Research on Network Scanning Strategy Based on Information Granularity

Futong Qin¹, Pengfei Shi², Jing Du³, Ruosi Cheng⁴ and Yunyan Zhou⁵
Luoyang Electronic Equipment Test Center of China, Luoyang, China
Email: †ftqin@163.com, ‡spf123000177010@163.com, §jdstarry@aliyun.com,
¶lincoln_cheng@foxmail.com, ‡‡zzyy@sina.com

Abstract. As the basic mean to obtain the information of the targets network, network scanning is often used to discover the security risks and vulnerabilities existing on the network. However, with the development of network technology, the scale of network is more and more large, and the network scanning efficiency put forward higher requirements. In this paper, the concept of network scanning information granularity is proposed, and the design method of network scanning strategy based on information granularity is proposed. Based on single information granularity and hybrid information granularity, four network scanning strategies were designed and verified experimentally. Experiments show that the network scanning strategies based on hybrid information granularity can improve the efficiency of network scanning.

1. Introduction

With the continuous development of computer and network technology, the computer network technology is widely used in people's lives and work in all aspects. However, the hidden dangers of the network itself and the flaws of the network communication protocols make the network vulnerable to various types of attacks [1][2]. Moreover, the openness of network architecture, the sharing of resources, the diversity of access methods, the irregularity of terminal distribution and the unpredictability of network boundary make the network technology to provide people with many conveniences but also bring many problems and hidden dangers.

Study the security of computer network, especially how to improve the security of computer network must first know both the enemy and us. Network scanning is an important way to understand the status of computers in network security research. Through network scanning, it can obtain some characteristics of the computer on the network, such as the operating system running on the computer, the vulnerability information in the system and other fingerprint information, which is very valuable for network security researchers. Therefore, network scanning plays an important role in network security research.

Network scanning as the basic means to obtain system and network information, the size of the target scan and the level of information they want to obtain, largely determines the time required to scan the network. If the scanning strategy is configured improperly, it may result in a small amount of information but a long scanning time, greatly reducing the scanning efficiency. Common network scanning strategies generally include random scanning, target list-based scanning [3], routing scanning [4] and partition scanning [5] and so on. Random scanning is randomly select computers on the network to scan. Target list-based scanning first collect a list of potentially vulnerable host addresses before scanning, and then scan the hosts in a specific target list. Routing scanning using BGP route prefix to reduce the scan space and selectively scan the IP address space. Partition scanning makes the target list
divided into several sub-lists before scanning, and then conducts multiple scanning tasks according to the sub-list at the same time. All of these scanning strategies are to improve the efficiency of network scanning from the point of reducing the network scanning IP address space, and the influence of information granularity on scanning efficiency is not considered.

The purpose of this paper is to analyze the network scanning principle and define the granularity of network scanning information, so as to discuss the method of designing efficient network scanning strategy, and to provide guidance for computer network scanning. Aiming at the problem of network scanning efficiency, we first propose a network scanning strategy design method based on information granularity. The experimental results show that this method can significantly improve the network scanning efficiency.

2. Network Scanning Technology
Network scanning is a kind of information system security protection means, it uses the method of simulated hacker attack to detect and evaluate the security of the target system. The basic principle of network scanning is through a series of means to simulate the attack behavior of the system and get the attack results. Based on these, analyzing the results so as to understand the security configurations and the application services in the target, analyze the existing security vulnerabilities, assess the network risk level objectively, guide the network administrator corrects the network security vulnerabilities and system errors in the configuration according to the results, make prevention before the launch of hacker attack.

Network scanning technology can be divided into three categories according to a complete scanning process. The first is target survivability scanning technology, that is to determine whether the target exists in the network, and in the on-line state. The second is target information collection technology, including detection of the target host open ports, the target operating system type, open services, etc. The third is vulnerability scanning technology, according to the information collected to determine or further test the system for the existence of security risks.

2.1 Target Survivability Scanning Technology
Target survivability scanning is to discover whether the target host survives, the basic principle is to use packets generated by a variety of network protocol and the inherent nature of the network protocol itself for scanning. A commonly used scanning method is PING scanning.

PING scanning usually adopts methods that send various types of ICMP or TCP, UDP request message, through the results of the message sending to determine whether the target is alive. Several major approaches include ICMP broadcast, ICMP sweep, ICMP non-echo, TCP and UDP sweep. The ICMP broadcast speed is fast but cannot find the Windows system host, ICMP sweep is easy to use, but slower. Compared with the detection of ICMP protocol, TCP sweep is more effective. It uses the famous "three-way handshake" to scan, in the handshake process to send SYN packets to the target, regardless of the target host is to respond to SYN/ACK message or RST message, you can determine the target host survival. It is now the most effective method of target discovery.

2.2 Target information scanning technology
Target information include the information of the target host port open, operating system type and open information service, etc. The main technologies of target information scanning are port scanning technology, operating system identification technology and system service identification technology.

2.2.1 Port scanning technology. Port scanning [6] is mainly divided into full connection, half connection and secret scanning, etc.

Full Connection Scanning: Full connection scanning is the most basic TCP scanning method, also known as the "TCP connect scanning", which principle is to use the connect() system call provided by the operating system to connect to a specific TCP port on the target host. If the target port is listening, connect() can connect successfully, else the connection will be prompted to fail. The advantage is that any user in the system can call the connect() function without any special permissions, and scanning speed. The disadvantage is easy to be detected and filtered by the security software, but also in the
target host's log file will produce a series of error message about the service connection is established and immediately disconnect.

a) Half Connection Scanning: Half connection scanning also known as "TCP SYN scanning". TCP SYN scanning sends a SYN packet, if return to the SYN/ACK packet said target port in the listening state, if return to RST packet indicates the port is not open. If a SYN/ACK packet is received, the scanner then sends an RST packet to terminate the connection. This method has the advantage of being covert because it does not need to establish a complete TCP connection, even if the log has been recorded on the scanning, but the record on try to connect is much less than full scanning. The disadvantage is that in most operating systems, the sending host needs to construct the IP packets, this is too much trouble, and constructing SYN packets needs super or authorized users to access the special system call.

b) Secret Scanning: A typical secret scanning technique is the TCP FIN scanning. TCP FIN scanning is subtler than TCP SYN scanning, its packets can through the firewall and packet filtering systems without any hassle. TCP FIN scanning method is based on the principle that the port that is not opened will give RST response to the FIN packet, but the open port will ignore the FIN packet directly, do not answer. This method is associated with the type of the target system, some system reply RST regardless of whether the port is opened or not, at this point, this scanning method can be used to distinguish the target system is a UNIX or Windows.

2.2.2 Operating System Identification Technology. Each operating system, or even each kernel revision, has small differences in the TCP/IP stack, and these differences make the response to the corresponding packet not the same. Some network scanning tools can provide a response list, by comparing with it, the operating system running on the target host can be identified. Operating system identification technology mainly contains active identification and passive identification.

a) Active identification: Active contract, many times test, screening of different information, for example, according to the returned ACK value, some systems will return the sequence number of the acknowledged TCP packet, while others return the sequence number plus 1. Some operating systems use some fixed TCP window or set the DF header of the IP header to improve performance, which is the basis for judging the corresponding operating system. Active identification technology to determine the accuracy of Windows is relatively poor, can only determine the approximate range, it is difficult to determine the exact version, but UNIX, CISCO network equipment more accurate. However, the more hops between the target host and the source host, the worse the accuracy. This is mainly because many of the eigenvalues in the data packet have been modified or blurred in the transmission process.

c) Passive identification: Passively monitoring the network data packets, using the data in the packet header of DF, TOS, window size and the TTL value to determine the operating system. Because this method does not need to actively send detection packets, only need to crawl the network packets, so called passive identification.

2.2.3 System Service Identification Technology. The detection of system service is mainly based on the scanned port to determine, or the use of HTTP response analysis, binary information detection and other means to achieve.

According to the port to identify system services is to directly use the port and the corresponding services relationship, this approach is an earlier way, for a wide range of assessment, it has some value, but its accuracy is low, it often produces injustice in actual use. For example, it thinks that as long as open port 80 is to open the HTTP protocol. However, this is not the case, this is the fundamental flaws of using port scan technology in the service decision.

Relatively speaking, the binary information detection method is relatively accurate, access to the service banner, is a more mature technology, can be used to determine the current operation of the service, the judgments are more accurate. Moreover, it not only can determine the service, but also can determine the specific version of the service, but it is easy to be cheated by banner camouflage methods and produce false positives.
2.3 Vulnerability Scanning Technology
Vulnerability scanning technology is a kind of security technology developed to make system administrators keep abreast of the security vulnerabilities in the system and take corresponding precautions to reduce the security risk of the system. Using vulnerability scanning technology, you can scan the security vulnerabilities of the LAN, Web sites, host operating systems, system services, and firewall systems, etc.

Vulnerability scanning can be mainly divided into passive and active scanning according to the scanning strategy. Passive scanning based on the hosts, the general system architecture adopts C/S mode, through installing agents in the hosts to collect vulnerability information. There are certain limitations in this mode, which are caused by the software compatibility problems. Active scanning is network-based, sending specific packets through the network to detect the existence of security vulnerabilities of the target system, this vulnerability scanning system more widely applicable, the use of such scanning system is more common especially in internet companies.

3. Design Method of Network Scanning Strategy Based on Information Granularity

3.1 The Definition of Information Granularity

3.1.1 Definition 1: Information Granularity. Information granularity refers to the extent to which the information obtained by network scanning can clearly reflect the security problem of the target system. The smaller the granularity of information, the more clearly the security problem of the target system.

According to the degree of detail of information obtained by network scanning, we defined four information granularity levels, as shown in figure 1.

![Figure 1. The levels of information granularity of network](image)

3.1.2 Definition 2: Target Survivability Information Level. The information obtained by network scanning is the survivability of the hosts in the target network, and the information granularity is the largest, which can be used to statistic the distribution of surviving hosts in the network. Using the same scanning tool, the time spent in this granularity scan is minimal.

3.1.3 Definition 3: Port Open Information Level. The information obtained by the network scanning is the open ports of the target hosts. The information granularity is smaller than the target survivability level. It is based on the target survivability detection and further obtains the host open port information.

3.1.4 Definition 4: Service Information Level. The information obtained by network scanning is information such as open service, application version and host operating system. The granularity of the information is smaller than the port open level. The port open level scanning can only obtain the status of the port open, while the service information scanning can determine whether the service corresponding to the port is the default service that the port has, and can obtain information such as the application version of the service and the target host operating system.

3.1.5 Definition 5: Vulnerability Information Level. The information obtained by the network scanning is the vulnerability information of the target host, and the information granularity is the smallest. Vulnerability information level scanning results can be directly used by the attacker to attack the target host, or as a "security scanning", can directly find the system security option setting error, in order to improve system security.
3.2 Design of Network Scanning Strategy Based on Hybrid Information Granularity

Network attack and defense has always been a game problem. For the attacker, the sooner the problems found in the network, the more able to break through the target system. For the defender, the sooner the problems found in the network, the more timely to prevent possible attacks.

By the second section of this paper on the analysis of network scanning, it is not difficult to find that the larger the scanning network scanning information granularity is, the less data is needed to judge the scanning result. The smaller the network scanning information granularity is, the more data the scanning result needs to be obtained. If the network scanning information granularity is the target surviving level, the scanning host only needs to send the SYN message to the target. The feedback information obtained is that the target host responds to the SYN/ACK or RST message. Regardless of the message information, the time is only the sum of the sending, feedback and normal network transmission times. If the network scan information granularity is the service information level, such as obtaining the system service information, regardless of whether the target host is alive, the scanning host will send detection information to it and wait for the detection target feedback result, and then judge the target system service information according to the feedback result. The time of the whole process takes is the sum of the time to send probe packets, feedback packets, packet analysis and network delay. If the target host survival, the network delay is the normal network delay, if the target host is not alive, the network delay consists of the normal network delay and the timeout period of the host to wait for detection of target feedback. From the above analysis, we can see that the smaller the granularity of network scanning information, the more complex the analysis and judgment of the scanning results, the more time is needed, especially when the target host is not alive.

When designing a network scanning strategy, in the same size of the scanning IP address space conditions, if we first using the coarse-grained scanning to reduce the IP address space, and then using the fine-grained scanning, these will greatly reduce the proportion of network overtime in the whole scanning time-consuming, will certainly be able to improve network scanning efficiency. In view of this, this paper designs the following four network scanning strategies.

C1: For the scan target IP address list A, performing L1 target survivability scanning.
C2: For the scan target IP address list A, performing L2 port open scanning.
C3: For the scan target IP address list A, performing L3 service information scanning.
C4: For the scan target IP address list A, first performing L1 target survivability scanning to obtain the survive hosts’ IP addresses, then make these IP addresses into a new list A1, and performing L3 service information scanning on A1.

Among them, scanning strategy C1, C2 and C3 are strategies based on single information granularity, C4 is a strategy based on hybrid information granularity.

4. Experiment Analysis

In this section, we will analyze the four C1–C4 network scanning strategies to verify the effectiveness of the network scanning strategy based on information granularity in improving the scanning efficiency. The target network is built in laboratory environment.

4.1 Experimental Setup

The target network consists of four subnets. Each subnet host sets the IP addresses to the address segments as in table 1.

| Subnet   | IP address segments |
|----------|---------------------|
| Subnet1  | 192.168.11.0/24     |
| Subnet2  | 192.168.12.0/24     |
| Subnet3  | 192.168.13.0/24     |
| Subnet4  | 192.168.14.0/24     |
Each subnet has five hosts, and each host is deployed the same FTP and Web services with the ports 20 and 80 are opened. Set up a scanning host connected with each subnet, and deploy the network scanning tool Nmap [8] in the scanning host. The specific experimental network topology as shown in Figure 2. In the experiment, we set up two groups of scanning targets. The group 1 includes two subnets, subnet 1 and subnet 2, the group 2 includes all four subnets in the target network.

![Figure 2. The specific experimental network topology](image)

### 4.2 Experimental results and analysis

Set the scanning host scanning the two groups of targets respectively in accordance with the scanning strategies C1 - C4. For each group of targets, five scans were performed using each scanning strategy. The time spent for each scanning was recorded separately. The average of five scanning times was taken as the experimental result. All the results were shown in table 2.

| Scanning Target | C1   | C2   | C3   | C4   |
|-----------------|------|------|------|------|
| group 1         | 25.886 | 72.124 | 137.744 | 56.306 |
| group 2         | 112.608 | 200.778 | 223.206 | 129.188 |

![Figure 3. The influence of IP address space on scanning time](image)
In order to show the influence of scanning target IP address space size and scanning information granularity on scanning time, we analyze the contents in table 1 in graphical form. Figure 3 shows the influence of IP address space on the network scanning time. It can be found that scanning the four subnets (group 2) takes more time than scanning two subnets (group 1) when using the same scanning strategy. In addition, it can be seen from figure 3 that using these three strategies C1, C2 and C3 to scan the same target consuming more and more time since the three strategies are based on the single information granularity and their information granularity are more and more smaller. Figure 4 shows the influence of scanning strategies on scanning time. In the experiment, scanning strategies C3 and C4 were used to get the service information of the target. C3 is scanning the target directly, and C4 is firstly using coarse-grained granularity scanning to scan the target, removing non-active targets to reduce scanning IP address space, and then using fine-grained granularity scanning to scan the new reduced IP address space. It can be seen from figure 4 that the time consumed by strategy C4 is significantly less than the time consumed by strategy C3, whether scanning two subnets or scanning four subnets. Thus, we can find that designing scanning strategies based on the hybrid information granularity can effectively improve the network scanning efficiency.

5. Conclusion
Network scanning often occurs when cybersecurity personnel conduct a census of network assets, or in the preparatory stage for hackers to conduct network attacks. Since the target network has a large IP address space, various types of equipment, and their ports, services, vulnerabilities vary widely, the network scanning often takes a lot of time when conducting assets census on a large-scale network. Based on the analysis of network scanning principle, in this paper, we defined the information granularity of network scanning, and designed the network scanning strategies based on the information granularity. The experiment proved that using hybrid information granularity design network scanning strategies could effectively improve the efficiency of network scanning.

6. Acknowledgment
In this paper, the research was sponsored by the National Natural Science Foundation of China with Grant No. 61303061 and the Open fund from HPCL with Grant No. 201513-01.

7. References
[1] Li Xiaoyan, Wang Qingxian, Yang Lin, Luo Xiangyang. The Research on Network Security Visualization Key Technology[C]. 2012 fourth International Conference on Multimedia Information Networking and Security (MINES), 2012:983-988.
[2] Fang Lan, Wang Chunlei, Ma Guoqing. A framework for network security situation awareness based on knowledge discovery [C]. 2010 2nd International Conference on Computer Engineering and Technology (ICCET), 2010:226-231.
[3] Weaver N. Potential Strategies for High Speed Active Worms: A Worst Case Analysis [EB/OL]. http://www.cs.berkeley.edu/~nweaver/worms.pdf, 2002.

[4] Zou CC, Towsley D, Gong W, et al. Routing worm: A fast selective attack worm based on IP address information [A]. In Proc. of 19th ACM/IEEE/SCS Workshop on Principles of Advanced and Distributed Simulation (PAD/05) [C].2005, 199~206.

[5] Zou CC, Towsley D, Gong W. On the performance of Internet Worm Scanning Strategies [J]. Performance Evaluation, 2006, 63(7):700~723.

[6] Fyodor. The Art of Port Scanning [EB/OL]. http://www.insecure.org/nmap/nmap_doc.html, 2002.

[7] Fyodor. Remote OS Detection Via TCP/IP Stack Fingerprinting [EB/OL]. www.insecure.org, 2002.

[8] Nmap[EB/OL], http://nmap.org/download.html.