2
Section 2: Reconnaissance

This section teaches the reader about the importance of conducting extensive reconnaissance on a target prior to launching any exploits or payloads. The reader will learn the various types of information-gathering techniques, along with a variety of online and offline tools to assist in retrieving information and specific details pertaining to a target. Upon completing this section, the reader will be equipped with the essentials of both the theory and hands-on experience of performing various information-gathering techniques as a cyber security professional.

This section comprises the following chapters:

- `Chapter 5`, *Passive Information Gathering*
- `Chapter 6`, *Active Information Gathering*

5
Passive Information Gathering

Beginning a career in ethical hacking and/or penetration testing can be very exciting and, most of the time, our minds will be a bit overwhelmed, causing us to visit only the chapters about exploiting a system in a book such as this. However, conducting a penetration test is like starting a new project at home. Before you build a pool in your backyard, there are a few things you must consider, such as the space that's available, the cost of materials, the contractor's fees, and other details. Information gathering is a very important phase of the hacking life cycle and penetration testing.

In this chapter, we will focus on passive information gathering techniques and methods. We will learn how to use the internet to get us the information and specific details we need about our target by using both online resources and tools on Kali Linux.

We will be covering the following topics:

- Reconnaissance and footprinting
- Understanding passive information gathering
- Understanding **open source intelligence** (**OSINT**)
- Using the top OSINT tools
- Identifying target technologies and security controls
- Finding data leaks in cloud resources
- Understanding whois and copying websites with HTTrack
- Finding subdomains using Sublist3r

Dear reader! Please ensure that you **do not** perform scans on any target organization, networks, or systems in the absence of appropriate **legal permission**.

Technical requirements

The following are the technical requirements for this chapter:

- **Kali Linux**: `https://www.kali.org/`
- **Maltego**: `www.paterva.com`
- **Recon-ng**: `https://bitbucket.org/LaNMaSteR53/recon-ng`
- **theHarvester**: `https://github.com/laramies/theHarvester`
- **OSRFramework**: `https://github.com/i3visio/osrframework`
- **HTTrack**: `www.httrack.com`
- **Sublist3r**: `https://github.com/aboul3la/Sublist3r`
- **S3Scanner**: `https://github.com/sa7mon/S3Scanner`

Reconnaissance and footprinting

The various phases of hacking include reconnaissance, scanning, gaining access, maintaining access, and clearing tracks. The reconnaissance phase is the most important phase of a penetration test since this is when the ethical hacker or penetration tester conducts extensive research into gathering as much information about the target as possible. Furthermore, footprinting will help create a profile of the target, gathering profiling information such as running services, open ports, and operating systems.

We will now look at both reconnaissance and footprinting in more detail.

Reconnaissance

From a military perspective, reconnaissance is the observation and research of an enemy target. In cybersecurity, as a penetration tester, we use various tools and techniques to gather detailed information about a target organization and its underlying infrastructure.

Reconnaissance is vital in the field of penetration testing. As a penetration tester, we definitely need to know about our target, as well as its vulnerabilities and operating systems, before we attempt to gain access via exploitation. The information gathered during the reconnaissance phase will help us to choose the right tools and techniques to successfully exploit the target.

Reconnaissance can be divided into two categories:

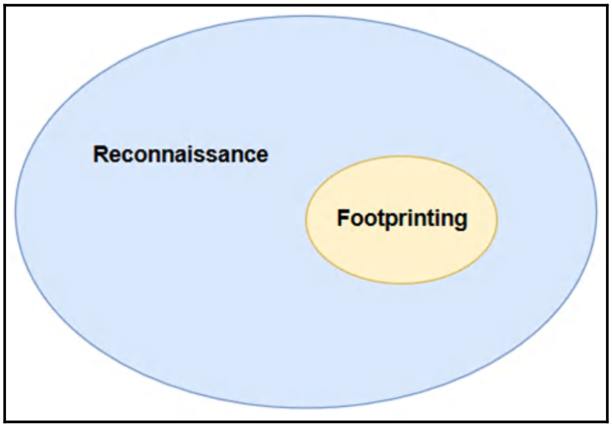
- **Passive**: Uses an indirect approach and does not engage the target
- **Active**: Directly engages the target to gather specific details

Next, we will dive into understanding footprinting.

Footprinting

Footprinting is the procedure whereby as much information as possible is gathered in relation to a target. In footprinting, the objective is to obtain specific details about the target, such as its operating systems and the service versions of running applications. The information that's collected can be used in various ways to gain access to the target system, network, or organization. Footprinting allows a penetration tester to understand the security posture of the target infrastructure, quickly identify security vulnerabilities on the target systems and networks, create a network map of the organization, and reduce the area of focus to the specific IP addresses, domain names, and the types of devices regarding which information is required.

Footprinting is part of the reconnaissance phase; however, since footprinting is able to provide more specific details about the target, we can consider footprinting to be a subset of the reconnaissance phase. The following diagram provides a visual overview of how reconnaissance and footprinting sit together:



The following are the main objectives of footprinting:

- Collecting network information (domain names, IP addressing schemes, and network protocols)
- Collecting system information (user and group names, routing tables, system names, and types)
- Collecting organization information (employee details, company directory, and location details)

To successfully obtain information about a target, I would recommend using the following footprinting methodology:

- Checking search engines such as Yahoo, Bing, and Google
- Performing Google hacking techniques (advanced Google searches)
- Information gathering through social media platforms such as Facebook, LinkedIn, Instagram, and Twitter
- Footprinting the company's website
- Performing email footprinting techniques
- Using the `whois` command
- Performing DNS footprinting
- Network footprint techniques
- Social engineering

You are now able to differentiate between reconnaissance and footprinting. Both reconnaissance and footprinting are required during penetration testing as each provides vital information about the target. In the next section, we will take a deep dive into passive information gathering.

Understanding passive information gathering

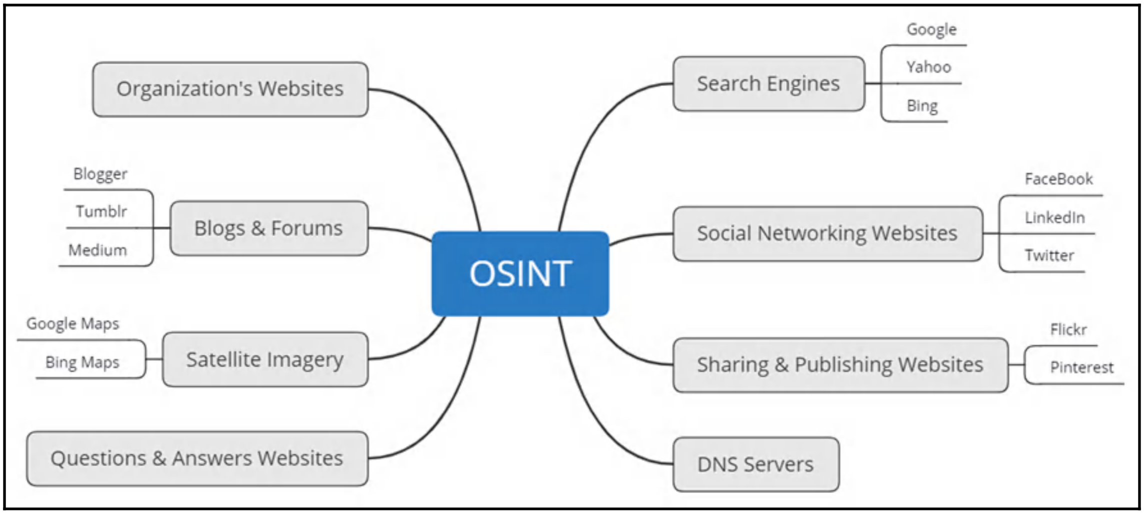
Passive information gathering is when you use an indirect approach to obtain information about your target. This method obtains information that's publicly available from many sources, thus eliminating direct contact with the potential target. Passive information gathering is usually fruitful, and a lot of organizations usually publish information and details about their organizations as a marketing strategy for their existing and potential customers. Sometimes, when organizations advertise a vacancy on a job recruiting website, the recruiter posts technical requirements for the potential candidate. From a penetration tester's point of view, the technical details can indicate the types of platforms and applications that are running within the organization's network infrastructure.

We have covered the concepts of passive information gathering. Now, let's take a deep dive into learning about OSINT in the next section.

Understanding OSINT

As mentioned previously, the first stage of a penetration test is to gather as much information as possible on a given target or organization. Gathering information prior to exploiting and gaining access to a network or system will help the penetration tester narrow the scope of the attack and design specific types of attacks and payloads that are suitable for the attack surface of the target. We will begin our information-gathering phase by utilizing the largest computer network in existence: the internet.

The following diagram provides a brief overview of the different areas where OSINT can be found on a target:



The internet has many platforms, ranging from forums and messaging boards to social media platforms. A lot of companies create an online presence to help market their products and services to potential clients. In doing so, the creation of a company's website, Facebook, Instagram, Twitter, LinkedIn, and so on ensures that their potential customers get to know who they are and what services and products are offered. The marketing department is usually responsible for ensuring that an organization's online presence is felt and that their digital portfolio is always up to date and eye-catching.

Organizations usually publish information about themselves on various internet platforms, such as blogs and recruitment websites. As the internet is so readily available and accessible, it's quite easy for someone to gather information on a target organization simply by using search engines and determining their underlying infrastructure. This technique is known as **OSINT.**

This is where a penetration tester or ethical hacker uses various tools and techniques that harness information that's publicly available on the internet to create a portfolio of the target. OSINT is a type of passive information gathering where the penetration tester does not make direct contact or a connection with the actual target, but rather asks legitimate and reliable sources about the target.

Over the years, I have noticed a lot of job-hunting websites where recruiters post vacancies for IT positions within a company, but the recruiter specifies that an ideal candidate should have experience with specific technologies. This can be a good thing for the company and the applicant; however, it can be bad as well. The following are the pros and cons of companies posting their technologies on recruitment websites:

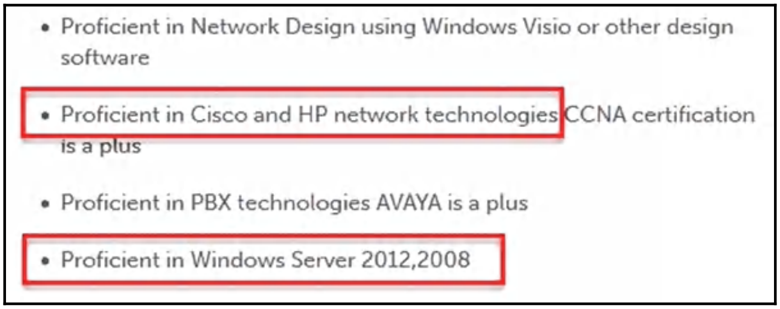
These are the pros:

- The potential candidate will know what type of environment to expect if they are hired.
- The potential candidate can determine beforehand whether they have the skill set required for the job.

These are the cons:

- The company is partially exposing their technologies to the general public.
- A hacker can determine the infrastructure and better select exploits and tools to perform a cyber attack.

Let's take a look at the following screenshot from a job site. Looking closely, we notice that the job poster has specified that they are using both Cisco and HP networking technologies, the company uses an AVAYA PBX system as their **Voice over Internet Protocol** (**VoIP**) system, and they are running Windows Server 2008 and/or 2012 in their network:



- Proficient in Network Design using Windows Visio or other design software
- Proficient in Cisco and HP network technologies CCNA certification is a plus
- Proficient in PBX technologies AVAYA is a plus
- Proficient in Windows Server 2012,2008

As a penetration tester, we can see that the company is using specific types of technologies within their IT infrastructure. From a penetration tester's point of view, if this organization were our target for a penetration test, we could now narrow our scope of attacks to these specific technologies.

Now that we have completed this section on better understanding OSINT, let's dive into the practical of using OSINT tools.

Using the top OSINT tools

In this section, I will demonstrate some of the most popular OSINT tools that are available for Kali Linux. Each tool will help us create a profile about a target using various sources of information that can be found on the internet.

Over the next few sub-sections, we will cover the following OSINT tools:

- Maltego
- Recon-ng
- theHarvester
- Shodan
- OSRFramework

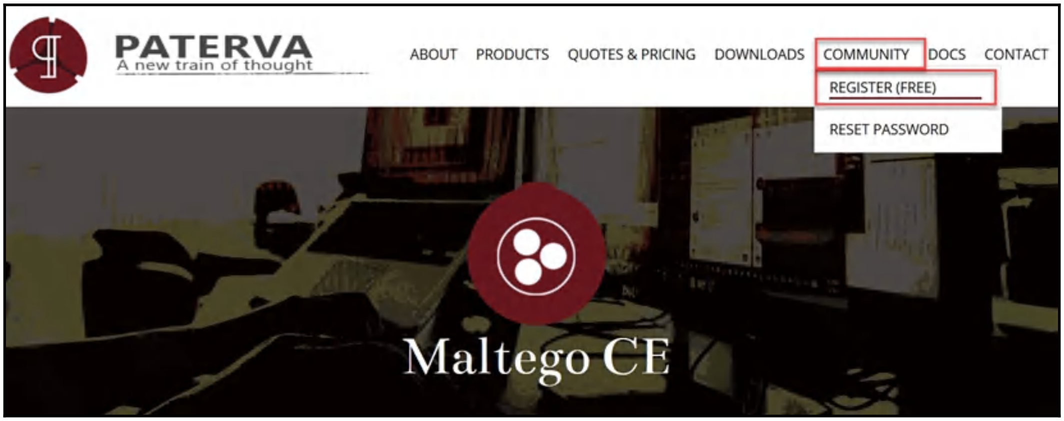
Now, let's take a deeper dive into each of these amazing tools.

Maltego

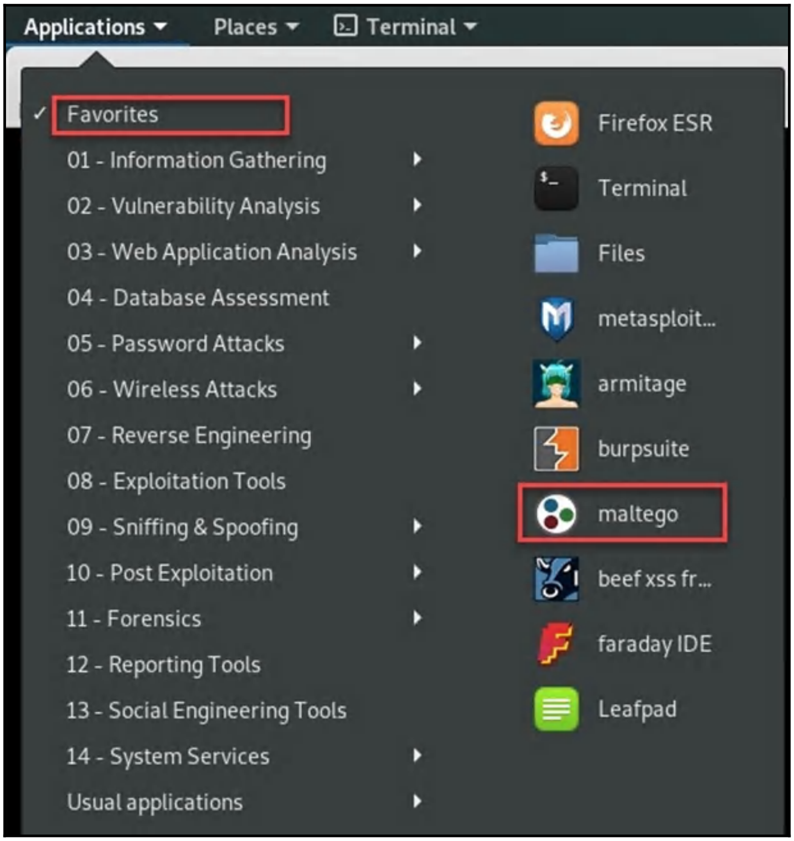
Maltego was created by **Paterva** (`www.paterva.com`) as a graphical interactive data mining application with the ability to query and gather information from various sources on the internet and present data in easy-to-read graphs. The graphs demonstrate the relationship between each entity and the target.

To get started, you need a user account to access the functions and features of Maltego:

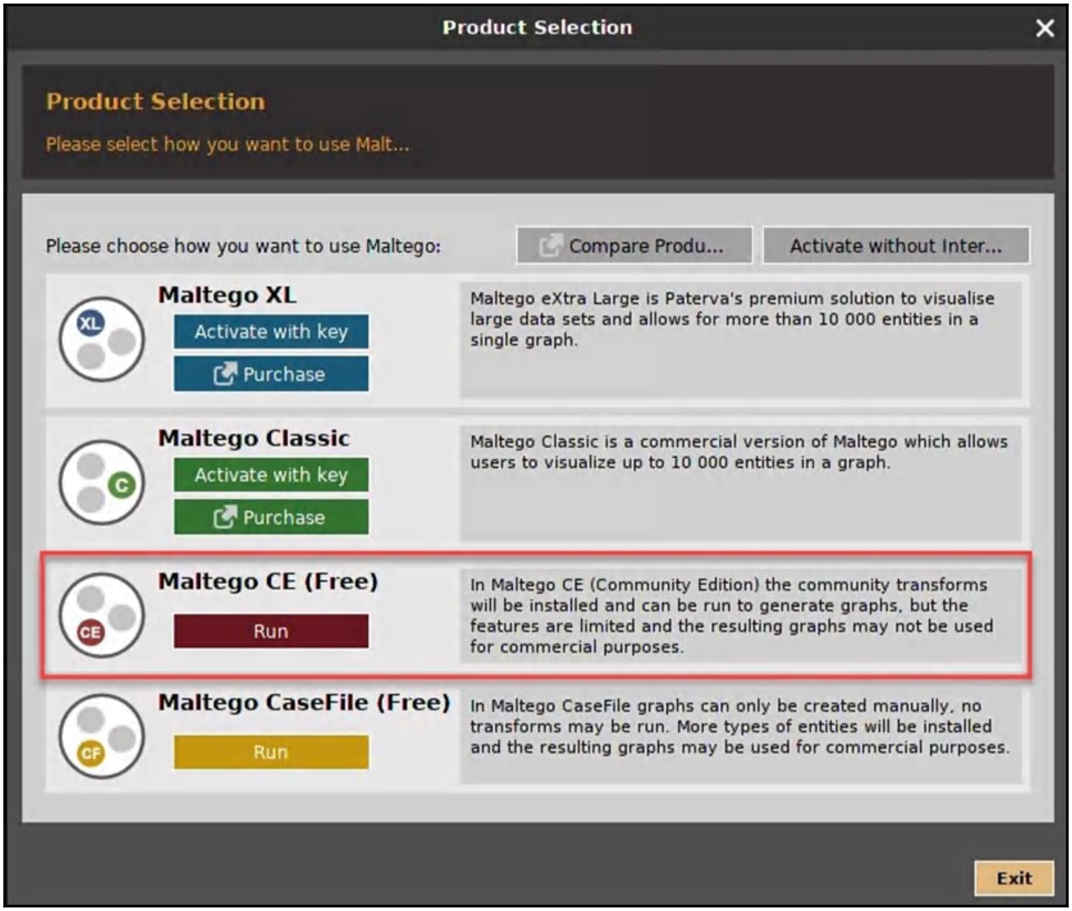
1. Go to `www.paterva.com` and click on **COMMUNITY**. A drop-down menu will be presented. Click on **REGISTER (FREE)** to create a user account:



2. After creating the user account, please ensure that you verify your email address prior to logging in. Once this step has been completed, head back to your Kali Linux desktop. Open the **Maltego** tool by clicking on **Applications** | **Favorites** | **maltego**, as shown in the following screenshot:



3. Once the application opens, click on **Maltego CE (Free)** to configure and run the community edition of Maltego:



Please select how you want to use Malt...

Please choose how you want to use Maltego:

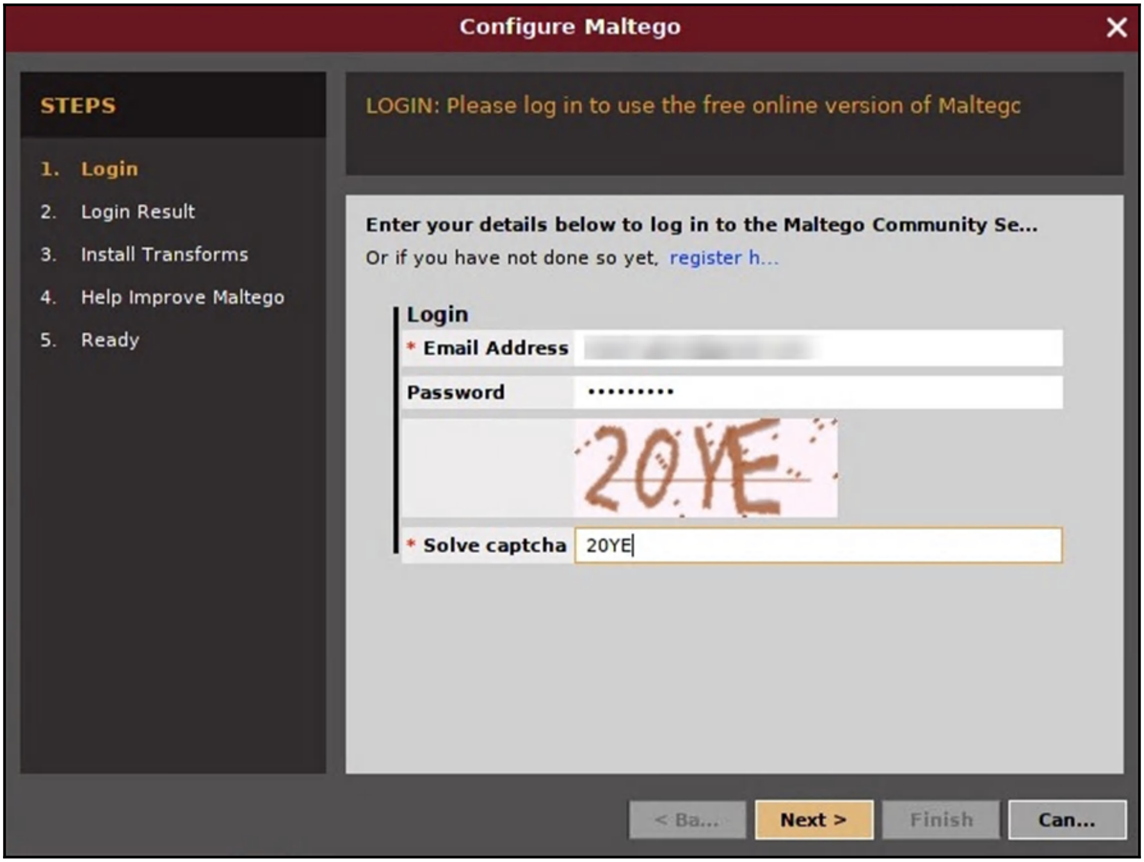
Maltego eXtra Large is Paterva's premium solution to visualise large data sets and allows for more than 10 000 entities in a single graph.

Maltego Classic is a commercial version of Maltego which allows users to visualize up to 10 000 entities in a graph.

In Maltego CE (Community Edition) the community transforms will be installed and can be run to generate graphs, but the features are limited and the resulting graphs may not be used for commercial purposes.

In Maltego CaseFile graphs can only be created manually, no transforms may be run. More types of entities will be installed and the resulting graphs may be used for commercial purposes.

4. Next, you should be presented with the Maltego configuration wizard. Ensure that you log in using the same user account you created previously on the Paterva website and click **Next**. Read and follow the instructions that appear in the next few steps of the configuration wizard. You can leave everything in the default state and click **Next** until the process ends:



Configure Maltego

STEPS

1. Login
2. Login Result
3. Install Transforms
4. Help Improve Maltego
5. Ready

LOGIN: Please log in to use the free online version of Maltegc

Enter your details below to log in to the Maltego Community Se...

Or if you have not done so yet, register h...

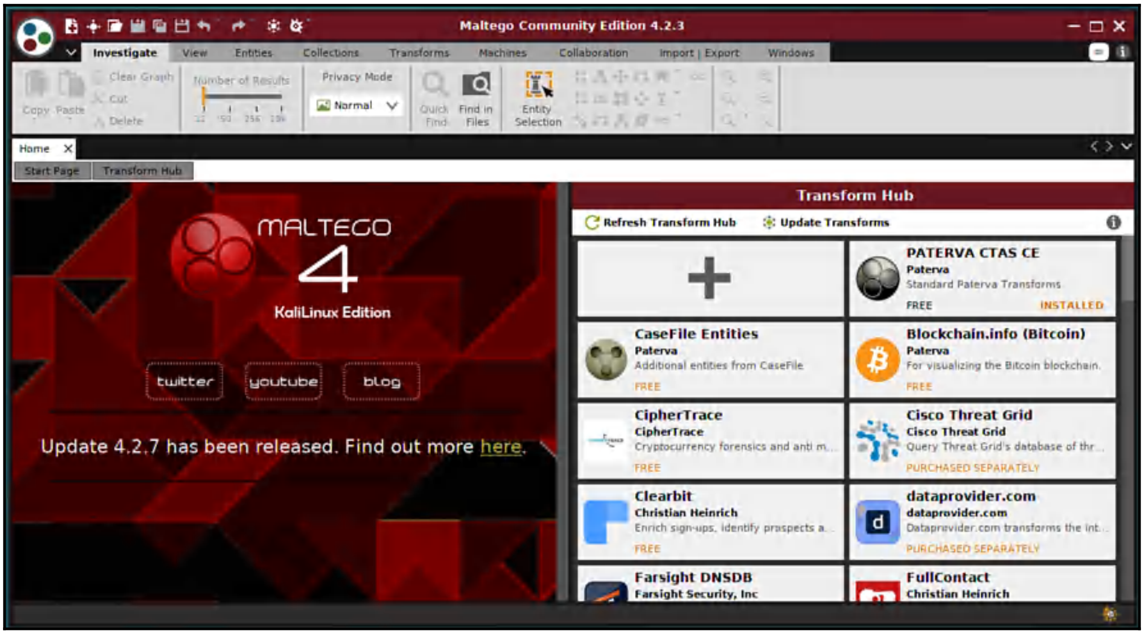
Login

* Email Address

Password ••••••••••

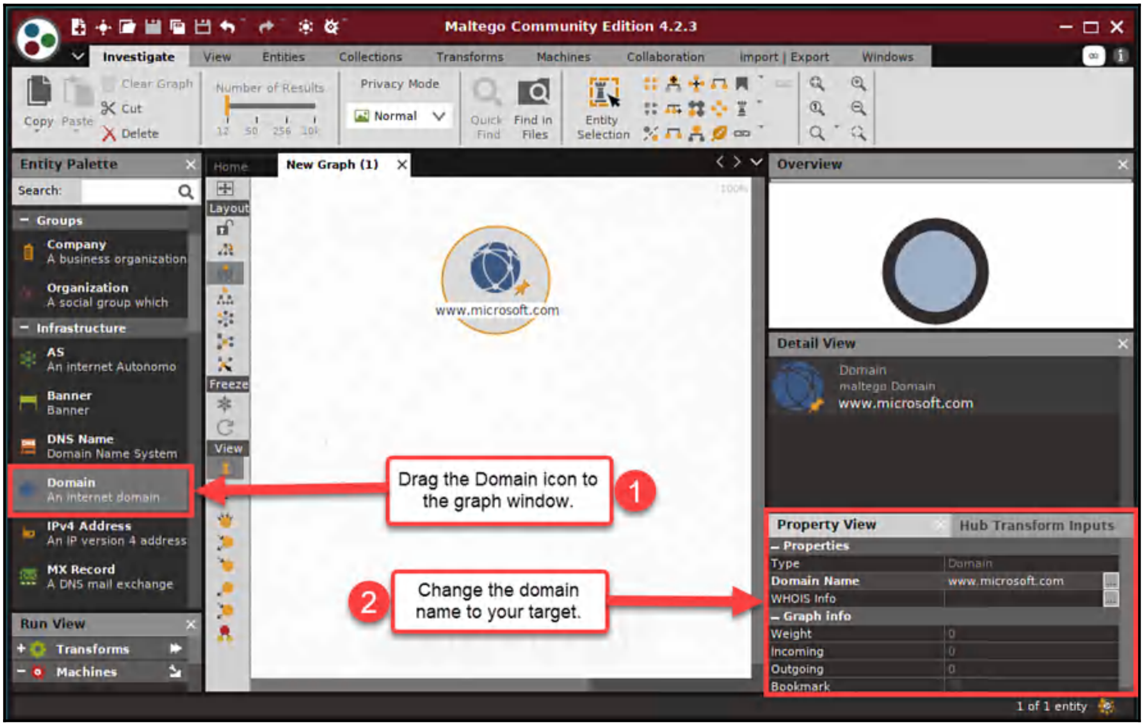
* Solve captcha 20YE

5. Once the configuration window closes, the general Maltego interface is displayed, as shown in the following screenshot. On the start page, there are many transform sets that can be added to Maltego. A **transform** is an open source resource that Maltego can query for information. Adding transforms is optional:



6. To begin gathering information on a target organization, we must first open a new graph. To do this, click on the Maltego icon in the top-left corner, and then click on **New**. Once a new graph has been created, you'll see various types of information (entities) on the left, while, on the right-hand side, you'll see **Overview**, **Detail View**, and **Property View**.

7. To add a domain of a company, click and drag the domain entity to the center of the graph. By default, `paterva.com` will appear as the target domain. Let's change the domain value to something else. On the left-hand side of the interface, click on **Property View**. You will be able to edit the value in the **Domain Name** field, as shown in the following screenshot:



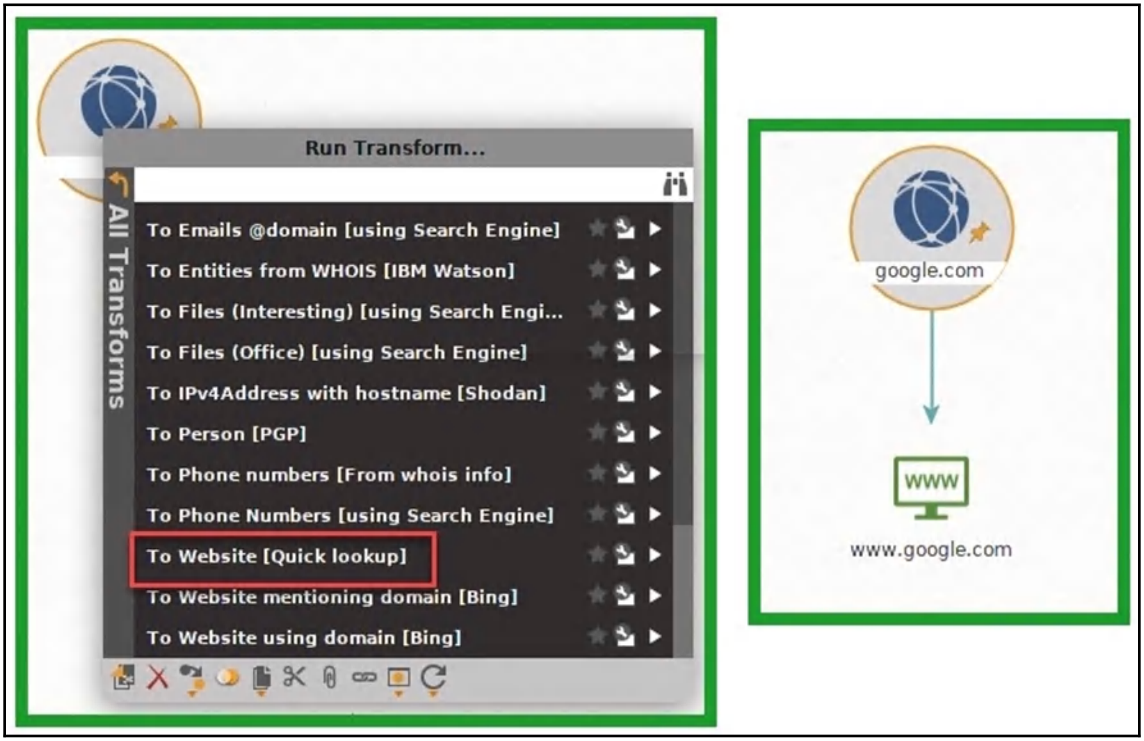
1 Drag the Domain icon to the graph window.

2 Change the domain name to your target.

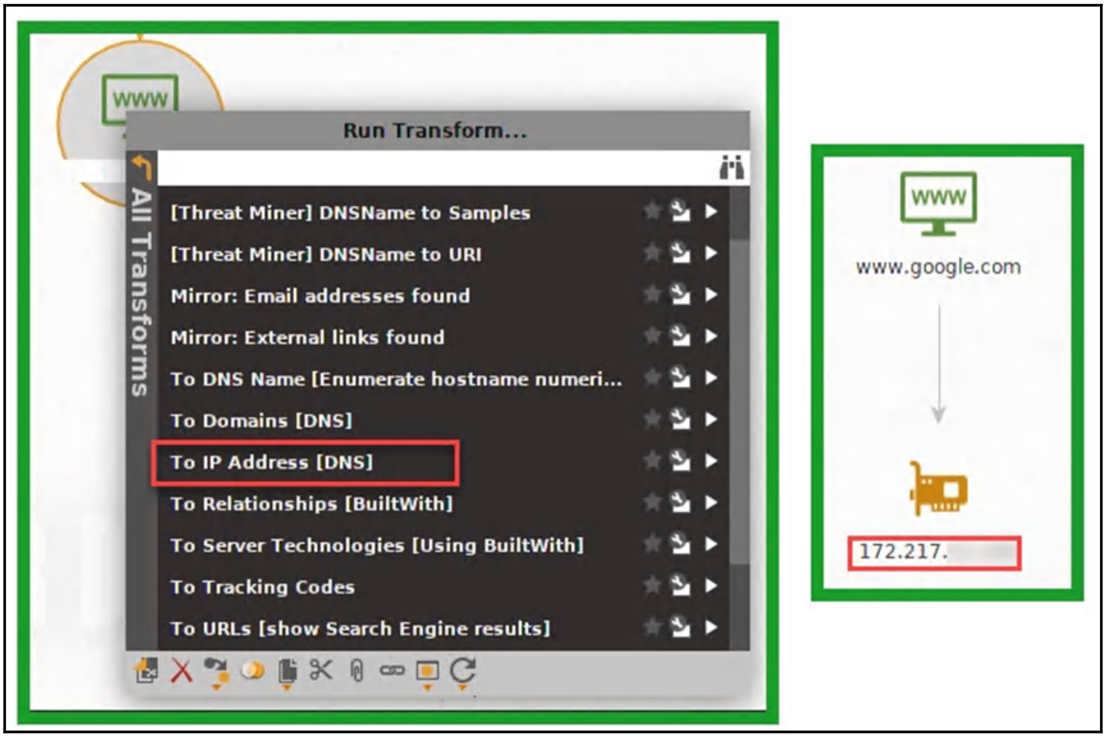
Property View

- Properties
  - Type: Domain
  - Domain Name: www.microsoft.com
  - WHOIS Info:
- Graph info
  - Weight: 0
  - Incoming: 0
  - Outgoing: 0
  - Bookmark:

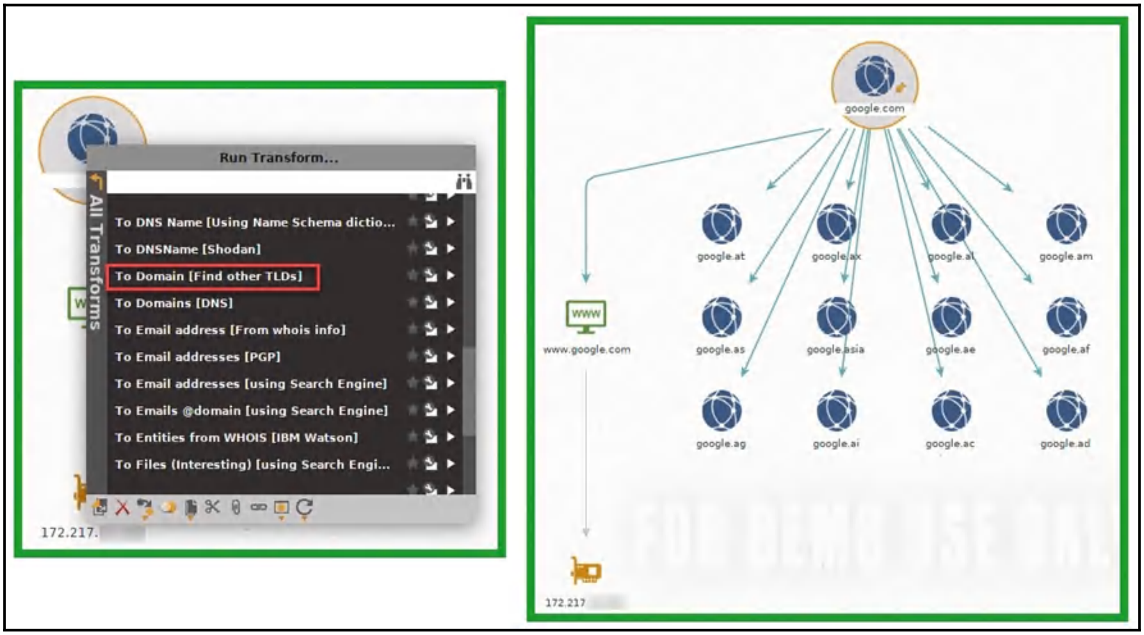
8. Once the domain name has changed, we can proceed and resolve the website URL for the domain. Right-click on the domain entity, click on **All Transforms**, and select **To Website [Quick lookup]**. This transform will simply discover the website address (refer to the screenshot on the left) and display the relationship (refer to the screenshot on the right):


Run Transform...
All Transforms
To Emails @domain [using Search Engine]
To Entities from WHOIS [IBM Watson]
To Files (Interesting) [using Search Engi...
To Files (Office) [using Search Engine]
To IPv4Address with hostname [Shodan]
To Person [PGP]
To Phone numbers [From whois info]
To Phone Numbers [using Search Engine]
To Website [Quick lookup]
To Website mentioning domain [Bing]
To Website using domain [Bing]
google.com
WWW
www.google.com

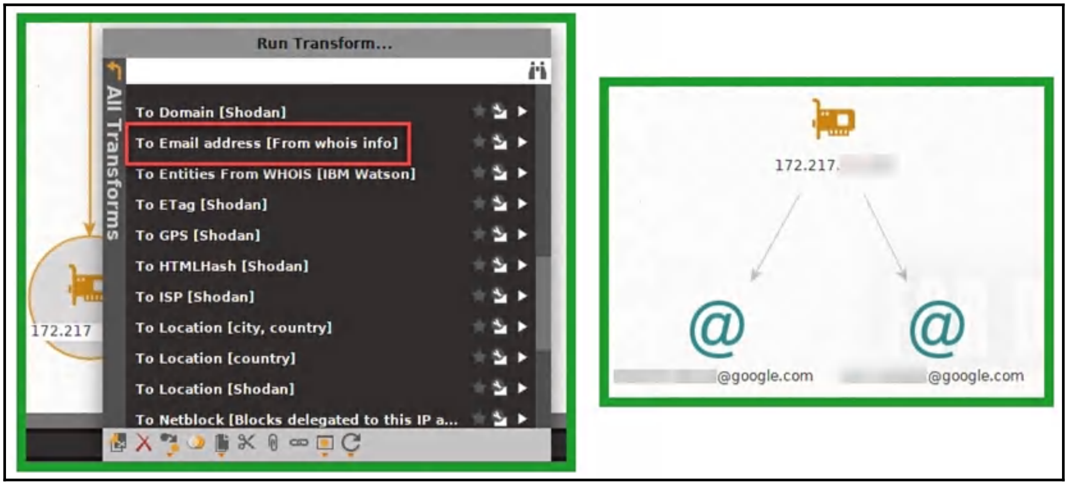
9. Next, we can attempt to obtain the IP address of the website address. Right-click on the website address (`www.google.com`) | **All Transforms** | **To IP Address (DNS)**. The following screenshot on the right displays the IP address that was resolved for `www.google.com`. Please note that this is one of many IP addresses that are used for the `www.google.com` URL:


WWW
Run Transform...
All Transforms
[Threat Miner] DNSName to Samples
[Threat Miner] DNSName to URI
Mirror: Email addresses found
Mirror: External links found
To DNS Name [Enumerate hostname numeri...
To Domains [DNS]
To IP Address [DNS]
To Relationships [BuiltWith]
To Server Technologies [Using BuiltWith]
To Tracking Codes
To URLs [show Search Engine results]
WWW
www.google.com
172.217.[redacted]

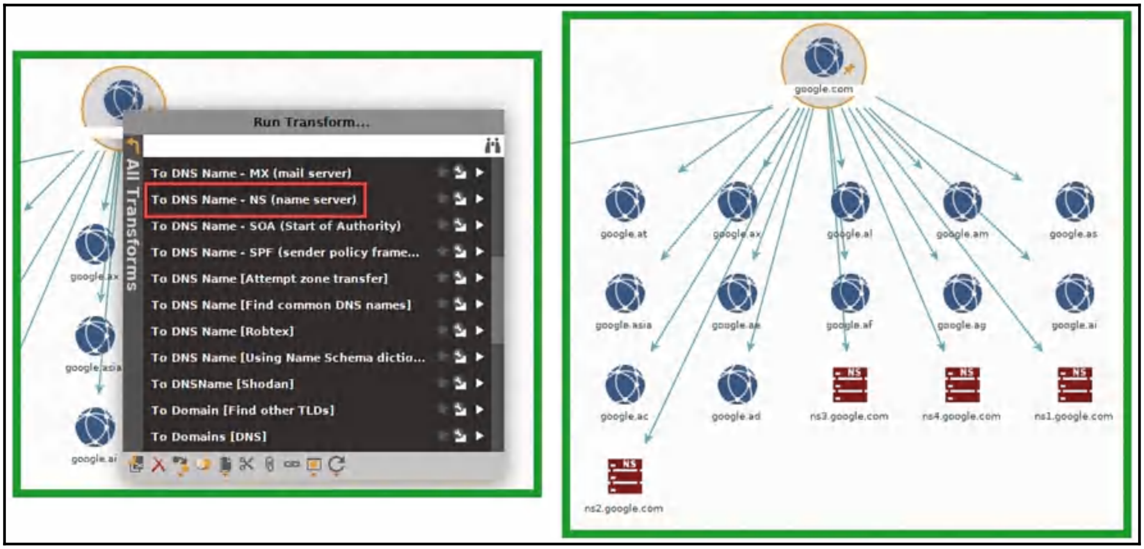
10. We can take this phase even further. How about discovering the **top-level domains** (**TLDs**) that are a part of the `google.com` domain? To complete this task, we begin by right-clicking on the domain entity (`google.com`) and selecting **To Domain [Find other TLDs]**, as shown in the following screenshot on the left. Once the transform has been completed, Maltego will present the information in a tree-like structure on the graph plane, as shown in the following screenshot on the right:


Run Transform...
All Transforms
To DNS Name [Using Name Schema dictio...
To DNSName [Shodan]
To Domain [Find other TLDs]
To Domains [DNS]
To Email address [From whois info]
To Email addresses [PGP]
To Email addresses [using Search Engine]
To Emails @domain [using Search Engine]
To Entities from WHOIS [IBM Watson]
To Files (Interesting) [using Search Engi...
172.217.
google.com
www.google.com
google.at
google.ax
google.al
google.am
google.as
google.asia
google.ae
google.af
google.ag
google.ai
google.ac
google.ad
172.217
FOR DEMO USE ONLY

11. Now, how about gathering the email addresses of contacts who are registered to a domain? We can right-click on the IP address entity | **All Transforms** | **To Email address [From whois info]**. If there are any email addresses, they will be displayed, as shown in the screenshot on the right:



12. Furthermore, we can attempt to obtain the **name servers** (**NSes**) for the domain. Right-click on the domain entity (`google.com`) | **All Transforms** | **To DNS Name – NS (name server)**. All the NSes for the domain will be presented, as shown in the screenshot on the right:.



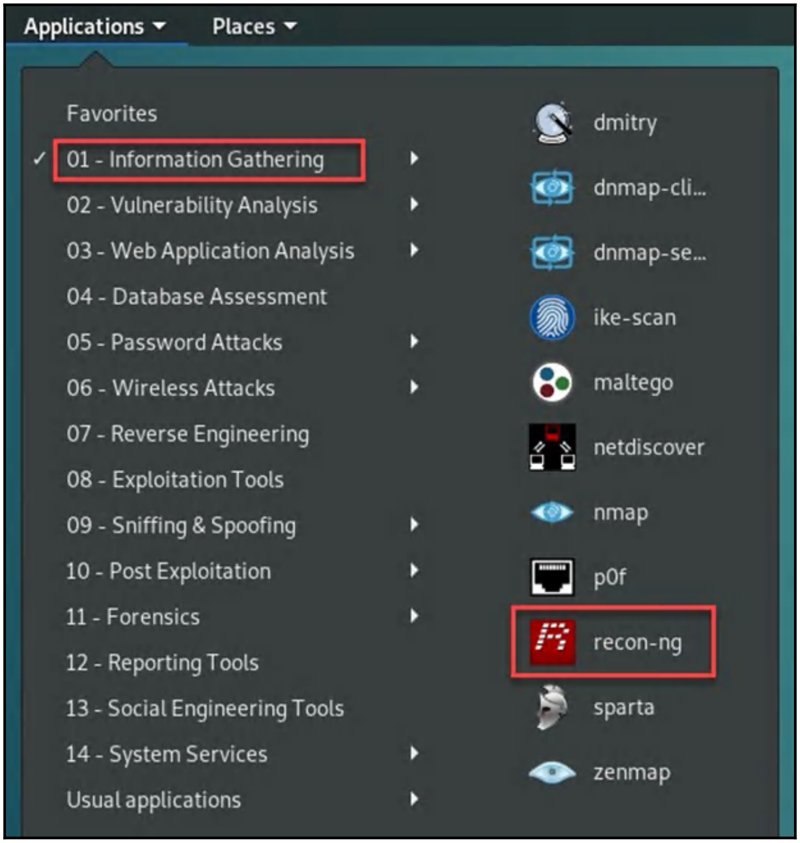
Now, you have a better idea of the functions of Maltego and how to navigate the various transforms. A nice feature of Maltego is the relationship mapping on the graph to help you analyze information and entities.

Having completed this section, you are now familiar with using Maltego to gather information. In the next section, we will use a Python-based tool to assist us in passive information gathering.

Recon-ng

Recon-ng is an OSINT reconnaissance framework written in Python. The tool itself is complete with its own modules, database, interactive help, and menu system, similar to Metasploit. Recon-ng is able to perform web-based, information-gathering techniques using various open source platforms.

Recon-ng is already part of the arsenal of tools in Kali Linux. To access the interface of Recon-ng, simply click on **Applications** in the top-left corner to expand the application menu, and then select **01 – Information Gathering**. You should see **recon-ng**. Click on it to open the framework:



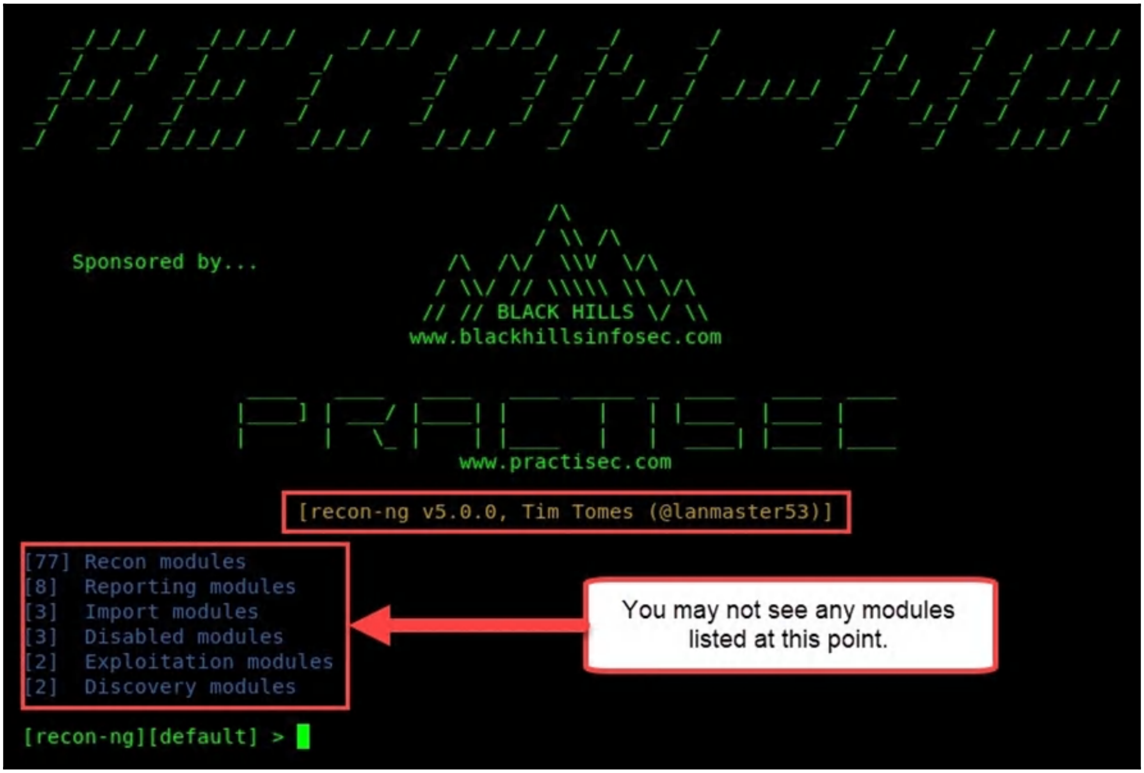
Additionally, you can open the Linux Terminal window and type `recon-ng` to run the framework.

To download and set up the latest version of Recon-ng, use the following instructions:

1. Execute the following configurations on your Terminal to download the latest version of Recon-ng and install it:

```
git clone https://github.com/lanmaster53/recon-ng.git
cd recon-ng
pip install -r REQUIREMENTS
./recon-ng
```

After you're finished, your screen should be similar to the following screenshot:



```
Sponsored by...
                                  // // BLACK HILLS \/ \\
                                 www.blackhillsinfosec.com

                                    www.practisec.com

                     [recon-ng v5.0.0, Tim Tomes (@lanmaster53)]

[77] Recon modules
[8]  Reporting modules
[3]  Import modules
[3]  Disabled modules
[2]  Exploitation modules
[2]  Discovery modules

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

You may not see any modules listed at this point.

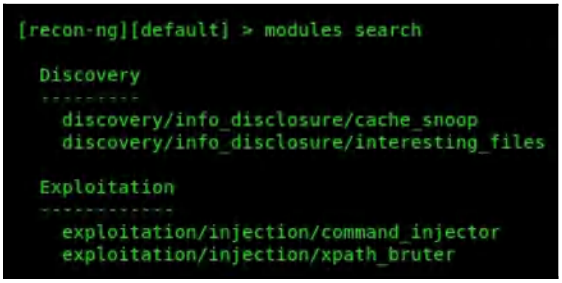
2. If there are no modules installed or enabled, use the following command to install all modules onto Recon-ng:

```
marketplace install all
```

Optionally, you can reload all the modules by using the `modules reload` command.

Now that the Recon-ng framework has opened, let's get familiar with its interface by using the following instructions:

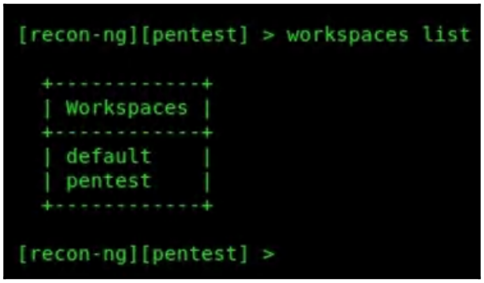
1. To view all the modules in **Recon-ng v5**, we must use the `modules search` command to view a list of all available modules. All the modules will be presented under their categories (`Discovery`, `Exploitation`, `Recon`, `Reporting`, and so on), as shown in the following screenshot:



```
[recon-ng][default] > modules search

  Discovery
  ---------
    discovery/info_disclosure/cache_snoop
    discovery/info_disclosure/interesting_files

  Exploitation
  ------------
    exploitation/injection/command_injector
    exploitation/injection/xpath_bruter
```

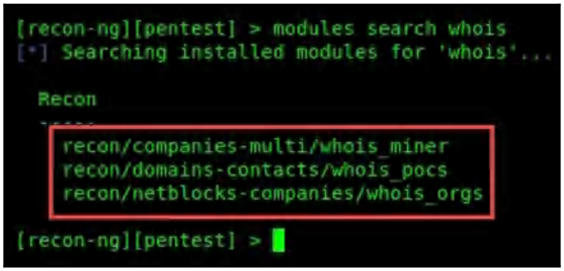
As a penetration tester, you may use the same tools for multiple organizations throughout your career. We can create workspaces within Recon-ng to help isolate our projects/data more efficiently. If you look closely at the command-line interface, you will see the word `default` between square brackets. This implies that we are currently within the default workspace of Recon-ng.

2. To create a workspace, we can use the `workspaces create <worksplace-name>` command. We are going to create a new workspace named `pentest`. Use the `workspaces create pentest` command within the Recon-ng interface.
3. To view all existing workspaces and to verify the creation of our new workspace, we can use the `workspaces list` command:



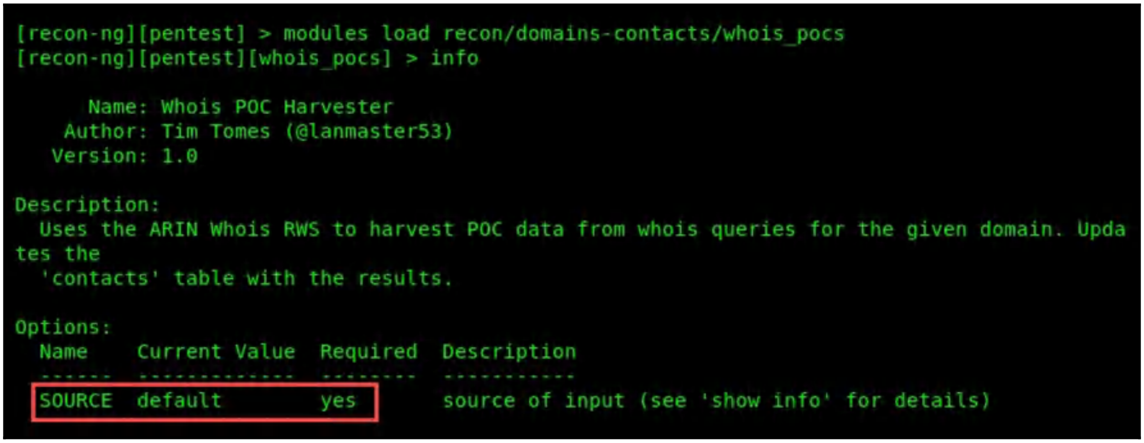
```
[recon-ng][pentest] > workspaces list

  +------------+
  | Workspaces |
  +------------+
  | default    |
  | pentest    |
  +------------+

[recon-ng][pentest] >
```

4. To use the `pentest` workspace, use the `workspaces select pentest` command. The workspace should change on your command-line interface.
5. Now that we have added a company and its domain to our database, let's search for a module to perform a `whois` lookup. We can use the `modules search whois` command, where `whois` is the keyword or search criteria:



```
[recon-ng][pentest] > modules search whois
[*] Searching installed modules for 'whois'...

  Recon
  -----
    recon/companies-multi/whois_miner
    recon/domains-contacts/whois_pocs
    recon/netblocks-companies/whois_orgs

[recon-ng][pentest] >
```

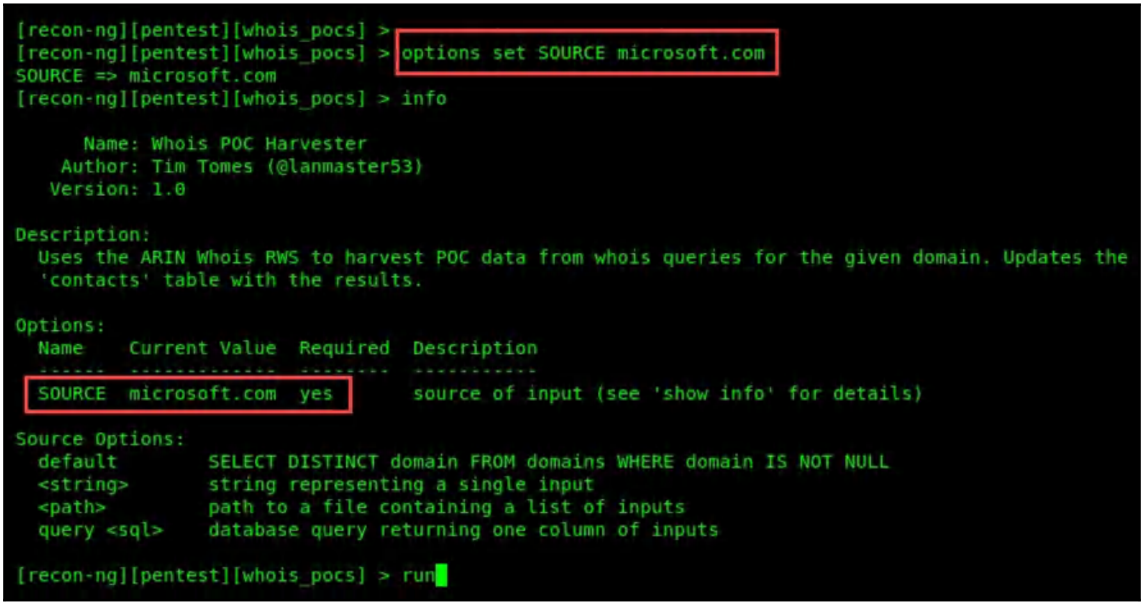
6. We are going to use the **point-of-contacts** (**POCS**) module to obtain further details of people related to the domain. To do this, execute the `modules load recon/domains-contacts/whois_pocs` command. Using the `info` command will provide you with a description of the module and whether additional parameters are required:



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

      Name: Whois POC Harvester
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Uses the ARIN Whois RWS to harvest POC data from whois queries for the given domain. Upda
tes the
  'contacts' table with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)
```

The `default` value is set by the developer of the module; set your own source for each module within Recon-ng. Additionally, you can use the `info` command to view details about a module. The `input` command will list the input values for the `SOURCE` component of a module. The `input` command is useful to verify what the `SOURCE` values are for a specific module.

7. Let's set the `SOURCE` value to `microsoft.com`; this can be done by using the `options set SOURCE microsoft.com` command, as shown in the following screenshot:



```
[recon-ng][pentest][whois_pocs] >
[recon-ng][pentest][whois_pocs] > options set SOURCE microsoft.com
SOURCE => microsoft.com
[recon-ng][pentest][whois_pocs] > info

      Name: Whois POC Harvester
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Uses the ARIN Whois RWS to harvest POC data from whois queries for the given domain. Updates the
  'contacts' table with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  microsoft.com  yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT domain FROM domains WHERE domain IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs

[recon-ng][pentest][whois_pocs] > run
```

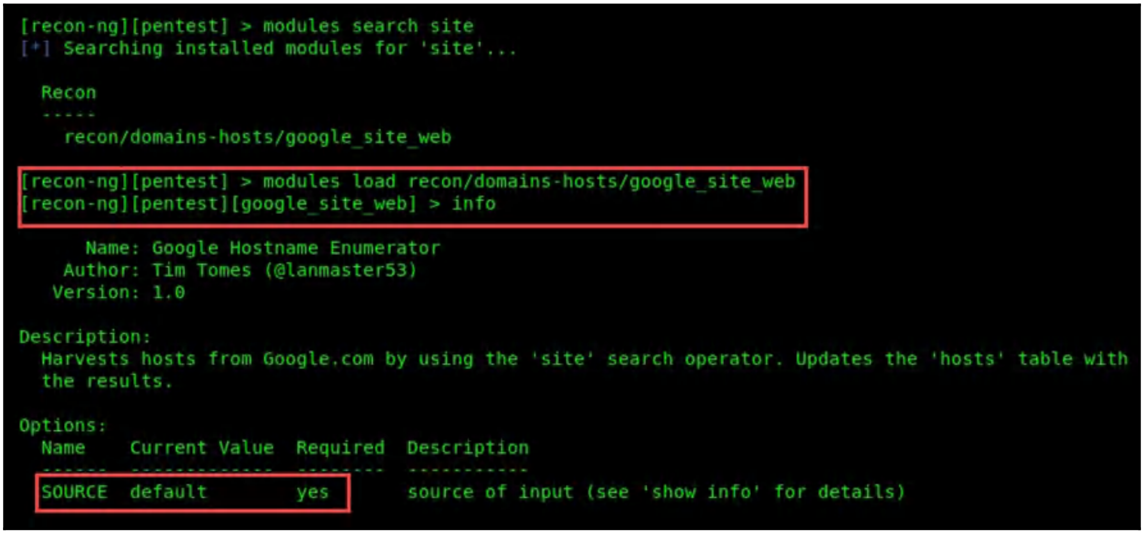
8. Once everything is set correctly, use the `run` command to execute the module against the domain. Once the module has completed its analysis, we can use the `show contacts` command to view the list of information, such as a person's first and last names, email addresses, region, and country.

As you can see, Recon-ng is a very powerful tool and is able to handle data management quite well. Organizations usually create subdomains for many purposes; some can be used as a login portal, or simply as another directory on a website.

To obtain a list of subdomains of a target, observe the following steps:

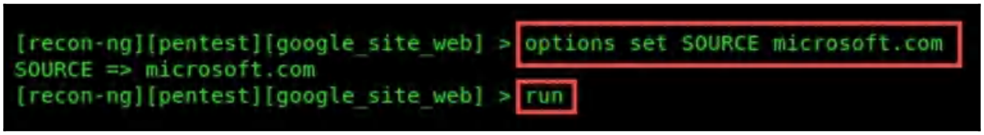
1. Let's start by searching for a suitable module by using the `modules search site` command. Recon-ng will return some of the modules that contain `site` as part of their names.

2. We're going to use the `google_site_web` module. Simply execute the `modules load recon/domains-hosts/google_site_web` command:



```
[recon-ng][pentest] > modules search site
[*] Searching installed modules for 'site'...

  Recon
  -----
    recon/domains-hosts/google_site_web

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

      Name: Google Hostname Enumerator
    Author: Tim Tomes (@lanmaster53)
   Version: 1.0

Description:
  Harvests hosts from Google.com by using the 'site' search operator. Updates the 'hosts' table with
  the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)
```

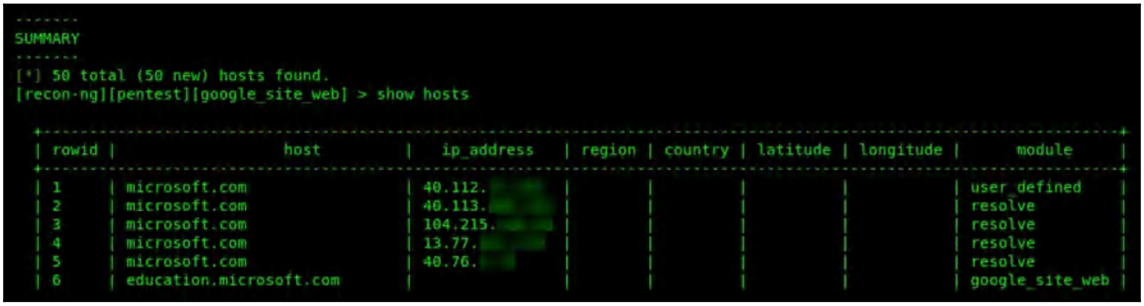
3. Next, let's change the `SOURCE` value for this module by using the `options set SOURCE microsoft.com` command, as shown in the following screenshot:



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

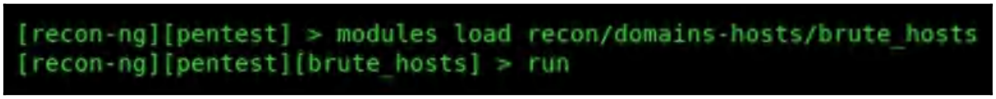
4. Use the `run` command to execute this module.

5. Once the module has finished its query, use the `show hosts` command to view the list of subdomains that were found for the `microsoft.com` domain:



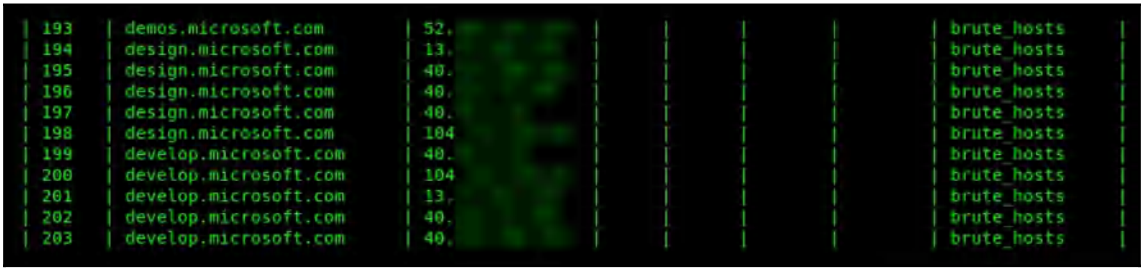
```
-------
SUMMARY
-------
[*] 50 total (50 new) hosts found.
[recon-ng][pentest][google_site_web] > show hosts

  +--------------------------------------------------------------------------------------------------------------+
  | rowid |             host              |   ip_address      | region | country | latitude | longitude |     module      |
  +--------------------------------------------------------------------------------------------------------------+
  | 1     | microsoft.com                 | 40.112.[redacted] |        |         |          |           | user_defined    |
  | 2     | microsoft.com                 | 40.113.[redacted] |        |         |          |           | resolve         |
  | 3     | microsoft.com                 | 104.215.[redacted]|        |         |          |           | resolve         |
  | 4     | microsoft.com                 | 13.77.[redacted]  |        |         |          |           | resolve         |
  | 5     | microsoft.com                 | 40.76.[redacted]  |        |         |          |           | resolve         |
  | 6     | education.microsoft.com       |                   |        |         |          |           | google_site_web |
```

6. Now, let's make this a bit more exciting. How about obtaining the IP addresses for the subdomains? To do this, we are going to run the `module load recon/domains-hosts/brute_hosts` command and set the `SOURCE` value to `microsoft.com`:



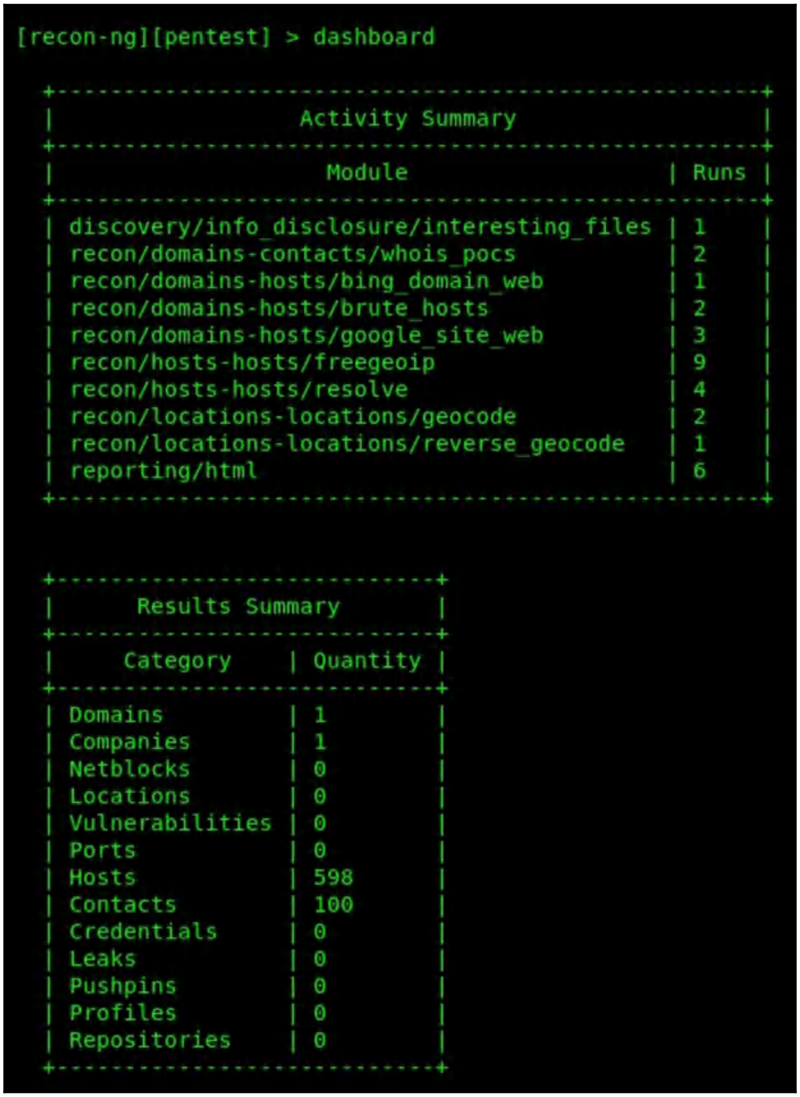
```
[recon-ng][pentest] > modules load recon/domains-hosts/brute_hosts
[recon-ng][pentest][brute_hosts] > run
```

7. Once the module is finished executing, use the `show host` command once more. You should see IP addresses corresponding to a subdomain:



```
| 193 | demos.microsoft.com     | 52.[redacted]  |  |  |  |  | brute_hosts |
| 194 | design.microsoft.com    | 13.[redacted]  |  |  |  |  | brute_hosts |
| 195 | design.microsoft.com    | 40.[redacted]  |  |  |  |  | brute_hosts |
| 196 | design.microsoft.com    | 40.[redacted]  |  |  |  |  | brute_hosts |
| 197 | design.microsoft.com    | 40.[redacted]  |  |  |  |  | brute_hosts |
| 198 | design.microsoft.com    | 104[redacted]  |  |  |  |  | brute_hosts |
| 199 | develop.microsoft.com   | 40.[redacted]  |  |  |  |  | brute_hosts |
| 200 | develop.microsoft.com   | 104[redacted]  |  |  |  |  | brute_hosts |
| 201 | develop.microsoft.com   | 13.[redacted]  |  |  |  |  | brute_hosts |
| 202 | develop.microsoft.com   | 40.[redacted]  |  |  |  |  | brute_hosts |
| 203 | develop.microsoft.com   | 40.[redacted]  |  |  |  |  | brute_hosts |
```

As a penetration tester, writing reports can be very overwhelming as a report is a summary of the actions performed and results obtained while you were performing various tasks and using a number of tools. Recon-ng has a few reporting modules that are able to generate reports in multiple formats. The `dashboard` command will provide a summary of tasks performed with Recon-ng, as shown here:

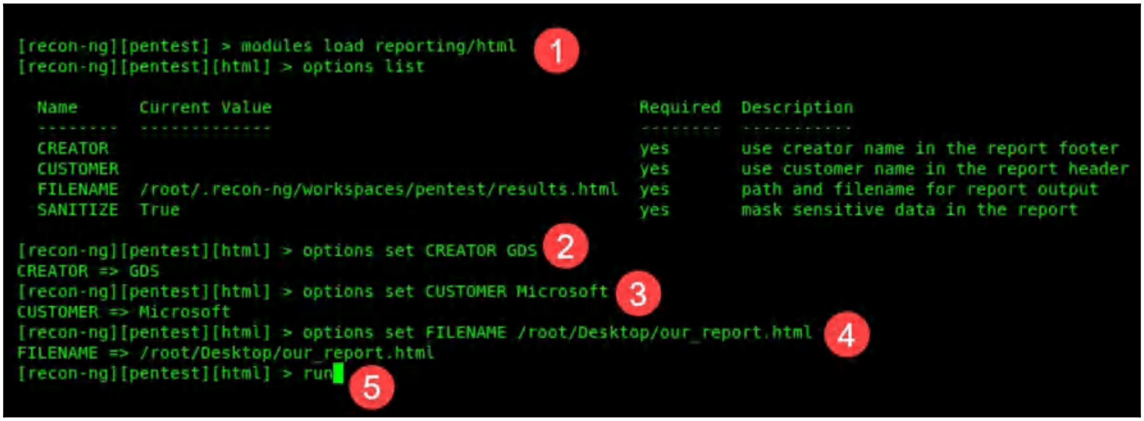


```
[recon-ng][pentest] > dashboard

  +------------------------------------------------------+
  |                  Activity Summary                    |
  +------------------------------------------------------+
  |                    Module                     | Runs |
  +------------------------------------------------------+
  | discovery/info_disclosure/interesting_files | 1    |
  | recon/domains-contacts/whois_pocs           | 2    |
  | recon/domains-hosts/bing_domain_web         | 1    |
  | recon/domains-hosts/brute_hosts             | 2    |
  | recon/domains-hosts/google_site_web         | 3    |
  | recon/hosts-hosts/freegeoip                 | 9    |
  | recon/hosts-hosts/resolve                   | 4    |
  | recon/locations-locations/geocode           | 2    |
  | recon/locations-locations/reverse_geocode   | 1    |
  | reporting/html                              | 6    |
  +------------------------------------------------------+


  +-----------------------------+
  |       Results Summary       |
  +-----------------------------+
  |     Category    | Quantity  |
  +-----------------------------+
  | Domains         | 1         |
  | Companies       | 1         |
  | Netblocks       | 0         |
  | Locations       | 0         |
  | Vulnerabilities | 0         |
  | Ports           | 0         |
  | Hosts           | 598       |
  | Contacts        | 100       |
  | Credentials     | 0         |
  | Leaks           | 0         |
  | Pushpins        | 0         |
  | Profiles        | 0         |
  | Repositories    | 0         |
  +-----------------------------+
```

To generate a report, you can use the `modules search report` command. This will let you view a list of reporting modules within the Recon-ng interface. We are going to create a report in HTML format. To create the report, perform the following steps:

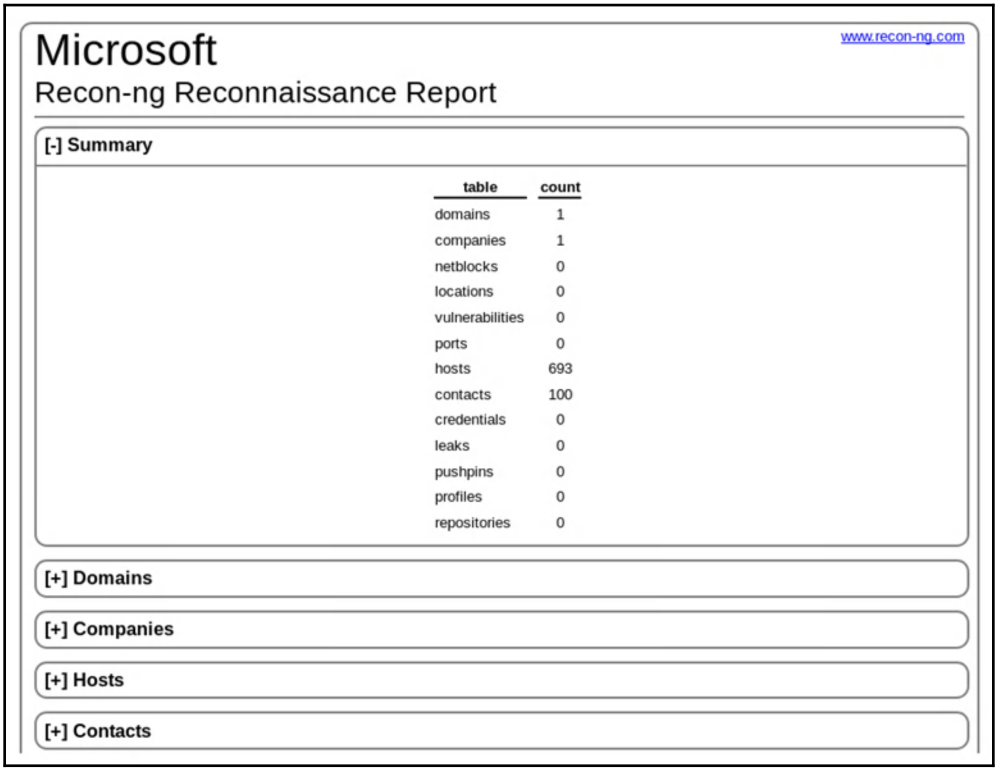
1. Execute the `modules load reporting/html` command.
2. Use the `options set CREATOR` command to set the creator of the report.
3. Use the `options set CUSTOMER` command to set the customer.
4. Use the `options set FILENAME` command to set the output location with the filename of the report.
5. Lastly, use the `run` command to execute the module and generate the report.

The following screenshot has labels that correspond to the aforementioned steps for generating and exporting a report file using `reporting/html module`:



```
[recon-ng][pentest] > modules load reporting/html
[recon-ng][pentest][html] > options list

  Name      Current Value                                       Required  Description
  --------  -------------                                       --------  -----------
  CREATOR                                                       yes       use creator name in the report footer
  CUSTOMER                                                      yes       use customer name in the report header
  FILENAME  /root/.recon-ng/workspaces/pentest/results.html     yes       path and filename for report output
  SANITIZE  True                                                yes       mask sensitive data in the report

[recon-ng][pentest][html] > options set CREATOR GDS
CREATOR => GDS
[recon-ng][pentest][html] > options set CUSTOMER Microsoft
CUSTOMER => Microsoft
[recon-ng][pentest][html] > options set FILENAME /root/Desktop/our_report.html
FILENAME => /root/Desktop/our_report.html
[recon-ng][pentest][html] > run
```

Finally, navigate to the output location on your Kali Linux machine and open the HTML file. The view should be similar to the following:



Microsoft

Recon-ng Reconnaissance Report

www.recon-ng.com

[-] Summary

table	count
domains	1
companies	1
netblocks	0
locations	0
vulnerabilities	0
ports	0
hosts	693
contacts	100
credentials	0
leaks	0
pushpins	0
profiles	0
repositories	0

[+] Domains

[+] Companies

[+] Hosts

[+] Contacts

Now, the information is nicely categorized and summarized for viewing.

In the next section, we'll take a look at using **theHarvester** to gather the email addresses of people associated with an organization.

theHarvester

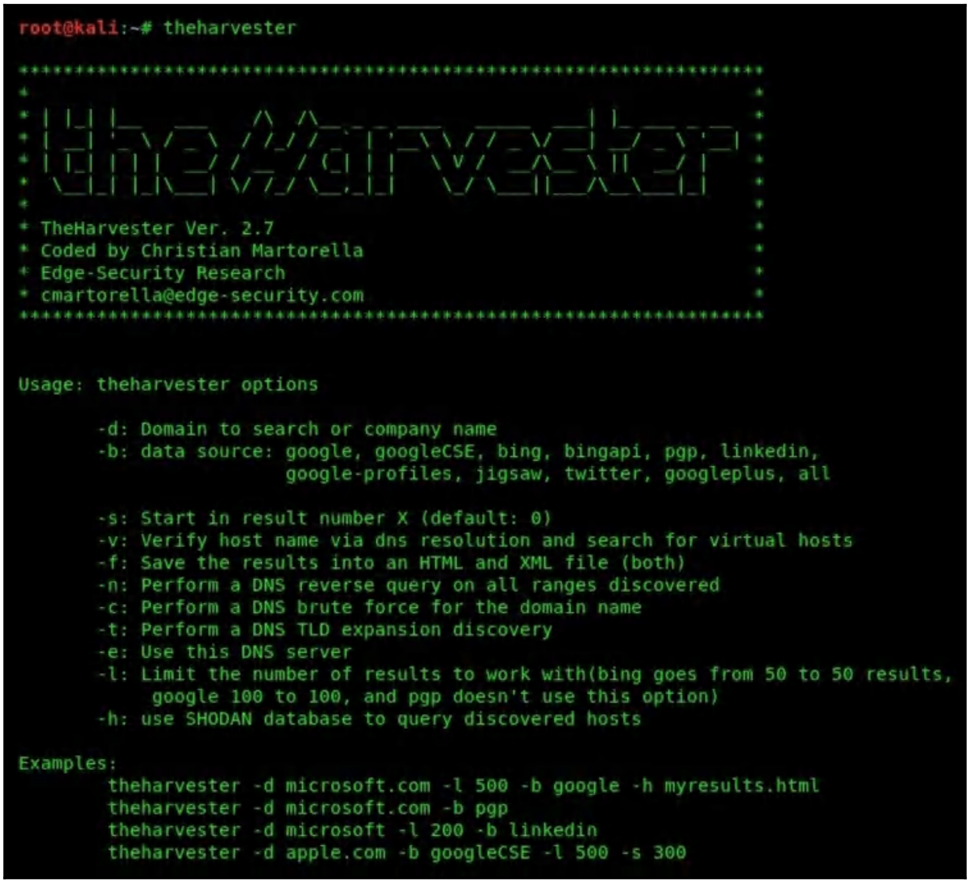
theHarvester is designed to gather email addresses, domains, and employee details for a given company. theHarvester uses multiple open sources on the internet, such as search engines, to piece together details in a readable format.

Further details on theHarvester can be found on its GitHub page:

`https://github.com/laramies/theHarvester`

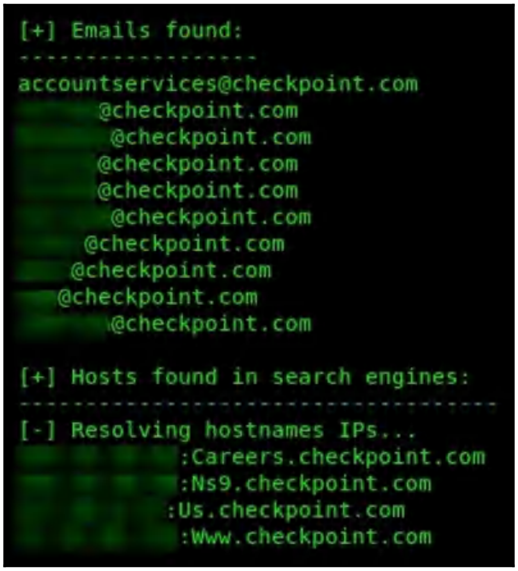
To get started with theHarvester, execute the following steps:

1. Open a Terminal window and execute `theharvester`. The description and usage of the tool will be presented on the Terminal, as shown in the following screenshot:



```
root@kali:~# theharvester

*******************************************************************
*                                                                 *
* | |_| |__   ___    /\  /\__ _ _ ____   _____  ___| |_ ___ _ __  *
* | __| '_ \ / _ \  / /_/ / _` | '__\ \ / / _ \/ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* TheHarvester Ver. 2.7                                           *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*******************************************************************


Usage: theharvester options

       -d: Domain to search or company name
       -b: data source: google, googleCSE, bing, bingapi, pgp, linkedin,
                        google-profiles, jigsaw, twitter, googleplus, all

       -s: Start in result number X (default: 0)
       -v: Verify host name via dns resolution and search for virtual hosts
       -f: Save the results into an HTML and XML file (both)
       -n: Perform a DNS reverse query on all ranges discovered
       -c: Perform a DNS brute force for the domain name
       -t: Perform a DNS TLD expansion discovery
       -e: Use this DNS server
       -l: Limit the number of results to work with(bing goes from 50 to 50 results,
            google 100 to 100, and pgp doesn't use this option)
       -h: use SHODAN database to query discovered hosts

Examples:
        theharvester -d microsoft.com -l 500 -b google -h myresults.html
        theharvester -d microsoft.com -b pgp
        theharvester -d microsoft -l 200 -b linkedin
        theharvester -d apple.com -b googleCSE -l 500 -s 300
```

2. Let's attempt to gather the email addresses of employees of a company whose email addresses are published publicly, such as on forums, websites, blogs, and social media platforms. We can use the `theharvester –d <domain> –b <data source>` command to do this. In our example, we'll search for email addresses of the `checkpoint.com` domain while using Google as the data source:



```
[+] Emails found:
------------------
accountservices@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com
[redacted]@checkpoint.com

[+] Hosts found in search engines:
------------------------------------
[-] Resolving hostnames IPs...
[redacted]:Careers.checkpoint.com
[redacted]:Ns9.checkpoint.com
[redacted]:Us.checkpoint.com
[redacted]:Www.checkpoint.com
```

The results have provided us with some corporate email accounts of some of the employees of the company and the IP addresses of some subdomains. I would recommend using various data sources to gather as much information as possible. One purpose of gathering the email addresses of a company is to perform phishing attacks.

Next, we are going to use the Shodan search engine, which indexes **Internet of Things** (**IoT**) and other online devices to retrieve information about a potential target.

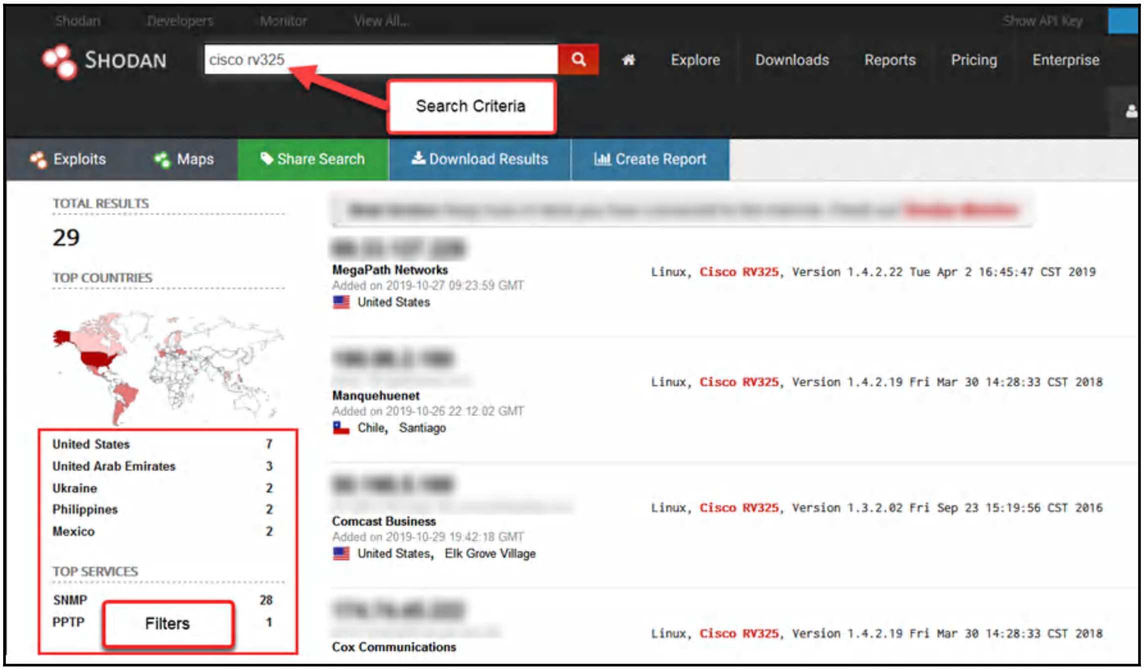
Shodan

Shodan (`www.shodan.io`) is a search engine that indexes various devices that are connected to the internet. What does this mean? To elaborate, let's take a real-life example of discovering devices of a certain vulnerability level. In January 2019, Hacker News (`https://thehackernews.com`) published an article indicating that over 9,000 Cisco SMB RV320 and RV325 routers were globally affected by a new exploit. The exploits were CVE-2019-1652 and CVE-2019-1653, and they allow a malicious person to obtain configuration files and gain control of devices.

This article can be found at the following URL:

`https://thehackernews.com/2019/01/hacking-cisco-routers.html`

Imagine that you're interested in discovering all the devices of this nature on the internet. Using Shodan as a regular search engine, we can quickly discover multiple devices that fit our search criteria of `cisco rv325`, as shown in the following screenshot. Look closely: we can see a list of online Cisco RV325 routers, their IP addresses, their hostnames, and their locations:



TOTAL RESULTS

29

TOP COUNTRIES

United States	7
United Arab Emirates	3
Ukraine	2
Philippines	2
Mexico	2

TOP SERVICES

SNMP	28
PPTP	1

[redacted]

[redacted]
MegaPath Networks
Added on 2019-10-27 09:23:59 GMT
United States

```
Linux, Cisco RV325, Version 1.4.2.22 Tue Apr 2 16:45:47 CST 2019
```

[redacted]
Manquehuenet
Added on 2019-10-26 22:12:02 GMT
Chile, Santiago

```
Linux, Cisco RV325, Version 1.4.2.19 Fri Mar 30 14:28:33 CST 2018
```

[redacted]
Comcast Business
Added on 2019-10-29 19:42:18 GMT
United States, Elk Grove Village

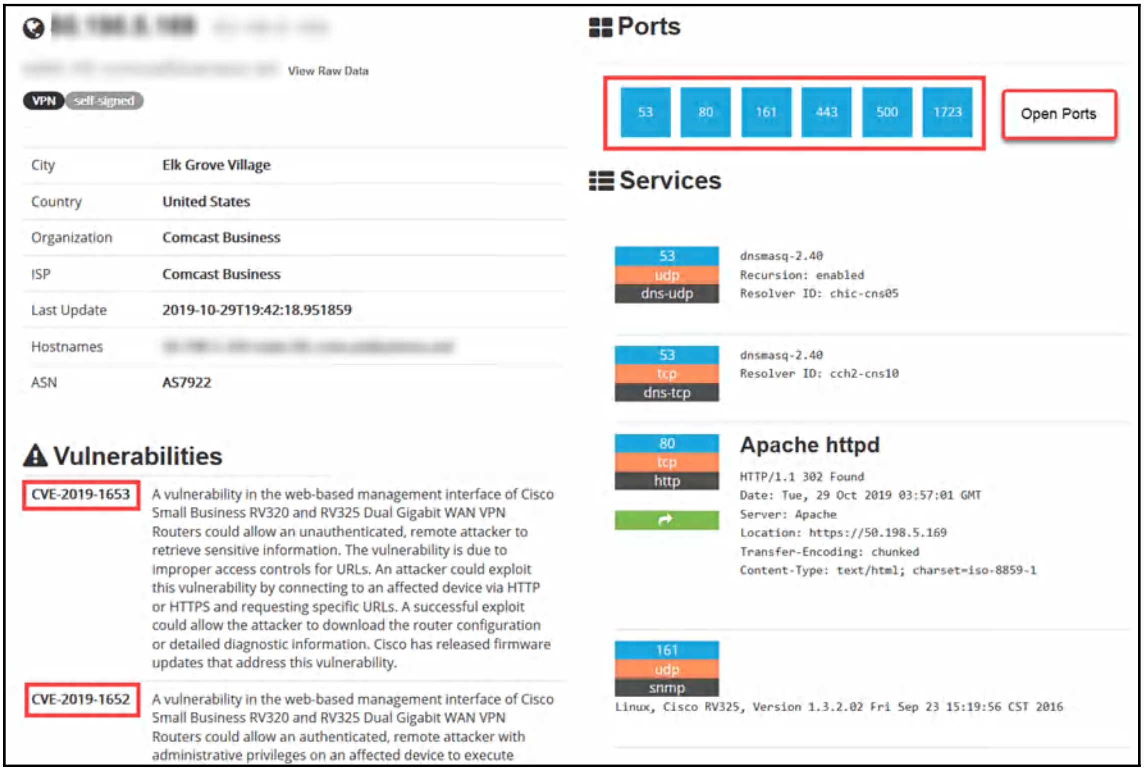
```
Linux, Cisco RV325, Version 1.3.2.02 Fri Sep 23 15:19:56 CST 2016
```

[redacted]
Cox Communications

```
Linux, Cisco RV325, Version 1.4.2.19 Fri Mar 30 14:28:33 CST 2018
```

The results provide geolocation information for the devices and IP addresses, and banner information such as firmware versions. On the left, we can see a global map indicating the number of internet-connected devices per country and organization. Simply clicking on a country, an organization, or even an IP address will filter that information for us.

Clicking on an IP address will provide greater insight into the selected device, such as the hostname, open ports, running services, organization, **Internet Service Provider** (**ISP**) details, and the vulnerabilities the device is susceptible to, as shown in the following screenshot:



[redacted]

[redacted] View Raw Data

VPN self-signed

City	Elk Grove Village
Country	United States
Organization	Comcast Business
ISP	Comcast Business
Last Update	2019-10-29T19:42:18.951859
Hostnames	[redacted]
ASN	AS7922

Vulnerabilities

CVE-2019-1653 A vulnerability in the web-based management interface of Cisco Small Business RV320 and RV325 Dual Gigabit WAN VPN Routers could allow an unauthenticated, remote attacker to retrieve sensitive information. The vulnerability is due to improper access controls for URLs. An attacker could exploit this vulnerability by connecting to an affected device via HTTP or HTTPS and requesting specific URLs. A successful exploit could allow the attacker to download the router configuration or detailed diagnostic information. Cisco has released firmware updates that address this vulnerability.

CVE-2019-1652 A vulnerability in the web-based management interface of Cisco Small Business RV320 and RV325 Dual Gigabit WAN VPN Routers could allow an authenticated, remote attacker with administrative privileges on an affected device to execute

Ports

53 80 161 443 500 1723

Open Ports

Services

53 udp dns-udp

```
dnsmasq-2.40
Recursion: enabled
Resolver ID: chic-cns05
```

53 tcp dns-tcp

```
dnsmasq-2.40
Resolver ID: cch2-cns10
```

80 tcp http

Apache httpd

```
HTTP/1.1 302 Found
Date: Tue, 29 Oct 2019 03:57:01 GMT
Server: Apache
Location: https://50.198.5.169
Transfer-Encoding: chunked
Content-Type: text/html; charset=iso-8859-1
```

161 udp snmp

```
Linux, Cisco RV325, Version 1.3.2.02 Fri Sep 23 15:19:56 CST 2016
```

Information that's gathered from Shodan can help you build a better profile of your target organization during a penetration test as it can provide you with possible operating system versions and other technical details that you can use to successfully exploit your target. Put simply, Shodan can help you identify the target's technologies and control systems in their organization and IT infrastructure.

In the next section, we'll learn about **OSRFramework**.

OSRFramework

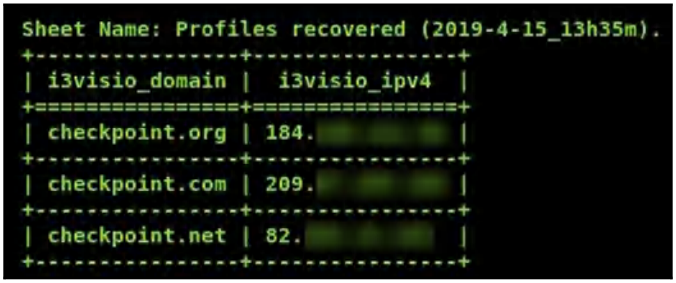
Another awesome OSINT tool is OSRFramework. This toolset performs lookups using usernames, DNS records, deep web searches, and much more.

To begin, we can execute the `osrf` command on the Terminal to provide a description of the usage of the tool itself. Let's imagine that we need to obtain a list of corporate email addresses of our target company on a social media platform. The following are the components of OSRFramework that we can utilize in order to gather information from various sources:

- `domainfy`: Checks whether domain names that use words and nicknames are available
- `entify`: Extracts entities using regular expressions from provided URIs
- `mailfy`: Gets information about email accounts
- `checkfy`: Verifies whether a given email address matches a pattern
- `phonefy`: Looks for information linked to spam practices by a phone number
- `searchfy`: Performs queries on several platforms
- `usufy`: Looks for registered accounts with given nicknames

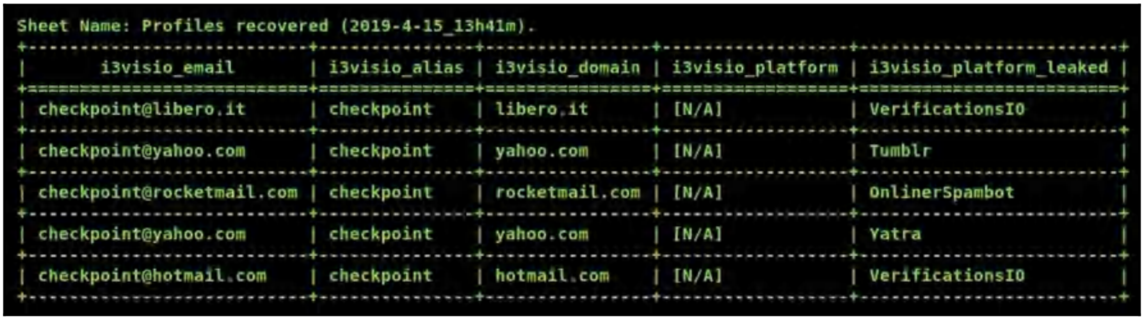
In our first example, we can attempt to obtain information about our target domain using the `whois` database. Using the `domainfy.py --whois –n <target>` syntax, the framework will begin querying `whois` and provide the results in a table format after a few minutes. In our example, I have used the `domainfy.py --whois –n checkpoint` command to specifically retrieve information for any domain that contains the name `checkpoint`.

The following are the results:



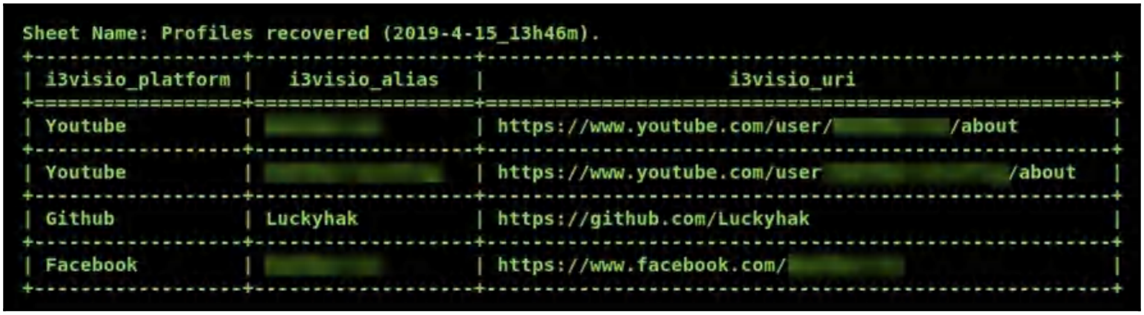
```
Sheet Name: Profiles recovered (2019-4-15_13h35m).
+---------------+---------------+
| i3visio_domain |  i3visio_ipv4  |
+===============+===============+
| checkpoint.org | 184.[redacted] |
+---------------+---------------+
| checkpoint.com | 209.[redacted] |
+---------------+---------------+
| checkpoint.net | 82.[redacted]  |
+---------------+---------------+
```

Next, we can attempt to obtain the email addresses of a given search string. In this second example, we are attempting to discover email addresses that contain the `checkpoint` string, which has been used on various websites on the internet. We can begin by using the `mailfy.py –n checkpoint` command. We will be presented with a table displaying the email addresses that fit our search criteria, domain, and platform location, as shown in the following screenshot:



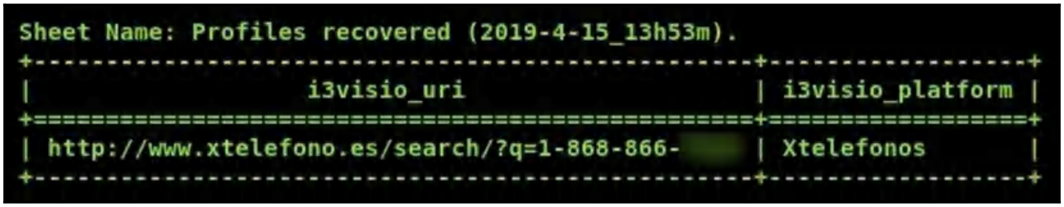
```
Sheet Name: Profiles recovered (2019-4-15_13h41m).
+---------------------------+---------------+----------------+------------------+-------------------------+
|       i3visio_email       | i3visio_alias | i3visio_domain | i3visio_platform | i3visio_platform_leaked |
+===========================+===============+================+==================+=========================+
| checkpoint@libero.it      | checkpoint    | libero.it      | [N/A]            | VerificationsIO         |
+---------------------------+---------------+----------------+------------------+-------------------------+
| checkpoint@yahoo.com      | checkpoint    | yahoo.com      | [N/A]            | Tumblr                  |
+---------------------------+---------------+----------------+------------------+-------------------------+
| checkpoint@rocketmail.com | checkpoint    | rocketmail.com | [N/A]            | OnlinerSpambot          |
+---------------------------+---------------+----------------+------------------+-------------------------+
| checkpoint@yahoo.com      | checkpoint    | yahoo.com      | [N/A]            | Yatra                   |
+---------------------------+---------------+----------------+------------------+-------------------------+
| checkpoint@hotmail.com    | checkpoint    | hotmail.com    | [N/A]            | VerificationsIO         |
+---------------------------+---------------+----------------+------------------+-------------------------+
```

In our third example, we are going to use a string to search across all the services of OSRFramework. To achieve this task, use the `seachfy.py –q string` command on your Terminal. Once completed, the results are displayed and tell you about the location that was found, any aliases, and the URLs, as shown in the following screenshot:



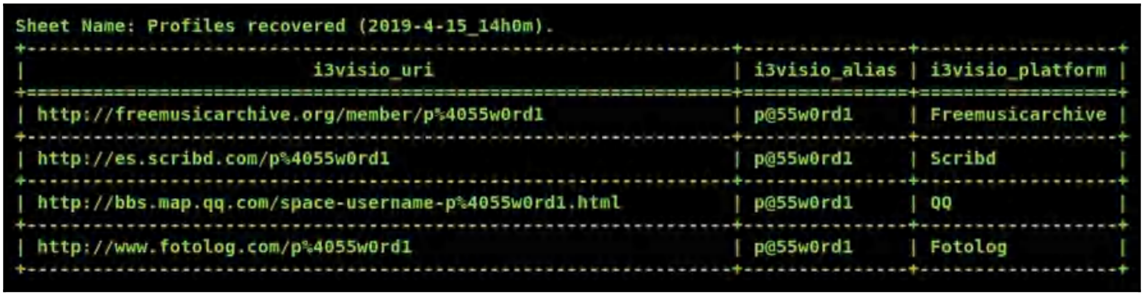
```
Sheet Name: Profiles recovered (2019-4-15_13h46m).
+------------------+-------------------+---------------------------------------------------------+
| i3visio_platform |   i3visio_alias   |                       i3visio_uri                       |
+==================+===================+=========================================================+
| Youtube          | [redacted]        | https://www.youtube.com/user/[redacted]/about           |
+------------------+-------------------+---------------------------------------------------------+
| Youtube          | [redacted]        | https://www.youtube.com/user[redacted]/about            |
+------------------+-------------------+---------------------------------------------------------+
| Github           | Luckyhak          | https://github.com/Luckyhak                             |
+------------------+-------------------+---------------------------------------------------------+
| Facebook         | [redacted]        | https://www.facebook.com/[redacted]                     |
+------------------+-------------------+---------------------------------------------------------+
```

Additionally, checking for telephone number leakage is simple with OSRFramework. Using the `phonefy.py –n number` command, OSRFramework will begin its search. The following screenshot displays the URL location and platform for a given telephone number:



```
Sheet Name: Profiles recovered (2019-4-15_13h53m).
+---------------------------------------------------+------------------+
|                    i3visio_uri                    | i3visio_platform |
+===================================================+==================+
| http://www.xtelefono.es/search/?q=1-868-866-[redacted] | Xtelefonos       |
+---------------------------------------------------+------------------+
```

In our final example, we can search for usernames. Using the `usufy.py –n string` command will allow OSRFramework to search for various online resources. In this example, I have searched for `p@55w0rd1` as the username, and the following are the results:



```
Sheet Name: Profiles recovered (2019-4-15_14h0m).
+---------------------------------------------------------------+---------------+------------------+
|                          i3visio_uri                          | i3visio_alias | i3visio_platform |
+===============================================================+===============+==================+
| http://freemusicarchive.org/member/p%4055w0rd1                | p@55w0rd1     | Freemusicarchive |
+---------------------------------------------------------------+---------------+------------------+
| http://es.scribd.com/p%4055w0rd1                              | p@55w0rd1     | Scribd           |
+---------------------------------------------------------------+---------------+------------------+
| http://bbs.map.qq.com/space-username-p%4055w0rd1.html         | p@55w0rd1     | QQ               |
+---------------------------------------------------------------+---------------+------------------+
| http://www.fotolog.com/p%4055w0rd1                            | p@55w0rd1     | Fotolog          |
+---------------------------------------------------------------+---------------+------------------+
```

As you have seen, OSRFramework is another very powerful tool within the Kali Linux platform. Using a tool such as this can save you a lot of time during your information-gathering process.

Having completed this section, you now have the skills to use multiple OSINT tools to gather specific and detailed information about a target organization. In the following section, we will discuss the topic of data leaks in cloud resources.

Identifying target technology and security controls

As a penetration tester, it's quite important to determine the technologies used by a target network or organization prior to performing an external network penetration test. Discovering technologies used by a target usually proves to be very useful prior to any sort of offensive attack on the target network or organization. It allows us, as penetration testers, to better equip ourselves with the appropriate tools to get the job done efficiently and successfully. Imagine starting a new job as a carpenter but you arrive on your first day without any tools—how can you expect to be successful?

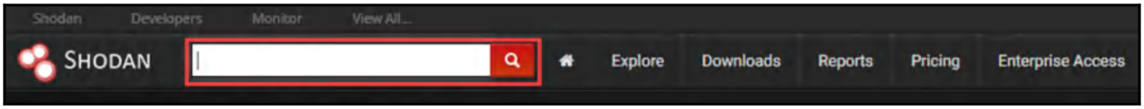
Furthermore, knowing about the technologies and security controls used on the target network beforehand will allow us to better prepare ourselves by researching and developing exploits to take advantage of known security weaknesses on the target system and network. We will look at this idea in more detail in the following sections.

Discovering technologies using Shodan

We are first going to use Shodan to help us discover technologies running on the target servers. Remember that Shodan is a search engine for IoT devices that provides in-depth information about devices connected to the internet.

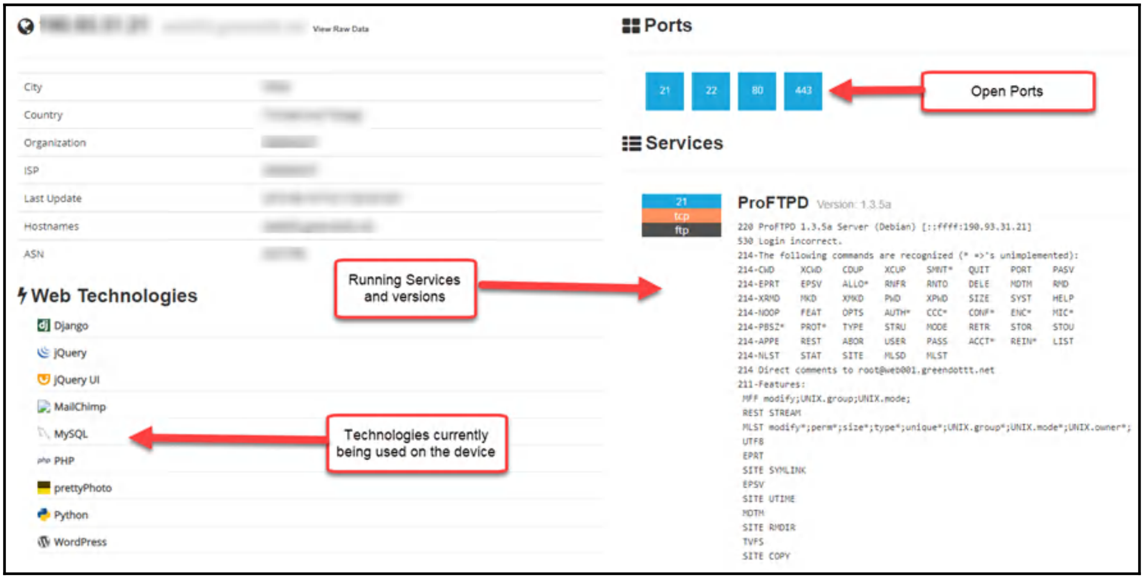
To get started, observe the following steps:

1. Using your web browser, go to `https://www.shodan.io`.
2. You may be required to register and create an account with Shodan to get better results.
3. In the search bar, enter an organization to search for a device. The following screenshot shows the search bar on Shodan:



4. Once the search is complete, click on a target from the search results to access the information found with Shodan.

5. On the target's page, you'll be presented with a list of open network ports, running services, and their versions, as well as any technologies being used:



[redacted] [redacted] View Raw Data

City	[redacted]
Country	[redacted]
Organization	[redacted]
ISP	[redacted]
Last Update	[redacted]
Hostnames	[redacted]
ASN	[redacted]

Web Technologies

- Django
- jQuery
- jQuery UI
- MailChimp
- MySQL
- PHP
- prettyPhoto
- Python
- WordPress

Ports

21 22 80 443

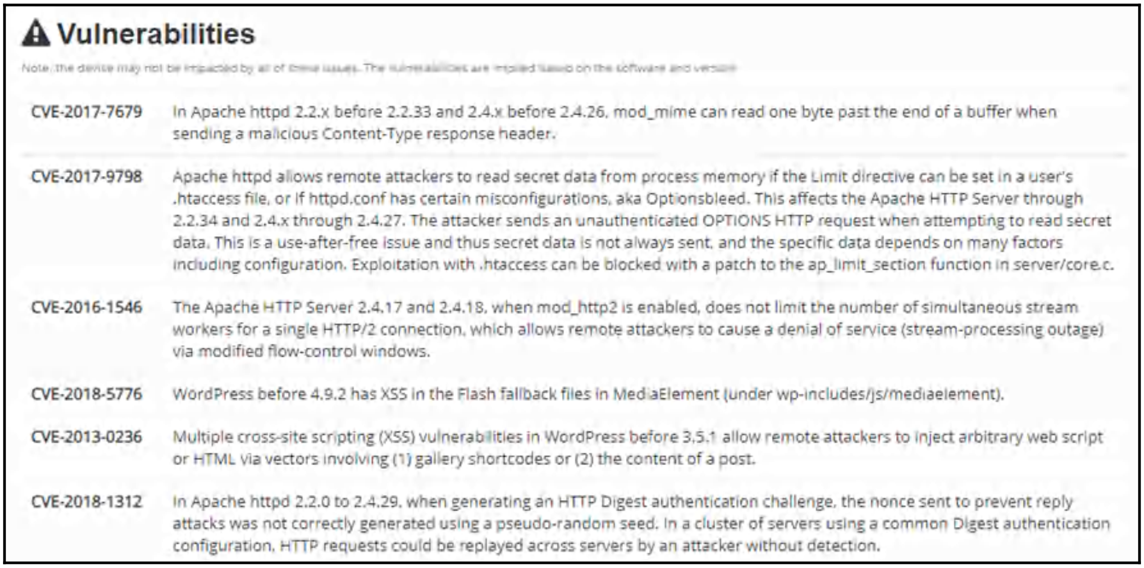
Services

Open Ports
Running Services and versions
Technologies currently being used on the device
21
tcp
ftp

ProFTPD Version: 1.3.5a

```
220 ProFTPD 1.3.5a Server (Debian) [::ffff:190.93.31.21]
530 Login incorrect.
214-The following commands are recognized (* =>'s unimplemented):
214-CWD     XCWD    CDUP    XCUP    SMNT*   QUIT    PORT    PASV
214-EPRT    EPSV    ALLO*   RNFR    RNTO    DELE    MDTM    RMD
214-XRMD    MKD     XMKD    PWD     XPWD    SIZE    SYST    HELP
214-NOOP    FEAT    OPTS    AUTH*   CCC*    CONF*   ENC*    MIC*
214-PBSZ*   PROT*   TYPE    STRU    MODE    RETR    STOR    STOU
214-APPE    REST    ABOR    USER    PASS    ACCT*   REIN*   LIST
214-NLST    STAT    SITE    MLSD    MLST
214 Direct comments to root@web001.greendottt.net
211-Features:
 MFF modify;UNIX.group;UNIX.mode;
 REST STREAM
 MLST modify*;perm*;size*;type*;unique*;UNIX.group*;UNIX.mode*;UNIX.owner*;
 UTF8
 EPRT
 SITE SYMLINK
 EPSV
 SITE UTIME
 MDTM
 SITE RMDIR
 TVFS
 SITE COPY
```

6. Scrolling down a bit, if there are any known vulnerabilities found on the target, Shodan will provide a list with descriptions:



Vulnerabilities

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

CVE-2017-7679	In Apache httpd 2.2.x before 2.2.33 and 2.4.x before 2.4.26, mod_mime can read one byte past the end of a buffer when sending a malicious Content-Type response header.
CVE-2017-9798	Apache httpd allows remote attackers to read secret data from process memory if the Limit directive can be set in a user's .htaccess file, or if httpd.conf has certain misconfigurations, aka Optionsbleed. This affects the Apache HTTP Server through 2.2.34 and 2.4.x through 2.4.27. The attacker sends an unauthenticated OPTIONS HTTP request when attempting to read secret data. This is a use-after-free issue and thus secret data is not always sent, and the specific data depends on many factors including configuration. Exploitation with .htaccess can be blocked with a patch to the ap_limit_section function in server/core.c.
CVE-2016-1546	The Apache HTTP Server 2.4.17 and 2.4.18, when mod_http2 is enabled, does not limit the number of simultaneous stream workers for a single HTTP/2 connection, which allows remote attackers to cause a denial of service (stream-processing outage) via modified flow-control windows.
CVE-2018-5776	WordPress before 4.9.2 has XSS in the Flash fallback files in MediaElement (under wp-includes/js/mediaelement).
CVE-2013-0236	Multiple cross-site scripting (XSS) vulnerabilities in WordPress before 3.5.1 allow remote attackers to inject arbitrary web script or HTML via vectors involving (1) gallery shortcodes or (2) the content of a post.
CVE-2018-1312	In Apache httpd 2.2.0 to 2.4.29, when generating an HTTP Digest authentication challenge, the nonce sent to prevent reply attacks was not correctly generated using a pseudo-random seed. In a cluster of servers using a common Digest authentication configuration, HTTP requests could be replayed across servers by an attacker without detection.

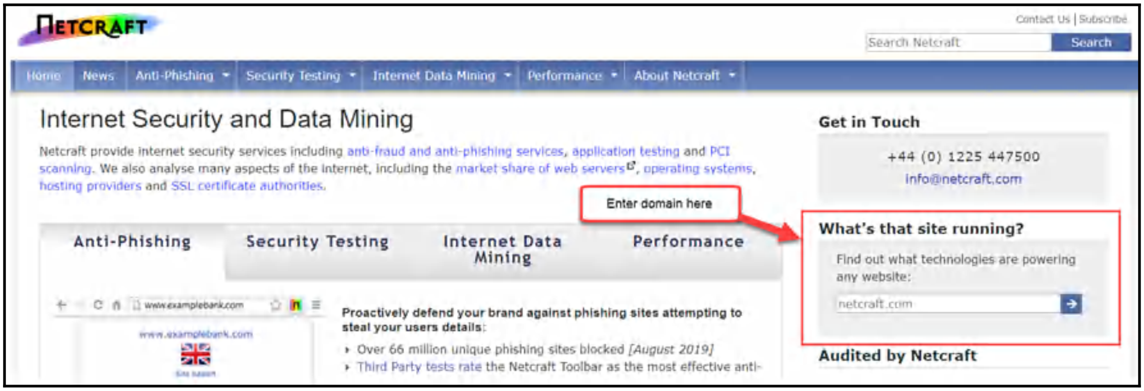
In the next section, we will take a look at using **Netcraft** to gather more detailed information about the underlying technologies of a target web server.

The power of Netcraft

Netcraft allows us to gather information about a target domain, such as network block information, registrar information, email contacts, the operating system of the hosting server, and the web platform.

To get started, use the following instructions:

1. Using your web browser, go to `https://www.netcraft.com/`.
2. In the search bar highlighted in the following screenshot, enter a domain:



Internet Security and Data Mining

Netcraft provide internet security services including anti-fraud and anti-phishing services, application testing and PCI scanning. We also analyse many aspects of the internet, including the market share of web servers, operating systems, hosting providers and SSL certificate authorities.

Proactively defend your brand against phishing sites attempting to steal your users details:

- Over 66 million unique phishing sites blocked *[August 2019]*
- Third Party tests rate the Netcraft Toolbar as the most effective anti-

Get in Touch

+44 (0) 1225 447500
info@netcraft.com

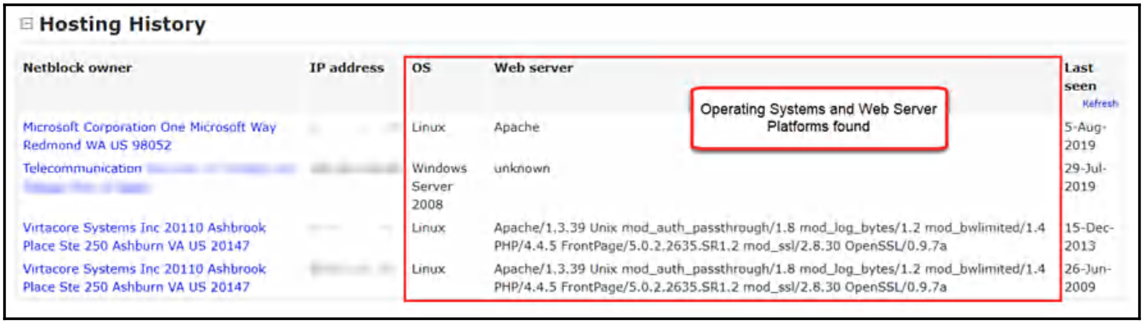
What's that site running?

Find out what technologies are powering any website:

Audited by Netcraft

If a search doesn't return anything, include `www` as part of the website address.

3. The results page will appear, providing network-related information about the target. Scroll down a bit until you see **Hosting History**. This section provides you with the hosting server's operating system, web server platform, and service versions, as shown here:

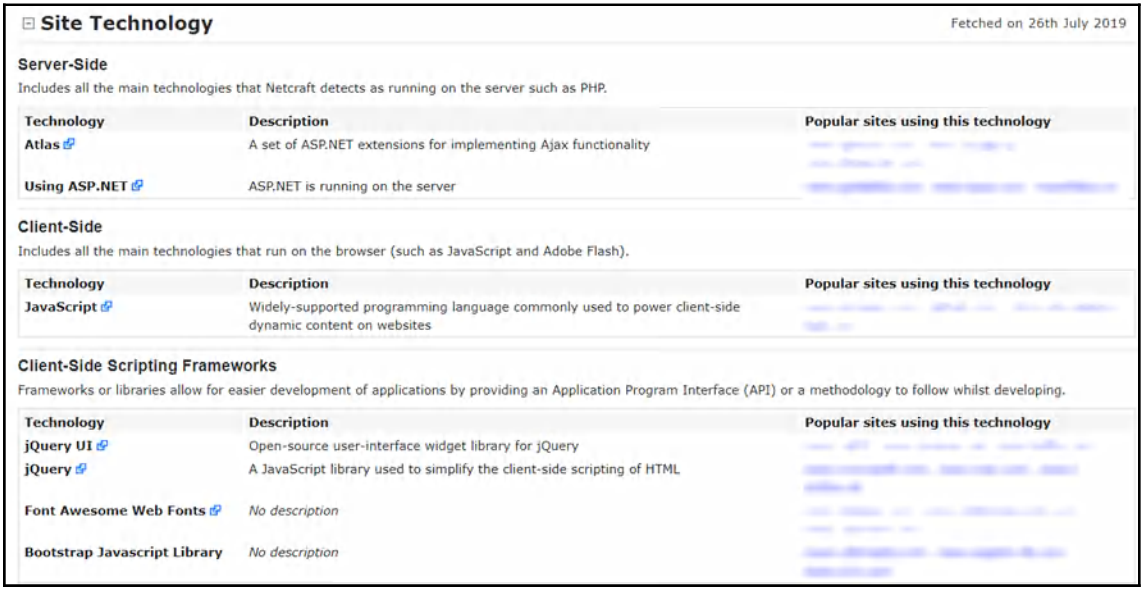


⊟ **Hosting History**

Netblock owner	IP address	OS	Web server	Last seen Refresh
Microsoft Corporation One Microsoft Way Redmond WA US 98052	[redacted]	Linux	Apache	5-Aug-2019
Telecommunication [redacted]	[redacted]	Windows Server 2008	unknown	29-Jul-2019
Virtacore Systems Inc 20110 Ashbrook Place Ste 250 Ashburn VA US 20147	[redacted]	Linux	Apache/1.3.39 Unix mod_auth_passthrough/1.8 mod_log_bytes/1.2 mod_bwlimited/1.4 PHP/4.4.5 FrontPage/5.0.2.2635.SR1.2 mod_ssl/2.8.30 OpenSSL/0.9.7a	15-Dec-2013
Virtacore Systems Inc 20110 Ashbrook Place Ste 250 Ashburn VA US 20147	[redacted]	Linux	Apache/1.3.39 Unix mod_auth_passthrough/1.8 mod_log_bytes/1.2 mod_bwlimited/1.4 PHP/4.4.5 FrontPage/5.0.2.2635.SR1.2 mod_ssl/2.8.30 OpenSSL/0.9.7a	26-Jun-2009

Operating Systems and Web Server Platforms found

4. Scroll down a bit more until you see **Site Technology**. This section informs you about the technologies Netcraft was able to identify on the target web server:



⊟ **Site Technology**

Fetched on 26th July 2019

Server-Side

Includes all the main technologies that Netcraft detects as running on the server such as PHP.

Technology	Description	Popular sites using this technology
Atlas	A set of ASP.NET extensions for implementing Ajax functionality	[redacted]
Using ASP.NET	ASP.NET is running on the server	[redacted]

Client-Side

Includes all the main technologies that run on the browser (such as JavaScript and Adobe Flash).

Technology	Description	Popular sites using this technology
JavaScript	Widely-supported programming language commonly used to power client-side dynamic content on websites	[redacted]

Client-Side Scripting Frameworks

Frameworks or libraries allow for easier development of applications by providing an Application Program Interface (API) or a methodology to follow whilst developing.

Technology	Description	Popular sites using this technology
jQuery UI	Open-source user-interface widget library for jQuery	[redacted]
jQuery	A JavaScript library used to simplify the client-side scripting of HTML	[redacted]
Font Awesome Web Fonts	*No description*	[redacted]
Bootstrap Javascript Library	*No description*	[redacted]

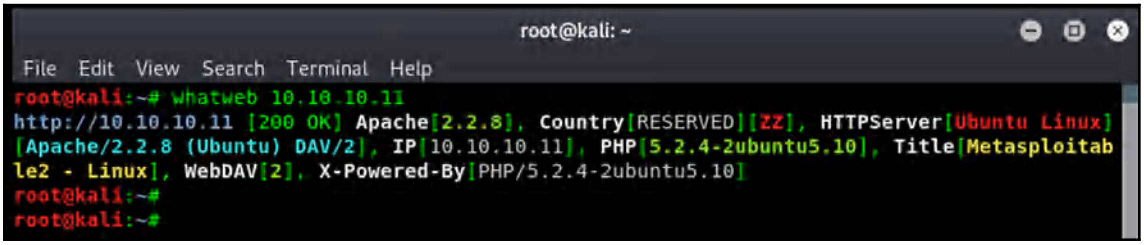
As you have seen, Netcraft is able to provide us with very useful information. We can now take the information found and perform further research on the technologies to find known vulnerabilities and exploits to compromise the target.

Recognizing technologies with WhatWeb

Lastly, we can take a look at using the **WhatWeb** tool in Kali Linux. WhatWeb is capable of recognizing the technologies used for websites, email addresses, web frameworks, and databases.

To get started using WhatWeb, observe the following steps:

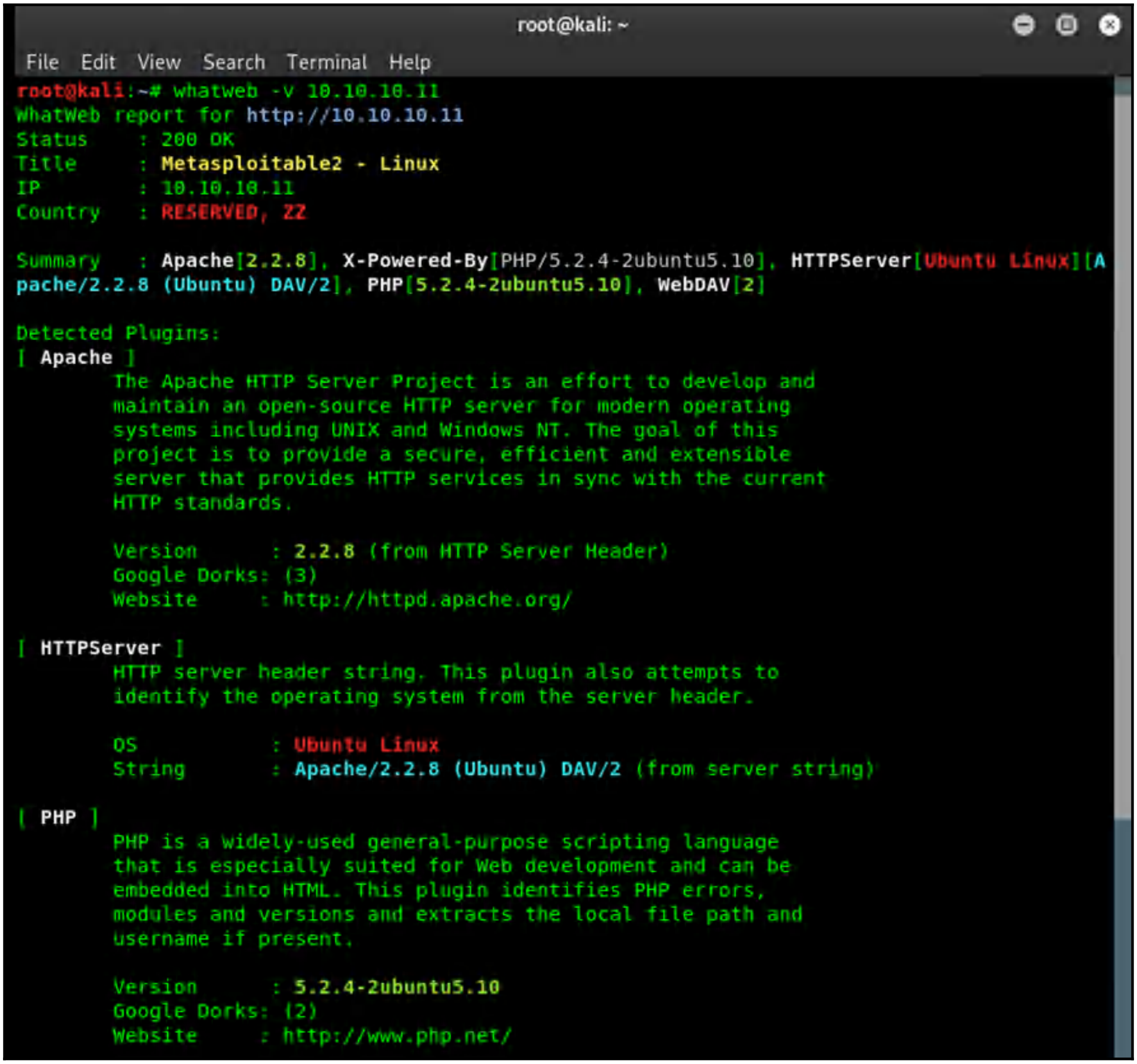
1. Open a new Terminal and enter the `whatsweb –h` command to get the help menu, displaying the syntax.
2. To run WhatWeb against a target such as Metasploitable in our lab, use the `whatweb <target>` command as shown here:



```
root@kali:~# whatweb 10.10.10.11
http://10.10.10.11 [200 OK] Apache[2.2.8], Country[RESERVED][ZZ], HTTPServer[Ubuntu Linux]
[Apache/2.2.8 (Ubuntu) DAV/2], IP[10.10.10.11], PHP[5.2.4-2ubuntu5.10], Title[Metasploitab
le2 - Linux], WebDAV[2], X-Powered-By[PHP/5.2.4-2ubuntu5.10]
root@kali:~#
root@kali:~#
```

WhatWeb was able to provide us with the technologies used on the web server platform on our Metasploitable virtual machine.

3. Next, using the `whatweb –v <target>` command, provide a verbose output as shown here:



```
root@kali:~# whatweb -v 10.10.10.11
WhatWeb report for http://10.10.10.11
Status    : 200 OK
Title     : Metasploitable2 - Linux
IP        : 10.10.10.11
Country   : RESERVED, ZZ

Summary   : Apache[2.2.8], X-Powered-By[PHP/5.2.4-2ubuntu5.10], HTTPServer[Ubuntu Linux][A
pache/2.2.8 (Ubuntu) DAV/2], PHP[5.2.4-2ubuntu5.10], WebDAV[2]

Detected Plugins:
[ Apache ]
        The Apache HTTP Server Project is an effort to develop and
        maintain an open-source HTTP server for modern operating
        systems including UNIX and Windows NT. The goal of this
        project is to provide a secure, efficient and extensible
        server that provides HTTP services in sync with the current
        HTTP standards.

        Version      : 2.2.8 (from HTTP Server Header)
        Google Dorks: (3)
        Website      : http://httpd.apache.org/

[ HTTPServer ]
        HTTP server header string. This plugin also attempts to
        identify the operating system from the server header.

        OS           : Ubuntu Linux
        String       : Apache/2.2.8 (Ubuntu) DAV/2 (from server string)

[ PHP ]
        PHP is a widely-used general-purpose scripting language
        that is especially suited for Web development and can be
        embedded into HTML. This plugin identifies PHP errors,
        modules and versions and extracts the local file path and
        username if present.

        Version      : 5.2.4-2ubuntu5.10
        Google Dorks: (2)
        Website      : http://www.php.net/
```

In the verbose output shown here, WhatWeb provides us with much more detail about the descriptions of plugins used and their results.

In this section, we have completed various exercises designed to help us discover technologies that are being used on a target network or system. In the next section, we will take a deep dive into learning about cloud resources.

Finding data leaks in cloud resources

Over the past few years, cloud computing has become one of the fastest-growing trends in the IT industry. Cloud computing allows companies to migrate and utilize computing resources within a cloud provider's data center. Cloud computing providers have a pay-as-you-go model, which means that you only pay for the resources you use. Some cloud providers allow pay-per-minute schemes, while others use a pay-per-hour structure.

There are three big cloud providers:

- **Amazon Web Services** (**AWS**): Amazon's cloud service
- **Microsoft Azure**: Microsoft's cloud service
- **Google Cloud Platform** (**GCP**): Google's cloud service

A service that cloud providers usually offer to customers is a storage facility. The AWS storage facility is known as **Simple Storage Service** (**S3**). Whenever a customer enables the S3 service, a bucket is created. A bucket is a storage unit within the AWS platform where the customer can add or remove files. In Microsoft Azure, the file storage facility is known as **Azure Files**. Additionally, on GCP, the storage facility is known as **Google Cloud Storage**.

In the field of information security, we must remember that when a company is using a cloud platform, the data on the cloud platform must be secured, just like it should be when stored on premises (that is, when stored locally). Sometimes, administrators forget to enable security configurations or lack knowledge regarding the security of a cloud solution. This could lead to, say, an attacker discovering a target organization's AWS S3 buckets. Let's look at a simple example of doing that now.

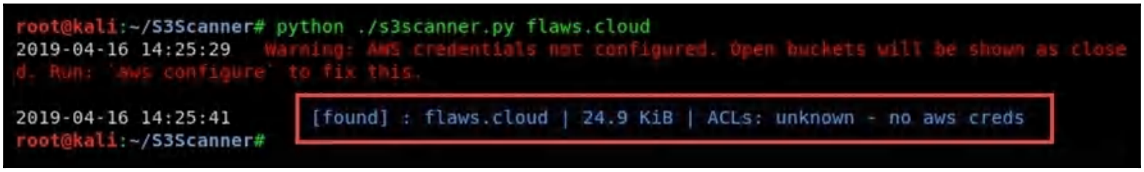
`http://flaws.cloud/` is a website that you can use to learn about cloud security vulnerabilities.

In our exercise, we are going to use the S3Scanner tool. Follow these steps to get started:

1. This tool is not pre-installed in Kali Linux, so we will need to create a clone of the GitHub repository by using the following command:

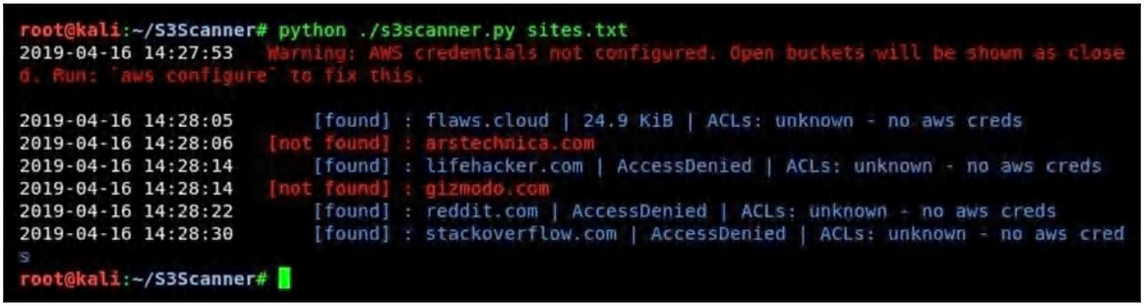
```
git clone https://github.com/sa7mon/S3Scanner.git
```

2. Next, change directories to the `S3Scanner` folder using the `cd S3Scanner` command.
3. Now, you'll need to install additional dependencies for this tool. Don't worry—the developers made this step very easy for us. To complete this step, use the `pip install -r requirements.txt` command.
4. Once completed, we can now use our tool on a target's domain. Using the `python ./s3scanner.py domain` syntax, the tool scanner will create a domain for an AWS S3 bucket and determine whether it's accessible.
5. The following screenshot shows the use of S3Scanner to check for any AWS S3 buckets on the `flaws.cloud` domain:



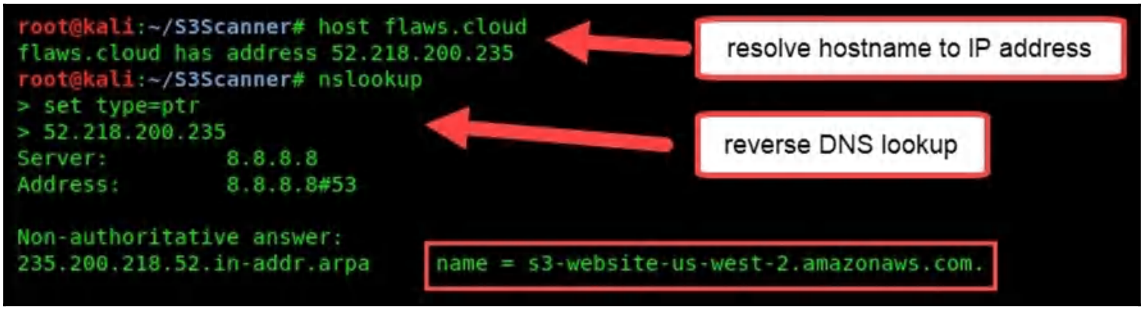
```
root@kali:~/S3Scanner# python ./s3scanner.py flaws.cloud
2019-04-16 14:25:29   Warning: AWS credentials not configured. Open buckets will be shown as close
d. Run: `aws configure` to fix this.

2019-04-16 14:25:41     [found] : flaws.cloud | 24.9 KiB | ACLs: unknown - no aws creds
root@kali:~/S3Scanner#
```

6. One bucket has been found on the domain. Additionally, you can create a list of domains in a text file and query the entire file at once. The following is an example of querying multiple domains that are stored in the `sites.text` file:



```
root@kali:~/S3Scanner# python ./s3scanner.py sites.txt
2019-04-16 14:27:53   Warning: AWS credentials not configured. Open buckets will be shown as close
d. Run: `aws configure` to fix this.

2019-04-16 14:28:05       [found] : flaws.cloud | 24.9 KiB | ACLs: unknown - no aws creds
2019-04-16 14:28:06   [not found] : arstechnica.com
2019-04-16 14:28:14       [found] : lifehacker.com | AccessDenied | ACLs: unknown - no aws creds
2019-04-16 14:28:14   [not found] : gizmodo.com
2019-04-16 14:28:22       [found] : reddit.com | AccessDenied | ACLs: unknown - no aws creds
2019-04-16 14:28:30       [found] : stackoverflow.com | AccessDenied | ACLs: unknown - no aws cred
s
root@kali:~/S3Scanner# 
```

7. Furthermore, we can use the `host` command to resolve the IP address of the domain. Then, by using the `nslookup` utility with the `ptr` parameter, a reverse lookup can be performed, which will result in us getting the actual name of the AWS S3 bucket, as shown in the following screenshot:



```
root@kali:~/S3Scanner# host flaws.cloud
flaws.cloud has address 52.218.200.235
root@kali:~/S3Scanner# nslookup
> set type=ptr
> 52.218.200.235
Server:         8.8.8.8
Address:        8.8.8.8#53

Non-authoritative answer:
235.200.218.52.in-addr.arpa     name = s3-website-us-west-2.amazonaws.com.
```

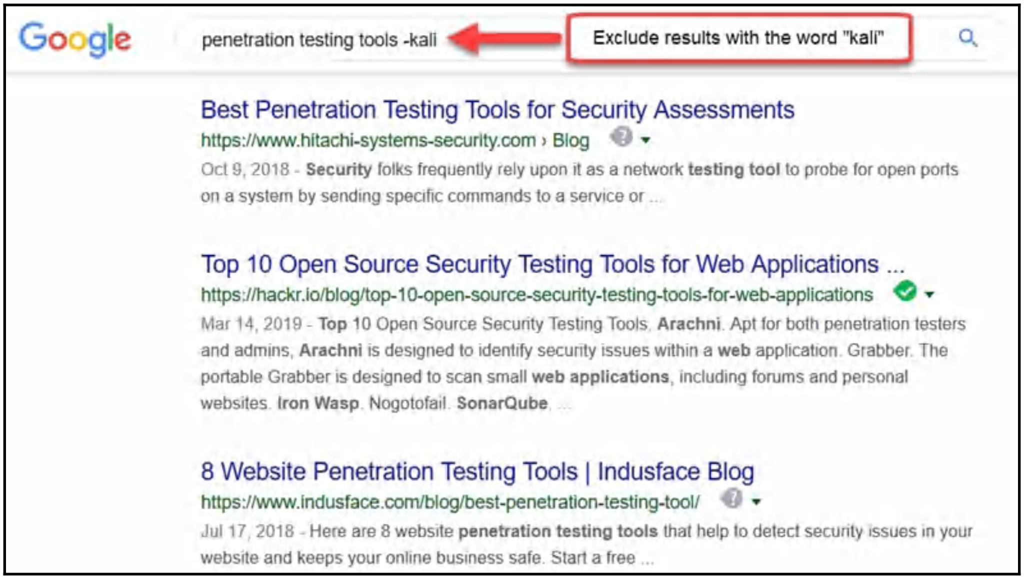
Data leaks can happen on any platform and to any organization. As an upcoming penetration tester and cybersecurity professional, you must have knowledge of how to find them before an actual hacker does and exploits them. Companies can store sensitive data on cloud platforms, or even leave other data completely unprotected on a cloud service provider network. This can lead to the successful retrieval of data and accounts.

Understanding Google hacking and search operators

The concept of Google hacking is not actually hacking into Google's network infrastructure or systems, but rather using advanced search parameters within the Google search engine. We can use Google to help us find vulnerable systems, hidden information, and resources on the internet by simply inserting special search operators in the Google search bar.

Let's imagine that you would like to use the Google search engine to look for various websites, but you don't want to see results that contain certain keywords or phrases. We can use the `<string of text here> -<keyword>` syntax to do this. The keyword is the phrase or text that you want to exclude.

Let's look at the following example:



penetration testing tools -kali

Exclude results with the word "kali"

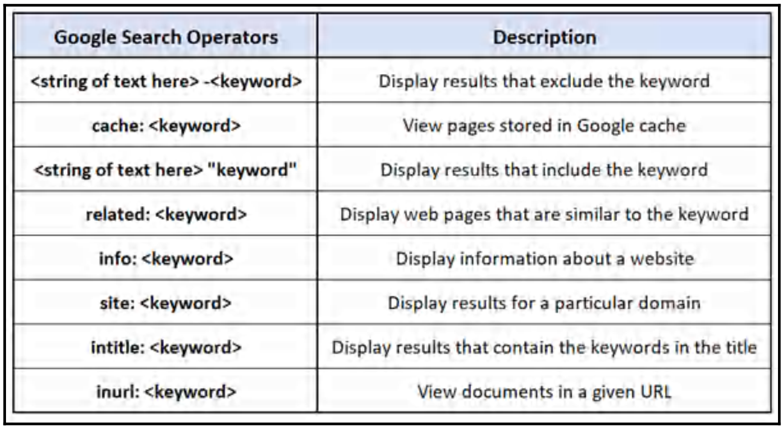
Best Penetration Testing Tools for Security Assessments
https://www.hitachi-systems-security.com › Blog
Oct 9, 2018 - **Security** folks frequently rely upon it as a network **testing tool** to probe for open ports on a system by sending specific commands to a service or ...

Top 10 Open Source Security Testing Tools for Web Applications ...
https://hackr.io/blog/top-10-open-source-security-testing-tools-for-web-applications
Mar 14, 2019 - **Top** 10 Open Source Security Testing Tools. **Arachni**. Apt for both penetration testers and admins, **Arachni** is designed to identify security issues within a **web** application. Grabber. The portable Grabber is designed to scan small **web applications**, including forums and personal websites. **Iron Wasp**. Nogotofail. **SonarQube**. ...

8 Website Penetration Testing Tools | Indusface Blog
https://www.indusface.com/blog/best-penetration-testing-tool/
Jul 17, 2018 - Here are 8 website **penetration testing tools** that help to detect security issues in your website and keeps your online business safe. Start a free ...

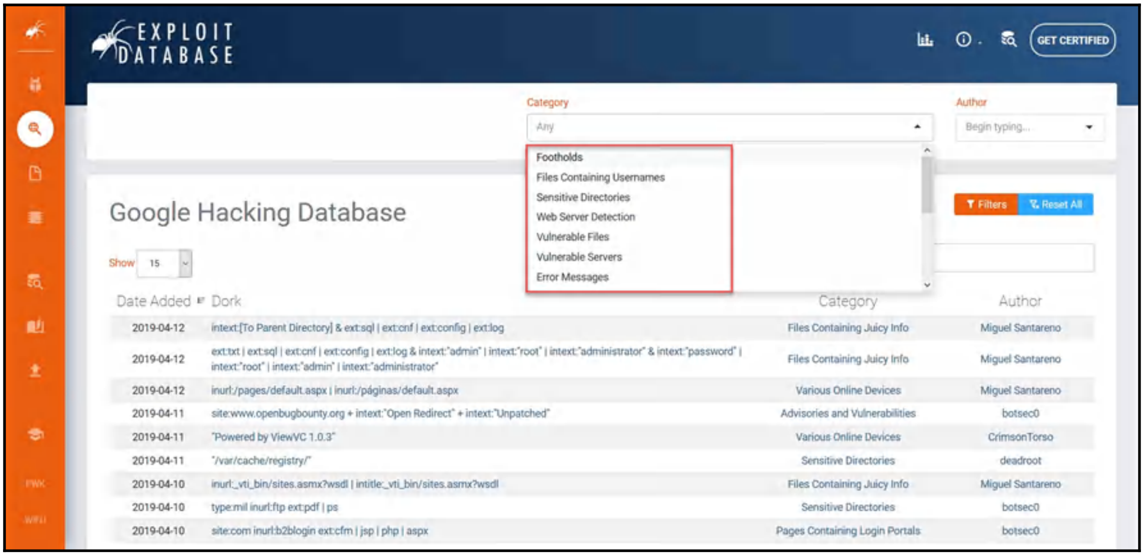
In our example, we are searching for penetration testing tools. At the same time, we are telling the Google search algorithm to not display any results that contain the word `kali`. Additionally, we can use the `<string of text here> "keyword"` syntax to view results that do contain the keyword.

The following table is a brief list of Google search operators, also known as **Google dorks**, that can help you find sensitive information on the internet:



Google Search Operators	Description
<string of text here> -<keyword>	Display results that exclude the keyword
cache: <keyword>	View pages stored in Google cache
<string of text here> "keyword"	Display results that include the keyword
related: <keyword>	Display web pages that are similar to the keyword
info: <keyword>	Display information about a website
site: <keyword>	Display results for a particular domain
intitle: <keyword>	Display results that contain the keywords in the title
inurl: <keyword>	View documents in a given URL

Furthermore, the team at Offensive Security (`www.offensive-security.com`) maintains the Exploit Database (`www.exploit-db.com`), which has a dedicated section known as the **Google Hacking Database** (**GHD**) (`https://www.exploit-db.com/google-hacking-database`). The GHD is constantly updated by community members and contains search parameters in many categories, as shown in the following screenshot:

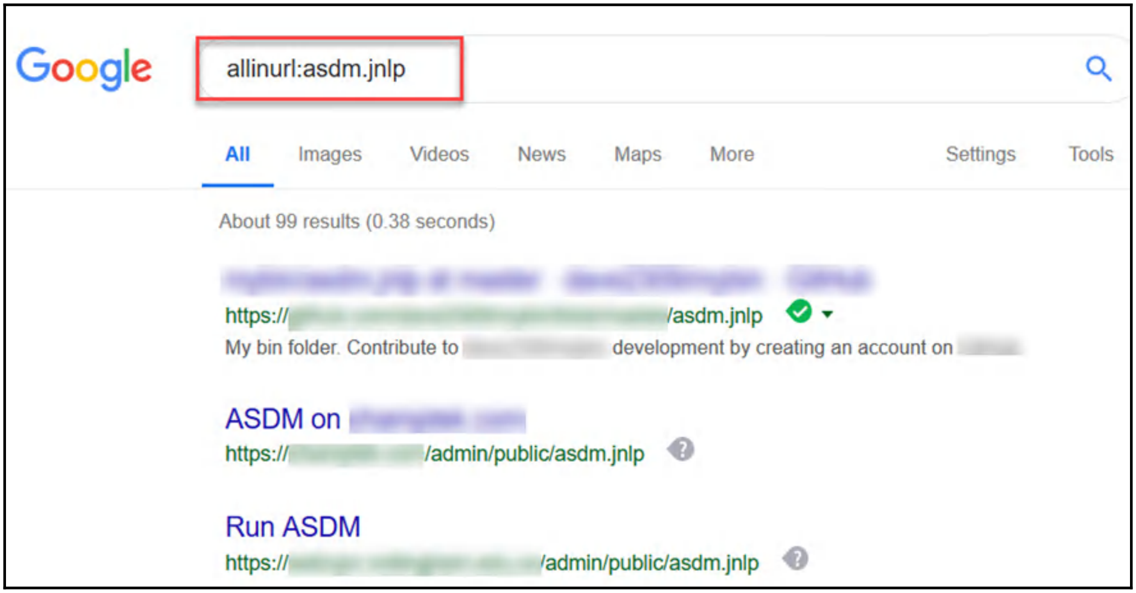


Google Hacking Database

<table>
<tr><th>Date Added</th><th>Dork</th><th>Category</th><th>Author</th></tr>
<tr><td>2019-04-12</td><td>intext:[To Parent Directory] & ext:sql | ext:cnf | ext:config | ext:log</td><td>Files Containing Juicy Info</td><td>Miguel Santareno</td></tr>
<tr><td>2019-04-12</td><td>ext:txt | ext:sql | ext:cnf | ext:config | ext:log & intext:"admin" | intext:"root" | intext:"administrator" & intext:"password" | intext:"root" | intext:"admin" | intext:"administrator"</td><td>Files Containing Juicy Info</td><td>Miguel Santareno</td></tr>
<tr><td>2019-04-12</td><td>inurl:/pages/default.aspx | inurl:/páginas/default.aspx</td><td>Various Online Devices</td><td>Miguel Santareno</td></tr>
<tr><td>2019-04-11</td><td>site:www.openbugbounty.org + intext:"Open Redirect" + intext:"Unpatched"</td><td>Advisories and Vulnerabilities</td><td>botsec0</td></tr>
<tr><td>2019-04-11</td><td>"Powered by ViewVC 1.0.3"</td><td>Various Online Devices</td><td>CrimsonTorso</td></tr>
<tr><td>2019-04-11</td><td>"/var/cache/registry/"</td><td>Sensitive Directories</td><td>deadroot</td></tr>
<tr><td>2019-04-10</td><td>inurl:_vti_bin/sites.asmx?wsdl | intitle:_vti_bin/sites.asmx?wsdl</td><td>Files Containing Juicy Info</td><td>Miguel Santareno</td></tr>
<tr><td>2019-04-10</td><td>type:mil inurl:ftp ext:pdf | ps</td><td>Sensitive Directories</td><td>botsec0</td></tr>
<tr><td>2019-04-10</td><td>site:com inurl:b2blogin ext:cfm | jsp | php | aspx</td><td>Pages Containing Login Portals</td><td>botsec0</td></tr>
</table>

Each search parameter can be copied and pasted into Google Search, and the results will be displayed accordingly. Each entry within the GHD contains a brief description of the search operator.

The following is a search parameter that's used to discover the Cisco **Adaptive Security Appliance** (**ASA**), which has a publicly available login page:



allinurl:asdm.jnlp

About 99 results (0.38 seconds)

[redacted]
https://[redacted]/asdm.jnlp
My bin folder. Contribute to [redacted] development by creating an account on [redacted]

ASDM on [redacted]
https://[redacted]/admin/public/asdm.jnlp

Run ASDM
https://[redacted]/admin/public/asdm.jnlp

Such sensitive information and hidden directories can be a hacker's playground; similarly, for a penetration tester, it's a gold mine just waiting to be exploited.

We have completed our discussion of Google hacking. In the next section, we'll take a took at copying websites locally using Kali Linux.

Leveraging whois and copying websites with HTTrack

In this section, we are going to take a look at two particular resources. We'll use whois to help us gather contact information from a domain registrar for a target domain, and we'll use **HTTrack** to copy a website locally.

Let's dive in and look at the functions of whois and how it can be beneficial to us.

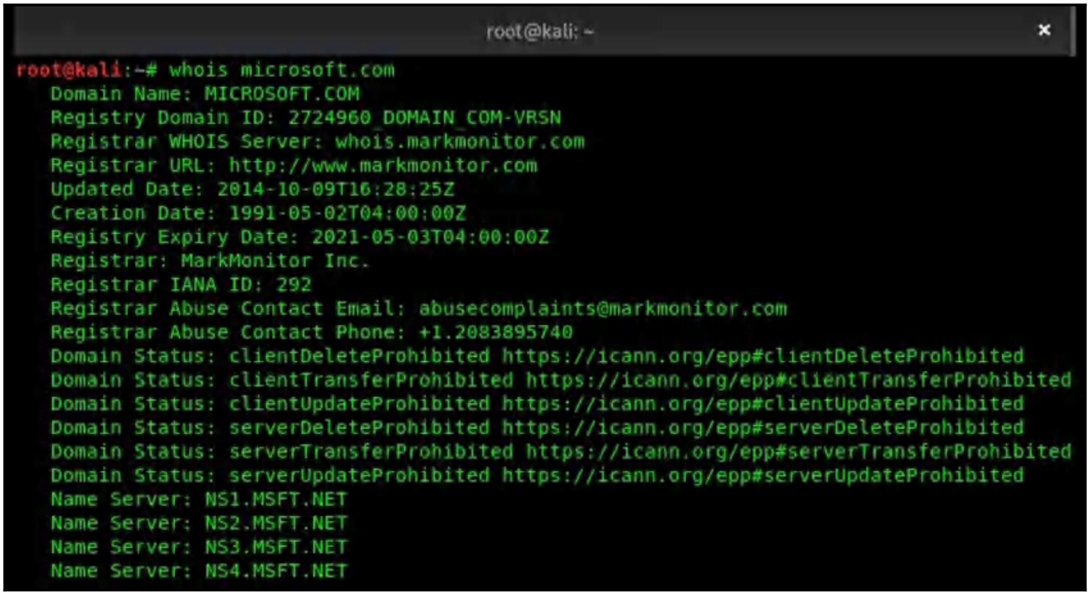
whois

whois is a database that keeps a record of the registry information of all registered domains. The following is a brief list of some information types that are usually stored for public records:

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

Accessing a whois database is quite simple: you can use the Google search engine to find various databases. Some whois websites include `www.whois.net`, `whois.domaintools.com`, `who.is`, and `www.whois.com`. However, Kali Linux contains a built-in whois tool. To perform a `whois` lookup on a domain, implement the following steps:

1. Open the Terminal and use the `whois <domain-name>` syntax, as shown in the following screenshot:



```
root@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: 2014-10-09T16:28:25Z
   Creation Date: 1991-05-02T04:00:00Z
   Registry Expiry Date: 2021-05-03T04:00:00Z
   Registrar: MarkMonitor Inc.
   Registrar IANA ID: 292
   Registrar Abuse Contact Email: abusecomplaints@markmonitor.com
   Registrar Abuse Contact Phone: +1.2083895740
   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.MSFT.NET
   Name Server: NS2.MSFT.NET
   Name Server: NS3.MSFT.NET
   Name Server: NS4.MSFT.NET
```

2. The output is presented on the Terminal for the domain. The information that's obtained can be leveraged by a penetration tester for various types of attacks on a target organization.

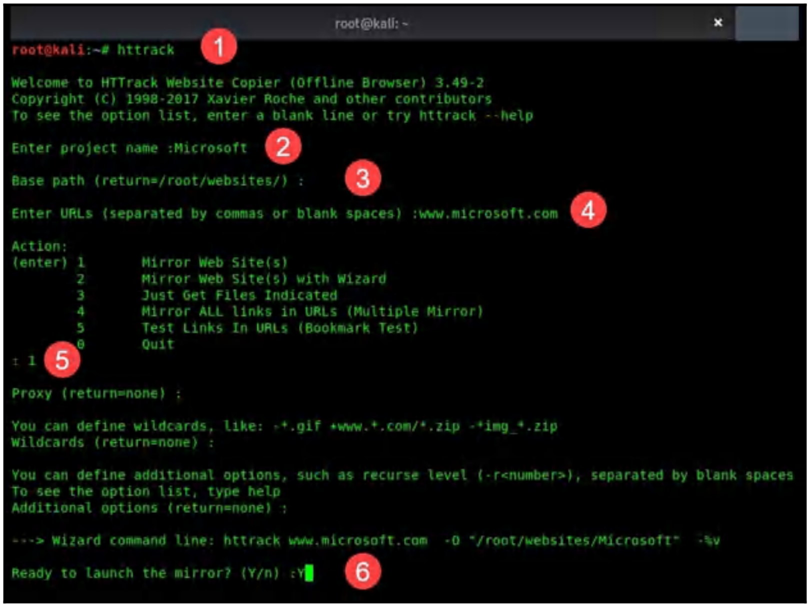
In the next section, you will learn how to use HTTrack to copy a website locally.

HTTrack

HTTrack (`www.httrack.com`) allows us to view an entire website offline. It does this by creating a clone copy of an online website and storing it locally on our computer. To use HTTrack, simply open a new Terminal window and perform the following steps:

1. Execute the `httrack` command to invoke the interactive wizard.
2. Enter a name for the project.
3. Set the destination path to store the offline copy of the target website. Hitting *Enter* will use the defaults in the brackets.
4. Specify the URL.
5. Choose an appropriate action.
6. Confirm the details and launch HTTrack to mirror the website.

The following is a screenshot indicating the steps that we have just outlined:



```
root@kali:~# httrack

Welcome to HTTrack Website Copier (Offline Browser) 3.49-2
Copyright (C) 1998-2017 Xavier Roche and other contributors
To see the option list, enter a blank line or try httrack --help

Enter project name :Microsoft

Base path (return=/root/websites/) :

Enter URLs (separated by commas or blank spaces) :www.microsoft.com

Action:
(enter) 1       Mirror Web Site(s)
        2       Mirror Web Site(s) with Wizard
        3       Just Get Files Indicated
        4       Mirror ALL links in URLs (Multiple Mirror)
        5       Test Links In URLs (Bookmark Test)
        0       Quit
: 1

Proxy (return=none) :

You can define wildcards, like: -*.gif +www.*.com/*.zip -*img_*.zip
Wildcards (return=none) :

You can define additional options, such as recurse level (-r<number>), separated by blank spaces
To see the option list, type help
Additional options (return=none) :

---> Wizard command line: httrack www.microsoft.com  -O "/root/websites/Microsoft"  -%v

Ready to launch the mirror? (Y/n) :Y
```

Cloning a website can be very useful as you'll be able to discover and access hidden resources and files that weren't accessible via the online version. As a penetration tester, you can explore each offline directory thoroughly; normally, webmasters tend not to always perform any cleanup of old data and files, so there'll be lots to explore.

Having completed this section, you have the knowledge required to copy a target website onto your Kali Linux machine. In the next section, we will attempt to retrieve subdomains for a target domain.

Finding subdomains using Sublist3r

As a user of the internet, you will have realized that multiple search engines such as MSN, Google, Yahoo, and Bing frequently learn and index new and existing websites to improve their search results. If you search for a company's website, you are most likely to discover the main domain name, such as `company.com`. A lot of organizations create subdomains for various reasons, however. As penetration testers, we would like to discover all the possible subdomains of a target organization as they can lead to login portals and sensitive corporate directories, which may contain confidential files and resources.

We can leverage the power of search engines for this task using the **Sublist3r** tool. Sublist3r is a Python-based tool that is used to enumerate (extract/obtain) the subdomains of a given website using OSINT, such as search engines and other internet indexing platforms.

The Sublist3r tool is not natively installed on Kali Linux, and so we will need to download it from its GitHub repository.

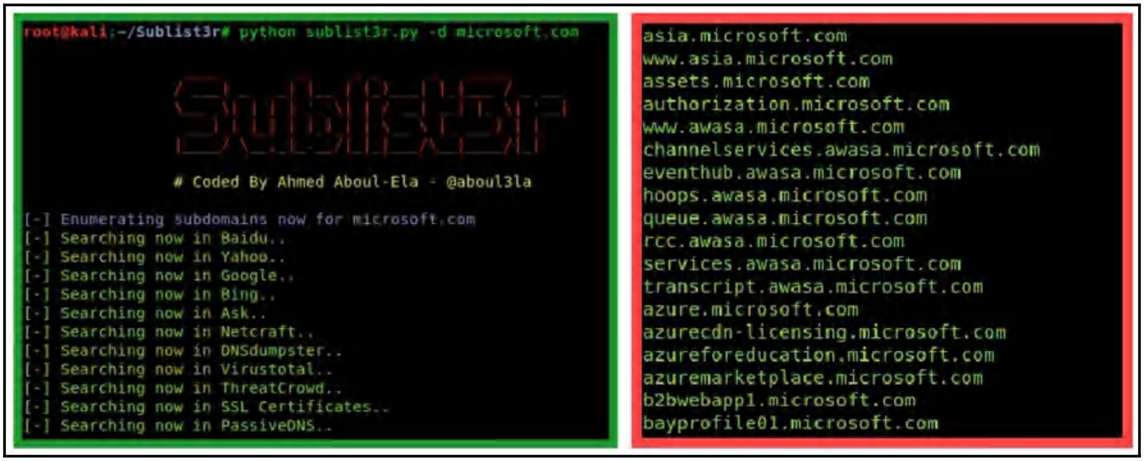
To get started, execute the following steps:

1. Open the Terminal on your Kali Linux machine and execute the following command:

   ```
   git clone https://github.com/aboul3la/Sublist3r.git
   ```

2. Once the cloning process has been completed, change directory to the `Sublist3r` folder using the `cd Sublist3r` command.

3. At this point, we can use the Sublist3r tool to search for the subdomains of a target domain (company) using the `python sublist3r.py –d domain-name` command. The screenshot to the left shows the successful invocation of the tool, while the right-hand screenshot shows the results being populated on the Terminal:



```
root@kali:~/Sublist3r# python sublist3r.py -d microsoft.com

                # Coded By Ahmed Aboul-Ela - @aboul3la

[-] Enumerating subdomains now for microsoft.com
[-] Searching now in Baidu..
[-] Searching now in Yahoo..
[-] Searching now in Google..
[-] Searching now in Bing..
[-] Searching now in Ask..
[-] Searching now in Netcraft..
[-] Searching now in DNSdumpster..
[-] Searching now in Virustotal..
[-] Searching now in ThreatCrowd..
[-] Searching now in SSL Certificates..
[-] Searching now in PassiveDNS..
```

```
asia.microsoft.com
www.asia.microsoft.com
assets.microsoft.com
authorization.microsoft.com
www.awasa.microsoft.com
channelservices.awasa.microsoft.com
eventhub.awasa.microsoft.com
hoops.awasa.microsoft.com
queue.awasa.microsoft.com
rcc.awasa.microsoft.com
services.awasa.microsoft.com
transcript.awasa.microsoft.com
azure.microsoft.com
azurecdn-licensing.microsoft.com
azureforeducation.microsoft.com
azuremarketplace.microsoft.com
b2bwebapp1.microsoft.com
bayprofile01.microsoft.com
```

Using this tool can save us a lot of time that would otherwise have been spent manually searching the internet.

You have now learned how to efficiently discover subdomains for a target website using the Sublist3r tool on Kali Linux.

Summary

In this chapter, we have discussed the importance of conducting reconnaissance prior to attacking a target. The information gathered during this phase is vital to the later phases of penetration testing. Information gathering helps create the foundation for the penetration tester's research of a target's security vulnerabilities, which is required for system and network exploitation.

We outlined the differences between reconnaissance and footprinting, and we also took a look at how to use various OSINT tools to obtain information about various targets.

In the next chapter, *Active Information Gathering*, we will be covering further topics on information gathering.

Questions

1. What is the purpose of footprinting?
2. What OSINT tools can be used to gather information?
3. Using Google Search, how can you view results for a particular domain?
4. Name a reputable online source for researching exploits.
5. You are interested in gathering the domain registry information of a target company. What resources would you use?
6. How can you discover the subdomains of a company's website?

Further reading

- **Web Application Information Gathering**: `https://hub.packtpub.com/web-application-information-gathering/`
- **Open Source Intelligence**: `https://hub.packtpub.com/open-source-intelligence/`