic:*** a computer virus can be found in an executable program or Word document, which is started when the virus is executed. And before starting the program, it is just an ordinary file, not easy to be found[5]. This is the most basic feature of the virus.

2) ***Infectious:*** computer virus infectivity is that computer viruses can replicate, and the replication of the virus attached to other programs that are not infected, or more serious may be to replace disk boot record, which carry the virus program or a magnetic disc into a new source of the virus, which again is viral replication, repeat the original infection process. The most essential difference between computer viruses and other programs is that computer viruses can infect other programs. Computer viruses infect other programs by means of all the entities that can transmit computer information.

3) ***Destructive:*** the harm of computer virus to the system completely depends on the purpose of the virus. Some computer virus is just a trick, and some will destroy the system data. In a word, the consequences of the destruction of the virus are unpredictable. Because the computer virus is a malicious program, so for all the computer resources used in ordinary procedures, computer viruses are likely to be destroyed[6]. According to statistics, after virus attack, resulting in the destruction of mainly makes the user's partial loss of data, resulting in the remote control system cannot be used, malicious modification of configuration of the browser, cannot use the network or using a constrained, by attackers and. According to statistical analysis, the browser configuration is modified, data loss, the network cannot be used is the most common.

4) ***Latent:*** many computer viruses infected with normal computer, generally not immediately attack, but wait to meet the conditions set by some attackers before the implementation of the computer viruses, malicious functions, so as to achieve the purpose of destruction. The most common trigger condition for a computer virus is a specific date.

Despite the recent Internet happened more and more malicious code security incidents, and the harm is also growing, but not what new things and malicious code. In recent years, an attacker using the latest network technology has been in the malicious code to study attack ability and survival ability stronger. Malicious code development, making malicious code from very simple game of infection of virus to the complex operating system kernel virus and today active communication and destructive worms. The malicious code has achieved a great success in the rapid propagation mechanism and survivability technology research. Because of the network technology, malicious code first appeared on the network technology requirements higher, slower development, but with the rapid development of Internet, the Internet has become malicious code is released and the rapid spread of the platform, many hackers malicious code attack tools released to the network, making a lot of technical ability of the attacker is easily obtained strong attack tools, thus causing serious harm to the network. Especially in the past few years, the rapid growth of malicious code, confirmed this point. From the virus to e-mail worms, to take advantage of loopholes in the system take the initiative to attack by malicious code: malicious code to early destruction ability is not strong, route of transmission is limited, harm scope is not wide, the most aggressive behavior is by the virus and infected can be caused by the executable file. However, in recent years, the vulnerability of the system and network spread and infection on the network caused more and more harm.

2.2. Malicious Code Detection

Malicious code detection based on the computer network or computer systems in a number of key points to collect information and analysis, the discovery system network, or whether there is a breach of security strategy and the attacks were signs. The combination of hardware and software for the detection of malicious code is the malicious code detection system. Code detection is detection and recognition in computer systems and networks, or the broader sense of the information system of illegal attack or, in violation of the security policy[7]. It collects the data from the computer system or network environment, analyzes the data, finds the suspicious attack behavior or abnormal events, and takes certain measures to intercept the attack behavior and reduce the possible loss. Malicious code detection system is divided into various types according to the source of the original data, data analysis tools, the system architecture.

With the popularity of the network environment, a large number of malicious code detection system based on network. Based on network malicious code detection system using the raw network data packet as the data source, often using a running in promiscuous mode network adapter to real-time monitoring and analysis through all communication business network, and by pattern matching, statistical analysis, protocol analysis technology to detect malicious code behavior. As soon as the attack is detected, the response module of the malicious code detection system provides a variety of options to inform, alert and respond to the attack. Take response measures for different manufacturers of products and different, but usually include notify an administrator, interruption of the joint, and to collect evidence, records of malicious code attacks. The malicious code detection system based on network has many rely solely on based on host malicious code detection system cannot provide the function[8]. It has the following advantages: the network has a global view, through the view of the data flow to multiple hosts, malicious code detection system detectors can be obtained with the network view against our network attacks. If someone is scanning multiple hosts in our network, this information can easily be found by the detector. Less monitor, it only needs to run in the detector and controller platform on a limited number of. These platforms can be selected to meet specific performance requirements. In addition to cannot be a stranger in the monitored network, these devices can also easily be reinforced to prevent attacks by itself, because they only in the network perform specific tasks. Operating system independence, the network based malicious code detection system does not need to run on each operating system in the network.

With the emergence of a variety of advanced network equipment, network transmission speed is greatly improved, network traffic increases, the traditional centralized malicious code detection system often cannot keep up the improvement of network speed, which affect the testing result, and this malicious code detection system cannot solve the current relatively common DDOS attacks. Therefore, it is more and more important to collect and deal with the growing network traffic by using the distributed structure, and the research of the distributed malicious code detection system is becoming more and more important. The traditional malicious code detection system is generally restricted to a single host or network architecture, so traditional malicious code detection system has not fully protected the network security. Therefore, it is necessary sufficient cooperation between various malicious code detection systems, even many kinds of network security products combined with each other, using all kinds of network security products of expertise to against malicious code attacks were detected, so as to achieve the purpose of protecting network security. So, it need to solve the problem is to establish a unified standard for the detection of malicious code, technology relating to study all kinds of network security products, improve the accuracy of detecting malicious code, network maintenance of normal use. With malicious code attacks kinds of diversification and malicious code number to grow exponentially, it is necessary and important to study the new detection algorithm to improve the detection efficiency. At present, neural network, immune, genetic algorithm of machine learning algorithm has preliminary results, but these are just the research work, also need to further study to solve the self-learning and adaptive ability, so as to achieve the practical use of malicious code detection system in the target.

3. Network Behavior Analysis

3.1. Network Infrastructure

Many different manufacturers produce various models of computers, and the operating systems they run may be quite different, but the TCP/IP protocol family allows them to communicate with each other. TCP/IP originated in the late 60's the United States government funded a packet switched network research project, to the 90's has been developed into the most common form of networking between computers. It is a real open system, because the definition of the protocol family and its various implementations can be obtained without money or a little money. It becomes the basis of being called global

Internet or Internet. Network protocol is usually divided into different levels of development; each layer is responsible for different communication functions. TCP/IP protocol family is a set of different levels of the combination of a number of protocols. TCP/IP is generally considered to be a four layer protocol system

1) ***Link layer:*** sometimes called a data link layer or network interface layer, usually includes a device driver in the operating system and a corresponding network interface card in the computer. They deal with the physical interface details of the cable (or any other transmission medium).

2) ***Network layer:*** sometimes also called the Internet layer, to deal with the activities of the group in the network, such as the selection of the road and so on. In the TCP/IP protocol family, the network layer protocol includes IP protocol (Internet Protocol), ICMP protocol, and IGMP protocol (Internet group management protocol).

3) ***Transport layer:*** providing end to end communication for applications on two host computers. In the TCP/IP protocol family, there are two different transmission protocols: TCP (transmission control protocol) and UDP (user data protocol). TCP provides high reliability data communication for two hosts. The work done by it includes the application of the program to its data into the appropriate small pieces to the following network layer, to confirm the receipt of the packet, set to send the final confirmation packet timeout clock and so on. The application layer can ignore all these details because of the high reliability of the end to end communication provided by the transport layer.

4) ***Application layer:*** responsible for handling specific application details. Almost all kinds of different TCP / IP implementation will provide these generic applications: Hypertext Transfer Protocol HTTP; telnet remote login; FTP file transfer protocol; SMTP (Simple Mail Transfer Protocol; SNMP (simple network management protocol.

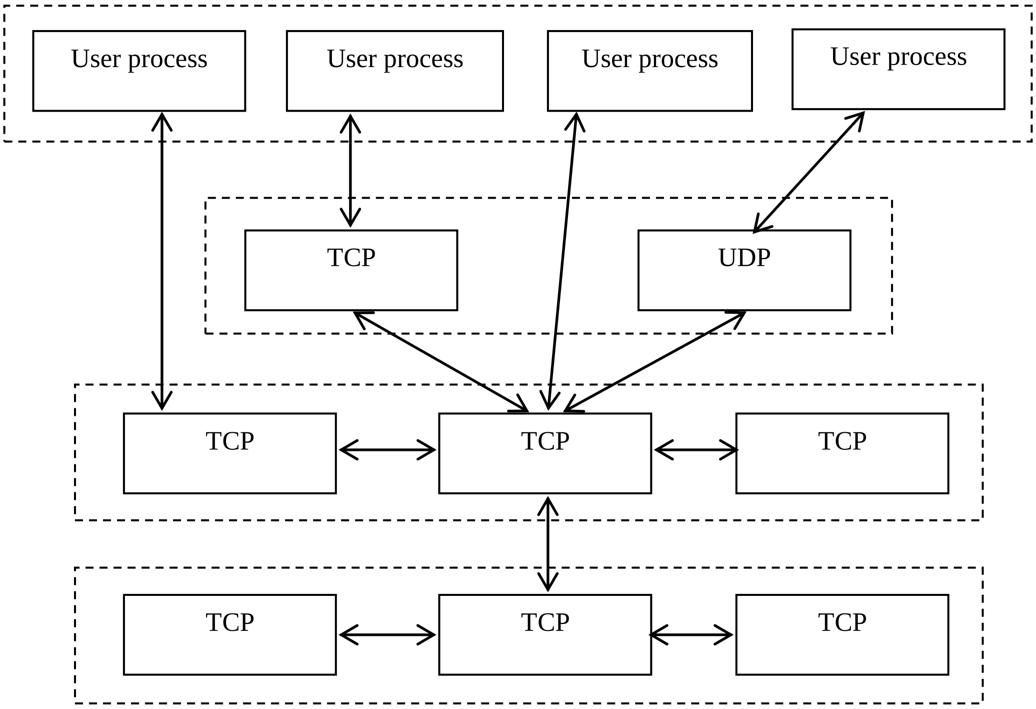

Figure 1. Different Levels of Protocols in TCP/IP Protocol

In the TCP/IP protocol family, there are many kinds of protocols. In Figure 1, some important protocols of different levels of TCP/IP protocol are given. In order to analyze the data in the network, it is necessary to analyze the data packets transmitted in the network according to different protocols.

3.2. Protocol Analysis

ARP protocol is an IP address to the corresponding hardware address to provide dynamic mapping. We use the word is dynamic because the process is automated, general application users or system administrators do not have to care about. When a computer network in the Ethernet data frame is sent to another computer on the same LAN, is to determine the destination interface according to the 48 bit Ethernet address. The device driver is not going to check the IP datagram to IP address. Address resolution provides a map for the two different forms of address: address any type of IP address and data link layer using 32bit. In recent years, ARP spoofing Trojan is very common in the network, caused a great loss to the normal network users. ARP Trojan poisoning phenomenon in general: LAN will suddenly dropped, after a period of time will automatically return to normal. For example, the client state frequently turned red, user's frequently broken network, IE browser error frequently, and some commonly used software always appear fault. If the LAN is access to the Internet through authentication, suddenly appeared while certification, but is unable to access the Internet phenomenon, restart the computer or run under MS-DOS window command ARP - D refresh ARP cache later, they can restore the Internet. ARP deception attacks only need to successfully infect a computer, it may cause the entire local area network cannot be online, serious or even cause the entire network paralysis. The attack in addition to lead to other users in the same LAN Internet intermittent phenomenon, it also will steal the user's password. For example purloin QQ software password and steal all kinds of online game password and account to do money transactions, even steal online banking account to do illegal trading activities etc.. This is the modus operandi of the Trojan, often cause great inconvenience and huge economic losses to the user. ARP attack is through forged IP address and MAC address of ARP spoofing, can in the network have a large number of ARP traffic congestion on the network to achieve the purpose, the attacker lasted for as long as the continuous issued forged ARP response packet will be able to change the target computer ARP buffer storage IP-MAC entry, resulting in network outages or man in the middle attack. ARP attack is mainly present in the local area network, local area network (LAN) if a person is infected with ARP Trojan, infected the ARP Trojan system will be trying to through "ARP spoofing" means to capture in the network where the other computer communication information, and thus caused the other computers in the net communication failures.

3.3. Main Methods of Network Behavior Analysis

Network behavior analysis method and operation system is relatively independent, strong versatility, regardless of is the windows operating system is Linux operating system in the network transmission of data package is based on TCP / IP protocol, so the for protocol analysis as the foundation of the network behavior analysis method were detected can be for all TCP / IP based operating system achieve the purpose of detecting malicious code. It can achieve the detection of the previously did not have the effect of malicious code, and the traditional signature based detection is only on the already appeared malicious code to check. However, it is difficult to establish a general behavior analysis rule, because the network behavior of different users is different, so the result is very high. The network behavior analysis method mainly has the following several.

1) ***Probability statistics method:*** Based on the probability statistics method is the first method in the network behavior analysis method, and it is also one of the main technologies in the malicious code detection method. Network users within a certain time exhibit certain statistical regularity, users of statistics of short-term or long-term activities, detection system calculated statistical information of all network users, and pre designed rules of conduct were compared, if the behavior does not conform to the rules set, and computing statistical information obtained result exceeds a specified threshold is considered abnormal.

2) ***Machine learning method:*** Based on machine learning method through machine learning to achieve network behavior analysis, the user network behavior can be attributed to the training samples to learn the behavior characteristics of users, systems and networks. The main methods of study include experiential induction learning, analysis of learning, analogical learning, genetic algorithm, etc.. Empirical induction learning uses some data intensive empirical methods to study the examples. The examples and the results of the study are usually represented by the symbolic representation of the attribute, predicate, relation and so on. It is equivalent to the inductive learning based on the classification of learning strategies, but the part of deduction of join learning, genetic algorithm and reinforcement learning.

3) ***Data mining:*** data mining is from a lot of, completely, noisy, fuzzy, random data extraction implicit in which people do not know in advance, but is potentially useful information and knowledge process. The introduction of data mining technology, applied to the malicious code detection system, completes the process of automatically extracted from a large number of data. In the process of establishing the attack detection system, it can eliminate the artificial factors and specific factors, which can develop a set of automatic tools to generate the attack detection model from various audit data. By using correlation analysis and sequential pattern analysis, we find the relation between the features and the time sequence, so as to complete the collection process of the user's network behavior information data.

4) ***Neural network:*** neural network as an effective method to deal with nonlinear systems, it has been successfully applied in the field of function approximation, classification, pattern recognition and probability density estimation. The anomaly detection technology of malicious code detection system is essentially a pattern recognition or classification problem. Therefore, the application of neural network technology to malicious code detection system has become a hot research topic in this year. Compared with the statistical theory, the malicious code detection method based on neural network can better express the non-linear relationship between variables, and can automatically learn and update.

4. System Design and Implementation

4.1. Malicious Code Detection System

The unknown malicious code detection system based on network behavior analysis is through the network behavior analysis of the data in the network; the system is located on the protected network data export in order to obtain all the data in the network transmission, figure 2 is the system in the network operating environment.

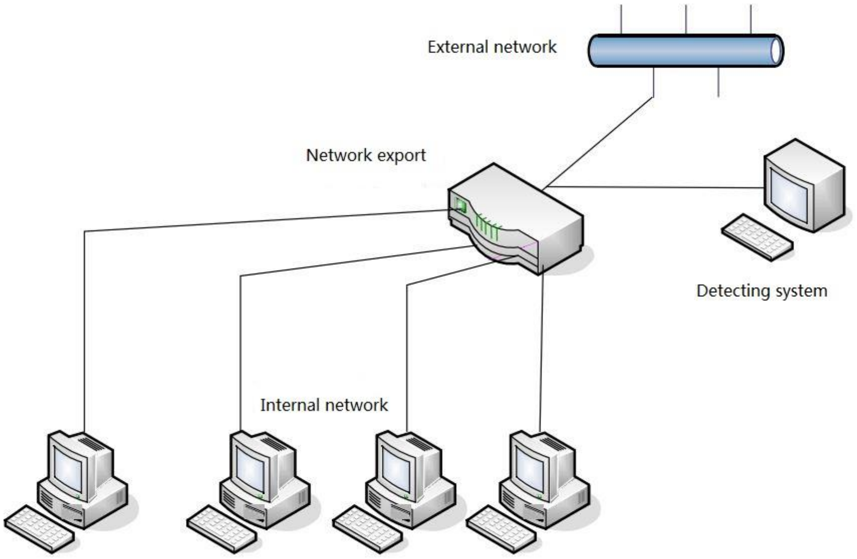

Figure 2. Operation Environment of Unknown Malicious Code Detection System

Based on the network behavior analysis of unknown malicious code detection system is located in between internal network and external network, for all flows through export network data collection and analysis, the Fedora 8 operating system, using Libpcap software to obtain the network transmission of data packet. Packet capture module main function is using Libpcap software to capture data packets, the Libpcap software installation to detection system where the host can obtain all through export network data packets, to obtain data packets using thread pool technology and memory zero copy technology in order to avoid data packet loss phenomenon, in the acquisition of data packet, the data packet information is transmitted to the protocol analysis and information statistics module for processing.

Network behavior analysis and detection module is the core module of the network behavior analysis of unknown malicious code detection system based on and the main function of this module is to network transmission of data information analysis and detection, malicious code behavior is in the module is detected and sent to the alarm module for processing, network behavior analysis and detection module specific analysis and detection module and ARP attack behavior analysis module, DDOS attack behavior analysis module, Trojan horse attack behavior analysis module eight modules.

(1) Data acquisition module

(2) Protocol analysis and information statistics module

(3) Network behavior analysis and detection module

Subdivided into eight modules:

<1> ARP attack behavior analysis module

<2> DDOS attack behavior analysis module

<3> Trojan attack behavior analysis module

<4> IP protocol behavioral analysis module

<5> TCP protocol behavior analysis module

<6> UDP protocol behavioral analysis module

<7> Application layer protocol behavioral analysis module

<8> specific user behavior analysis module

(4) alarm module

The behavior of the Trojan horse attack analysis module is the main function is used to detect the Trojan horse attack, after access to the TCP protocol and UDP protocol data to the entire internal network of all hosts for flow rate, time and scale of the behavior analysis, Trojan main intention is to to the secret to steal information, so performed above all hosts on the internal network behavior analysis, can be a very good detection out which hosts a large amount of information is stolen, the information flow duration. If detected in the internal network there are some of the host's behavior in line with the behavior of the Trojan horse attack immediately do alarm processing, as shown in Figure 3.

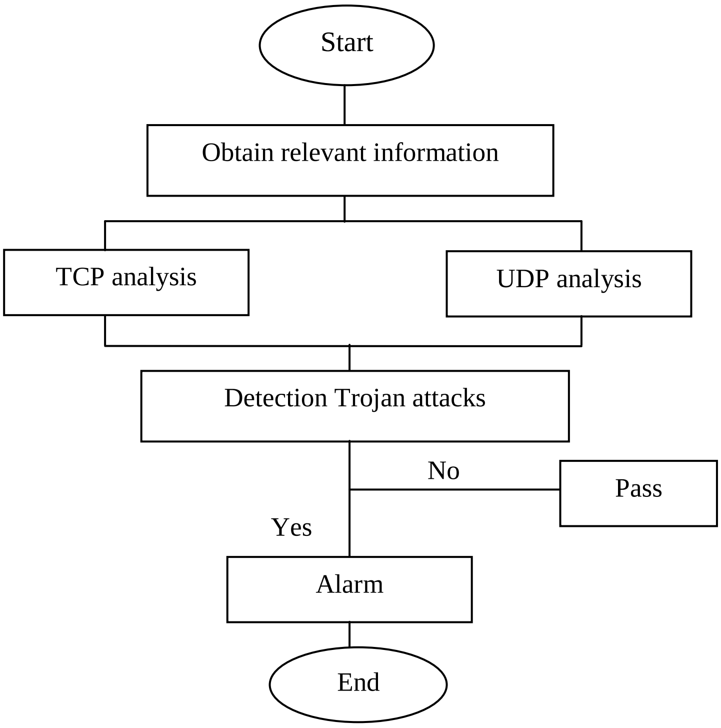

Figure 3.Trojan Attack Behavior Analysis Module Design

4.2. Test Result

The use of ARP attack software WinArpAttacker on all the hosts in the internal network of attack: the ARP attack software WinArpAttacker installed on a host of internal attack 202.115.27.33. Of all the hosts in the internal network of ARP attack, respectively, using IP address spoofing, IP address conflict and MAC address spoofing method sends an ARP attack data packets to all hosts in the network. The hosts in the internal network are 3 kinds of ARP attacks, IP address deception, IP address conflict, MAC address to verify the ARP address of the system to verify the function of the analysis module. testing procedure as:

1 set up test environment

2 using the ARP attack software IP on the internal attack host 202.115.27.33 on the internal network to carry out WinArpAttacker address to deceive the attack

3 using the ARP attack software IP on the internal attack host 202.115.27.33 on the internal network WinArpAttacker address conflict attacks

4 using the ARP attack software MAC on the internal attack host 202.115.27.33 on the internal network to carry out WinArpAttacker address to deceive the attack

5 observation and detection status

Test results of unknown malicious code detection system of 3 kinds of ARP attack were issued a warning, in all the ARP attack detection, based on network behavior analysis of unknown malicious code detection system were detected in the ARP attack and normal alarm information, to achieve the desired effect. Using the SYN Flood attack, Ping flood attack, UDP flood attack, a land attack, Ping of death attack and teardrop attacks the six kinds of DDoS attack methods from external network attack the host to all the hosts in the network attack. In the detection of these six attacks, the unknown malicious code detection system based on the analysis of the network behavior has detected the attack, and generated the normal alarm information, achieved the expected effect. After the network behavior rules of a particular user is set up, the unknown malicious code detection system based on network behavior analysis can effectively detect does not conform to the specific user network behavior rules of communication behavior, and the normal work of the alarm information, to achieve the desired effect. Based on network behavior analysis of unknown malicious code detection system in the 100m network normal operation, packets in the data transmission rate from 10Mbps up to 100Mbps does not appear phenomenon of packet loss, so as to ensure the integrity of data detection, provided the guarantee for the correctness of the test results, thus validating the good performance of the system.

5. Conclusions

Computer malicious code origin has been exist for long time, prevention of malicious code have also made a lot of research, but with the development of information technology, malicious code attacks methods and hidden methods are constantly changing. In particular, the popularity of Internet, the widespread existence of system vulnerabilities and the strong demand for information sharing provides an unprecedented convenience for the spread of malicious code. The form of a variety of malicious code, the number of malicious code exponentially growing, to enterprises and institutions, government agencies and individuals have brought great losses. Therefore, it is significant to study the detection of malicious code in the new form. In this paper, based on the analysis of the principle of malicious code attacks and malicious code detection technology and detection system, we design and implement an unknown malicious code detection system based on network behavior analysis. Study of malicious code detection system definition, classification, detection technology and the direction of development of the, for a variety of malicious code detection system classification do detailed introduction, and a variety of malicious code detection system of the main working principle, advantages and disadvantages were described in detail. This paper studies the main malicious code detection technology, introduces the principle and the advantages and disadvantages of the main detection technology, and summarizes the development direction of the malicious code detection system. In this paper, the design and Realization of unknown malicious code detection system based on network behavior analysis, given the detection system running environment, structure design, module design and each module of the detection system achieve the function flow and has carried on the detailed elaboration.

Network behavior analysis method also needs to be further improved, in order to achieve better analysis results, to provide more reliable data analysis results for the detection system.

References

[1] K.Kumar, "Securing communication using function extraction technology for malicious code behavior analysis",Computers and Security, Vol.28, **(2009)**, pp.77-84.

[2] A.Fiskiran,L.Ruby, "Runtime Execution Monitoring (REM) to detect and prevent malicious code execution",IEEE International Conference on Computer Design , Vol.4, **(2014)**,pp.452-457.

[3] T.Wan-Hui, F.Chin-Feng, "Handling malicious code on control systems",Lecture Notes in Computer Science. Vol.4, **(2007)**, pp.68-74.

[4] W.Wei-Ping, Q.Si-Han, "Mechanism and defense on malicious code", Wuhan University Journal of Natural Sciences,Vol.10, **(2005)**, pp.83-88.

[5] E.Yuval, S.Asaf, "Applying machine learning techniques for detection of malicious code in network traffic",Lecture Notes in Computer Science, Vol.6, **(2007)**, pp.44-50.

[6] M.Christodorescu, "Testing malware detectors", International Symposium on Software Testing and Analysis,Vol.3, **(2014)**,pp.25-28.

[7] L.Ming, C.Xianyi, "Intelligent Agent distributed intrusion detection system based on improved BP neural network",Computer applications and software,Vol.1, **(2014)**,pp.105-107.

[8] Z.Sihai, H.Shen, "Statistical analysis method of network script virus",Chinese Journal of computers, ,Vol.29, **(**2**015)**,pp.969-975.

Author

Xiaoyong Yu, 1979.8,Suzhou, Anhui, China

Current position, grades: the lecturer of Computer Science, Suzhou University,Anhui,China.

Scientific interest: His research interestfields include Computer architecture and Network security.

Publications: more than 5 papers published

Experience: He has teaching experience of 12years, has completed one scientific research projects.