Module 1 / Unit 2
Reconnaissance Techniques

Objectives

On completion of this unit, you will be able to:

- Understand the range of environmental and network reconnaissance techniques that may be used to test security systems.
- Identify social engineering techniques and web search tools to perform reconnaissance.
- Use topology discovery, host discovery, and OS fingerprinting tools.

This unit covers the following exam domain objectives and content examples:

- 1.1 Given a scenario, apply environmental reconnaissance techniques using appropriate tools and processes
  Procedures / common tasks (Topology discovery, OS fingerprinting, Service discovery, Email harvesting, Social media profiling, Social engineering, DNS harvesting, Phishing) • Variables (Wireless vs. wired, Virtual vs. physical, Internal vs. external, On-premises vs. cloud) • Tools (Nmap, Host scanning, Network mapping, netstat)

- 4.5 Compare and contrast the general purpose and reasons for using various cybersecurity tools and technologies
  Network scanning (Nmap), Command line / IP utilities (netstat, ping, tracert / traceroute, ipconfig / ifconfig, nslookup / dig)

The Kill Chain

There are several models for describing the general process of an attack on systems security. These steps are often referred to as a kill chain, following the influential white paper Intelligence-Driven Computer Network Defense commissioned by Lockheed Martin (gtsgo.to/zprf). The security company Mandiant's APT1 report into Chinese cyber espionage units (gtsgo.to/hx9p2) was also hugely influential in shaping the language and understanding of modern cyber-attack lifecycles.

Mandiant is now owned by FireEye and the company produces many free incident analysis tools and cyber security best practice guides.
Stages of the Kill Chain

The following stages conflate some of these models into a general overview:

■ **Planning / scoping** - in this stage the attackers determine what methods they will use to complete the phases of the attack. One significant issue here is that the attacker will not want to draw attention to himself so will try to identify stealthy methods to proceed. The attacker also needs to establish resources to launch the attack. To evade detection, this will normally mean a botnet of compromised home computers, which can be used as unwitting zombies to facilitate scans, Denial of Service (DoS) attacks, and exploits and mask their origin.

■ **Reconnaissance / discovery** - in this phase the attacker discovers what he can about how the target is organized and what security systems it has in place. This phase may use both passive information gathering and active scanning of the target network. The outcome of the phase, if successful, will be one or more potential exploits.

■ **Weaponization** - in this phase the attacker utilizes an exploit to gain access. The point of access (a compromised computer or user account for instance) is referred to as a **pivot point**. The aim will be for the attacker to install tools to maintain covert access to the system. The installation of such tools is referred to as an **Advanced Persistent Threat (APT)**. This phase would normally comprise a number of steps:
  
  o **Exploit** - run code on the target system to exploit a vulnerability and gain elevated privileges.
  
  o **Callback** - establish a covert channel to an external **Command and Control (C2 or C&C)** network operated by the attacker.
  
  o **Tool download** - install additional tools to the pivot to progress the attack.

■ **Post-exploitation / lateral discovery / spread** - if the attacker obtains a pivot point, the next phase is typically to perform more privileged network scans with a view to discovering more of the network topology, locating and exploiting additional pivot points, and identifying assets of interest.

■ **Action on objectives** - in this phase, the attacker typically uses the access he has achieved to covertly copy information from target systems (**data exfiltration**). An attacker may have other goals or motives however.

■ **Retreat** - once the attacker has achieved his initial aims without being detected, he may either maintain an APT or seek to withdraw from the network, removing any trace of his presence to frustrate any subsequent attempt to identify the source of the attack.
Reconnaissance Phase Techniques

In the rest of this unit, we will focus on the techniques used to facilitate the reconnaissance / discovery phase. There are three main categories of information gathering for an attacker to use:

- Open Source Intelligence (OSINT) - this refers to using web search tools and social media to obtain information about the target. It requires almost no privileged access as it relies on finding information that the company makes publicly available, whether intentionally or not.

- Social engineering - this refers to obtaining information, physical access to premises, or even access to a user account through the art of persuasion.

- Scanning - this refers to using software tools to obtain information about a host or network topology. Scans may be launched against web hosts or against wired or wireless network segments, if the attacker can gain physical access to them.

Reconnaissance Phase Objectives

The objective of the reconnaissance phase is to obtain information that will be useful for identifying valuable data assets (something the attacker would like to obtain) and potential pivot points. The attacker will therefore attempt to obtain as much information as possible in the following general categories:

- Organization - what does the business or institution do? What use of IT does it make? What is its attitude to security and risk? How accessible is the organization, physically and "digitally"?

- Employees - who works at the organization? What are the departmental or hierarchical functions? Who has privileged access? What outside interests might be exploitable?

- IT systems - what operating systems and network appliances are deployed? What software applications are in use? What security systems are deployed?

- Suppliers and customers - what access do other organizations have to the target?
Open Source Intelligence

Most companies and the individuals that work for them publish a huge amount of information about themselves on the web and on social media sites. Some of this information is published intentionally; quite a lot is released unintentionally or can be exploited in ways that the company or individual could not foresee. An attacker can "cyberstalk" his victim to discover information about them via Google Search or by using other web or social media tools.

This information gathering is also referred to as passive reconnaissance. Publicly available information and tools for aggregating and searching it are referred to as Open Source Intelligence (OSINT).

If an attacker is already thinking about covering their tracks, they will not use an account that can be linked back to them to perform this type of reconnaissance. This might mean the use of a public workstation, an anonymized proxy or VPN, or a compromised host. Another approach is to use false credentials to set up a temporary web server instance. There are also "bulletproof" hosting providers and ISPs that specialize in providing "no questions asked, anonymity guaranteed" services.

Google Hacking and Search Operators

To perform "Google hacking" (meaning hacking information via Google Search rather than trying to hack Google’s servers) you will need to be familiar with the search engine’s advanced syntax, though you can also build queries using the advanced search page (google.com/advanced_search). Some of the most important operators are as follows:

- Quotes - use double quotes to specify an exact phrase and make a search more precise.
- NOT - use the minus sign in front of a word or quoted phrase to exclude results that contain that string.
- AND / OR - search strings use a logical OR between terms automatically. You can use the keyword AND to force results to contain both strings. You must type the operator in caps or you can use the pipe (|) character for OR. You may also want to use the AND and OR keywords but with parentheses. For example, compare:
  
  user account password AND database
  (user OR account) AND password AND database

- Scope - a multitude of keywords can be used to target the search. Examples include site: (within a domain or TLD), filetype:, related: (return results from sites that Google identifies as similar to the one specified), and allintitle: / allinurl: / allinanchor: (match terms in a specific part of the page).
URL modifiers - you can add these to the results page URL to affect the results returned. Some examples include &pws=0 (do not personalize), &filter=0 (do not filter), and &tbs=li:1 (do not autocorrect search terms).

Google Dorks

As well as researching people, Google hacking can also be performed to identify vulnerable web servers and web applications or to obtain information from a web server that may not have been intended for publication. The Google Hacking Database maintained by Offensive Security (gtsgo.to/s0gc9) contains a list of search strings to locate such "Google Dorks" who are running vulnerable web application versions, have made files containing passwords available, or left a webcam publicly accessible.

You can use this database to learn the search operators that return fruitful results.

Running these searches is likely to trigger Google's CAPTCHA mechanism. This requires you to answer a challenge (identifying parts of an image for instance) and is designed to prevent the abuse of search by automated software.

Email Harvesting

The general purpose of email harvesting is to identify who works at a company. Most companies use real names for email addresses. This makes it possible for the attacker to identify social media or personal web accounts operated by an employee and from there try to identify an exploit.

An attacker will also want to try to match email addresses to job roles. In many circumstances a company may just publish information about senior staff and their job roles on its website or in promotional material such as a shares prospectus or the information filed with regulatory authorities (the SEC's Edgar database [gtsgo.to/s0gc9] for instance).

There are many methods of email harvesting:

- Trading lists from spammers or obtaining legitimate sales lead databases.
- Use a Google search against "*@target.com" or use an automated "scraper" tool that scans pages and social media for email addresses.
- Test the email system for bouncebacks against a dictionary of potentially valid addresses. Note that this is likely to alert the organization, if they are running any sort of intrusion detection.

theHarvester (gtsgo.to/r48nr) is a command-line tool for gathering subdomain information and email addresses included with the Kali pen testing Linux distribution.
Social Media Profiling

Once an attacker obtains a list of names of people that work at a company they can set about using social media to build up a profile of each employee to determine whether there are any vulnerabilities to social engineering attempts. To obtain private information an attacker would need to become a contact or hack the account of one of the target's existing contacts. Your online privacy may only be as good as your friends' passwords...

Remember that an indirect approach may also be fruitful. Rather than investigate a company directly, the attacker may identify a supplier or customer with weaker security controls and use them as a means of obtaining access to the target.

Even without private access, an unwary user might have made a large amount of information about themselves publicly available, especially on a business networking site such as LinkedIn. Social media analytics and OSINT software (such as pipl.com, peekyou.com or echosec.net) can aggregate and process the metadata from multiple sites to build up surprisingly detailed pictures of user's interests and even their habits and geographic location at a particular point in time.

OSINT can allow an attacker to develop any number of strategies for compromising a target. Locating an employee on a dating site might expose opportunities for blackmail or entrapment; finding an employee looking for a second-hand laptop or mobile on an auction site might allow an attacker to get a compromised device into the employee's home or workplace. Knowing the target's routine or present location might facilitate break-in or theft or create an opportunity for some sort of social engineering.
DNS Harvesting

An attacker might be able to obtain useful information by examining a company's domain registration records by running a `whois` lookup against the appropriate registry.

```
whois -v gtslearning.com
```

The `whois` command is part of Linux and for Windows users is available as one of the utilities in the Sysinternals suite. See Unit 2.1 and Unit 4.3 for more information about Sysinternals.

An attacker may also test a network to find out if the DNS service is misconfigured. A misconfigured DNS may allow a zone transfer, which will give the attacker the complete records of every host in the domain, revealing a huge amount about the way the network is configured. You can use the `nslookup` command in interactive mode to attempt a zone transfer:

```
set type=any
ls-d gtslearning.com
```

```
C:\WINDOWS\system32\cmd.exe - nslookup - ns1.clok.net
```

Testing whether the name server for gtslearning.com will allow a zone transfer
You can also use the `dig` command from a UNIX or Linux machine:

```
dig axfr NameServer Target
```

A zone transfer is often called an "axfr" after this switch sequence. For example, the following command queries the name server "ns1.isp.com" for the zone records for the "widget.com" domain:

```
dig axfr ns1.isp.com widget.com
```

If DNS harvesting is successful, you will obtain IP addresses for servers in the target domain. You can use a `geoip` tool to identify the approximate geographic location of the servers.

The `netcraft.com` site also contains a useful domain analysis tool.

![Netcraft Site Report](image)

**Website Ripping**

A **website ripper (or copier)** is a tool that caches the code behind a website. A tool such as httrack ([httrack.com](http://httrack.com)) recurses through each directory of the local site and can follow links to third party sites to a specified depth. Analyzing the ripped site might reveal vulnerabilities in the code or the web application used to deliver the content. There might be old or forgotten orphaned pages with useful information. Website ripping is also a means of harvesting email addresses.

See [Unit 3.3](#) for more information about application code vulnerability testing.
Social Engineering

The information gleaned from searching publicly available resources can be used to launch a social engineering attack. This type of attack will attempt to exploit a person as a "pivot point" either to obtain user credentials, to compromise an authorized device, such as a phone or laptop, or obtain access to the company premises.

Social engineering attacks may be launched face-to-face or over the phone. The attacker will use the information gathered about the company and the employee target to make the attack convincing. Attacks may also be launched over email or instant messaging or via social media, all of which can be conceived as some sort of phishing attack. Often the term spear phishing is used when a specific employee is targeted or the term whaling appears when an attacker goes after a senior employee.

An attack might proceed using some detail of something that is happening at the company. Perhaps there is an IT upgrade project ongoing and the attacker poses as a contractor. Alternatively, the attacker might exploit the target's social interests or hobbies. Perhaps they can use a third-party website to deliver an exploit or craft an email with a malicious payload that appeals to the target in a non-work context.

The Social Engineer Toolkit (gtsgo.to/zi2hp) is a modularized framework for developing phishing and spoofing attacks and for exploiting other social engineering vectors. It is based on the Metasploit Framework pen testing tool. See Unit 3.2 for more information about Metasploit.
Topography Discovery

Topography discovery (or “footprinting”) is the part of the discovery phase where the attacker starts to identify the structure of the target network. Organizations will also use topology discovery as an auditing technique to build an asset database and identify non-authorized hosts or network configuration errors.

Variables in Topography Discovery

Networks can be built in many different ways and an attacker might have only external access to the network systems rather than internal access.

- **Web / remote access** - if the attacker only has public access to the company’s web server infrastructure, they will try to identify information about where the servers are hosted, what web server and web application versions are running, what application development environment is in place, what backend database servers are used, and so on. As well as web infrastructure, the attacker might be able to identify remote access methods used by the company (such as VPNs or modems) and attempt to exploit those.

- **Wireless** - an attacker may be able to intercept wireless traffic from a company’s private network. If the attacker is very lucky, the network may be unencrypted or based on WEP, in which case they will be able to obtain actual data. Even when faced with secure encryption (WPA / WPA2), the attacker may be able to identify active access points and stations (client devices) and possibly host names and the IP scheme in use.

- **Wired** - a sniffer attached to a network segment can identify hosts communicating over the network.

- **Virtual versus physical** - many networks implement servers as virtualized hosts with networking infrastructure also implemented as virtual network ports, switches, and routers. This can make the hosts less accessible to intrusive scanning.

- **On-premise versus cloud** - extending the idea of virtual hosts, an organization might host its entire network infrastructure in the cloud, using either a hosting provider or as private infrastructure. To probe cloud infrastructure, the adversary usually has to gain access through the Internet, unless they can infiltrate the cloud provider's systems somehow.
Network Mapping Tools

A network mapping tool performs host discovery and identifies how the hosts are connected together on the network.

For auditing, there are enterprise suites such as Microsoft's System Center products or HP's OpenView. Such suites can be provided with credentials to perform authorized scans and obtain detailed host information via management protocols such as the Simple Network Management Protocol (SNMP).

An attacker attempting to work out the network topology stealthily faces a number of problems:

■ Gaining access to the network - both the challenge of connecting to the physical wired or wireless network and of circumventing any access control or authentication mechanisms that could block his equipment from receiving network traffic.

■ Scanning stealthily - to prevent the network owner detecting and blocking the scans and being alerted to an intrusion event.

■ Gaining access to the wider network from the local segment - this may involve defeating access control lists on routers and firewalls.

A couple of basic Windows and Linux commands can be used to facilitate host discovery.

**ipconfig / ifconfig**

The ipconfig (Windows) or ifconfig (Linux) command can be used to report the configuration assigned to the network adapter. The attacker can identify whether the network uses DHCP or a static IP addressing scheme.

Identifying the current IP configuration with ipconfig
**ping**

The `ping` command can be used to detect the presence of a host on a particular IP address or that responds to a particular host name. You can use `ping` with a simple script to perform a ping sweep. The following example will scan the 10.1.0.0/24 subnet from a Windows machine:

```
for /l %i in (1,1,255) do @ping -n 1 -w 100 10.1.0.%i | find /I "reply"
```

Performing a ping sweep in Windows with a For loop - searching multiple octets requires nested loops

A machine’s ARP cache can also be examined for host entries (using the `arp -a` command). The ARP cache shows the hardware (MAC) address of the interface associated with each IP address the local host has communicated with recently.

**Nmap Host Discovery**

Scanning a network using these simple tools would be time-consuming and non-stealthy however. Most topology discovery is performed using a dedicated tool like the Nmap Security Scanner (nmap.org). Nmap can use diverse methods of host discovery, some of which can operate stealthily and serve to defeat security mechanisms such as firewalls and intrusion detection. The tool is open source software with packages for most versions of Windows, Linux, and OS X. It can be operated with a command line or via a GUI (Zenmap).

The basic syntax of an Nmap command is to give the IP subnet (or IP address) to scan. When used without switches like this, the default behavior of Nmap is to ping and send a TCP ACK packet to ports 80 and 443 to determine whether a host is present. On a local network segment, Nmap will also perform ARP and ND (Neighbor Discovery) sweeps. If a host is detected, Nmap performs a port scan against that host (to determine which services it is running).

This OS fingerprinting can be time-consuming on a large IP scope and is also non-stealthy. If you only want to perform host discovery, you can use Nmap with the `-sn` switch (or `-sP` in earlier versions) to suppress the port scan.
Alternative host discovery methods can be used to overcome specific challenges:

- **List scan** (`-sL`) - this lists the IP addresses from the supplied target range(s) and performs a reverse-DNS query to discover any host names associated with those IPs. This can be used to check that you have specified appropriate targets. No probes are directed at the actual hosts.

- **TCP SYN ping** (`-PS PortList`) - to defeat a firewall, the attacker might want to probe ports other than the default HTTP / HTTPS ones. There are numerous other host detection techniques, including TCP ACK, UDP, SCTP INIT, and IP protocol ping.

- **Sparse scanning** (`--scan-delay Time`) - one of the principal means of making a scan stealthy is to collect results over an extended period. You can set Nmap to issue probes with significant delays between each probe to try to defeat intrusion detection systems. Of course, this makes host discovery a lengthy process. You can also configure delays using a timing template (`-Tn`, where `n` is a number from 0 to 5, with 0 being slowest). Another IDS evasion technique is to scan the scope in a random order (`--randomize-hosts`).

- **TCP Idle scanning** (`-sI`) - another way to make a scan stealthy is to use a so-called "zombie" host to appear to initiate the scan, disguising the identity of the host used to launch the scan. This type of scan takes much longer to complete than ordinary ping detection. Another masking option is to use the `-d` switch to add a number of decoy source IP addresses.

- **Fragmentation** (`-f` or `--mtu`) - this technique splits the TCP header of each probe between multiple IP datagrams. The principle is that splitting the header will make it harder for intrusion detection software to analyze. If the sensor attempts to reassemble the packets, that will consume more CPU cycles so that option is sometimes disabled to improve performance. As security appliances become more powerful however, fragmentation is less likely to succeed as a tactic (and the IDS can be configured to look for unusual fragmentation patterns).
The result of a host discovery scan will be a list of IP addresses and for each address whether a response was received or not.

**tracert / traceroute and Nmap Topology Discovery**

When performing host discovery on an internetwork (a network of routed IP subnets), the attacker will want to discover how the subnets are connected by routers (and whether any misconfigured gateways between subnets exist!)

The *tracert* (Windows) or *traceroute* (Linux) command tools provide a simple means of probing the path from one end system (host) to another, listing the intermediate systems (routers) providing the link. Of course, a routed internetwork will provide multiple paths for redundancy and fault tolerance. You can use source routing options within *tracert* to pre-determine the path taken but to discover a complete internetwork topology you need a more advanced tool.

You can use Nmap with the `--traceroute` option to record the path to an IP target address. The Zenmap tool can use this information to display a graphic of the detected network topology.

![Using the --traceroute option and topology view in Zenmap](image)

The Masscan tool ([gtsgo.to/q18t6](gtsgo.to/q18t6)) is another good option for scanning a large network as it can perform scans very quickly. You should note that speed generally involves a tradeoff with accuracy however.