5

Exploring Open-Source Intelligence

Just a couple of decades ago, the internet was not readily available to many people and organizations around the world due to many constraints. However, as technologies continue to evolve and **Internet Service Providers (ISPs)** work continuously to expand their network infrastructure to ensure everyone can connect and access the internet, there are more users on the internet today than ever and the numbers are continuing to increase as many people and organizations are using the internet for their personal gain and business, such as education, banking, marketing, digitally connecting with others, and e-commerce. This means that people are continuously creating and uploading data in various forms on many platforms on the internet, making information easily available to anyone with access to the internet. Sometimes, people, and even employees of an organization, share too much sensitive information on the internet without realizing how adversaries can collect and analyze the data to create intelligence, which can then be used to plan a cyber-attack on an organization.

Ethical hackers and penetration testers collect and analyze **Open Source Intelligence (OSINT)** found in online data sources. In this chapter, you will learn how they create a profile of their target to better understand them before proceeding to develop or acquire an exploit to compromise targeted systems and networks. Here, you'll learn how to use passive reconnaissance techniques and procedures to efficiently collect and analyze OSINT of a target. You will learn how to

use Google hacking techniques to filter the search results to identify any unintentionally exposed assets, systems, and resources of a targeted organization. In addition, you will gain the hands-on skills used by threat actors to perform passive reconnaissance on a targeted domain and identify sub-domains of an organization. Furthermore, you will explore various internet search engines that are commonly used by penetration testers to identify the technical infrastructure of a company and how hackers are able to collect employees' data to plan and improve their operations.

In this chapter, we will cover the following topics:

- Google hacking techniques
- Domain reconnaissance
- Sub-domain harvesting
- Identifying organizational infrastructure
- Harvesting employees' data
- Social media reconnaissance

Let's dive in!

Technical requirements

To follow along with the exercises in this chapter, please ensure that you have met the following hardware and software requirements:

- Kali Linux – **https://www.kali.org/get-kali/**
- DNSmap – **https://github.com/resurrecting-open-source-projects/dnsmap**
- Sublist3r – **https://github.com/aboul3la/Sublist3r**
- Sherlock – **https://github.com/sherlock-project/Sherlock**

If you're unable to connect to the internet from Kali Linux, use the `cat /etc/resolv.conf` command to determine whether your **Domain Name System (DNS)** servers are set correctly on Kali Linux, then use the `sudo systemctl restart NetworkManager` command to restart the Network Manager stack. As a last resort, you can restart the Kali Linux operating system.

Google hacking techniques

The concept of Google hacking, commonly referred to as Google dorking, is not the process of hacking into Google's network infrastructure or systems, but rather leveraging the advanced search parameters within the Google search engine to filter specific results. Many organizations don't always pay close attention to which systems and resources they are exposing on the internet. Google Search is a very powerful search platform that crawls/indexes everything on the internet and filters most malicious websites. Since Google indexes everything, the search engine can automatically discover hidden online directories, resources, and login portals of many organizations. Keep in mind that while Google's search capabilities can be used for finding sensitive information, over recent years, Google has taken steps to prevent abuse of its platform.

Using Google dorking techniques is not illegal but there's a very fine line that you shouldn't cross; otherwise, you'll be in legal trouble. We can use Google dorking techniques to discover hidden and sensitive directories and web portals on the internet, but if you use such information with malicious intentions to perform a cyber-attack, then you can face legal action.

To get started with learning about Google dorking, let's take a look at the following scenarios:

- Imagine you are required to use passive reconnaissance techniques to identify domains and sub-domains of a targeted organization. A common technique is to use Google Search to discover public-facing assets of the target. To do this, use the `site:domain-name` syntax to filter all results for the specified domain as shown here:

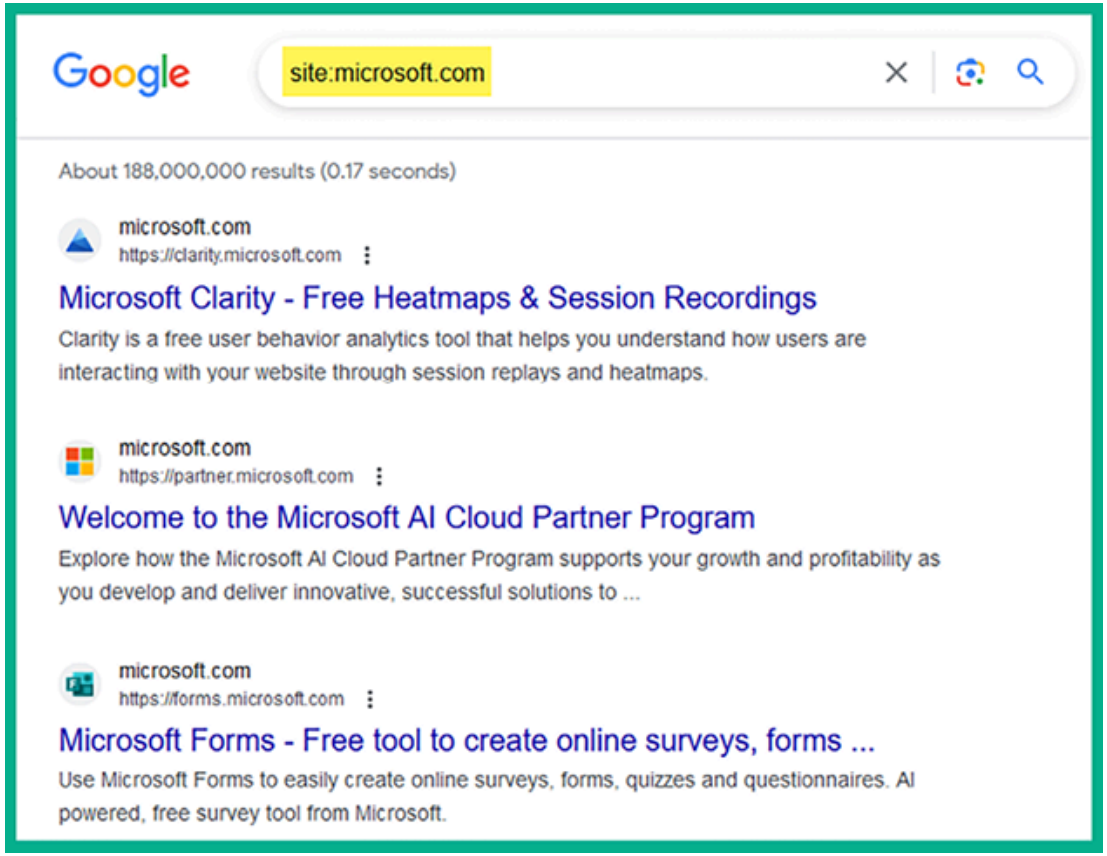


Figure 5.1: site:microsoft.com

As shown in the preceding screenshot, Google Search returned only the results that contained the targeted domain name.

- If you want to filter the search results based on a specific keyword for a targeted domain name, use the `keyword site:domain-name` syntax, as shown here:

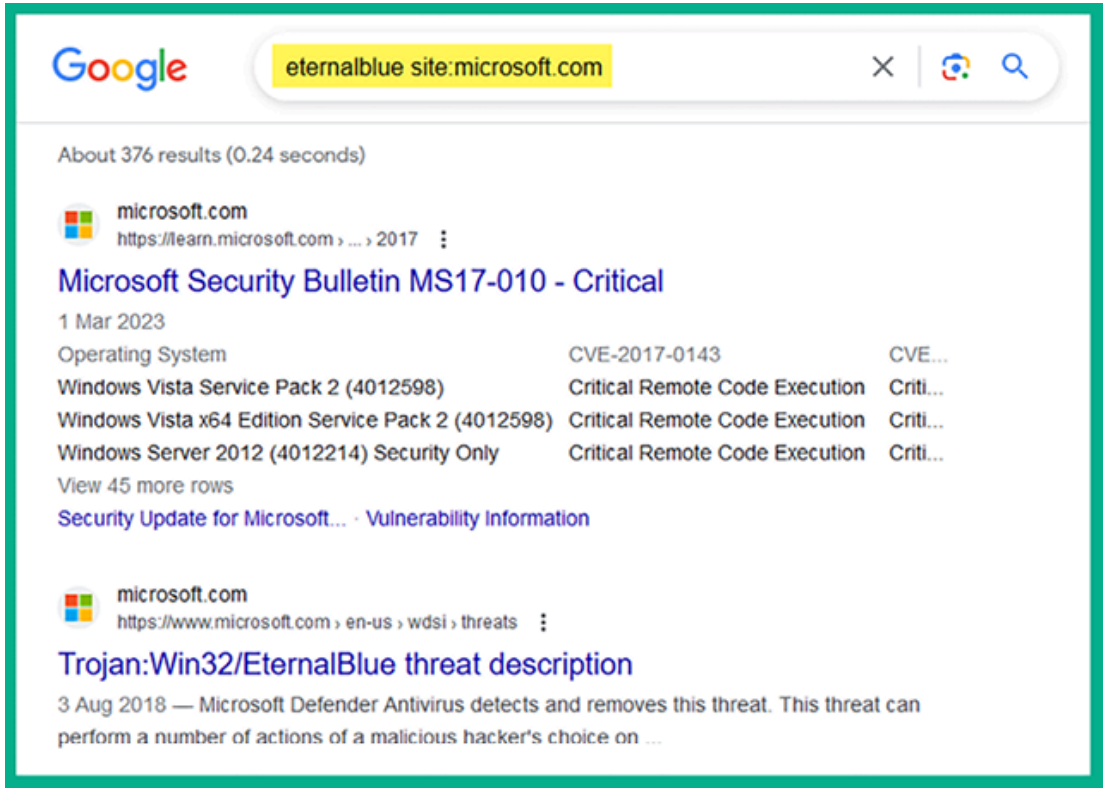


Figure 5.2: Using keyword

As shown in the preceding screenshot, the `eternalblue site:microsoft.com` syntax enables us to filter the search results to display all of Microsoft's domains and URLs that contain the `eternalblue` keyword. This is useful when performing research for security vulnerabilities and exploits on a targeted system based on the application, operating system, and vendor of the device.

- If you want to find all the domains of a targeted organization and filter the results based on two keywords, use the `keyword1 AND keyword2 site: domain-name` syntax, as shown here:

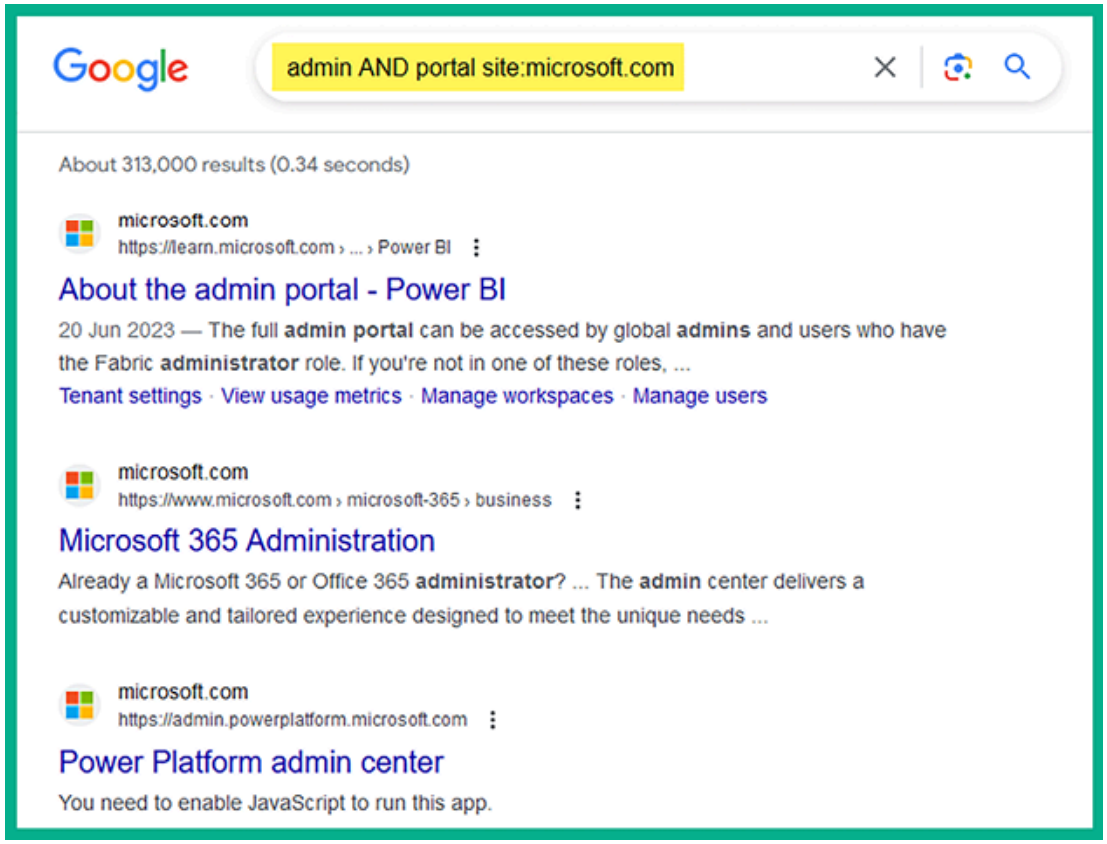


Figure 5.3: Multiple keywords

As shown in the preceding screenshot, using the `AND` operator with carefully chosen keywords helps us to find the login portals of a targeted domain.

You can use the `OR` syntax to specify keywords compared to using AND to include both keywords.

- If you're interested in searching for specific file types on a targeted domain, use the `site:domain-name filetype:file type` syntax, as shown here:

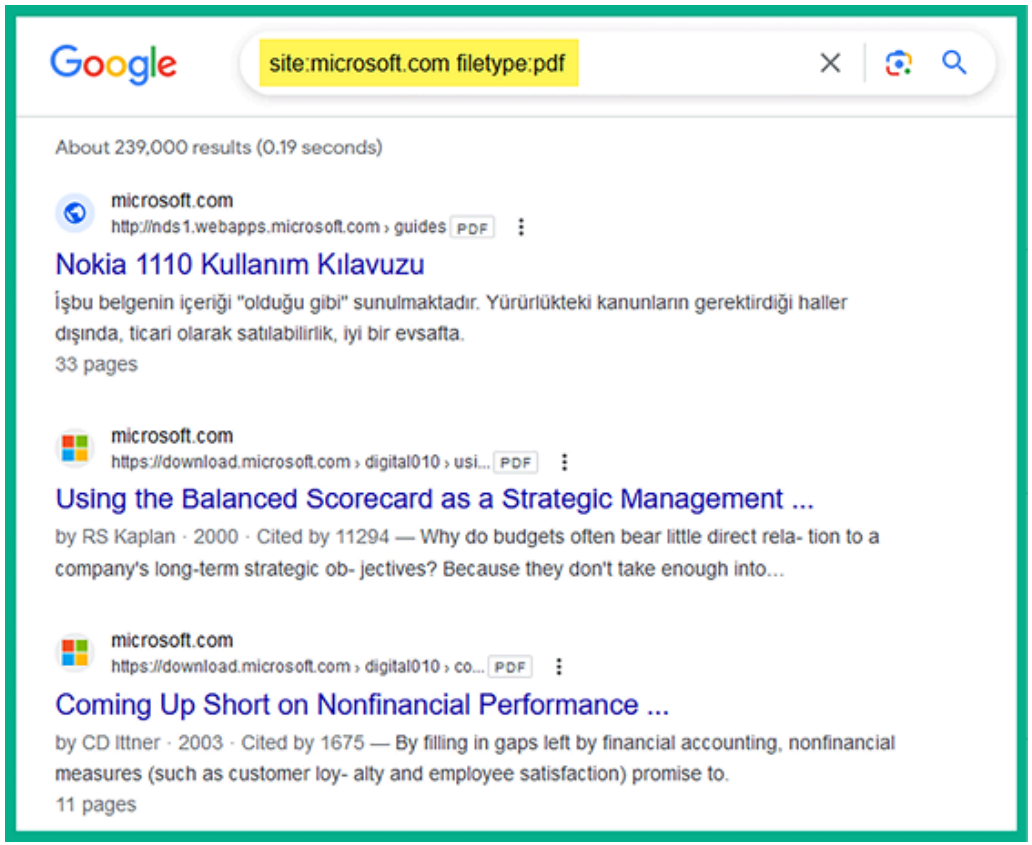


Figure 5.4: Specific filetype

As shown in the preceding screenshot, including the `filetype:` syntax helps us to filter the search results to display any files that are either intentionally or unintentionally leaked by the targeted organization.

- To discover specific directories that contain sensitive keywords on their title pages, use the `site:domain-name intitle:keyword` syntax, as shown here:

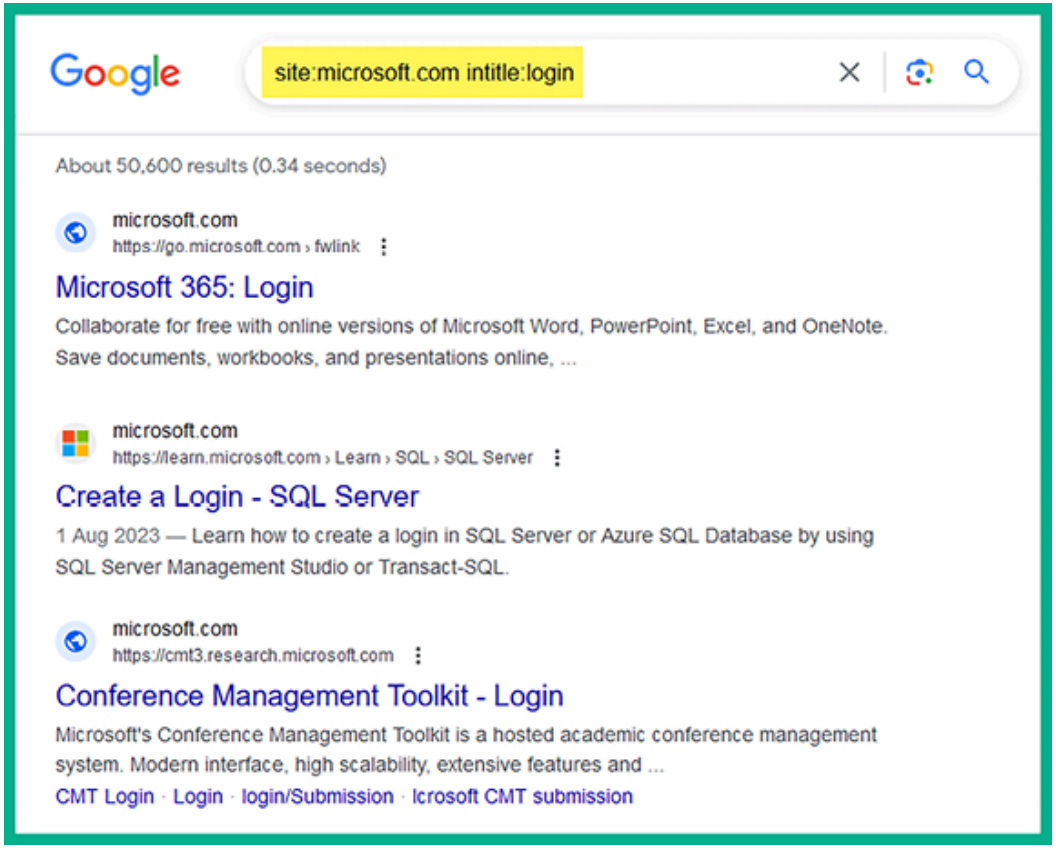


Figure 5.5: Using intitle keywords

As shown in the preceding screenshot, using the `login` keyword as the `intitle:` parameter is useful for displaying the login portals of the targeted domain.

- To find sub-domains of a targeted organization, use the `site:domain-name -www` syntax to exclude the `www` parameter, as shown here:

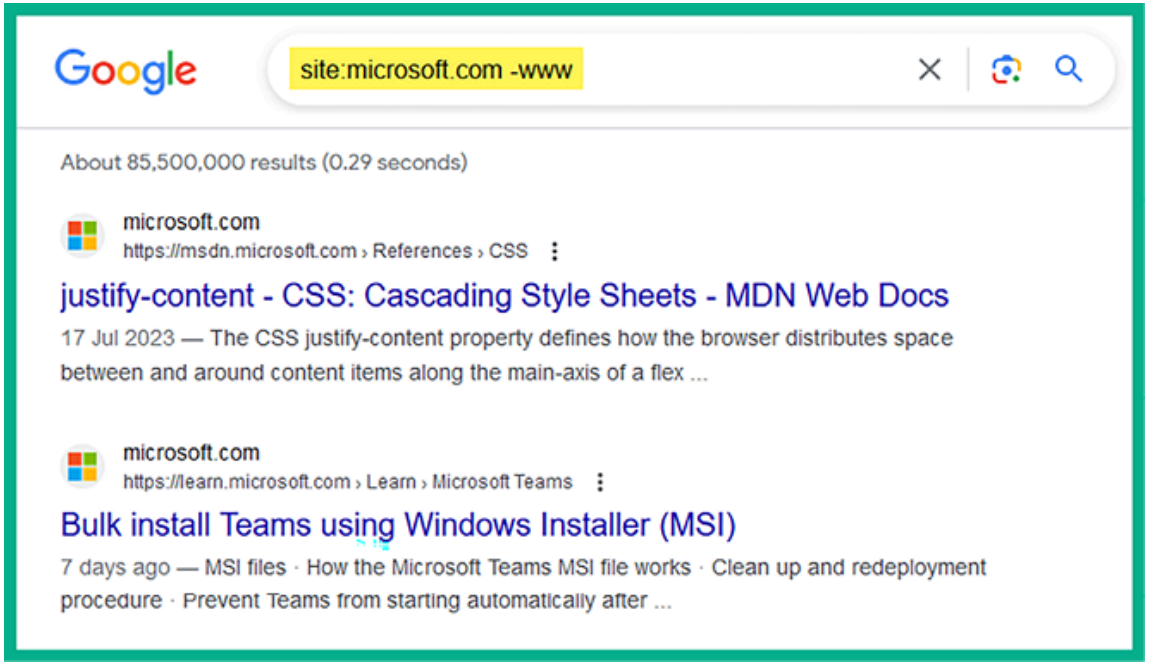


Figure 5.6: Excluding the www parameter

Using this technique is a good way to remove specific sub-domains and URLs from your search results.

In addition, you can use the `site:*.domain.com -site:www.domain.com` syntax to find sub-domains of the targeted organization.

Furthermore, if you're not too sure how to use the advanced search operators on the Google search engine, you can simply head on over to the Google home page and click on **Settings** | **Advanced search** to open the **Advanced Search** menu, as shown here:

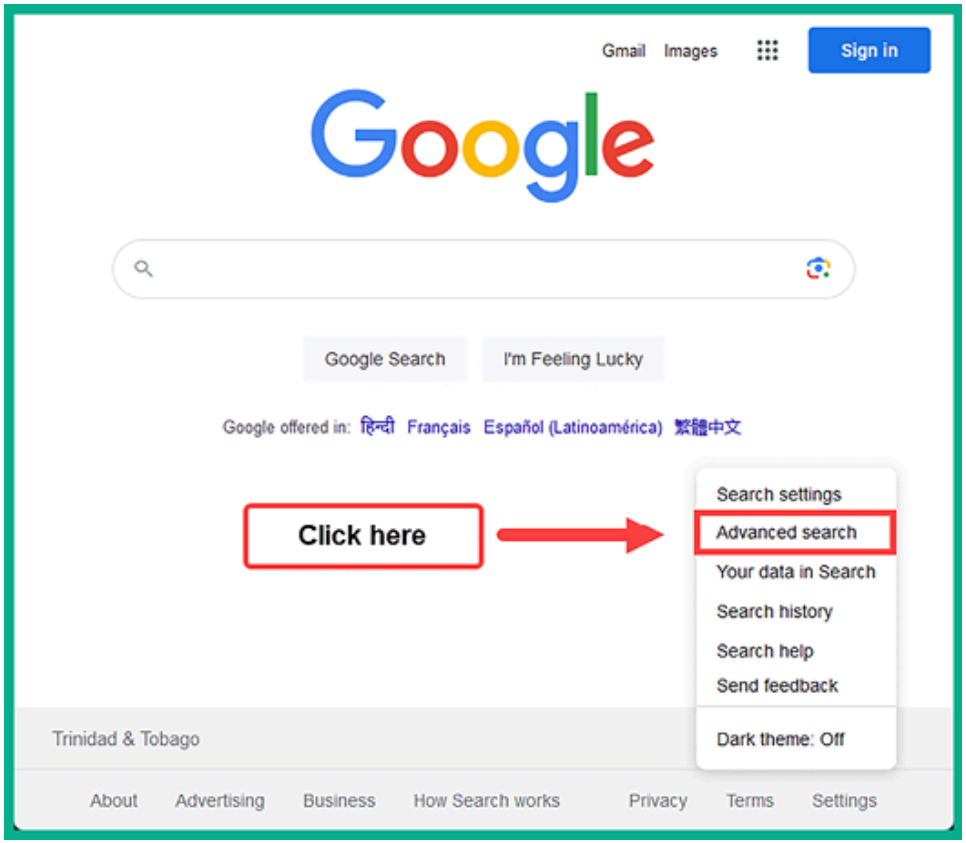


Figure 5.7: Advanced search

Google provides a very easy and simple method to enable users to perform advanced searching and filtering, without having to know the advanced search operators, as shown here:

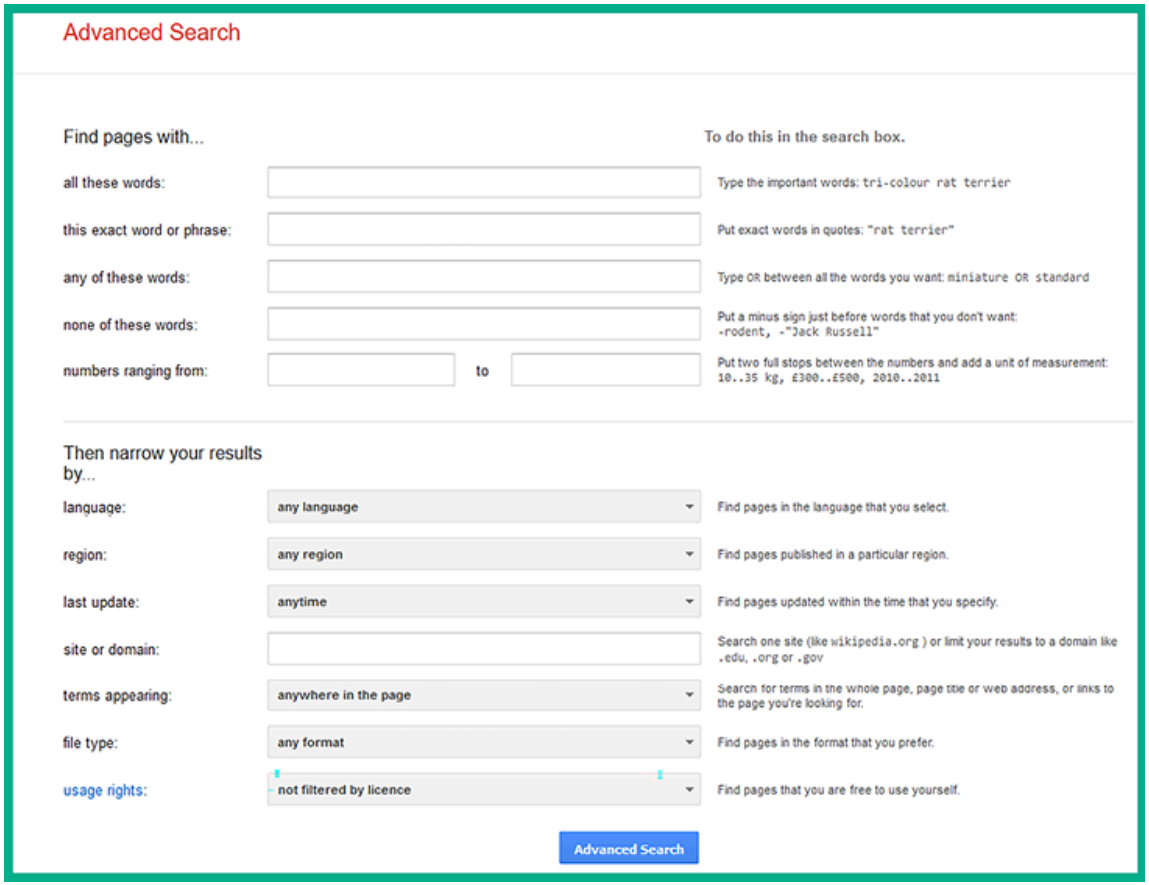


Figure 5.8: Advanced search details

Once you've filled in the necessary details and clicked on the **Advanced Search** button, Google will automatically insert the appropriate search operations needed to perform advanced searches.

While there are so many possibilities when using advanced Google search operators, it can be a bit overwhelming. The **Google Hacking Database** (**GHDB**) is maintained by the creators of Kali Linux, **Offensive Security** (**https://www.offsec.com/**), and can be found at **https://www.exploit-**

db.com/google-hacking-database. The GHDB is a website that contains a list of various Google dorks (advanced search operators), which are used to find very sensitive information and resources on the internet using Google Search:

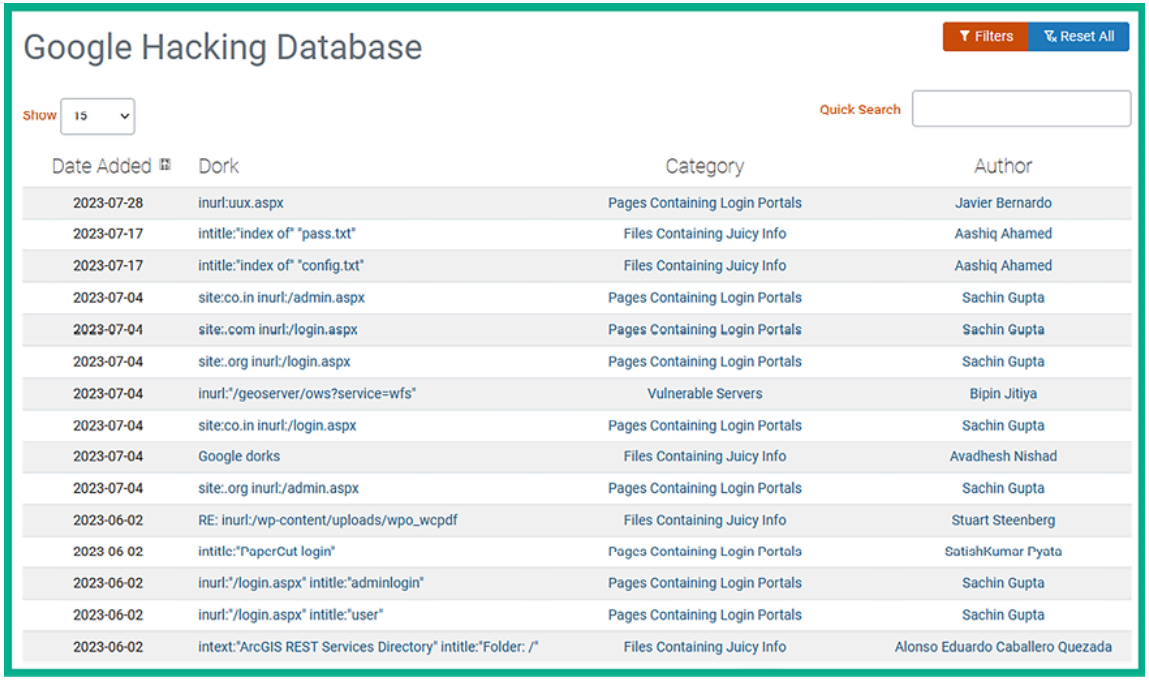


Figure 5.9: Advanced search results

As shown in the preceding screenshot, the GHDB is regularly updated with new search syntax to help users discover vulnerable services and sensitive directories. A word of caution, though – please be very mindful and careful when lurking around using Google hacking techniques. Do not use the information you find for malicious purposes or to cause harm to a system or network.

Having completed this section, you have learned how ethical hackers and penetration testers can leverage the power of Google Search to discover hidden direc-

tories and resources. In the next section, you will learn how to discover exposed assets owned by organizations.

Domain reconnaissance

Domain reconnaissance involves collecting information about a target-owned domain, which helps cybercriminals, ethical hackers, and penetration testers to identify whether the targeted organization has any exposed systems and network infrastructure that can be leveraged when planning a future attack. In addition, it helps ethical hackers and penetration testers to determine the external attack surface of an organization, that is, identifying all the internet-facing systems, their operating systems, open ports, and running services with the intention of discovering security vulnerabilities that can be exploited by real attackers. Domain reconnaissance can be classified as active reconnaissance if the ethical hacker or penetration tester is retrieving the domain records from a DNS server that's owned by the target. However, with passive information gathering, the information is collected from other trusted sources that are not directly linked to the target.

This helps ethical hackers to determine whether their targets are unintentionally exposing vulnerable systems, services, and applications on the internet, and how a threat actor can leverage the information to perform a cyber-attack. Over the following sub-sections, you'll learn how to collect domain registration details, enumerate DNS records, attempt to transfer zone records from a vulnerable DNS server, and automate domain reconnaissance techniques.

Collecting WHOIS data

What if you could access a database that contains the records of registered domains on the internet? Many domain registrars allow the general public to view publicly available information about registered domains. This information can be found on various WHOIS databases on the internet.

The following is a list of various types of information that can be collected from WHOIS databases:

- Registrant contact information
- Administrative contact information
- Technical contact information
- Name servers
- Important dates, such as registration, update, and expiration dates
- Registry domain ID
- Registrar information

Accessing a WHOIS database is quite simple: you can use your favorite internet search engine to find various WHOIS databases, such as the following:

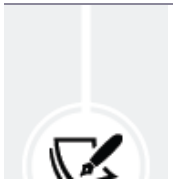
- **https://who.is/**
- **https://www.whois.com/**
- **https://lookup.icann.org/**
- **https://whois.domaintools.com/**

Within Kali Linux, you will find a pre-installed WHOIS tool, which enables penetration testers to perform a WHOIS lookup directly on the Terminal. To perform a WHOIS lookup on a targeted domain, open the Terminal on Kali Linux and execute the `whois <domain-name>` commands to begin a search, as shown here:

```
kali@kali:~$ whois microsoft.com
   Domain Name: MICROSOFT.COM
   Registry Domain ID: 2724960_DOMAIN_COM-VRSN
   Registrar WHOIS Server: whois.markmonitor.com
   Registrar URL: http://www.markmonitor.com
   Updated Date: 2023-08-18T16:15:54Z
   Creation Date: 1991-05-02T04:00:00Z
   Registry Expiry Date: 2025-05-03T04:00:00Z
   Registrar: MarkMonitor Inc.
   Registrar IANA ID: 292
   Registrar Abuse Contact Email: abusecomplaints@markmonitor.com
   Registrar Abuse Contact Phone: +1.2086851750
   Domain Status: clientDeleteProhibited https://icann.org/epp#clientDeleteProhibited
   Domain Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited
   Domain Status: clientUpdateProhibited https://icann.org/epp#clientUpdateProhibited
   Domain Status: serverDeleteProhibited https://icann.org/epp#serverDeleteProhibited
   Domain Status: serverTransferProhibited https://icann.org/epp#serverTransferProhibited
   Domain Status: serverUpdateProhibited https://icann.org/epp#serverUpdateProhibited
   Name Server: NS1-39.AZURE-DNS.COM
   Name Server: NS2-39.AZURE-DNS.NET
   Name Server: NS3-39.AZURE-DNS.ORG
   Name Server: NS4-39.AZURE-DNS.INFO
   DNSSEC: unsigned
   URL of the ICANN Whois Inaccuracy Complaint Form: https://www.icann.org/wicf/
>>> Last update of whois database: 2023-08-22T13:08:06Z <<<
```

Figure 5.10: The WHOIS search

As shown in the preceding screenshot, the WHOIS tool was able to retrieve publicly available information about the targeted domain by simply asking a trusted online source. Keep in mind that, as the need for online privacy increases around the world, domain owners are paying a premium fee to ensure their contact and personal information is not revealed by WHOIS databases to the general public. This means that you will not commonly find personal contact information for domains that are no longer being revealed on WHOIS databases if the domain owner pays the premium for additional privacy features.



WHOIS tools query databases that store information about domain ownership, registration dates, expiration dates, and contact details of domain owners. However, the level of detail available can vary sig-

nificantly based on the domain registrar's policies and privacy settings chosen by the domain owner.

However, do not pass this tool aside as there are still many organizations around the world that do not always value online privacy. Due to the lack of security awareness and negligence of many people and organizations, threat actors and penetration testers can exploit this vulnerability to collect OSINT on their targets.

Performing DNS enumeration

DNS is an application-layer protocol that enables a system such as a computer to resolve a hostname to an IP address. While there are so many devices on a network, especially on the internet, remembering the IP addresses of web servers can be quite challenging. Using DNS, a system administrator can configure each device with both an IP address and a hostname. Using a hostname is a lot easier to remember, such as `www.packtpub.com` or `www.google.com`. However, do you know the IP addresses of the servers that are hosting these websites for Packt and Google? You probably don't, and that's okay because, on the internet, there is a hierarchy of DNS servers that contain the records of public hostnames and their IP addresses. These are known as root DNS servers.

To learn more about root DNS servers, please visit `https://www.cloudflare.com/learning/dns/glossary/dns-root-server/`. More information on types of DNS servers can be found at `https://www.cloudflare.com/learning/dns/dns-server-types/`.

A DNS server is like a traditional telephone directory, with a list of people and their telephone numbers. On a DNS server, you can find records of the hostnames

of servers and devices, as well as their associated IP addresses. Many popular internet companies, such as Cisco, Google, and Cloudflare, have set up many public DNS servers around the internet, which contain the records of almost every public domain name on the internet.

To get a better understanding of how a client device such as a computer uses DNS to resolve a domain name, let's take a look at the following scenario:

1. Imagine you want to view the webpage on `www.example.com` on your computer, so you decide to open the web browser and enter `www.example.com` within the address bar and hit enter to connect to the web server.
2. Your computer will check the local DNS cache to determine whether the IP address of `www.example.com` is known already due to a previous connection. If the IP address of `www.example.com` is found within the local cache, the computer will establish a connection to the destination server.
3. If the IP address is not found within the local DNS cache of the client, the client sends a **DNS Query** message to the DNS server, requesting the IP address of the hostname (`www.example.com`), as shown here:

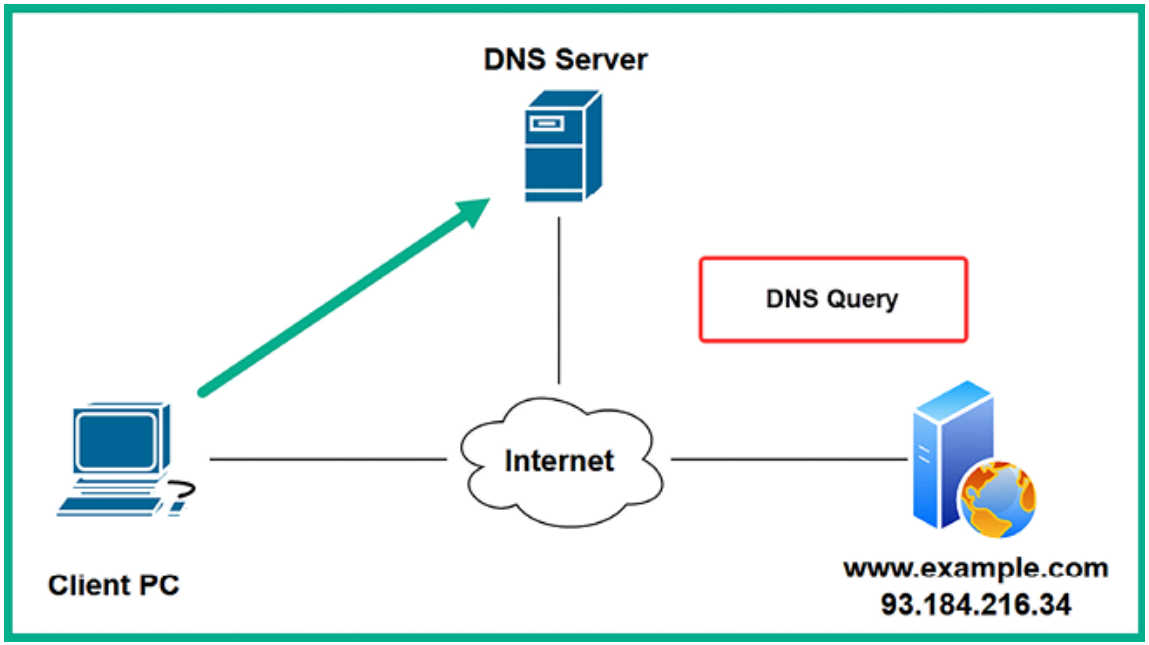


Figure 5.11: DNS query

4. The DNS server will check its records and respond to the client with a non-authoritative **DNS reply** message, providing the client with the IP address of the hostname, as shown here:

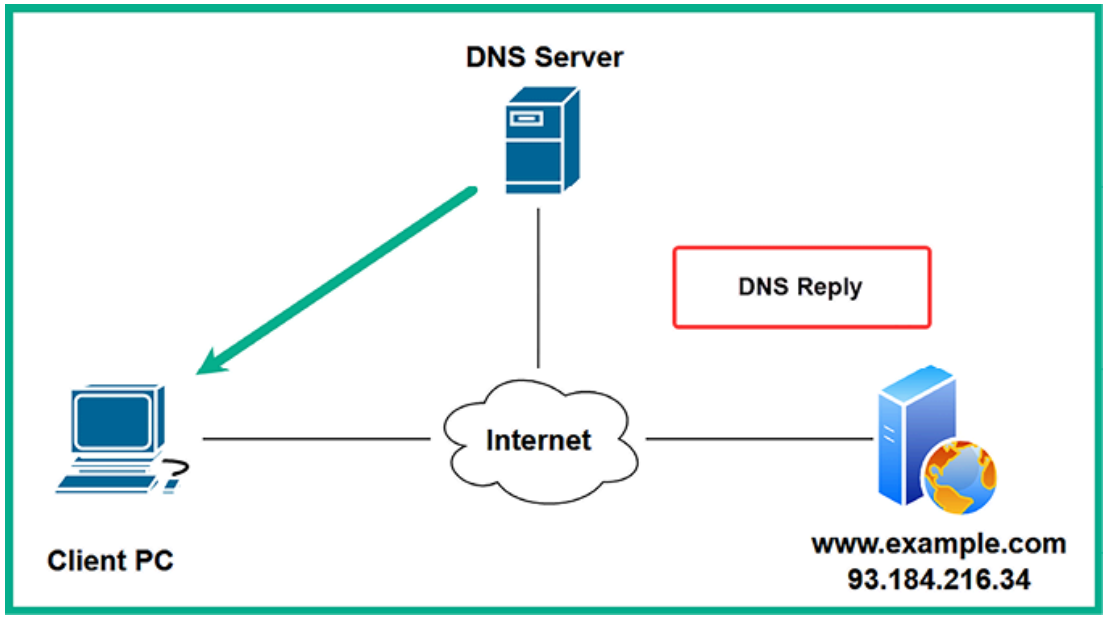


Figure 5.12: DNS response

If the DNS server does not have the requested records for the hostname, it performs a recursive DNS lookup to retrieve the DNS records either from other DNS servers on the internet or the root DNS server. When the client receives the IP address from the DNS reply from the DNS server, the client stores the IP address-to-host-name mapping within the local DNS cache for future reference.

5. The client uses the IP address from the **DNS reply** to connect to `www.example.com` on the internet, as shown here:

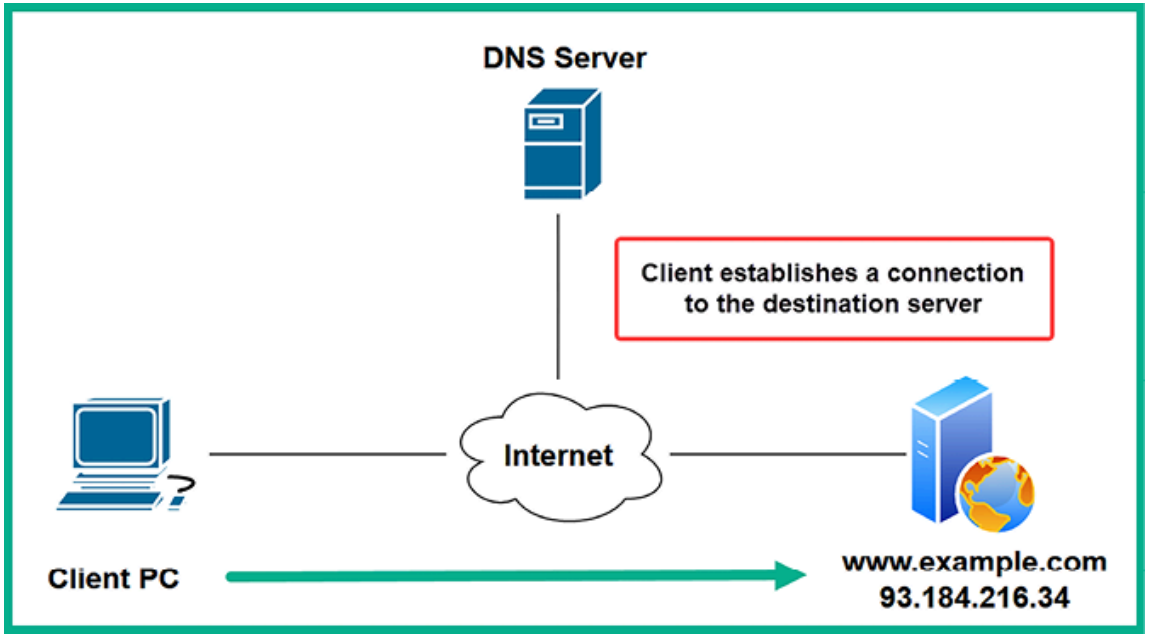


Figure 5.13: Client establishing a connection

There are many public DNS servers on the internet; some are created by threat actors with malicious intentions, such as redirecting unaware users to malicious websites for social engineering purposes. As a result, I recommend using a trusted DNS provider on all of your networking devices, security appliances, servers, and computers to improve your organization's online safety.

The following are some popular DNS servers on the internet:

- Cloudflare: **https://1.1.1.1/**
- Quad 9: **https://www.quad9.net/**
- Cisco OpenDNS: **https://www.opendns.com/**
- Google Public DNS: **https://developers.google.com/speed/public-dns**

Additionally, DNS servers not only resolve a hostname to an IP address, but they also contain various types of records with information about a domain, such as the following:

- `A` : This record maps a hostname to an IPv4 address.
- `AAAA` : This is used to map a hostname to an IPv6 address.
- `NS` : This is used for specifying the name servers for a domain.
- `MX` : This specifies the mail exchange or email servers for the domain.
- `PTR` : This record maps an IPv4 or IPv6 address to a hostname.
- `CNAME` : This is used to specify an alias for another record.
- `RP` : This report contains the responsible person for the domain.
- `SOA` : This record specifies the authority for the domain.
- `SRV` : This record contains the service records such as port numbers for specific services on the domain.
- `TXT` : This record allows the domain owner to specify a text record. Commonly used for verification of ownership for a domain.

You're probably wondering what learning about DNS has to do with passive reconnaissance and OSINT as a penetration tester. As an aspiring penetration tester, DNS enumeration is the technique of probing specific DNS records for a targeted domain to retrieve information about an organization's internet-facing assets and identify any security vulnerabilities that can assist in planning a cyber-attack. Performing DNS enumeration is simply requesting the DNS records of a targeted domain from a public DNS server on the internet, and then analyzing the collected information to create intelligence and better understand how an adversary can leverage the intelligence to compromise the targeted organization.

Within Kali Linux, you will find many DNS analysis tools to help ethical hackers and penetration testers efficiently collect and analyze DNS records of a targeted

domain. While the choice of tool usually depends upon the personal preference of the penetration tester, I strongly urge you to try all the available tools to better understand which ones work best for you.

To get started with using **DNSRecon** for DNS enumeration, please use the following instructions:

1. Power on the **Kali Linux** virtual machine and log in.
2. Next, open the **Terminal** and use the following commands to retrieve the DNS records for a targeted domain:

```
kali@kali:~$ dnsrecon -d microsoft.com -n 1.1.1.1
```

The following screenshot shows DNSRecon was able to retrieve the public DNS records for the Microsoft.com domain from Cloudflare's public DNS server:

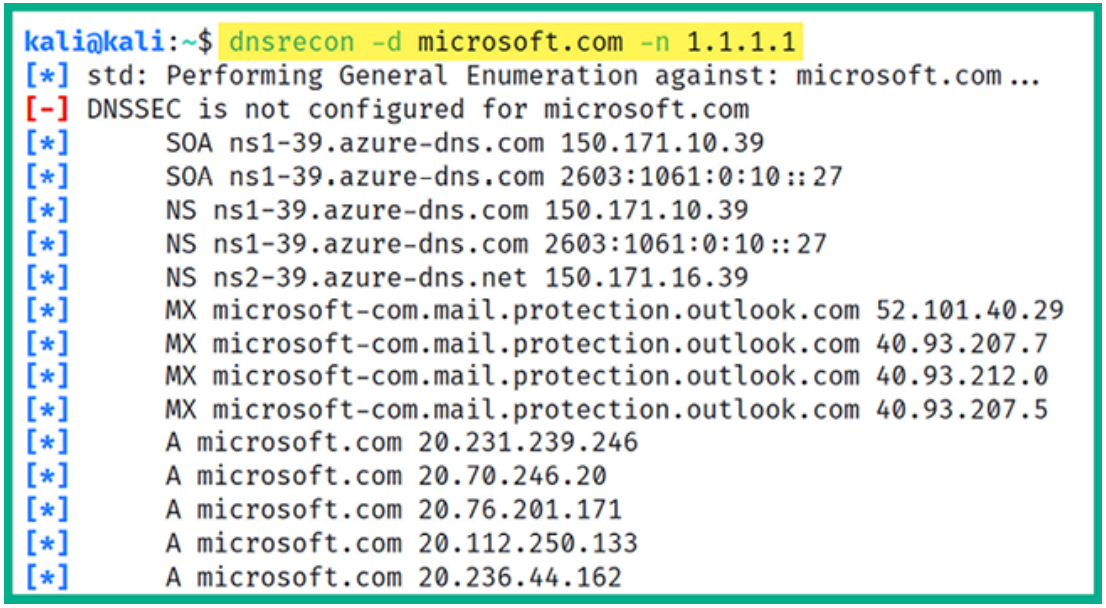


Figure 5.14: DNS enumeration

Using the `-d` syntax enables you to specify the targeted domain, while the `-n` syntax enables you to specify a name server to query.

As shown in the preceding screenshot, DNSrecon was able to retrieve various DNS records for the targeted domain, such as the `A`, `NS`, `MX`, and `SOA` records. An ethical hacker and penetration tester can leverage the information collected to identify the public IP address of additional assets owned by the target.

3. In addition, DNSrecon was able to enumerate the SRV records of the targeted domain, as shown here:

```
[*] Enumerating SRV Records
[+]      SRV _sipfederationtls._tcp.microsoft.com sipfed.online.lync.com 52.112.127.17 5061
[+]      SRV _xmpp-server._tcp.microsoft.com sipdog3.microsoft.com 131.107.1.47 5269
[+]      SRV _sip._tls.microsoft.com sipdir.online.lync.com 52.112.64.11 443
[+]      SRV _sip._tls.microsoft.com sipdir.online.lync.com 2603:1037::b 443
[+] 4 Records Found
```

Figure 5.15: Enumerated SRV records

As shown in the preceding screenshot, the end of each line indicates the open port number for each service. Identifying open ports helps penetration testers determine running services and points of entry into a targeted system.

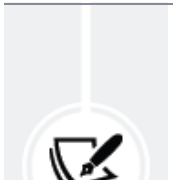
To learn more about DNSrecon and its additional features, use the `dnsrecon -h` and `man dnsrecon` commands on Kali Linux.

Having completed this exercise, you have learned how to enumerate DNS records from public DNS servers for a targeted domain. Next, you will learn how to exploit a vulnerable DNS server to extract sensitive DNS records.

Exploiting DNS zone transfer

DNS zone transfer allows the zone records from one DNS server to be copied from a master DNS server onto another DNS server over a network. DNS zone transfers provide redundancy such that the DNS records are replicated between a primary and secondary DNS server on a network and load-balancing DNS queries between multiple DNS servers with the same zone records. Sometimes, an IT professional may forget to secure their DNS server and implement security controls to prevent the zone records from being copied to unauthorized DNS servers. If a threat actor were to successfully perform a DNS zone transfer on a targeted organization, the adversary would be able to retrieve both public and private DNS records, which helps the attacker to identify critical systems on the internal network of the target.

In another scenario, the targeted organization may not separate their internal and external namespaces from each other on their DNS servers for the company. This type of misconfiguration on DNS servers can lead to a future DNS zone transfer attack. While nowadays, it's less likely to discover a target's DNS server with this security vulnerability, it's still important for both ethical hackers and penetration testers to understand how adversaries are able to discover and exploit this security flaw.



To learn more about the security vulnerability within DNS zone transfer, please visit **https://www.cisa.gov/news-**

events/alerts/2015/04/13/dns-zone-transfer-axfr-requests-may-leak-domain-information.

However, as security training is applied to almost every field within IT courses and certifications, the upcoming generation of IT professionals is usually made aware of this security flaw to ensure their systems and networks are always secure. Hence, the possibility of a poorly configured DNS server may be almost nonexistent since, as an aspiring penetration tester, you should leave no stone unturned and always test for everything within your scope of a penetration test on your target.

The awesome folks at **DigiNinja** (https://digi.ninja/) set up an amazing environment to better understand how to test for DNS zone transfer vulnerabilities. In addition, they have made their online platform free to the public so anyone can learn more about the security vulnerabilities of misconfigured DNS servers.

To get started with this exercise, please use the following instructions:

1. Power on your **Kali Linux** virtual machine and log in.
2. Open the **Terminal** and use the `host` command to retrieve the DNS records of `zonetransfer.me`, as shown here:

```
kali@kali:~$ host zonetransfer.me
zonetransfer.me has address 5.196.105.14
zonetransfer.me mail is handled by 20 ASPMX4.GOOGLEMAIL.COM.
zonetransfer.me mail is handled by 10 ALT1.ASPMX.L.GOOGLE.COM.
zonetransfer.me mail is handled by 20 ASPMX3.GOOGLEMAIL.COM.
zonetransfer.me mail is handled by 0 ASPMX.L.GOOGLE.COM.
zonetransfer.me mail is handled by 10 ALT2.ASPMX.L.GOOGLE.COM.
zonetransfer.me mail is handled by 20 ASPMX2.GOOGLEMAIL.COM.
zonetransfer.me mail is handled by 20 ASPMX5.GOOGLEMAIL.COM.
```

Figure 5.16: Gathering DNS records

As shown in the preceding screenshot, various DNS records were retrieved, such as the `A` and `MX` records.

3. Next, let's attempt to retrieve the `NS` records for the targeted domain. To do so, use the `host -t ns zonetransfer.me` commands, as shown here:

```
kali@kali:~$ host -t ns zonetransfer.me
zonetransfer.me name server nsztm2.digi.ninja.
zonetransfer.me name server nsztm1.digi.ninja.
```

Figure 5.17: Retrieving name servers

As shown in the preceding screenshot, the targeted domain has two name servers, which are `nsztm1.digi.ninja` and `nsztm2.digi.ninja`. We can proceed to check each of these name servers to determine whether they are misconfigured for unauthorized zone transfer.

4. Next, let's query the `nsztm1.digi.ninja` name server to identify whether it's vulnerable to DNS zone transfer and retrieve the zone records. Use the following command:

```
kali@kali:~$ host -l zonetransfer.me nsztm1.digi.ninja
```

The following is a screenshot of all the DNS records that were obtained from the `nsztm1.digi.ninja` name server for the targeted domain:

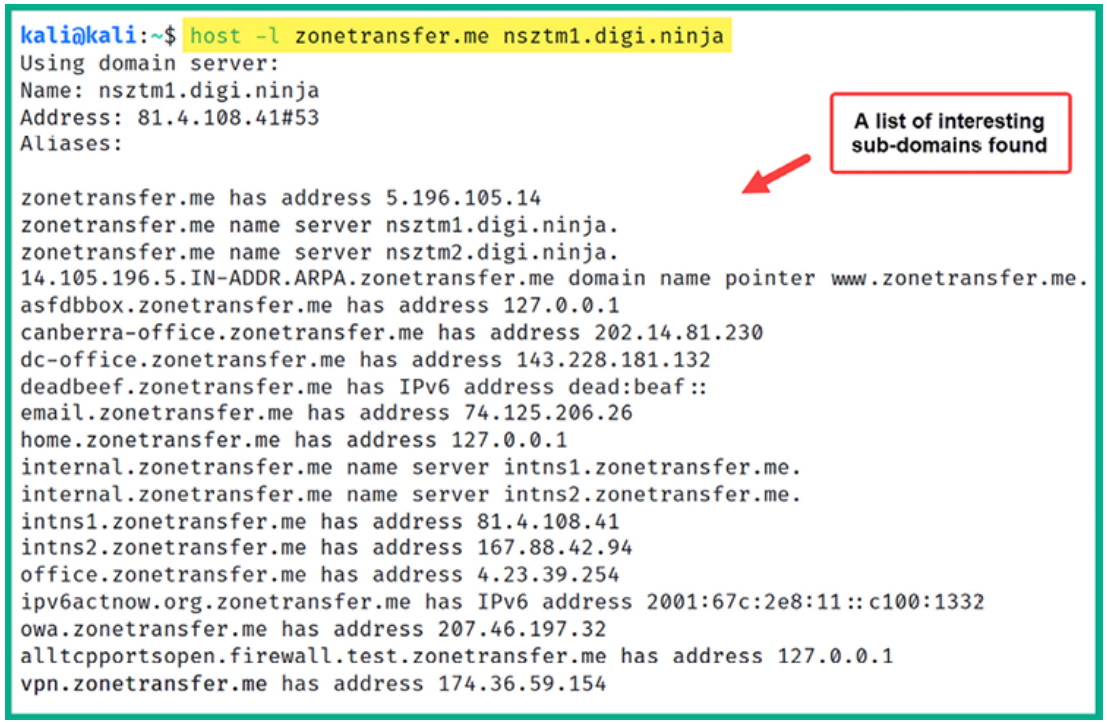


Figure 5.18: Retrieving DNS records

As shown in the preceding screenshot, there are many interesting hostnames, and their corresponding IP addresses were retrieved. These hostnames may not be intentionally exposed to the internet by the targeted organization but as a result of poorly configured DNS server settings, they were.

Be sure to query all the name servers for a given domain – sometimes, one server may be misconfigured even though the others

are secured.

5. Next, to automate the DNS analysis and perform DNS zone transfer on a targeted domain, use the **DNSenum** tool with the following commands:

```
kali@kali:~$ dnsenum zonetransfer.me
```

The DNSenum tool will attempt to retrieve all DNS records for the targeted domain and will attempt to perform DNS zone transfer using all the name servers that are found. The following screenshot shows that DNSenum was able to retrieve the zone records for the targeted domain:

```
Trying Zone Transfer for zonetransfer.me on nsztm2.digi.ninja ...
zonetransfer.me.                         7200     IN    SOA                (
zonetransfer.me.                         300      IN    HINFO        "Casio
zonetransfer.me.                         301      IN    TXT                (
zonetransfer.me.                         7200     IN    MX                 0
zonetransfer.me.                         7200     IN    MX                10
zonetransfer.me.                         7200     IN    A        5.196.105.14
zonetransfer.me.                         7200     IN    NS       nsztm1.digi.ninja.
zonetransfer.me.                         7200     IN    NS       nsztm2.digi.ninja.
_acme-challenge.zonetransfer.me.         301      IN    TXT                (
_acme-challenge.zonetransfer.me.         301      IN    TXT                (
_sip._tcp.zonetransfer.me.               14000    IN    SRV                0
14.105.196.5.IN-ADDR.ARPA.zonetransfer.me. 7200     IN    PTR      www.zonetransfer.me.
asfdbauthdns.zonetransfer.me.            7900     IN    AFSDB              1
asfdbbox.zonetransfer.me.                7200     IN    A         127.0.0.1
asfdbvolume.zonetransfer.me.             7800     IN    AFSDB              1
canberra-office.zonetransfer.me.         7200     IN    A        202.14.81.230
cmdexec.zonetransfer.me.                 300      IN    TXT                ";
contact.zonetransfer.me.                 2592000  IN    TXT                (
dc-office.zonetransfer.me.               7200     IN    A        143.228.181.132
deadbeef.zonetransfer.me.                7201     IN    AAAA     dead:beaf::
dr.zonetransfer.me.                      300      IN    LOC                53
```

Figure 5.19: Zone transfer using DNSenum

DNSenum was able to retrieve additional zone records, as shown here:

```
email.zonetransfer.me.              2222    IN    NAPTR            (
email.zonetransfer.me.              7200    IN    A        74.125.206.26
Hello.zonetransfer.me.              7200    IN    TXT             "Hi
home.zonetransfer.me.               7200    IN    A         127.0.0.1
Info.zonetransfer.me.               7200    IN    TXT              (
internal.zonetransfer.me.           300     IN    NS       intns1.zonetransfer.me.
internal.zonetransfer.me.           300     IN    NS       intns2.zonetransfer.me.
intns1.zonetransfer.me.             300     IN    A        81.4.108.41
intns2.zonetransfer.me.             300     IN    A        52.91.28.78
office.zonetransfer.me.             7200    IN    A        4.23.39.254
ipv6actnow.org.zonetransfer.me.     7200    IN    AAAA     2001:67c:2e8:11::c100:1332
owa.zonetransfer.me.                7200    IN    A        207.46.197.32
```

Figure 5.20: DNS records

As you can imagine, the collected information can be leveraged by both adversaries and ethical hackers to discover additional assets that are owned by the targeted organization, and identify their hostnames and IP addresses.

Having completed this exercise, you have learned how to perform DNS enumeration and zone transfer as an ethical hacker and penetration tester. Next, you will learn how to automate OSINT collection using SpiderFoot.

Automation using SpiderFoot

SpiderFoot is a popular OSINT tool that helps ethical hackers, penetration testers, and cybersecurity researchers automate their processes and workloads when gathering domain intelligence about their targets. Rather than running multiple tools or spending a lot of time using manual domain reconnaissance techniques, SpiderFoot can reduce the time on collecting and analyzing domain-related information about a target. This tool provides excellent visualization of all data gath-

ered in the form of graphs and tables (as we will demonstrate in this section), which helps you easily read and interpret the data that's been collected.

To get started with SpiderFoot, please use the following instructions:

1. Power on the **Kali Linux** virtual machine and ensure it has an active internet connection.
2. Next, open the **Terminal** and use the following commands to launch the SpiderFoot web interface:

```
kali@kali:~$ spiderfoot -l 0.0.0.0:1234
```

The following screenshot shows the execution of the preceding commands:

Figure 5.21: Running SpiderFoot

As shown in the preceding screenshot, the `-l` syntax specifies the IP address and port number for the SpiderFoot web interface, where `0.0.0.0` specifies

all interfaces and `1234` is the open port for incoming connections to the SpiderFoot web interface.

3. Next, open the web browser within Kali Linux and go to `http://127.0.0.1:1234/` to access the SpiderFoot web interface, as shown here:

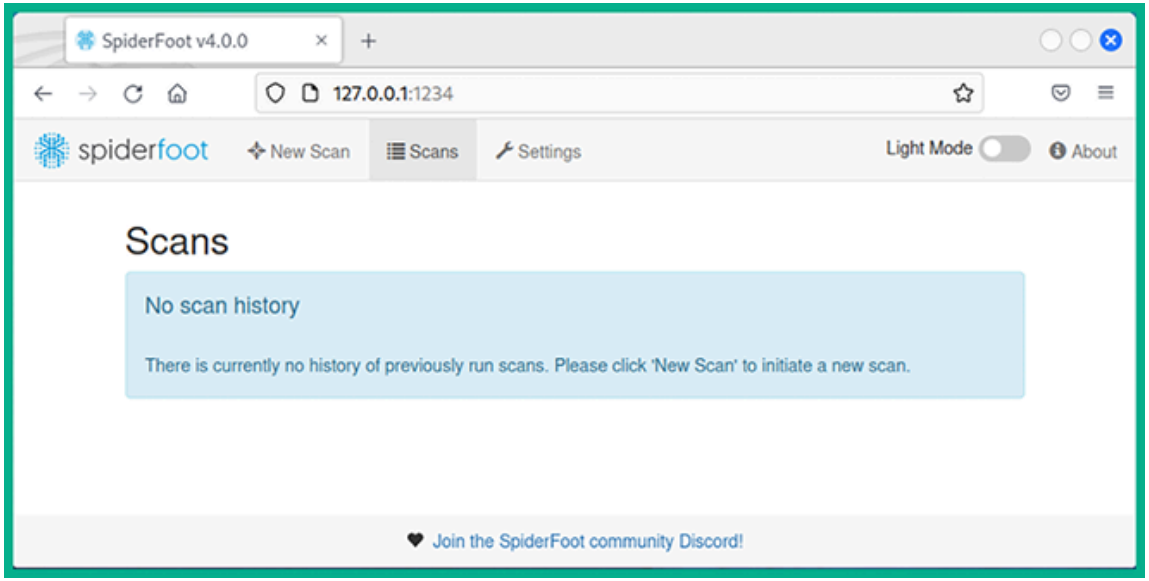


Figure 5.22: SpiderFoot web interface

4. Next, to automate the OSINT data collection and analysis, click on **New Scan**, set a **Scan Name** with a **Scan Target** and use the **Passive** option, then click on **Run Scan Now**:

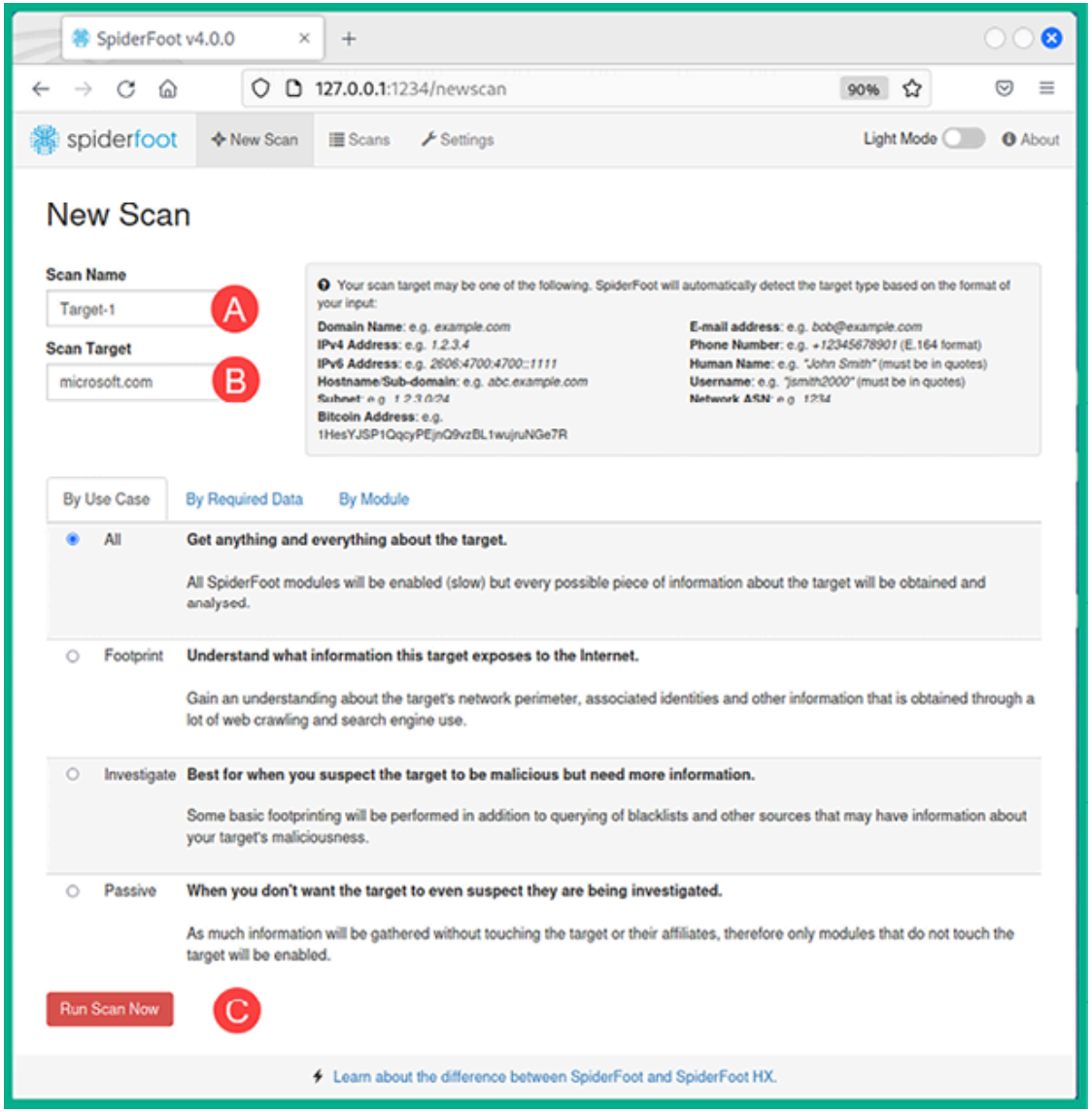


Figure 5.23: New scan

5. SpiderFoot begins to collect and analyze data from multiple data sources on the internet about the targeted organization or domain, as shown here:

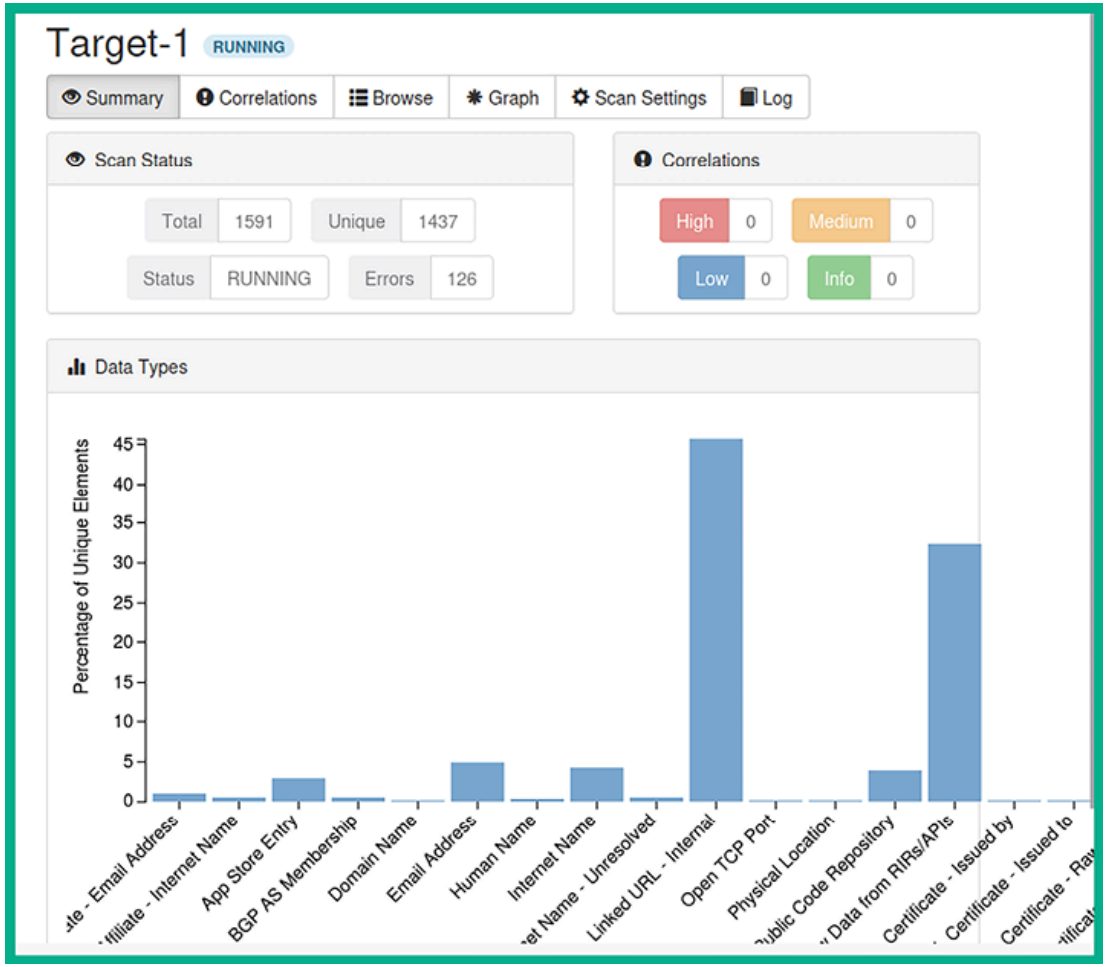


Figure 5.24: SpiderFoot summary

As shown in the previous screenshot, the bars within the graph increase as more data is collected for a specific category. For instance, as more email addresses are found on OSINT data sources, the bar that represents the email address category will rise.

6. Select the **Graph** tab to view how each data point is interconnected to the target domain, as shown here:

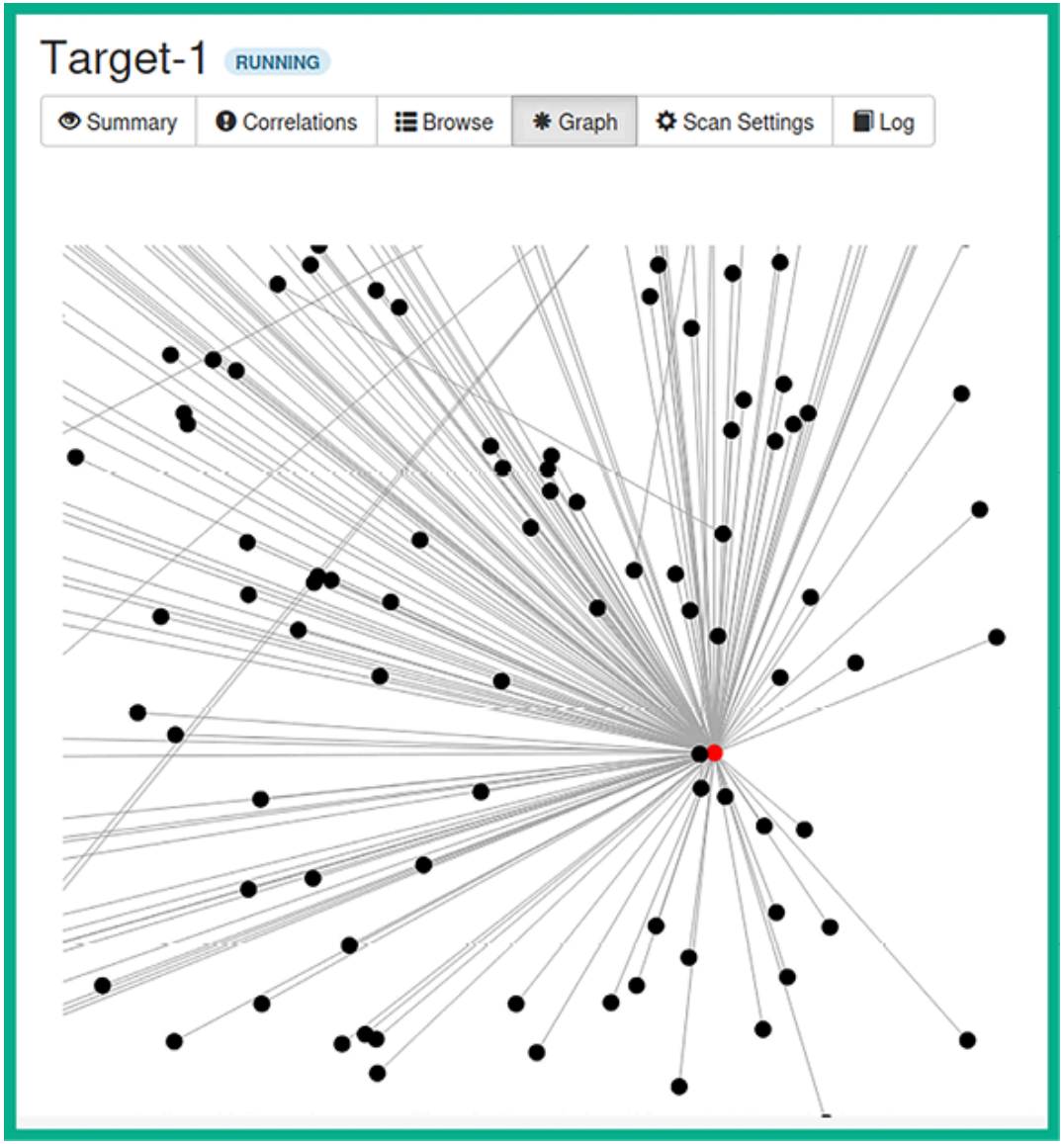


Figure 5.25: SpiderFoot graph

Clicking a data point shown in the preceding screenshot reveals a domain name, sub-domain, hostname, email address, or URL that's associated with the target. Each data point within the **Graph** section is commonly used by ethical

hackers and penetration testers to gain a better visualization of the attack surface of the target and to understand what type of data is being leaked on the internet that can be leveraged by a threat actor.

7. Next, to view the data that was collected based on categories, click on **Browse**, as shown in the following screenshot:

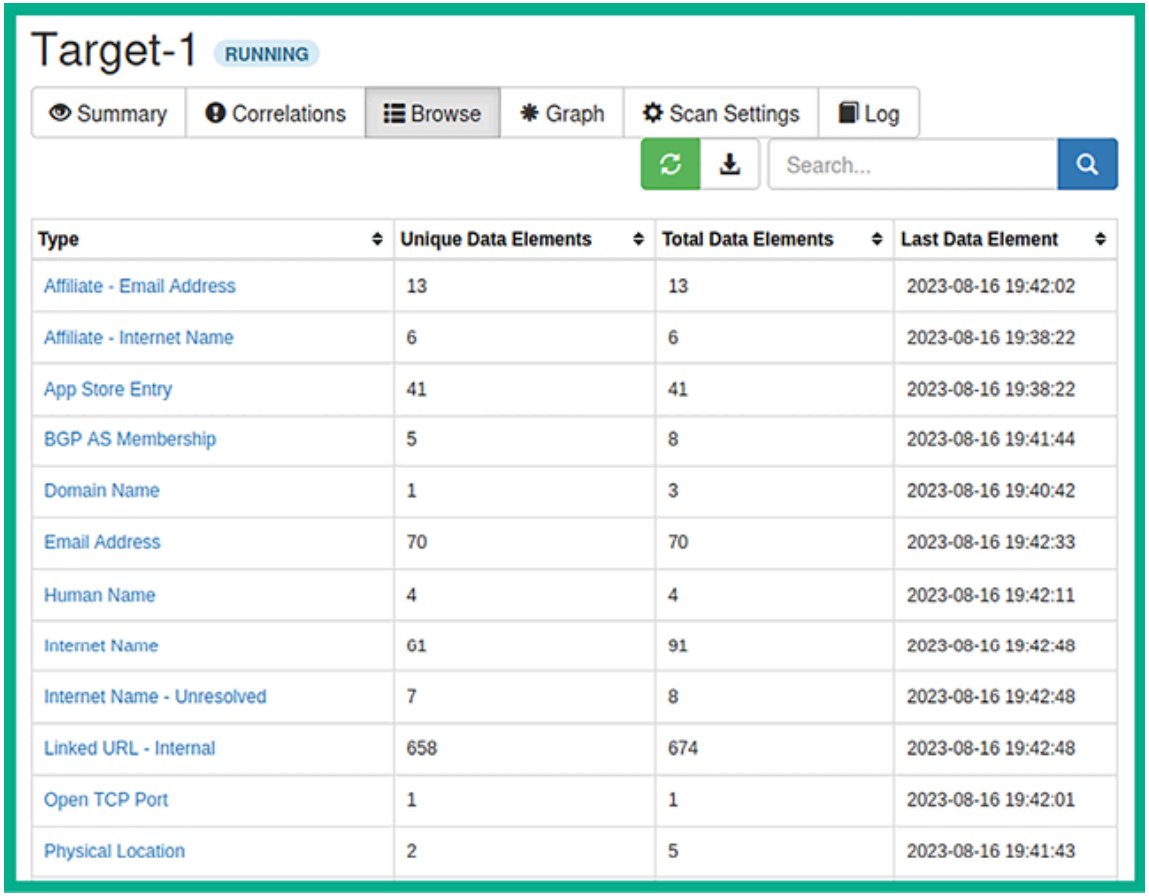


Figure 5.26: Viewing data

8. Next, click on the **Internet Name** category to see the data that was collected, as shown here:

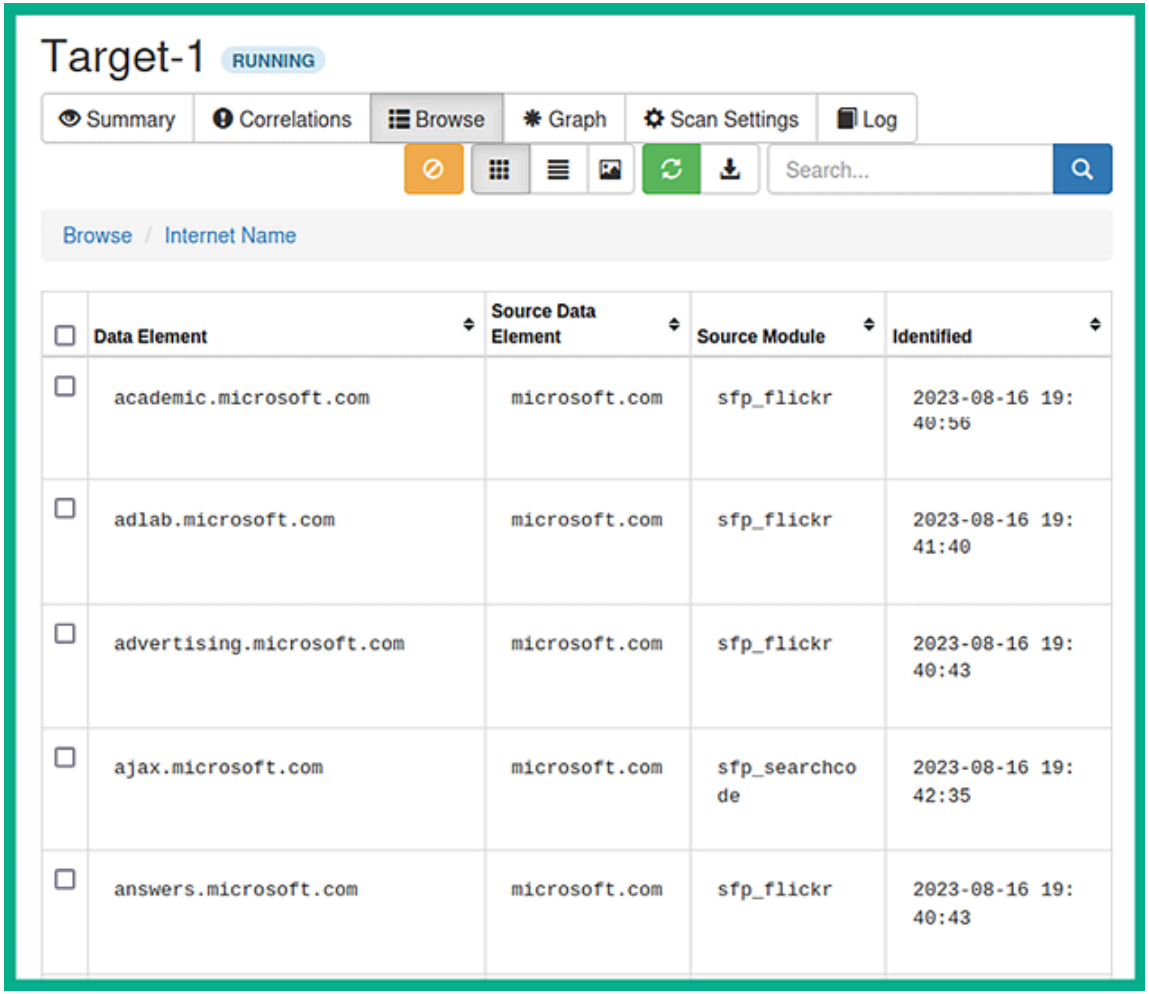


Figure 5.27: Viewing Intenet name category

As you can imagine, SpiderFoot can dig deeper until it gathers all the data about a targeted domain, inclusive of DNS information, formats the data into information,

and converts it into intelligence that can be leveraged by ethical hackers and penetration testers.

Having completed this section, you have gained the hands-on experience and skills to perform domain and DNS reconnaissance. In the next section, you will learn how to discover sub-domains using OSINT techniques.

Sub-domain harvesting

Every day, search engines such as Bing, Google, and Yahoo frequently learn and index new and existing websites to improve their search results. If a person searches for a company's website, you're likely to discover the primary domain, such as `example.com`. A lot of organizations create sub-domains for various reasons, but as an aspiring ethical hacker and penetration tester, discovering all the possible sub-domains of a targeted organization can lead to finding sensitive locations and resources, such as login portals and unintentionally exposed corporate directories, which may contain confidential files and resources.

In this section, you'll learn how to identify sub-domains using DNSMap and Sublist3r.

Enumeration with DNSMap

DNSMap works a bit differently from the tools we looked at in the previous sections. DNSMap attempts to enumerate the sub-domains of a targeted parent domain by querying a built-in wordlist within Kali Linux. DNSMap also has the capability of querying custom wordlists to identify sub-domains of a target. Once a sub-domain is found, DNSMap will also attempt to resolve the IP address automatically.

To get started using DNSMap, please use the following instructions:

1. Power on the **Kali Linux** virtual machine and ensure it has internet connectivity.
2. Next, open the **Terminal** and use the following commands to install the latest version of DNSMap on Kali Linux:

```
kali@kali:~$ sudo apt update
kali@kali:~$ sudo apt install dnsmap
```

3. Next, use the following commands to automate the discovery of sub-domains for a target using DNSMap:

```
kali@kali:~$ dnsmap microsoft.com
```

The following screenshot shows DNSMap is identifying the sub-domains of a targeted organization and is resolving each hostname/sub-domain to an IP address:

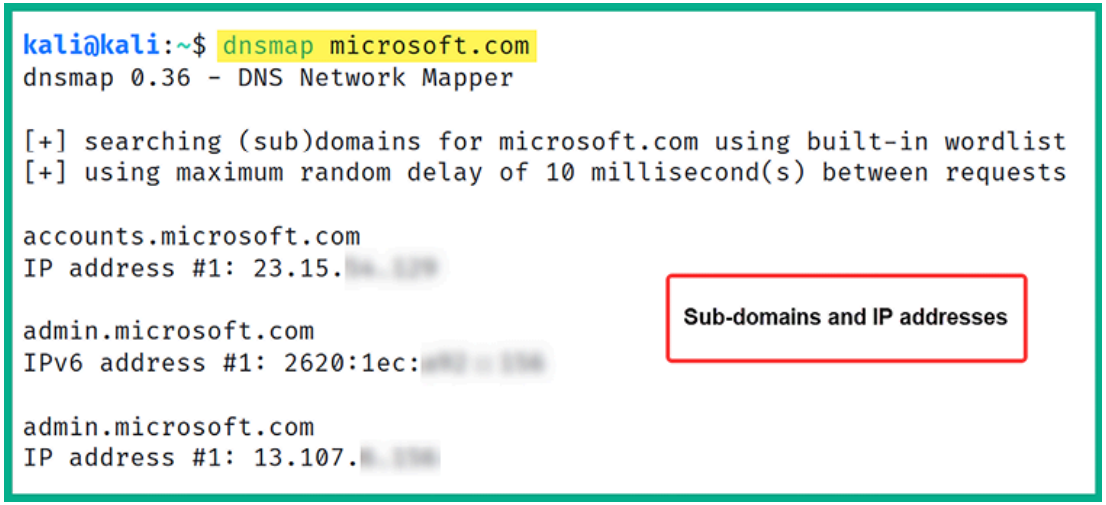


Figure 5.28: Discovering sub-domains

As a penetration tester, discovering the sub-domains of your target can lead to finding vulnerable web applications and even systems. Furthermore, such information can be used to build a better profile of your target.

Next, you will learn how to use another popular tool that leverages OSINT to gather the sub-domains of a targeted organization.

Sub-domain discovery with Knockpy

You can leverage the power of search engines to discover sub-domains by using the **Knockpy** tool. Knocky is a Python-based tool that is used to enumerate (extract/obtain) the sub-domains of a targeted public domain using OSINT techniques and data sources, such as search engines and other internet indexing platforms.

To get started using Knockpy, please use the following instructions:

1. Firstly, power on the **Kali Linux** virtual machine and ensure it has internet connectivity.
2. Next, open the **Terminal** and use the following commands to download and install the **Knockpy** application:

```
kali@kali:~$ sudo apt update
kali@kali:~$ sudo apt install -y knockpy
```

3. Next, use the following commands to perform sub-domain discovery on a targeted domain:

```
kali@kali: knockpy --recon --dns 8.8.8.8 -d microsoft.com
```

   The following screenshot shows the sub-domain discovery process with Knockpy:

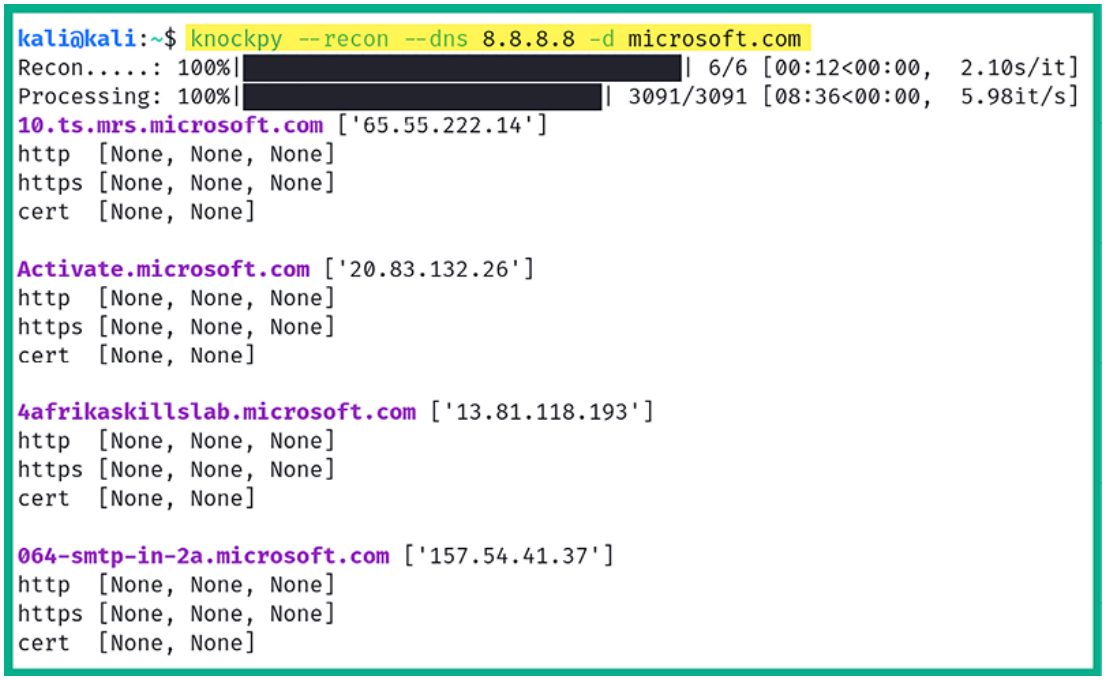


Figure 5.29: Sub-domain discovery with Knockpy

The `--recon` syntax specifies to perform sub-domain enumeration, `--dns` syntax enables you to specify a custom DNS server to query, and `-d` specifies the targeted domain.

Using the information that was found regarding sub-domains, penetration testers will need to check these sub-domains to determine where they lead, such as to a vulnerable web application or even a login portal for employees or customers.

Having completed this section, you have learned how to efficiently discover the sub-domains of a targeted organization. In the next section, you will learn how to use OSINT to identify the technical infrastructure of an organization.

Identifying organizational infrastructure

While many organizations think their network infrastructure is hidden behind their public IP address and that threat actors are unable to determine their internal infrastructure, threat actors use various OSINT techniques and tools to identify the systems and applications that are running within a targeted organization.

Over the next sub-sections, you will learn how organizations are leaking technical details about their internal network and how they can be leveraged by threat actors to improve their cyber-attacks.

Data leakage on job websites

Over the years, I've noticed many organizations leak a lot of data about their internal infrastructure and systems, which can help adversaries improve their plan of attack and identify security vulnerabilities within an organization by simply analyzing public information. For instance, a recruiter may post a vacancy on a job board or their company's website for job seekers. Quite often, the recruiter or job poster provides specific technical details about the organization's internal systems to help the job seeker determine whether the position is a good fit for their career development.

The following are the advantages of companies posting their technologies on recruitment websites:

- The potential candidate will have an idea of the environment and technologies they will be working with if they are successful during the interviewing process.
- The potential candidate can determine whether they have the skillset required for the job beforehand.

However, a threat actor can leverage the technical details found in a job post to determine the type of operating systems, applications and versions, networking, and security solutions that are running within the company. In addition, such information is usually public information and OSINT, which can be leveraged by adversaries to determine the attack surface and security vulnerabilities of the company. Accordingly, the following are the disadvantages of companies posting their technologies on recruitment websites:

- The company is leaking details about its technologies to the public, and this information can be leveraged by a threat actor.
- A hacker can determine the infrastructure and select exploits and tools to perform a cyber-attack on the targeted organization.

As a penetration tester, when recruiters reveal such information, we can easily create a portfolio of the targeted organization's internal infrastructure by identifying the operating systems of clients and servers, the vendor of networking devices, and the vendor of security appliances and technologies within the company's network.

To get a better understanding of developing a hacker mindset as a penetration tester, let's look at the following screenshot:

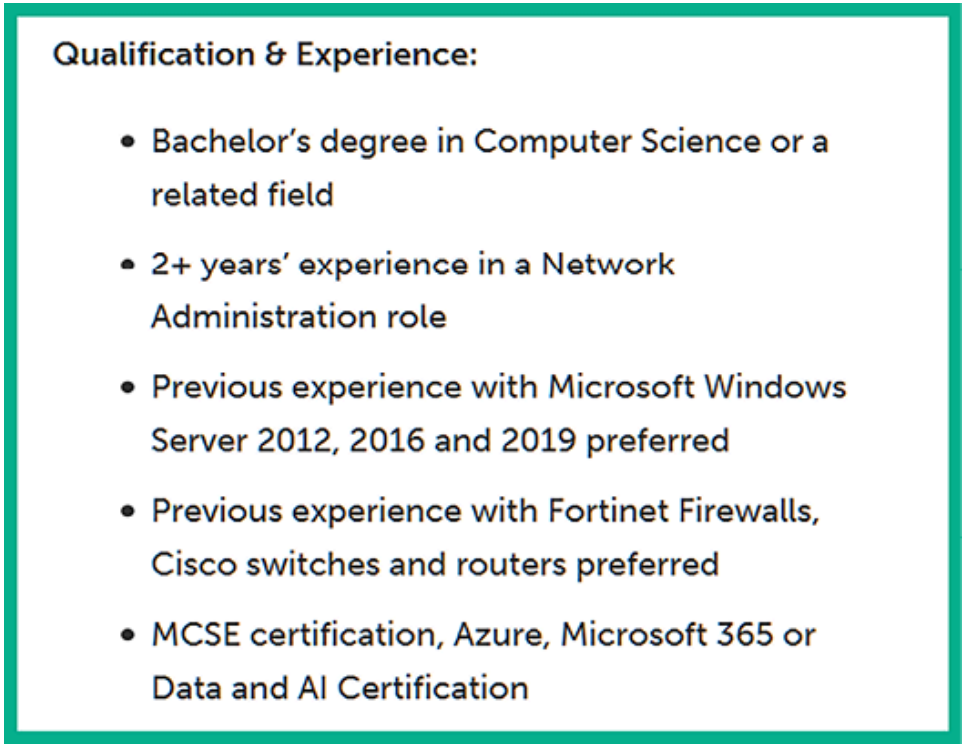


Figure 5.30: The main qualifications of the ideal penetration tester candidate

As shown in the preceding screenshot, the recruiter listed the main qualifications of the ideal candidate. Let's analyze the information provided by taking a closer look at the desired experience. The job poster is looking for someone who's experienced in Microsoft Windows Server 2012, 2016, and 2019.

The following can be derived from this information:

- The hiring organization has a Microsoft Windows environment with some older versions of Windows Server, specifically 2012 and 2016.
- There is the possibility that either the older systems or all Windows servers within the organization are not fully patched and contain security

vulnerabilities.

- The organization may not have rolled out Windows Server 2019 within their network yet or is planning to roll out the newer version of Windows Server soon.
- The hiring company specified the vendors for their existing networking devices and security solutions, which are Cisco routers and switches and Fortinet firewalls. This gives the attacker a clear idea of the threat prevention systems that are in place.
- The organization is also using Microsoft cloud computing services, such as Azure. Their cloud-based servers and applications are likely to not be secure.

As an aspiring penetration tester, using your favorite search engine, you can search for known security vulnerabilities and learn how to exploit each of these technologies. As you have seen, the recruiter leaked too much data about the organization, which can also be used against that same organization by threat actors for malicious purposes, but also by ethical hackers and penetration testers who have been hired to simulate a real-world cyber-attack, who can help the organization identify how they are leaking data and the potential impact if the information is leveraged by a real attacker.

Next, you will learn how to use a special internet search engine to find exposed systems from many organizations around the world.

Finding vulnerable systems using Shodan

Shodan is a search engine for the **internet of things** (**IoT**), systems, and networks that are directly connected to the internet. Ethical hackers, penetration testers, and even threat actors use Shodan to identify their organization's or target's assets, and they check whether they have been publicly exposed on the internet.

This online tool helps cybersecurity professionals quickly determine whether their organization's assets have been exposed on the internet.

To provide some additional insight, imagine that you want to determine whether your organization has any systems, such as servers that are accessible over the internet. These servers may include open service ports, vulnerable running applications, and services. Imagine that your organization has a legacy system running an older operating system that isn't patched with the latest security updates from the vendor and is directly connected to the internet.

A penetration tester or threat actor can use an online tool, such as Shodan, to discover such systems without even sending a probe of any kind directly from the penetration tester's system to the targeted server, simply because Shodan scans the internet by sending requests to a wide range of IP addresses, indexing the responses. This is an active scanning process conducted by Shodan itself, not the users of Shodan; it detects it automatically.

To get started using Shodan, please use the following instructions:

1. Using your web browser, go to **https://www.shodan.io/** and register for an account. You can perform searches on Shodan without an account but the results will be very limited.
2. After creating your account, log in and use the Shodan search field to enter some keywords such as `windows server 2008`, as shown in the following screenshot:

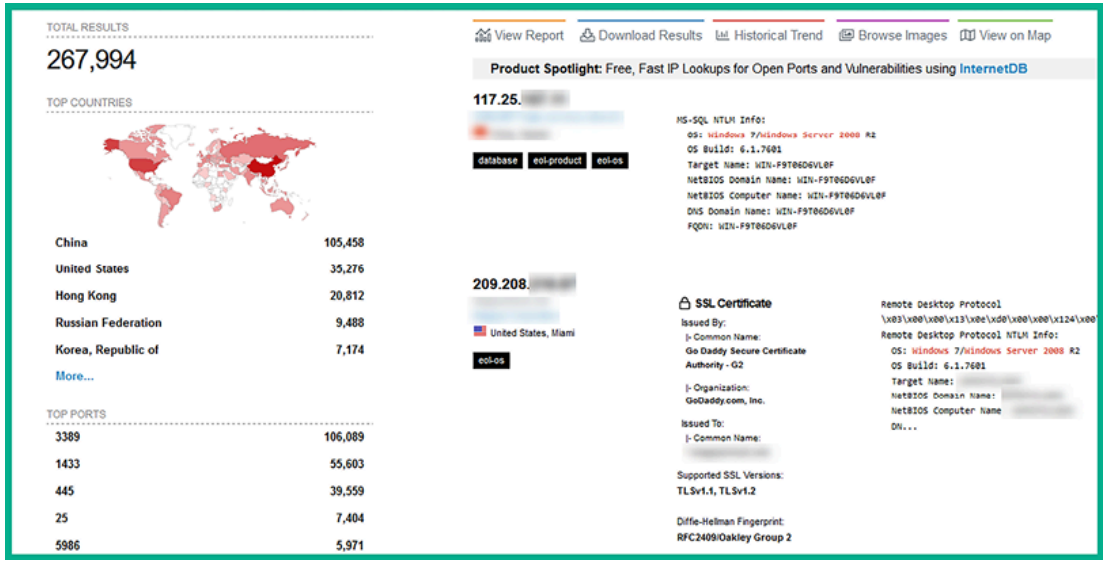


Figure 5.31: Shodan search

As shown in the preceding screenshot, there are over 200,000 devices around the world that are still running the Microsoft Windows Server 2008 operating system, which are identified using Shodan, and these systems are directly connected to the internet. As an ethical hacker and penetration tester, you can use Shodan to find exposed assets that are owned by the targeted organization to determine the attack surface of the company. Furthermore, once you're able to identify the operating systems of the target, you can research known security vulnerabilities for these systems.

3. Clicking on any of these systems will provide additional information about the system, such as open ports, running services, banners, and locale details. The following screenshot shows the local information of a system:

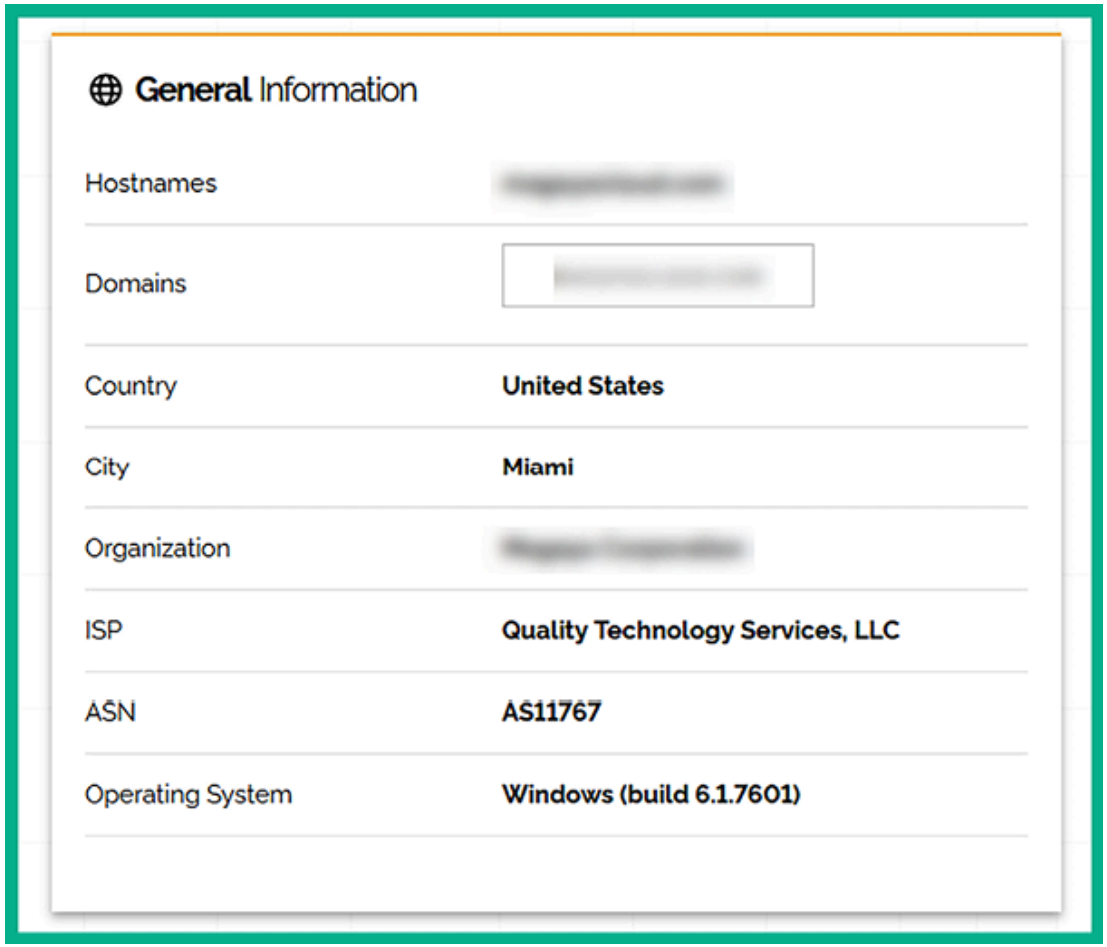


Figure 5.32: Local information of a system

As shown in the preceding screenshot, Shodan was able to retrieve the hostname, domain name, ISP details, **autonomous system number** (**ASN**), and locale information. Such information helps cybercriminals and ethical hackers determine the locality of the targeted organization during their reconnaissance phase.

4. Additionally, Shodan provides details on the open ports and their associated services that are running on the targeted system, as shown here:

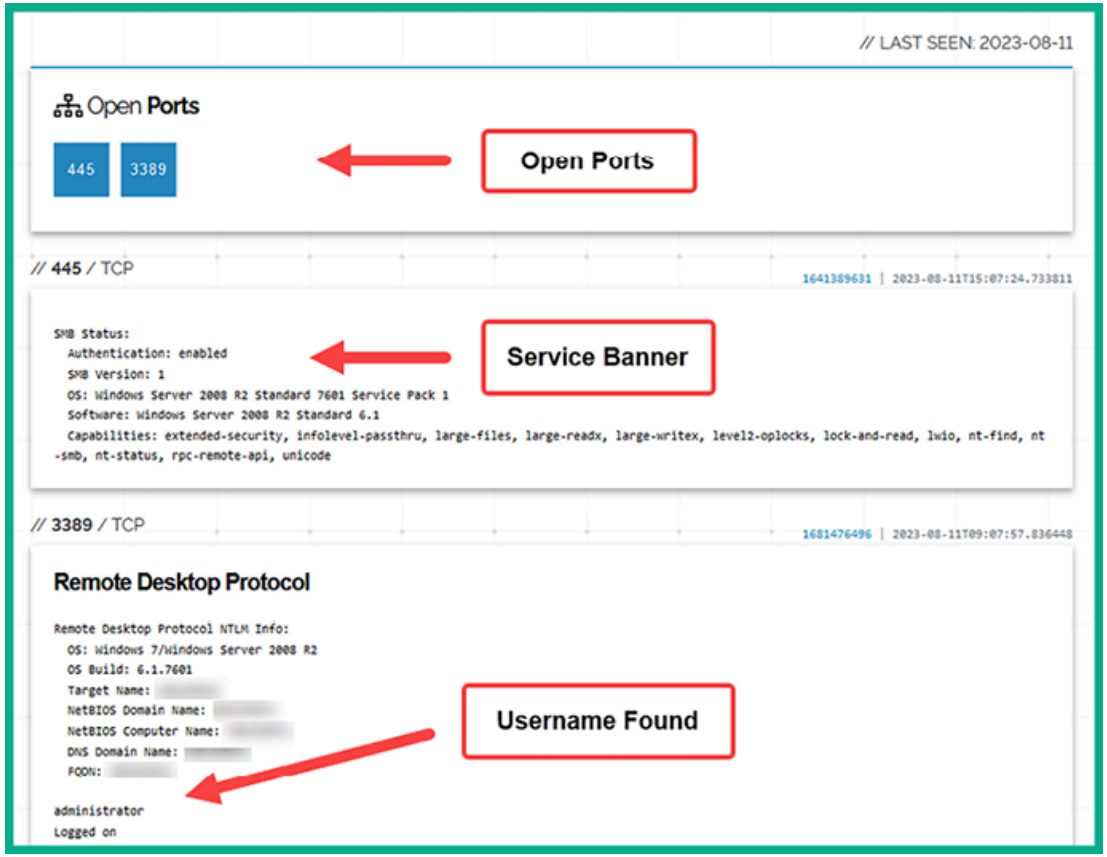


Figure 5.33: Details of open ports and their associated services

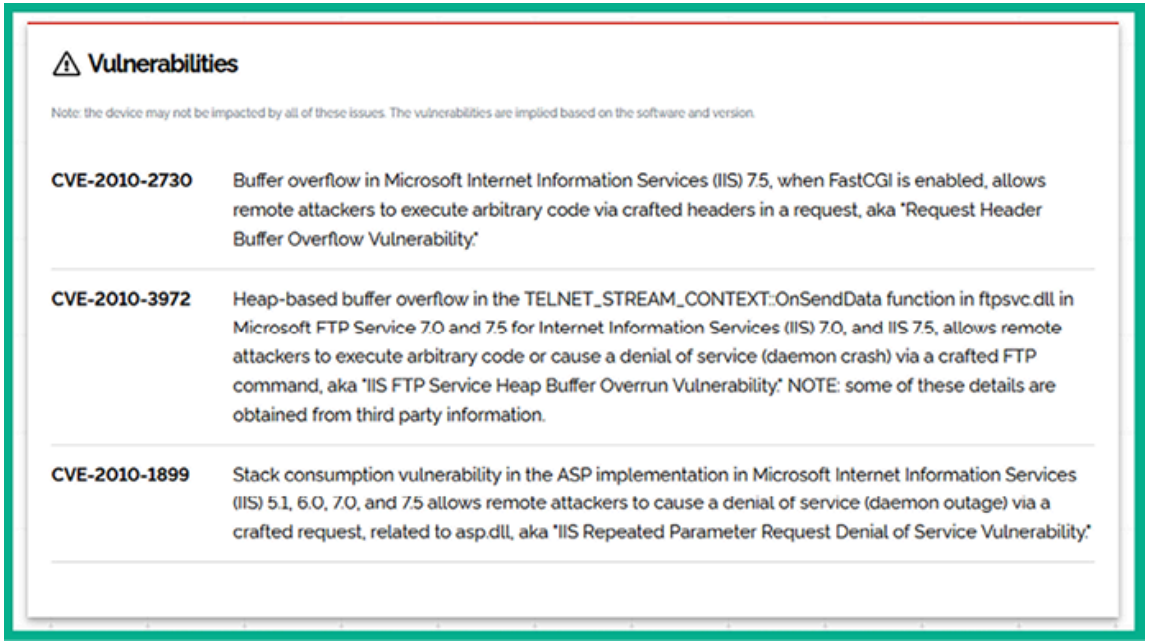
As shown in the preceding screenshot, Shodan was able to identify exposed services and ports on this system. Whether these ports were intentionally exposed by the organization or not, cybercriminals and ethical hackers can use this information to determine which services are running on the targeted system and determine whether there are any security vulnerabilities.

To learn more about service names and port numbers, please see https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml.

For instance, the system is running **Server Message Block (SMB)** version 1, which is known to contain security flaws that enable an attacker to perform **remote code execution (RCE)** on the targeted system. Furthermore, Shodan was able to enumerate a valid username for the **Remote Desktop Protocol (RDP)** service that's running on the device. Such details help the ethical hacker and penetration tester improve their attack and future operations aimed at gaining access to the target.

As an aspiring ethical hacker and penetration tester, you should know that ports are open on a system to allow ingress and egress traffic. Identifying open ports helps you to determine the entry points on a targeted system.

5. Additionally, if Shodan detects any known security vulnerabilities on the system, it will provide the details as shown here:



⚠ Vulnerabilities

Note: the device may not be impacted by all of these issues. The vulnerabilities are implied based on the software and version.

CVE-2010-2730	Buffer overflow in Microsoft Internet Information Services (IIS) 7.5, when FastCGI is enabled, allows remote attackers to execute arbitrary code via crafted headers in a request, aka "Request Header Buffer Overflow Vulnerability."
CVE-2010-3972	Heap-based buffer overflow in the TELNET_STREAM_CONTEXT::OnSendData function in ftpsvc.dll in Microsoft FTP Service 7.0 and 7.5 for Internet Information Services (IIS) 7.0, and IIS 7.5, allows remote attackers to execute arbitrary code or cause a denial of service (daemon crash) via a crafted FTP command, aka "IIS FTP Service Heap Buffer Overrun Vulnerability." NOTE: some of these details are obtained from third party information.
CVE-2010-1899	Stack consumption vulnerability in the ASP implementation in Microsoft Internet Information Services (IIS) 5.1, 6.0, 7.0, and 7.5 allows remote attackers to cause a denial of service (daemon outage) via a crafted request, related to asp.dll, aka "IIS Repeated Parameter Request Denial of Service Vulnerability."

Figure 5.34: Security vulnerabilities

As shown in the preceding screenshot, Shodan provides a list of known security vulnerabilities with a brief description and their associated **Common Vulnerabilities and Exposure** (**CVE**) numbers.

The CVE database allows cybersecurity professionals and researchers to report and track security vulnerabilities at https://cve.mitre.org/. Furthermore, cybersecurity professionals use the CVE details to create **cyber threat intelligence** (**CTI**) to improve their cyber defenses and mitigate new and emerging threats.

CTI contains any information about an attack or threat actor, such as **indicators of compromise** (**IoCs**) and the **Tactics, Techniques, and Procedures** (**TTPs**) used to perform the attack. To learn more about CTI, please visit https://www.techtarget.com/whatis/definition/threat-intelligence-cyber-threat-intelligence.

As shown in the preceding steps, ethical hackers and penetration testers can leverage the search algorithm of Shodan to passively collect information to identify the attack surface of their targets. Shodan can help you gather OSINT data without having to directly engage a target. In the next section, you will discover how to use another well-known tool within the industry to gather in-depth intelligence on systems on the internet.

Discovering exposed systems with Censys

Censys is another internet search engine that helps cybersecurity professionals and researchers collect and analyze information about internet-facing systems and identify their attack surface to better understand how a cybercriminal leverages public information such as the domain names, IP addresses, and digital certificates that are associated with a targeted organization.

To get started working with Censys, please use the following instructions:

1. Firstly, go to **https://search.censys.io/** and register for a free user account on the platform:

Figure 5.35: Censys search page

2. After registering for an account, log in and use the **Search** field to enter the name, IP address, or domain name of your targeted organization and click on **Search** to perform a lookup, as shown here:

Figure 5.36: Censys search

As shown in the preceding screenshot, a lookup was performed on Cloudflare's DNS address. The results show information about the network and IP addressing, running services and open ports on the server, and the geolocation of the server. Such information is useful when trying to find the geolocation or locale data for a targeted organization.

3. Next, the **Explore** tab provides additional information such as associations with other domain names, IPv4 and IPv6 addresses, and other assets owned by the organization, as shown here:

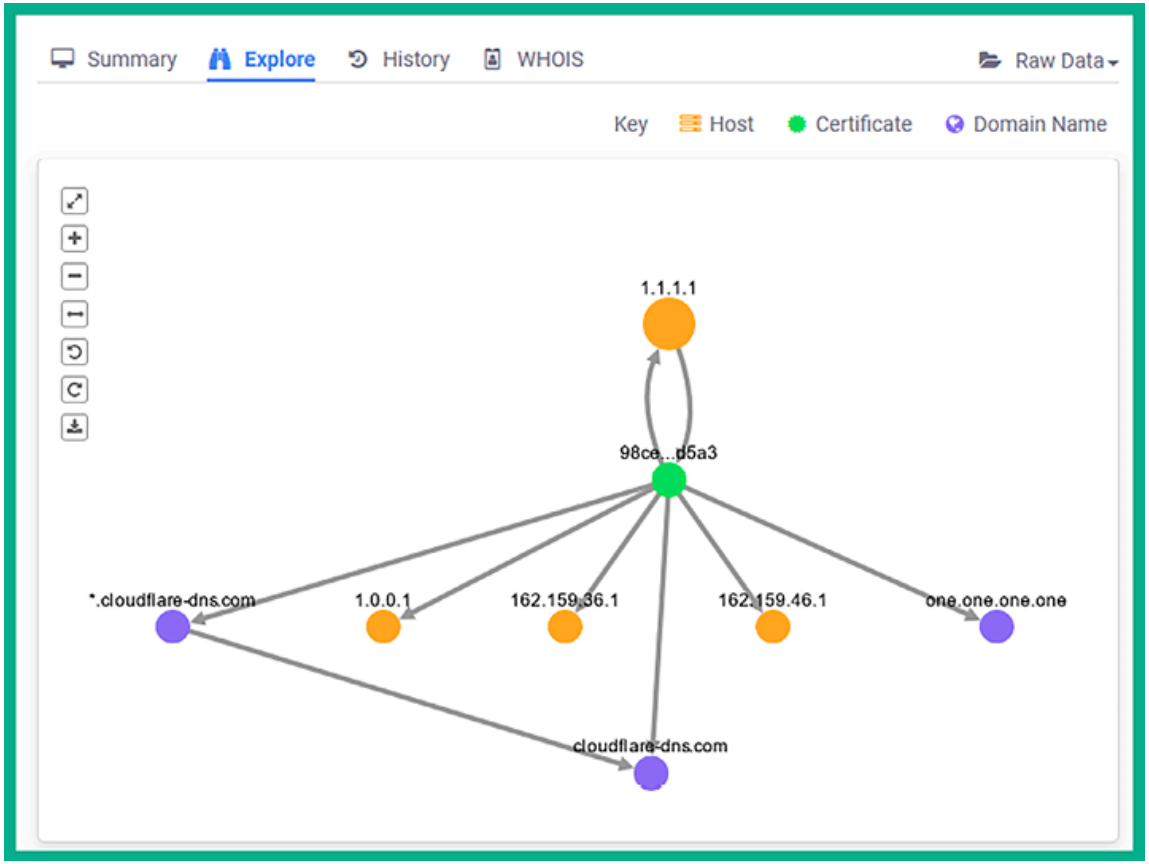


Figure 5.37: Explore view

4. The **History** tab allows you to view the change that occurred on the targeted system. Understanding what has changed helps ethical hackers and penetration testers determine whether there's a vulnerability within an application or

configuration. For instance, installing a new plugin on a web application can introduce new security vulnerabilities that can be exploited by a threat actor.

5. The **WHOIS** table provides the domain registration details and contact information about the owner of the domain. Sometimes, a domain owner does not pay an additional fee during the domain registration process to conceal their personal information. It's common for threat actors and ethical hackers to identify the domain registration details to determine the owner's contact details and geolocation of the organization.

Using the information gathered from Censys, ethical hackers and penetration testers can create a profile of systems that are publicly available through the internet and their open ports. Such information can be leveraged to research security vulnerabilities and techniques to compromise those systems.

In the next section, you will learn how to automate the mapping of external systems using a passive reconnaissance tool such as Maltego, as it helps us to easily locate OSINT data.

Mapping external systems using Maltego

Maltego is a graphical OSINT tool created and maintained by Maltego Technologies. This tool helps ethical hackers and penetration testers collect intelligence on a targeted organization's infrastructure by using a graphical interactive data mining application. This application provides the ability to query and gather information from multiple data sources on the internet and present data in easy-to-understand graphs. These graphs provide visualizations of the relationships between each entity and the target, therefore helping penetration testers to identify the external attack surface or a targeted system, network, and organization.

To get started with using Maltego for data harvesting, please use the following instructions:

1. Go to **https://www.maltego.com/ce-registration/** to register for a free **Community Edition** (**CE**) user account for the Maltego application.
2. Next, power on the **Kali Linux** virtual machine and log in.
3. On the **Kali Linux** desktop, open the **Terminal** and use the following commands to update the local package repository list and install Maltego:

```
kali@kali:~$ sudo apt update
kali@kali:~$ sudo apt install maltego
```

4. Next, click the Kali Linux icon (top-left corner) to expand the applications menu, select **1 – Information Gathering** | **OSINT Analysis** | **maltego**, as shown here:

Figure 5.38: Information gathering

5. Next, the Maltego **Product Selection** window will appear, select **Maltego CE (Free)** | **Run** to launch the community edition, as shown here:

Figure 5.39: Product selection

6. Next, the **Configure Maltego – License Agreement** window will appear, accept the license agreement, and click on **Next**.
7. Next, the **Configure Maltego – Login** window appears; confirm your user credentials that were created during *step 1*, and click on **Next** to continue.
8. Click **Next** on the **Login Results**, **Install Transforms**, **Help Improve Maltego**, **Web Browser Options**, **Privacy Mode Options**, and **Ready** windows.
9. To start gathering information on a targeted organization, open a new graph. To do this, click on the **Maltego icon** (top-left corner), and then click on **New**:

Figure 5.40: New graph

As shown in the preceding screenshot, once a new graph is created, you'll see various types of entities on the left pane, while on the right side, you'll see the **Overview**, **Detail View**, and **Property View** panes.

10. Next, to start collecting infrastructure information about a targeted organization, from the **Entity Palette** section, drag and drop the **Domain** entity onto the middle of the graph pane, as shown here:

Figure 5.41: Domain entity on the graph pane

11. Next, double-click on the **Domain** entity on the graph pane to open the **Details** window, enter an organization's domain name within the **Domain Name** field, and click on **OK** as shown here:

Figure 5.42: Domain name

12. To retrieve the target's public DNS records, right-click on the **Domain** entity on the graph pane and select **All Transforms** | **To DNS Name – NS (name server)**, as shown here:

Figure 5.43: To DNS name - NS transform

Once the transform executes and retrieved data, Maltego populates the graph pane to show the name servers of the target, as shown here:

Figure 5.44: Name servers of the target

13. To retrieve the **mail exchange (MX)** records to identify a target's email servers, right-click on the **Domain** entity and select **All Transforms** | **To DNS Name – MX (mail server)**, as shown here:

Figure 5.45: To DNS name - MX transform

Once Maltego retrieves the MX records from public DNS servers, the graph pane is updated to show the email servers of the targeted organization:

Figure 5.46: Email servers of the target

14. To retrieve the public IP address of the name servers or email servers, right-click on one of the entities on the graph pane and select **All Transforms** | **To IP Address [DNS]**, as shown here:

Figure 5.47: IP addresses of the target

15. Next, to discover whether there's a website that's associated with the targeted domain, right-click on the **Domain** entity and select **All Transforms** | **To Website [Quick lookup]**.
16. To retrieve the public IP addresses that are associated with the website address, right-click on the **Website** entity and select **All Transforms** | **To IP Address [DNS]**, as shown here:

Figure 5.48: To Website transform

17. To retrieve a list of publicly known email addresses, which are associated with the targeted domain, right-click on the **Domain** entity and select **All Transforms** | **To Email addresses [PGP]**, as shown here:

Figure 5.49: To Email addresses transform

18. Lastly, you can save the information collected by Maltego by clicking on the Maltego icon on the left corner and selecting the **Save** option.

The relation-mapping feature on Maltego helps you analyze information and understand how one component is connected to another. Using the information that's been collected from Maltego, you can determine publicly available servers, IP addresses, employees' email addresses, linked URLs on web pages, and more. As you have seen, using a tool such as Maltego can help automate the process of gathering various types of OSINT data from multiple data sources on the internet, which helps ethical hackers and penetration testers reduce spending during the reconnaissance phase.

Next, you will learn how to use Netcraft to identify the external attack surface, assets, and technologies of a targeted organization.

Identifying infrastructure with Netcraft

Netcraft enables ethical hackers and penetration testers to collect OSINT on targeted organizations. It provides capabilities that allow them to better understand the technologies, operating systems, applications, and locations of their internet-facing devices.

Netcraft provides the following data types:

- Network and IP information
- IP geolocation information
- Website technologies and applications

To get started using Netcraft to profile a targeted organization/domain, please use the following instructions:

1. Using a standard web browser, go to **https://sitereport.netcraft.com/**, then enter a targeted domain name within the domain field, and click on **LOOK UP**, as shown here:

Figure 5.50: Using Netcraft

2. After a few seconds, Netcraft will automatically display all of the information it knows about the targeted domain and its technologies, as shown here:

Figure 5.51: Data retrieved by netcraft

As shown in the preceding screenshot, ethical hackers and penetration testers use the information to determine the owner of the domain name, the name servers, the hosting company, and the public addresses of the target.

3. Next, to determine the geolocation of the targeted organization, scroll down to the **SSL/TLS** section, as shown here:

Figure 5.52: SSL/TLS section

As shown in the preceding screenshot, Netcraft was able to collect and analyze the information found within the digital certificate for the targeted domain

and provide the organization, state, and country. Such information is useful for ethical hackers when planning a physical penetration test. In addition, the **Subject Alternative Name** field provides additional sub-domains, which are permitted to use this digital certificate; this data helps penetration testers identify additional assets that are owned by the target.

4. Next, the **Site Technology** section provides valuable information such as identifying the server-side and client-side technologies. This information is useful when planning a web application penetration test, as shown here:

Figure 5.53: Site technology section

Having completed this exercise, you have gained the knowledge of how ethical hackers are able to leverage Netcraft to identify the public infrastructure of a targeted organization. Next, you will learn how to use Recon-ng to automate data collection and analysis as an ethical hacker.

Using Recon-ng for data harvesting

Recon-ng is an OSINT reconnaissance framework written in Python. The tool itself contains a lot of modules for additional capabilities, a database for storing OSINT, interactive help, and a menu system, similar to Metasploit. Recon-ng can perform web-based, information-gathering techniques to collect OSINT from multiple online data sources, and it's one of the must-have tools for any aspiring ethical hacker or penetration tester to have within their arsenal.

To get started using Recon-ng for data harvesting, please use the following instructions:

1. Power on the **Kali Linux** virtual machine and execute the following command within the **Terminal** to start Recon-ng:

```
kali@kali:~$ recon-ng
```

2. Recon-ng uses various modules that are designed to collect and analyze data from multiple data sources. By default, there are no modules pre-installed on Recon-ng, therefore, use the following commands to install all modules from the Recon-ng marketplace:

```
[recon-ng][default] > marketplace install all
```

The following screenshot shows Recon-ng is downloading and setting up the modules:

Figure 5.54: Running Recon-ng

After the modules are installed, Recon-ng will automatically reload the newly installed modules and there will be a lot of warning messages that are written in red, as shown here:

Figure 5.55: Module reload

The preceding screenshot shows there are various Recon-ng modules that require an **application programming interface** (**API**) key to authenticate and allow Recon-ng to retrieve OSINT from the data source.

The `modules search` command is used to display all current modules with Recon-ng and their categories, such as Discovery, Exploitation, Import, Recon, and Reporting.

3. Next, to view a list of supported API keys on Recon-ng, use the following commands:

```
[recon-ng][default] > keys list
```

As shown in the following screenshot, the `keys list` commands allow us to view which API keys are supported and whether there's already an API in use:

Figure 5.56: Keys list

4. Next, to get a supported API key, simply go to the data source such as **BuiltWith** at **https://builtwith.com/** and create a free user account. Once an account is created, log in and go to **Tools** | **API Access** to find the API key. Feel free to acquire as many API keys for each supported module from the list of supported APIs.
   Consider getting an API key from the following data sources:
   1. Hunter – **https://hunter.io**
   2. Censys – **https://search.censys.io**

3. VirusTotal – **https://www.virustotal.com**
4. Shodan – **https://www.shodan.io/**

5. Once you've acquired your API keys, the next step is to add each API to their API-supported modules. Use the `keys add <API-module-name> <API key value>` command. For instance, the following commands are used to add an API key for the `builtwith_api`:

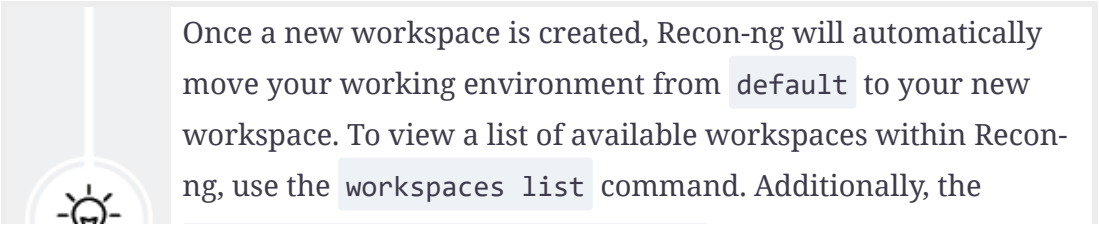
```
[recon-ng][default] > keys add builtwith_api 12345
```

6. After adding your API keys, use the `keys list` command to verify whether the keys were added successfully, as shown here:

Figure 5.57: Verifying API keys

7. As an ethical hacker and penetration tester, you may be working on multiple projects at a time, Recon-ng enables you to create multiple, virtual workspaces to help you better manage the collection and analysis of data. To create a new workspace, use the following commands:

```
[recon-ng][default] > workspaces create myfirstproject
```



Once a new workspace is created, Recon-ng will automatically move your working environment from `default` to your new workspace. To view a list of available workspaces within Recon-ng, use the `workspaces list` command. Additionally, the

`workspaces load <workspace-name>` command allows you to select and work within a specific workspace, while the `workspaces remove <workspace-name>` command removes a workspace from Recon-ng.

8. Next, the `modules search <keyword>` commands enable you to search for specific modules based on a keyword. For instance, use the `modules search whois` command to view all Recon-ng modules that contain the `whois` keyword, as shown here:

Figure 5.58: modules search whois

9. Next, to use a specific module within Recon-ng, use the `modules load <module-name>` command. For instance, to gather a list of **point of contacts (POCs)** for a targeted domain on the internet, use the following commands:

```
[recon-ng][myfirstproject] > modules load recon/domains-contacts/whois_pocs
[recon-ng][myfirstproject][whois_pocs] > info
```

As shown in the following screenshot, the `info` command prints the description and required options for the selected module:

Figure 5.59: modules load

10. To set the required options for the module, use the following commands to set `microsoft.com` as the **SOURCE** for our targeted domain:

```
[recon-ng][myfirstproject][whois_pocs] > options set SOURCE microsoft.com
```

To unset a value within a module, use the `options unset <parameter/value>` command. Ensure that you execute the `info` command afterward to verify the value is unset/removed.

11. Next, use the `run` command to execute the module, as shown here:

Figure 5.60: Executing the module

12. Next, use the `back` command to exit a module and `modules search bing` command for modules that can leverage the Bing search engine, as shown here:

Figure 5.61: modules search bing

13. Next, use the following command to load the `bing_domain_web` module, display its information, set the targeted domain, and execute the module:

```
[recon-ng][myfirstproject] > modules load recon/domains-hosts/google_site_web
[recon-ng][myfirstproject][google_site_web] > info
```

```
[recon-ng][myfirstproject][google_site_web] > options set SOURCE microsoft.com
[recon-ng][myfirstproject][google_site_web] > run
```

14. Use the `show hosts` command to view a list of sub-domains and hostnames that were collected about the target, as shown here:

Figure 5.62: Showing hosts

15. Next, use the `show contacts` command to view a list of contact information that was collected, as shown here:

Figure 5.63: Showing contacts

The `show` command can be used with `show [ companies ] [ credentials ] [ hosts ] [ locations ] [ ports ] [ pushpins ] [ vulnerabilities ] [ contacts ] [ domains ] [ leaks ] [ netblocks ] [ profiles ] [ repositories ]` to view specific information that was obtained by Recon-ng. Additionally, the `dashboard` command provides a summary of all activities in Recon-ng such as showing the number of times a module was executed and how much data was collected.

16. To view a summary of your activities within the `myfirstproject` workspace, use the `dashboard` command, as shown here:

Figure 5.64: The dashboard command

17. Next, collecting all the data can be overwhelming to process and analyze, however, Recon-ng has various reporting modules to help us. Use the `modules search report` command to view a list of all reporting modules, as shown here:

Figure 5.65: modules search report

18. To generate an HTML-format report, use the following commands to set the required parameters and specify the output location for the final report:

```
[recon-ng][myfirstproject] > modules load reporting/html
[recon-ng][myfirstproject][html] > info
[recon-ng][myfirstproject][html] > options set CREATOR GLEN
[recon-ng][myfirstproject][html] > options set CUSTOMER ACME_Enterprises
[recon-ng][myfirstproject][html] > options set FILENAME /home/kali/Desktop/myfirstproject_report
[recon-ng][myfirstproject][html] > run
```

The following screenshot shows how the preceding commands were applied on the module:

Figure 5.66: Generating HTML-format report

19. To view the report, simply go to the output directory such as `/home/kali/Desktop` and open the report HTML file using the web browser, as shown here:

Figure 5.67: Viewing the report

This report provides a very easy-to-understand summary of all the data that was collected using Recon-ng. The reporting module plays an excellent role in helping ethical hackers correlate data collected during the reconnaissance phase when using Recon-ng to develop a profile about the target and identify security vulnerabilities.

20. Next, to access the web interface of Recon-ng, use the following command on a new Terminal:

```
kali@kali:~$ recon-web
```

21. Once the workspace has been initialized, open the web browser within Kali Linux and go to `http://127.0.0.1:5000/`, as shown here:

Figure 5.68: Workspace page at 127.0.0.1:5000

22. As shown in the preceding screenshot, ethical hackers and penetration testers can improve their data collection and analysis using the web interface of Recon-ng.

To learn more about Recon-ng and its features, please visit the official GitHub repository at https://github.com/lanmaster53/recon-ng.

Having completed this exercise, you've learned how to leverage Recon-ng to efficiently collect and analyze OSINT from multiple data sources. Next, you will learn how to use theHarvester for data harvesting.

Data collection with theHarvester

Using a tool such as **theHarvester** enables to you efficiently collect OSINT to identify sub-domains and additional exposed assets of a targeted organization. This tool helps ethical hackers and penetration testers to automate the collection of email addresses, sub-domains, hostnames, and employees' names, and identify open ports and banners of systems that are associated with the target.

To get started using theHarvester for data collection, please use the following instructions:

1. Firstly, power on the **Kali Linux** virtual machine and ensure it has internet connectivity.
2. Next, open the **Terminal** and use the following command to display the menu for theHarvester:

```
kali@kali:~$ theHarvester -h
```

The preceding command displays the help menu and provides a list of various syntaxes and how they can be used to retrieve OSINT from online sources. In

addition, the help menu provides a list of various data sources using the `-b` command.

3. Next, to retrieve a list of sub-domains of a targeted domain, use the following commands:

```
kali@kali:~$ theHarvester -d microsoft.com -b duckduckgo
kali@kali:~$ theHarvester -d microsoft.com -b dnsdumpster
kali@kali:~$ theHarvester -d microsoft.com -b bing
kali@kali:~$ theHarvester -d microsoft.com -b yahoo
kali@kali:~$ theHarvester -d microsoft.com -b crtsh
```

The following screenshot shows theHarvester was able to collect multiple sub-domains for the targeted domain:

Figure 5.69: Data collection with theHarvester

To learn more about the features of theHarvester, please visit the official GitHub repository at **https://github.com/laramies/theHarvester**. Some sources require an API key to retrieve data from the online database. To learn more about how to add an API key to theHarvester, please see **https://github.com/laramies/theHarvester/wiki/Installation#api-keys**.

Having completed this section, you have gained the skills needed to collect OSINT information on targeted organizations to identify how they are leaking data such as their internal infrastructure to anyone on the internet. In the next section, you will learn how to gather employee OSINT.

Harvesting employees' data using Hunter

Around the world, employees of many organizations commonly leak and share too much information about themselves and their organization without realizing how a threat actor or adversary can collect and analyze such information to plan a cyber-attack or improve a threat towards their organizations and themselves. Quite often, you'll notice that many employees of the leadership team for an organization commonly share their contact details on professional social networking platforms, such as the following types of information:

- Full name and job title
- Company's email address
- Telephone number
- Roles and responsibilities
- Recent projects with technical details
- Pictures of their employee badges

As a penetration tester, it's quite simple to create an account that will function as a sock puppet on a site such as LinkedIn, populate some false information on the account, such as information stating you're an employee who is working at another branch office, and then add some low-level employees from the targeted organization as connections. Therefore, other employees of the targeted organization will notice your sock puppet profile has mutual connections and may reduce suspicions.

Sock puppets are covered in *Chapter 4*, *Passive Reconnaissance*.

There's a possibility the employees will automatically accept the connection/friend request because they will see that you're a fellow employee at their company. This will provide some leverage for you to connect with the high-profile employees of the targeted organization and enable you to collect contact details to plan various social engineering attacks and identify your targets.

Hunter is an online data source that harvests both employee and organizational data from public sources on the internet. As an ethical hacker and penetration tester, this is a must-have resource for gathering employees' names, telephone numbers, email addresses, and even their job titles when planning a social engineering attack.

To get started using this tool, please use the following instructions:

1. Firstly, you'll need to register for a free account at **https://hunter.io/** and complete the registration process.
2. Once the registration process is completed, log in to the online platform using your user credentials.
3. Next, you'll be presented with the **Domain Search** field. Here, simply enter a targeted domain, as shown here:

Figure 5.70: Domain Search

4. While entering a domain within the **Domain Search** field, Hunter will provide suggestions for your search. I've used `microsoft.com` as an example, as shown here:

Figure 5.71: Suggestions provided by Hunter

As shown in the preceding screenshot, Hunter can provide a list of employees' information, such as their names, email addresses, telephone numbers, and other sources of information. In addition, collecting email addresses for targeted organizations helps you to determine the format of employees' email addresses.

Therefore, if an adversary or ethical hacker knows the names of employees, then it's easy to guess the email addresses of various employees. This information is useful when planning social engineering, password spraying, and credential stuffing attacks.

To learn more about password spraying and credential stuffing attacks, please see `https://attack.mitre.org/techniques/T1110/003/` and `https://attack.mitre.org/techniques/T1110/004/`.

The following screenshot shows all the sources that Hunter used to collect the data for a specific person:

Figure 5.72: Data collection by Hunter

While employees will provide their contact details on various online platforms, including their company's website, such information can be leveraged by a threat actor and a penetration tester to perform social engineering attacks against the organization.

Automating social media reconnaissance with Sherlock

Employees of an organization often leak too much information about themselves and their company. While many employees are very happy to be working in their organizations, sometimes, they share information that can be leveraged by threat actors to improve their attack on a target. As an aspiring ethical hacker and penetration tester, collecting and analyzing information from social media platforms can be useful in finding employee profiles with weak privacy, which are not secure, and collecting any sensitive data from their profiles.

The following is some information that's commonly leaked:

- Employee contact information, such as telephone numbers and email addresses, which can be used during social engineering and account takeover attacks.
- Sharing photos with their employee badges, which can be used by a threat actor to create a fake ID for impersonation for physical penetration testing.
- Pictures of an employee's computing systems and desktop, which can inform a threat actor about the available device vendors and operating systems.
- Projects that have been completed by the employee may contain specific technical details, which can allow a threat actor to profile the internal network

infrastructure.

These are just some of the many types of information that are commonly posted on social media platforms such as LinkedIn. As a penetration tester, you can create a sock puppet, impersonate someone on social media, and trick the employees of the targeted organization into performing an action or revealing sensitive information (social engineering). Furthermore, imagine performing a physical penetration test, where you can print a fake employee ID badge and dress like a typical employee by using the information found on the targeted organization's social media page.

Sherlock is an OSINT tool that helps penetration testers quickly determine whether their target has any social media accounts and which platforms the accounts may exist on. This tool supports over 200 social media websites, automates the process of checking each site, and generates a report of the results.

To get started using Sherlock for social media reconnaissance, please use the following instructions:

1. Power on the **Kali Linux** virtual machine, open the **Terminal**, and use the following commands to download **Sherlock** from its official GitHub repository:

```
kali@kali:~$ sudo apt update
kali@kali:~$ git clone https://github.com/sherlock-project/Sherlock
```

2. Next, use the following commands to install the requirements for Sherlock:

```
kali@kali:~$ cd sherlock
```

```
kali@kali:~/sherlock$ python3 -m pip install -r requirements.txt
```

3. Next, to search for a targeted organization's social media presence on the internet, use the `python3 sherlock <username>` command, as shown here:

```
kali@kali:~/sherlock$ python3 sherlock microsoft --timeout 5
```

Notice the `--timeout` command was used to instruct Sherlock to not spend more than five seconds on any of the social media sites, as shown here:

Figure 5.73: Running Sherlock with the timeout command

When Sherlock completes the task, the results will be stored in a text file within the present working directory, as shown here:

Figure 5.74: File organization by Sherlock

Be sure to check each site within the output file to ensure it is valid and provides meaningful information about your target. A penetration tester can use the information that's been collected to easily identify the social media accounts owned by a targeted organization or user. Such information can be also used to gather further intelligence on the target.

To learn more about Sherlock, please visit **https://github.com/sherlock-project/sherlock** and

https://www.kali.org/tools/sherlock/.

Having completed this section, you have learned how to automate the data collection process of finding user accounts for a targeted organization or person using Sherlock.

Summary

During this chapter, you have learned how to apply various Google hacking techniques to perform advanced search and filtering to identify sensitive directories and exposed resources on the internet. In addition, you have gained the hands-on skills needed to perform domain reconnaissance to collect and analyze DNS records, perform zone transfer, and identify the sub-domains of a target. Furthermore, you have learned how to leverage specialized internet search engines to identify exposed assets of companies around the world and gained a better understanding of how OSINT helps ethical hackers and penetration testers to develop a profile about their targets.

I trust that the knowledge presented in this chapter has provided you with valuable insights, supporting your path toward becoming an ethical hacker and penetration tester in the dynamic field of cybersecurity. May this newfound understanding empower you in your journey, allowing you to navigate the industry with confidence and make a significant impact. In the next chapter, *Exploring Active Reconnaissance*, you will learn how to perform active reconnaissance techniques to identify live systems, open ports, and running services.

Further reading

- OSINT – https://www.imperva.com/learn/application-security/open-source-intelligence-osint/
- Top OSINT tools – https://www.csoonline.com/article/567859/what-is-osint-top-open-source-intelligence-tools.html
- What is WHOIS? – https://www.domaintools.com/support/what-is-whois-information-and-why-is-it-valuable/

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

https://packt.link/SecNet