Chapter 3

Information Gathering

OUTLINES

- Footprinting and Enumeration
- Location and Organizational Data
- Electronic Documents
- Financial Data
- Employees
- Domains
- DNS and Traceroute Information
- Zone Transfers
- IP Ranges
- Security Search Engines
- Active Reconnaissance and Enumeration
    - Hosts
    - Services
    - Port Scanning
    - Common ports and services
    - Service and Version Identification
    - Operating System Fingerprinting
    - Nmap
    - Network Topology
    - Eavesdropping and Packet Capture
    - Email
    - Social Networking Sites
    - Relationships
    - Shares
    - Web Pages and Servers
    - Fingerprinting
- Defenses Against Active Reconnaissance
- Preventing Passive Information Gathering

Scenario, Part 1: Plan for a Vulnerability Scanning

Black box penetration test against MCDS, LLC. You have worked out the scope of work and rules of engagement and know that your engagement includes the organization's website and externally accessible services, as well as all systems on both wired and wireless networks in their main headquarters location. Third-party providers, services, and off-site locations are not included in the scope of the test.

This is a black box test; you must first identify the organization's domains, IP ranges, and other information, then build and execute an information-gathering plan.

Footprinting and Enumeration

The first step in many penetration tests is to gather information about the organization via passive intelligence gathering methods. Passive methods are those that do not actively engage the target organization's systems, technology, defenses, people, or locations. The information gathered through this process is often called OSINT, or open-source intelligence. Among other data that can be gathered, OSINT is often used to determine the organization's footprint: a listing of all of the systems, networks, and other technology that an organization has. Of course, if you are conducting a white box test, you may already have all of this information in the documentation provided by the target organization.

OSINT

OSINT Includes data from publicly available sources, such as DNS registrars, web searches, security-centric search engines like Shodan and Censys, and a myriad of other information sources. It also includes information beyond technology-centric organizational information. Social media, corporate tax filings, public information, and even the information found on an organization's website can be part of open-source intelligence gathering.

Open Source Security Testing Methodology Manual (OSSTM)

National Institute of Standards and Technology (NIST)

NIST

The National Institute of Standards and Technology (NIST) provides standards, resources, and frameworks for cybersecurity. From a penetration tester's viewpoint, SP 800-115, the Technical Guide to Information Security Testing and Assessment, is a critical guidance document, particularly if you do work with the US government or a government contractor.

MITRE

The MITRE Corporation is a US not-for-profit corporation that performs federally funded research and development. Among the tools it has developed or maintains are a number of classification schemes useful to penetration testers

- **The Common Attack Pattern Enumeration and Classification (CAPEC):**

  Is a resource intended to help identify and document attacks and attack patterns. It allows users to search attacks by their mechanism or domain and then breaks down each attack by various attributes and prerequisites. It also suggests solutions and mitigations, which mean it, can be useful for identifying controls when writing a penetration test report.
  https://capec.mitre.org

- **The Common Vulnerabilities and Exposures (CVE):**

  Is Identifies vulnerabilities by name, number, and description. This makes the job of a penetration tester easier, as vendors, exploit developers, and others can use a common scheme to refer to vulnerabilities.
  https://www.cve.mitre.org

- **The Common Weakness Enumeration (CWE):**

  Is another community-developed list. CWE tackles a broad range of software weaknesses and breaks them down by research concepts, development concepts, and architectural concepts. Like CAPEC, it describes each weakness and how it can be introduced to code, what platforms it applies to, and what happens when something goes wrong.
  https://cwe.mitre.org

Scenario Part 2: Scoping the Penetration Test

To scope the penetration test that you are performing for MCDS, you need to determine the following items:

- domain names does MCDS own
- IP ranges does MCDS use for its public services
- email addresses can you gather
- What does the physical location look like, and what is its address?
- What does the organization's staff list and org chart look like?
- What document metadata can you gather?
- What technologies and platforms does MCDS use?
- Does MCDS provide remote access for staff?
- What social media and employee information can you find?

Location and Organizational Data

Tester may need to know more about the physical locations and defenses that a target has in place. Testers will typically start by working to understand what buildings and property the target organization uses. A black box test can make this harder, but public records can help by providing ownership and tax records. These records provide contact persons, whose details could help later. Additional physical location information that a tester will look for usually includes the physical security design, including locations of cameras, entrances and exits, guards, fences, and other physical security controls like badges or entry access systems. At this point in the information-gathering process, it isn't uncommon to find out that the organization has other locations, subsidiaries, or remote sites.

Methods:

Social engineering engagements or in-person security control testing, wireless or wired network penetration, or even dumpster diving to see what type of paper records and other information the tester can recover.

Electronic Documents

Help you understand how an organization is structured. They can also provide a wealth of other information, ranging from technologies used to staff names and email addresses, as well as internal practices and procedures. In addition to the information that is contained in the documents, many penetration testers will also carefully review the document metadata to identify additional useful information.

Methods:

(ExifTool): designed to allow you to quickly and easily view document metadata

(FOCA) & (metagofile): can be used to find metadata. FOCA scans using a search engine—Google, Bing, or DuckDuckGo—and then compile metadata information from files like Microsoft Office documents, PDF files, and other file types like SVG and InDesign files.

Financial Data

Financial disclosures, tax information, and other financial documents can provide additional information for motivated pen-testers. The US Securities and Exchange Commission provides the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, a service that allows you to look up SEC filings.

Methods:

(EDGAR): Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, a service that allows you to look up SEC filings.

Employees

Finding out who is employed by an organization can sometimes be as simple as using an online directory or checking its posted organizational charts. In most cases, identifying employees will take more work.

Methods:

LinkedIn and Facebook, as well as reviewing corporate email addresses, publications, and public records. Social engineering techniques can also be useful, particularly when searching for information on a specific individual or group.

Domains

Domain names are managed by domain name registrars. Domain registrars are accredited by generic top-level domain (gTLD) registries and/or country code top-level domain (ccTLD) registries. This means that registrars work with the domain name registries to provide registration services—the ability to acquire and use domain names. Registrars provide the interface between customers and the domain registries and handle purchase, billing, and day-to-day domain maintenance, including renewals for domain registrations.

Methods:

Target: Netflix.com

(WHOIS): Domain ownership and registration is maintained by registrars, with each registrar covering a specific portion of the world. Allows you to search databases of registered users of domains and IP address blocks and can provide useful information about an organization or individual based on their registration information.

DNS and Traceroute Information

The DNS converts domain names like google.com to IP addresses and IP addresses to domain names.

Methods:

Target: Netflix.com

(Nslookup): is command for this on Windows, Linux, and MacOS systems

Zone Transfers

A DNS zone transfer (AXFR) is a transaction that is intended to be used to replicate DNS databases between DNS servers. Of course, this means that the information contained in a zone transfer can provide a wealth of information to a penetration tester and that most DNS servers will have zone transfers disabled or well protected.

Methods:

Target: Netflix.com

- (Host tool): host -t axfr domain.name dns-server
- (Dig tool): dig axfr @target.nameserver.com domain.name
- Nmap (using the Nmap scripting engine or NSE): nmap –script dns-zone-transfer.nse –script-args dns-zone-transfer.domain <domain> -p53 <hosts>

A zone transfer will show you quite a bit of data, including the name server, primary contact, serial number, time between changes, the minimum time to live for the domain, MX records, name servers, latitude and longitude, and other TXT records.

IP Ranges

Once you know the IP address that a system is using, you can look up information about the IP range it resides in. That can provide information about the company or about the hosting services it uses. The IP address or hostname can also be used to gather information about the network topology around the system or device that has a given IP address. One of the first stops once you have an IP address is to look up who owns the IP range.

Methods:

(Whois): ex: Netflix.com

(Traceroute & tracert): explore the route to the IP

Now that we know who owns it, we can also explore the route to the IP. Using Traceroute (or tracert on Windows systems), you can see the path packets take to the host. Since the Internet is designed to allow traffic to take the best path, you may see multiple different paths on the way to the system, but you will typically find that the last few responses stay the same. These are often the local routers and other network devices in an organization's network, and knowing how traffic gets to a system can give your insight into their internal network topology.

Security Search Engines

A quick way to search for exposed systems belonging to an organization by domain or IP address is to use a security search engine. These search engines provide a way to review hosts, services, and other details without actively probing networks yourself.

Shodan

Shodan is one of the most popular security search engines and provides pre-built searches as well as categories of search for industrial control systems, databases, and other common search queries.

Censys

Much like Shodan, Censys is a security-oriented search engine. When you dig into a host in Censys, you will also discover geoIP information if it is available, a comprehensive summary of the services the host exposes, and drill-down links for highly detailed information.

Active Reconnaissance and Enumeration

Building a list of all of the resources or potential targets of a specific type is important in this state of a penetration test. Once sufficient open-source intelligence has been gathered, testers typically move on to an active reconnaissance stage with the goal of first building, then narrowing down the list of hosts, networks, or other targets. Techniques for each of these vary, so you will need to be familiar with each of the following methods.

Hosts

Enumerating hosts on a network may be the first task that most penetration testers think of when they prepare to assess a target. Active scans can identify many hosts, and it can be tempting to just rely on port scanners to identify hosts, but there are quite a few other ways to identify hosts on a network, and combining multiple methods can help to ensure that you didn't miss systems

- Leveraging central management systems like SCCM, Jamf Pro, or other tools that maintain an inventory of systems, their IP addresses, and other information.
- Network logs and configuration files can provide a wealth of information about systems on a network. Logs from DHCP servers can be particularly valuable, as most modern networks rely heavily on DHCP to issue addresses to network connected systems. Router logs, ARP tables, and other network information can also be very valuable.

Services

Service identification is one of the most common tasks that a penetration tester will perform while conducting active reconnaissance. Identifying services provides a list of potential targets, including vulnerable services and those you can test using credentials you have available, or even just to gather further information from. Service identification is often done using a port scanner.

Port Scanning

Port scanning tools are designed to send traffic to remote systems and then gather responses that provide information about the systems and the services they provide. Therefore, port scans are often one of the first steps in a penetration test of an organization.

Features:

- **Host discovery**
- **Port scanning and service identification**
- **Service version identification**
- **Operating system identification**

Important notes:

- **Number of ports is 65535**
- **Ports 0–1023 are known as "well-known ports" or "system ports," there are quite a few higher ports that are commonly of interest when conducting port scanning.**
- **Ports ranging from 1024 to 49151 are registered ports and are assigned by IANA when requested.**

Common ports and services

Port	TCP/UDP	service
20	TCP/UDP	FTP data
21	TCP/UDP	FTP control
22	TCP/UDP	SSH
23	TCP/UDP	Telnet
25	TCP/UDP	SMTP (email)
53	UDP	DNS
67	TCP/UDP	DHCP (server)
68	TCP/UDP	DHCP (client)
69	TCP/UDP	TFTP
80	TCP/UDP	HTTP
88	TCP/UDP	Kerberos
110	TCP/UDP	POP3
123	TCP/UDP	NTP
135	TCP/UDP	Microsoft EPMAP
136-139	TCP/UDP	NetBIOS
143	TCP	IMAP
161	UDP	SNMP
162	TCP/UDP	SNMP traps
389	TCP/UDP	LDAP
443	TCP/UDP	HTTPS
445	TCP	Microsoft AD and SMB
500	TCP/UDP	ISAKMP, IKE
515	TCP	LPD print services
1433	TCP	Microsoft SQL Server
1434	TCP/UDP	Microsoft SQL Monitor
1521	TCP	Oracle database listener
1812, 1813	TCP/UDP	RADIUS

Service and Version Identification

The ability to identify a service can provide useful information about potential vulnerabilities as well as verifying that the service that is responding on a given port matches the service that typically uses that port. Service identification is usually done in one of two ways: either by connecting and grabbing the banner or connection information provided by the service or by comparing its responses to the signatures of known services.

Operating System Fingerprinting

The ability to identify an operating system based on the network traffic that it sends is known as operating system fingerprinting, and it can provide useful information when performing reconnaissance. This is typically done using TCP/IP stack fingerprinting techniques that focus on comparing responses to TCP and UDP packets sent to remote hosts.

Nmap

A scan technique, like TCP SYN, Connect, ACK, or other methods. By default, Nmap uses a TCP SYN scan (-sS), allowing for fast scans that tend to work through most firewalls. In addition, sending only a SYN (and receiving a SYN/ACK) means that the TCP connection is not fully set up. TCP connect (sometimes called "full connect") scans (-sT) complete the TCP three-way handshake and are usually used when the user account using Nmap doesn't have the privileges needed to create raw packets—a common occurrence for penetration testers who may not have gained a privileged account yet during a test. A final common scan technique flag is the (-sU) flag, used to conduct a UDP-only scan. If you just need to scan for UDP ports, this flag allows you to do so. Nmap provides a multitude of features, and many flags. You'll need to know quite a few of the common ones, as well as how a typical Nmap command line is constructed, for the exam. Make sure you practice multiple types of scans and understand what their results look like and how they differ.

- A port range, either specifying ports or including the full 1–65535 range.
- Service version detection using the –sV flag.
- OS detection using the –O flag.
- Disabling Ping using the - Pn flag.
- The aggressiveness of the scan via the - T timing flag. The timing flag can be set either using a numeric value from 0 to 5 or via the flag's text representation name. If you use a number, 0 will run an exceptionally slow scan, while 5 is a very fast scan. The text representation of these flags, in order, is paranoid |sneaky | polite | normal | aggressive | insane. Some testers will use a paranoid or sneaky setting to attempt to avoid intrusion detection systems or to avoid using bandwidth. As you might suspect, -T3 or normal, is the default speed for Nmap scans.
- Input from a target file using -IL .
- Output to a variety of formats. You will want to be familiar with the -oX XML output flag, the - oN "normal" output mode, and even the outdated -oG greppable (searchable) format, which XML has almost entirely replaced. The -oA file, or "all" output mode, accepts a base filename and outputs normal, XML, and greppable formats all at the same time as basename.nmap , basename.xml , and basename.gmap . If you use multiple tools to interface with your Nmap results, this can be a very useful option!

Network Topology

Understanding the topology, or layout, of a network helps a penetration tester design their scanning and attack process. A topology map can provide information about what systems and devices are likely to be accessible, thus helping you make decisions about when to pivot to a different target to bypass security controls. Topology diagrams can be generated using tools like the Zenmap GUI for Nmap as well as purpose-built network topology mapping programs.

Methods:

(Zenmap): GUI for Nmap as well as purpose-built network topology mapping programs

Eavesdropping and Packet Capture

In addition to actively scanning for hosts and gathering topology information, penetration testers will also gather information using eavesdropping with packet capture or sniffer tools. Tools like Wireshark are often used to passively gather information about a network, including IP addresses, MAC addresses, and time to live for packets, and even data about services and the content of traffic when it is unencrypted.

Methods:

(Wireshark): are often used to passively gather information about a network, including IP addresses, MAC addresses, and time to live for packets, and even data about services and the content of traffic when it is unencrypted.

Email

Gathering valid email addresses commonly occurs prior to a phishing campaign or other penetration testing activity. In addition to more manual options, theHarvester is a program designed to gather emails, employee names, subdomains, and host information, as well as open ports and banners from search engines (including Shodan) and other sources.

Methods:

- TheHarvester: is a program designed to gather emails, employee names, subdomains, and host information, as well as open ports and banners from search engines

- Metasploit module:search_email_collector

Social Networking Sites

Social media enumeration focuses on identifying all of an individual's or organization's social media accounts. These are sometimes targeted in the exploit phase for password attacks, social engineering attacks, or attempts to leverage password resets or other compromised accounts to gain access.

Groups

Groups come in many forms, from Active Directory groups in an AD domain to group management tools built into identity management suites. Groups also exist in applications and service management interfaces. As a penetration tester, you need to understand both which groups exist and what rights, roles, or permissions they may be associated with.
If your target supports SNMP, and you have the appropriate community string, you can use *snmpwalk* to enumerate users as shown below using public as the community string and 10.0.0.1 as the target host. The grep and cut commands that the snmpwalk output is piped into will provide the user with information from the overall snmpwalk output.

Snmpwalk public -v1 10.0.0.1 1 | grep 77.1.2.25 | cut -d "" -f4

Relationships

Understanding how users relate to each other can be very useful when attempting to understand an organization. Fortunately, tools like the MIT Media Lab's Immersion tool (https://immersion.media.mit.edu/) can help you figure out which users connect frequently with others. Other relationship visualization tools are starting to become widely available, making big data techniques approachable for penetration testers.

Shares

Enumerating Samba (SMB) shares seeks to find all available shares, which are readable and writeable, and any additional information about the shares that can be gathered. SMB scanners are built into a variety of vulnerability scanning tools, and there are also purpose built SMB scanners like SMBMap. Nmap includes the smb-enum-shares and smb-enumusers NSE scripts as well.

Web Pages and Servers

Web pages and servers can be crawled and enumerated using a variety of tools. Dedicated web application assessment tools like w3af, Burp Suite, and many others can make this easier once you have identified web servers.
Many devices provide embedded web interfaces, so you may find a multitude of web servers during an active scan of a larger organization. One of the first tasks a penetration tester must perform is to narrow down the list of targets to a set of useful initial targets.

Fingerprinting

Application assessments rely on knowing information about the applications, such as the name, version number, underlying web server and application stack, host operating system, and any other details that can be gathered. This information is sometimes known as a fingerprint. Fingerprinting applications typically starts with banner grabbing.

GET / HTTP/3.0

Defenses Against Active Reconnaissance

- Limiting external exposure of services to those that absolutely must be exposed
- Using an IPS or similar defensive technology that can limit or stop probes to prevent scanning
- Using monitoring and alerting systems to alarm on events that continue despite these preventative measures

Preventing Passive Information Gathering

- Blacklisting systems or networks that abuse the service
- Using CAPTCHAs to prevent bots.
- Providing privacy services that use third-party registration information instead of the actual person or organization registering the domain.
- Implementing rate limiting to ensure that lookups are not done at high speeds.
- Not publishing zone files if possible, but gTLDs are required to publish their zone files, meaning this only works for some ccTLDs.

Questions

1. Mika runs the following Nmap scan:
Nmap -sU -sT -p 1-65535 example.com What information will she NOT receive?
A. TCP services
B. The state of the service
C. UDP services
D. MOD

2. What technique is being used in the following command:
Host -t axfr domain.com dns1.domain.com
A. DNS query
B. Nslookup
C. Dig scan
D. Zone transfer

3. After running an Nmap scan of a system, Lauren discovers that TCP ports 139, 443, and 3389 are open. What operating system is she most likely to discover running on the system?
A. Windows
B. Android
C. Linux
D. iOS

4. Charles runs an Nmap scan using the following command:
Nmap -sT -sV -T2 -p 1-65535 example.com
After watching the scan run for over two hours, he realizes that he needs to optimize the scan. Which of the following is not a useful way to speed up his scan?
A. Only scan via UDP to improve speed.
B. Change the scan timing to 3 or faster.
C. Change to a SYN scan.
D. Use the default port list.

5. Karen identifies TCP ports 8080 and 8443 open on a remote system during a port scan. What tool is her best option to manually validate running on these ports?
A. SSH
B. SFTP
C. Telnet
D. A web browser

6. Angela recovered a PNG image during the early intelligence-gathering phase of a penetration test and wants to examine it for useful metadata. What tool could she most successfully use to do this?
A. ExifTool
B. Grep
C. PsTools
D. Nginx

7. During an Nmap scan, Casey uses the -O flag. The scan identifies the host as follows:
Running: Linux 2.6.X
OS CPE: cpe:/o:linux:linux_kernel:2.6
OS details: Linux 2.6.9 - 2.6.33
What can she determine from this information?
A. the Linux distribution installed on the target
B. The patch level of the installed Linux kernel
C. The date the remote system was last patched
D. That the system is running a Linux 2.6 kernel between .9 and .33

8. What is the full range of ports that a UDP service can run on?
A. 1–1024
B. 1–16,383
C. 1–32,767
D. 1–65,535

9. Steve is working from an un-privileged user account that was obtained as part of a penetration test. He has discovered that the host he is on has Nmap installed and wants to scan other hosts in his subnet to identify potential targets as part of a pivot attempt. What Nmap flag is he likely to have to use to successfully scan hosts from this account?
A. -sV
B. -u
C. -oA
D. –sT

10. Which of the following tools provides information about a domain's registrar and physical location?
A. Nslookup
B. Host
C. WHOIS
D. Traceroute

11. Chris runs an Nmap scan of the 10.10.0.0/16 network that his employer uses as an internal network range for the entire organization. If he uses the -T0 flag, what issue is he likely to encounter?
A. The scan will terminate when the host count reaches 0.
B. The scan will not scan IP addresses in the .0 network.
C. The scan will progress at a very slow speed.
D. The scan will only scan for TCP services.

12. Which of the following Nmap output formats is unlikely to be useful for a penetration tester?
A. -oA
B. -oS
C. -oG
D. –oX

13. During an early phase of his penetration test, Mike recovers a binary executable file that he wants to quickly analyze for useful information. Which of the following tools will quickly give him a view of potentially useful information in the binary?
A. Netcat
B. strings
C. Hashmod
D. Eclipse

14. Jack is conducting a penetration test for a customer in Japan. What NIC is he most likely to need to check for information about his client's networks?
A. RIPE
B. ARIN
C. APNIC
D. LACNIC

15. After running an SNMP sweep, Greg finds that he didn't receive any results. If he knows there are no network protection devices in place and that there are devices that should respond to SNMP queries, what problem does he most likely have?
A. the SNMP private string is set.
B. There is an incorrect community string.
C. SNMP only works on port 25.
D. SNMP sweeps require the network to support broadcast traffic.

16. Charles uses the following hping command to send traffic to a remote system. Hping remotesite.com -S -V -p 80 What type of traffic will the remote system see?
A. HTTP traffic to TCP port 80
B. TCP SYNs to TCP port 80
C. HTTPS traffic to TCP port 80
D. A TCP three-way handshake to TCP port 80

17. What does a result of * * * mean during a traceroute?
A. No route to host.
B. All hosts queried.
C. No response to the query, perhaps a timeout, but traffic is going through.
D. A firewall is blocking responses.

18. Rick wants to look at the advertised routes to his target. What type of service should he look for to do this?
A. A BGP looking glass
B. A RIP-off
C. An IGRP relay
D. A BGP tunnel

19. Why would a penetration tester look for expired certificates as part of an information gathering and enumeration exercise?
A. They indicate improper encryption, allowing easy decryption of traffic.
B. They indicate services that may not be properly updated or managed.
C. Attackers install expired certificates to allow easy access to systems.
D. Penetration testers will not look for expired certificates; they only indicate procedural issues.

20. John has gained access to a system that he wants to use to gather more information about other hosts in its local subnet. He wants to perform a port scan but cannot install other tools to do so. Which of the following tools isn't usable as a port scanner?
A. Hping
B. Netcat
C. Telnet
D. ExifTool