Port Scan Detection

Jayant Gadge and Anish Anand Patil

Abstract- **Port scanning is a phase in footprinting and scanning; this comes in reconnaissance which is considered as the first stage of a computer attack. Port scanning aims at finding open ports in a system. These open ports are exploited by attackers to carry out attacks and exploits. There are a number of tools to scan for open ports. However, very few tools are present to detect port scanning attempts.**

The goal of this project is to identify port scan attempts and find out information about the machine from where port scan attempts were made. If an attack takes place after the port scan, the collected information would help in bringing the criminal to justice. We hope that this work will add an additional layer of defense by identifying port scan attempts thereby indicating that an attack may follow.

I. INTRODUCTION

With an increase in computer literacy people are becoming aware about the loopholes present in the Operating Systems, networking protocols, software applications which are used on a daily basis. Many easy-to-use tools are freely available on the Internet which can take advantage of these loopholes to gain unauthorized access to a system. To further complicate the things most of us do not follow good security practices, making the job of computer criminals even easier.

Computer crimes have increased over the years. They are not limited to trivial acts such as guessing the login password of a system, they are much more dangerous. Studies indicate that the first stage of an attack is reconnaissance [5]. In this stage the prime objective is to get information about the target system. One critical piece of information is the list of open ports of the system. Open ports of a system can be exploited in a number of ways. To identify open ports a number of tools are available [4].

Currently a number of solutions are in place to deal with attacks. However, most of the solutions such as Antivirus and Intrusion Detection Systems indicate occurrence of an attack or an un-authorized activity when it happens. Having a system which predicts occurrence of attacks in the near future is advantageous. As port scans are usually performed before an actual attack, identification of port scan attempts gives precautionary indication that attacks might follow in the near future.

The goal of this project is to identify port scan attempts. This would make it possible to take precautionary steps to strengthen the defenses of the system. It would be very useful to have information about the machine from where the scans are coming. Information such as the Operating System being used, the possible location from where the scan came, information from WHOIS database and Traceroute would help in providing clues about the scanner. This information can be used against him if an attack takes place in future.

II. SCAN DETECTION METHODOLOGY

An attack typically goes through the following phases:

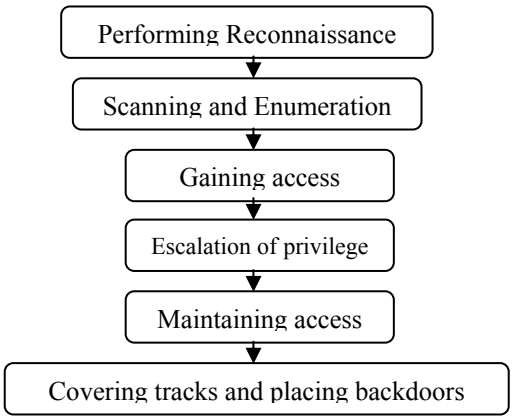

Figure 1. Steps in an attack.

Reconnaissance is considered the first pre-attack phase and is a systematic attempt to locate, gather, identify, and record information about the target. The hacker seeks to find out as much information as possible about the victim. This first step is considered a passive information gathering. This involves activities such as information gathering, determining the network range, identifying active machines, finding open ports and access points, OS fingerprinting, fingerprinting services and mapping the network.

The port scanners are mainly of two types.

- Brute force scanners.
- Stealth scanners.

Brute force scanners essentially perform scans in an aggressive manner by scanning one port after another for the specified range. They establish a full connection to the target machine and inspect whether the port is open. Owing to the full connection establishment, it is possible to detect their presence. Thus when a large number of SYN packets arrive to request for a connection from a single IP address at multiple ports of the target machine, it indicates that a brute force scanner is being used to look for open ports.

Stealth scanners get their name from their pattern of not establishing a full connection with the target. They send a single packet with a particular flag set at the target, based on the response it can be understood whether the ports are open or not. There are various types of scans; patterns for each scan can be identified as follows:

978-1-4244-3805-1/08/$25.00 ©2008 IEEE

A. SYN Scans

In this scan a large number of packets with only the SYN flag set arrive at the destination. This scan does not complete the 3-way TCP connection establishment handshake and tears down the connection after the victim replies with a SYN/ACK indicating an open port. This scan can be easily identified if there are a large number of packets with the SYN flag set in them coming from a single host.

B. TCP Connect Scan

In this scan a large number of connections are established with the victim at different ports. Establishment of a connection at a port indicates that the corresponding port is open. Once the connection is established and the open ports identified, the connection is closed. As in this scan complete connection is established, TCP options such as timestamp and sequence acknowledgement are present. Thus if from a particular host a large number of connection are established at multiple ports in a very short span of time it can be inferred that a TCP CONNECT scan is coming from that machine.

C. ACK Scan

In this scan a large number of packets with only the ACK flag set arrive at the destination. This scan does not complete the 3-way TCP connection establishment handshake and tears down the connection after the victim replies with a SYN/ACK indicating an open port. This scan can be easily identified if there are a large number of packets with the ACK flag set in them coming from a single host.

D. FIN Scan

In this scan a large number of packets with only the FIN flag set arrive at the destination. If the victim replies with a RST it indicates that the port is closed, open ports simply ignore these packets. This scan can be easily identified if there are a large number of packets with the FIN flag set in them come from a single host.

E. NULL Scan

In this scan a large number of packets with no flags set arrive at the destination. The open ports ignore these packets whereas closed ports reply back with a RST. This scan can be easily identified if there are a large number of packets with the no flag set in them coming from a single host.

F. XMAS Scan

In this scan the flags FIN, PSH and URG are set. Open ports ignore these packets whereas closed ports reply with a RST. This scan can be easily identified if there are a large number of packets with the FIN, PSH and URG flag set in them coming from a single host.

G. UDP Scan

In this scan a large number of UDP packets arrive at the destination. This scan does not complete the 3-way TCP connection.

H. ICMP Scan

This includes sending ICMP echo request to the specified IP addresses to see if they are alive.

I. Fragmentation Attack

In order to overcome rules set by firewalls, attackers split the packets into small fragments and send these individual pieces over the network. These packets pass the firewall as rules meant for these individual pieces are not present This can be detected if there are a large number of packets with a 'header too short' string in them.

III. FRAMEWORK OF THE SYSTEM

The system is designed to help detect a possible port scan, getting additional information about the scanner such as his probable location, Operating System being used by attacker would help in uncovering the identity of the scanner.

The following figure shows the framework of the proposed system.

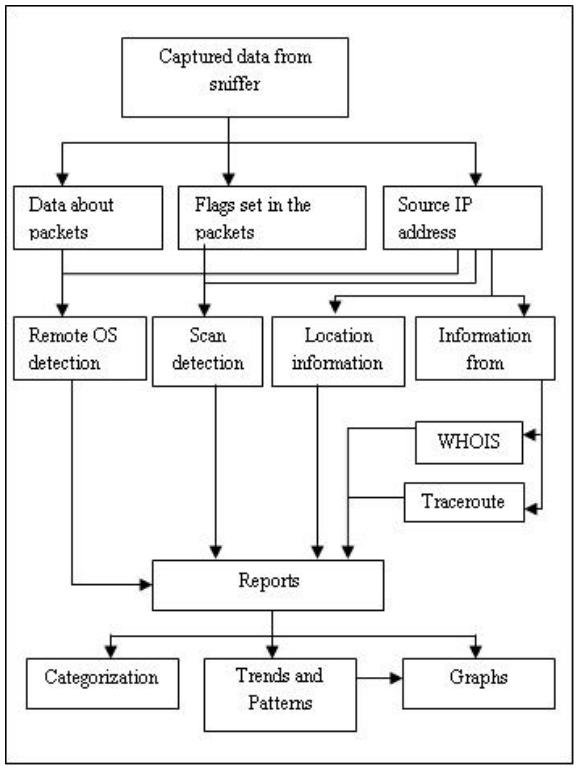

Figure 2. Framework of the System

For an administrator it would be very helpful to have knowledge that someone is performing a port scan over a system. This gives a hint that some sort of attack might follow. Having an understanding of which ports are being scanned, it is possible to predict what kind of attack may follow. This helps the Administrator in taking precautionary steps before an attack takes place.

A. Scan Detection

Using the information provided by the sniffer about the incoming packets, in particular the TCP flags present, patterns similar to general and stealth scans are found within the incoming packets. If a patterns is identified it will be marked. If the number of marked entries reaches a pre-specified threshold, it indicates that a scan is being performed.

Filters are used to filter out the unnecessary traffic and concentrate only on packets which might indicate a port scan attempt. For instance, a filter ' tcp[tcpflags] (tcp-syn|tcp-fin|tcp-ack) !=0 ' captures data about packets which have Ack, Syn and Fin flags set. Thus, using appropriate filters enables capturing of relevant packets with regards to different types of scans.

Packets may be coming from many different sources; every packet is associated with the machine from which it is coming with the help of IP address.

Online activities such as checking the E-mail, internet messengers and surfing web-pages generate packets which might be captured by the filters used. But the number of packets captured within a short span of time for these activities are very few compared to those captured when a scan is coming from a port scanner. Moreover, port scanners generally send a large number of packets with a particular flag set to a large number of ports on the target machine; this further distinguishes general online activity from port scans. This approach allows detection of scans coming from brute force scanners as well as stealth scans.

IV. INFORMATION GATHERING

A packet sniffer provides information about incoming packets. The incoming packets coming from a machine give a very crucial piece of information, the IP address of the machine from where the scans came from. This IP address now acts as the identity of the scanning machine. Apart from detecting scans, it is advantageous to have a system which provides information about the machine from where scans come from. This information gives clues which if put together may be sufficient enough to uncover the real identity of the person performing the scans.

Some of the information which can be found about the remote machine using IP address is as follows.

A. Operating System Being Used

Operating System being used on a remote machine can be guessed once the IP address of the corresponding machine is known. Every Operating System has default values for certain parameters such as time to live (TTL), type of service (Tos) and window size (Wsize). For packets coming from a machine if we know the values for these parameters, the Operating System being used can be understood. The values considered for the system developed are as follows:

Type Of Service (TOS)

The TOS bits specifies how the network should make trade-offs between throughput, delay, reliability, and cost.

Don't Fragment Flag

This flag can be set to 1 by a transmitting device to specify that a datagram not be fragmented in transit. This may be used in certain circumstances where the entire message must be delivered intact as pieces may not make sense.

Window Size (Wsize)

Window size specifies the maximum amount of received data, in bytes, that can be buffered at one time on the receiving side of a connection. The sending host can send only that amount of data before waiting for an acknowledgment and window update from the receiving host.

Time To Live (TTL)

Time to Live originally involved a sense of time. It is now used as a simple, but very effective; count to prevent routing errors and loops. Every router that handles the packet decrements the TTL value and if it reaches zero the packet is returned with an ICMP Time Exceeded message.

Maximum Segment Size (MSS)

Maximum segment size has to do with defining the largest amount of data that a computer system or other type of communications device can efficiently deal with without breaking the data into smaller components. Generally, the maximum segment size is calculated as the number of bytes that the device can handle at one time.

Timestamp

Timestamps were conceived to assist TCP in accurately measuring round trip time (RTT) in order to adjust retransmission time-outs.

Sequence Acknowledge (SackOK)

Prior to SACK, a receiver could only acknowledge the latest sequence number of contiguous data that had been received, or the left edge of the receive window. With SACK enabled, the receiver continues to use the ACK number to acknowledge the left edge of the receive window, but it can also acknowledge other blocks of received data individually.

No Operation (NOP)

The nop TCP Option means "No Option" and is used to separate the different options used within the TCP Option field. The implementation of the nop field depends on the operating system used. Nop option occupies 1 byte.

Signatures can be prepared for different Operating Systems based on values for these parameters. For instance, the signatures for Operating Systems such as Windows and Linux families are as follows:

Operatign System : TOS, DF, Wsize, MSS, Timestamp, SackOK, nops.

Windows : 0x0, [none], 30000-90000, 1, 0, 1, 2.

Linux : 0x0, [DF], 5000-9000, 1, 1, 1, 1.

These values are present in the packets itself, as a result this information is provided by the sniffer. However, value for

parameters such as Timestamp, Nop, MSS are present in a packet only if a TCP connection is established between the sender and receiver. These parameters are collectively called TCP Options. These options are not present in packets which do not establish a complete connection. Thus, for packets belonging to stealth scans, these parameters are not present. As a result, Operating System detection would be done on the basis of remaining parameters.

For the system developed, two approaches have been used to detect the remote Operating System.

- Active Operating System detection
- Passive Operating System detection

Active Operating System Detection

In this method, services such as ping, telnet and file transfer protocol (FTP) are used to generate a response from the remote machine. From the incoming response packets, values for the different parameters are obtained. Once the values are obtained a simple correlation with signatures for different Operating Systems provides the possible Operating System being used on the remote machine.

This method may not be successful each time as it entirely depends on the reply packets coming from the remote machine. Good security measures such as a firewall with robust rules would block the ping and telnet packets sent to the remote machine. As a result no replies would be generated.

Passive Operating System Detection

In this method, values for the parameters are obtained from the incoming packets captured from the sniffer. The 'Verbose' option of TCPDump provides more detailed information about the packets. Among the information provided about the packets, values for the parameters required to detect remote Operating System are also provided. As a result, in contrast to the active method of remote Operating System detection, there is no dependency on reply packets from the remote machine. Operating System detection is achieved from the initial packets coming from the remote machine.

Being completely passive in nature, there is very little chance that the person on the other side ever comes to know that the Operating System being used by him is being detected. This method is more reliable compared to the active method as there is very little dependency.

B. Probable Location from IP Address

A number of websites on the internet provide information about an IP address, specifically the location of the machine to which this address belongs and the service provider. The system developed uses the services of such websites to provide the user with the probable location of the machine. Earlier such websites used to give information about the location in plain text and numbers signifying latitude and longitude. But over the years they have become more content rich. For instance, most of the websites are integrated with Google maps. Thus they show information about the location on a map making it easier to understand. Providing information about the location from a number of websites is advantageous as the results can be correlated for improved accuracy.

C. Information from WHOIS Database

WHOIS provides a very useful set of information such as the administrative contact, owner of a domain, the Internet Service Provider. Websites such as arin.net provide information from WHOIS. WHOIS query is also present in most of today's Operating System so one just needs to type 'whois' followed by the IP address or domain name to get information.

As WHOIS provides information about owner of a domain, it is especially useful in cases where scans are coming from private organizations as information provided by them during domain registration is most likely to be present in the WHOIS database. Another valuable piece of information provided by WHOIS is the Internet Service Provider for a particular IP address. If a large number of scans are observed to be coming from a particular IP address, further details about the owner can be obtained from the ISP indicated by WHOIS.

D. Information from Traceroute

Traceroute is the program that shows the route over the network which packets take between two systems. It lists all the intermediate routers a connection must pass through to get to its destination. It is useful to have information about the possible route which the incoming packets from the scanning machine might have taken. Traceroute gives information about the various routers through which the packets travel from source to destination. This also gives clues about the location of the routers. The country to which the router belongs can be understood through the IP address of that router. Thus, traceroute gives clues about the location of the scanning machine.

E. Self DIagnosis

A very basic precaution which should be taken to secure a system is to close all the unnecessary ports. Operating Systems by default keep some ports open. In order to ensure maximum security the open ports which are not being used for any services or by any applications should be closed. Performing port scanning over our own machines is a very effective security measure.

F. Identify Attacks Based On Ports Scanned

Many attacks and exploits are performed on open ports. The pattern usually followed is to find out if a particular port is open and then execute the attack or carry out an exploit on that port. If a number of scans are observed to be coming on a particular port, it indicates that an attack or exploit on that port may be performed. Thus, analysing the ports on which repeated scans are being performed gives clues about what kind of an attack may follow on that particular port. The system developed shows the number of times a scan has been performed on a particular port and based on this the corresponding attack or exploit which the attacker may try.

G. Trends and Patterns

It has been observed that in most of the crimes patterns are present, computer crimes are no different. Security systems should be capable of identifying trends or patterns followed by attackers in performing attacks. This helps in taking precautionary steps to avoid attacks or scans in future. Also, this gives a chance to nab the attacker by recording his activities and gathering sufficient information about him.

V. EXPERIMENTAL SETUP

The system has been developed in the Linux environment using Fedora Core 5. Code for the system has been written in Perl.

Data was collected in a LAN environment by performing port scans on a machine with the system installed. Well known port scanners such as Nmap (both windows and Linux version), Angry IP, Megaping were used.

VI. RESULTS

Results are shown for scans performed over a period of time. Both Windows and Linux port of Nmap were used along with other por scanners such as AngryIP and MegaPing.

Time Interval	Number of packets
14:40 - 14:41	15
14:41 - 14:42	13
14:42 - 14:43	3744
14:43 - 14:44	4379
14:44 - 14:45	3180
14:45 - 14:46	3515
14:46 - 14:47	19
14:47 - 14:48	12
14:48 - 14:49	3508
14:49 - 14:50	3722
14:50 - 14:51	3725
14:50 - 14:51	3420

Figure 3. Number of incoming packets at different time intervals.

The table above shows sudden increase in network activity when normal operations were performed. The sudden surge in incoming packets is a case when scans are being performed.

A. Scan Detection

The system has successfully identified scans coming from most of the port scanners available today. This includes scans from popular port scanners such as Angry IP, Nmap, MegaPing. Scans which establish full connection between the two hosts as well as stealth scans which open a half connection are detected. Most of the different scans supported by Nmap are identified based on the type of flag which is set in the incoming packets.

The scans from port scanners were performed when the machine running the port scan detector was online. Basic tasks such as checking the E-mail, Internet messaging using messengers were also performed; the system did not show these activities as port scans.

Scan Type	Count
Syn	129
Fin	97
Ack	63
Tcp-Connect	107
Xmas	49

Figure 4. Count of different scans detected over a month.

B. Information Gathering

Remote Operating System Detection

Operating System being used on the remote machine are being identified using the two methods mentioned in Section 3. Results given by the passive methods are more accurate compared to that given by active methods. The TTL is one of the parameters used in active method to identify the Operating System. Some of the current generation port scanners give the ability to manipulate the TTL which would be sent in the outgoing packets. The TTL can also be modified using other methods, for instance, the default value of TTL can be changed in Windows from the registry entry using Regedit. Thus, sometimes the results may not be conclusive.

Location from IP Address

Using the services of websites the probable location of a machine is being shown with the help of IP address. Websites such as www.ip2location, www.melissadata.com have been used to show this information. Clues about the information are also obtained from Traceroute by examining the IP address of the routers through which the packet travels in order to reach the destination. The information provided by WHOIS also gives clues about the location of the machine to which the IP address belongs.

Information from WHOIS

The default WHOIS application in Fedora 5 has been used. It gives valuable information such as the Domain owner, the contact number of a person from the organization. This information is especially useful when scans come from an organization. As organizations usually have Internet presence in the form of websites, they are most likely to have some sort of information in the WHOIS database. As a result it is easier to identify the source from where scans are coming. Another useful piece of information which is provided is the Internet Service Provider

Information from Traceroute

The default Traceroute application in Fedora 5 has been used. It shows the IP address of the routers through which the packets pass in order to reach the destination machine. Asteric in the hop count gives an indication that the ICMP packets might be blocked by a firewall or the host is unreachable. The

hop count and the IP address of the routers give an indication about the location of the destination machine.

Self Diagnosis

The system is capable of showing the ports open on the host system. The Perl module IO::Socket has been used to find out whether a specified port or range of ports are open or not.

Trends and Patterns

The system developed stores a report of the results provided by scan detection. Analysis of data such as time and day a particular scan was performed, from which IP the scan was performed, different ports on which the scans were performed helps in understanding the behaviour of the attacker. The system presents this information in graphical format in the form of bar graphs. Combination of different parameters such as the number of times scans came from a particular IP address, the types of scans performed on different days of the week give indication of the patterns followed by attackers.

Figure 5 shows a trend that more scans come on Saturdays than other days of the week. This gives the administrators a hint of when scans might be expected.

Syn	Sunday	30
	Monday	19
	Tuesday	0
	Wednesday	18
	Thursday	2
	Friday	12
	Saturday	48
Fin	Sunday	25
	Monday	8
	Tuesday	0
	Wednesday	0
	Thursday	13
	Friday	14
	Saturday	37
Xmas	Sunday	14
	Monday	0
	Tuesday	5
	Wednesday	7
	Thursday	0
	Friday	0
	Saturday	23

Figure 5. Count of number of times scans were detected on corresponding days of the week over a month.

Date	Scan Type	Number of Packets
19.04.2008	Syn	35640
19.04.2008	Ack	17223
19.04.2008	Fin	43463
20.04.2008	Syn	5203
21.04.2008	Syn	6542

Figure 6. Scan types and number of packets from 19.04.2008 to 21.04.2008.

IP Address	Day	Count
192.168.23.3	Tuesday	3
192.168.23.7	Saturday	9
192.168.23.8	Monday	39
192.168.23.10	Thursday	4
202.159.228.80	Tuesday	12
202.194.16.5	Monday	23

Figure 7. shows a trend that a large number of scans came from IP address 192.168.23.8 which is an internal address. In an organizational environment this would indicate that an employee from within the organization is performing scans.

VII. CONCLUSION

The system is designed to detect a possible port scan, getting additional information about the scanner such as his probable location, Operating System being used by attacker would help in uncovering the identity of the scanner.

For administrator it would be very helpful to have knowledge that someone is performing a port scans over a system. This gives a hint that some sort of attack might follow. Having an understanding of which ports are being scanned, it is possible to predict what kind of attack may follow. This helps the Administrator in taking precautionary steps before an attack takes place.

REFERENCES

[1] Fyodor, "The Art of Port Scanning", *Phrack Magazine, Volume 7, Issue 51, September 01 1997, Article 11 of 17.*

[2] Shaun Jamieson, "The Ethics and Legality of Port Scanning", October 8, 2001. http://www.sans.org/reading_room/whitepapers/legal/71.php

[3] Fyodor, "Remote OS Detection using TCP/IP Fingerprinting (2nd Generation)", Jan 2007. http://nmap.org/osdetect/index.html#id287961

[4] Roger Christopher, "Port Scanning Techniques and the Defense Against Them", October 5, 2001. http://www.sans.org/reading_room/whitepapers/auditing/70.php

[5] Kimberly Graves, "Official Certified Ethical Hacker Review Guide", Wiley Publishing, pp.15-65, February 2007.

[6] Kocher, J.E.; Gilliam, D.P., "Self port scanning tool: providing a more secure computing environment through the use of proactive port scanning", 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, 13-15 June 2005.

[7] Nmap Reference Guide (Man Pages) http://nmap.org/man/

[8] TCPDump Reference Guide (Man Pages) http://www.tcpdump.org/tcpdump_man.html