Module 02

Footprinting and Reconnaissance

EC-Council
Official Curricula

EC-Council C|EH v13 Certified Ethical Hacker

This page is intentionally left blank.

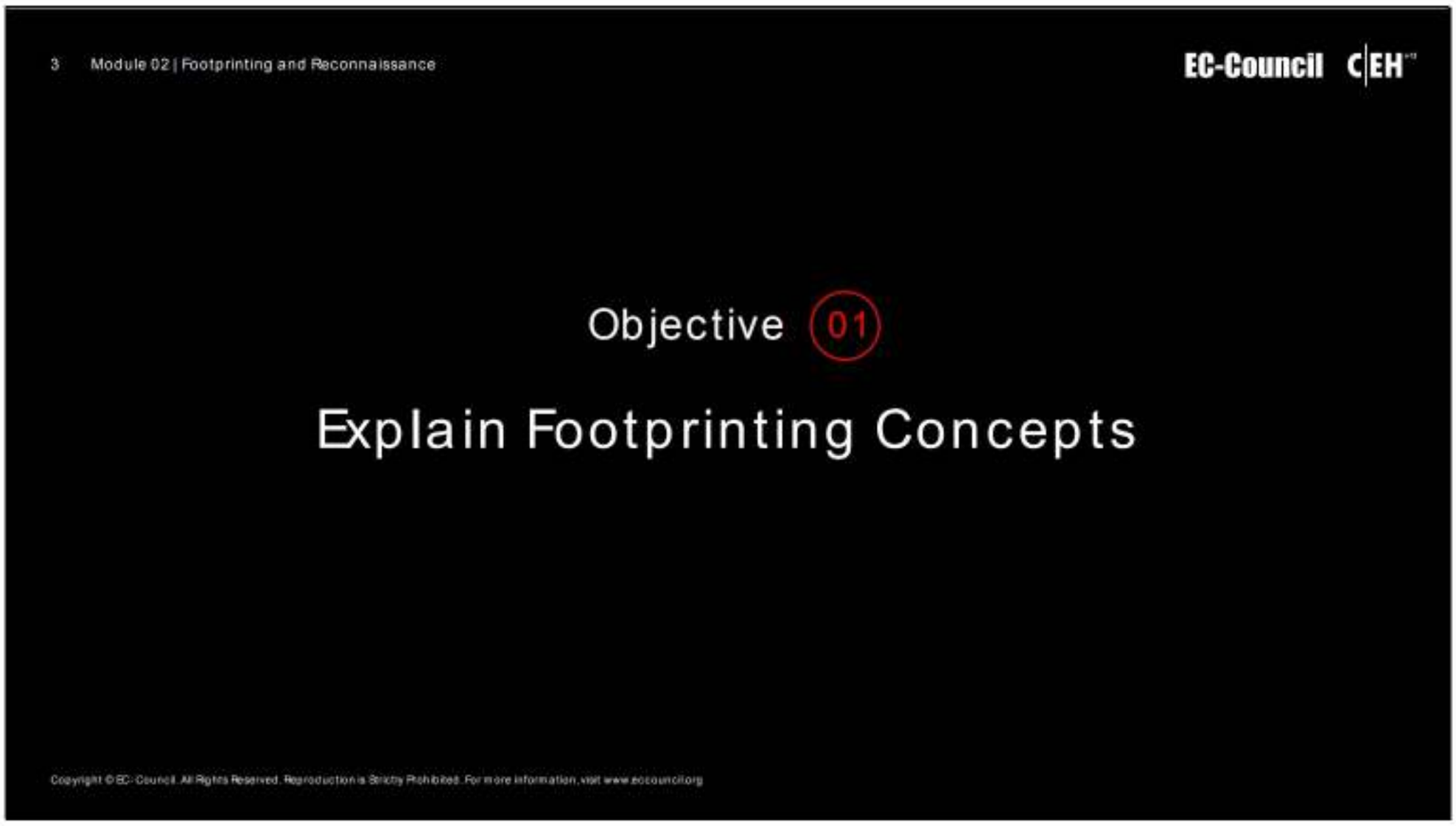
Learning Objectives

Footprinting is the first step in the evaluation of the security posture of the IT infrastructure of a target organization. Through footprinting and reconnaissance, one can gather maximum information about a computer system or a network and about any device connected to that network. In other words, footprinting provides a security profile blueprint for an organization and should be undertaken in a methodological manner.

This module starts with an introduction to footprinting concepts and provides insights into the footprinting methodology. The module ends with an overview of footprinting tools and countermeasures.

At the end of this module, you will be able to:

- Describe footprinting concepts
- Perform footprinting through search engines and using advanced Google hacking techniques
- Perform footprinting through Internet research services and social networking sites
- Perform Whois, DNS, network, and email footprinting
- Perform footprinting through social engineering
- Use different footprinting tools
- Apply footprinting best practices



Footprinting Concepts

This step acts as a preparatory phase for the attacker, who needs to gather as much information as possible to easily find ways to intrude into the target network.

This section aims to familiarize you with footprinting, why it is necessary, and its objectives.

Reconnaissance

Reconnaissance (also known as footprinting) refers to the preparatory phase where an attacker seeks to **gather as much information as possible about a target of evaluation** prior to launching an attack

Types of Reconnaissance

Passive

Gathering information about the target **without direct interaction**

It involves:

- Open-source Intelligence (OSINT) gathering
- Proprietary databases and paid services
- Sharing intelligence with partner organizations or industry groups

Active

Gathering information about the target **with direct interaction**

It involves:

- DNS interrogation
- Social engineering
- Network/port scanning
- User and service enumeration

Reconnaissance

Reconnaissance (also known as footprinting) refers to the preparatory phase where an attacker seeks to gather as much information as possible about a target of evaluation prior to launching an attack. An essential aspect of footprinting is identifying the level of risk associated with the organization's publicly accessible information. Footprinting, the first step in ethical hacking, refers to the process of collecting information about a target network and its environment. Using footprinting, you can find a number of opportunities to penetrate and assess the target organization's network.

After you complete the footprinting process in a methodological manner, you will obtain the blueprint of the security profile of the target organization. Here, the term "blueprint" refers to the unique system profile of the target organization acquired by footprinting.

There is no single methodology for footprinting, as information can be traced in a number of ways. However, the activity is important, as you need to gather all the crucial information about the target organization before beginning the hacking phase. For this reason, footprinting needs to be carried out in an organized manner. The information gathered in this step helps in uncovering vulnerabilities existing in the target network and in identifying different ways of exploiting these vulnerabilities.

Types of Footprinting/Reconnaissance

Footprinting can be categorized into passive footprinting and active footprinting.

- **Passive Footprinting**

  Passive footprinting involves gathering information about the target without direct interaction. It is mainly useful when the information gathering activities are not to be detected by the target. Performing passive footprinting is technically difficult, as active

traffic is not sent to the target organization from a host or anonymous hosts or services over the Internet. We can only collect archived and stored information about the target using search engines, social networking sites, and so on.

It involves:

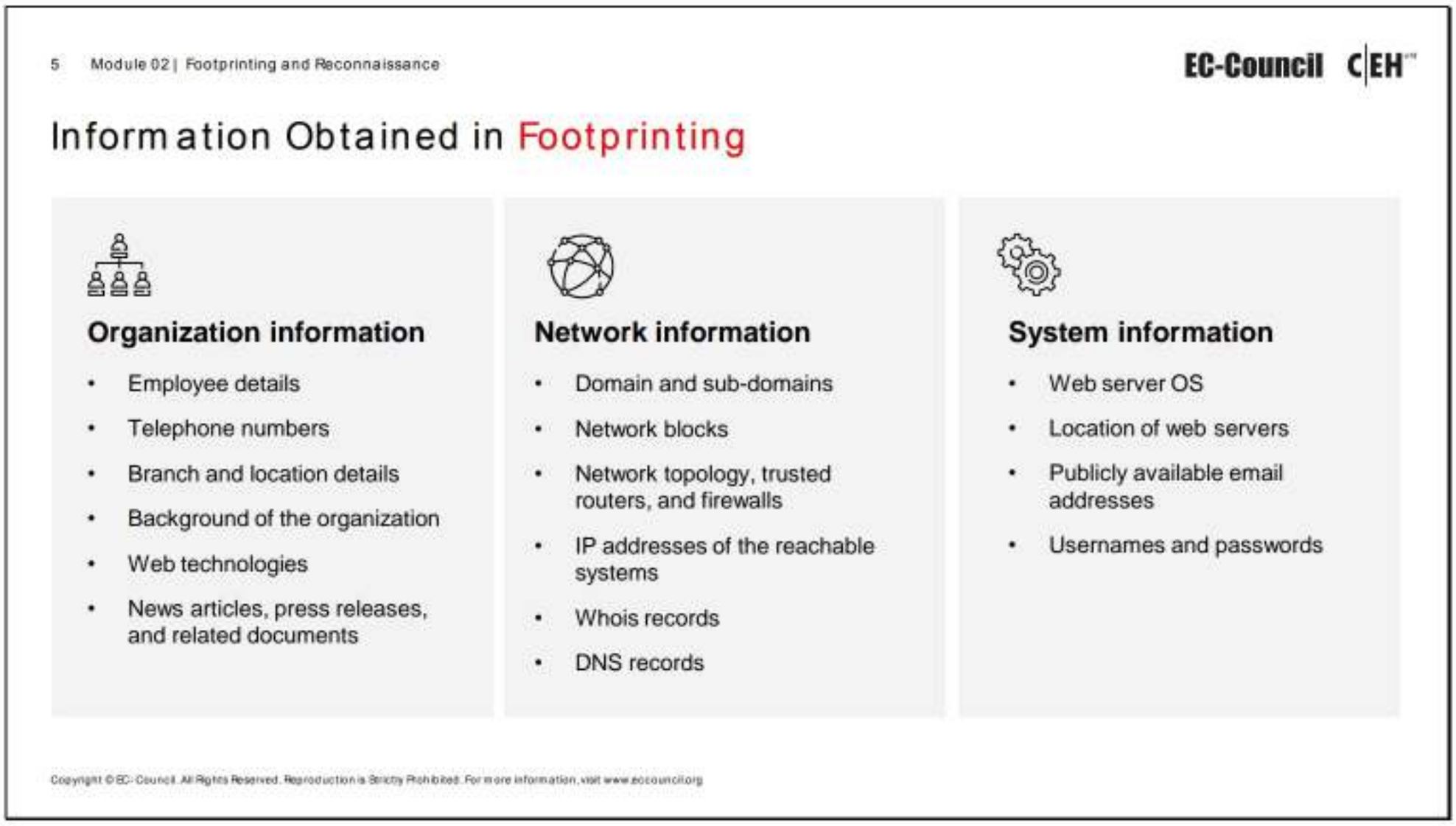
- Open-source Intelligence (OSINT) gathering
- Proprietary databases and paid services
- Sharing intelligence with partner organizations or industry groups

- **Active Footprinting**

  Active footprinting involves gathering information about the target with direct interaction. In active footprinting, the target may recognize the ongoing information gathering process, as we overtly interact with the target network. Active footprinting requires more preparation than passive footprinting, as it may leave traces that may alert the target organization.

  It involves:

  - DNS interrogation
  - Social engineering
  - Network/port scanning
  - User and service enumeration



Information Obtained in Footprinting

The major objectives of footprinting include collecting the network information, system information, and organizational information of the target. By conducting footprinting across different network levels, you can gain information such as network blocks, specific IP addresses, employee details, and so on. Such information can help attackers in gaining access to sensitive data or performing various attacks on the target network.

- **Organization Information**: The information about an organization is available from its website. In addition, you can query the target's domain name against the Whois database and obtain valuable information.

  The information collected includes:

  - Employee details (employee names, contact addresses, designations, and work experience)
  - Addresses and mobile/telephone numbers
  - Branch and location details
  - Partners of the organization
  - Web links to other company-related sites
  - Background of the organization
  - Web technologies
  - News articles, press releases, and related documents
  - Legal documents related to the organization
  - Patents and trademarks related to the organization

Attackers can access organizational information and use such information to identify key personnel and launch social engineering attacks to extract sensitive data about the entity.

- **Network Information**: You can gather network information by performing Whois database analysis, trace routing, and so on.

  The information collected includes:

  - Domain and sub-domains
  - Network blocks
  - Network topology, trusted routers, and firewalls
  - IP addresses of the reachable systems
  - Whois records
  - DNS records and related information

- **System Information**: You can gather system information by performing network footprinting, DNS footprinting, website footprinting, email footprinting, and so on.

  The information collected includes:

  - Web server OS
  - Location of web servers
  - Publicly available email addresses
  - Usernames, passwords, and so on.

Objectives of Footprinting

To build a hacking strategy, attackers must gather information about the target organization's network. They then use such information to identify the easiest way to break through the organization's security perimeter. As mentioned previously, the footprinting methodology makes it easy to gather information about the target organization and plays a vital role in the hacking process.

Footprinting provides an outline of the security posture, such as the placement of firewalls, proxies, and other security solutions. Hackers can analyze the footprinting report to identify loopholes in the security posture of the target organization and build a hacking plan accordingly.

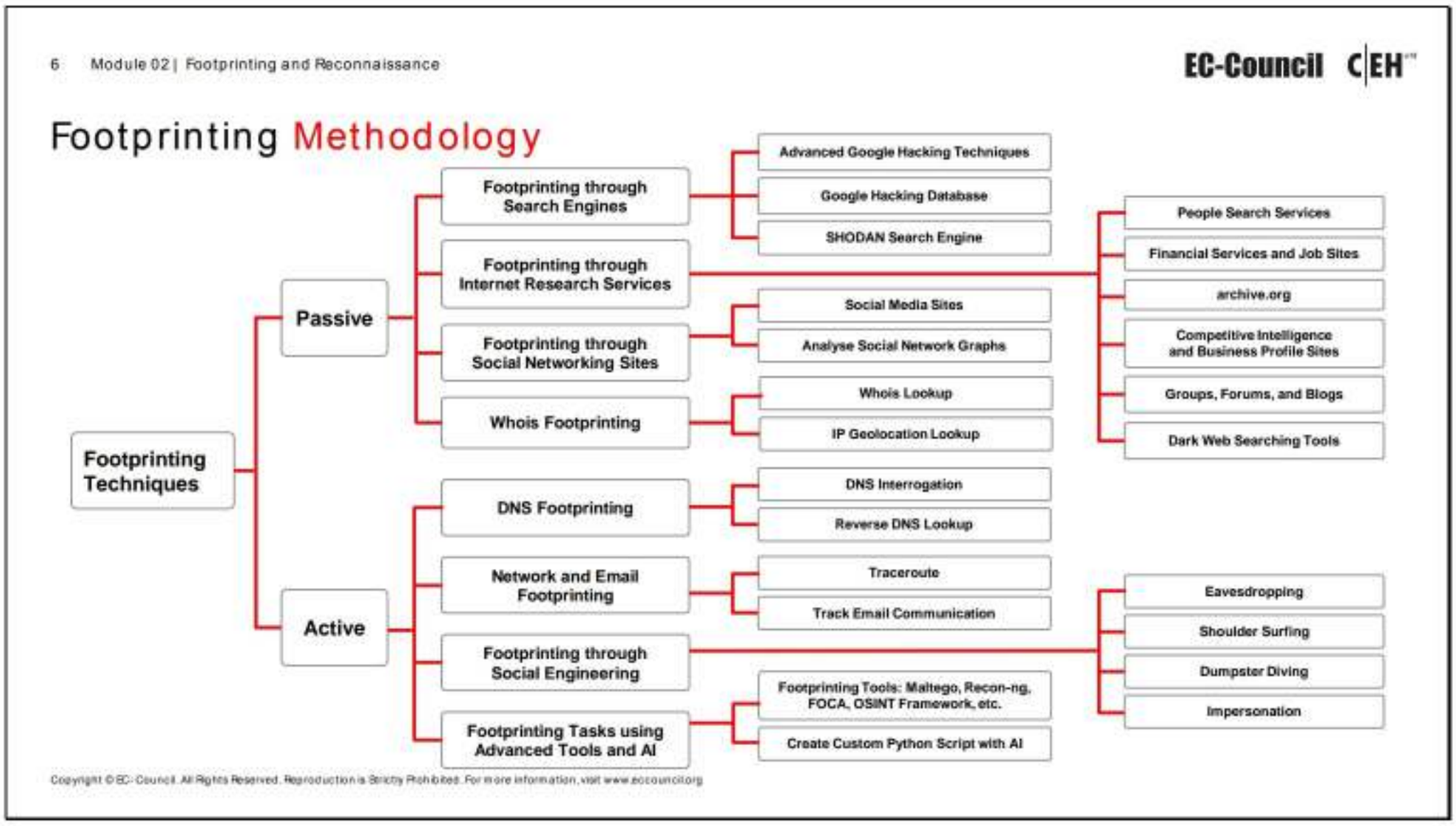
By using a combination of tools and techniques, attackers can take an unknown entity (for example, XYZ Organization) and reduce it to a specific range of domain names, network blocks, and individual IP addresses of systems directly connected to the Internet, in addition to other details pertaining to its security posture.

A detailed footprint provides maximal information about the target organization, allowing the attacker to identify vulnerabilities in the target systems to select appropriate exploits. Attackers can build their own information database regarding the security weaknesses of the target organization. Such a database can then help in identifying the weakest link in the organization's security perimeter.

Footprinting Threats

The following are assorted threats made possible through footprinting:

- **Social Engineering**: Without using any intrusion methods, hackers directly and indirectly collect information through persuasion and other means. Hackers gather crucial information from willing employees who are unaware of the hackers' intent.
- **System and Network Attacks**: Footprinting enables an attacker to perform system and network attacks. Thus, attackers can gather information related to the target organization's system configuration, the operating system running on the machine, and so on. Using this information, attackers can find vulnerabilities in the target system and then exploit such vulnerabilities. They can then take control of a target system or the entire network.
- **Information Leakage**: Information leakage poses a threat to any organization. If sensitive information of an entity falls into the hands of attackers, they can mount an attack based on the information or alternatively use it for monetary benefit.
- **Privacy Loss**: Through footprinting, hackers can access the systems and networks of the organization and even escalate the privileges up to admin levels, resulting in the loss of privacy for the organization as a whole and for its individual personnel.
- **Corporate Espionage**: Corporate espionage is a central threat to organizations, as competitors often aim to attempt to acquire sensitive data through footprinting. Through this approach, competitors can launch similar products in the market, alter prices, and generally undermine the market position of a target organization.
- **Business Loss**: Footprinting can have a major effect on organizations such as online businesses and other e-commerce websites as well as banking and finance-related businesses. Billions of dollars are lost every year due to malicious attacks by hackers.



Footprinting Methodology

Now that you are familiar with footprinting concepts and potential threats, we will discuss the footprinting methodology. The footprinting methodology is a procedure for collecting information about a target organization from all available sources. It involves gathering information about a target organization, such as URLs, locations, establishment details, number of employees, specific range of domain names, contact information, and other related information. Attackers collect this information from publicly accessible sources such as search engines, social networking sites, Whois databases, and so on. The diagram given below illustrates the common techniques used to collect information about the target organization from different sources.

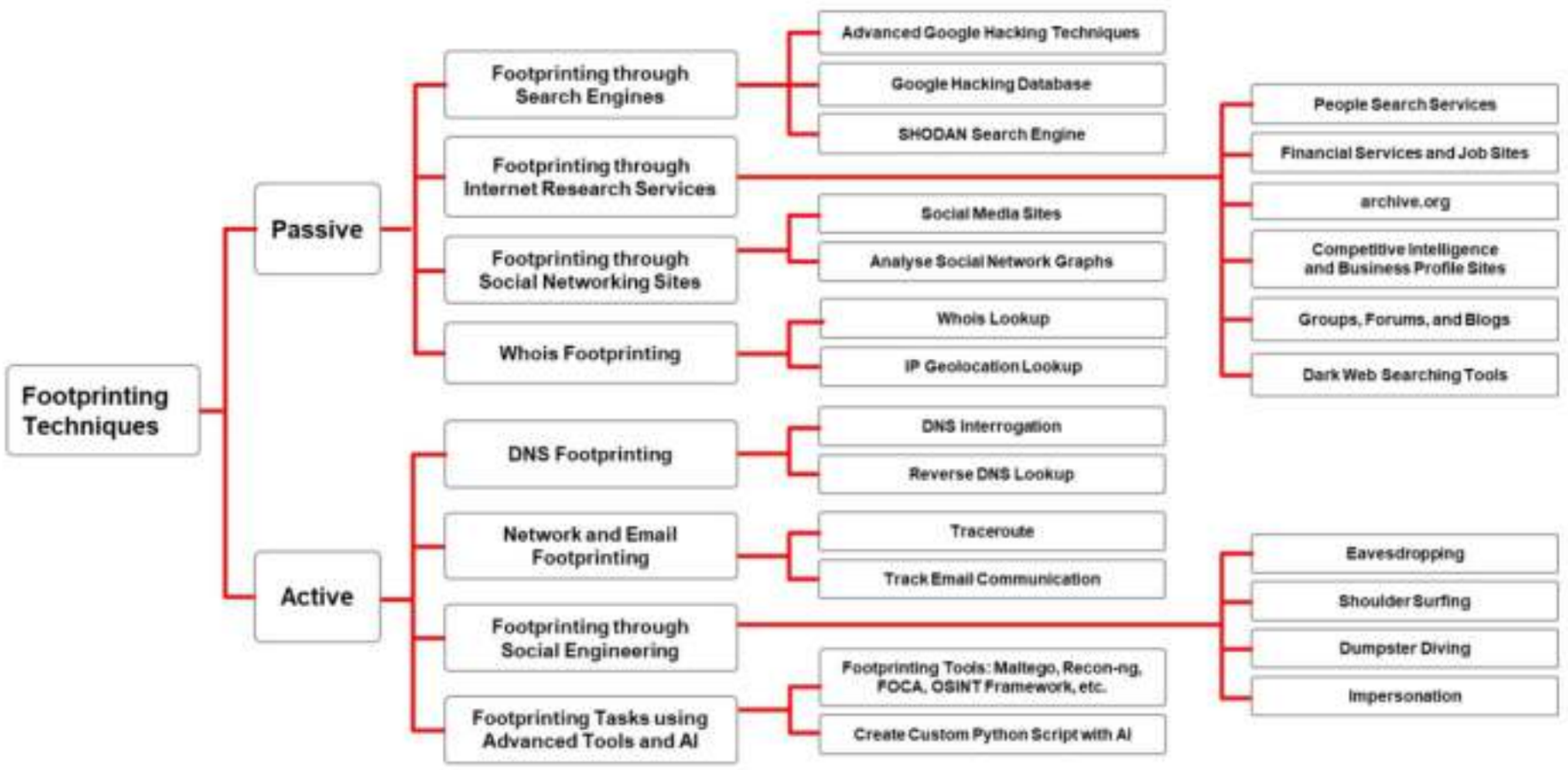


Figure 2.1: Footprinting Techniques

Footprinting through Search Engines

Search engines are the main sources of key information about a target organization. They play a major role in extracting critical details about a target from the Internet. Search engines use automated software, i.e., crawlers, to continuously scan active websites and add the retrieved results in the search engine index that is further stored in a massive database. When a user queries the search engine index, it returns a list of Search Engine Results Pages (SERPs). These results include web pages, videos, images, and many different file types ranked and displayed according to their relevance. Many search engines can extract target organization information such as technology platforms, employee details, login pages, intranet portals, contact information, and so on. The information helps the attacker in performing social engineering and other types of advanced system attacks.

Footprinting Using Advanced Google Hacking Techniques

- Attackers use search engines to **extract information about a target**, such as employed technology platforms, employee details, login pages, and intranet portals, which help the attacker to perform social engineering and other types of advanced system attacks
- Google hacking refers to the use of advanced Google search operators for **creating complex search queries** to extract sensitive or hidden information that helps attackers **find vulnerable targets**

Popular Google advanced search operators

Search Operator	Purpose
[cache:]	Displays the web pages stored in the Google cache
[link:]	Lists web pages that have links to the specified web page
[related:]	Lists web pages that are similar to the specified web page
[info:]	Presents some information that Google has about a particular web page
[site:]	Restricts the results to those websites in the given domain

Search Operator	Purpose
[allintitle:]	Restricts the results to those websites containing all the search keywords in the title
[intitle:]	Restricts the results to documents containing the search keyword in the title
[allinurl:]	Restricts the results to those containing all the search keywords in the URL
[inurl:]	Restricts the results to documents containing the search keyword in the URL
[location:]	Finds information for a specific location

Footprinting Using Advanced Google Hacking Techniques

A Google search could reveal submissions to forums by security personnel, disclosing the brands of firewalls or antivirus software used by the target. This information helps the attacker in identifying vulnerabilities in such security controls.

For example, consider an organization, perhaps Microsoft. Type **Microsoft** in the **Search** box of a search engine and press **Enter**; this will display the results containing information about Microsoft. Browsing the results often provides critical information such as physical location, contact addresses, services offered, number of employees, and so on, which may prove to be a valuable source for hacking.

Examples of major search engines include Google, Bing, Yahoo, Ask, Aol, Baidu, Yandex, WolframAlpha, and DuckDuckGo.

Attackers can use advanced search operators available with these search engines and create complex queries to find, filter, and sort specific information regarding the target. Search engines are also used to find other sources of publicly accessible information. For example, you can type "top job portals" to find major job portals that provide critical information about the target organization.

As an ethical hacker, if you find any deleted pages/information about your company in SERPs or the search engine cache, you can request the search engine to remove the pages/information from its indexed cache.

Google hacking refers to the use of advanced Google search operators for creating complex search queries to extract sensitive or hidden information. The accessed information is then used by attackers to find vulnerable targets. Footprinting using advanced Google hacking techniques

involves locating specific strings of text within search results using advanced operators in the Google search engine.

Advanced Google hacking refers to the art of creating complex search engine queries. Queries can retrieve valuable data about a target company from Google search results. Through Google hacking, an attacker tries to find websites that are vulnerable to exploitation. Attackers can use the Google Hacking Database (GHDB), a database of queries, to identify sensitive data. Google operators help in finding the required text and avoiding irrelevant data. Using advanced Google operators, attackers can locate specific strings of text such as specific versions of vulnerable web applications. When a query without advanced search operators is specified, Google traces the search terms in any part of the webpage, including the title, text, URL, digital files, and so on. To confine a search, Google offers advanced search operators. These search operators help to narrow down the search query and obtain the most relevant and accurate output.

The syntax to use an advanced search operator is as follows: operator**: search_term**

Note: Do not enter any spaces between the operator and the query.

Some popular Google advanced search operators include:

Source: *https://www.googleguide.com*

- **site**: This operator restricts search results to the specified site or domain.

  For example, the [games site: www.certifiedhacker.com] query gives information on games from the certifiedhacker site.

- **allinurl**: This operator restricts results to only the pages containing all the query terms specified in the URL.

  For example, the [allinurl: google career] query returns only pages containing the words "google" and "career" in the URL.

- **inurl**: This operator restricts the results to only the pages containing the specified word in the URL.

  For example, the [inurl: copy site:www.google.com] query returns only Google pages in which the URL has the word "copy."

- **intext:** This operator displays the results containing the specific keyword within the body of the webpage.

  For example, the [intext:"vpn configuration"] query returns the pages containing the phrase "vpn configuration" in their body text.

- **allintitle**: This operator restricts results to only the pages containing all the query terms specified in the title.

  For example, the [allintitle: detect malware] query returns only pages containing the words "detect" and "malware" in the title.

- **intitle**: This operator restricts results to only the pages containing the specified term in the title.

  For example, the [malware detection intitle:help] query returns only pages that have the term "help" in the title, and the terms "malware" and "detection" anywhere within the page.

- **inanchor**: This operator restricts results to only the pages containing the query terms specified in the anchor text on links to the page.

  For example, the [Anti-virus inanchor:Norton] query returns only pages with anchor text on links to the pages containing the word "Norton" and the page containing the word "Anti-virus."

- **allinanchor**: This operator restricts results to only the pages containing all query terms specified in the anchor text on links to the pages.

  For example, the [allinanchor: best cloud service provider] query returns only pages for which the anchor text on links to the pages contains the words "best," "cloud," "service," and "provider."

- **cache**: This operator displays Google's cached version of a web page instead of the current version of the web page.

  For example, [cache:www.eff.org] will show Google's cached version of the Electronic Frontier Foundation home page.

- **link**: This operator searches websites or pages that contain links to the specified website or page.

  For example, [link:www.googleguide.com] finds pages that point to Google Guide's home page.

  Note: According to Google's documentation, "you cannot combine a link: search with a regular keyword search."

  Also note that when you combine link: with another advanced operator, Google may not return all the pages that match.

- **related**: This operator displays websites that are similar or related to the URL specified.

  For example, [related:www.microsoft.com] provides the Google search engine results page with websites similar to microsoft.com.

- **info**: This operator finds information for the specified web page.

  For example, [info:gothotel.com] provides information about the national hotel directory GotHotel.com home page.

- **location:** This operator finds information for a specific location.

  For example, [location: 4 seasons restaurant] will give you results based on the term "4 seasons restaurant."

- **filetype:** This operator allows you to search for results based on a file extension.

  For Example, [jasmine:jpg] will provide jpg files based on jasmine.

- **source:** This operator displays information from a specific website in Google News.

  For example, [Malware news source:"Hacker News"] returns articles from Hacker News containing the word "Malware".

- **phonebook:** This operator finds the residential and business phone numbers of a person or organization.

  For example, [phonebook:Sundar Pichai] will provide Sundar Pichai's phone number.

- **before:** This operator filters search results to include only content published before a specified date.

  For example, [ransomware before:2020-06-29] will give results about the ransomware that occurred before June 29, 2020.

- **after:** This operator finds information that was published after a certain date.

  For example, [site:wikipedia.org after:2023-01-01 artificial intelligence] will retrieve Wikipedia articles about artificial intelligence published after January 1, 2023.

What can a Hacker Do with Google Hacking?

An attacker can create complex search-engine queries to filter large amounts of search results to obtain information related to computer security. The attacker can use Google operators to locate specific strings of text within search results. Thus, the attacker can not only detect websites and web servers that are vulnerable to exploitation but also locate private and sensitive information about the target. Once a vulnerable site is identified, attackers attempt to launch various possible attacks, such as buffer overflow and SQL injection, which compromise information security.

Examples of sensitive information on public servers that an attacker can extract with the help of Google Hacking Database (GHDB) queries include:

- Error messages that contain sensitive information
- Files containing passwords
- Sensitive directories
- Pages containing logon portals
- Pages containing network or vulnerability data, such as IDS, firewall logs, and configurations
- Advisories and server vulnerabilities
- Software version information
- Web application source code
- Connected IoT devices and their control panels, if unprotected
- Hidden web pages such as intranet and VPN services

Example: Use Google Advance Operator syntax [**intitle:intranet inurl:intranet +intext:"human resources"**] to find sensitive information about a target organization and its employees. Attackers use the gathered information to perform social engineering attacks.

The screenshot below shows a Google search engine results page displaying the results for the query mentioned above.

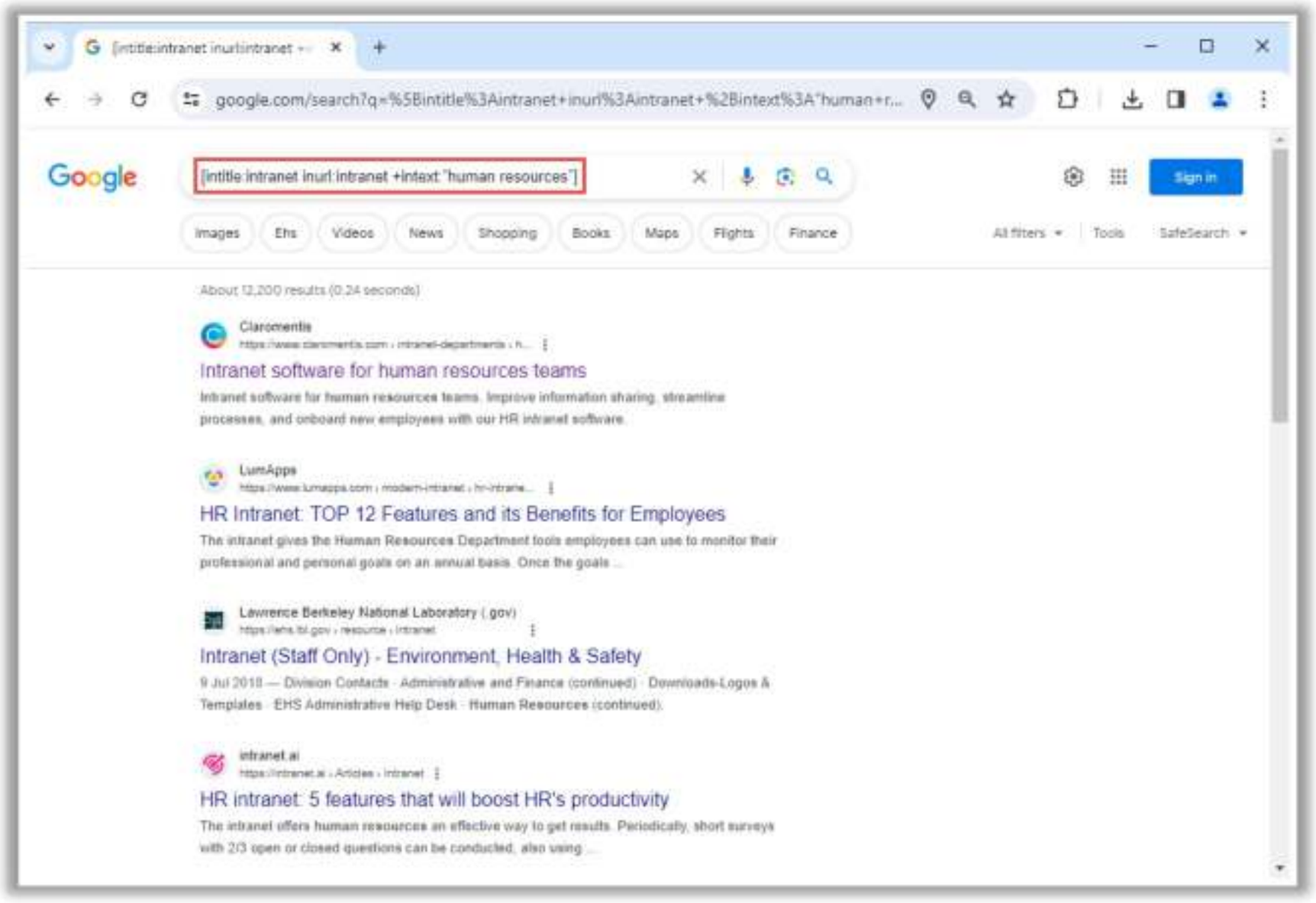
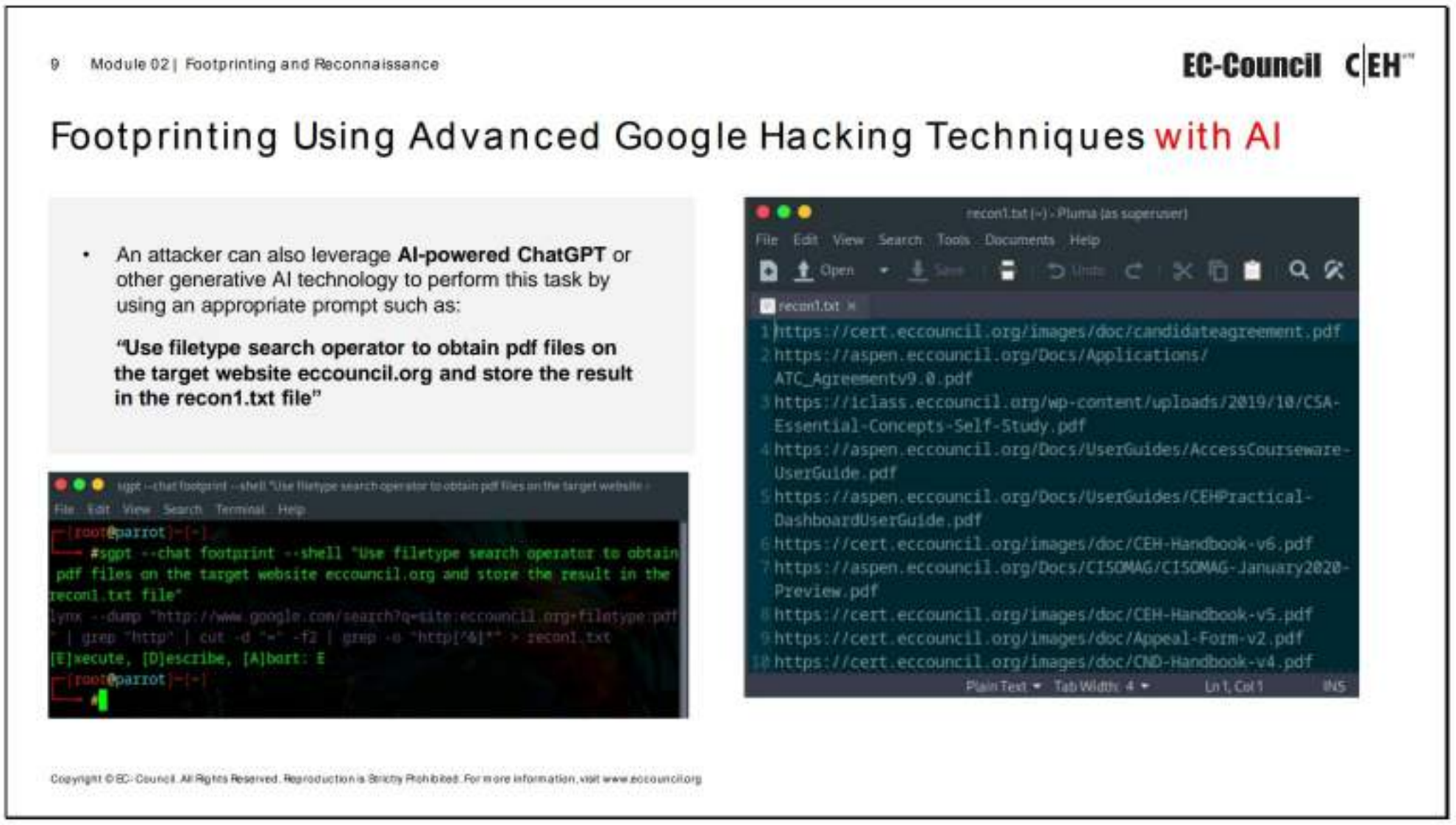


Figure 2.2: Search engine results for given Google Advance Operator syntax



Footprinting Using Advanced Google Hacking Techniques with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly use advanced Google hacking techniques to acquire valuable insights about their target.

For instance,

An attacker can use AI-powered technologies such as ShellGPT to perform this task by using an appropriate prompt such as:

"Use filetype search operator to obtain pdf files on the target website eccouncil.org and store the result in the recon1.txt file"

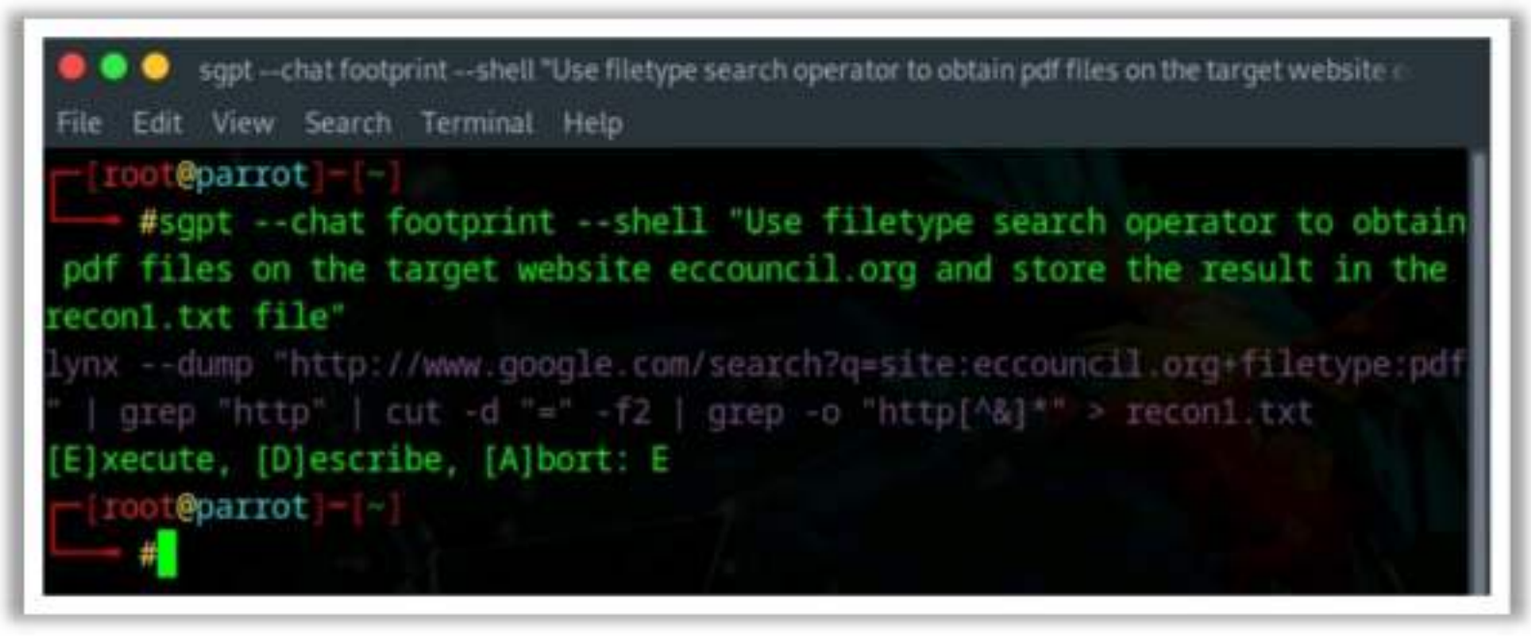


Figure 2.3: Prompt for Advanced Google Hacking with AI

The following shell command is designed to conduct advanced Google hacking using the "filetype" operator to specifically target PDF files within the eccouncil.org domain. The command then saves the obtained results to a file named "recon1.txt":

```
lynx --dump "http://www.google.com/search?q=site:eccouncil.org+filetype:pdf"
| grep "http" | cut -d "=" -f2 | grep -o "http[^&]*" > recon1.txt
```

- `` `lynx --dump "http://www.google.com/search?q=site:eccouncil.org+filetype:pdf"` ``: Initiates the Lynx web browser in dump mode to access Google's search results for PDF files within the eccouncil.org domain.
- `` `| grep "http"` ``: Filters out lines containing the string "http" from the Lynx output.
- `` `| cut -d "=" -f2` ``: Splits each line using the "=" delimiter and selects the second field.
- `` `| grep -o "http[^&]*"` ``: Searches for patterns starting with "http" followed by any characters except "&".
- `` `> recon1.txt` ``: Redirects the final output to a file named "recon1.txt" for storage.

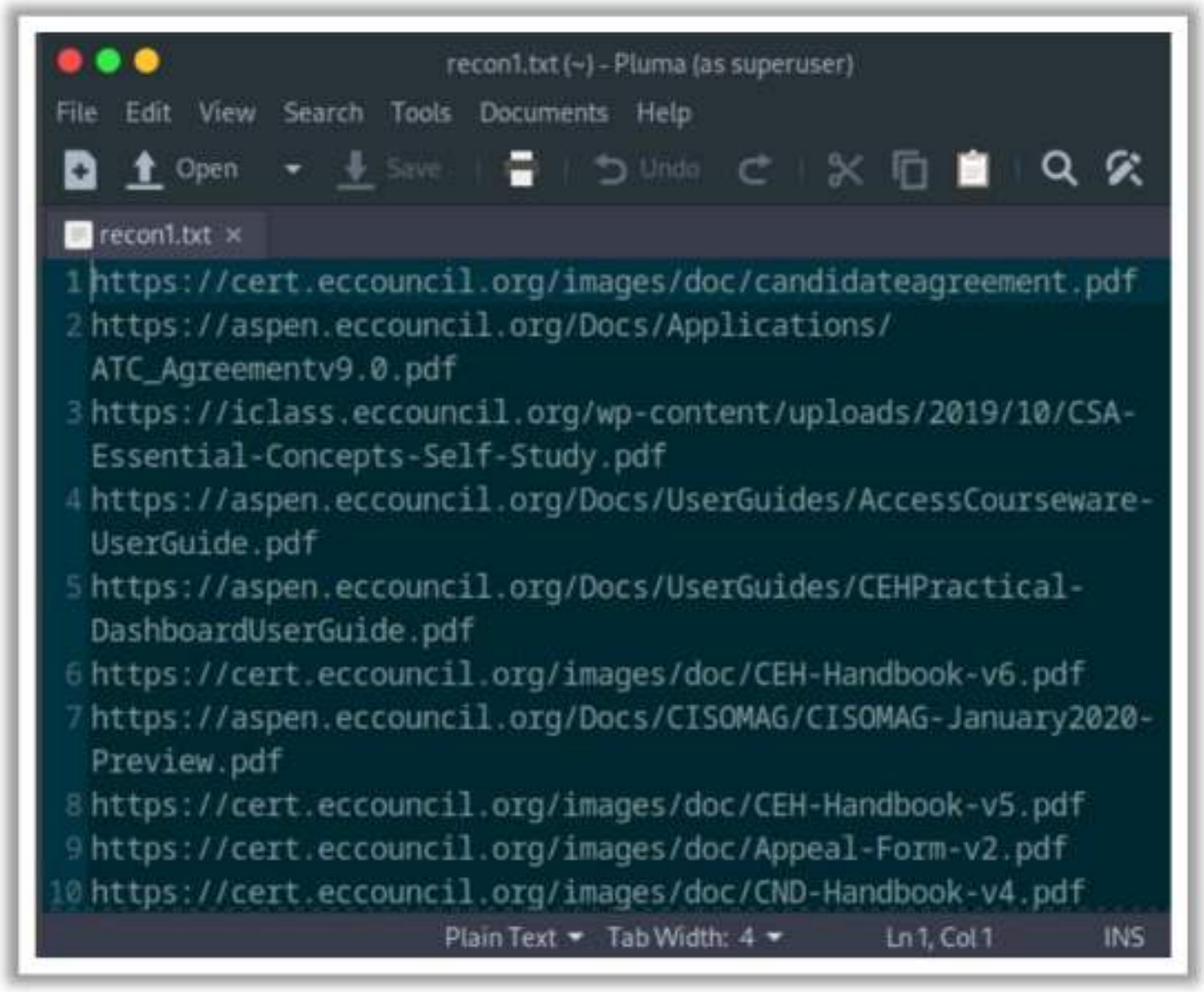
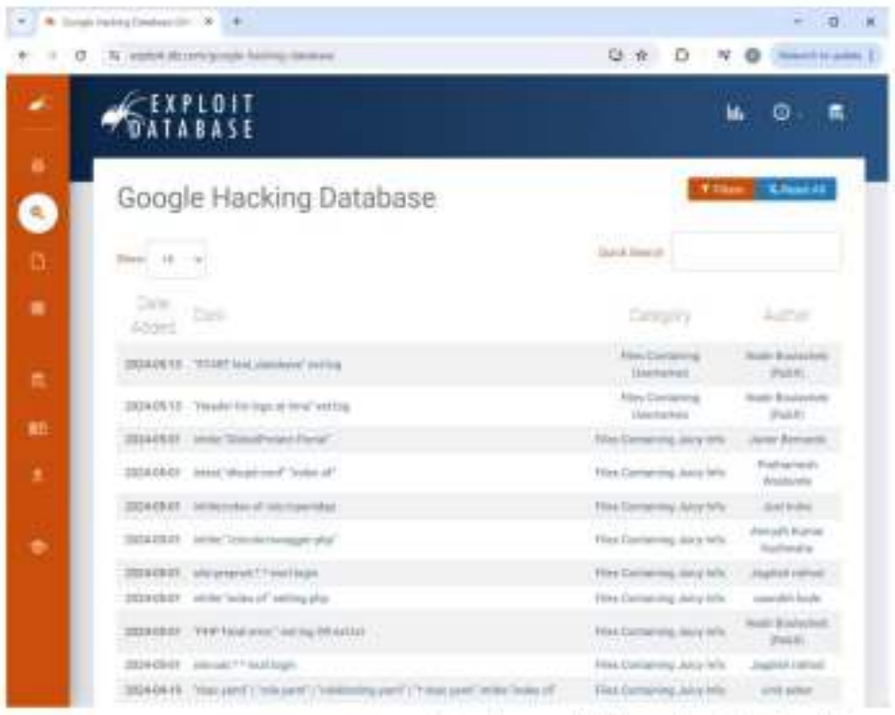


Figure 2.4: Output for Advanced Google Hacking with AI

Google Hacking Database

- The Google Hacking Database (GHDB) is an authoritative source for **querying the ever-widening reach of the Google search engine**
- Attackers use **Google dorks** in Google advanced search operators to extract sensitive information about their target, such as exposed files, directories, and devices that could be exploited, vulnerable servers, error messages, sensitive files, login pages, and websites



https://www.exploit-db.com/google-hacking-database

Google Hacking Database

Source: *https://www.exploit-db.com/google-hacking-database*

The GHDB is a subset of the Exploit-DB and focuses on using Google search queries (often referred to as "Google Dorks") to find sensitive information inadvertently exposed on the web.

These queries exploit advanced Google search operators to uncover the following:

- **Sensitive files**: Such as configuration files, database dumps, and log files that may contain usernames, passwords, or other confidential data.
- **Exposed directories**: Open directories on web servers that might contain sensitive information.
- **Error messages**: Web server or application error messages that may reveal server configurations or vulnerabilities.
- **Vulnerable devices**: Identifying specific types of devices or software versions known to have vulnerabilities.

Google Hacking Database Categories:

- Footholds
- Files Containing Usernames
- Sensitive Directories
- Web Server Detection
- Vulnerable Files
- Vulnerable Servers
- Error Messages
- Files Containing Juicy Info
- Files Containing Passwords
- Sensitive Online Shopping Info
- Network or Vulnerability Data
- Pages Containing Login Portals
- Various Online Devices
- Advisories and Vulnerabilities

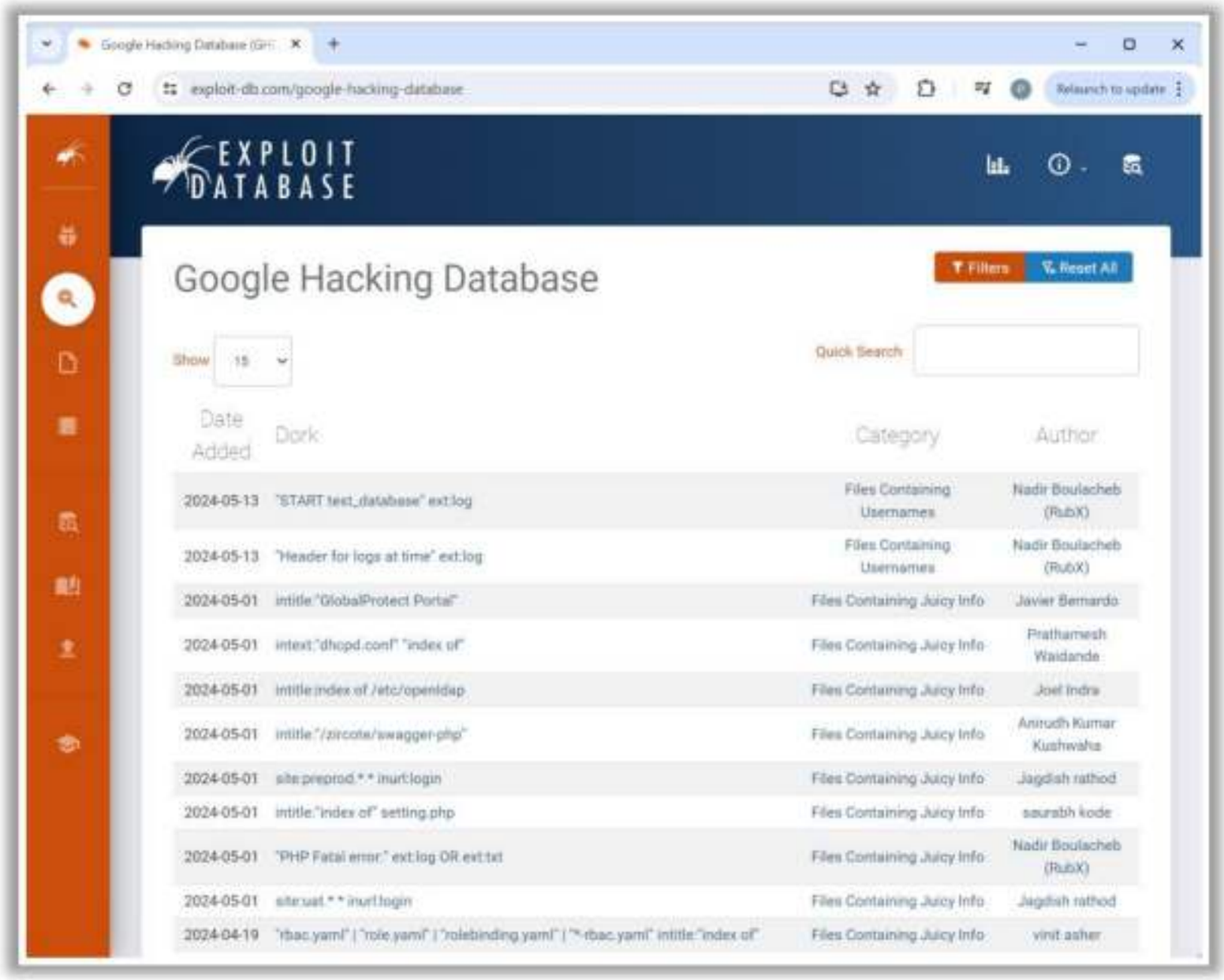


Figure 2.5: Screenshot of Google Hacking Database

Attackers can leverage the GHDB in various ways to identify and exploit vulnerabilities:

- **Reconnaissance**: Attackers use GHDB queries to gather information about potential targets, including exposed files, directories, and devices that could be exploited.
- **Exploiting Misconfigurations**: By identifying sensitive information exposed through misconfigured web servers or services, attackers can exploit these misconfigurations to gain unauthorized access.
- **Finding Vulnerable Systems**: Using GHDB, attackers can locate systems running outdated or vulnerable software versions, providing a starting point for further exploitation.
- **Credential Harvesting**: Sensitive information found through GHDB queries can include usernames and passwords, which attackers can use for credential stuffing or brute force attacks.
- **Identifying Open Ports and Services**: Some GHDB queries can reveal open ports and services on a network, giving attackers a map of potential entry points.

Attackers can also use SearchSploit, which is a command-line search tool for Exploit-DB that allows taking a copy of the Exploit database for remote use. It allows attackers to perform detailed offline searches through their locally checked-out copy of the repository. This capability is particularly useful for security assessments of segregated or air-gapped networks without Internet access.

VPN Footprinting through Google Hacking Database

Google hacking operators or Google dorks can be used for footprinting virtual private networks (VPNs). They provide information such as pages containing login portals and directories with keys of VPN servers.

The following tables summarize some Google hacking operators or Google dorks that are used to obtain specific information for VPN footprinting.

Google Dork	Description
inurl:"/sslvpn_logon.shtml" intitle:"User Authentication" "WatchGuard Technologies"	Finds pages containing login portals
inurl:/sslvpn/Login/Login	Finds VPN login portals
site:vpn.*.*/ intitle:"login"	
inurl:weblogin intitle:("USG20-VPN"|"USG20W-VPN"|USG40|USG40W|USG60|USG60W|USG110|USG210|USG310|USG1100|USG1900|USG2200|"ZyWALL110"|"ZyWALL310"|"ZyWALL1100"|ATP100|ATP100W|ATP200|ATP500|ATP700|ATP800|VPN50|VPN100|VPN300|VPN000|"FLEX")	Finds hosts with the Zyxel hardcoded password vulnerability
intext:Please Login SSL VPN inurl:remote/login intext:FortiClient	Finds Fortinet VPN login pages
site:vpn.*.*/ intext:"login" intitle:"login"	Retrieves various VPN login pages
intitle:"index of" /etc/openvpn/	Retrieves juicy information and sensitive directories
"-----BEGIN OpenVPN Static key V1-----" ext:key	Finds OpenVPN static keys
intitle:"index of" "vpn-config.*"	Retrieves juicy information about the vpn-config file
Index of / *.ovpn	Finds OpenVPN configuration files, some certificates, and keys
inurl:"/vpn/tmindex.html" vpn	Finds Netscaler and Citrix Gateway VPN login portals
intitle:"SSL VPN Service" + intext:"Your system administrator provided the following information to help understand and remedy the security conditions:"	Finds Cisco Adaptive Security Appliance (ASA) login web pages

Table 2.1: Google search queries for VPN footprinting

VPN Footprinting through Google Hacking Database with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can perform VPN footprinting effortlessly and acquire valuable insights.

For instance,

An attacker can use ChatGPT to perform this task by using an appropriate prompt such as:

"Use inurl search operator to obtain the Fortinet VPN login pages"

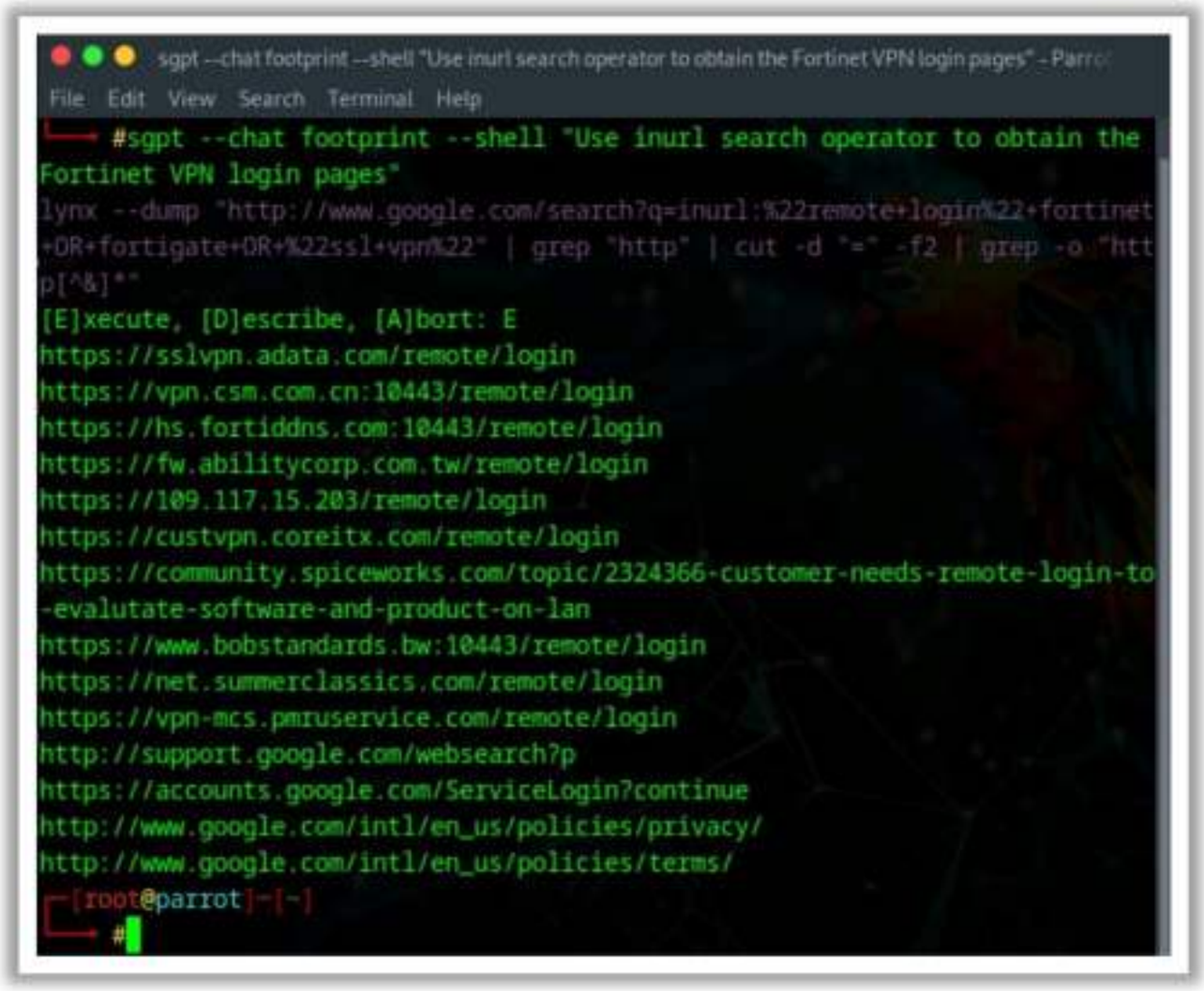


Figure 2.6: Use inurl search operator to obtain the Fortinet VPN login pages

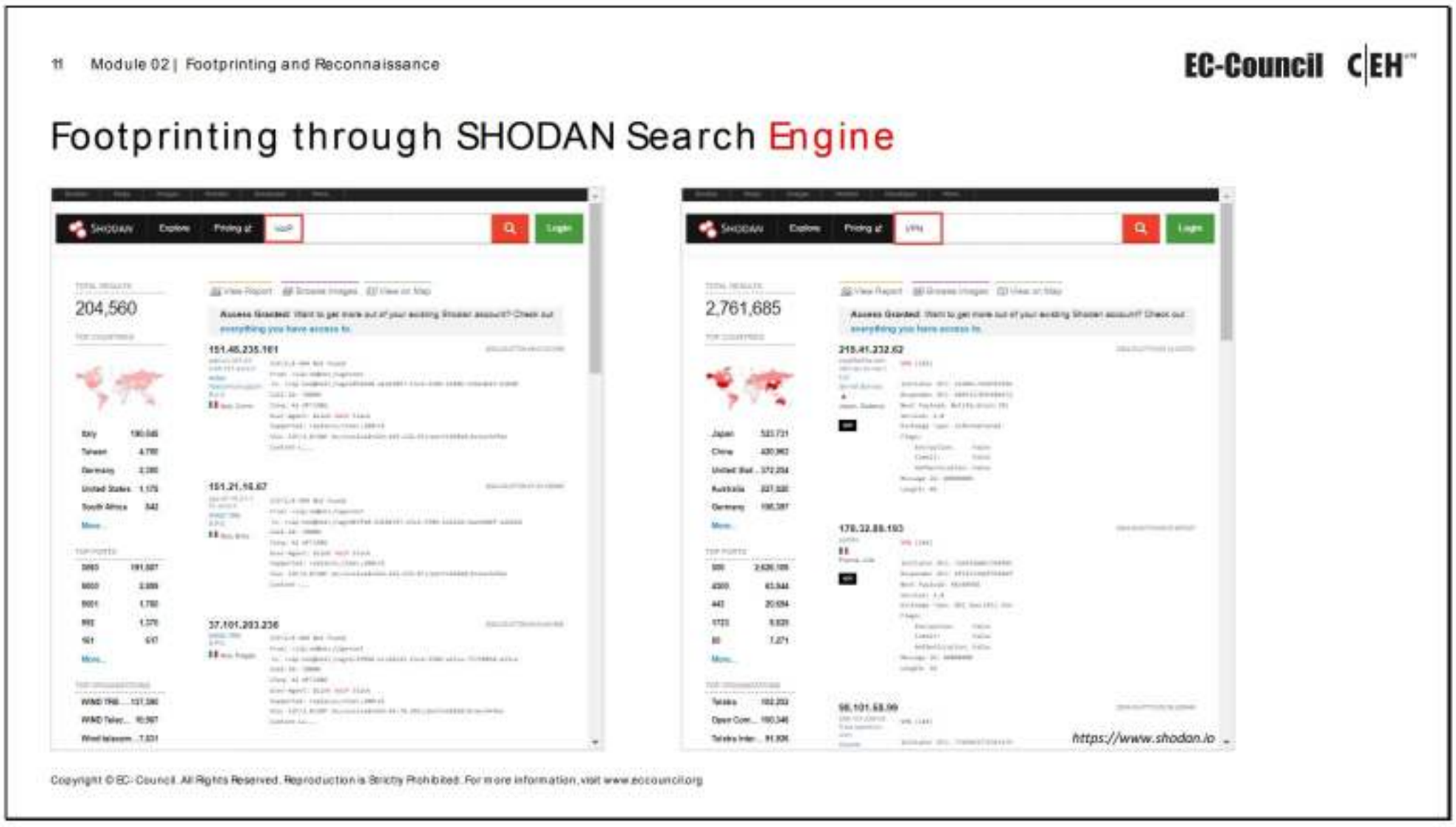
This command is designed to search for Fortinet VPN login pages using Lynx, a text-based web browser, along with Google search queries.

```
Lynx -dump
http://www.google.com/search?q=inurl:%22remote+login%22+fortinet+OR+fortigate
+OR+%22ssl+vpn%22 | grep "http | cut -d "" -f2 | grep -o "http[^&]*"
```

It fetches Google search results based on the specified query, extracts the URLs from the results, and then filters and formats the URLs before displaying them. The overall goal is to identify Fortinet VPN login pages from the search results.

- `` `lynx -dump "http://www.google.com/search?q=inurl:%22remote+login%22+fortinet+OR+fortigate+OR+%22ss1+vpn%22"` ``: This part of the command uses Lynx to access the Google search results page for the specified query. The query searches for web pages containing the phrase "remote login" in the URL and either "Fortinet", "Fortigate", or "ss1 vpn" in the content.
- `` `| grep "http"` ``: This part of the command pipes the output of the Lynx command to the `grep` command, which filters out lines containing the string "http".
- `` `| cut -d "=" -f2` ``: This part of the command pipes the output of the previous command to the `cut` command, which splits each line using the "=" delimiter and selects the second field.

- `` `| grep -o "httP[^&]*"` ``: This part of the command pipes the output of the previous command to another `grep` command, which searches for patterns starting with "http" followed by any characters except "&".



Footprinting through SHODAN Search Engine

Source: *https://www.shodan.io*

Shodan is a search engine that enables attackers to perform footprinting at various levels. It is used to detect devices and networks with vulnerabilities. A search in Shodan for VoIP and VPN footprinting can deliver various results, which will help gather VPN- and VoIP-related information. The following screenshots show some of the VPN and VoIP footprinting search results obtained through Shodan:

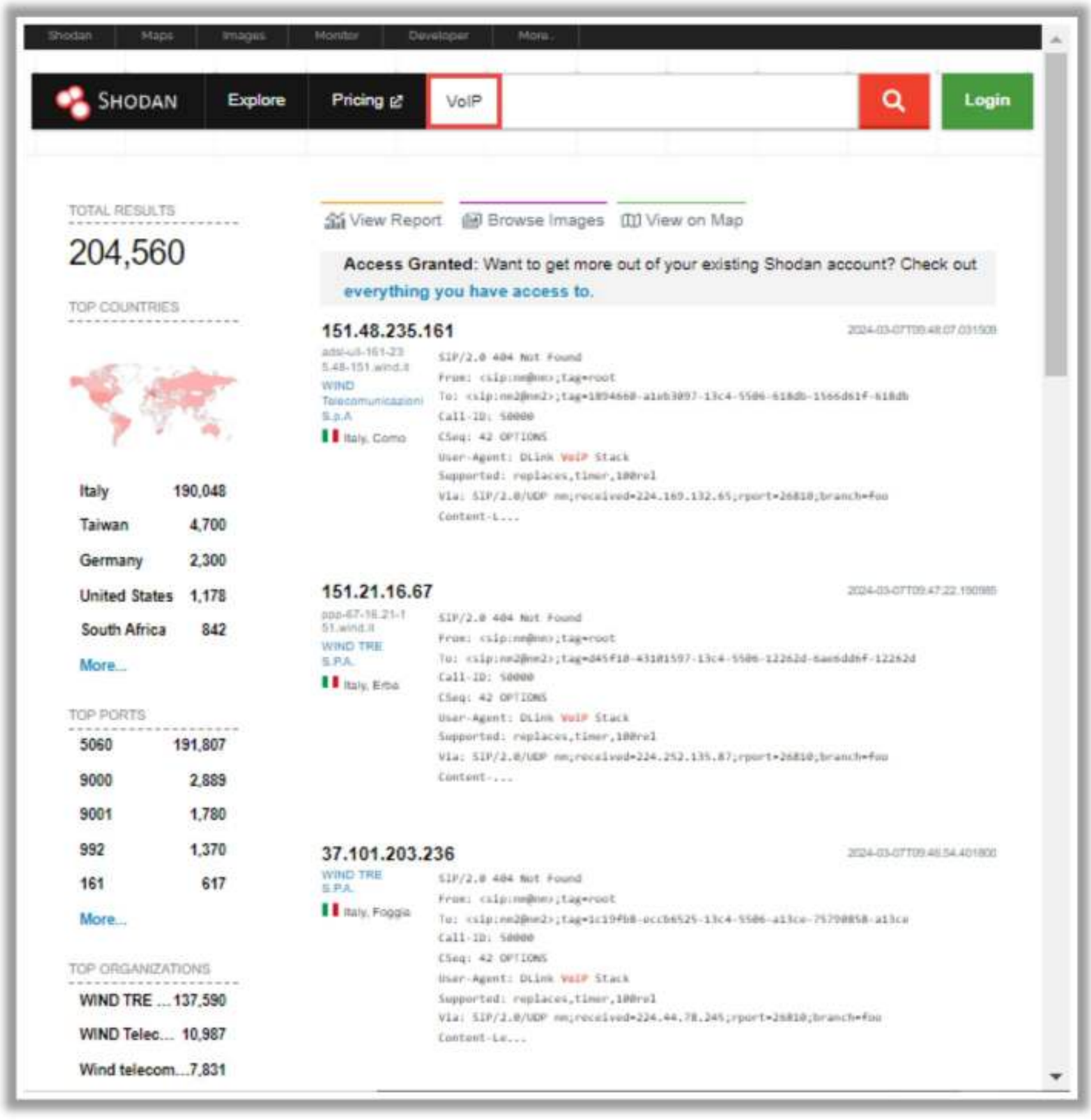


Figure 2.7: Screenshot of SHODAN search engine showing VoIP results

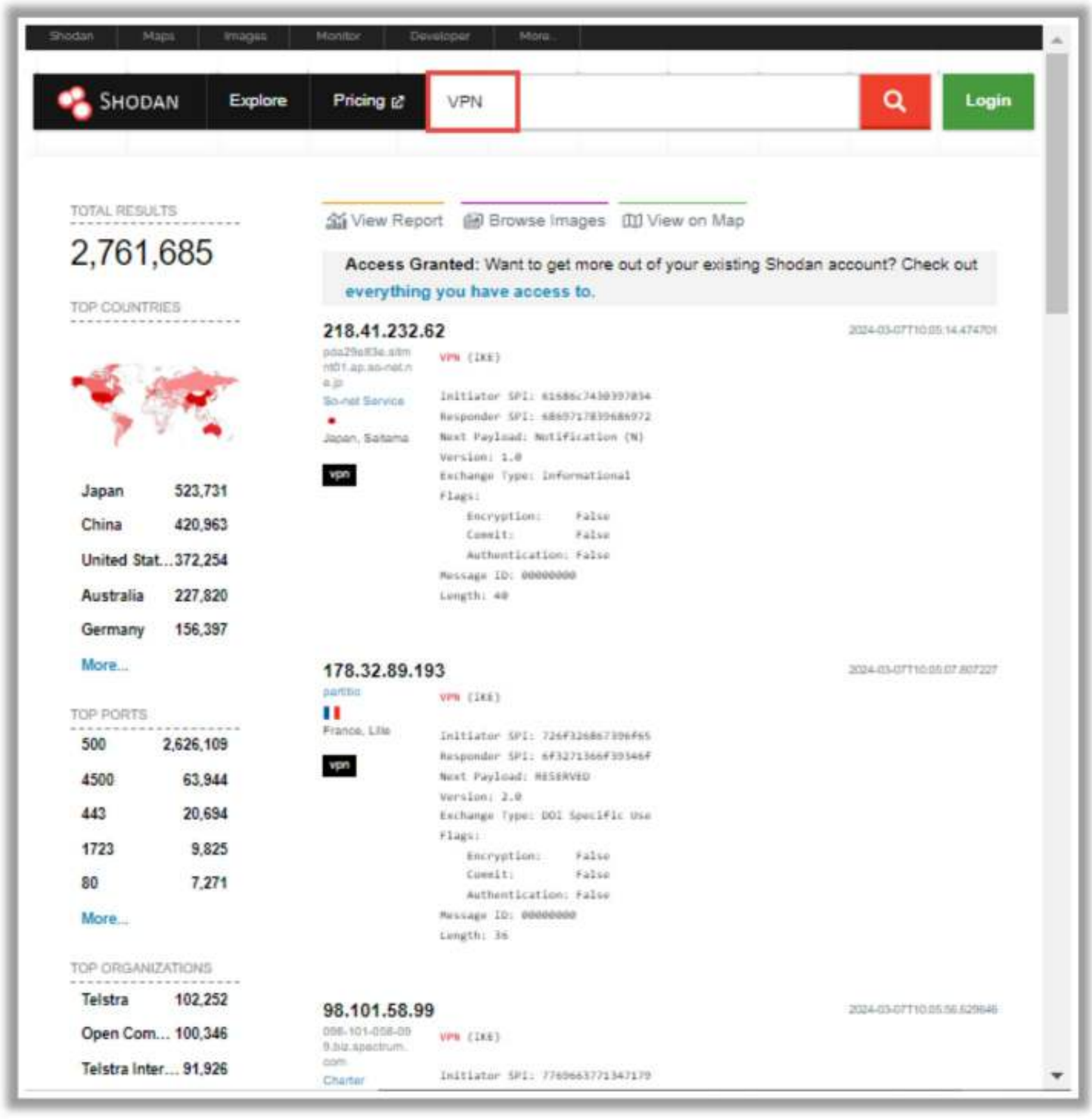


Figure 2.8: Screenshot of SHODAN search engine showing VPN results

Other Techniques for Footprinting through Search Engines

- **Gathering Information Using Google Advanced Search, Advanced Image Search, and Reverse Image Search**

  An attacker cannot always gather information easily from an information-rich site using only a normal search box. A complicated search involves a number of interrelated conditions.

  Google's Advanced search feature helps an attacker to perform complex web searching. With **Google Advanced Search** and **Advanced Image Search**, one can search the web more precisely and accurately. You can use these search features to achieve the same precision as that achieved using the advanced operators but without typing or

remembering the operators. Using Google's Advanced Search option, you can find sites that may link back to the target organization's website. This helps to extract information such as partners, vendors, clients, and other affiliations of the target website. You can use Google Advanced Image Search to acquire images of the target, its location, employees, and so on.

To perform an advanced search in Google, click **Settings** at the bottom-right of the **Google** home page, and then choose **Advanced search** in the menu or directly type ***https://www.google.com/advanced_search*** in the address bar. Advanced search allows you to specify any number of criteria that the search must match, as this pattern builds on the search box pattern by adding more search options. To do this, you choose a field. Then, enter the string you want to search for in the field's text box and click on the **Advanced Search** button. By default, various values are joined together with "and" (meaning all of them need to match) except for sets, blocks, and formats, which are joined together with "or" (meaning any of them can match).

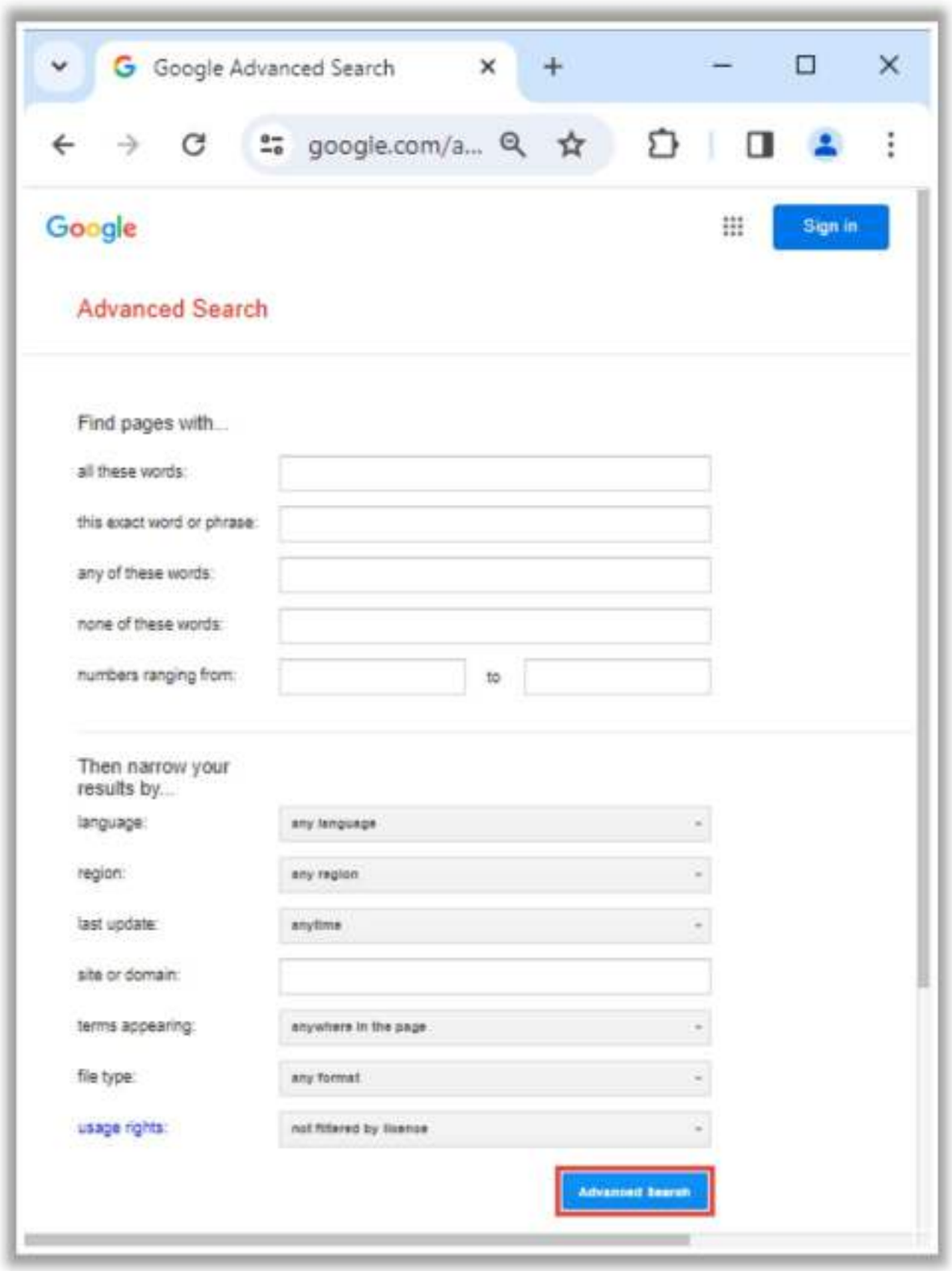


Figure 2.9: Google Advance Search

To perform an advanced image search in Google, type ***https://www.google.com/advanced_image_search*** in the address bar. Advanced image search allows you to tweak your image search in a number of ways. You can search based on image color, domain, file type, size, keyword, and so on. To do this, you choose a field. Then, enter the string you want to search for in the field's text box and click on the **Advanced Search** button.

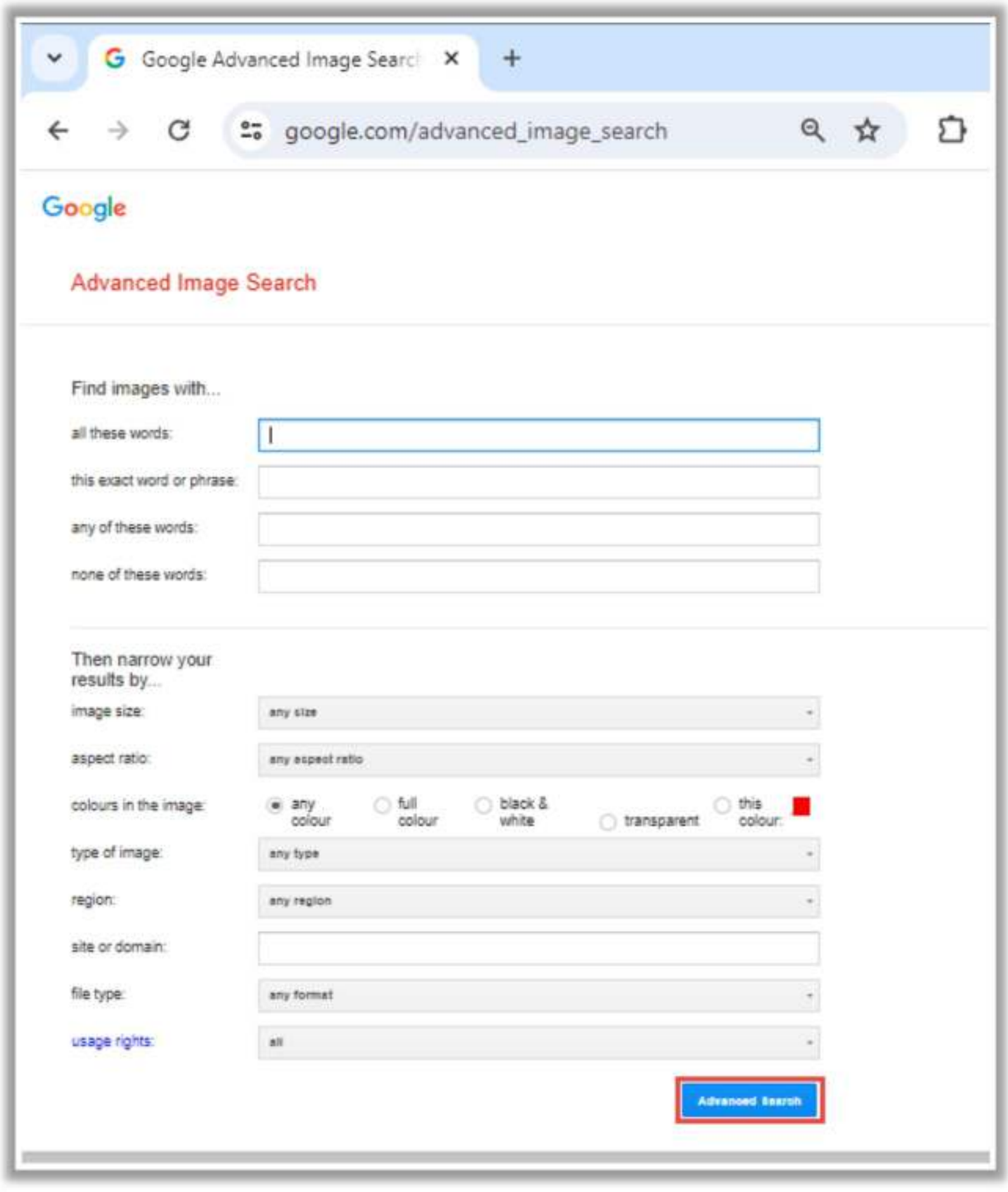


Figure 2.10: Google Advance Image Search

To perform a reverse image search in Google, type ***https://www.google.com/imghp*** in the address bar. Reverse image search allows you to use an image as a search query. You can upload an image or paste the URL of the image in the reverse image search engine. The search engine verifies the search engine index and displays all the online locations of the image in the search results page. The results obtained can help you in tracking the original source and details of the images, such as photographs, profile pictures, and memes.

Attackers use online tools such as Google Image Search, TinEye Reverse Image Search, Yahoo Image Search, Bing Image Search, and Pinterest Reverse image search to perform a reverse image search.

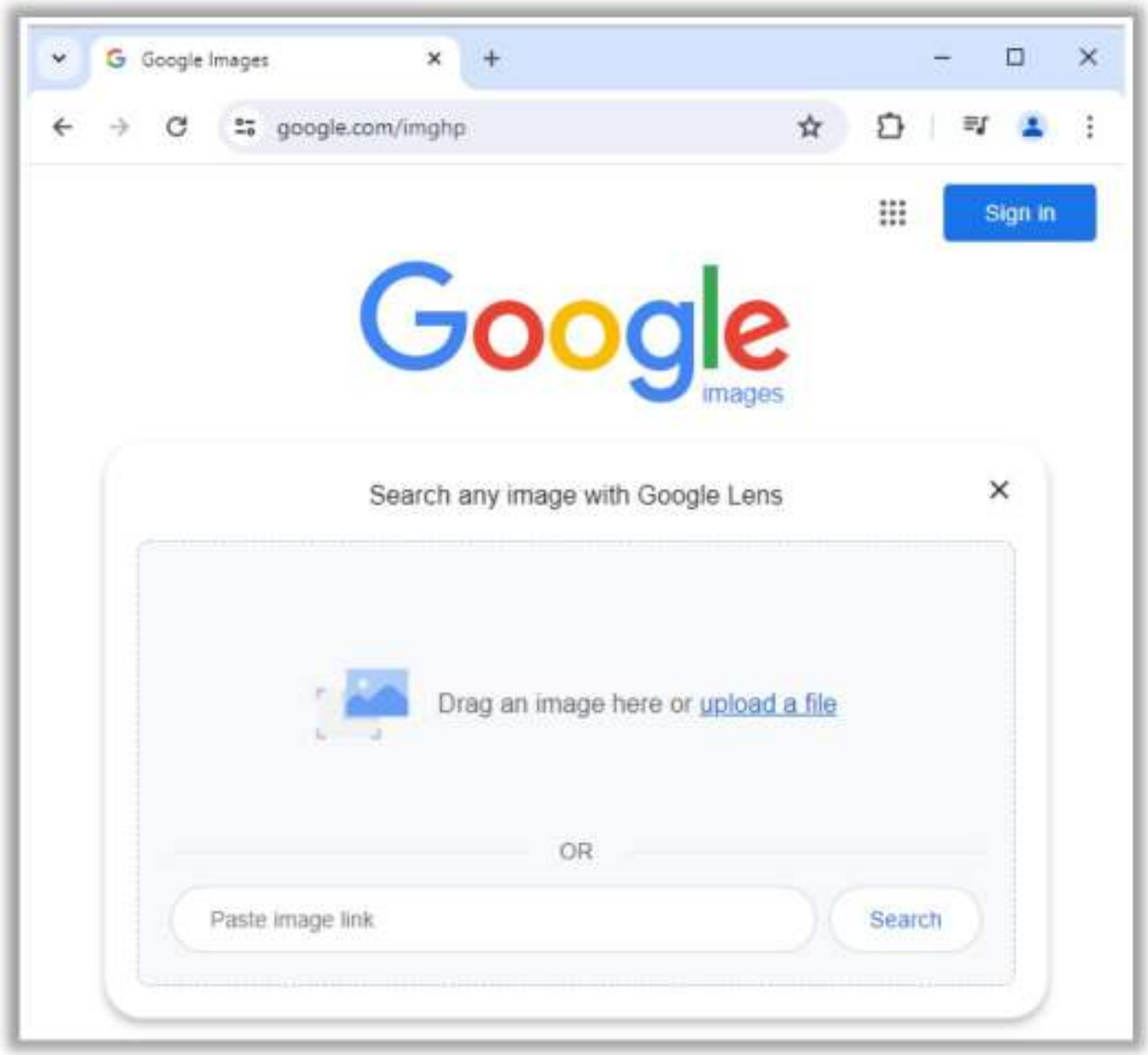


Figure 2.11: Reverse Image Search using Google

- **Gathering Information from Video Search Engines**

  Video search engines are Internet-based search engines that crawl the web for video content. These video search engines either provide the functionality of uploading and hosting video content on their own web servers or parse video content that is hosted externally. The video content obtained from video search engines is of high value, as it can be used for gathering information about the target. Video search engines such as YouTube, Google videos, Yahoo videos, and Bing videos allow attackers to search for video content based on the format type and duration.

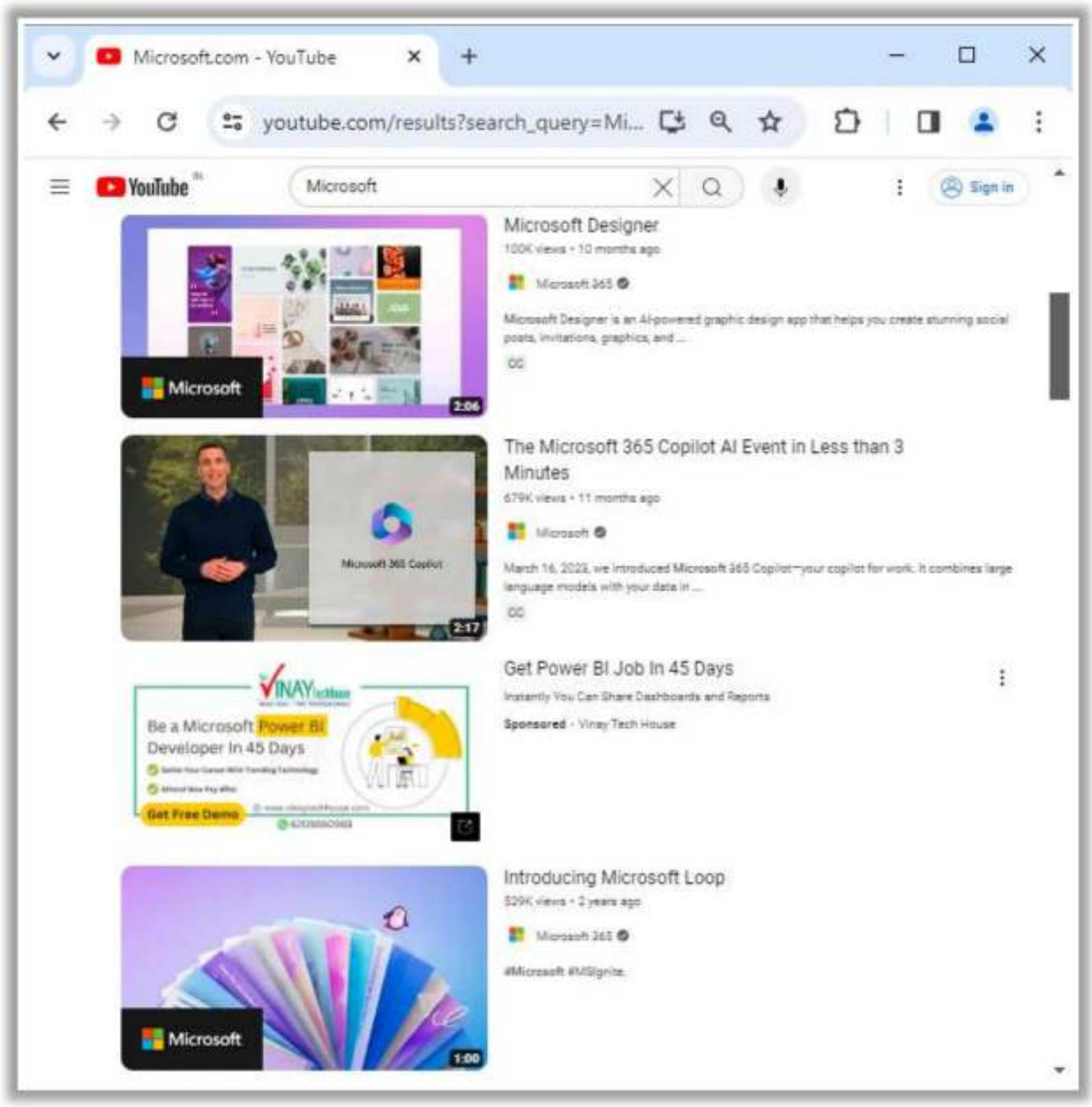


Figure 2.12: Screenshot of YouTube showing search results for Microsoft

After searching for videos related to the target using video search engines, an attacker can further analyze the video content to gather hidden information such as the time/date and thumbnail of the video. Using video analysis tools such as YouTube Metadata, YouTube DataViewer, MW Metadata, EZGif, and VideoReverser.com, an attacker can reverse a video or convert a video into text and other formats to extract critical information about the target.

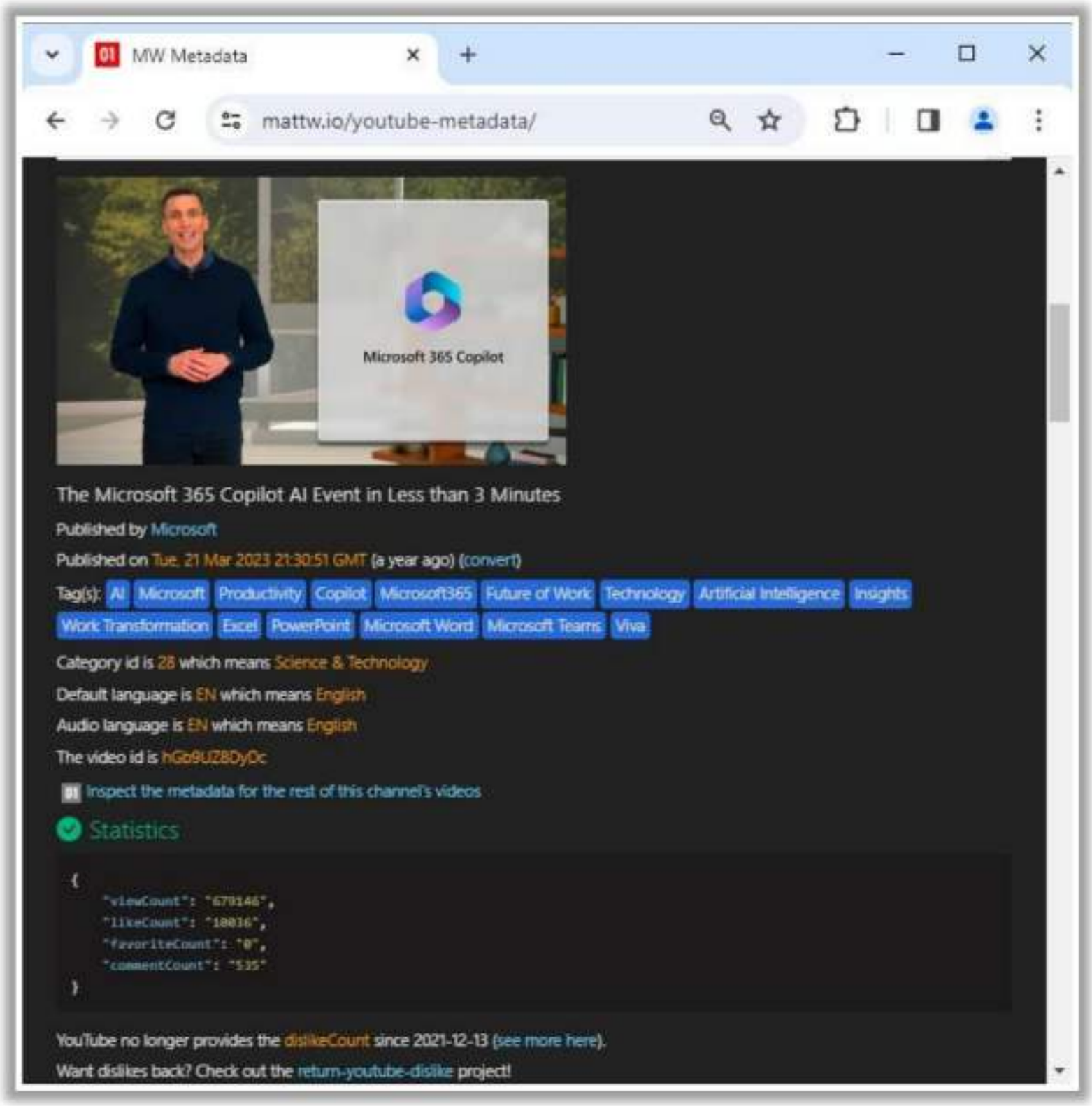


Figure 2.13: Screenshot of YouTube Metadata showing video analysis result

- **Gathering Information from Meta Search Engines**

  Meta search engines are a different type of search engines that use other search engines (Google, Bing, Ask.com, etc.) to produce their own results from the Internet in a very short time span. These search engines do not have their own search indexes; instead, they take the inputs from the users and simultaneously send out the queries to the third-party search engines to obtain the results. Once sufficient results are gathered, they are ranked according to their relevance and presented to the user through the web interface. Meta search engines also include a functionality whereby identical search results are filtered out so that if the user searches the same query again, then it will not display the same results twice. A meta search engine is advantageous compared to simple search engines, as it can retrieve more results with the same amount of effort.

  Using meta search engines, such as Startpage, MetaGer, and eTools.ch, attackers can send multiple search queries to several search engines simultaneously and gather substantially detailed information such as information from shopping sites (Amazon, eBay, BestBuy, etc.), images, videos, blogs, news, and articles from different sources.

Further, meta search engines also provide privacy to the search engine user by hiding the user's IP address.

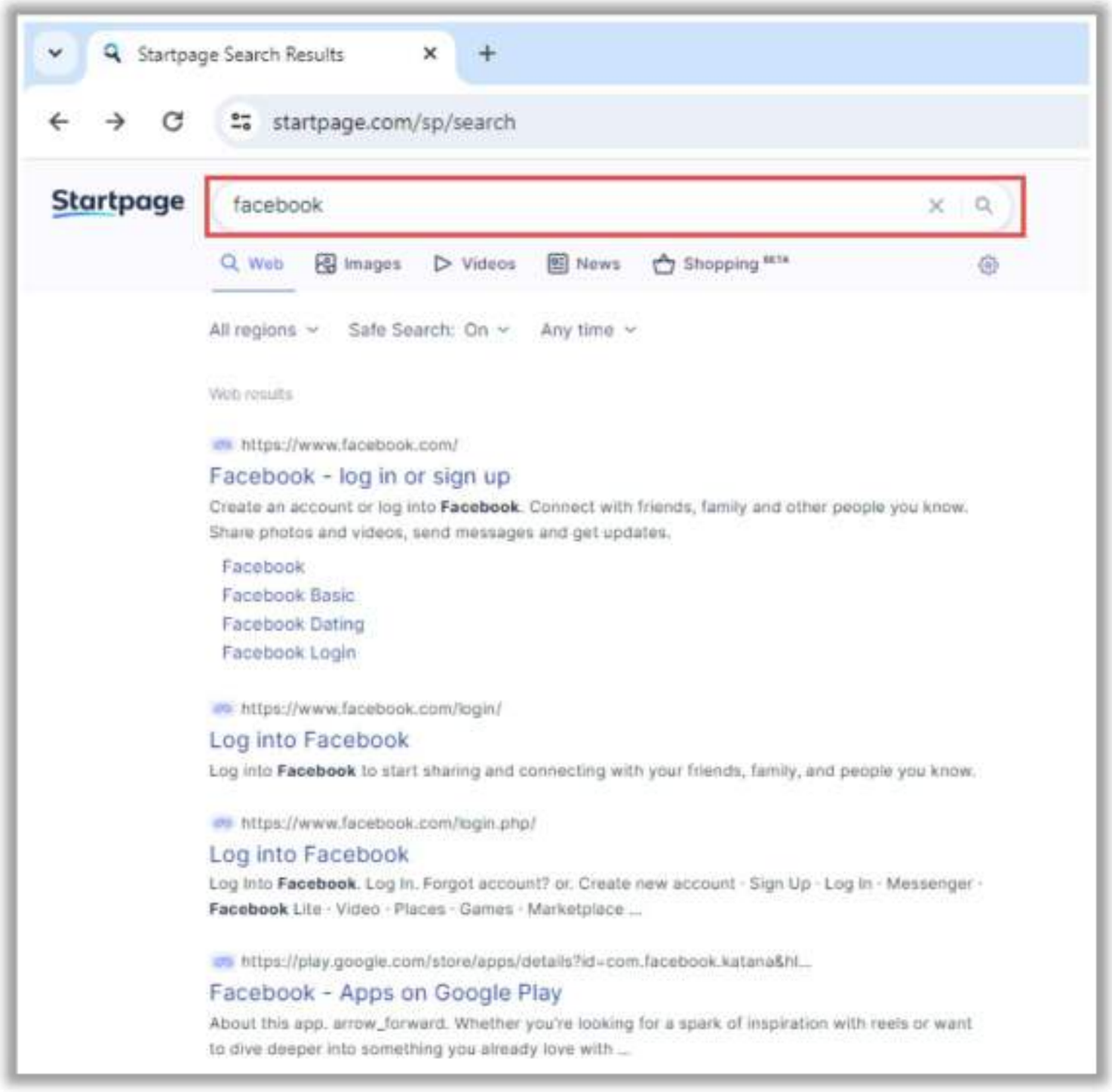


Figure 2.14: Screenshot of Meta Search Engine StartPage.com showing search results for Facebook

- **Gathering Information from File Transfer Protocol (FTP) Search Engines**

  FTP search engines are used to search for files located on FTP servers that contain valuable information regarding the target organization. Many industries, institutions, companies, and universities use FTP servers to store large file archives and other software shared among their employees. A special client such as FileZilla (*https://filezilla-project.org*) can be used to access FTP accounts; it also supports functionalities such as uploading, downloading, and renaming files. Although FTP servers are usually protected with passwords, many servers are left unsecured and can be accessed directly through web browsers.

  Using FTP search engines such as NAPALM FTP Indexer, FreewareWeb FTP File Search, Mamont, and Globalfilesearch.com, attackers can search for critical files and directories containing valuable information such as business strategies, tax documents, personal employee records, financial records, licensed software, and other confidential information.

Some of the important advanced Google search queries for finding FTP servers are listed in the below table.

Google Dork	Description
site:.in | .com | .net intitle:"index of" ftp	Finds files containing juicy information
intitle:"index of" "*/ftp.txt"	
intext:"index of" "ftp"	
inurl:WS_FTP.log	
intitle:index.of /cftp /robots.txt	
intitle: "Index of ftp passwords"	Finds files containing passwords
inurl: /ftp intitle:"office"	Detects the web server
inurl:/web-ftp.cgi	Finds pages containing login portals
site:sftp.*.*/ intext:"login" intitle:"server login"	
intitle:"Index of" ws_ftp.ini	Finds the "ws_ftp.ini" file, which contain usernames and passwords of FTP users
inurl:ftp -inurl:(http|https) intext:"@gmail.com" intext:subject fwd|confidential|important|CARD|cvv	Finds archived email conversations, at times revealing full credit-card numbers and customer information as well as private company emails
allintitle:"CrushFTP WebInterface"	Detects various pages of CrushFTP WebInterface, which includes login portals as well password reset/recovery page
"ws_ftp.log" ext:log	Finds sensitive directories
intitle:"Monsta ftp" intext:"Lock session to IP"	Shows websites that use the FTP service of Monsta FTP
"index of" /ftp/logs	Finds potential log files
intitle:"index of" inurl:ftp intext:admin	Lists admin folders on FTP servers

Table 2.2: Google search queries to find FTP servers

As shown in the screenshot, attackers can use NAPALM FTP Indexer, an online tool, to search for critical files and documents related to the target domain.

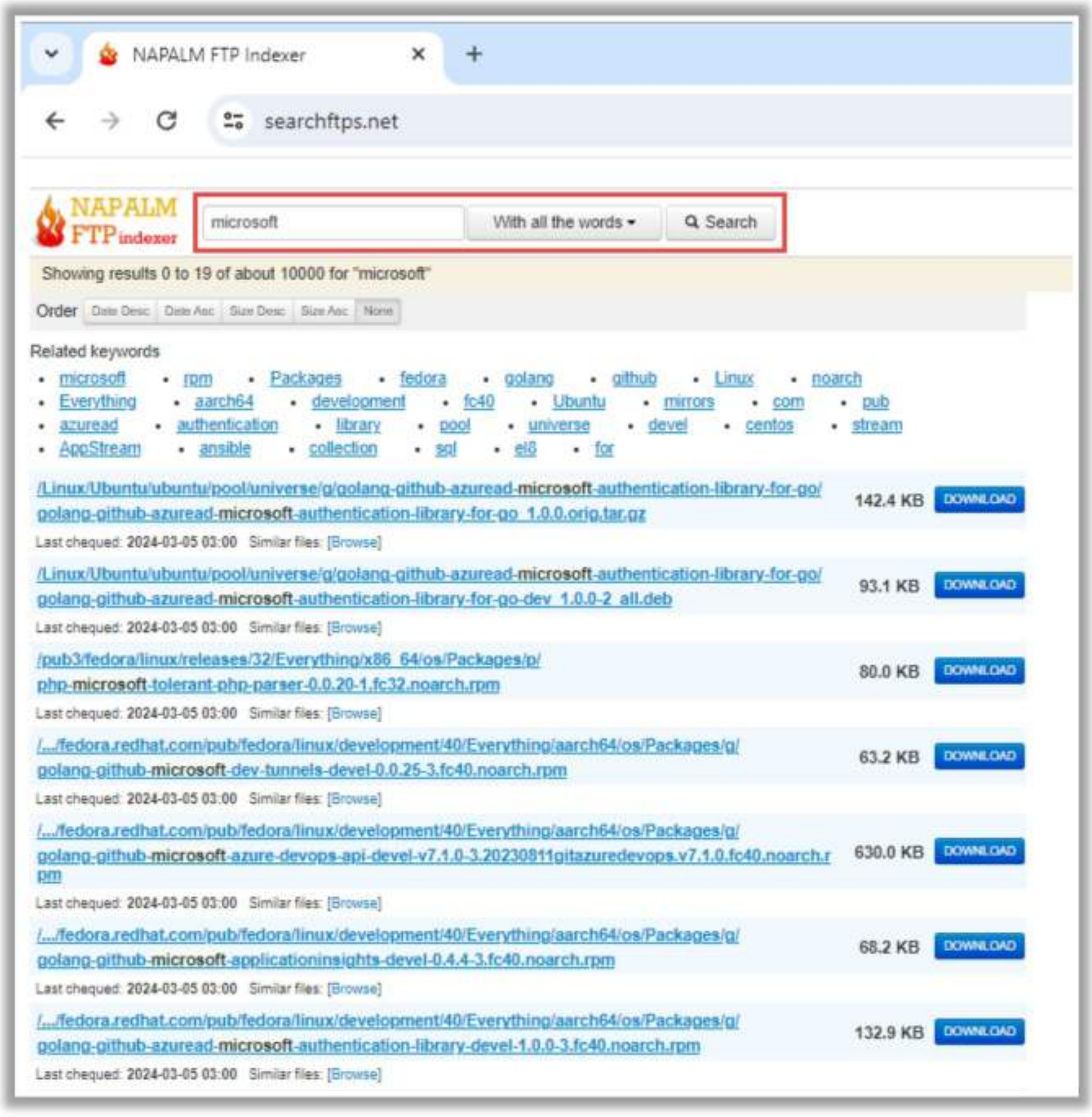


Figure 2.15: Screenshot of the FTP search engine NAPALM FTP Indexer showing search results for "microsoft"

- **Gathering Information from IoT Search Engines**

  Internet of Things (IoT) search engines crawl the Internet for IoT devices that are publicly accessible. Through a basic search on these search engines, an attacker can gain control of Supervisory Control and Data Acquisition (SCADA) systems, traffic control systems, Internet-connected household appliances, industrial appliances, CCTV cameras, etc. Many of these IoT devices are unsecured, i.e., they are without passwords or they use the default credentials, which can be exploited easily by attackers.

  With the help of IoT search engines such as Shodan, Censys, and ZoomEye, attackers can obtain information such as the manufacturer details, geographical location, IP address, hostname, and open ports of the target IoT device. Using this information, the attacker

can establish a back door to the IoT devices and gain access to them to launch further attacks.

As shown in the screenshot, attackers can use Shodan to find all the IoT devices of the target organization that are having open ports and services.

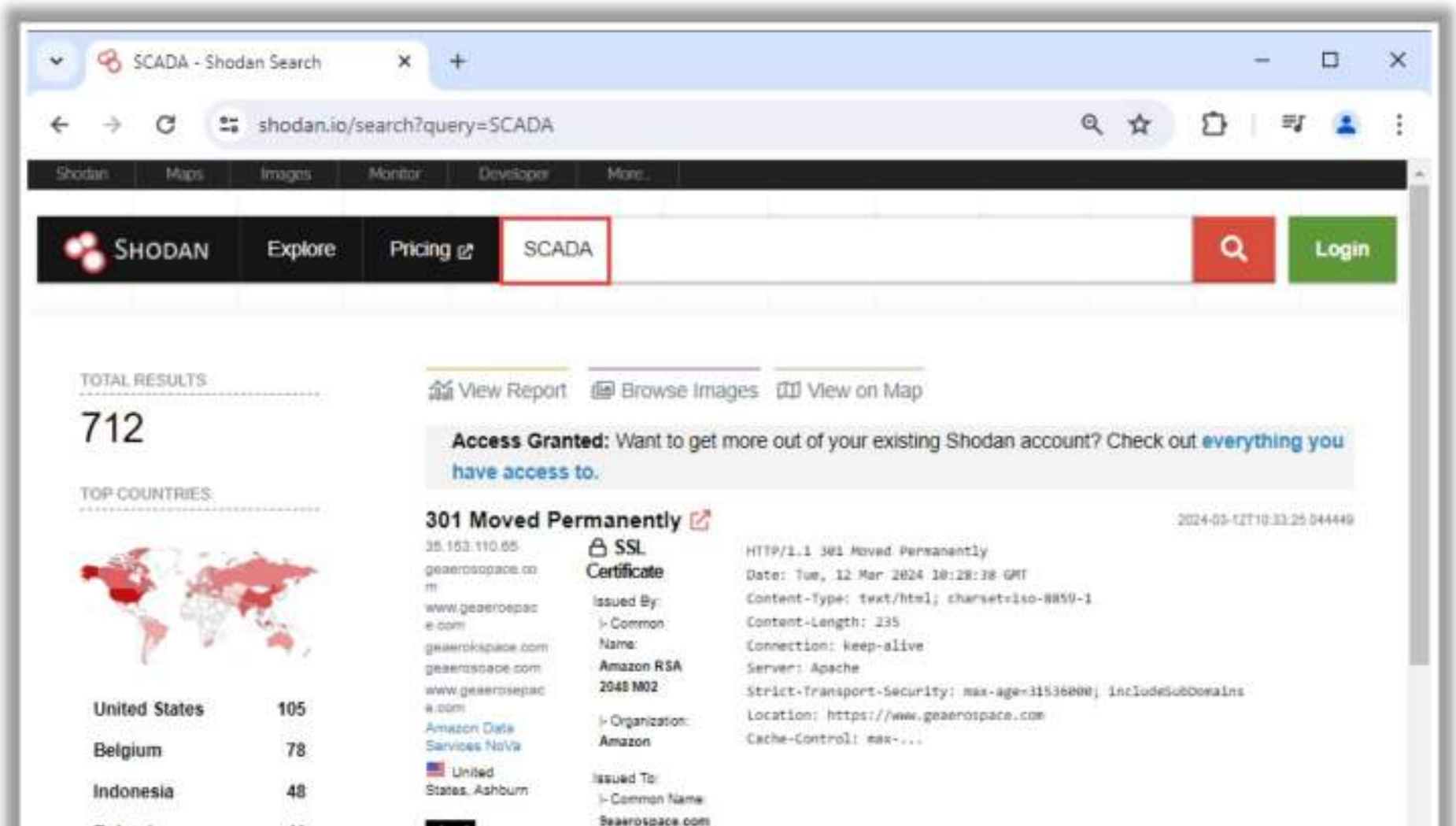


Figure 2.16: Screenshot of Shodan showing search results for SCADA devices

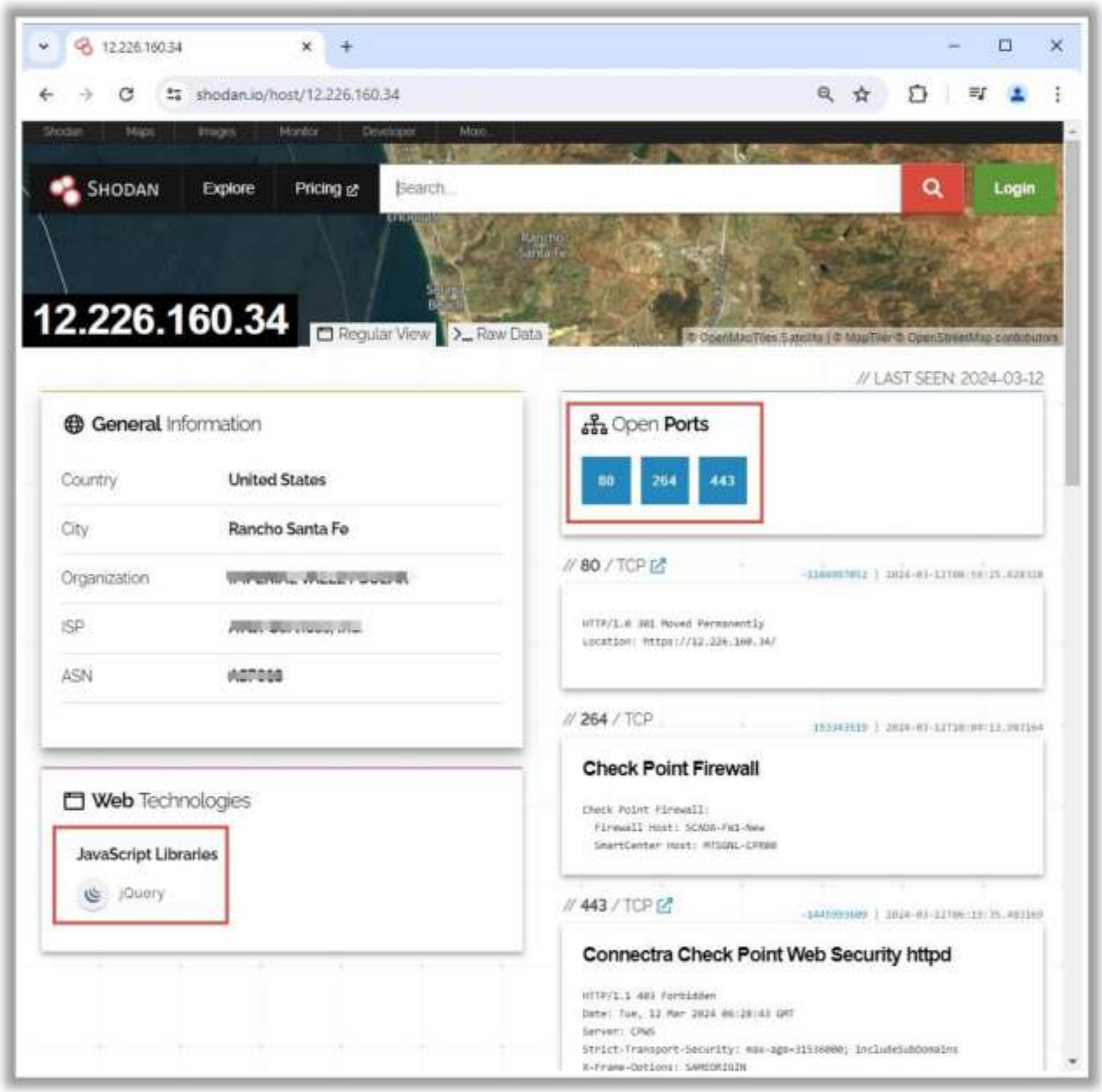
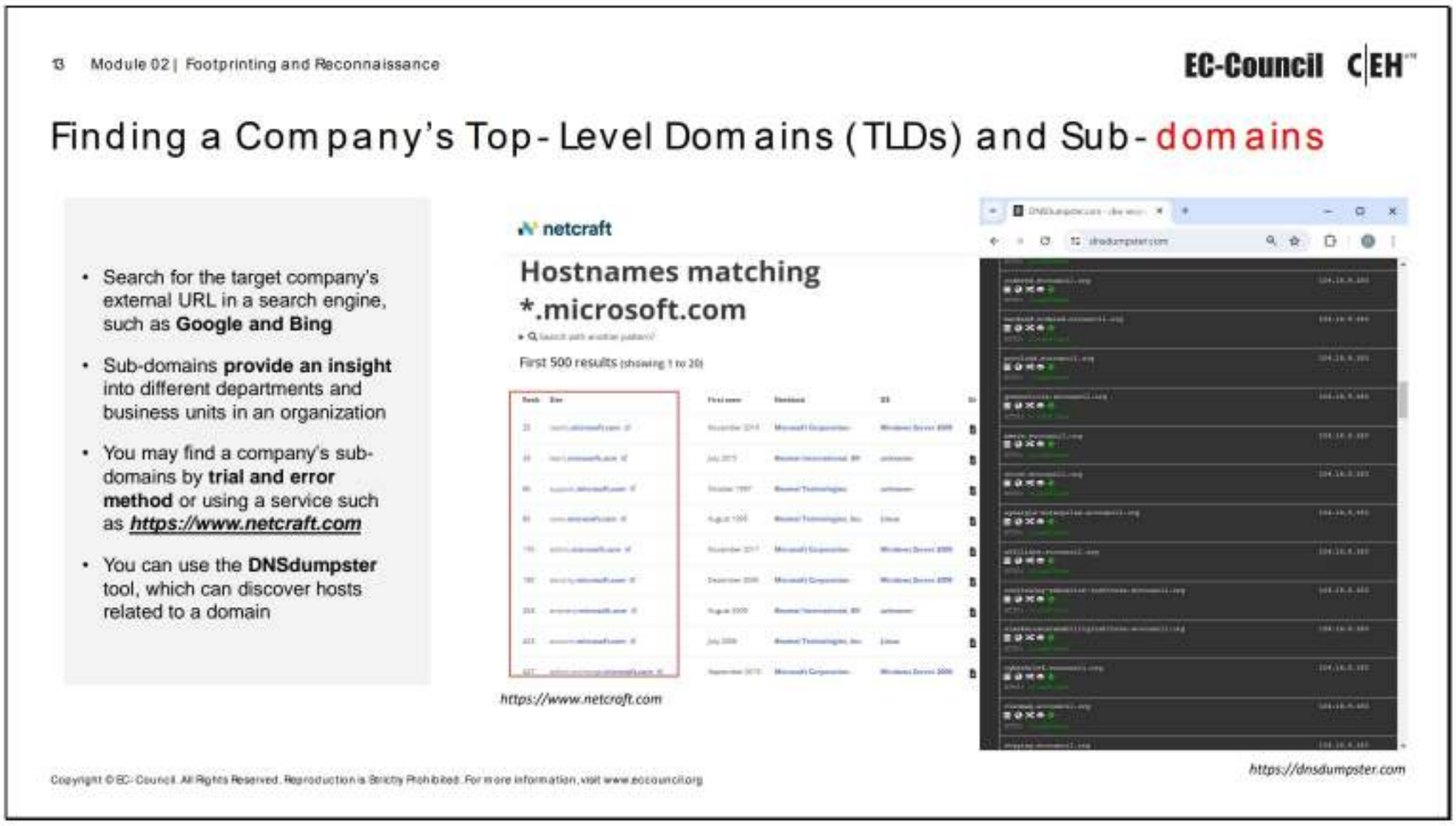


Figure 2.17: Screenshot of Shodan showing open ports and services of a SCADA system

Footprinting through Internet Research Services

Internet research services such as people search services can provide sensitive information about the target. People search services, alerting services, financial services, and job sites provide information about a target such as infrastructure details, physical location, and employee details. Using this information, an attacker may build a hacking strategy to break into the target organization's network and carry out other types of advanced system attacks.

This section aims to familiarize you with finding the target company's top-level domains, sub-domains, and geographical location, performing people search on people search services, gathering information from job sites, financial services, third-party data repositories, performing dark web footprinting, gathering competitive intelligence, etc.



Finding a Company's Top-Level Domains (TLDs) and Sub-domains

A company's top-level domains (TLDs) and sub-domains can provide a large amount of useful information to an attacker. A public website is designed to show the presence of an organization on the Internet. It is available for free public access. It is designed to attract customers and partners. It may contain information such as organizational history, services and products, and contact information. The target organization's external URL can be located with the help of search engines such as Google and Bing.

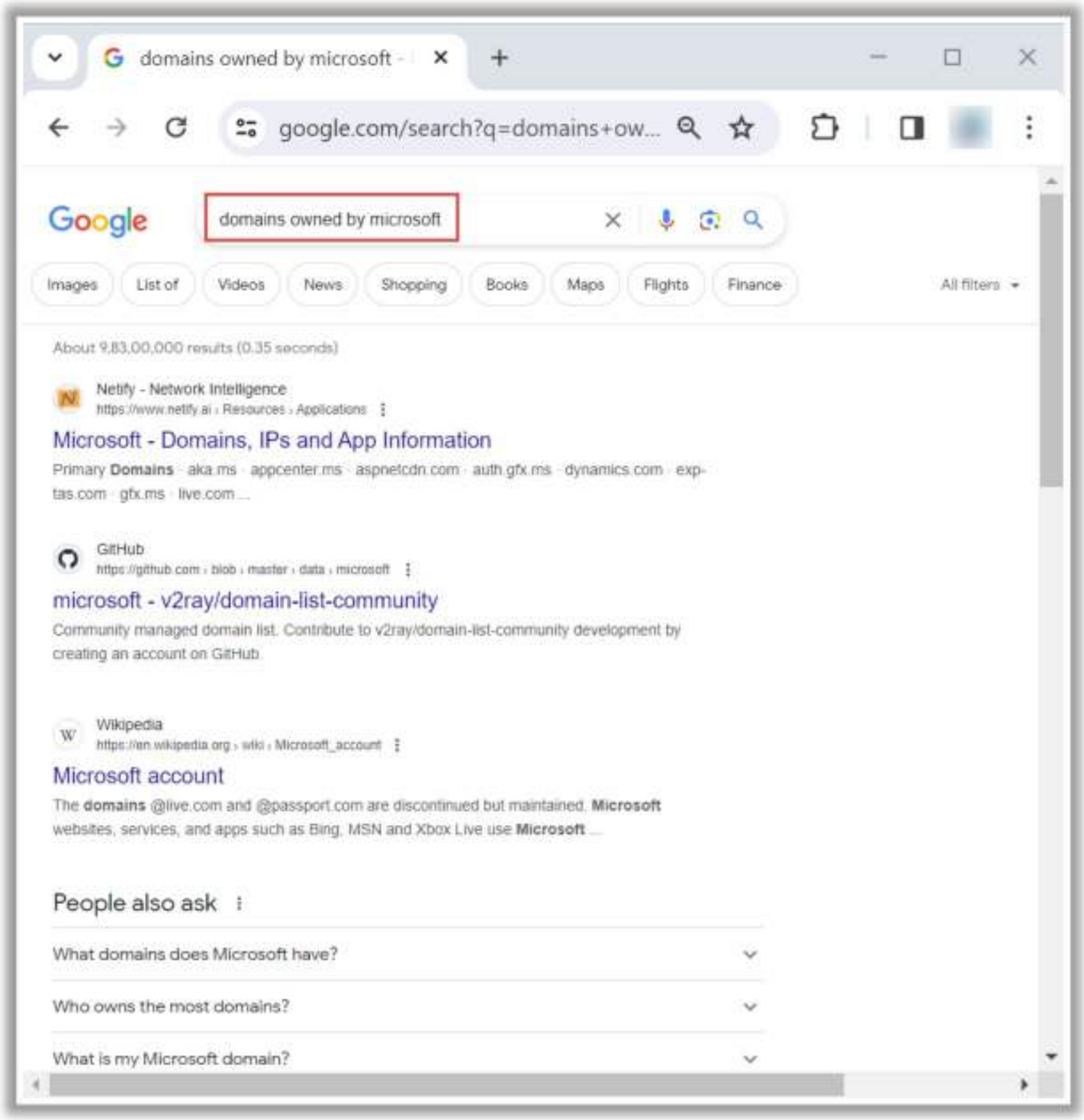


Figure 2.18: Google search engine showing search results for domains owned by Microsoft

The sub-domain is available to only a few people. These persons may be employees of an organization or members of a department. In many organizations, website administrators create sub-domains to test new technologies before deploying them on the main website. Generally, these sub-domains are in the testing stage and are insecure; hence, they are more vulnerable to various exploitations. Sub-domains provide insights into the different departments and business units in an organization. Identifying such sub-domains may reveal critical information regarding the target, such as the source code of the website and documents on the webserver. Access restrictions can be applied based on the IP address, domain or subnet, username, and password. The sub-domain helps to access the private functions of an organization. Most organizations use common formats for sub-domains. Therefore, a hacker who knows the external URL of a company can often discover the sub-domain through trial and error, or by using a service such as Netcraft.

You can also use the advanced Google search operator shown below to identify all the sub-domains of the target:

```
site:microsoft.com -inurl:www
```

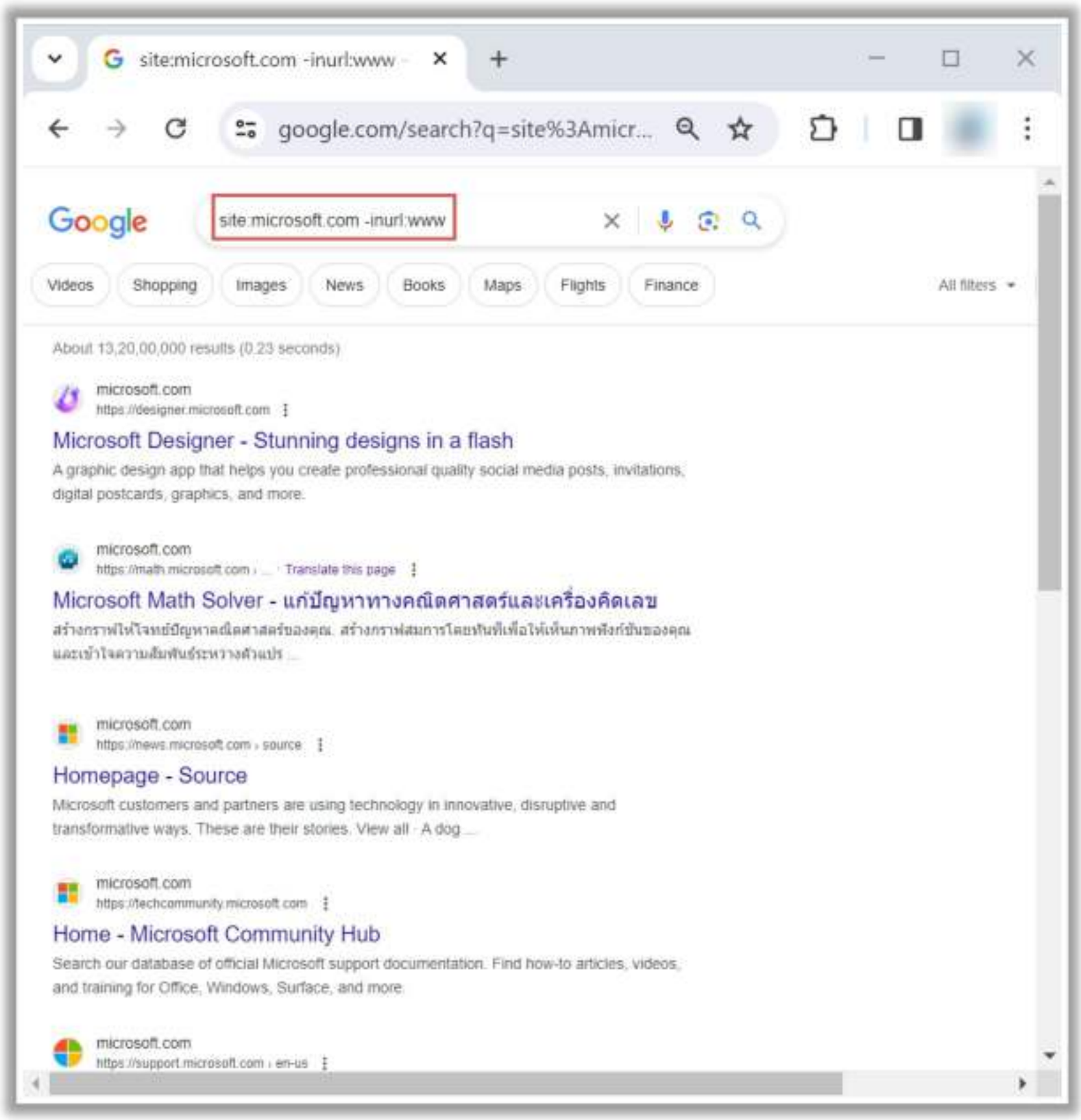


Figure 2.19: Finding sub-domains using Google Advanced Search Operator

Tools to Search Company's Sub-domains

- **Netcraft**

  Source: *https://www.netcraft.com*

  Netcraft provides Internet security services, including anti-fraud and anti-phishing services, application testing, and PCI scanning. They also analyze the market share of web servers, operating systems, hosting providers and SSL certificate authorities, and other parameters of the Internet.

  As shown in the screenshot below, attackers can use Netcraft to obtain all the sub-domains related to the target domain.

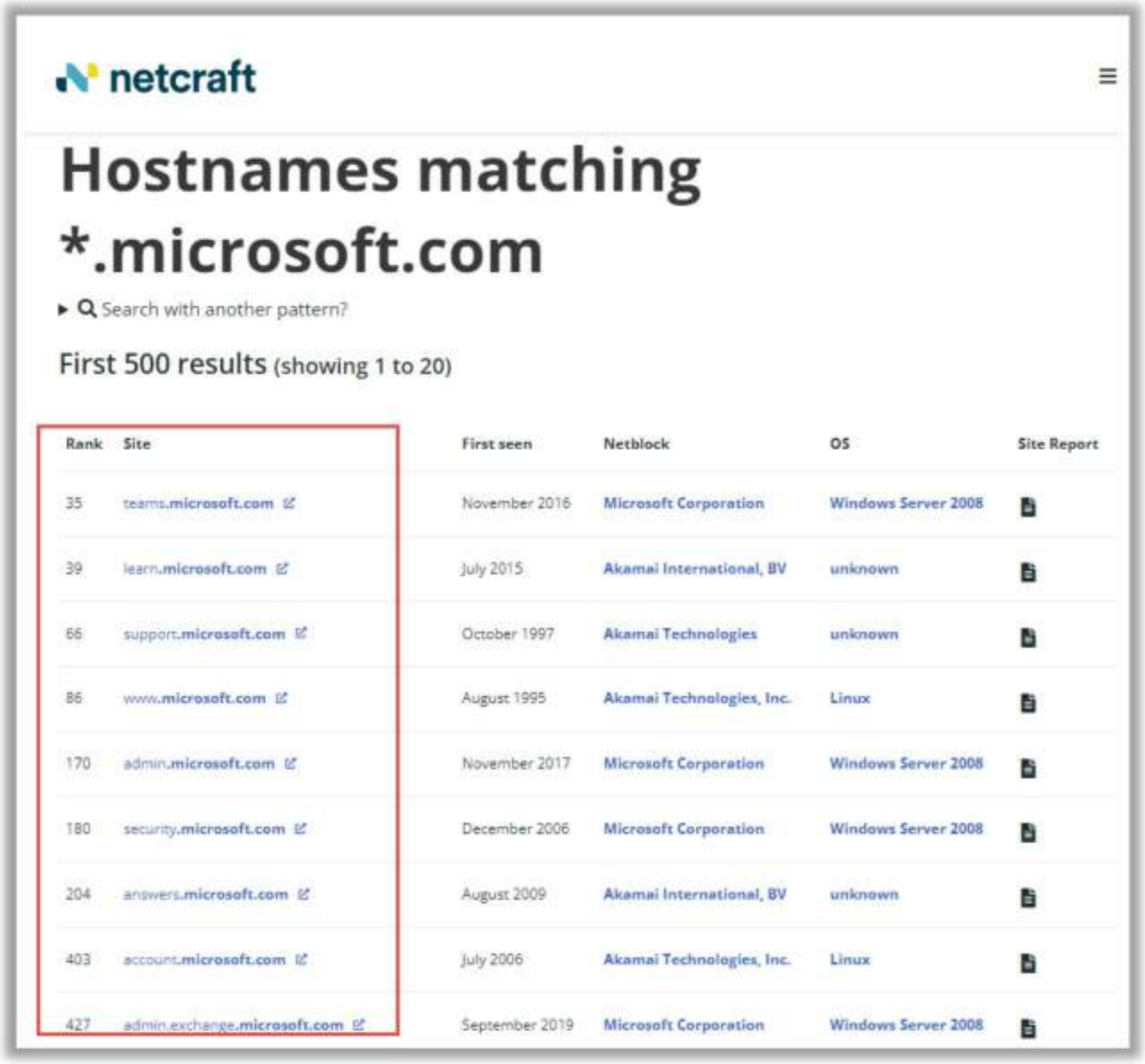


Figure 2.20: Screenshot of Netcraft displaying sub-domains of microsoft.com

- **DNSdumpster**

  Source: *https://dnsdumpster.com*

  DNSdumpster.com is a domain research tool that can be used by attackers to discover hosts related to a domain.

  As shown in the screenshot, attackers search for subdomains related to microsoft.com to obtain critical information about the target company domain, such as subdomains, IP addresses, DNS servers used, etc.

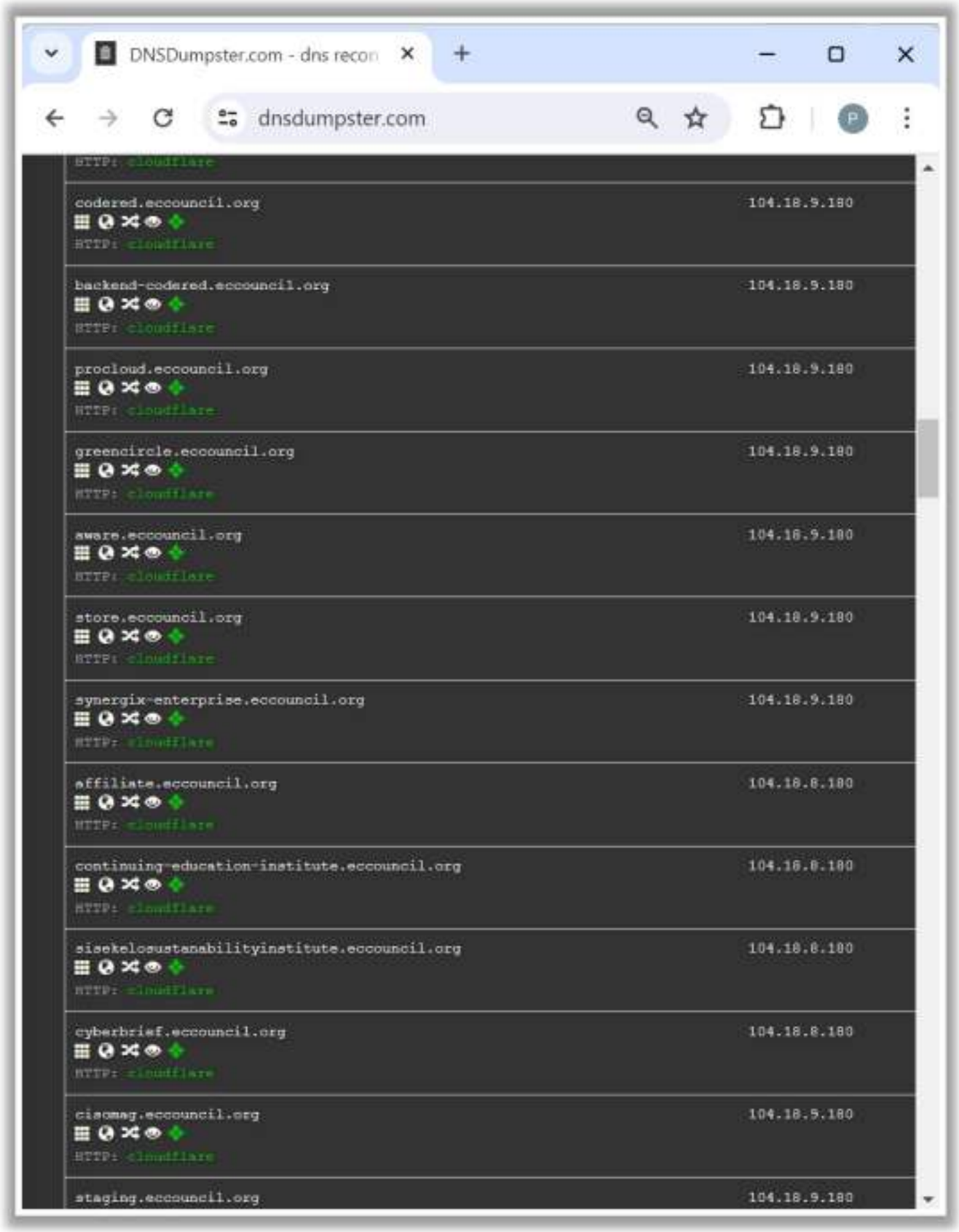


Figure 2.21: Screenshot of DNSdumpster tool displaying sub domains of eccouncil.org

- **Pentest-Tools Find Subdomains**

  Source: *https://pentest-tools.com*

  Pentest-Tools Find Subdomains is an online tool used for discovering subdomains and their IP addresses, including network information and their HTTP servers.

  As shown in the screenshot, attackers search for sub-domains related to microsoft.com to obtain critical information about the target company domain, such as sub-domains, IP addresses, operating systems, servers used, technology used, web platform, and page titles.

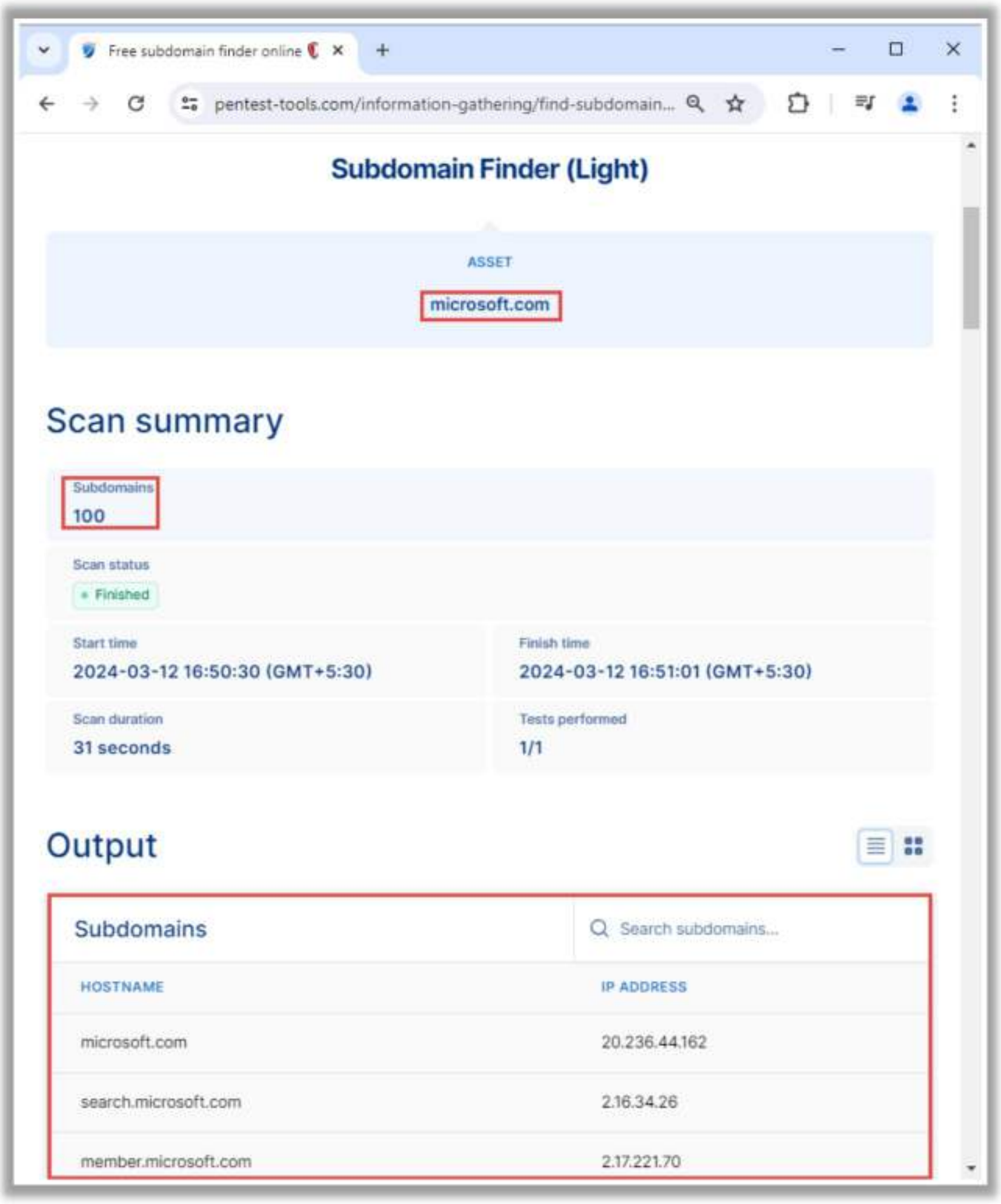
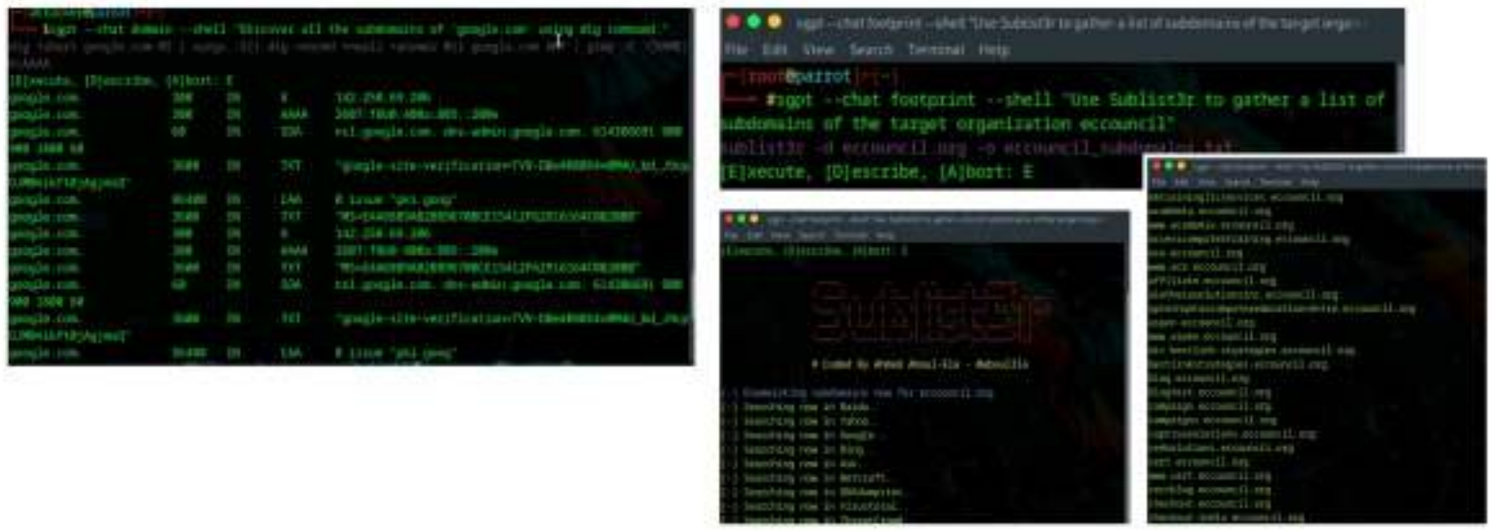


Figure 2.22: Screenshot of Pentest-Tools displaying sub-domains of microsoft.com

Finding a Company's Top-Level Domains (TLDs) and Sub-domains with AI

An attacker can also leverage **AI-powered ChatGPT** or other generative AI technology to perform this task by using appropriate prompts such as

1. ***"Discover all the subdomains of 'google.com' using dig command"***
2. ***"Use Sublist3r to gather a list of subdomains of the target organization eccouncil"***



Finding a Company's Top-Level Domains (TLDs) and Sub-domains with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly find the top-level domains and subdomains of the target.

Example #1

An attacker can use ChatGPT to perform this task by using an appropriate prompt such as:

"Discover all the subdomains of 'google.com' using dig command."

```
┌──[attacker@parrot]─[~]
└─$sgpt --chat domain --shell "Discover all the subdomains of 'google.com' using dig command."
dig +short google.com NS | xargs -I{} dig +nocmd +noall +answer @{} google.com ANY | grep -E 'CNAME|
A|AAAA'
[E]xecute, [D]escribe, [A]bort: E
google.com.             300     IN      A       142.250.69.206
google.com.             300     IN      AAAA    2607:f8b0:400a:805::200e
google.com.             60      IN      SOA     ns1.google.com. dns-admin.google.com. 614386691 900
900 1800 60
google.com.             3600    IN      TXT     "google-site-verification=TV9-DBe4R80X4v0M4U_bd_J9cp
OJM0nikft0jAgjmsQ"
google.com.             86400   IN      CAA     0 issue "pki.goog"
google.com.             3600    IN      TXT     "MS=E4A68B9AB2BB9670BCE15412F62916164C0B20BB"
google.com.             300     IN      A       142.250.69.206
google.com.             300     IN      AAAA    2607:f8b0:400a:805::200e
google.com.             3600    IN      TXT     "MS=E4A68B9AB2BB9670BCE15412F62916164C0B20BB"
google.com.             60      IN      SOA     ns1.google.com. dns-admin.google.com. 614386691 900
900 1800 60
google.com.             3600    IN      TXT     "google-site-verification=TV9-DBe4R80X4v0M4U_bd_J9cp
OJM0nikft0jAgjmsQ"
google.com.             86400   IN      CAA     0 issue "pki.goog"
```

Figure 2.23: Discover all the subdomains of 'google.com' using dig command

This command instructs **dig** to query the authoritative name servers (NS records) for "google.com" and display the response. The output will list all the name servers associated with "google.com", which often include the subdomains as well.

```
dig +short google.com NS | xargs I() dig +nocmd +noall +answer @() google.com
| grep -E 'CNAME A|AAAA
```

- `` `dig +short google.com NS` ``: This part retrieves the name server (NS) records for "google.com" in a shortened format.
- `` `| xargs -I{}` ``: This part takes each line of the output (which represents a name server) and passes it as an argument to the next `dig` command.
- `` `dig +nocmd +noall +answer @{} google.com A +noall +answer` ``: This part performs a DNS lookup for the A records of "google.com" using each name server obtained from the previous command. It retrieves the IP addresses associated with the subdomains of "google.com".

Example #2

An attacker can use ChatGPT to perform this task by using an appropriate prompt such as:

"Use Sublist3r to gather a list of subdomains of the target organization eccouncil"

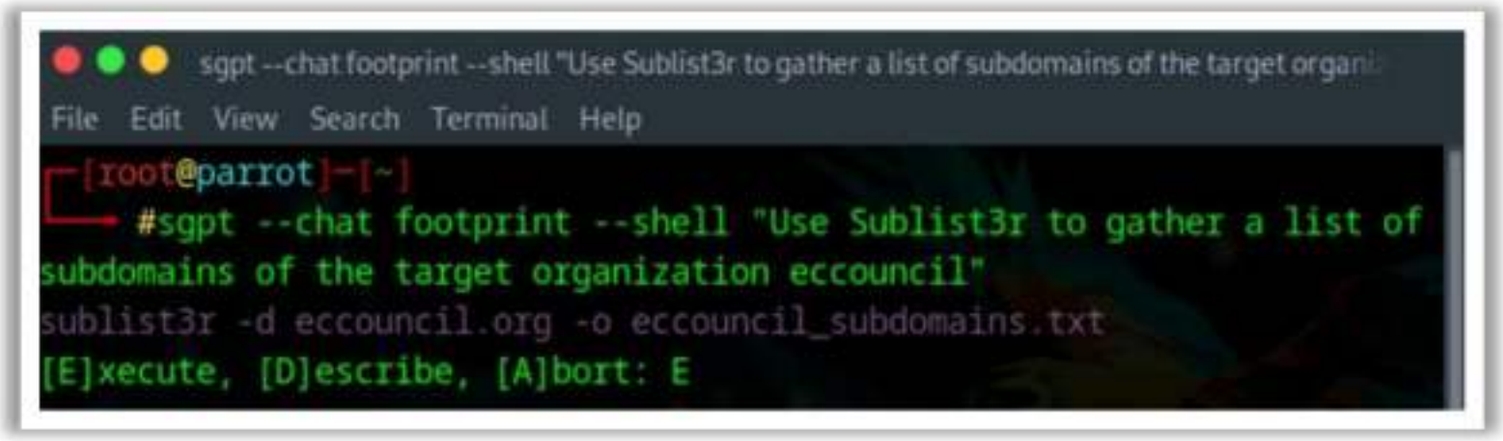


Figure 2.24: Use Sublist3r to gather a list of subdomains of the target organization eccouncil

The command sublist3r -d eccouncil.org -o eccouncil_subdomains.txt is used to execute the Sublist3r tool with specific parameters

```
sublist3r -d eccouncil.org -o eccouncil_subdomains.txt
```

- `` `sublist3r` ``: This is the command to invoke the Sublist3r tool.
- `` `-d eccouncil.org` ``: This parameter specifies the target domain, in this case, "eccouncil.org", for which we want to enumerate subdomains.
- `` `-o eccouncil_subdomains.txt` ``: This parameter specifies the output file where the enumerated subdomains will be saved. In this case, the file is named "eccouncil_subdomains.txt".

Figure 2.25: Subdomains associated with target

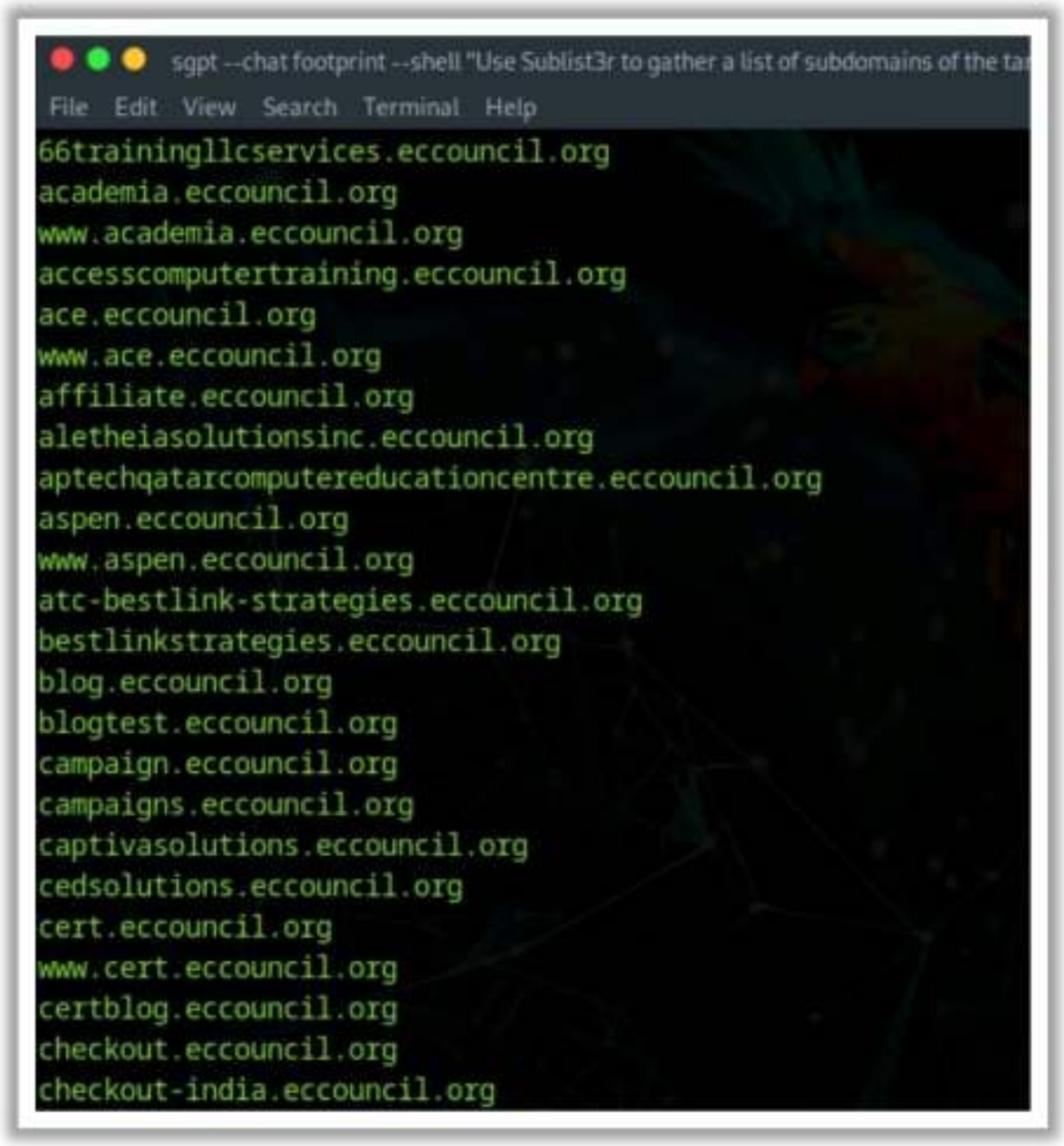
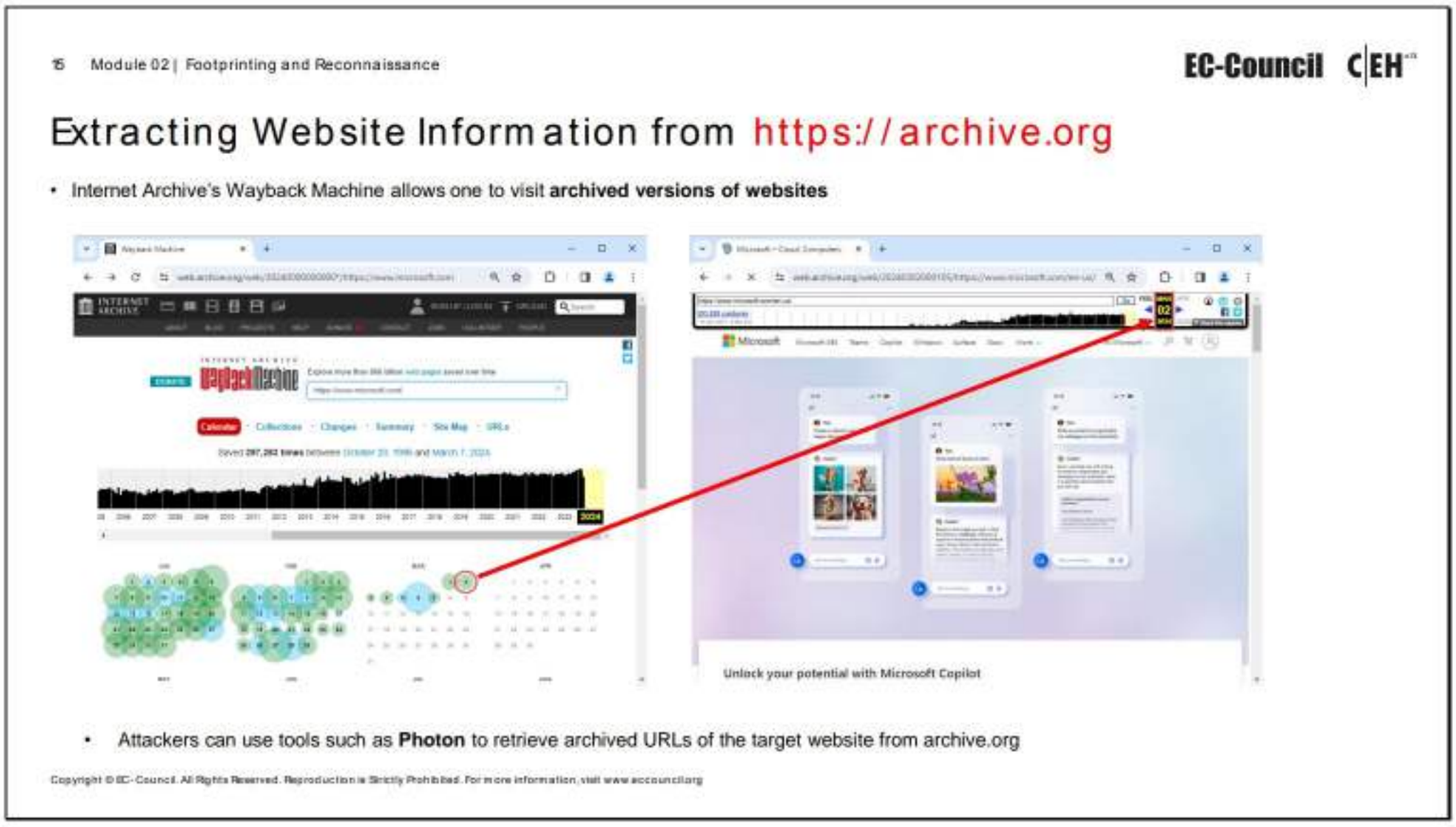


Figure 2.26: Subdomains associated with target

Overall, this command instructs Sublist3r to search for subdomains associated with the "eccouncil.org" domain and save the results to a text file named "eccouncil_subdomains.txt".



Extracting Website Information from https://archive.org

Source: *https://archive.org*

Archive is an Internet Archive Wayback Machine that explores archived versions of websites. Such exploration allows an attacker to gather information on an organization's web pages since its creation. As the website ***https://archive.org*** keeps track of web pages from the time of their creation, an attacker can retrieve even information removed from the target website, such as web pages, audio files, video files, images, text, and software programs. Attackers use this information to perform phishing and other types of web application attacks on the target organization.

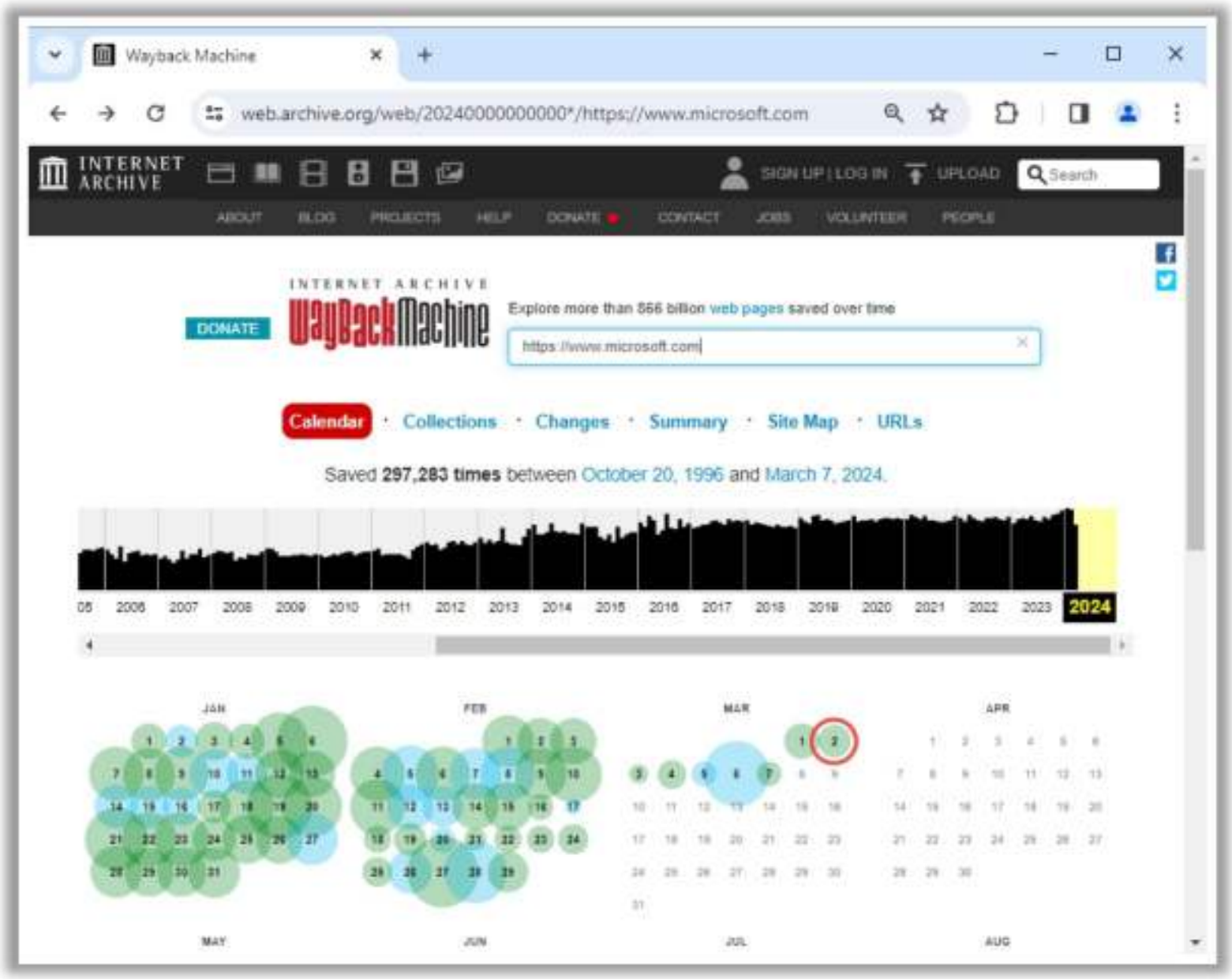


Figure 2.27: Screenshot of Archive showing archived versions of microsoft.com

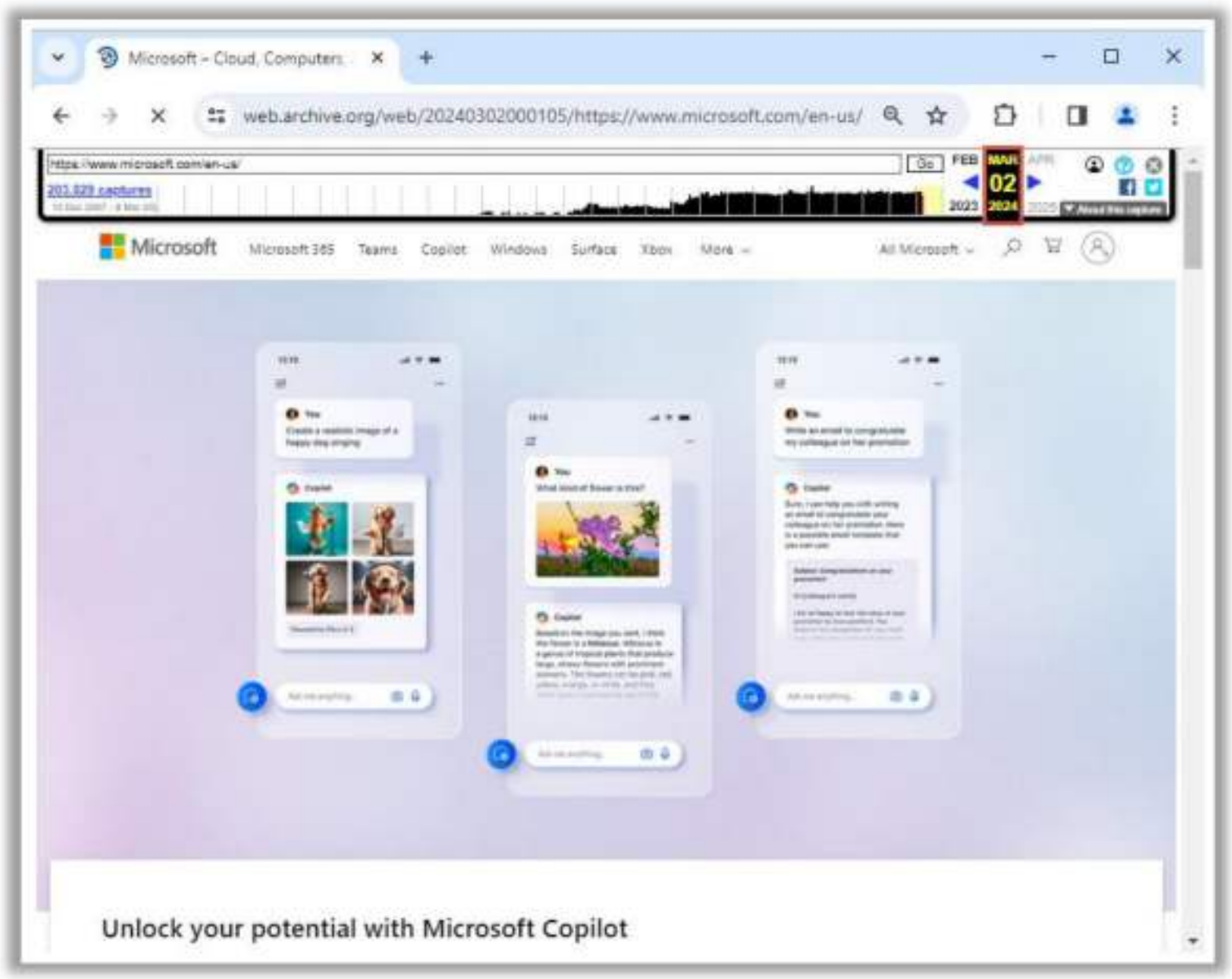


Figure 2.28: Screenshot of Archive showing archived web pages of microsoft.com

Attackers can use tools such as Photon to retrieve archived URLs of the target website from archive.org. Run the following command to retrieve the archive.org links of the target website:

photon.py -u <URL of the Target Website> -l 3 -t 200 --wayback

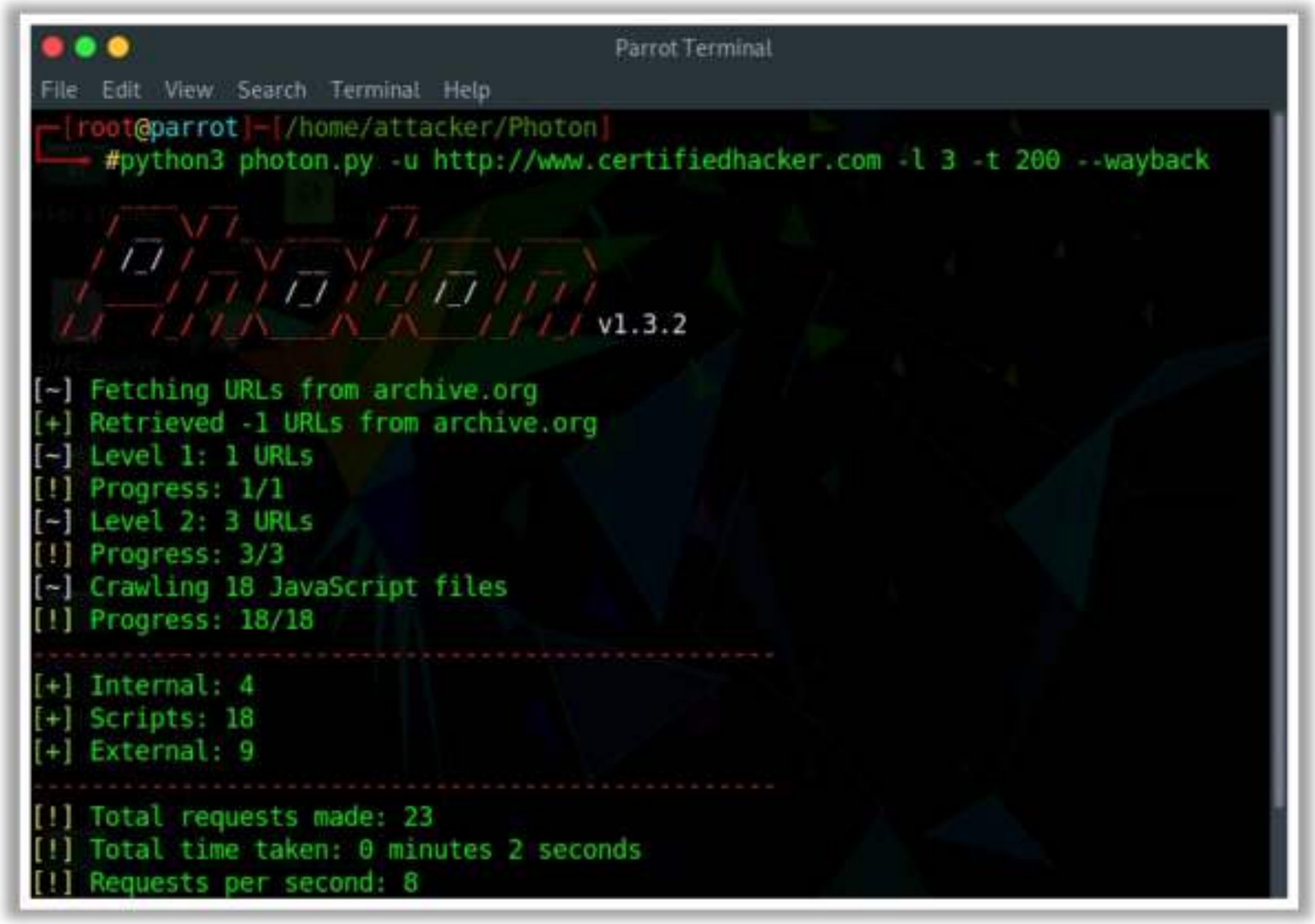


Figure 2.29: Screenshot of Photon showing the output for the command to retrieve archive.org links

Run the following command to retrieve archived URLs of the target website:

python photon.py -u <URL of the Target Website> -l 3 -t 200 -only-urls

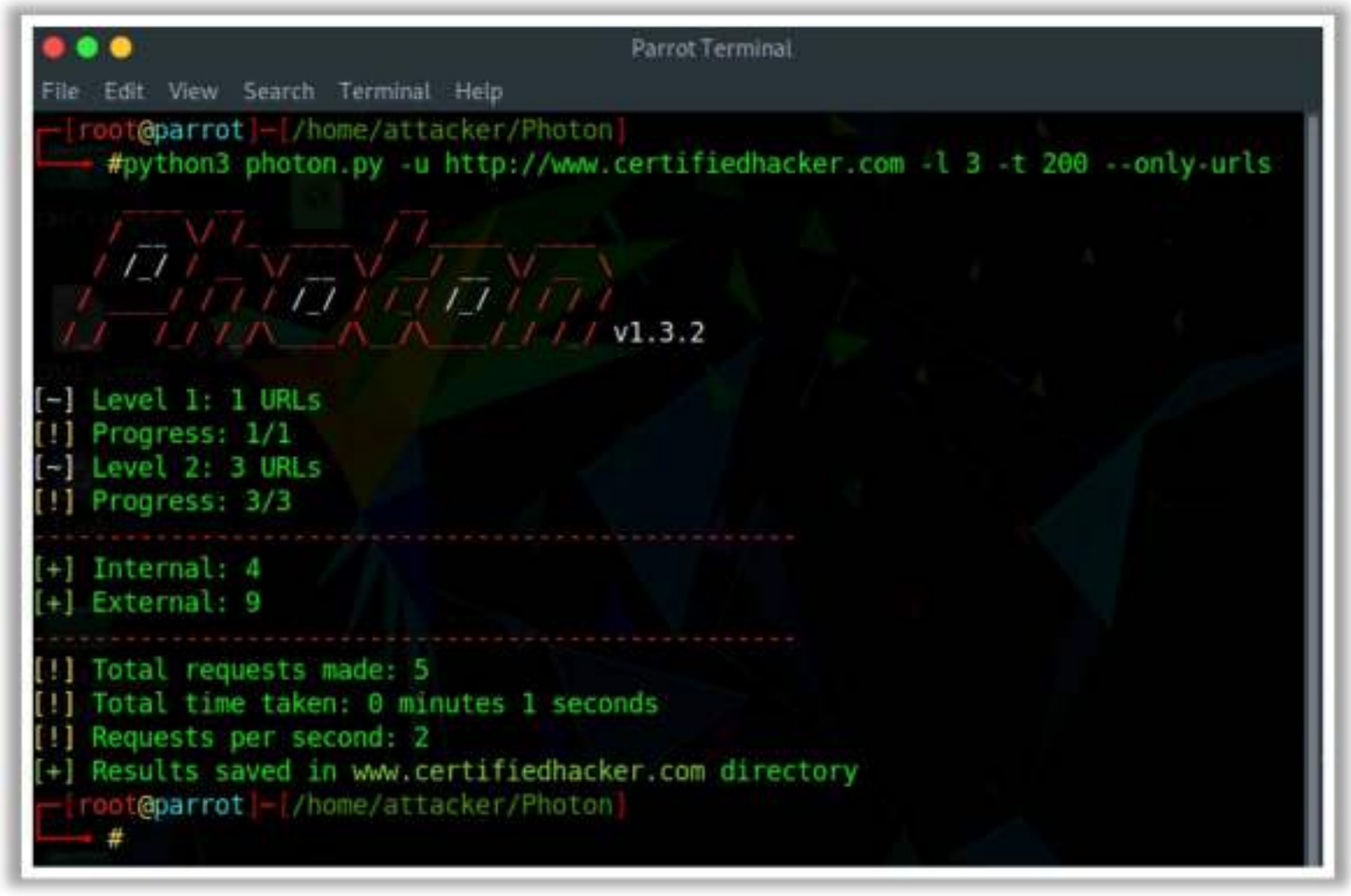
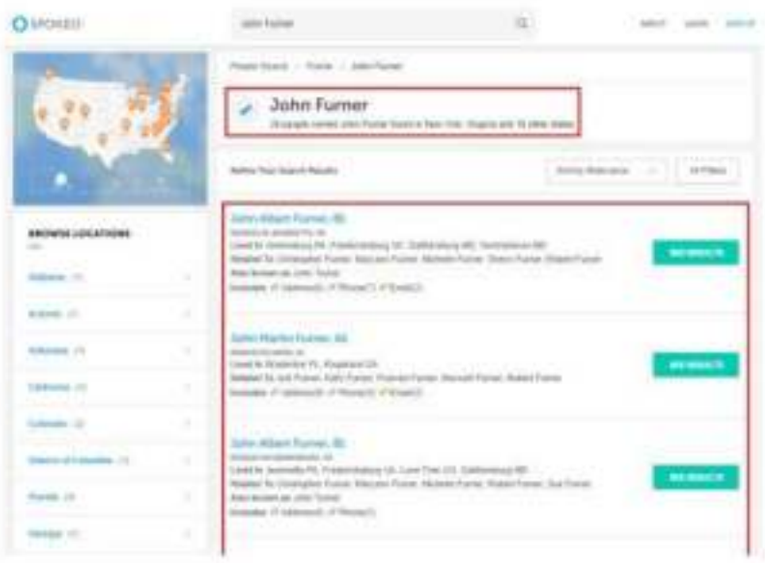


Figure 2.30: Screenshot of Photon showing the output for the command to retrieve archived URLs

EC-Council CEH

Footprinting through People Search Services and Job Sites

- People search services, such as **Spokeo**, **Intelius**, and **pipl** can provide critical **information about a person or an organization**, including location, emails, websites, blogs, contacts, important dates, address, etc.



https://www.spokeo.com

- Job sites such as Dice, LinkedIn, and Glassdoor can reveal details about a **company's infrastructure**, potentially aiding attackers in **identifying vulnerabilities within the target's IT environment**

https://www.glassdoor.com

Footprinting through People Search Services

You can use public record websites to find information about email addresses, phone numbers, house addresses, and other information. Many individuals use online people search services to find information about other people. Generally, online people search services such as Spokeo, Intelius, pipl, BeenVerified, Whitepages, Instant Checkmate, and PeekYou provide people's names, addresses, contact details, date of birth, photographs, videos, profession, details about their family and friends, social networking profiles, property information, and optional background on criminal checks**.** Further, online people search services may often reveal the profession of an individual, businesses owned by a person, upcoming projects and operating environment, websites and blogs, contact numbers, important dates, company email addresses, cell phone numbers, fax numbers, and personal e-mail addresses. Using this information, an attacker can try to obtain bank details, credit card details, past history, and so on. This information proves to be highly beneficial for attackers to launch attacks. There are many available online people search services that help in obtaining information regarding people.

- **People Search Service - Spokeo**

  Source: *https://www.spokeo.com*

  Attackers can use the Spokeo people search online service to search for people belonging to the target organization. Using this service, attackers obtain information such as phone numbers, email addresses, address history, age, date of birth, family members, social profiles, and court records.

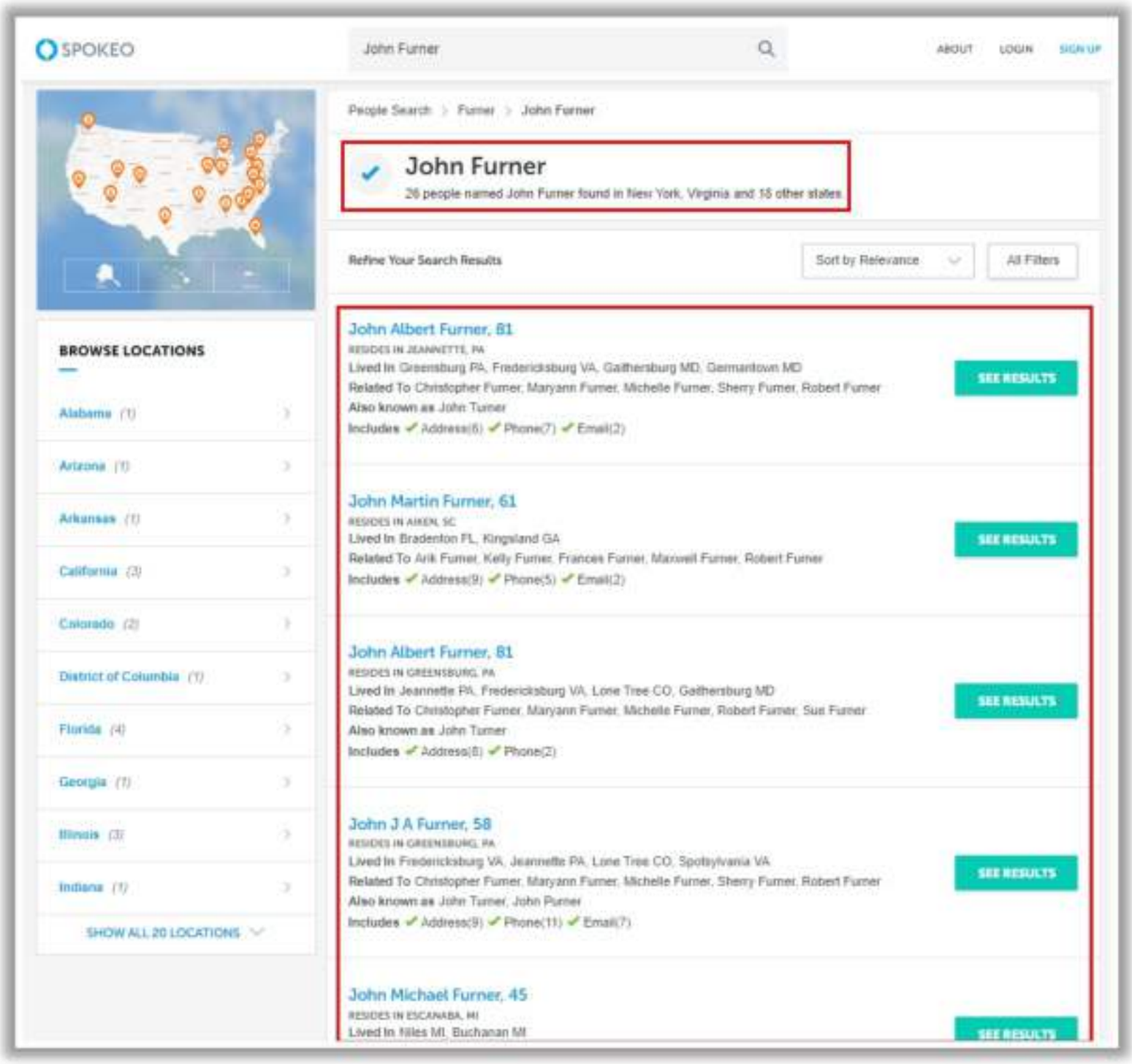


Figure 2.31: Screenshot of Spokeo people Search service

Footprinting through Job Sites

Attackers can gather valuable information about the operating system, software versions, company's infrastructure details, and database schema of an organization through footprinting job sites using different techniques. Many organizations' websites provide recruiting information on a job posting page that, in turn, reveals hardware and software information, network-related information, and technologies used by the company (e.g., firewall, internal server type, OS used, network appliances, hypervisors, VMs, and so on.). In addition, the website may have a key employee list with email addresses. Such information may prove to be beneficial for an attacker. For example, if an organization advertises a Network Administrator job, it posts the requirements related to that position.

Further, attackers can go through employee resumes posted on job sites and extract information such as an individual's expertise, educational qualifications, and job history. The job history of an employee can reveal technical information about the target organization. Attackers can use the technical information obtained through job sites such as Dice, LinkedIn, Glassdoor, and Simply Hired to detect underlying vulnerabilities in the target IT infrastructure.

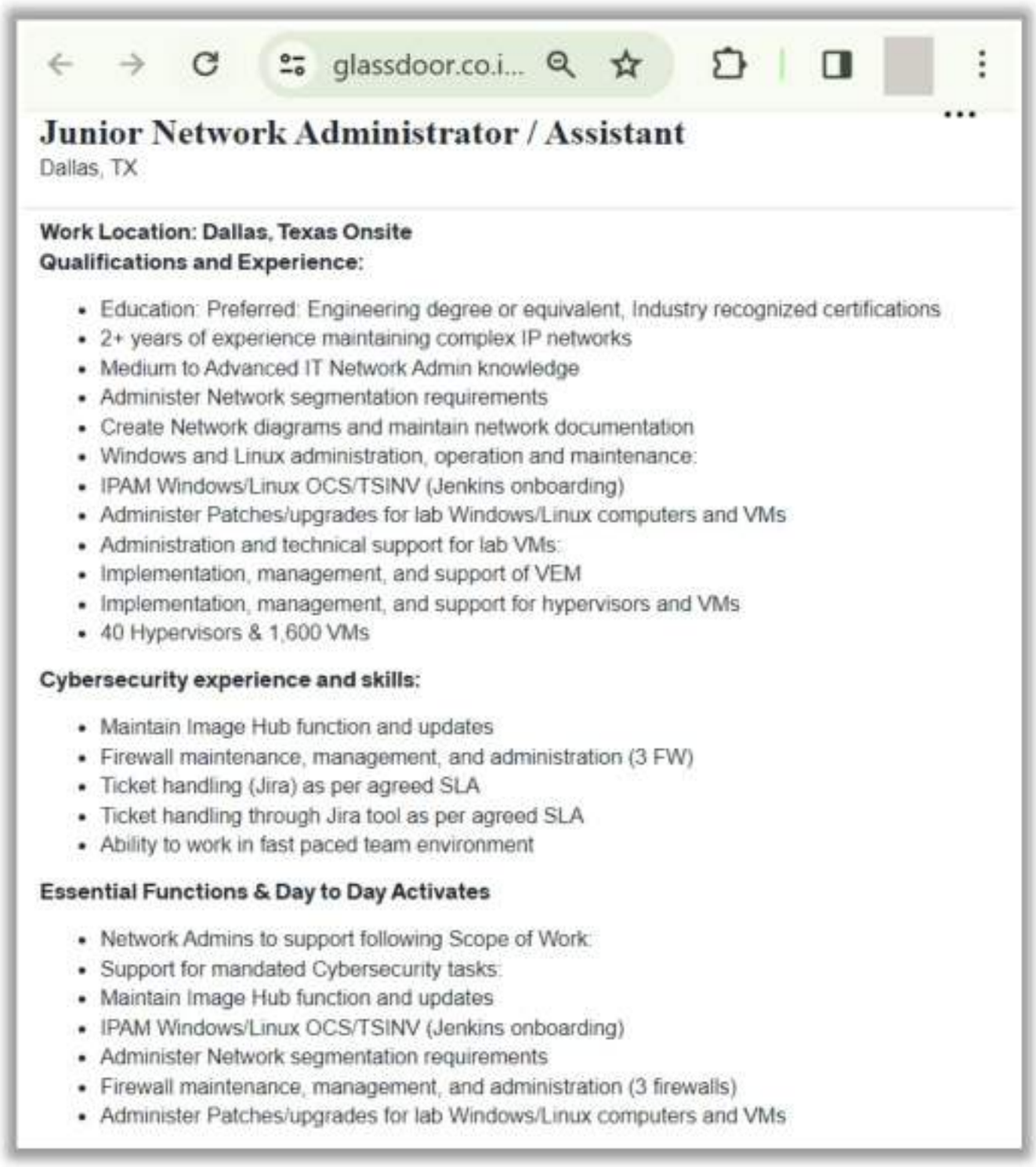
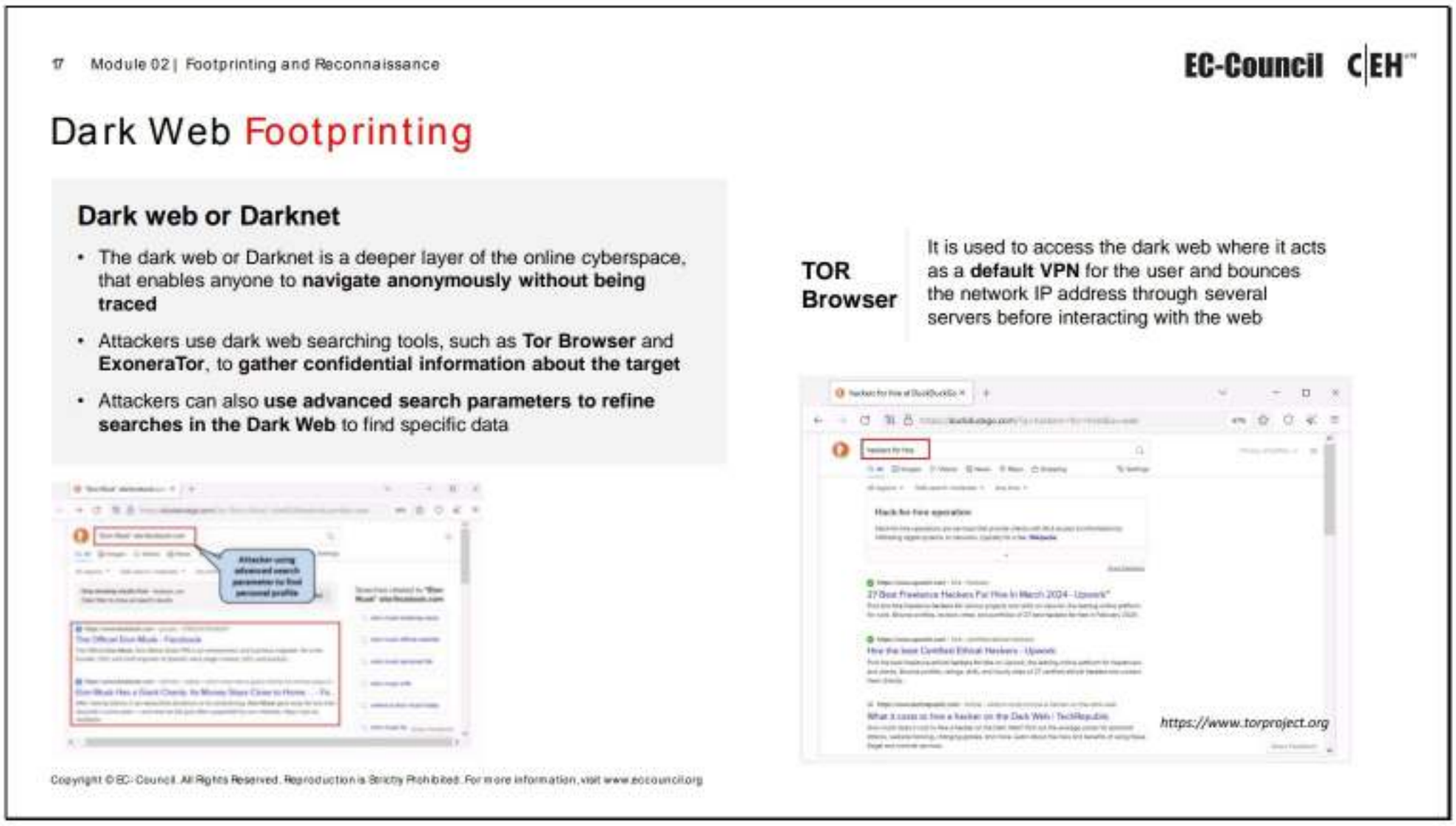


Figure 2.32: Screenshot of job posting showing valuable information



Dark Web Footprinting

The surface web is the outer layer of the online cyberspace that allows the user to find web pages and content using regular web browsers. Search engines use crawlers that are programmed bots to access and download web pages. The surface web can be accessed by browsers such as Google Chrome, Mozilla Firefox, and Opera.

The deep web is the layer of the online cyberspace that consists of web pages and content that are hidden and unindexed. Such content cannot be located using traditional web browsers and search engines. The size of the deep web is incalculable, and it expands to almost the entire World Wide Web. The deep web does not allow the crawling process of basic search engines. It consists of official government or federal databases and other information linked to various organizations. The deep web can be accessed using search engines such as Tor Browser and the WWW Virtual Library. It can be used for both legal and illegal activities.

The dark web or Darknet is a deeper layer of the online cyberspace, and it is the subset of the deep web that enables anyone to navigate anonymously without being traced. The dark web can be accessed only through specialized tools or darknet browsers. Attackers primarily use the dark web to perform footprinting on the target organization and launch attacks. The dark web can be accessed using search engines such as Tor Browser and ExoneraTor.

Attackers can use dark web searching tools such as Tor Browser, ExoneraTor, and OnionLand Search engine to gather confidential information about the target, such as credit card details, passports information, identification card details, medical records, social media accounts, and Social Security Numbers (SSNs). With the help of this information, they can launch further attacks on the targets.

- **Tor Browser**

  Source: *https://www.torproject.org*

  Tor Browser is used to access the dark web, where it acts as a default VPN for the user and bounces the network IP address through several servers before interacting with the web. Attackers use this browser to access hidden content, unindexed websites, and encrypted databases present in the dark web.

  As shown in the screenshot, by using Tor Browser, attackers can obtain more detailed and hidden information about the target organization.

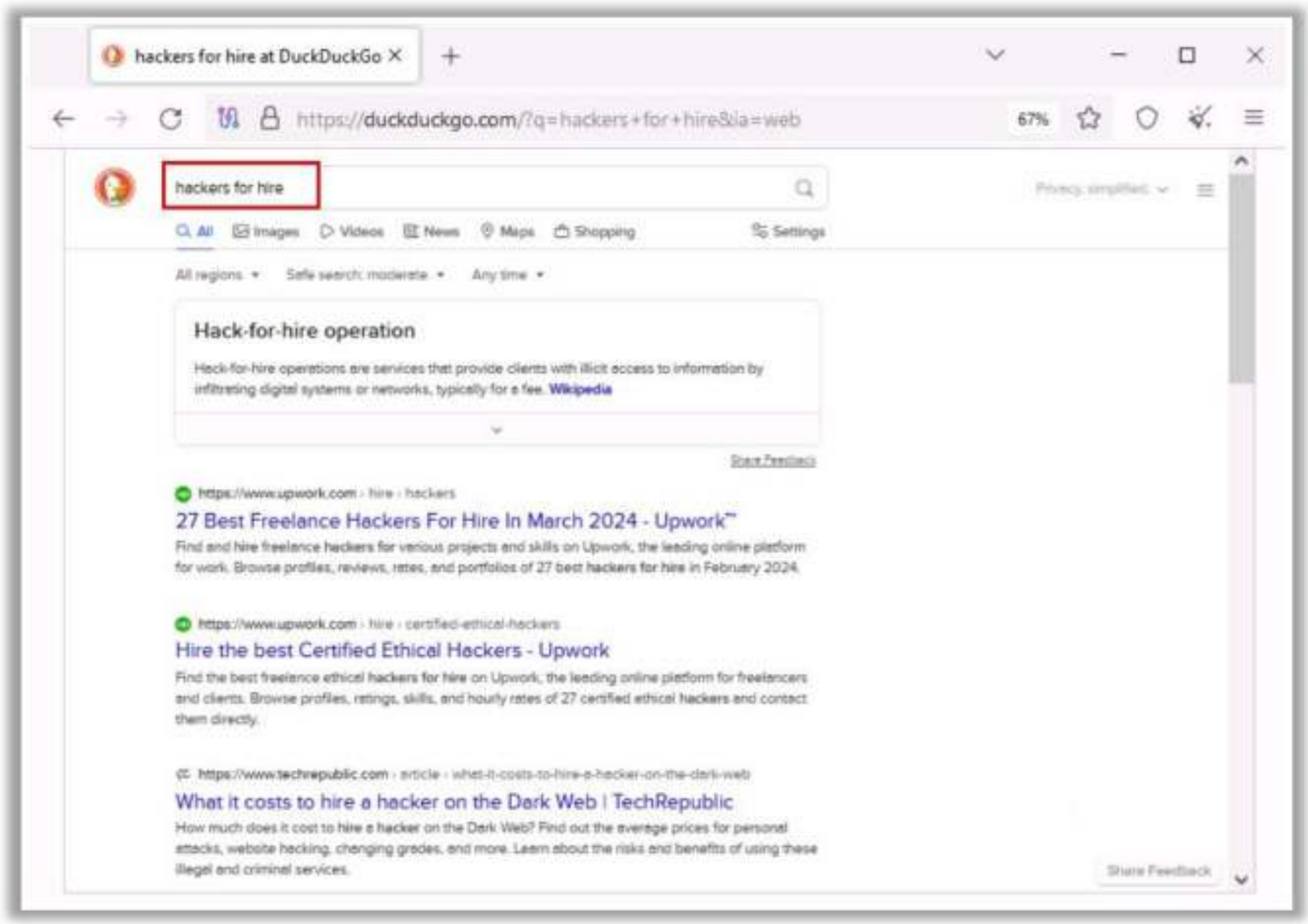


Figure 2.33: Screenshot of Tor Browser

Searching the Dark Web with Advanced Search Parameters

Attackers can also use advanced search parameters to refine searches in the Deep Web to find specific data by using Dark Web searching tools such as Tor Browser. They can find crucial information using advanced searching techniques such as filtering by file type or setting domain restrictions.

For instance, attackers can use advanced search parameters to find sensitive documents such as financial records or login credentials. These advanced searching techniques enable attackers to quickly find and gain access to valuable information and use it to perform malicious activities.

Attackers can refine their searches and focus on the specific data they are searching for by using the following parameters:

- **Personal profiles:** Search for information related to the victim's personal profiles such as profiles of social media or personal websites.

  For example, **`"John Doe" site:facebook.com OR site:linkedin.com`**

- **Scientific publications:** Search for publications on specific publications such as academic or scientific research papers, and articles.

  For example, **`"John Doe" site:scholar.google.com`**

- **Court records:** Search for legal documents related to court records or cases.

  For example, **`"John Doe" court records`**

- **Member directories:** Search for directories of members employed to the organizations.

  For example, **`"John Doe" site:example.com "employee directory"`**

- **Medical records:** Search for medical information or health history of the victim.

  For example, **`"John Doe" medical records`**

- **Location records:** Search for location information such as location history or GPS information of the victim.

  For example, **`"John Doe" location history`**

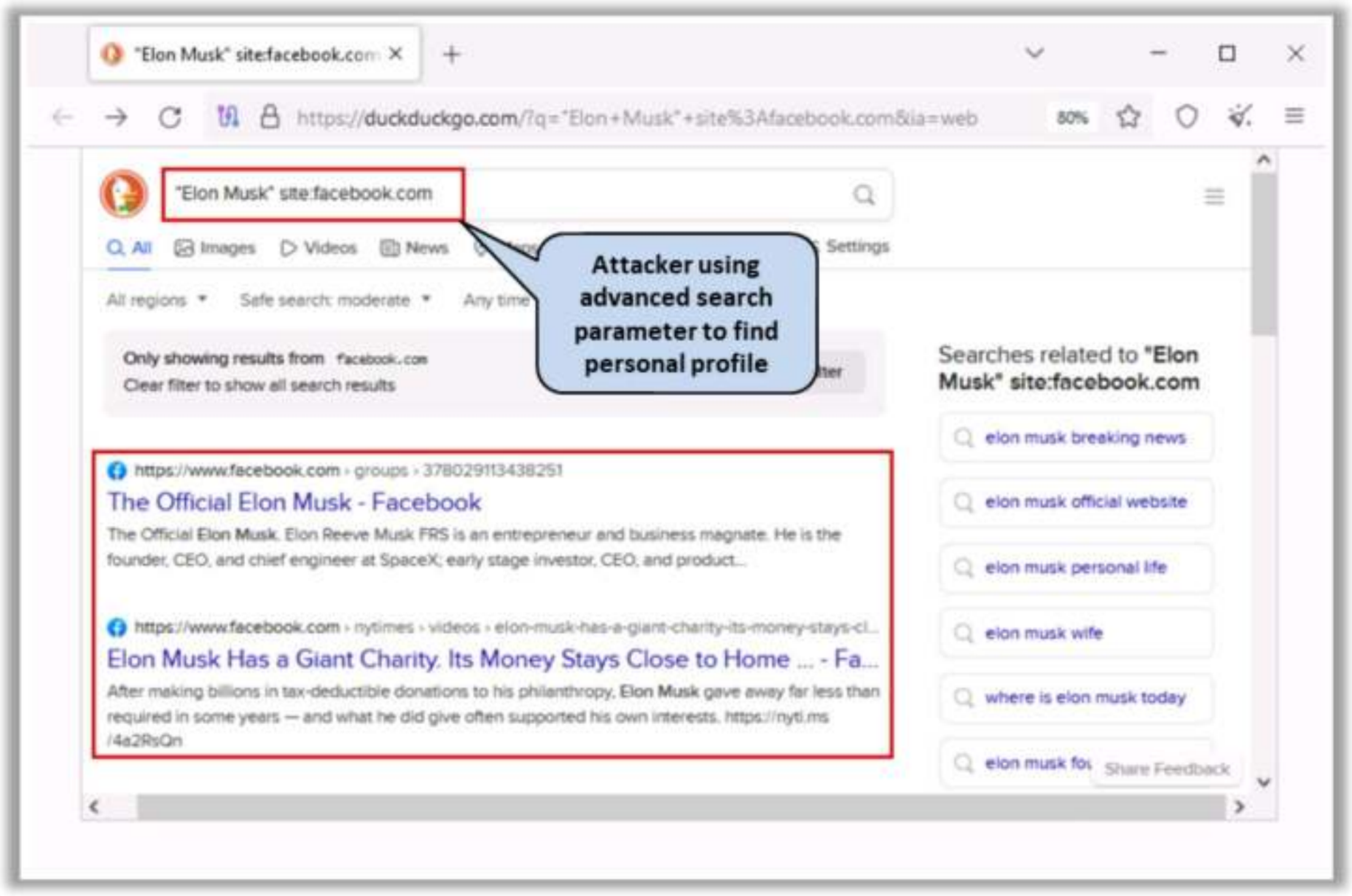


Figure 2.34: Screenshot of Tor Browser showing advanced search parameter results

Here's a table of advanced search queries designed for performing footprinting on the dark web to find sensitive information using the Tor Browser. These queries leverage specific search parameters and operators suitable for dark web search engines.

Type of Information	Search Query	Explanation
Sensitive PDFs	filetype:pdf site:onion confidential	Finds PDF documents marked as confidential on .onion sites.
Passwords in Config Files	inurl:config filetype:txt password	Searches for text files in configuration URLs containing passwords.
Financial Documents	filetype:xlsx site:onion financial	Locates Excel files related to financial data on .onion sites.
Database Dumps	filetype:sql site:onion dump	Finds SQL database dump files on .onion sites.
Email Lists	filetype:csv site:onion email	Searches for CSV files containing email lists on .onion sites.
Login Credentials	intitle:"login credentials" filetype:docx	Locates Word documents with login credentials in the title.
Server Configurations	filetype:xml inurl:config server	Finds XML files related to server configurations.
Private Keys	filetype:key site:onion private	Searches for private key files on .onion sites.
Medical Records	filetype:pdf site:onion "medical records"	Locates PDF documents containing medical records on .onion sites.
Business Plans	filetype:ppt site:onion "business plan"	Finds PowerPoint files with business plans on .onion sites.
Source Code	filetype:py site:onion "def "	Searches for Python source code files on .onion sites.
Legal Documents	filetype:docx site:onion "legal document"	Locates Word documents related to legal matters on .onion sites.
Bank Statements	filetype:pdf site:onion "bank statement"	Finds PDF documents containing bank statements on .onion sites.
Intellectual Property	filetype:pdf inurl:patent confidential	Searches for patent documents marked as confidential in PDFs.
Security Vulnerabilities	filetype:txt inurl:exploit "security vulnerability"	Finds text files detailing security vulnerabilities and exploits.

Table 2.3: Table of search queries for footprinting on the dark web

Determining the Operating System

Attackers use various online tools such as Netcraft, Shodan, and Censys to detect the operating system used at the target organization. These tools search the Internet for detecting connected devices such as routers, servers, and IoT devices belonging to the target organization. Using these tools, attackers obtain information such as the city, country, latitude/longitude, hostname, operating system, and IP address of the target organization. Such information further helps attackers in identifying potential vulnerabilities and finding effective exploits to perform various attacks on the target.

- **Netcraft**

  Source: *https://www.netcraft.com*

  The technique of obtaining information about the target network operating system is called OS fingerprinting. Open ***https://www.netcraft.com/tools/*** in the browser and type the URL of the target website in the **What's that site running?** field. Attackers use the Netcraft tool to identify all the sites associated with the target domain along with the operating system running at each site.

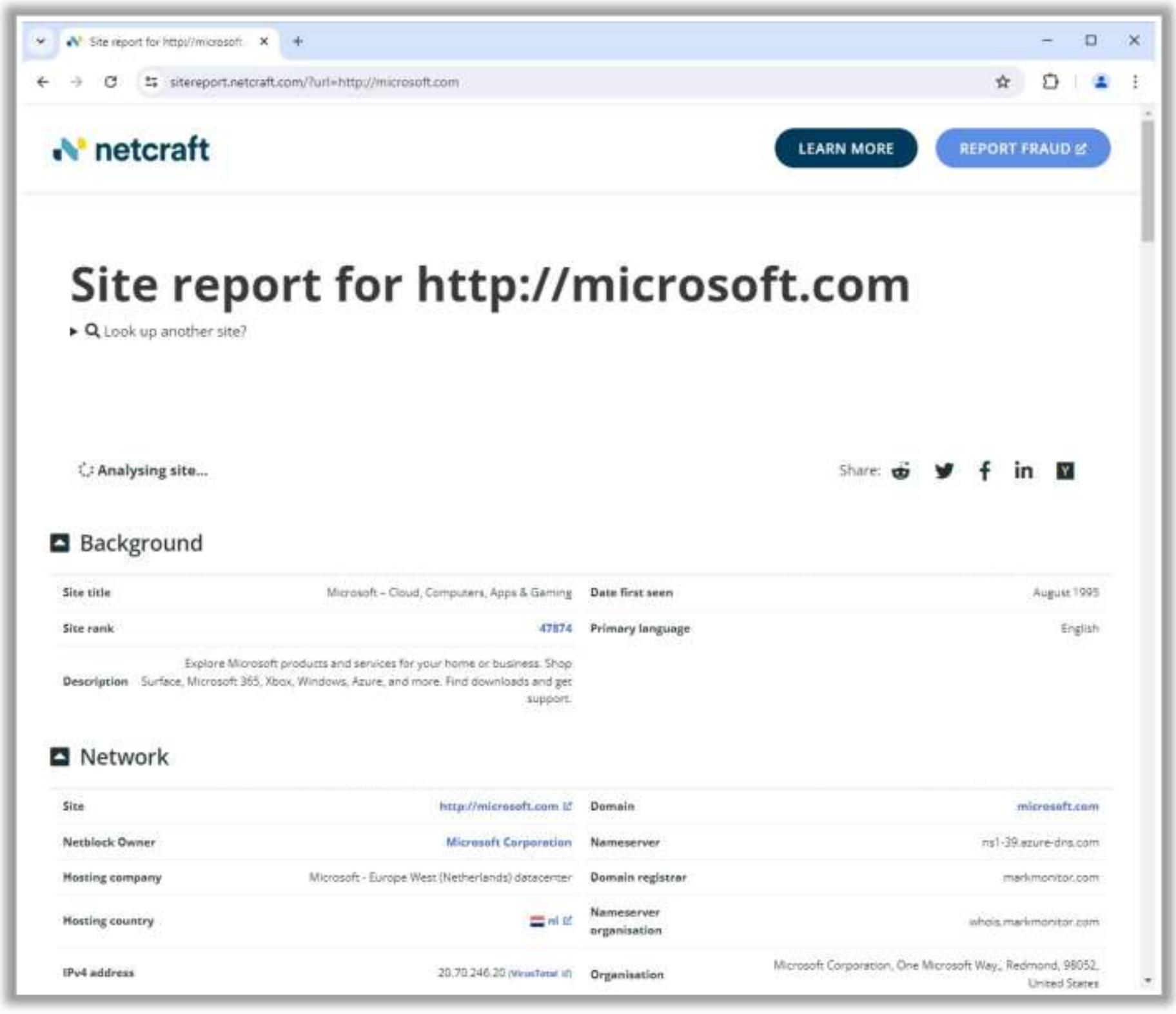


Figure 2.35: Screenshot of Netcraft showing results for Microsoft

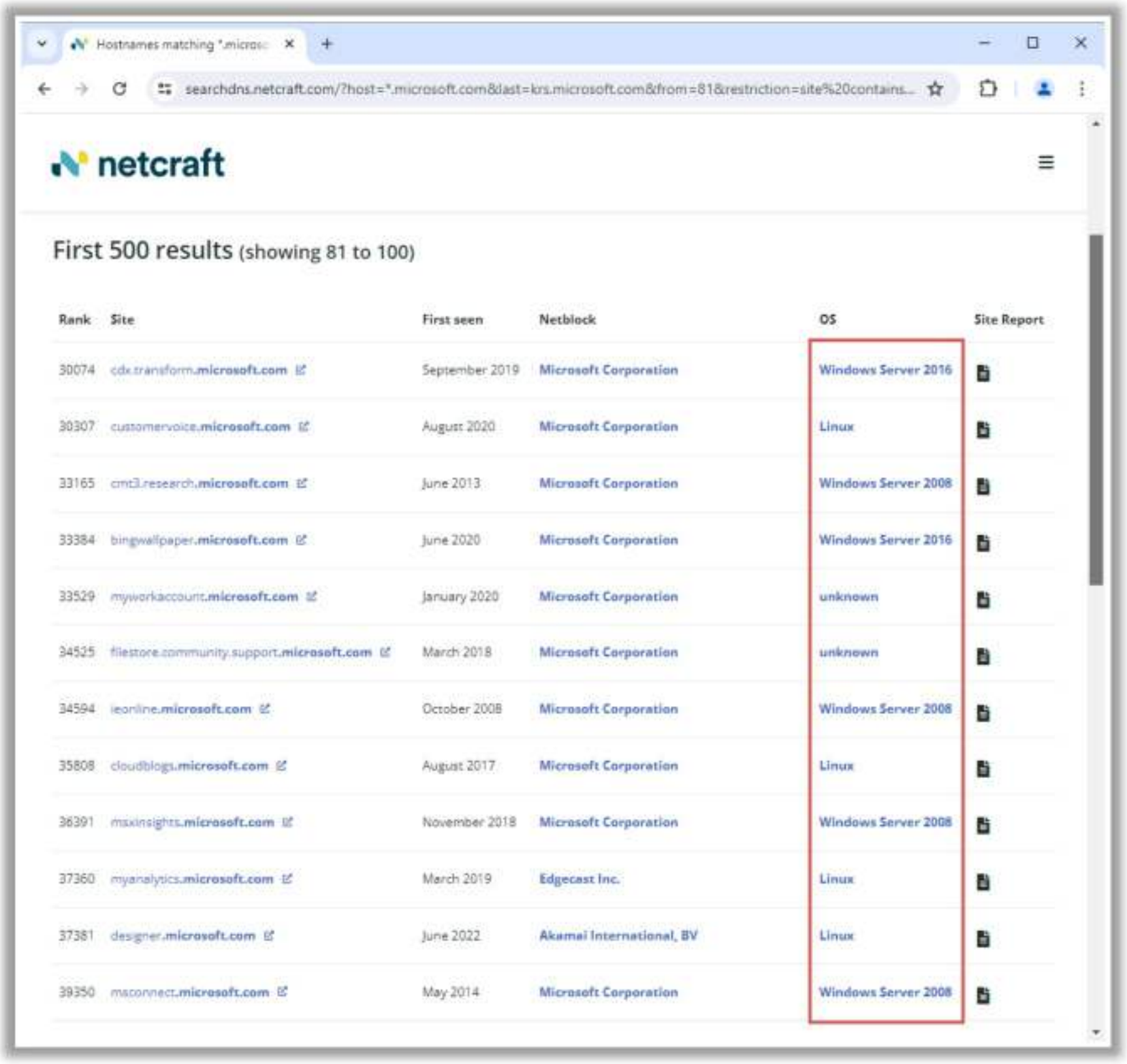


Figure 2.36: Screenshot of Netcraft showing target operating system

- **SHODAN Search Engine**

  Source: *https://www.shodan.io*

  Shodan is a computer search engine that searches the Internet for connected devices (routers, servers, and IoT.). You can use Shodan to discover which devices are connected to the Internet, where they are located, and who is using them.

  It helps attackers to keep track of all the devices on the target network that are directly accessible from the Internet. It also allows the attacker to find devices based on the city, country, latitude/longitude, hostname, operating system, and IP address. Further, it helps the attacker to search for known vulnerabilities and exploits across Exploit DB, Metasploit, CVE, OSVDB, and Packetstorm with a single interface.

  As shown in the screenshot, attackers use this tool to detect various target devices connected to the Internet along with the operating system used.

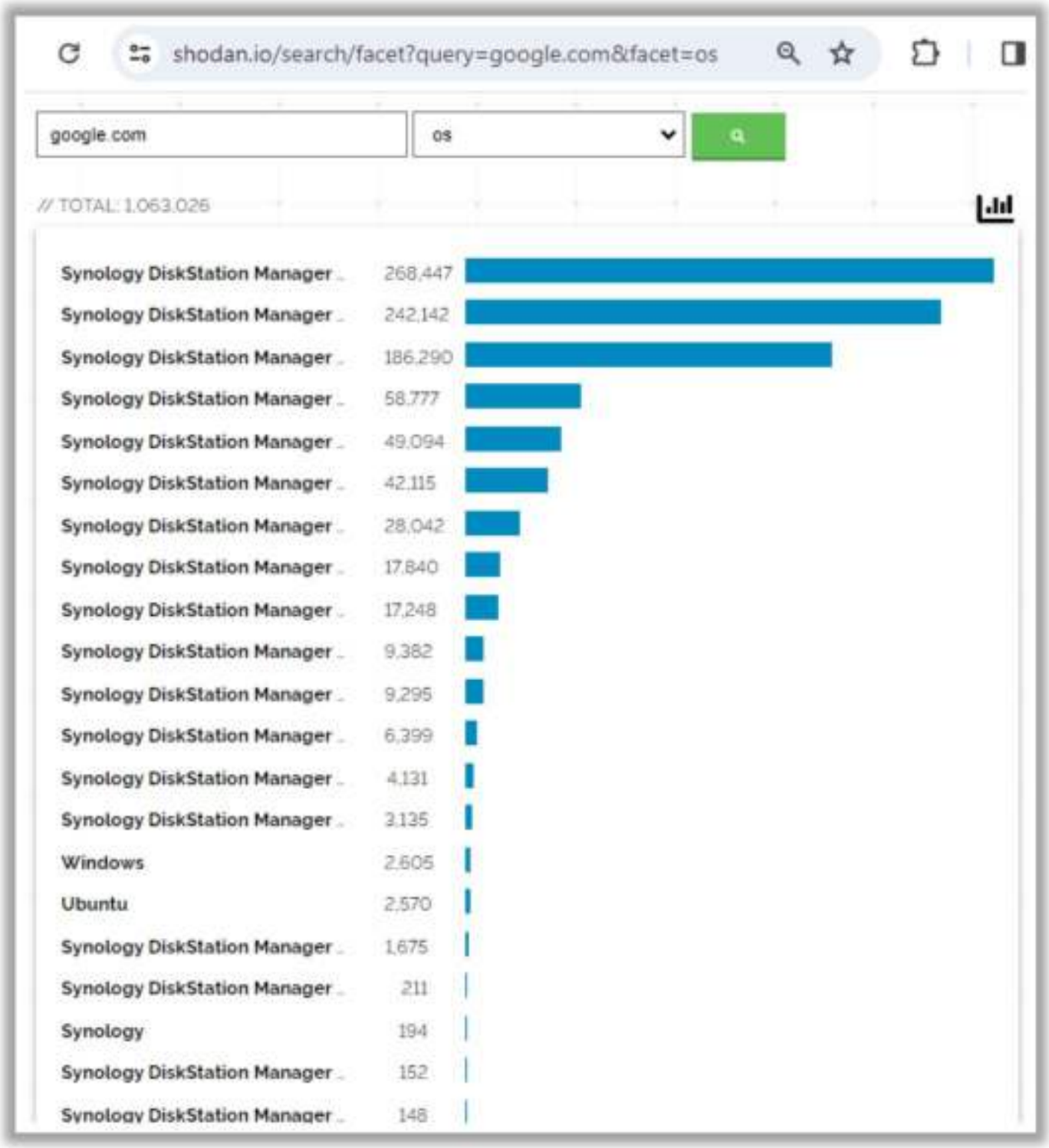


Figure 2.37: Screenshot of SHODAN Search Engine showing target operating system

- **Censys**

  Source*: https://censys.io*

  Censys monitors the infrastructure and discovers unknown assets anywhere on the Internet. It provides a full view of every server and device exposed to the Internet.

  Attackers use this tool to monitor the target IT infrastructure to discover various devices connected to the Internet along with their details such as the operating system used, IP address, protocols used, and geographical location.

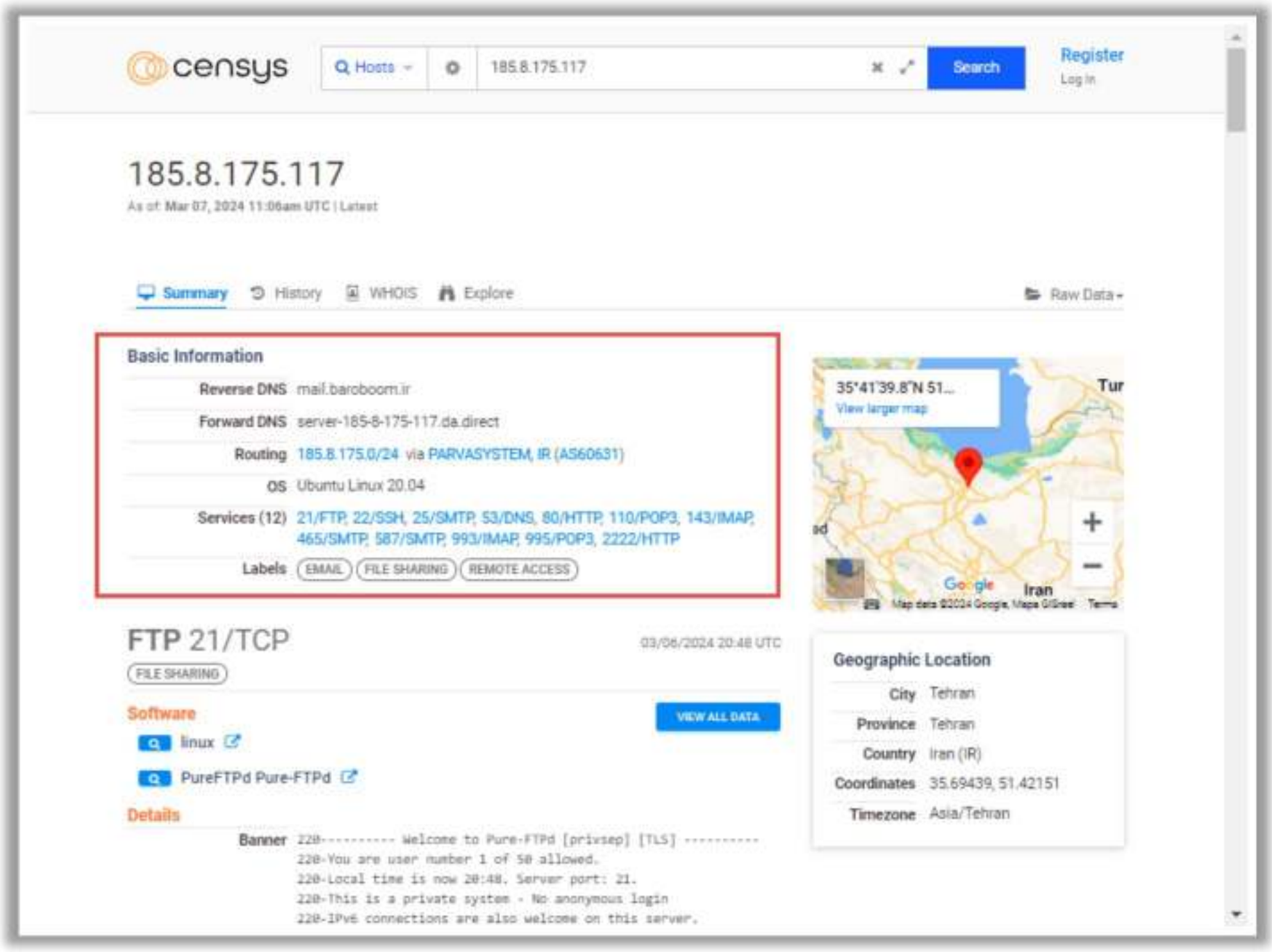
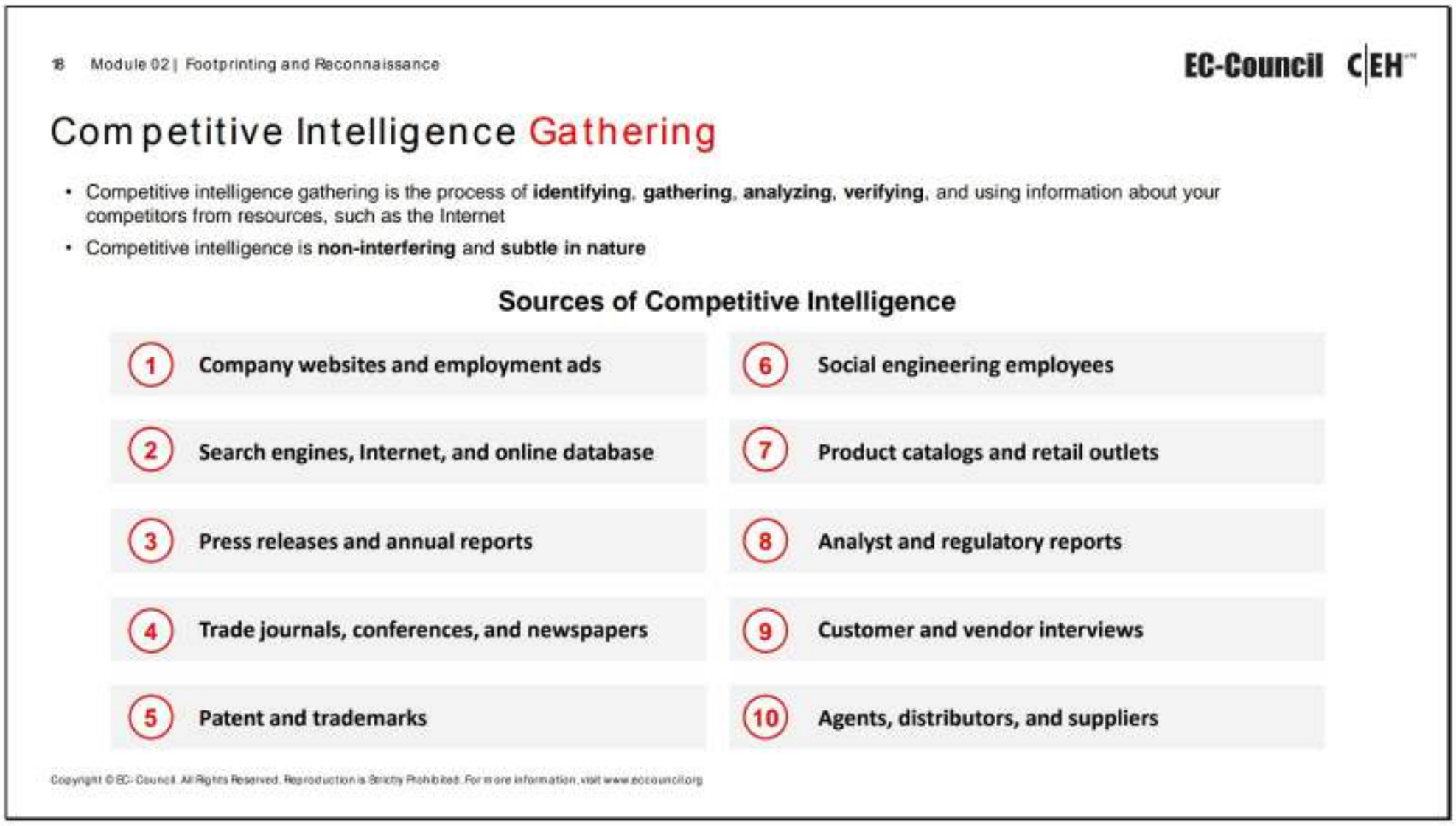


Figure 2.38: Screenshot of Censys Search Engine showing target operating system



Competitive Intelligence Gathering

Competitive intelligence gathering is the process of identifying, gathering, analyzing, verifying, and using information about your competitors from resources such as the Internet. Competitive intelligence means understanding and learning about other businesses to become as competitive as possible. It is non-interfering and subtle in nature compared to direct intellectual property theft carried out via hacking or industrial espionage. It focuses on the external business environment. In this method, professionals gather information ethically and legally instead of gathering it secretly.

Competitive intelligence helps in determining:

- What the competitors are doing?
- How competitors are positioning their products and services?
- What customers are saying about competitors' strengths and weaknesses?

Companies carry out competitive intelligence either by employing people to search for information or by utilizing a commercial database service, which involves lower costs. The information that is gathered can help the managers and executives of a company make strategic decisions.

Sources of Competitive Intelligence

Competitive Intelligence gathering can be performed using a direct or indirect approach.

- **Direct Approach**

  The direct approach serves as the primary source for competitive intelligence gathering. Direct approach techniques include gathering information from trade shows, social engineering of employees and customers, and so on.

- **Indirect Approach**

  Through an indirect approach, information about competitors is gathered using online resources. Indirect approach techniques include:

  - Company websites and employment ads
  - Support threads and reviews
  - Search engines, Internet, and online database
  - Social media postings
  - Press releases and annual reports
  - Trade journals, conferences, and newspapers
  - Patent and trademarks
  - Product catalogs and retail outlets
  - Analyst and regulatory reports
  - Customer and vendor interviews
  - Agents, distributors, and suppliers
  - Industry-specific blogs and publications
  - Legal databases, e.g., LexisNexis
  - Business information databases, e.g., D&B Hoovers
  - Online job postings
  - Financial filings
  - Technology solutions, e.g., Crunchbase
  - Intellectual property analysis

Competitive Intelligence - When Did this Company Begin? How Did it Develop?

Gathering competitor documents and records helps to improve productivity and profitability, which in turn stimulates the growth of the company. It helps in determining answers to the following:

- **When did it begin?**

  Through competitive intelligence, companies can collect the history of a particular company, such as its establishment date. Sometimes, they gather crucial information that is not often available to others.

- **How did it develop?**

  What are the various strategies that the company uses? Development intelligence can include advertisement strategies, customer relationship management, and so on.

- **Who leads it?**

  This information helps a company learn about the competitor's decision-makers.

- **Where is it located?**

  Competitive intelligence also includes the location of the company and information related to various branches and their operations.

Attackers can use the information gathered through competitive intelligence to build a hacking strategy.

Information Resource Sites

Information resource sites that help to gain competitive intelligence include:

- **EDGAR Database**

  Source: *https://www.sec.gov/edgar*

  The Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) performs automated collection, validation, indexing, acceptance, and forwarding of submissions by companies and others who are required by law to file with the U.S. Securities and Exchange Commission (SEC). Its primary purpose is to increase the efficiency and fairness of the securities market for the benefit of investors, corporations, and the economy by accelerating the receipt, acceptance, dissemination, and analysis of time-sensitive corporate information filed with the agency.

- **D&B Hoovers**

  Source: *https://www.dnb.com*

  D&B Hoovers leverages a commercial database of 120 million business records and analytics to deliver a sales intelligence solution that enables sales and marketing professionals to focus on the right prospects so that they can generate immediate growth for their business.

- **LexisNexis**

  Source: *https://www.lexisnexis.com*

  LexisNexis provides content-enabled workflow solutions designed specifically for professionals in the legal, risk management, corporate, government, law enforcement, accounting, and academic markets. It maintains an electronic database of information related to legal and public records. It enables customers to access documents and records of legal, news, and business sources. It is beneficial for companies and government agencies seeking data analytics supporting compliance, customer acquisition, fraud detection, health outcomes, identity solutions, investigation, receivables management, risk decisioning, and workflow optimization.

- **Business Wire**

  Source: *https://www.businesswire.com*

  Business Wire focuses on press release distribution and regulatory disclosure. This company distributes full-text news releases, photos, and other multimedia content from various organizations across the globe to journalists, news media, financial markets, investors, information website, databases, and general audiences. It has its own patented electronic network through which it releases news.

- **Factiva**

  Source: *https://www.dowjones.com*

  Factiva is a global news database and licensed content provider. It is a business information and research tool that gets information from licensed and free sources and provides capabilities such as searching, alerting, dissemination, and business information management. Factiva products provide access to more than 33,000 sources such as licensed publications, influential websites, blogs, images, and videos. Its resources are made available from nearly every country worldwide in 28 languages, including more than 600 continuously updated newswires.

Competitive Intelligence - What Are the Company's Plans?

Information resource sites that help attackers gain a company's business plans include:

- **MarketWatch**

  Source: *https://www.marketwatch.com*

  MarketWatch tracks the pulse of markets for engaged investors. The site is an innovator in business news, personal finance information, real-time commentary, and investment tools and data, with journalists generating headlines, stories, videos, and market briefs.

- **The Wall Street Transcript**

  Source: *https://www.twst.com*

  The Wall Street Transcript is a website as well as a paid subscription-based publication that publishes industry reports. It expresses the views of money managers and equity

analysts of different industry sectors. The site also publishes interviews with CEOs of companies.

- **Euromonitor**

  Source: *https://www.euromonitor.com*

  Euromonitor provides strategy research capabilities for consumer markets. It publishes reports on industries, consumers, and demographics. It provides market research and surveys focused on the organization's needs.

- **Experian**

  Source: *https://www.experian.com*

  Experian provides insights into competitors' search, affiliate, display, and social marketing strategies and metrics to improve marketing campaign results. It allows the user to:

  - Benchmark the effectiveness of existing customer acquisition strategies
  - Determine what is driving competitors' success
  - Use historical consumer data to forecast future trends and quickly respond to changing behaviors
  - Measure website's performance against industry or specific sites

- **The Search Monitor**

  Source: *https://www.thesearchmonitor.com*

  The Search Monitor provides competitive intelligence to monitor brand and trademark use, affiliate compliance, and competitive advertisers on paid search, organic search, local search, social media, mobile, and shopping engines worldwide. It helps interactive agencies, search marketers, and affiliate marketers to track ad rank, ad copy, keyword reach, click rates and CPCs, monthly ad spending, market share, trademark use, and affiliate activity.

- **USPTO**

  Source: *https://www.uspto.gov*

  The United States Patent and Trademark Office (USPTO) provides information related to patent and trademark registration. It provides general information concerning patents and search options for patents and trademark databases.

Competitive Intelligence - What Expert Opinions Say About the Company?

Information resource sites that help the attacker to obtain expert opinions about the target company include:

- **SEMRush**

  Source: *https://www.semrush.com*

  SEMRush is a competitive keyword research tool. It can provide a list of Google keywords and AdWords for any site, as well as a competitor list in the organic and paid Google

search results. It enables an approach for gaining in-depth knowledge about what competitors are advertising and their budget allocation to specific Internet marketing tactics.

- **ABI/INFORM Global**

  Source: *https://www.proquest.com*

  ABI/INFORM Global is a business database. ABI/INFORM Global offers the latest business and financial information for researchers. With ABI/INFORM Global, users can determine business conditions, management techniques, business trends, management practice and theory, corporate strategy and tactics, and the competitive landscape.

- **SimilarWeb**

  Source: *https://www.similarweb.com*

  SimilarWeb aggregates data from multiple sources to estimate traffic, geography, and referral data for a company's websites and mobile apps. It also provides a panel through a browser extension that allows refining other data sources by anonymously tracking browser activity across millions of browsers worldwide.

- **SERanking**

  Source: *https://seranking.com*

  SERanking is an online competitor analysis tool that provides a complete view of the target organization's website traffic dynamics. The tool also helps enterprises or users study their websites' major competitors, find newcomers, and compare semantics to rivals' keywords. It enables an approach to conduct pay-per-click (PPC) competitor research for tracing competitors' tactics to enhance one's own advertisement strategies.

Other Techniques for Footprinting through Internet Research Services

Footprinting Technique	Description	Information Gathered	Tools Used
Finding the Geographical Location of the Target	Obtain the physical location of the target	Entrances to buildings, security cameras, gates, places to hide, weak spots in perimeter fences, etc.	Google Earth, Google Maps, and Wikimapia
Gathering Information from Financial Services	Search for financial data such as stock quotes and charts, financial news, and portfolios	Market value of a company's shares, company profile, and competitor details	Google Finance, MSN Money, and Yahoo! Finance
Gathering Information from Business Profile Sites	Retrieve business information of companies located in a particular region	Location, addresses, contact information, and employee database of the target organization	opencorporates, Crunchbase, and corporationwiki
Monitoring Targets Using Alerts	Obtain up-to-date information of the target, usually via email or SMS	Mentions of the organization's name, member names, website, or any of its people or projects	Google Alerts, X Alerts, and Giga Alerts
Tracking the Online Reputation of the Target	Monitor a company's reputation on the Internet	Search engine ranking information, email notifications when a company is mentioned online, and social news about the company	Mention, ReviewPush, and Reputology
Gathering Information from Groups, Forums, and Blogs	Join the target organization's employee groups, where they share personal and company information	Public network information, system information, and personal information	Google Groups and LinkedIn Groups
Gathering Information from Public Source-Code Repositories	Identify information about the developers and technologies used	Configuration files, private SSH and SSL keys, source-code files, dynamic libraries, and software tools developed by contributors	Recon-ng

Other Techniques for Footprinting through Internet Research Services

- **Finding the Geographical Location of the Target**

  Information such as the physical location of an organization plays a vital role in the hacking process. Attackers with the knowledge of a target organization's location may attempt dumpster diving, surveillance, social engineering, and other non-technical attacks to gather more information.

  Tools for Finding the Geographical Location

  The tools for finding the geographical location allow you to find and explore most locations on the earth. They provide information such as images of buildings, as well as their surroundings, including Wi-Fi networks. These tools provide interactive maps, outline maps, satellite imagery, and information on how to interact with and create one's own maps. Google Maps, Apple Maps, Waze, and other tools provide driving directions, traffic conditions, landmarks, and detailed address and contact information. The attackers can use this information to gain unauthorized access to buildings, wired and wireless networks, and systems.

  Attackers may use tools such as Google Earth, Google Maps, and Wikimapia, to find or locate entrances to buildings, security cameras, gates, places to hide, weak spots in perimeter fences, and utility resources such as electricity connections, to measure the distance between different objects, and so on.

- **Google Earth**

  Source: *https://earth.google.com*

  Attackers use the Google Earth tool to find the exact location of a target. Using this tool, attackers can even access 3D images that depict most of the populated Earth's surface with a high resolution. The detail allows attackers to obtain street views, altitude information, and even coordinates.

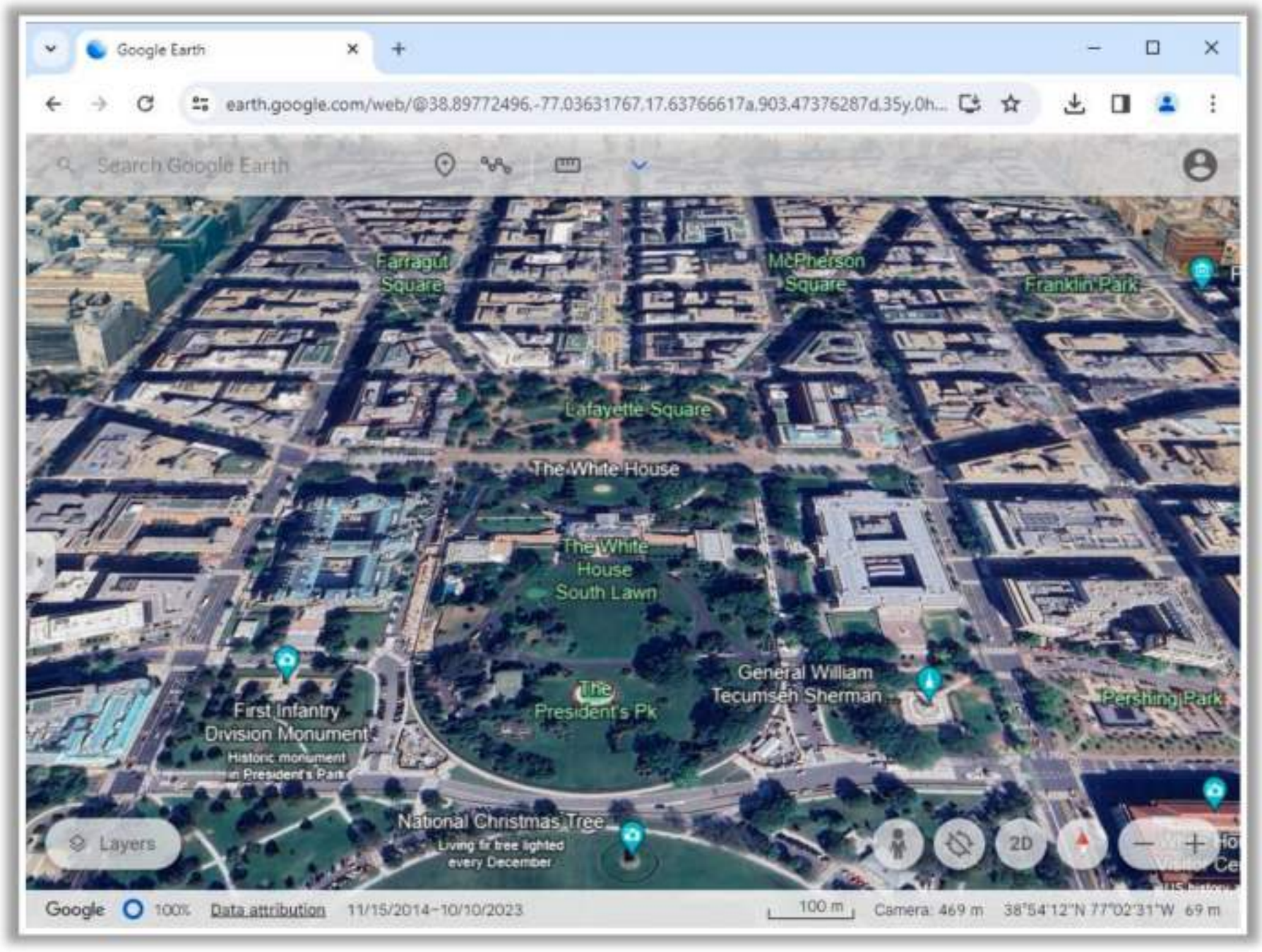


Figure 2.39: Screenshot of Google Earth

- **Gathering Information from Financial Services**

  Attackers who seek access to personal information or financial information often target financial data such as stock quotes and charts, financial news, and portfolios. Financial services such as Google Finance, MSN Money, Yahoo Finance, and Investing.com can provide a large amount of useful information such as the market value of a company's shares, company profile, competitor details, stock exchange rates, corporate press releases, financial reports along with news, and blog search articles about corporations. The information provided varies from one service to the other. Financial firms rely on web services to perform transactions and grant users access to their accounts. Attackers can obtain sensitive and private information regarding these firms by using malware, exploiting software design flaws, breaking authentication mechanisms, service flooding, and performing brute force attacks and phishing attacks.

- **Google Finance**

  Source: *https://www.google.com/finance*

  The Google finance service features business and enterprise headlines for many corporations, including their financial decisions and major news events. Stock information is also available, as are stock price charts that contain marks for major news events and corporate actions. The site also aggregates Google news and Google blog search articles about each corporation.

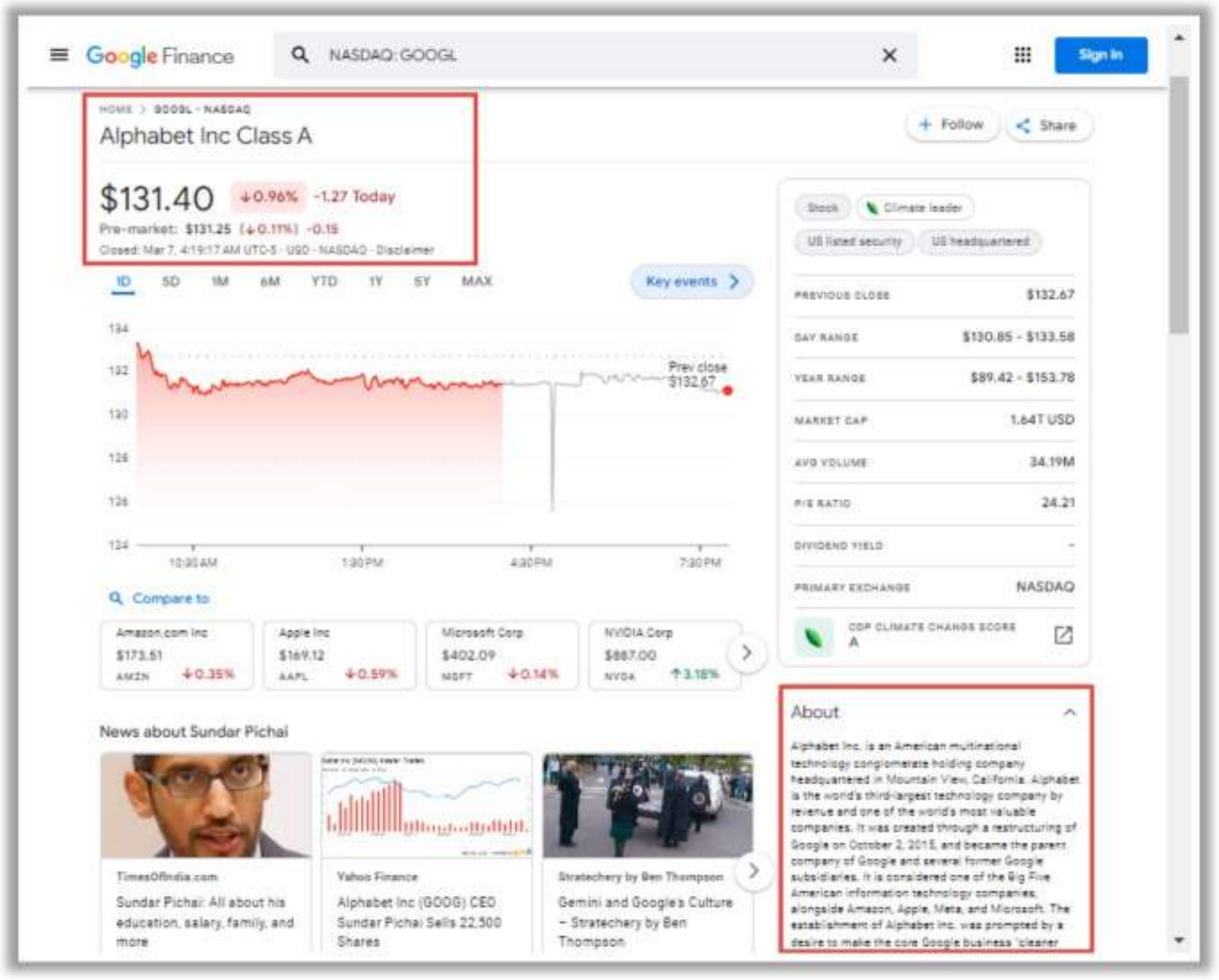


Figure 2.40: Screenshot of Google Finance Service

- **Gathering Information from Business Profile Sites**

  Finding useful information from corporate websites is a necessary step in the information gathering phase. These business profile sites contain business information of companies located in a particular region with their contact information, which can be viewed by anyone.

  Attackers use business profile sites such as opencorporates, Crunchbase, and corporationwiki to gather important information about the target organizations, such as their location, addresses, contact information (such as phone numbers, email addresses), employee database, department names, type of service provided, and type of industry.

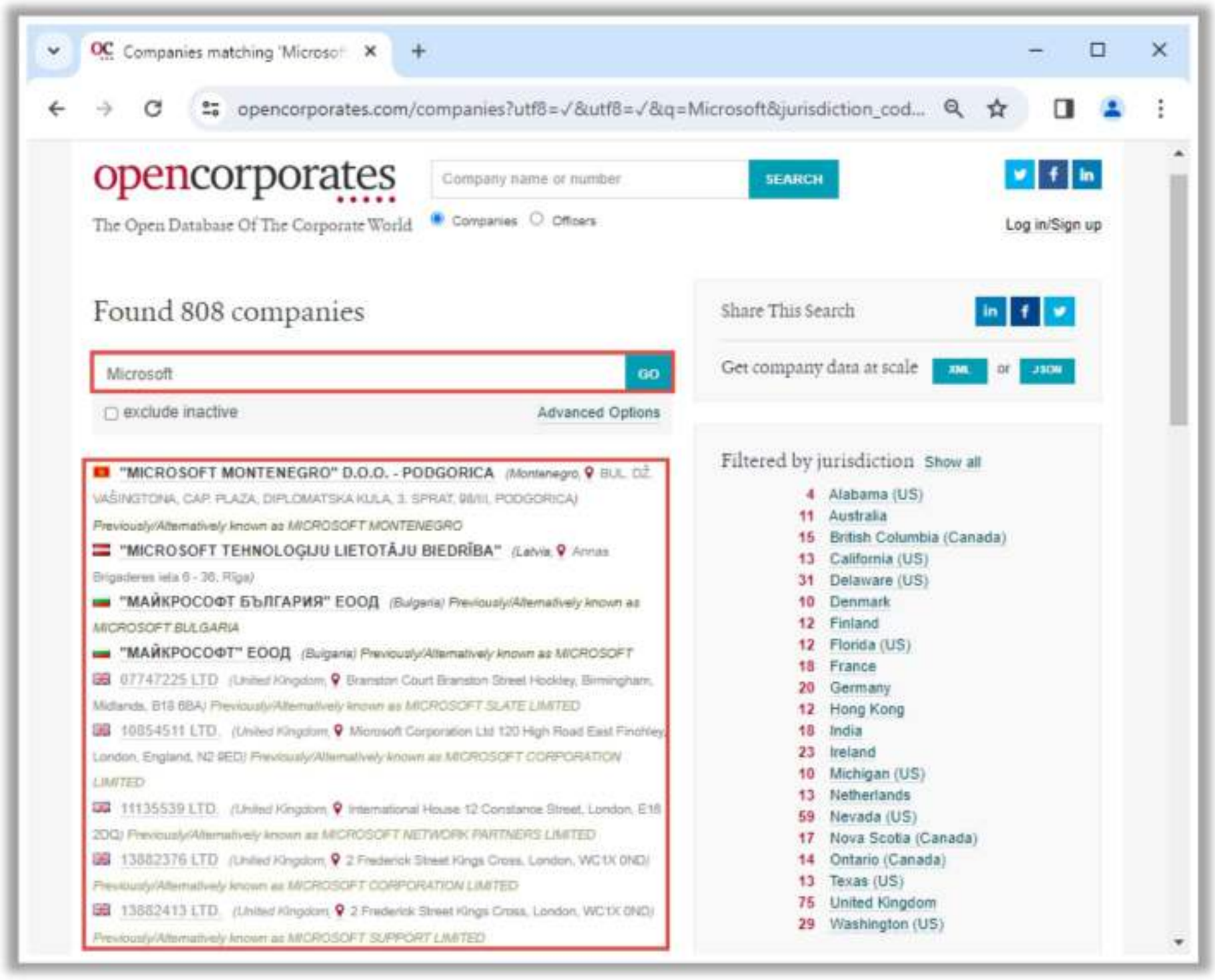


Figure 2.41: Screenshot of opencorporates showing search results of Microsoft

- **Monitoring Targets Using Alerts**

  Alerts are content monitoring services that provide automated, up-to-date information based on user preference, usually via email or SMS. To receive alerts, a user must register on the website and provide either an email address or a phone number. Online alert services automatically notify users when new content from news, blogs, and discussion groups matches a set of search terms selected by the user. These services provide up-to-date information about competitors and the industry.

  Tools such as Google Alerts, X Alerts, and Giga Alerts help attackers to track mentions of the organization's name, member names, website, or any people or projects that are important. Attackers can gather updated information about the target periodically from the alert services and use it for further attacks.

- **Google Alerts**

  Source: *https://www.google.com/alerts*

  Google Alerts automatically notifies users when new content from news, websites, blogs, videos, and/or discussion groups matches a set of search terms selected by the user and stored by the Google Alerts service.

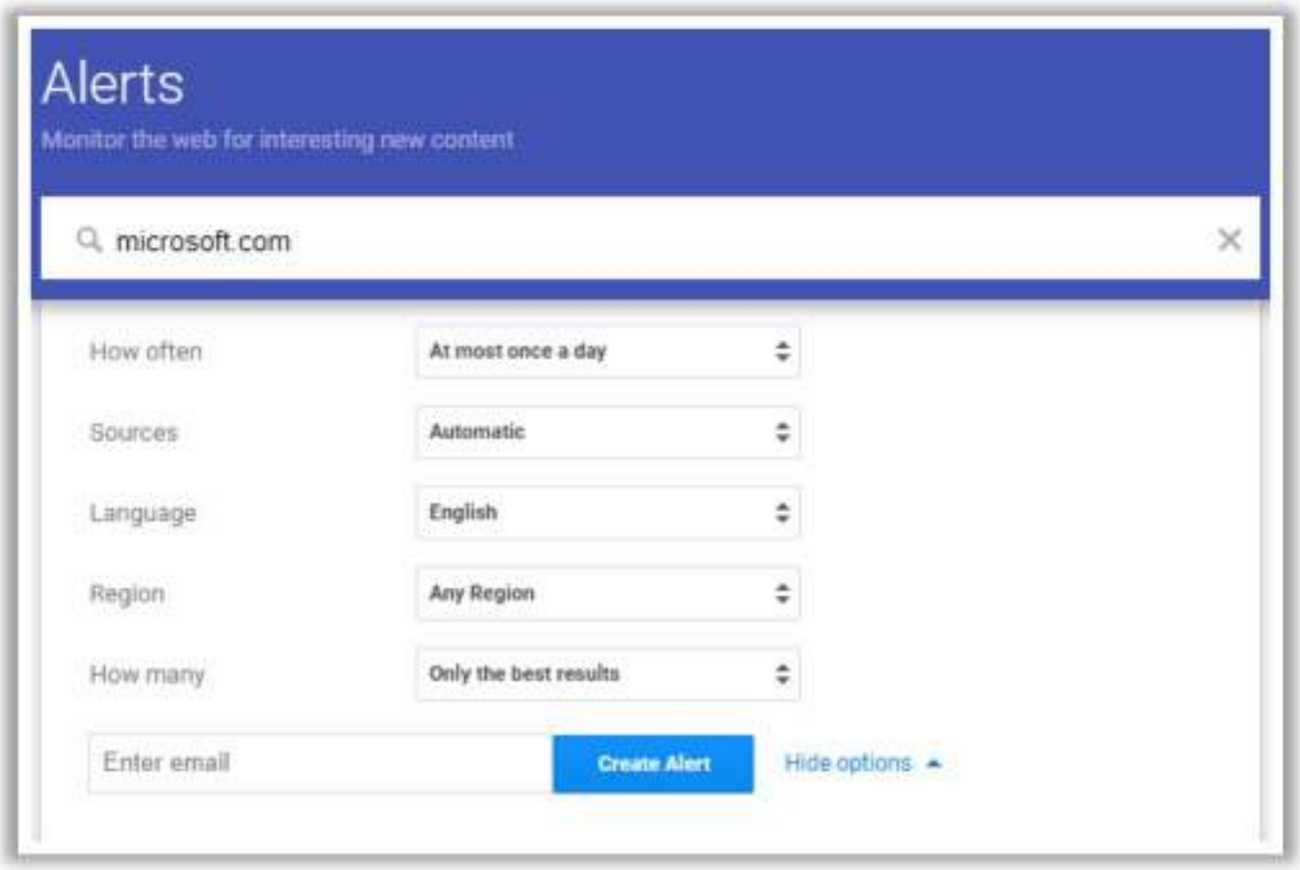


Figure 2.42: Screenshot of Google Alert

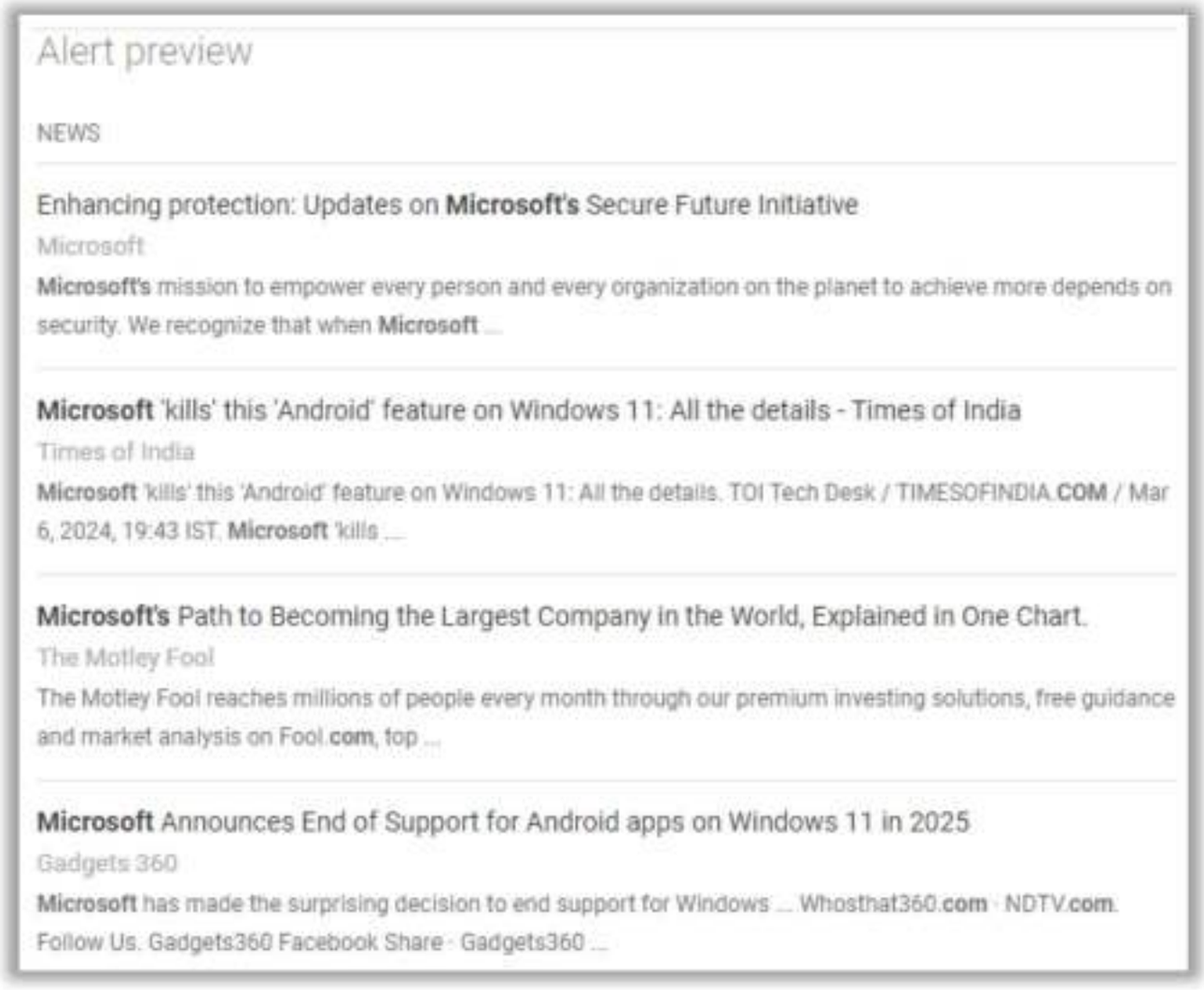


Figure 2.43: Screenshot of Google Alert Preview

- **Tracking the Online Reputation of the Target**

  Online Reputation Management (ORM) is a process of monitoring displays when someone searches for your company's reputation on the Internet. ORM then takes measures to minimize negative search results or reviews. The process helps to improve brand reputation.

  Companies often track the public feedback given to them using ORM tracking tools and then take measures to improve their credibility and retain their customers' trust. For positive online reputation management, organizations will often try to be more transparent over the Internet. This transparency may help the attacker to collect genuine information about the target organization.

  Online Reputation Tracking Tools

  Online reputation tracking tools help us to discover what people are saying online about the company's brand in real time across the web, social media, and news. They help in monitoring, measuring, and managing one's reputation online. Attackers can use tools such as Mention, ReviewPush, and Reputology to track the online reputation of companies.

  An attacker may use ORM tracking tools to:

  - Track a company's online reputation
  - Collect a company's search engine ranking information
  - Obtain email notifications when a company is mentioned online
  - Track conversations
  - Obtain social news about the target organization

  Mention

  Source: *https://mention.com*

  Mention is an online reputation tracking tool that helps attackers in monitoring the web, social media, forums, and blogs to learn more about the target brand and industry. As shown in the screenshot, this tool helps attackers in tracking online conversations as they happen, wherever they happen. Using Mention, attackers can have live, up-to-date reports delivered to any email address in real time.

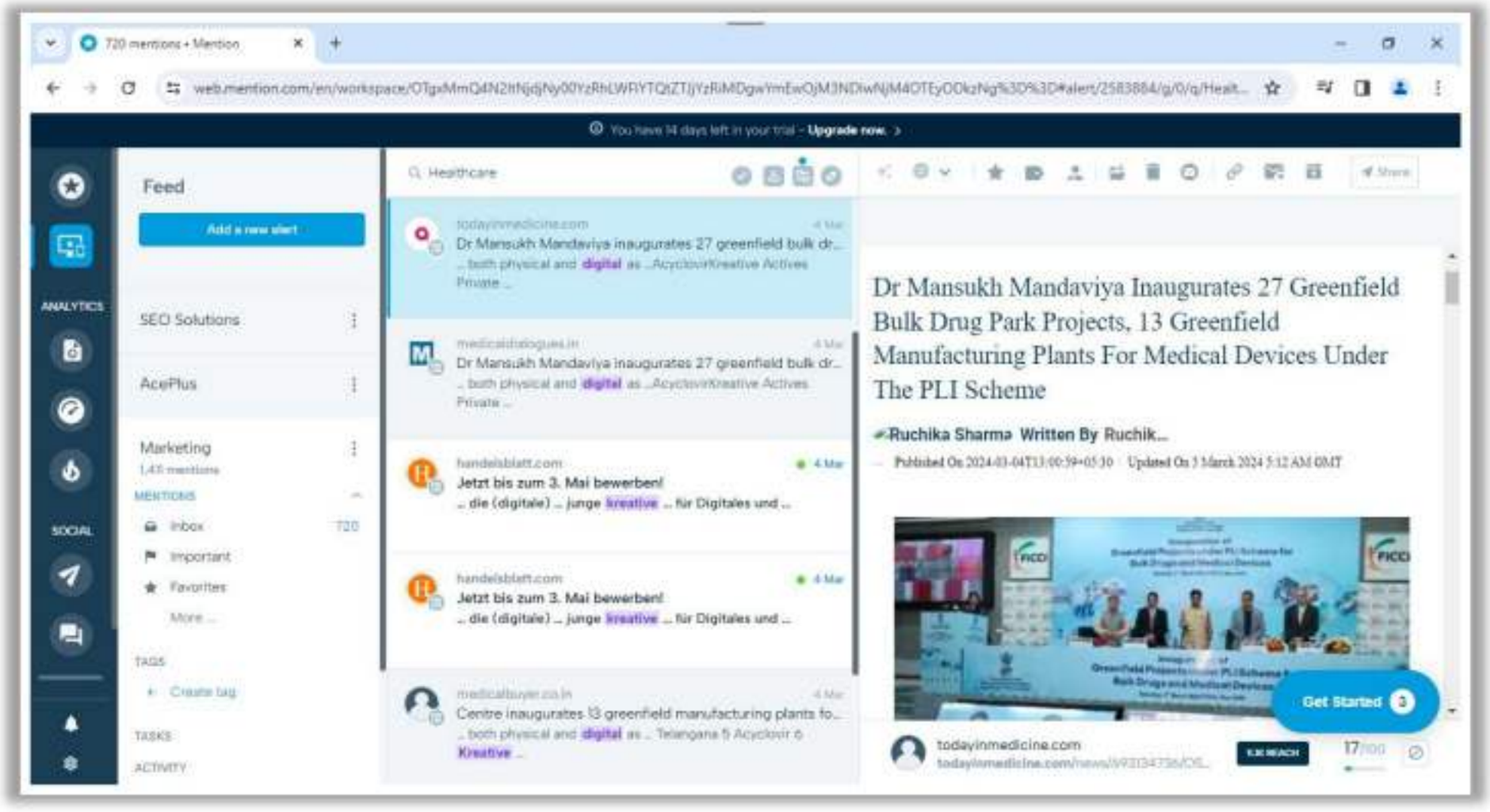


Figure 2.44: Screenshot of Mention

- **Gathering Information from Groups, Forums, and Blogs**

  Many Internet users use blogs, groups, and forums for knowledge sharing purposes. For this reason, attackers often focus on groups, forums, and blogs to find information about a target organization and its people. Organizations generally fail to monitor the exchange of information that employees reveal to other users in forums, blogs, and group discussions. Attackers take this as an advantage and collect sensitive information about the target, such as public network information, system information, and employee personal information. Attackers can register with fake profiles in Google Groups, LinkedIn Groups, and so on. They try to join the target organization's employee groups, where they can obtain personal and company information. Attackers can also search for information in groups, forums, and blogs by Fully Qualified Domain Names (FQDNs), IP addresses, and usernames.

  Employee information that an attacker can gather from groups, forums, and blogs may include:

  - Full name of the employee
  - Place of work and residence
  - Home telephone, cell number, or office number
  - Personal and organizational email address
  - Pictures of the employee residence or work location that include identifiable information
  - Pictures of employee awards and rewards or upcoming goals

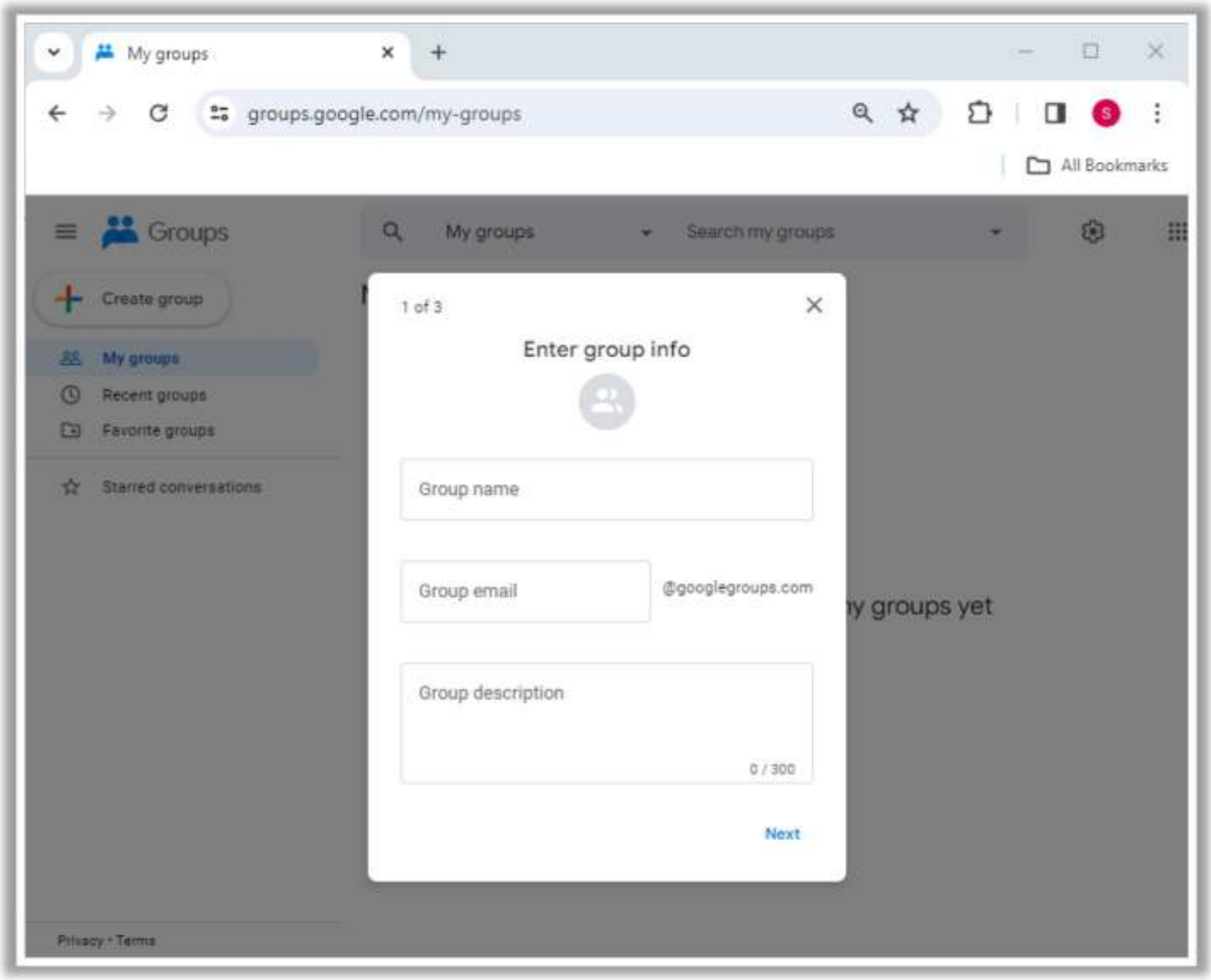


Figure 2.45: Screenshot of Google Groups

- **Gathering Information from Public Source-Code Repositories**

  Source code–based repositories are online services or tools available on internal servers or can be hosted on third-party websites such as GitHub, GitLab, SourceForge, and BitBucket. These sites contain sensitive data related to configuration files, private Secure Shell (SSH) and Secure Sockets Layer (SSL) keys, source-code files, dynamic libraries, and software tools developed by contributors, which can be leveraged by attackers to launch attacks on the target organization.

  Software professionals reposit a huge quantity of source code related to web pages and programs online, universally or privately, for future access. To meet deadlines and increase efficiency during product development, developers may also access data stored in the repositories to develop or patch their applications quickly and minimize the development time and cost.

  Source-code repository websites such as GitHub may contain security flaws that pave the way for attackers to target applications. The applications reposited on such public websites may also contain confidential files and information that allow attackers to search and identify the developers and technologies used.

The information collected from such public repositories may not be sufficient for performing direct attacks, but when this information is used with some active footprinting techniques, it may allow attackers to launch attacks such as spear phishing against specific users or employees of the target organization. Further, the security flaws in such repositories may allow attackers to accumulate constructive data to perform social engineering and infrastructure attacks on the target organization.

Attackers can use tools such as Recon-ng to discover public source-code repositories.

- **Recon-ng**

  Source: *https://github.com*

  Recon-ng is a full-featured reconnaissance framework designed to provide a powerful environment to conduct web-based reconnaissance quickly and thoroughly. It assists attackers in gathering information from public source-code repositories.

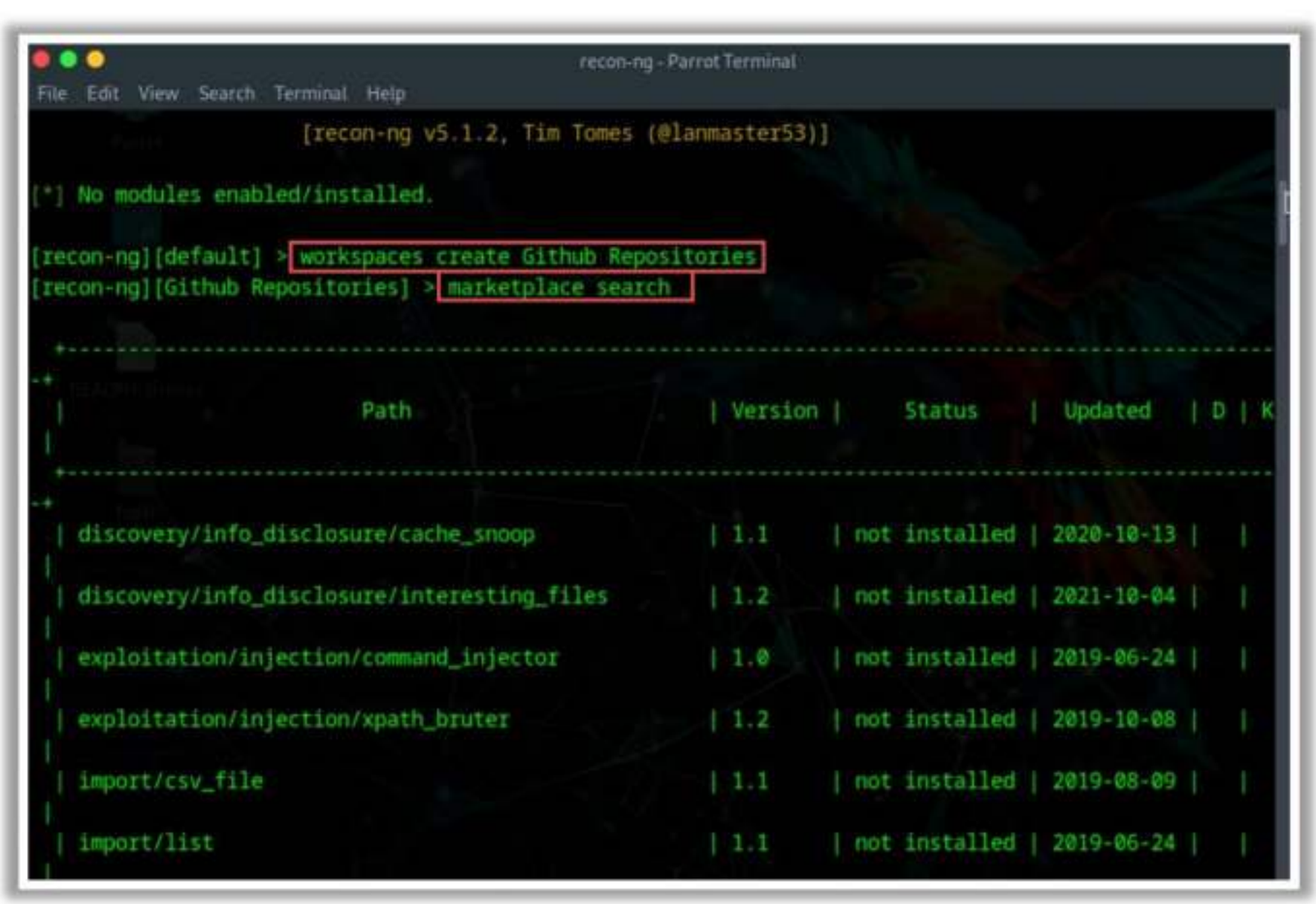


Figure 2.46: Screenshot of recon-ng showing the creation of a workspace and a marketplace search

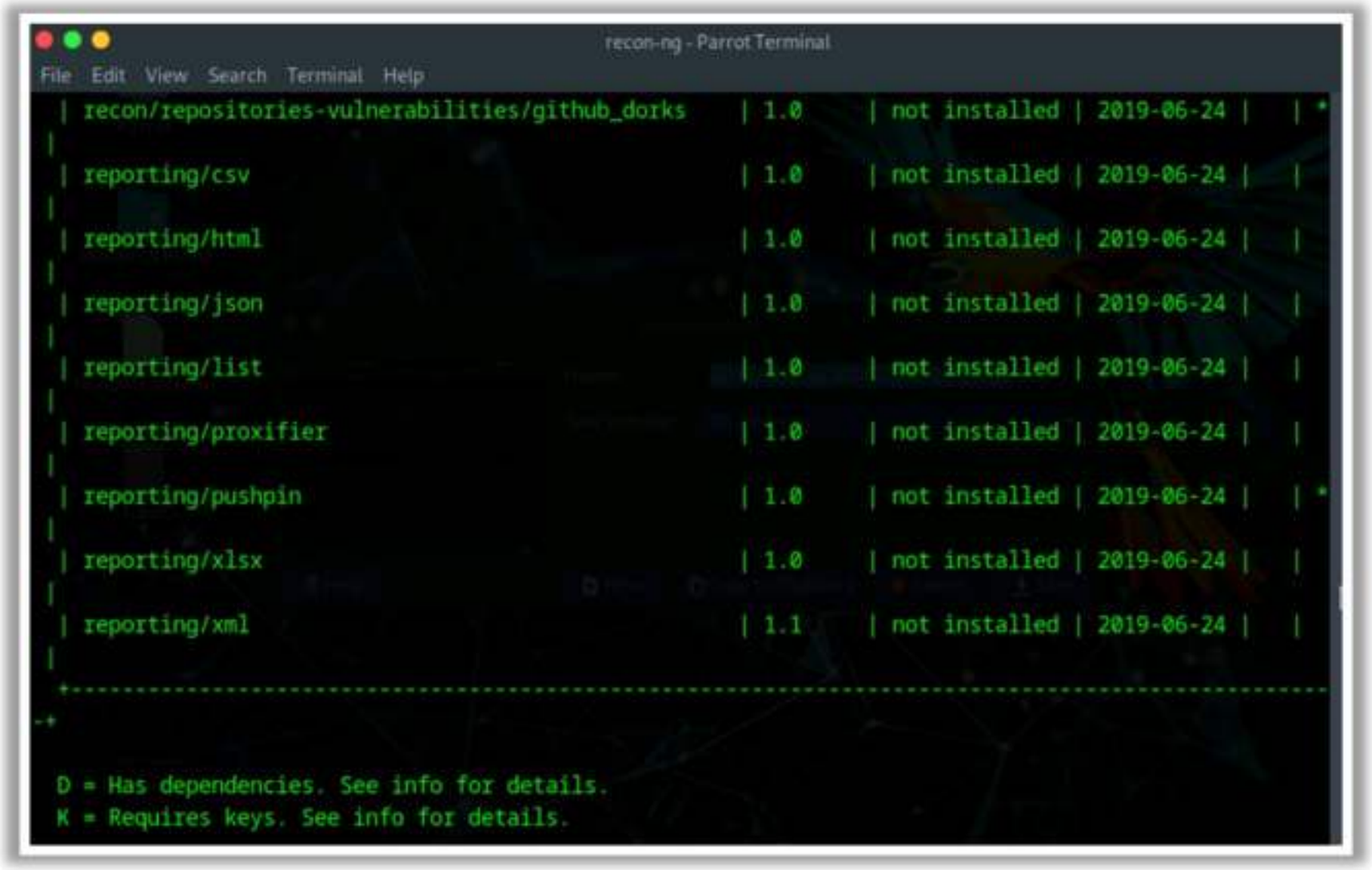


Figure 2.47: Screenshot of recon-ng displaying source-code repositories

Footprinting through Social Networking Sites

While footprinting through social networking sites may seem similar to footprinting through social engineering (which is discussed in greater detail later), there are some differences between the two methods. In footprinting through social engineering, the attacker tricks people into revealing information, whereas in footprinting through social networking sites, the attacker gathers information available on those sites. Attackers can even use social networking sites as a medium to perform social engineering attacks.

This section explains the type of information one can collect from social networking sites and how it can be obtained. It aims to familiarize you with locating information from social media sites using various online services and resources.

People Search on Social Networking Sites

Searching for a particular person on a social networking website is fairly easy. Social networking services are online services, platforms, or sites that focus on facilitating the building of social networks or social relations among people. These websites contain information provided by users in their profiles. They help relate people directly or indirectly to each other through various fields, such as common interests, work locations, and education.

Social networking sites allow people to share information quickly, as they can update their personal details in real time. Such sites allow users to update facts about upcoming or current events, recent announcements and invitations, and so on. Social networking sites are a great platform for finding people and their related information. Many social networking sites allow visitors to search for people without registering on the site; this makes people searching on social networking sites an easy and anonymous task. A user can search for a person using the name,

email, or address. Some sites allow users to check whether an account is active, which then provides information on the status of the person being searched.

Social networking sites such as Facebook, Twitter, LinkedIn, and Instagram allow you to find people by name, keyword, company, school, friends, colleagues, and the people living around them. Searching for people on these sites returns personal information such as name, position, organization name, current location, and educational qualifications. In addition, you can also find professional information such as company or business, current location, phone number, email ID, photos, videos and so on. Social networking sites such as Twitter are used to share advice, news, concerns, opinions, rumors, and facts. Through people searching on social networking services, an attacker can gather critical information that will help them in performing social engineering or other kinds of attacks.

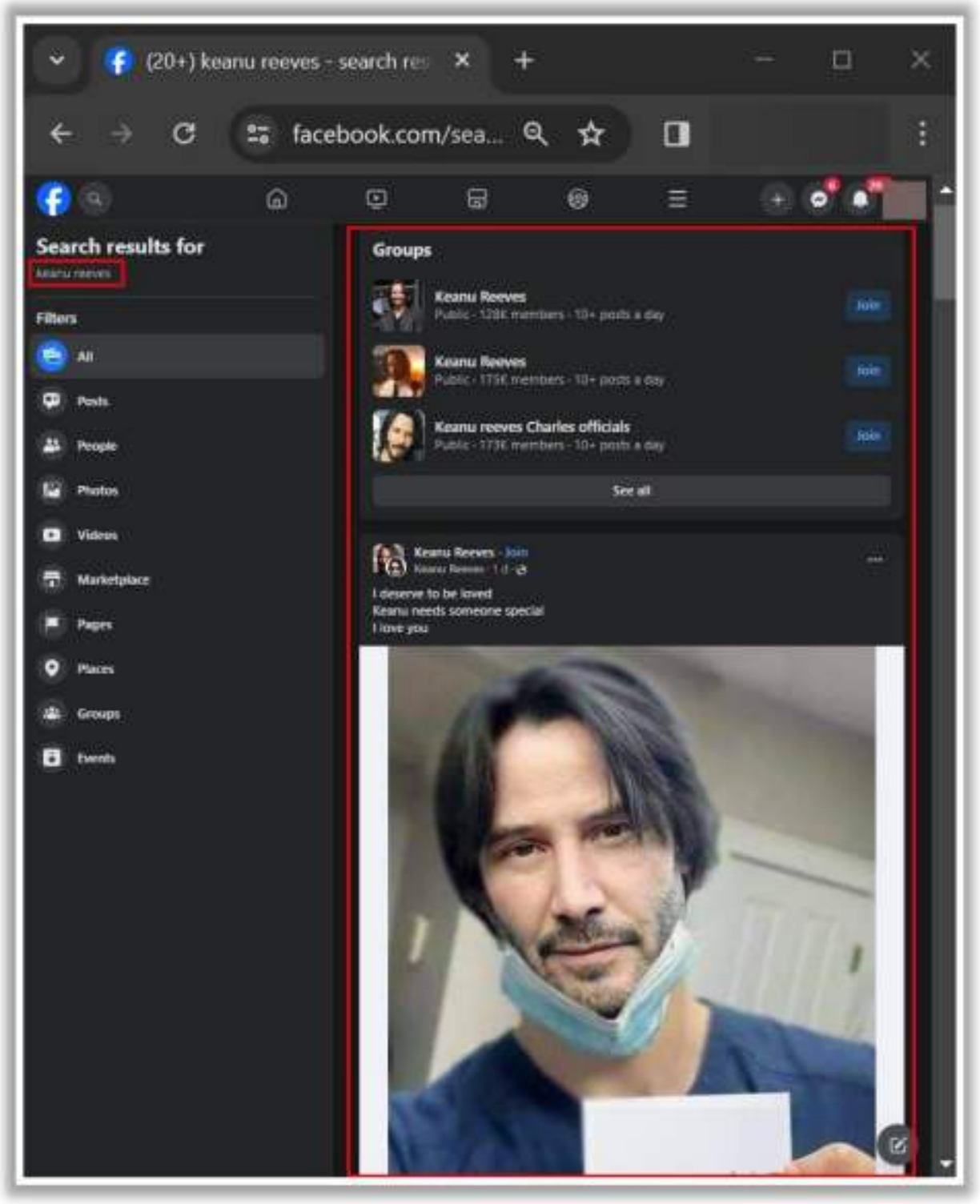
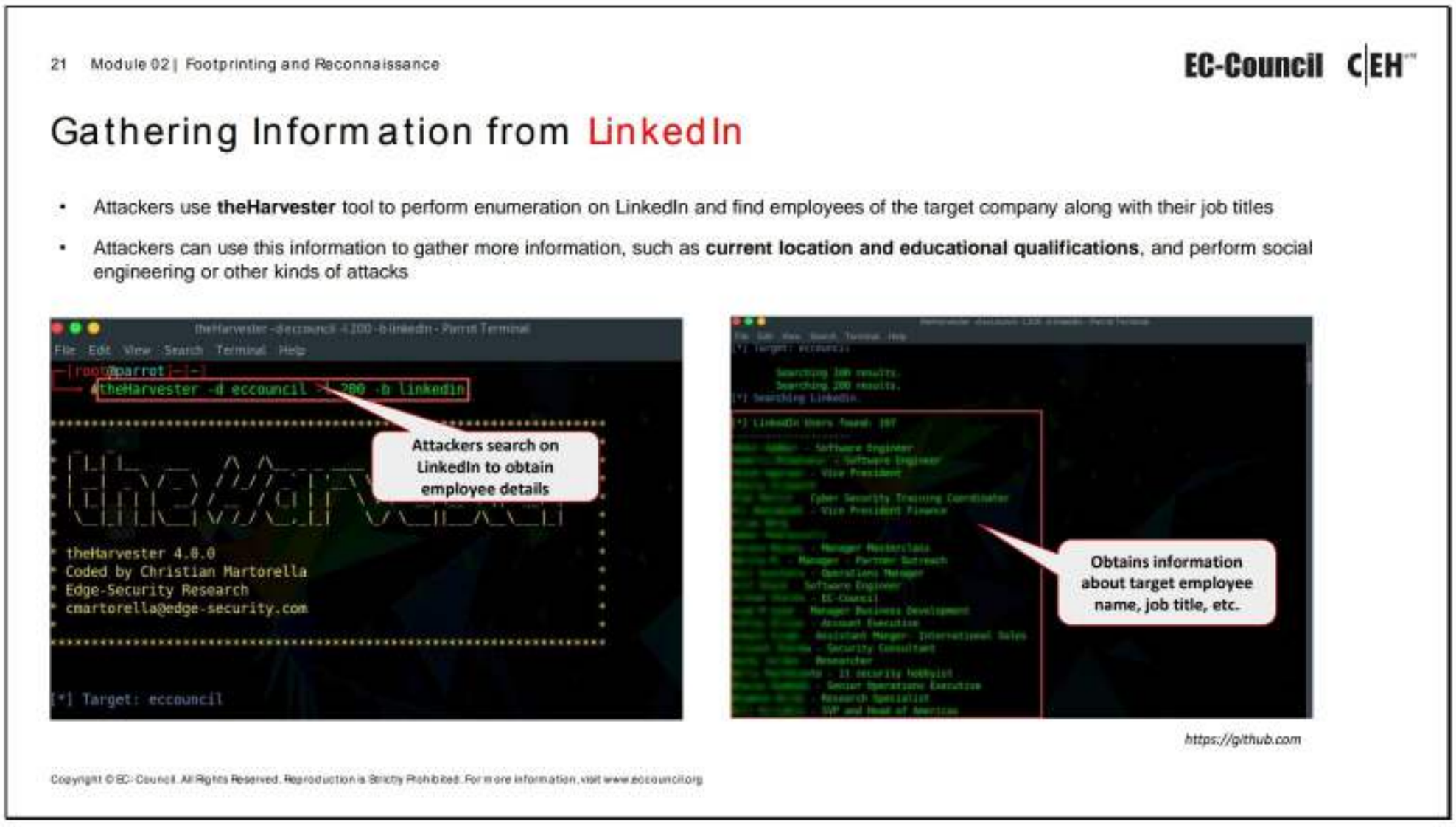


Figure 2.48: Screenshot of Facebook showing search results



Gathering Information from LinkedIn

LinkedIn is a social networking website for professionals. It connects the world's human resources to aid productivity and success. The site contains personal information such as name, position, organization name, current location, educational qualifications, and so on. Information gathered from LinkedIn helps an attacker in performing social engineering or other kinds of attacks.

Attackers can use theHarvester tool to gather information from LinkedIn based on the target organization name:

- **theHarvester**

  Source: *https://github.com*

  theHarvester is a tool designed to be used in the early stages of a penetration test. It is used for open-source intelligence gathering and helps to determine a company's external threat landscape on the Internet. Attackers use this tool to perform enumeration on the LinkedIn social networking site to find employees of the target company along with their job titles.

  As shown in the screenshot, the attacker uses the following command to enumerate users on LinkedIn:

  ```
  theHarvester -d microsoft -l 200 -b linkedin
  ```

  In the above command, `-d` specifies the domain or company name to search, `-l` specifies the number of results to be retrieved, and `-b` specifies the data source as LinkedIn.

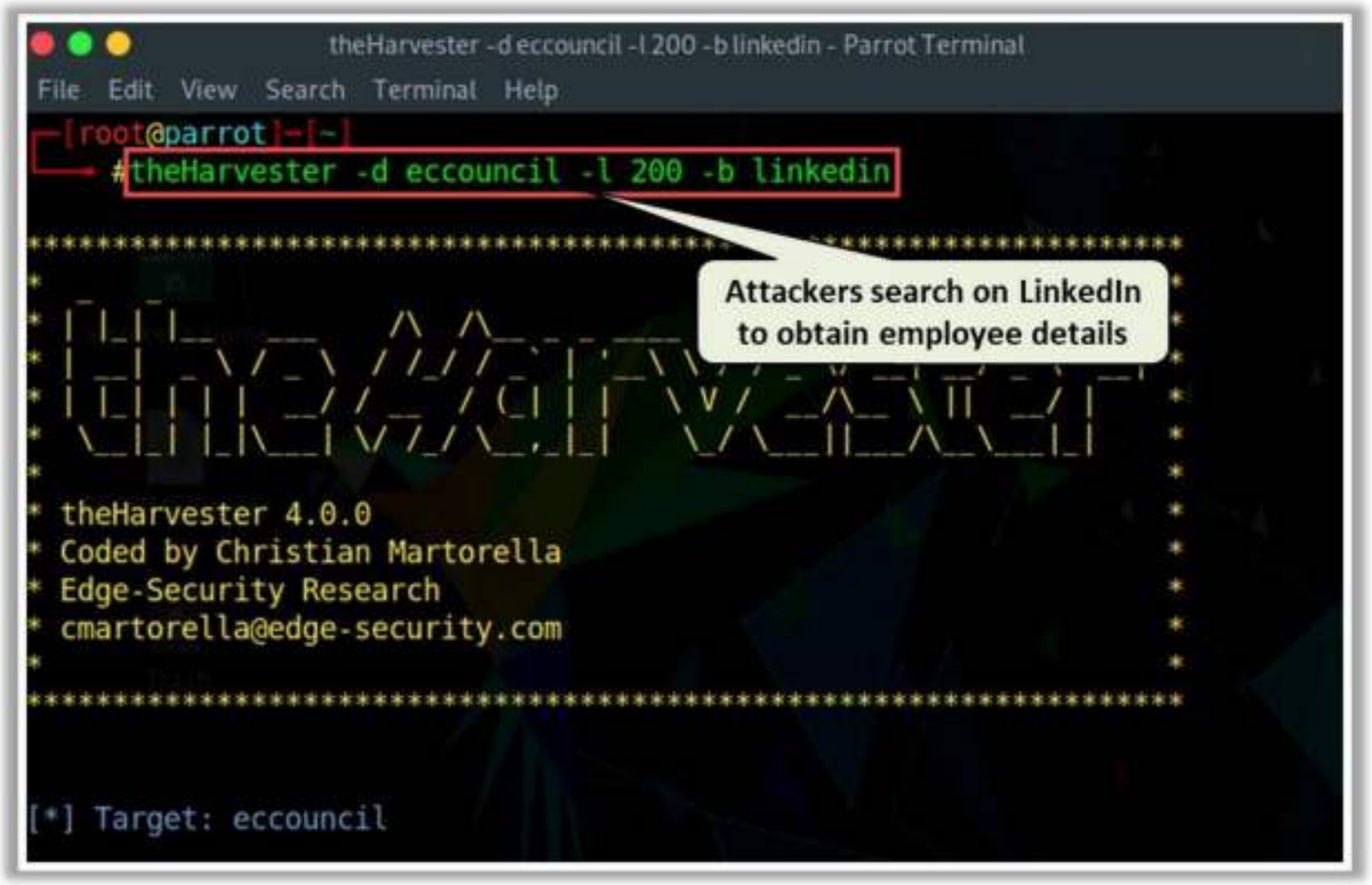


Figure 2.49: Screenshot showing theHarvester command to enumerate users on LinkedIn

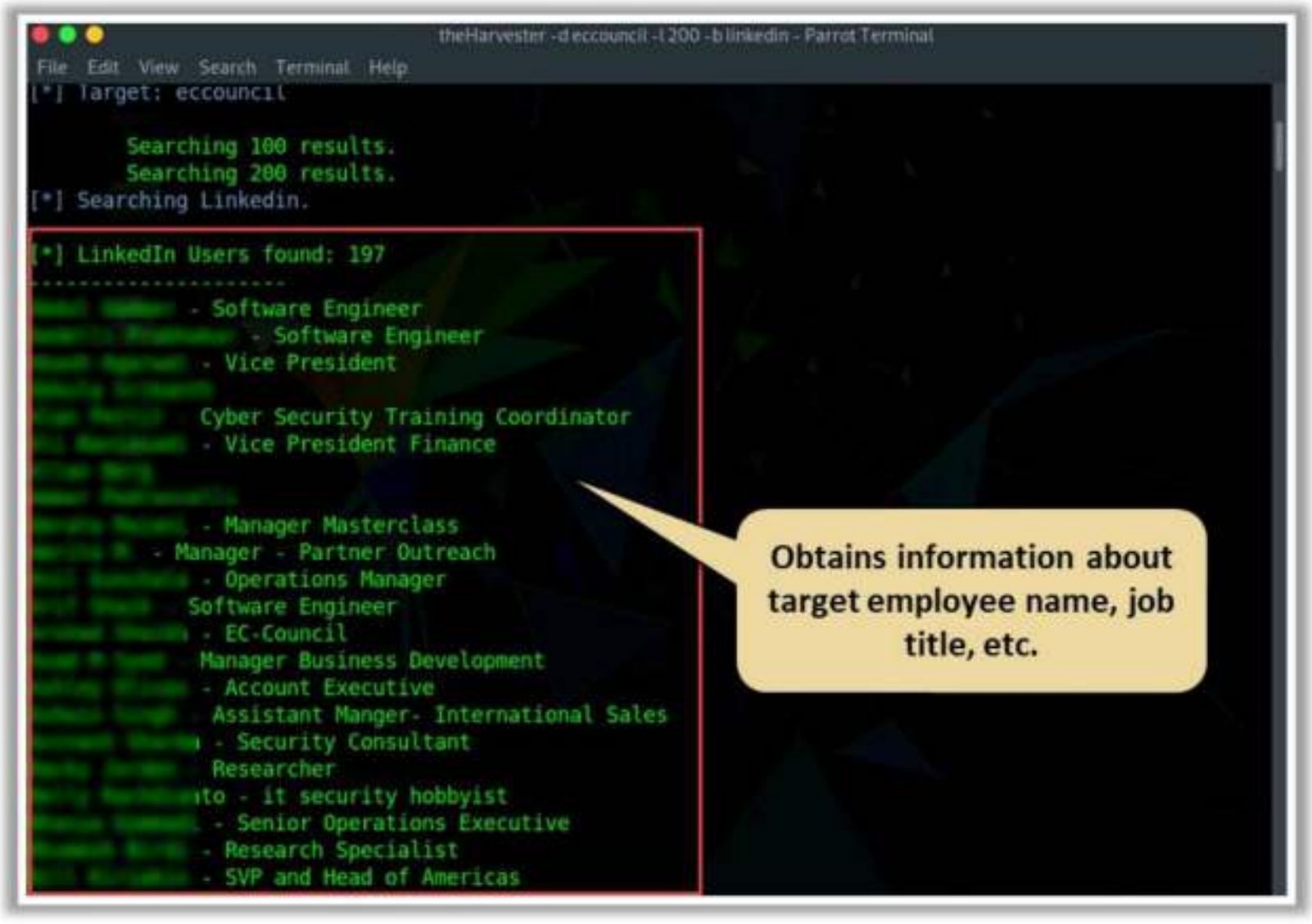


Figure 2.50: Screenshot showing theHarvester search results from LinkedIn

Harvesting Email Lists with AI

- Attackers use automated tools such as **theHarvester** and **Email Spider** to collect publicly available email addresses of the target organization that helps them perform social engineering and brute-force attacks
- An attacker can also leverage AI-powered ChatGPT or other generative AI technology to perform this task by using an appropriate prompt such as

  "Use theHarvester to gather email accounts associated with microsoft.com, limiting results to 200, and leveraging 'baidu' as a data source"

Harvesting Email Lists

Gathering email addresses related to the target organization acts as an important attack vector during the later phases of hacking. Attackers can use automated tools such as theHarvester and Email Spider to collect publicly available email addresses of the employees of the target organization. These tools harvest email lists related to a specified domain using search engines such as Google, Bing, and Yahoo. Attackers use these email lists and usernames to perform social engineering and brute force attacks on the target organization.

- **theHarvester**

  Source: *https://github.com*

  Attackers use theHarvester tool to extract email addresses related to the target domain. For example, attackers use the following command to extract email addresses of microsoft.com using the Baidu search engine:

  ```
  theharvester -d microsoft.com -l 200 -b baidu
  ```

  In the above command, -d specifies the domain used for harvesting the emails, -l will limit the results to 200, and -b tells theHarvester to extract the results from the Baidu search engine; alternatively, you can use Google, Bing, etc.

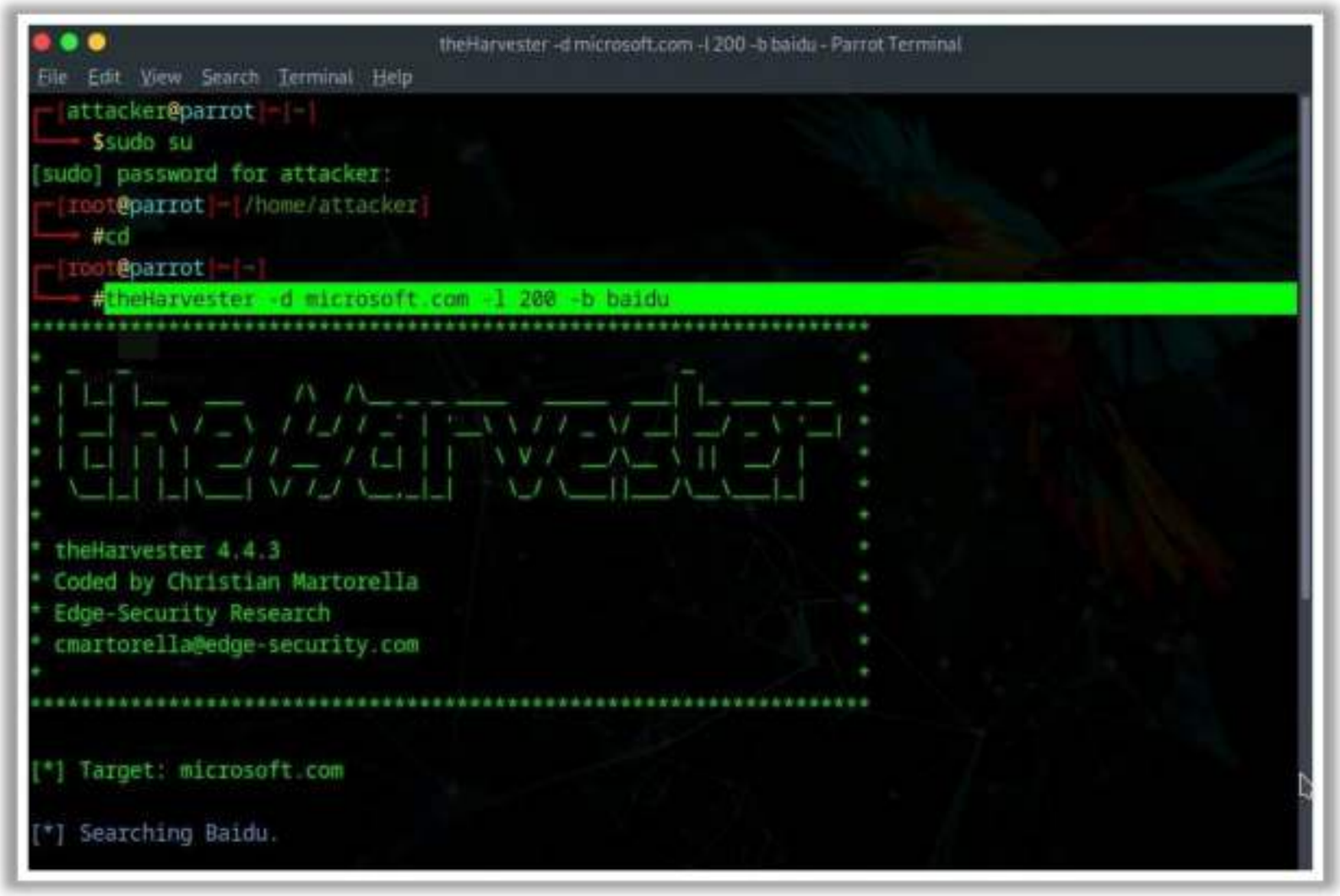


Figure 2.51: Screenshot showing theHarvester command to extract email addresses

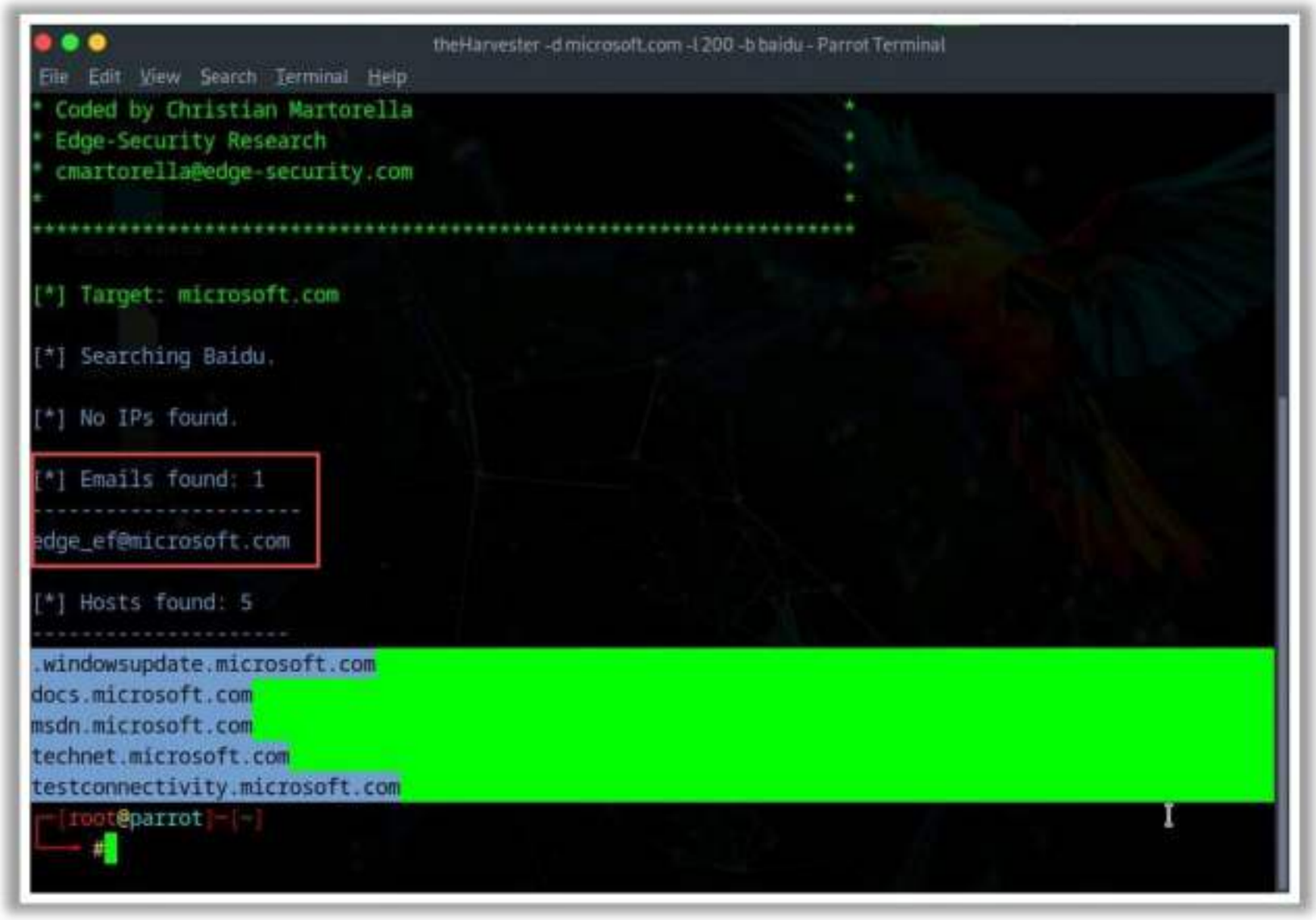


Figure 2.52: Screenshot showing the email list extracted by theHarvester

Harvesting Email Lists with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly find valuable email details of their target organization.

For example,

An attacker can use ChatGPT to perform this task by using an appropriate prompt such as:

"Use theHarvester to gather email accounts associated with 'microsoft.com', limiting results to 200, and leveraging 'baidu' as a data source"

Figure 2.53: Use theHarvester to gather email accounts associated with 'microsoft.com'

To gather email accounts associated with "microsoft.com" using theHarvester, limiting results to 200 and leveraging "Baidu" as a data source, you can use the following command:

```
theHarvester -d microsoft.com -l 200 -b Baidu -f Microsoft_emails.xml
```

- `` `theHarvester` ``: This is the command to invoke theHarvester tool.
- `` `-d microsoft.com` ``: This parameter specifies the target domain, in this case, "microsoft.com", for which we want to gather email accounts.
- `` `-l 200` ``: This parameter limits the number of results to 200, ensuring a manageable output.
- `` `-b baidu` ``: This parameter specifies "baidu" as the data source to be utilized for gathering email accounts associated with "microsoft.com". Baidu is a popular Chinese search engine and can provide valuable information in this context.

- `-f Microsoft_emails.xml`: This parameter specifies the output file name where the email accounts will be saved. In this case, the file is named "Microsoft_emails.xml", and the email accounts will be saved in XML format.

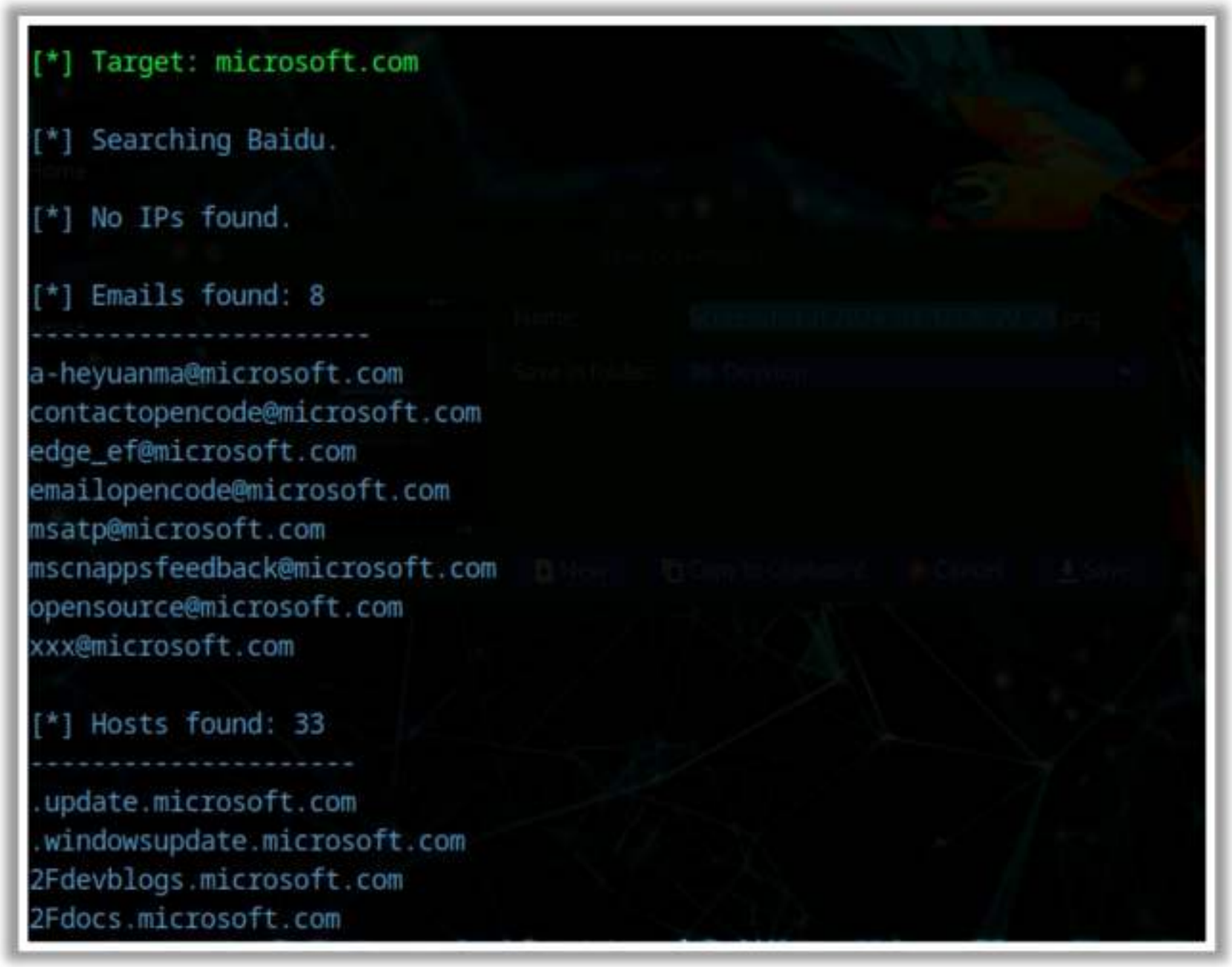
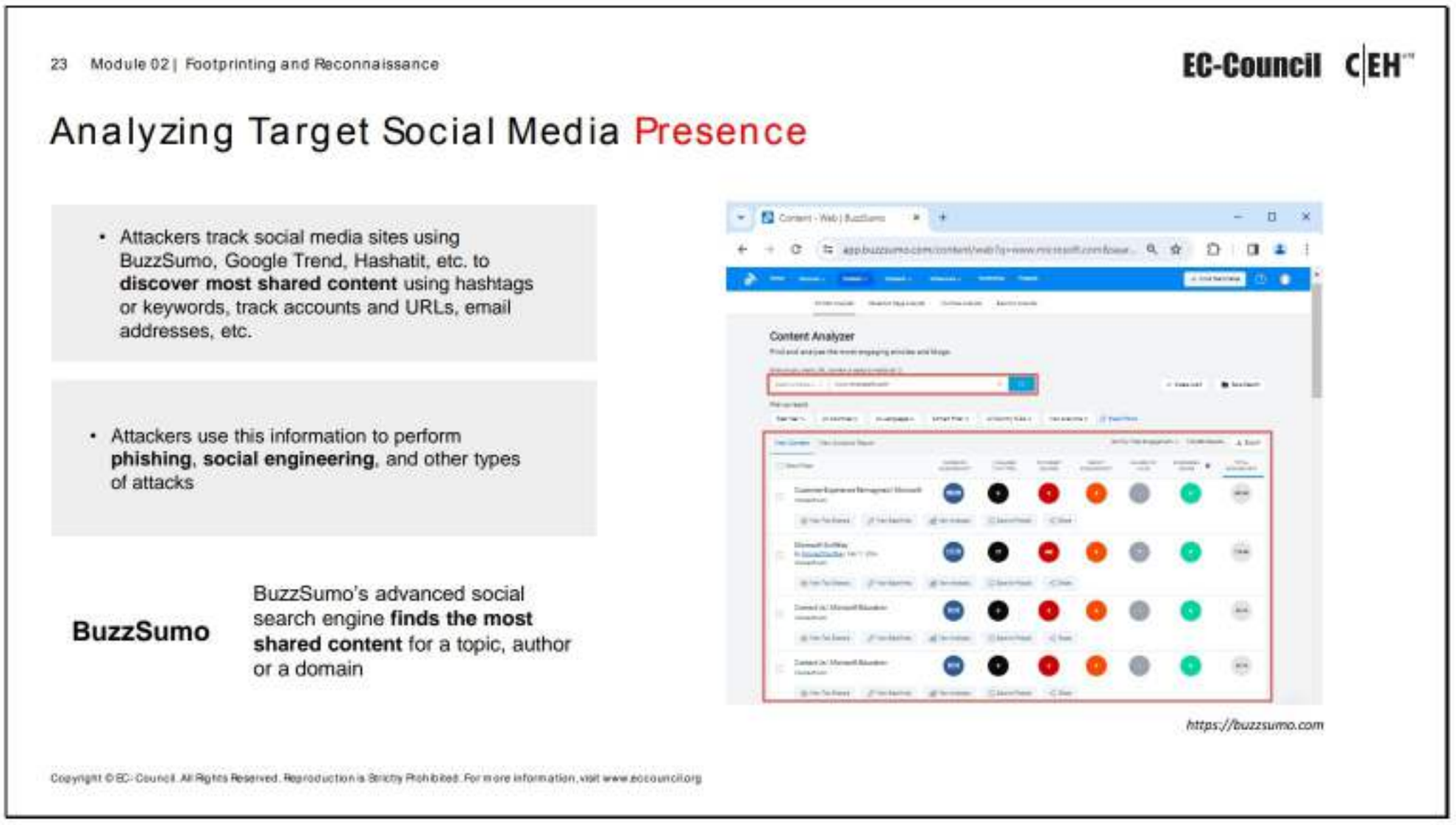


Figure 2.54: To gather email accounts associated with 'microsoft.com' using theHarvester



Analyzing Target Social Media Presence

Several online services and resources are available to gather valuable information about a target from one or more social media sites. These services allow attackers to discover most shared content across social media sites by using hashtags or keywords, track accounts and URLs on various social media sites, obtain a target's email address, etc. This information helps attackers to perform phishing, social engineering, and other types of attacks.

Attackers use tools such as BuzzSumo, Google Trends, Hashatit, and Ubersuggest to locate information on social media sites:

- **BuzzSumo**

  Source: *https://buzzsumo.com*

  BuzzSumo's advanced social search engine finds the most shared content for a topic, author, or domain. It shows the shared activity across all the major social networks including Twitter, Facebook, LinkedIn, Google Plus, and Pinterest.

  As shown in the screenshot, attackers use BuzzSumo to track the most shared content related to the target domain and obtain details such as social media account information, URLs, and email addresses.

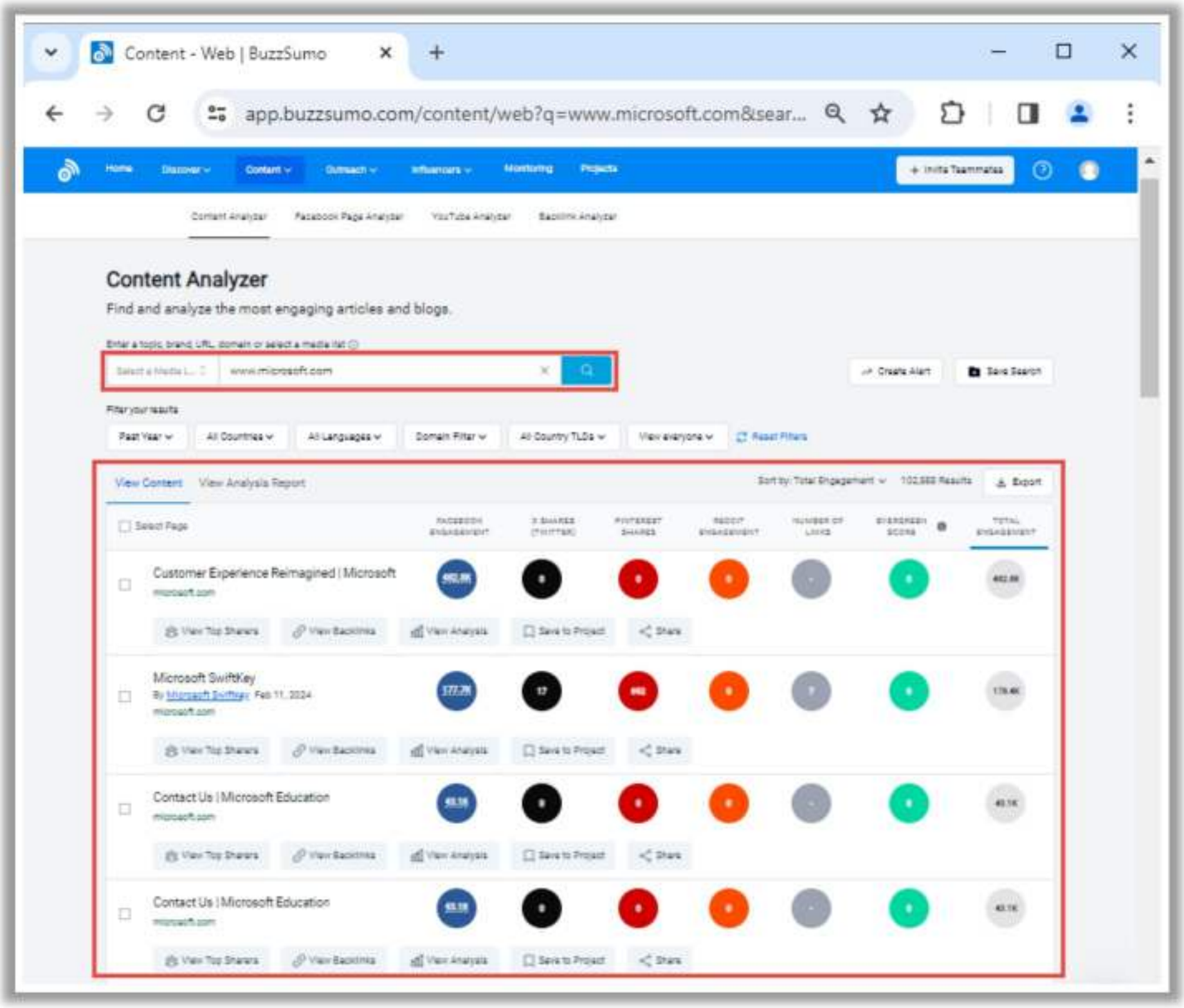
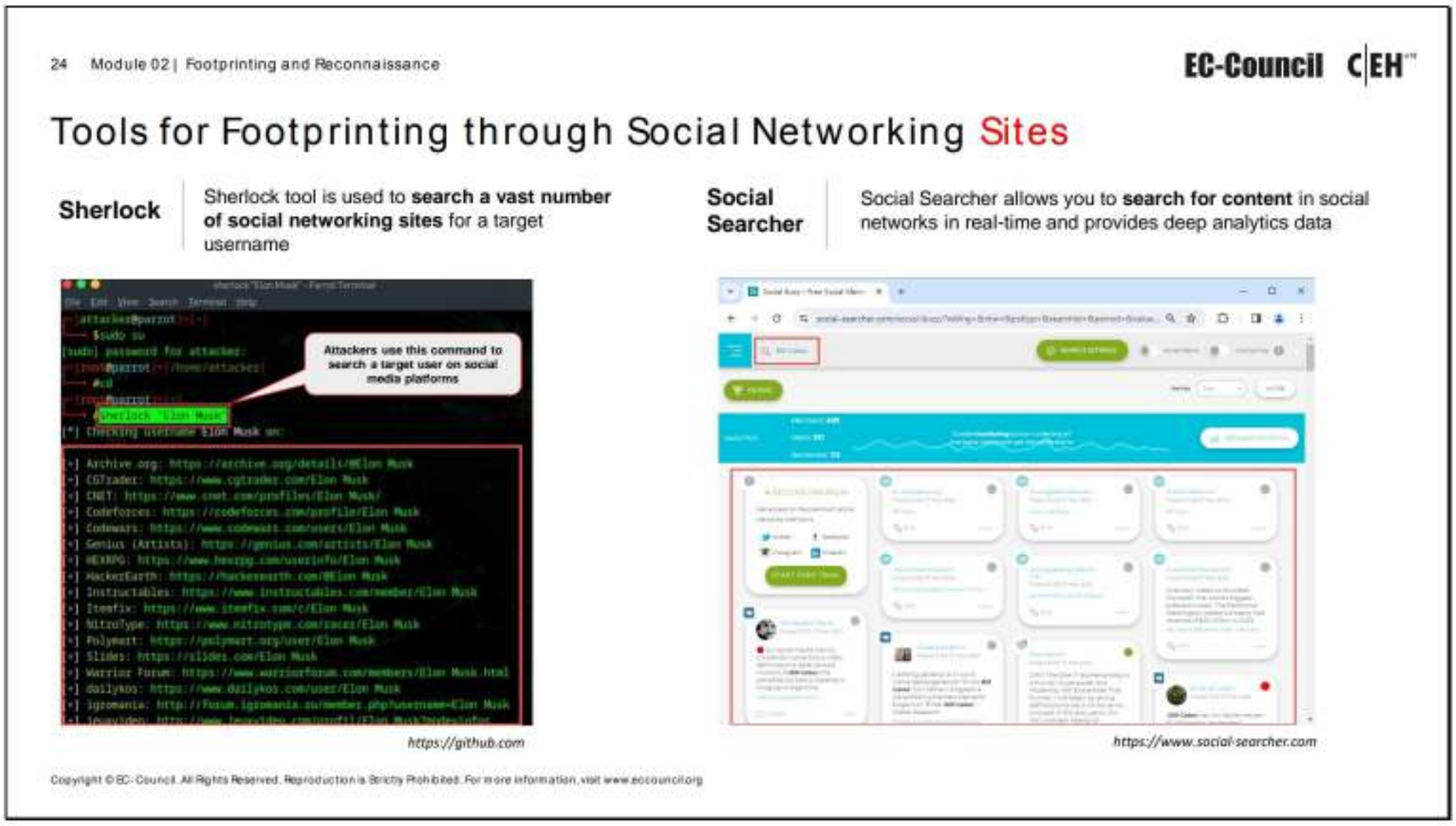


Figure 2.55: Screenshot of BuzzSumo showing the shared content



Tools for Footprinting through Social Networking Sites

Attackers use various tools such as Sherlock and Social Searcher to footprint social networking sites such as Twitter, Instagram, Facebook, and Pinterest to gather sensitive information about the target such as the date of birth, educational qualification, employment status, name of relatives, and information about the organization that they are working for, including the business strategy, potential clients, and upcoming project plans.

- **Sherlock**

  Source: *https://github.com*

  As shown in the screenshot, attackers use Sherlock to search a vast number of social networking sites for a target username. This tool helps the attacker locate the target user on various social networking sites, along with the complete URL.

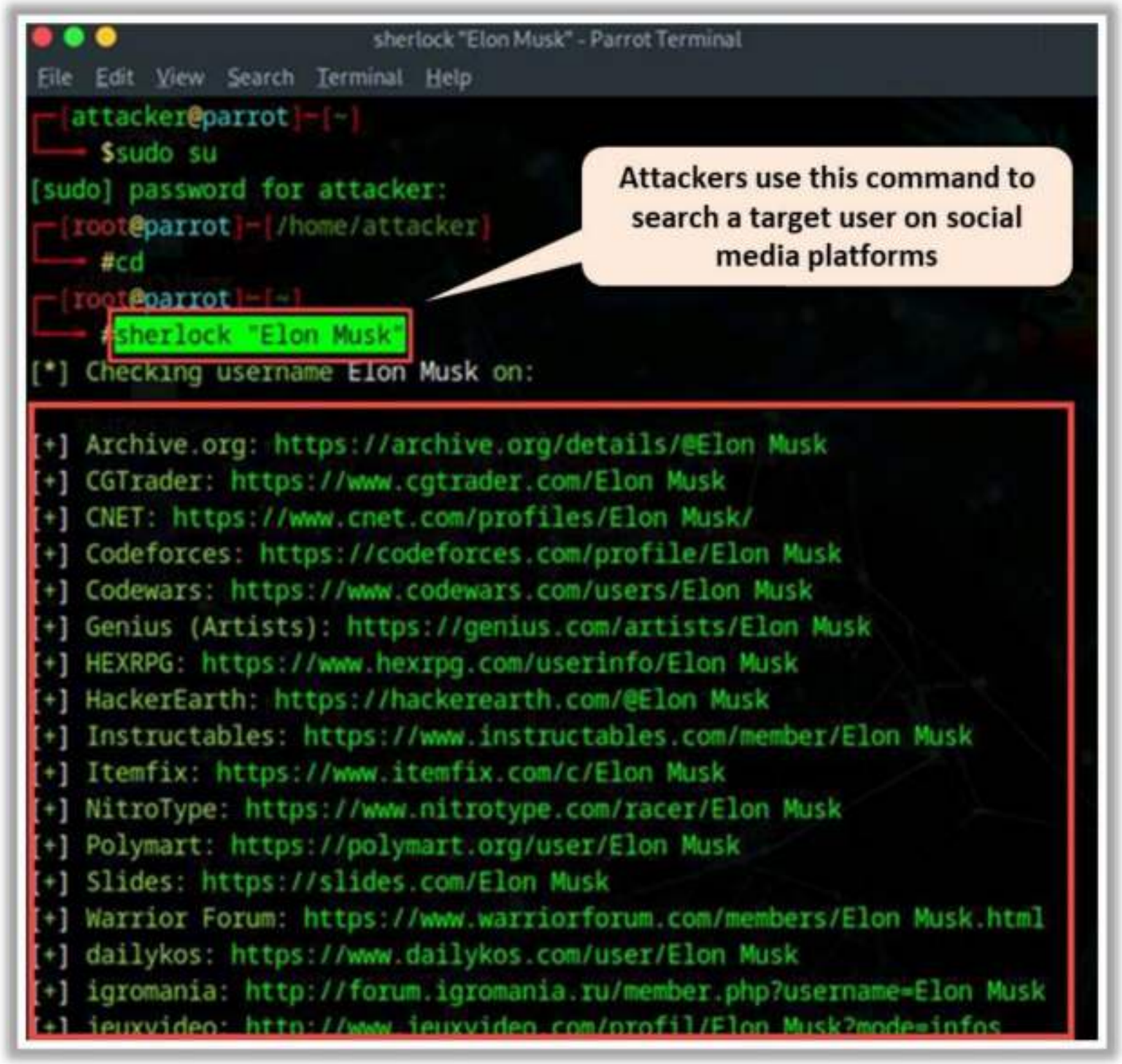


Figure 2.56: Screenshot showing the result of Sherlock tool

- **Social Searcher**

  Source: *https://www.social-searcher.com*

  Social Searcher allows attackers to search for content on social networks in real time and provides deep analytics data. Attackers use this tool to track a target user on various social networking sites and obtain information such as complete URLs to their profiles, their postings, and other personal information.

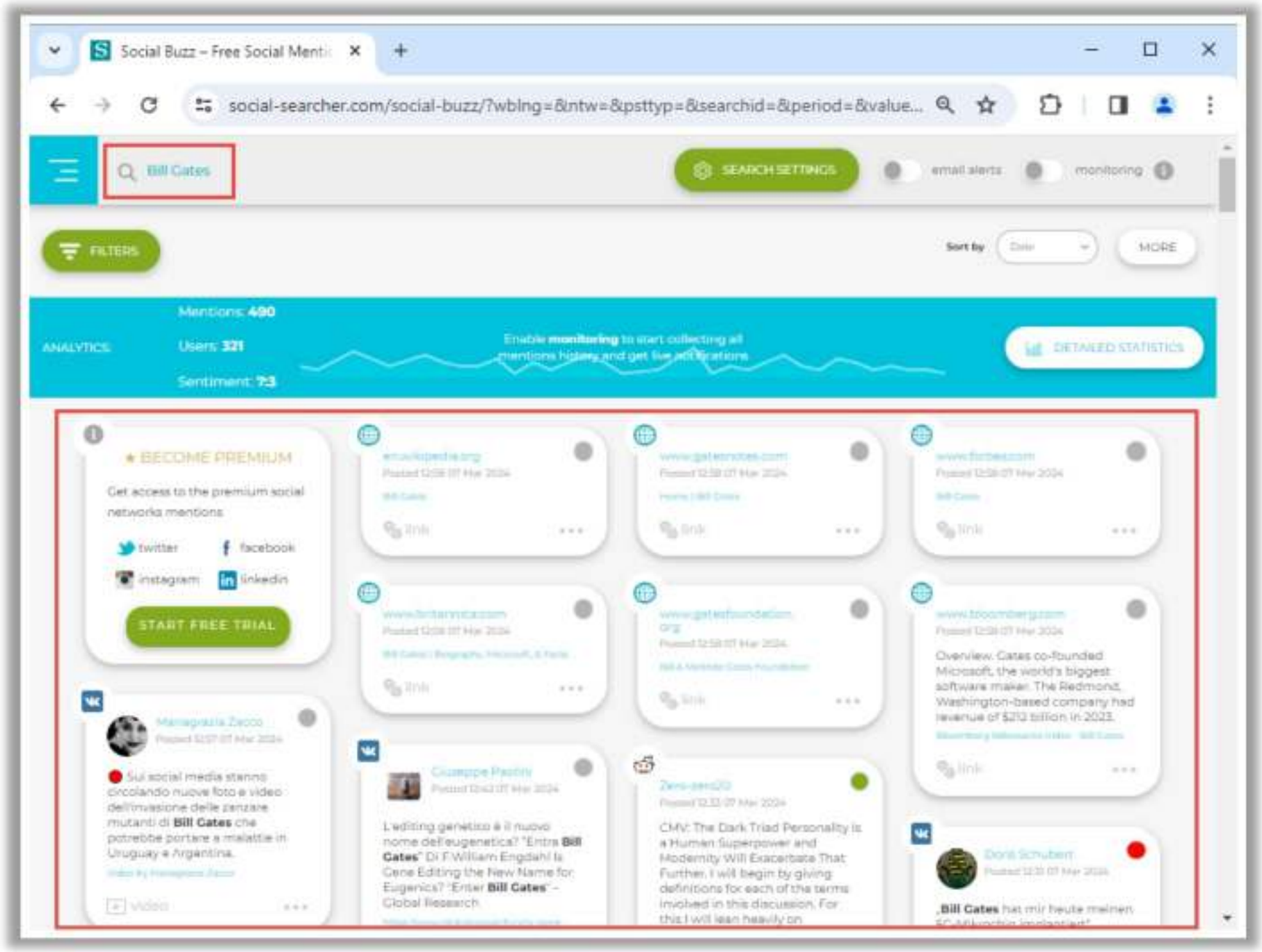


Figure 2.57: Screenshot of Social Searcher showing user content on social networks

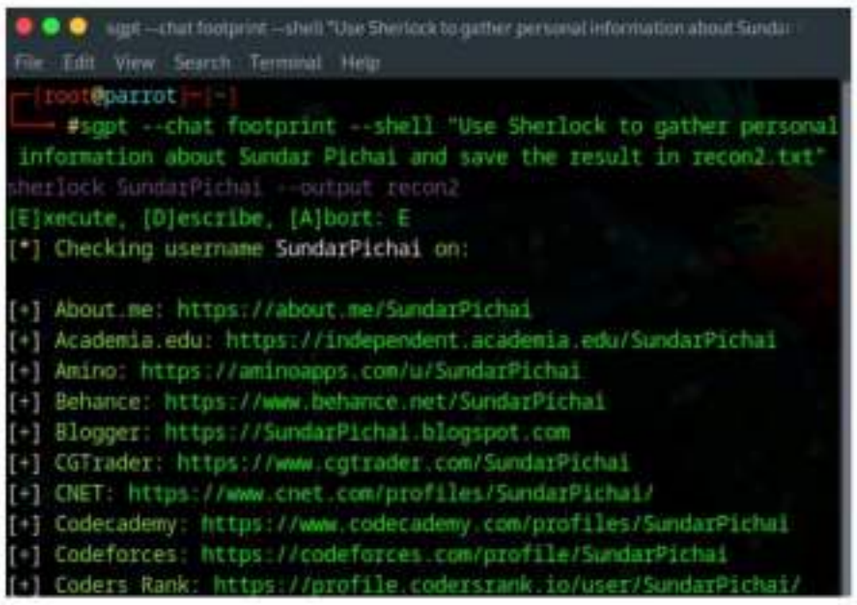
25 Module 02 | Footprinting and Reconnaissance

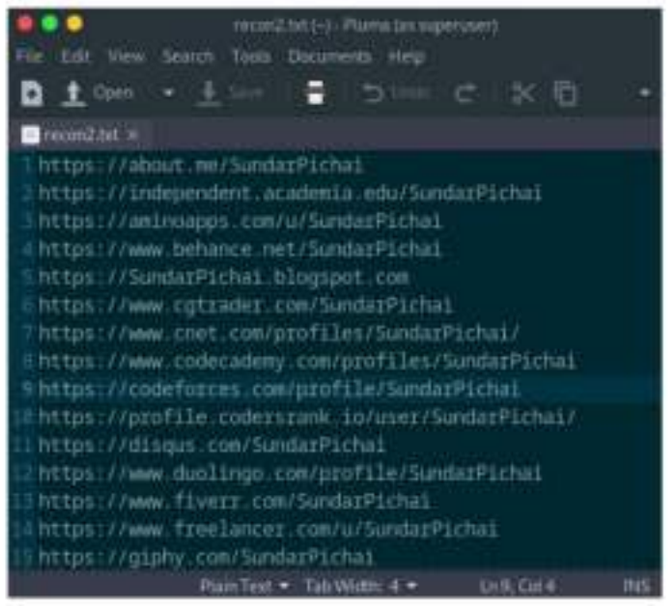
EC-Council CEH™

Footprinting through Social Networking Sites with AI

- An attacker can also leverage **AI-powered ChatGPT** or other generative AI technology to perform this task by using an appropriate prompt such as

  "Use Sherlock to gather personal information about Sundar Pichai and save the result in recon2.txt"





Footprinting through Social Networking Sites with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly use social networking sites to acquire valuable insights about their targets.

For example,

An attacker can use ChatGPT to perform this task by using an appropriate prompt such as:

"Use Sherlock to gather personal information about Sundar Pichai and save the result in recon2.txt"

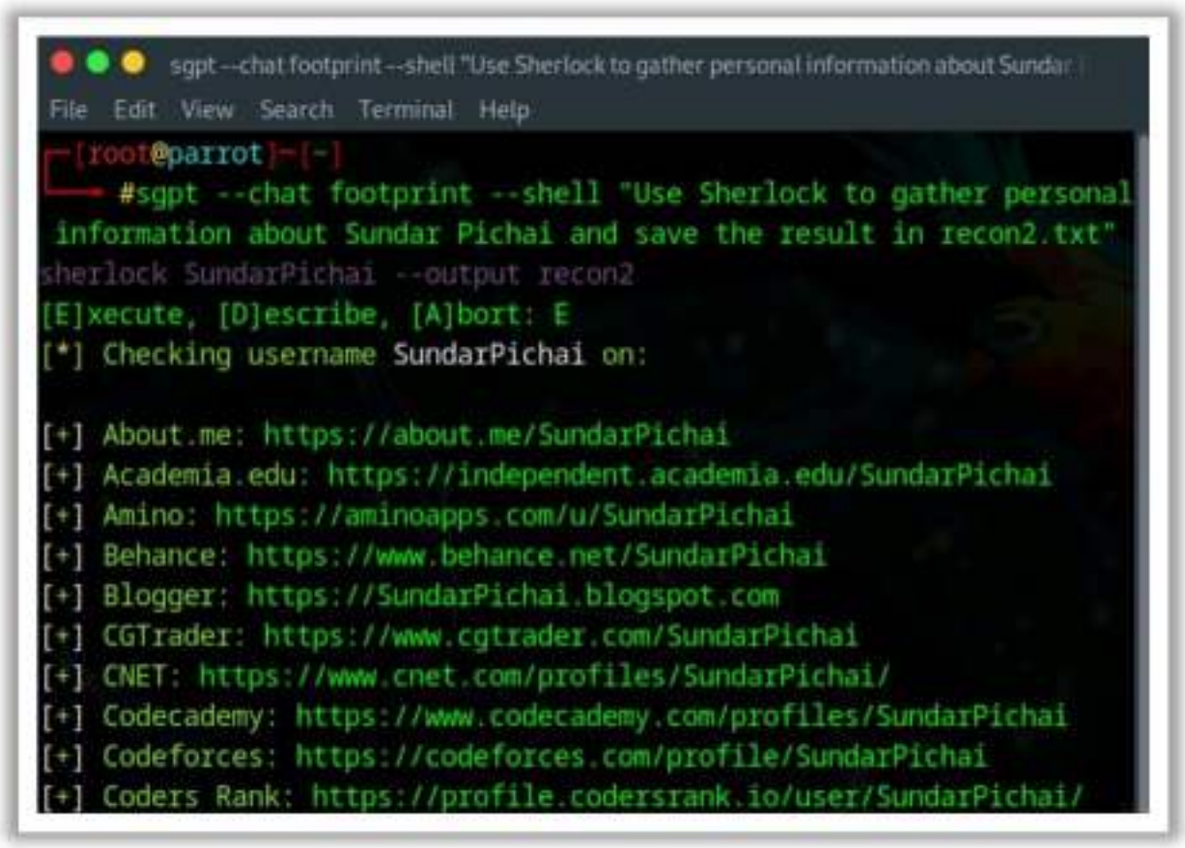


Figure 2.58: Search for online accounts associated with the name "Sundar Pichai"

To perform a search for online accounts associated with the name "Sundar Pichai" using Sherlock and save the output to a file named "recon2", you can use the following command:

```
sherlock SundarPichai - - output recon2
```

- `sherlock`: This command runs Sherlock, a tool for searching for online accounts across various platforms.
- `SundarPichai`: This is the name for which we want to search online accounts.
- `--output recon2`: This option specifies the output file where the results will be saved. In this case, the file is named "recon2".

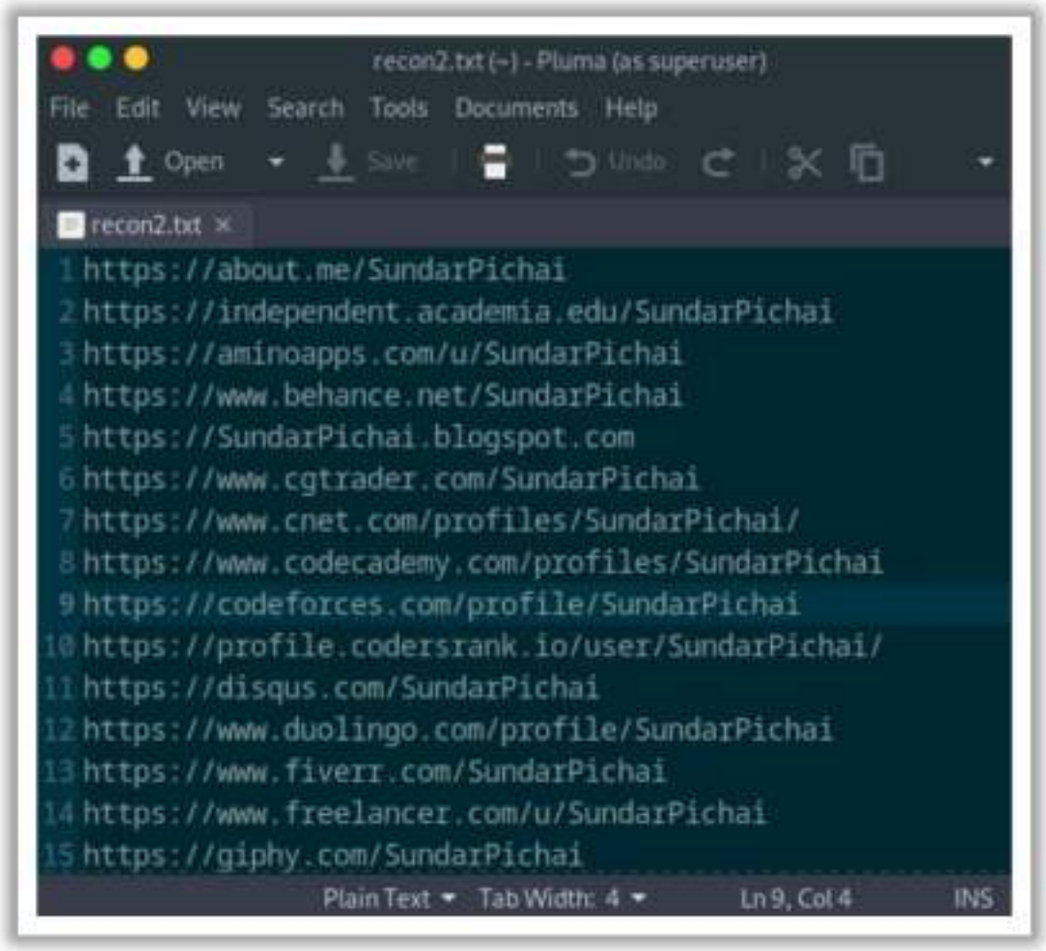


Figure 2.59: Output file

Whois Footprinting

Gathering network-related information such as "Whois" information about the target organization is important when planning an attack. In this section, we will discuss Whois footprinting, which helps in gathering domain information such as information regarding the owner of an organization, its registrar, registration details, its name server, and contact information. Whois footprinting focuses on how to perform a Whois lookup, analyze the Whois lookup results, and find IP geolocation information, as well as the tools used to gather Whois information.

Whois Lookup

Whois databases are maintained by **Regional Internet Registries** and contain **personal information of domain owners**

Whois query returns

- Domain name details
- Contact details of domain owners
- Domain name servers
- NetRange
- When a domain was created
- Expiry records
- Last updated record

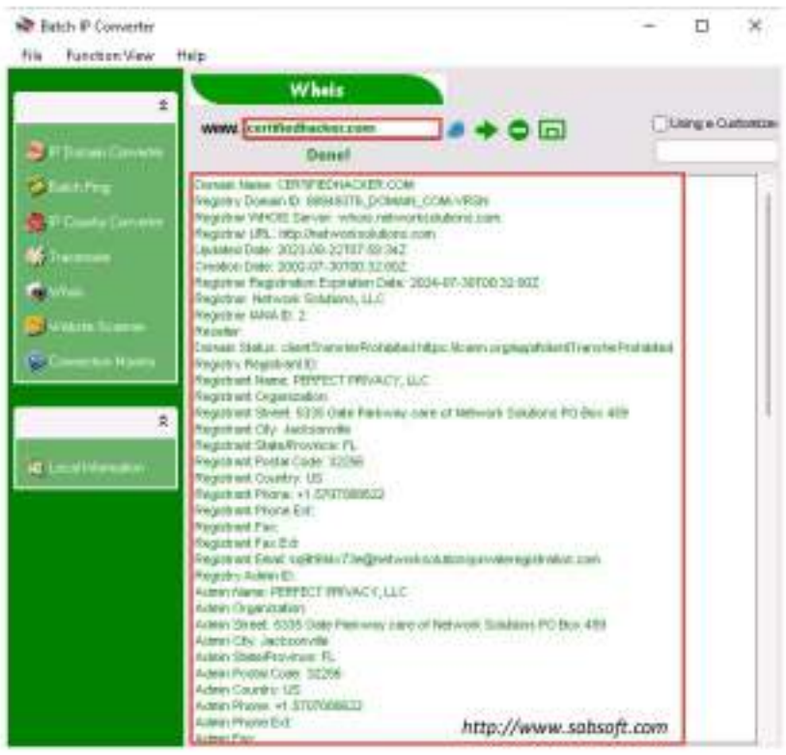
Information obtained from Whois database assists an attacker to

- Gather personal information that assists in social engineering
- Create a map of the target organization's network
- Obtain internal details of the target network

Regional Internet Registries (RIRs)

ARIN
AFRINIC
RIPE NCC
lacnic
APNIC

Whois Lookup (Cont'd)



Whois Lookup

Whois is a query and response protocol used for querying databases that store the registered users or assignees of an Internet resource, such as a domain name, an IP address block, or an autonomous system. This protocol listens to requests on port 43 (TCP). Regional Internet Registries (RIRs) maintain Whois databases, which contain the personal information of domain owners. For each resource, the Whois database provides text records with information about the

resource itself and relevant information regarding assignees, registrants, and administrative information (creation and expiration dates).

Three types of data models exist to store and lookup Whois information:

- **Thick Whois (Distributed Model)** - Stores the complete Whois information from all the registrars for a particular set of data.
- **Thin Whois (Centralized Model)** - Stores only the name of the Whois server of the registrar of a domain, which in turn holds complete details on the data being looked up.
- **Decentralized Whois -** Stores complete WHOIS information and has multiple independent entities to manage the WHOIS database.

Whois query returns the following information:

- Domain name details
- Domain registrar
- Contact details of the domain owner
- Domain name servers
- NetRange
- When a domain has been created
- Expiry records
- Records last updated
- Domain status (available, registered, or suspended)
- IP address information

An attacker can query a Whois database server to obtain information regarding the target domain, and the Whois server responds to the query with the requested information. Using this information, an attacker can create a map of the organization's network, mislead domain owners with social engineering, and obtain internal details of the network.

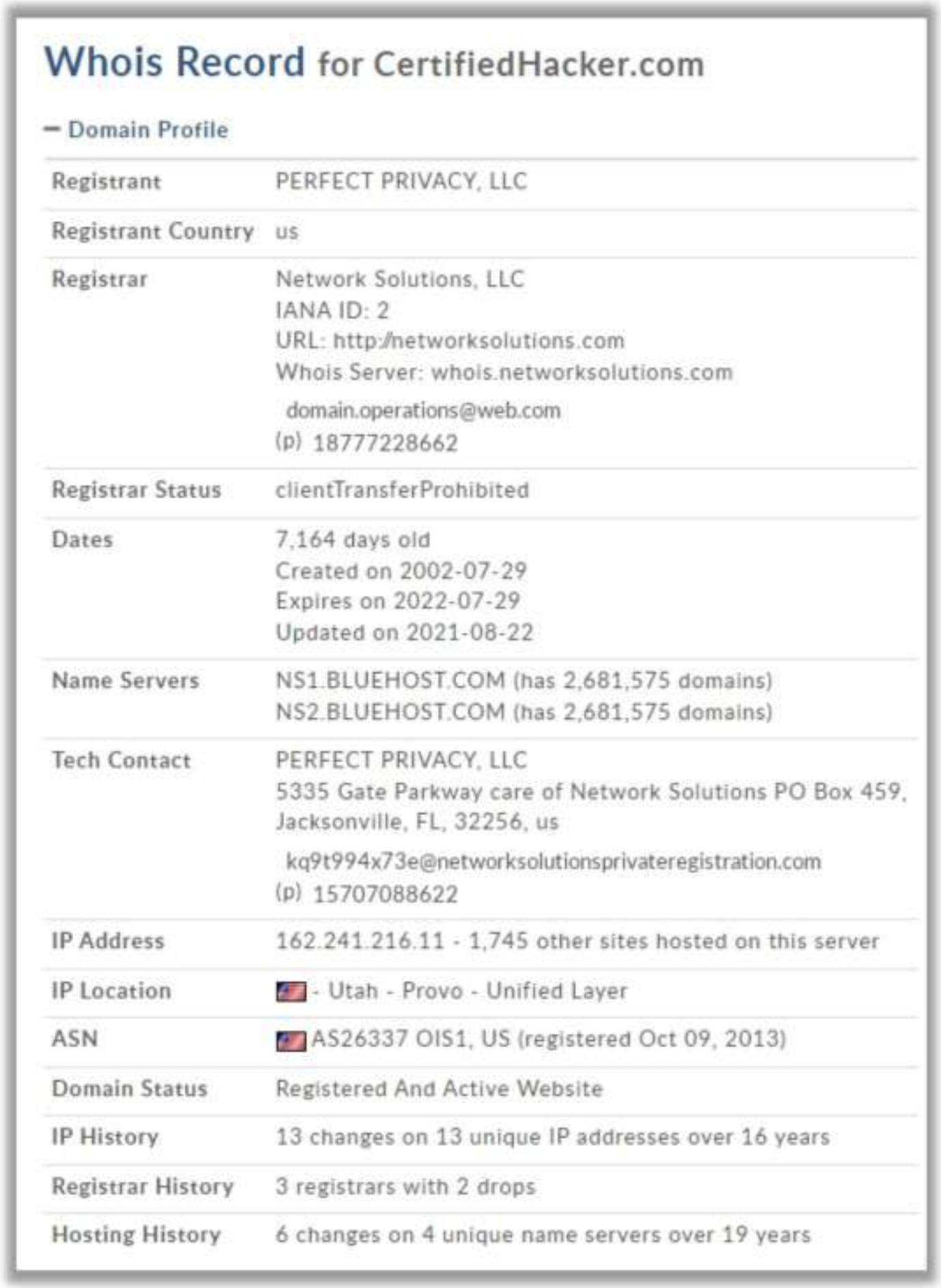
Regional Internet Registries (RIRs)

The RIRs include the following:

- American Registry for Internet Numbers (ARIN) (*https://www.arin.net*)
- African Network Information Center (AFRINIC) (*https://www.afrinic.net*)
- Asia Pacific Network Information Center (APNIC) (*https://www.apnic.net*)
- Réseaux IP Européens Network Coordination Centre (RIPE) (*https://www.ripe.net*)
- Latin American and Caribbean Network Information Center (LACNIC) (*https://www.lacnic.net*)

Whois Lookup Result

Whois services such as *https://whois.domaintools.com* and *https://www.tamos.com* can help perform Whois lookups. The screenshot below shows the result of a Whois lookup obtained with each of the two aforementioned Whois services. The services perform a Whois lookup by entering the target domain or IP address. Batch IP Converter, available at *http://www.sabsoft.com*, provides information about an IP address, hostname, or domain, including information about the country, state or province, city, phone number, fax number, name of the network provider, administrator, and technical-support contact information. It supports Internationalized Domain Names (IDNs), implying that one can query domain names that use non-English characters. It also supports IPv6 addresses.



Whois Record for CertifiedHacker.com

− Domain Profile

Registrant	PERFECT PRIVACY, LLC
Registrant Country	us
Registrar	Network Solutions, LLC IANA ID: 2 URL: http://networksolutions.com Whois Server: whois.networksolutions.com domain.operations@web.com (p) 18777228662
Registrar Status	clientTransferProhibited
Dates	7,164 days old Created on 2002-07-29 Expires on 2022-07-29 Updated on 2021-08-22
Name Servers	NS1.BLUEHOST.COM (has 2,681,575 domains) NS2.BLUEHOST.COM (has 2,681,575 domains)
Tech Contact	PERFECT PRIVACY, LLC 5335 Gate Parkway care of Network Solutions PO Box 459, Jacksonville, FL, 32256, us kq9t994x73e@networksolutionsprivateregistration.com (p) 15707088622
IP Address	162.241.216.11 - 1,745 other sites hosted on this server
IP Location	- Utah - Provo - Unified Layer
ASN	AS26337 OIS1, US (registered Oct 09, 2013)
Domain Status	Registered And Active Website
IP History	13 changes on 13 unique IP addresses over 16 years
Registrar History	3 registrars with 2 drops
Hosting History	6 changes on 4 unique name servers over 19 years

Figure 2.60: Screenshot of Whols

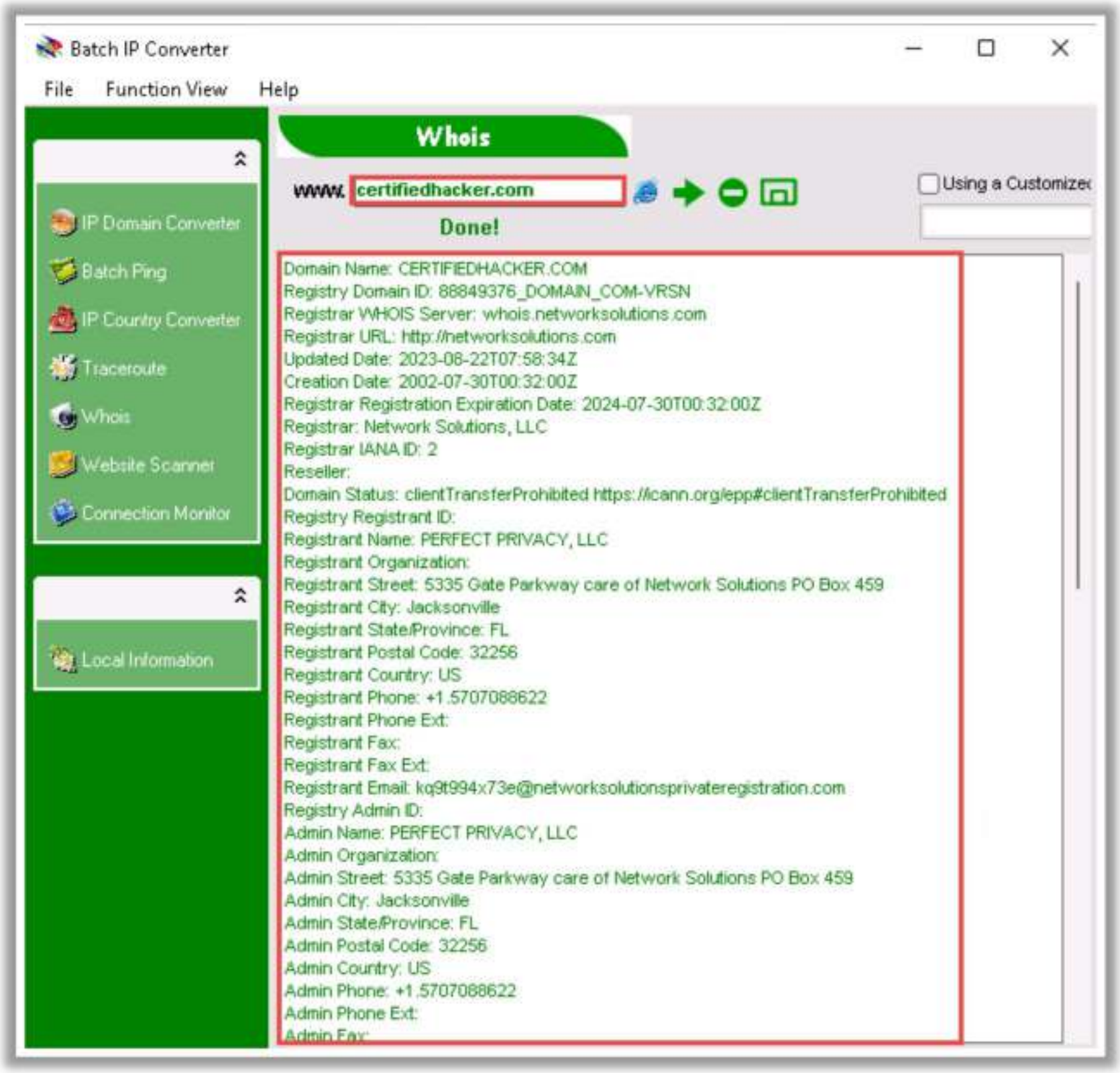
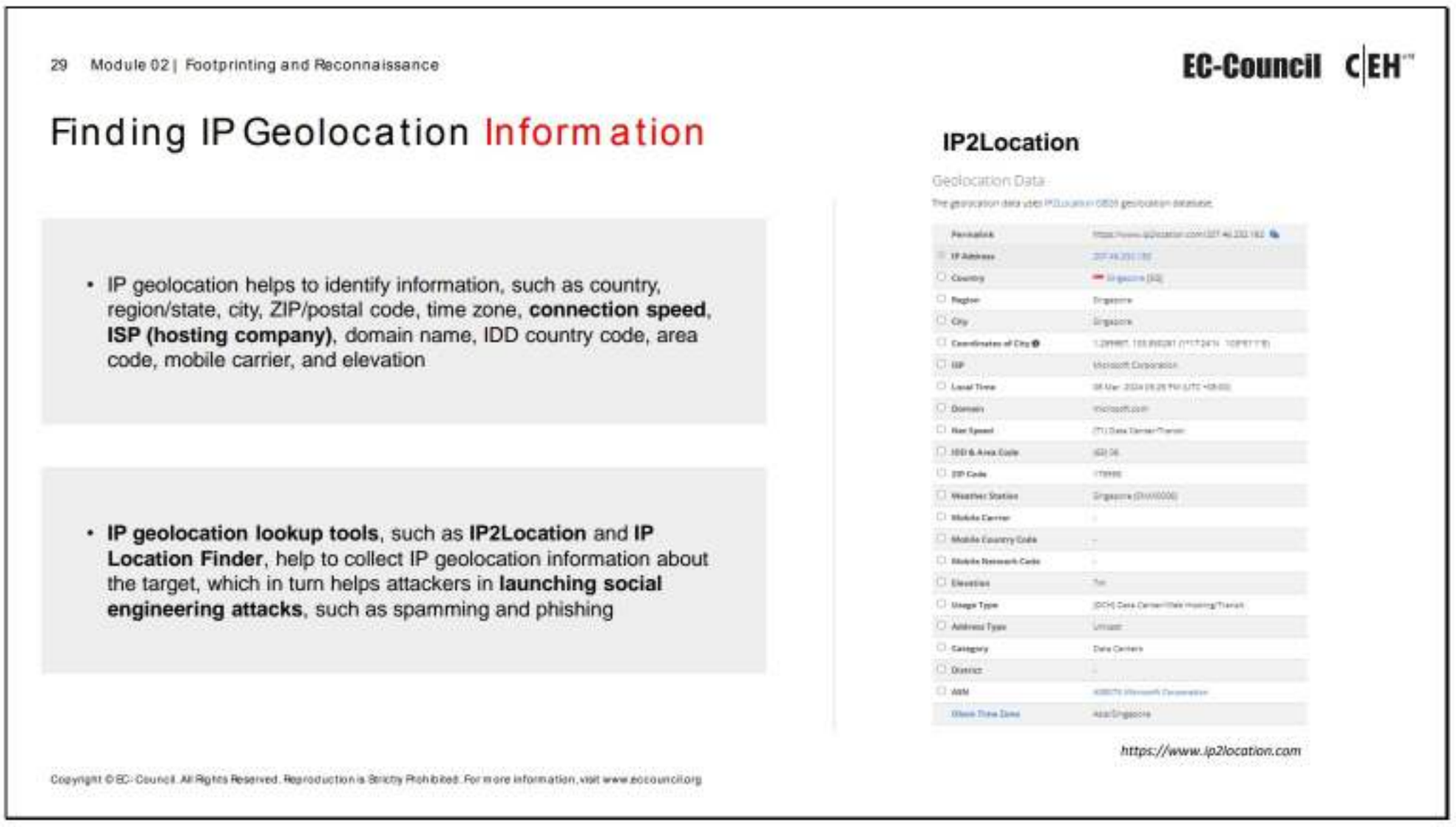


Figure 2.61: Screenshot of Batch IP Converter

Attackers also use Whois lookup tools such as WHOIS Domain Lookup and Active Whois to perform a Whois lookup on the target domain.



Finding IP Geolocation Information

IP geolocation helps to obtain information regarding a target such as its country, region/state, city, latitude and longitude of its city, ZIP/postal code, time zone, connection speed, ISP (hosting company), domain name, IDD country code, area code, weather station code and name, mobile carrier, and elevation.

Using the information obtained from IP geolocation, an attacker may attempt to gather more information about a target with the help of social engineering, surveillance, and non-technical attacks such as dumpster diving, hoaxing, or acting as a technical expert. With the help of the information obtained, an attacker can also set up a compromised web server near the victim's location, and if the exact location of the victim is detected, the attacker can perform malicious activities and infect the victim with malware designed for that specific area or gain unauthorized access to the target device or attempt to launch an attack on the target device.

IP geolocation lookup tools such as IP2Location, IP Location Finder, and IP Address Geographical Location Finder help to collect IP geolocation information about the target, which enables attackers to launch social engineering attacks such as spamming and phishing.

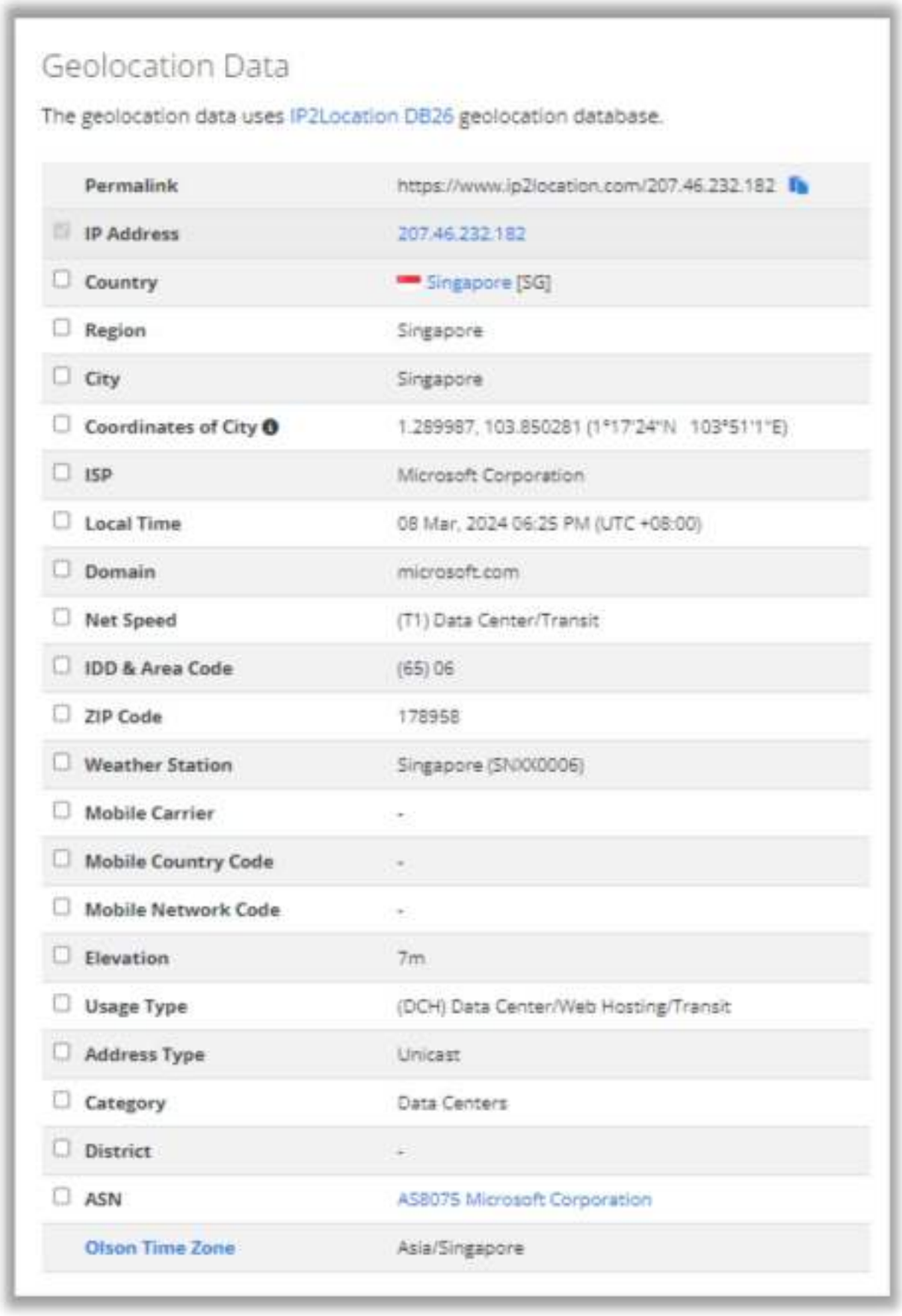
IP Geolocation Lookup Tools

- **IP2Location**

  Source: *https://www.ip2location.com*

  As shown in the screenshot, attackers use IP2Location tool to identify a visitor's geographical location, i.e., country, region, city, latitude and longitude of city, ZIP code, time zone, connection speed, ISP, domain name, IDD country code, area code, weather

station code and name, mobile carrier, elevation, and usage type information using a proprietary IP address lookup database and technology.

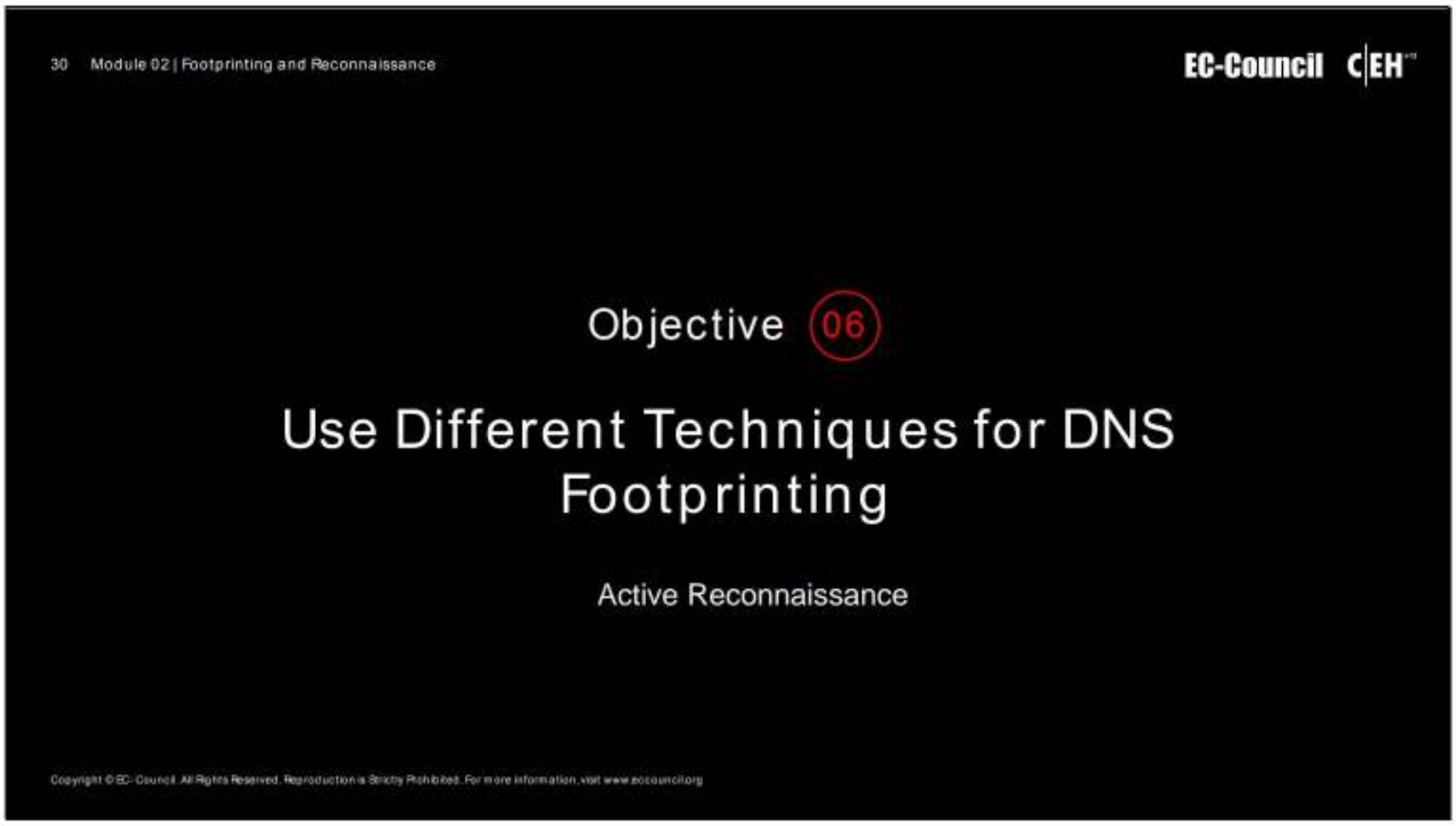


Geolocation Data

The geolocation data uses IP2Location DB26 geolocation database.

Permalink	https://www.ip2location.com/207.46.232.182
IP Address	207.46.232.182
Country	Singapore [SG]
Region	Singapore
City	Singapore
Coordinates of City	1.289987, 103.850281 (1°17'24"N 103°51'1"E)
ISP	Microsoft Corporation
Local Time	08 Mar, 2024 06:25 PM (UTC +08:00)
Domain	microsoft.com
Net Speed	(T1) Data Center/Transit
IDD & Area Code	(65) 06
ZIP Code	178958
Weather Station	Singapore (SN0000006)
Mobile Carrier	-
Mobile Country Code	-
Mobile Network Code	-
Elevation	7m
Usage Type	(DCH) Data Center/Web Hosting/Transit
Address Type	Unicast
Category	Data Centers
District	-
ASN	AS8075 Microsoft Corporation
Olson Time Zone	Asia/Singapore

Figure 2.62: Screenshot of IP2Location



DNS Footprinting

After collecting the Whois records of the target, the next phase of the footprinting methodology is Domain Name System (DNS) footprinting. Attackers perform DNS footprinting to gather information about DNS servers, DNS records, and the types of servers used by the target organization. This information helps attackers identify the hosts connected in the target network and further exploit the target organization. This section describes how to extract DNS information and perform reverse DNS lookups using various DNS interrogation tools.

EC-Council CEH™

Extracting DNS Information

- DNS records provide important information about the **location and types of servers**
- Attackers can gather DNS information to **determine key hosts in the network** and can perform social engineering attacks

Record Type	Description
A	Points to a host's IP address
MX	Points to domain's mail server
NS	Points to host's name server
CNAME	Canonical naming allows aliases to a host
SOA	Indicate authority for a domain
SRV	Service records
PTR	Maps IP address to a hostname
RP	Responsible person
HINFO	Host information record includes CPU type and OS
TXT	Unstructured text records

- Attackers query DNS servers using DNS interrogation tools, such as SecurityTrails, Fierce, DNSChecker, and zdns, to **retrieve the record structure** that contains information about the target DNS

Extracting DNS Information

DNS footprinting reveals information about DNS zone data. DNS zone data include DNS domain names, computer names, IP addresses, and much more information about a network. An attacker uses DNS information to determine key hosts in the network and then performs social engineering attacks to gather even more information.

DNS footprinting helps in determining the following records about the target DNS:

Record Type	Description
A	Points to a host's IP address
AAAA	Points to a host's IPv6 address
MX	Points to domain's mail server
NS	Points to host's name server
CNAME	Canonical naming allows aliases to a host
SOA	Indicate authority for a domain
SRV	Service records
PTR	Maps IP address to a hostname
RP	Responsible person
HINFO	Host information record includes CPU type and OS
TXT	Unstructured text records

Table 2.4: DNS records and their description

DNS Interrogation Tools

Attackers use DNS interrogation tools such as SecurityTrails, Fierce, DNSChecker, zdns and DNSdumpster.com to perform DNS footprinting. These tools can extract a range of IP addresses using IP routing lookup. If the target network allows unknown, unauthorized users to transfer DNS zone data, it is easy for an attacker to obtain DNS information with the help of a DNS interrogation tool.

When an attacker queries a DNS server using a DNS interrogation tool, the server responds with a record structure that contains information about the target DNS. DNS records provide important information regarding the locations and types of servers.

- **SecurityTrails**

  Source: *https://securitytrails.com*

  SecurityTrails is an advanced DNS enumeration tool capable of creating a DNS map of the target domain network. It can enumerate both current and historical DNS records such as A, AAAA, NS, MX, SOA, and TXT, which helps in building the DNS structure. It also enumerates all the existing subdomains of the target domain using brute-force techniques.

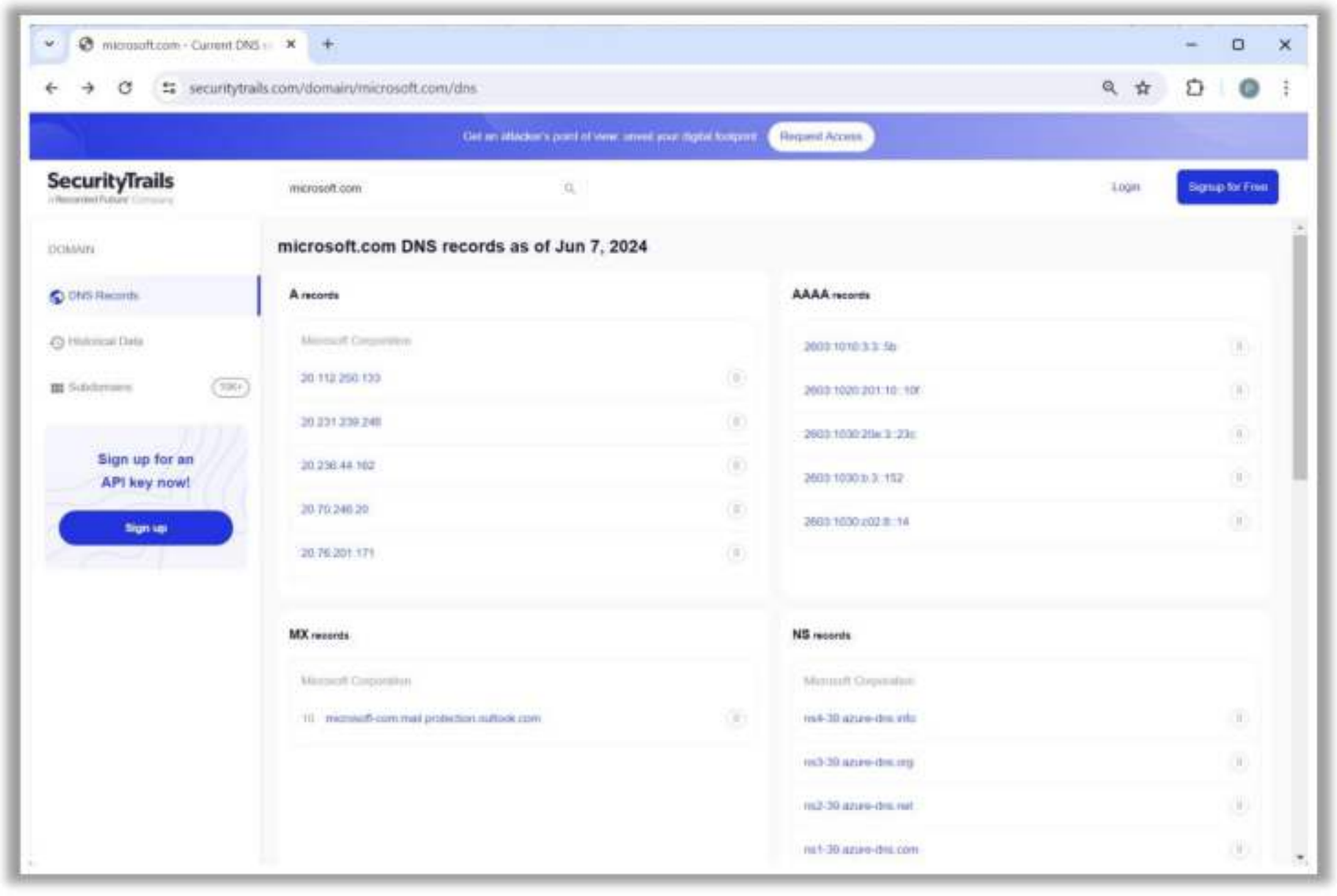


Figure 2.63: Screenshot of SecurityTrails

- **Fierce**

  Source: *https://github.com*

  Fierce is a DNS reconnaissance tool used for scanning and collecting crucial information about the target domain. Attackers can use this tool to enumerate subdomains related to the target domain. It also enables them to identify non-contiguous IP spaces and hostnames that are linked with specified domains or subdomains. By gathering this information, attackers can create a network environment and identify the potential targets for exploitation.

  Attackers can use the following commands to perform DNS reconnaissance using the Fierce tool:

  - Run the following command to start a basic scan on the target domain (certifiedhacker.com) without any additional options:

    `fierce --domain certifiedhacker.com`

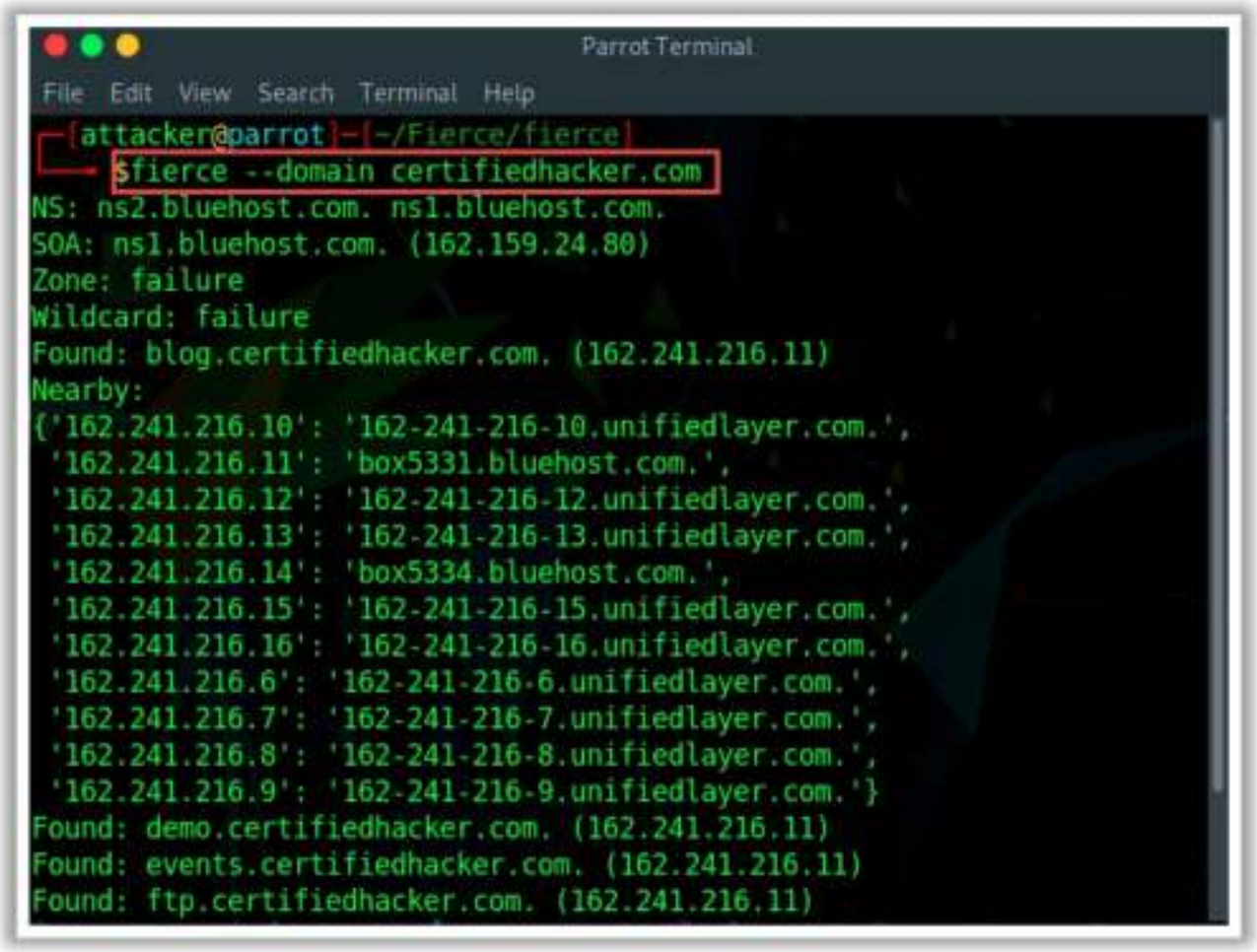


Figure 2.64: Screenshot showing reconnaissance of the targeted domain

  - Run the following command to scan the target domain to obtain specific subdomains (here, subdomains containing words such as write, admin, and mail):

    `fierce -domain certifiedhacker.com -subdomains write admin mail`

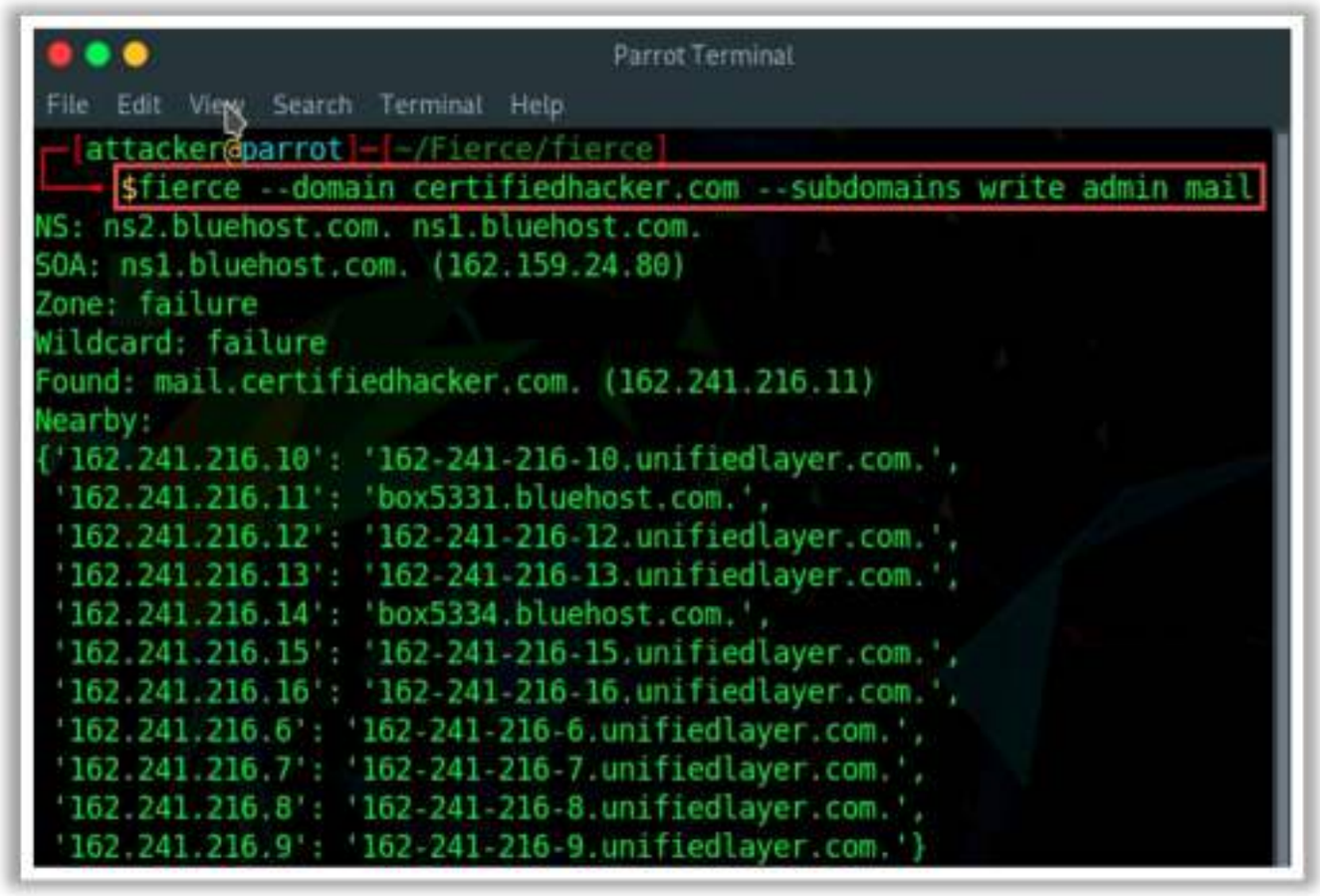


Figure 2.65: Screenshot showing a simple scan on targeted domain

- Run the following command to scan domains near the discovered records of the targeted domain:

  `fierce -domain certifiedhacker.com -subdomains mail -traverse 10`

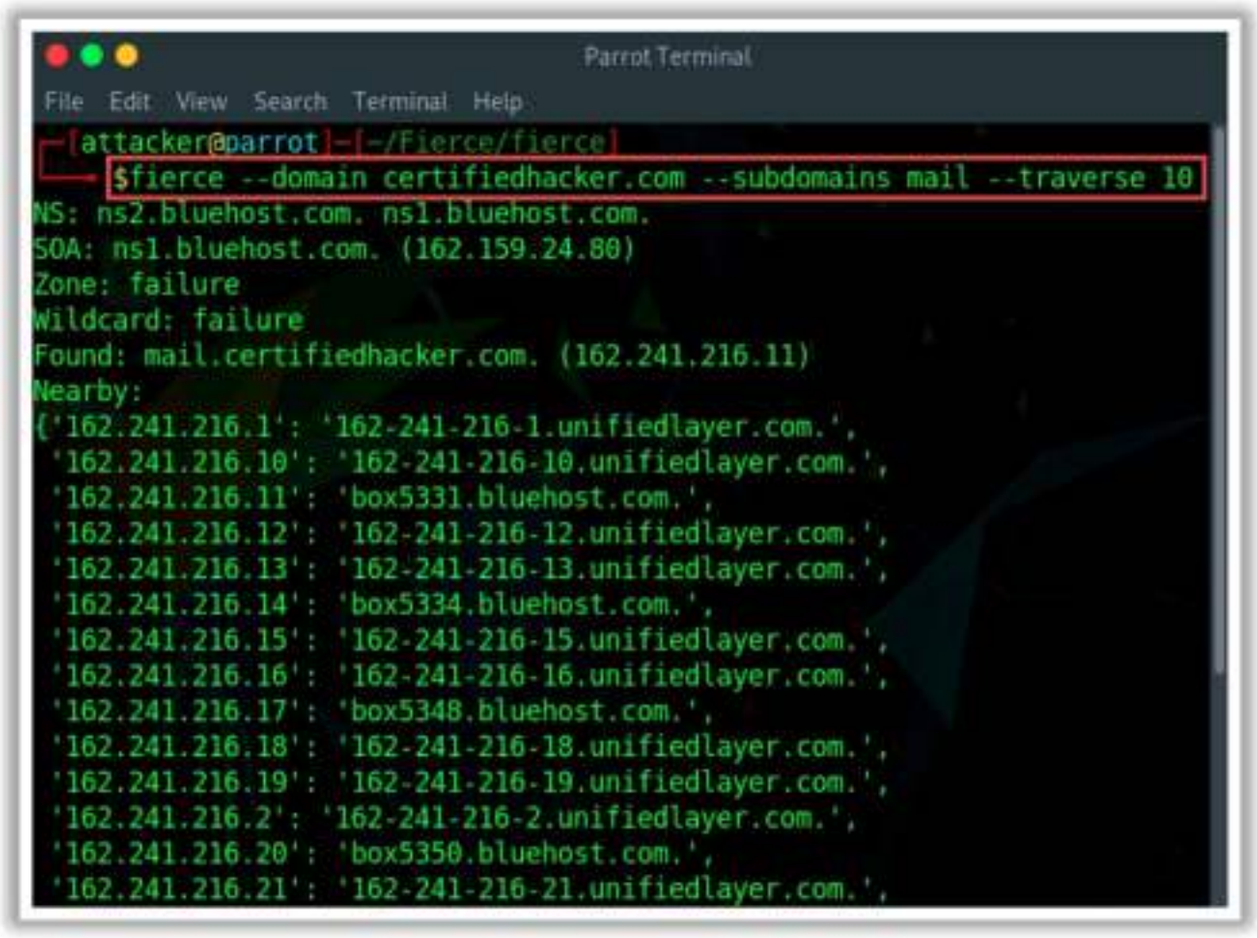


Figure 2.66: Screenshot showing the scanning of domains near discovered records

In the above command, the **`--traverse 10`** option instructs Fierce to search for contiguous blocks of IPs within a range of 10.

- Run the following command to attempt an HTTP connection on the discovered domains of the target:

  `fierce --domain certifiedhacker.com --subdomains mail --connect`

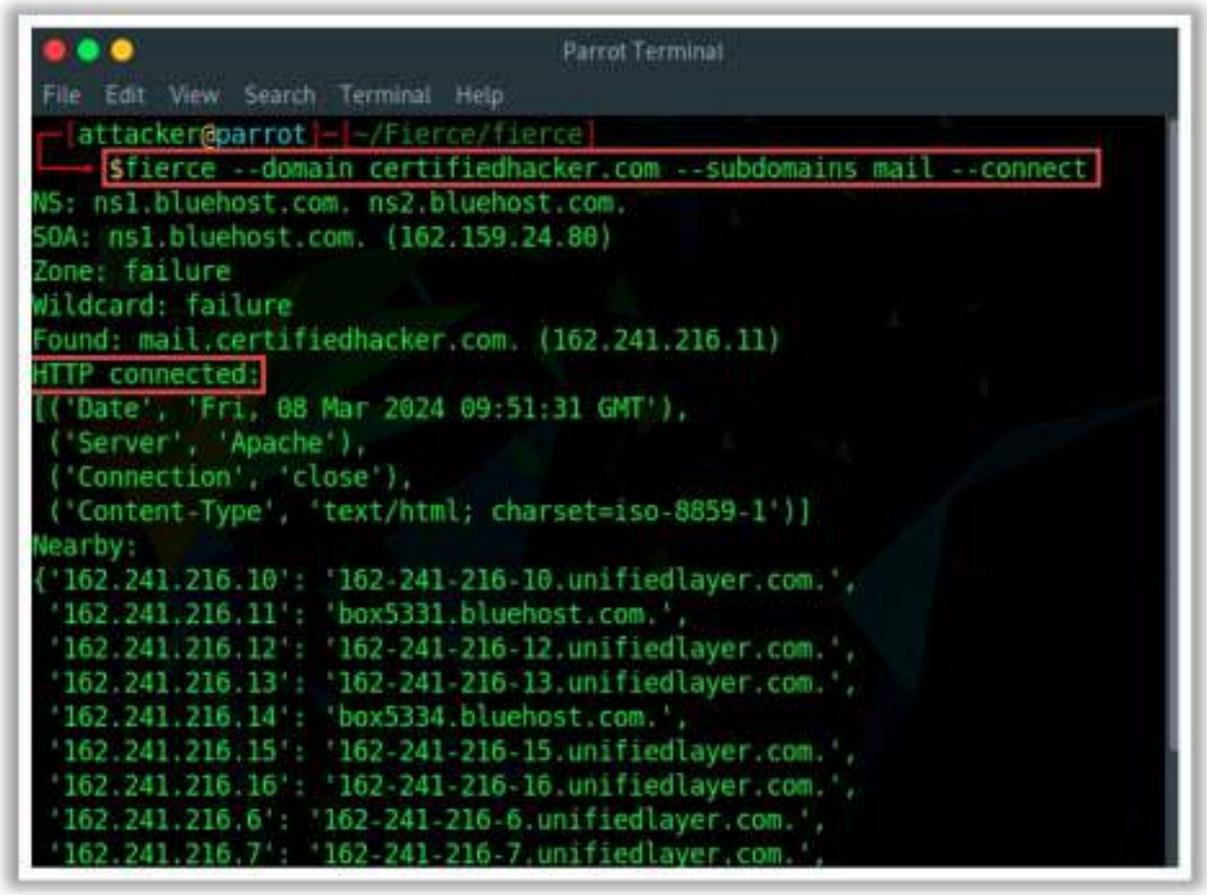


Figure 2.67: Screenshot showing the HTTP connection on discovered domains

- Run the following command to scan all the discovered records of the target domain, i.e., a full detailed scan:

  `fierce --domain certifiedhacker.com --wide`

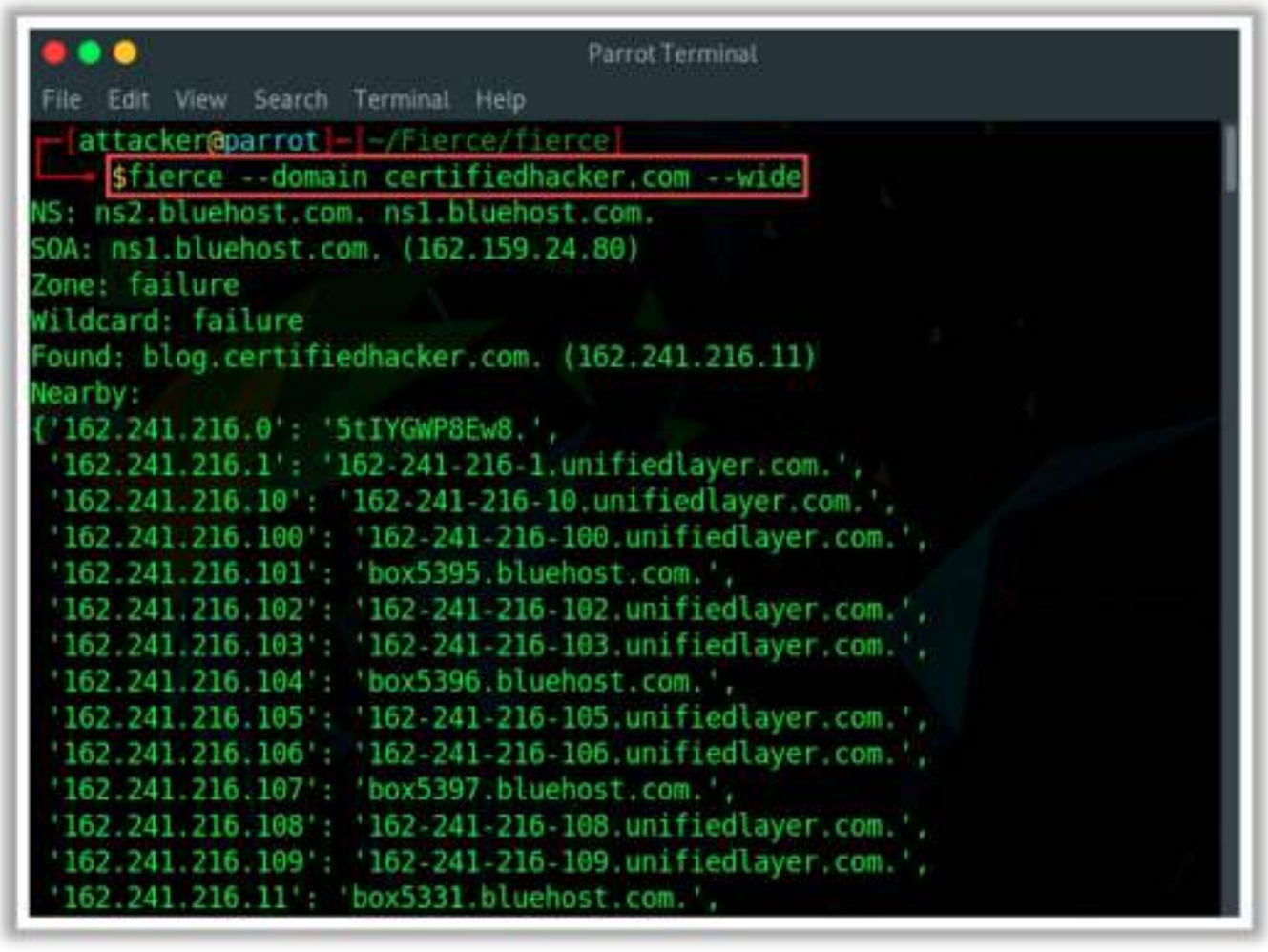
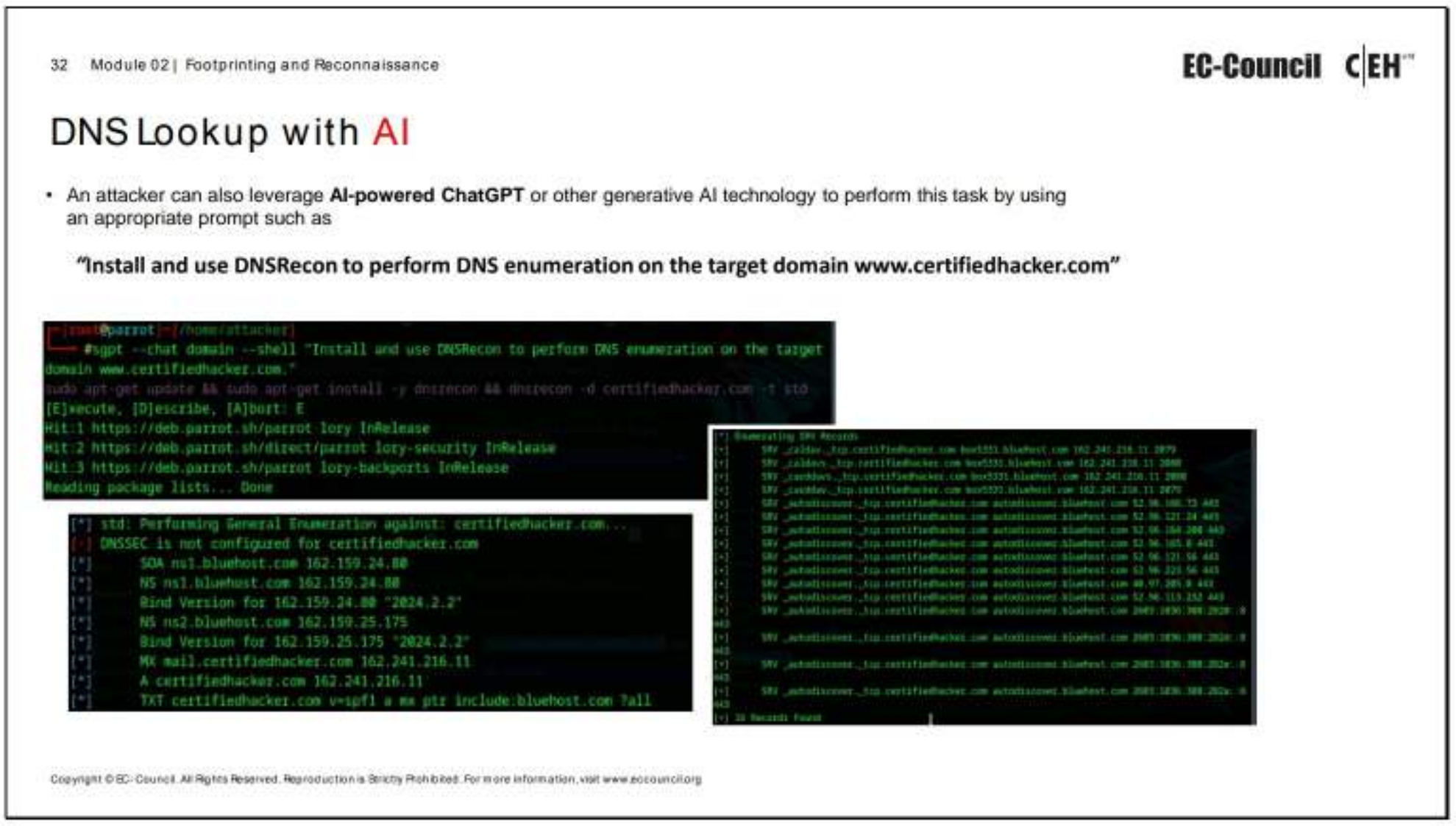


Figure 2.68: Screenshot showing the scanning of discovered records of all the classes



DNS Lookup with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly perform reverse DNS lookup activities on a target and acquire valuable insights.

For example,

Attackers can use ChatGPT to execute this task by using an appropriate prompt such as:

"Install and use DNSRecon to perform DNS enumeration on the target domain www.certifiedhacker.com"

Figure 2.69: Prompt for Installing and performing DNSRecon with AI

The following shell command is designed to perform DNS enumeration using the "dnsrecon" tool on the www.certifiedhacker.com domain:

```
sudo apt-get update && sudo apt-get install -y dnsrecon && dnsrecon -d certifiedhacker.com -t std
```

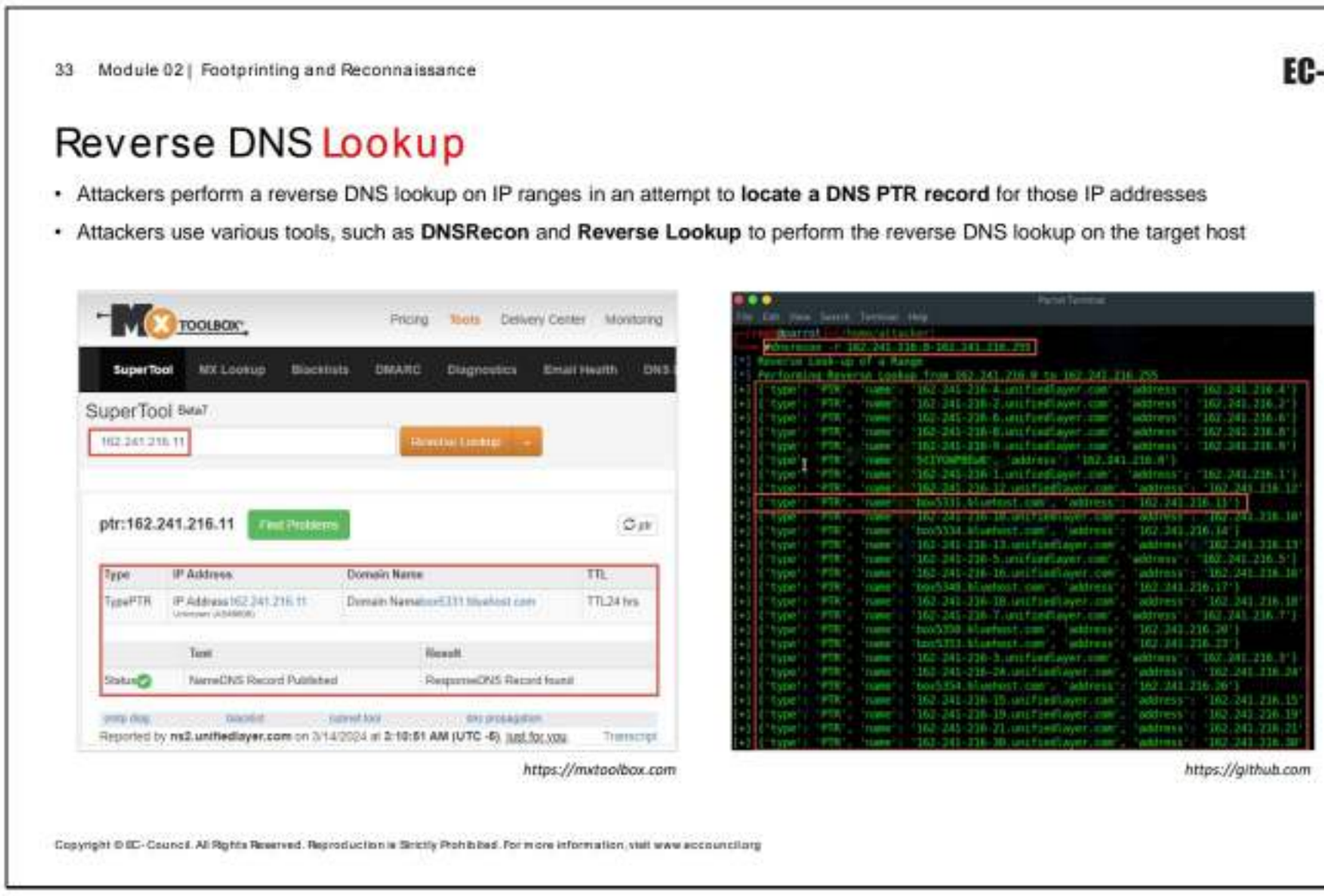
Explanation of the command:

- **`sudo apt-get update`**: Updates the package lists for upgrades and new package installations.
- **`&&`**: Concatenates commands to execute them sequentially.
- **`sudo apt-get install -y dnsrecon`**: Installs the dnsrecon tool with automatic "yes" to all prompts.
- **`dnsrecon -d certifiedhacker.com -t std`**: Initiates the dnsrecon tool to perform DNS enumeration on the certifiedhacker.com domain using standard enumeration techniques.

```
[*] std: Performing General Enumeration against: certifiedhacker.com...
[-] DNSSEC is not configured for certifiedhacker.com
[*]     SOA ns1.bluehost.com 162.159.24.80
[*]     NS ns1.bluehost.com 162.159.24.80
[*]     Bind Version for 162.159.24.80 "2024.2.2"
[*]     NS ns2.bluehost.com 162.159.25.175
[*]     Bind Version for 162.159.25.175 "2024.2.2"
[*]     MX mail.certifiedhacker.com 162.241.216.11
[*]     A certifiedhacker.com 162.241.216.11
[*]     TXT certifiedhacker.com v=spf1 a mx ptr include:bluehost.com ?all
```

```
[*] Enumerating SRV Records
[+]     SRV _caldav._tcp.certifiedhacker.com box5331.bluehost.com 162.241.216.11 2079
[+]     SRV _caldavs._tcp.certifiedhacker.com box5331.bluehost.com 162.241.216.11 2080
[+]     SRV _carddavs._tcp.certifiedhacker.com box5331.bluehost.com 162.241.216.11 2080
[+]     SRV _carddav._tcp.certifiedhacker.com box5331.bluehost.com 162.241.216.11 2079
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.166.72 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.121.24 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.164.200 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.165.8 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.121.56 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.223.56 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 40.97.205.8 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 52.96.113.232 443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 2603:1036:308:2820::8
443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 2603:1036:308:282d::8
443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 2603:1036:308:282e::8
443
[+]     SRV _autodiscover._tcp.certifiedhacker.com autodiscover.bluehost.com 2603:1036:308:282a::8
443
[+] 16 Records Found
```

Figure 2.70: Output for performing DNSRecon with AI



Reverse DNS Lookup

DNS lookup is used to find the IP addresses for a given domain name, and a reverse DNS operation is performed to obtain the domain name of a given IP address. When looking for a domain by entering the domain name in a browser, the DNS converts the domain name into an IP address and forwards the request for further processing. This conversion of a domain name into an IP address is performed using a record. Attackers perform a reverse DNS lookup on the IP range to locate a DNS PTR record for such IP addresses.

Attackers use various tools such as DNSRecon, Reverse Lookup, puredns, Reverse IP Domain Check, and Reverse IP Lookup to perform reverse DNS lookup on the target host. When we obtain an IP address or a range of IP addresses, we can use these tools to obtain the domain name.

- **DNSRecon**

  Source: *https://github.com*

  As shown in the screenshot, attackers use the following command to perform a reverse DNS lookup on the target host:

  ```
  dnsrecon -r 162.241.216.0-162.241.216.255
  ```

  In the above command, the **-r** option specifies the range of IP addresses (first to last) for a reverse lookup by brute force.

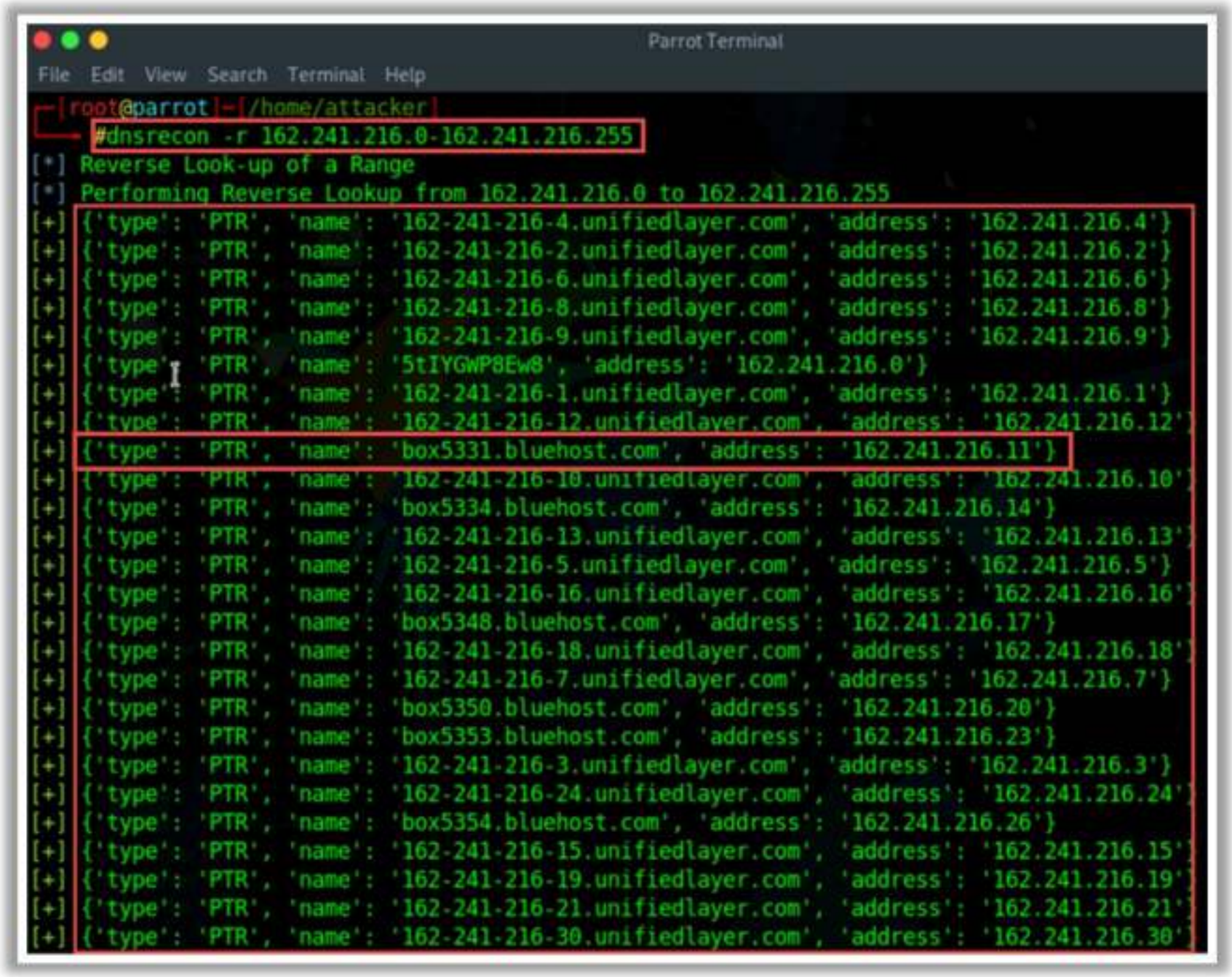


Figure 2.71: Screenshot of DNSRecon showing reverse DNS lookup information

- **Reverse Lookup**

  Source: *https://mxtoolbox.com*

  As shown in the screenshot, the Reverse Lookup tool performs a reverse IP lookup by taking an IP address and locating a DNS PTR record for that IP address.

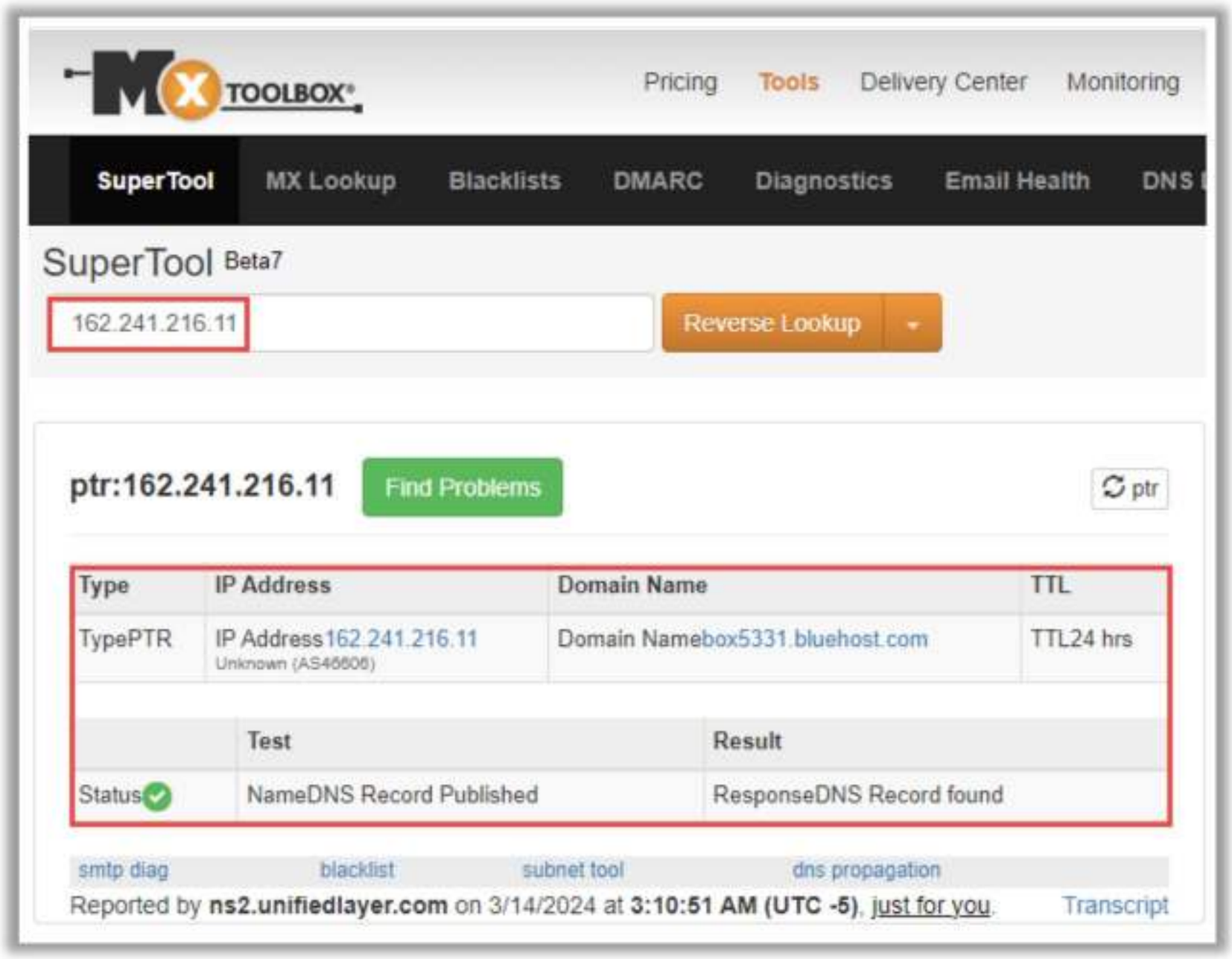
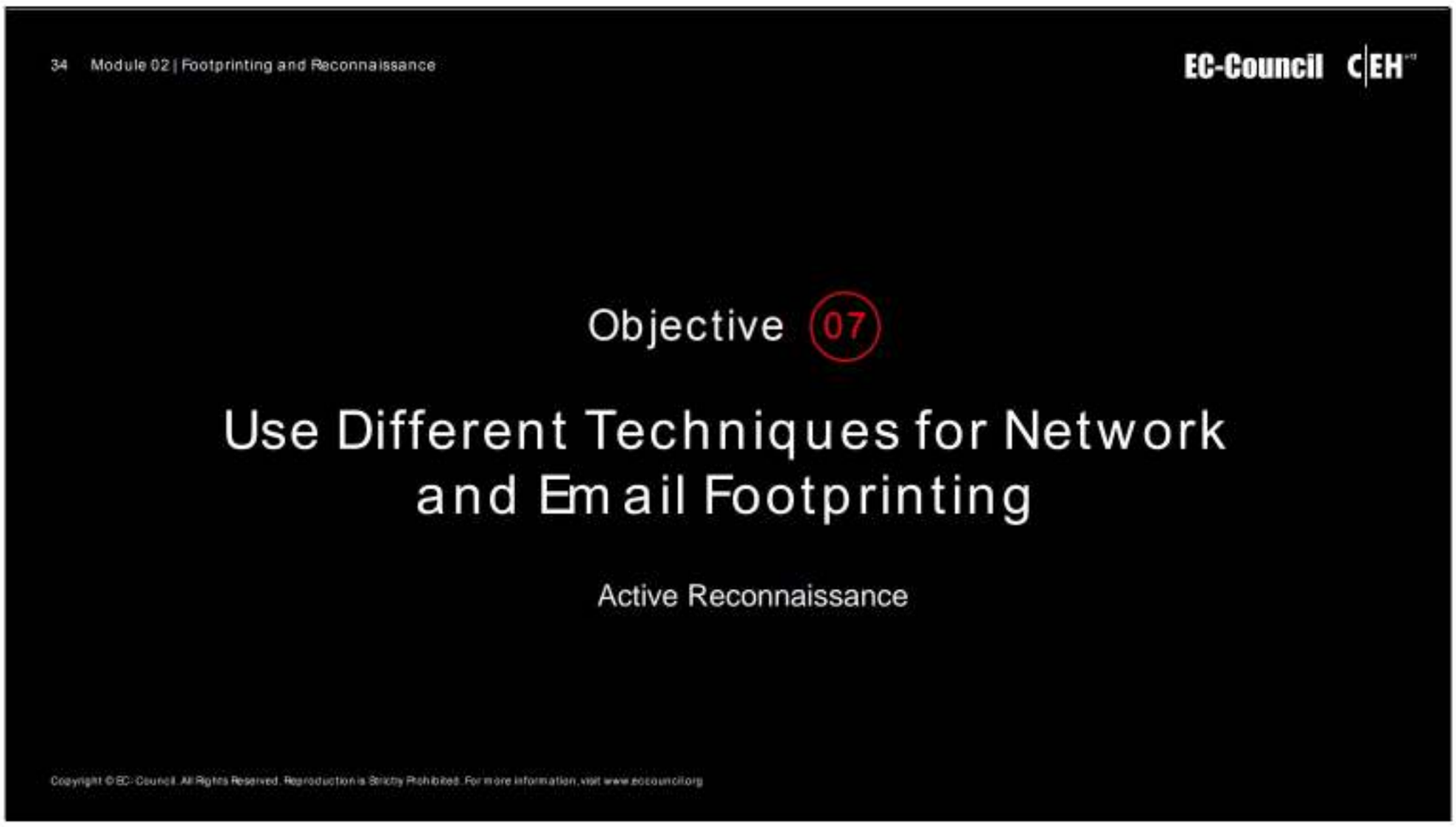


Figure 2.72: Screenshot of the Reverse Lookup tool



Network and Email Footprinting

The next step after retrieving DNS information is to gather network-related information and track email communications. This section describes the method to locate the network range, traceroute analysis, and traceroute tools. It also describes how to track email communications, how to collect information from email headers, and email tracking tools.

Locate the Network Range

To perform network footprinting, one needs to gather basic and important information about the target organization, such as what the organization does, who works there, and what type of work it does. The answers to these questions provide information that helps identify the internal structure of the target network.

After gathering the information, an attacker can determine the network range of the target system. Detailed information regarding IP allocation and the nature of allocation is available with the appropriate regional registry database. An attacker can also determine the subnet mask of the domain and trace the route between the system and target system. Widely used traceroute tools include NetScanTools Pro and PingPlotter.

Obtaining private IP addresses can be useful to attackers. The Internet Assigned Numbers Authority (IANA) has reserved the following three blocks of IP address space for private internets: 10.0.0.0–10.255.255.255 (10/8 prefix), 172.16.0.0–172.31.255.255 (172.16/12 prefix), and 192.168.0.0–192.168.255.255 (192.168/16 prefix).

Using the network range, an attacker can obtain information about how the network is structured and which machines in the network are alive. The network range also helps identify the network topology, access control device, and OS used in the target network. To find the network range of the target network, one must enter the server IP address (gathered in Whois footprinting) in the

ARIN Whois database search tool. A user can also visit the ARIN website (*https://www.arin.net/about/welcome/region*) and enter the server IP into the **SEARCH Site or Whois** text box. This yields the network range of the target network. Improperly set-up DNS servers offer attackers a good chance of obtaining a list of internal machines in the network. Additionally, if an attacker traces a route to a machine, it could be possible to obtain the internal IP address of the gateway, which can be useful.

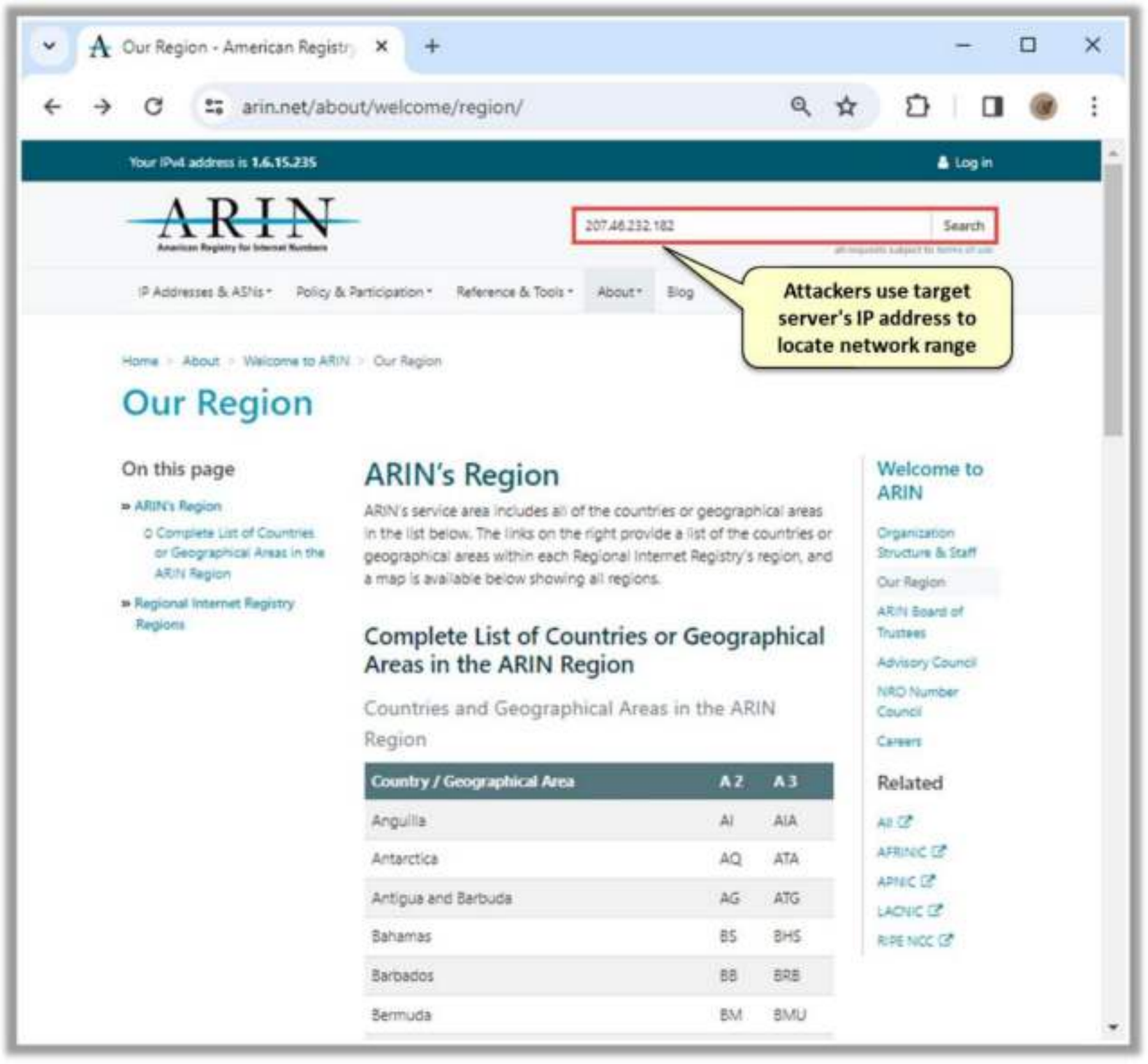


Figure 2.73: Screenshot of ARIN's Region

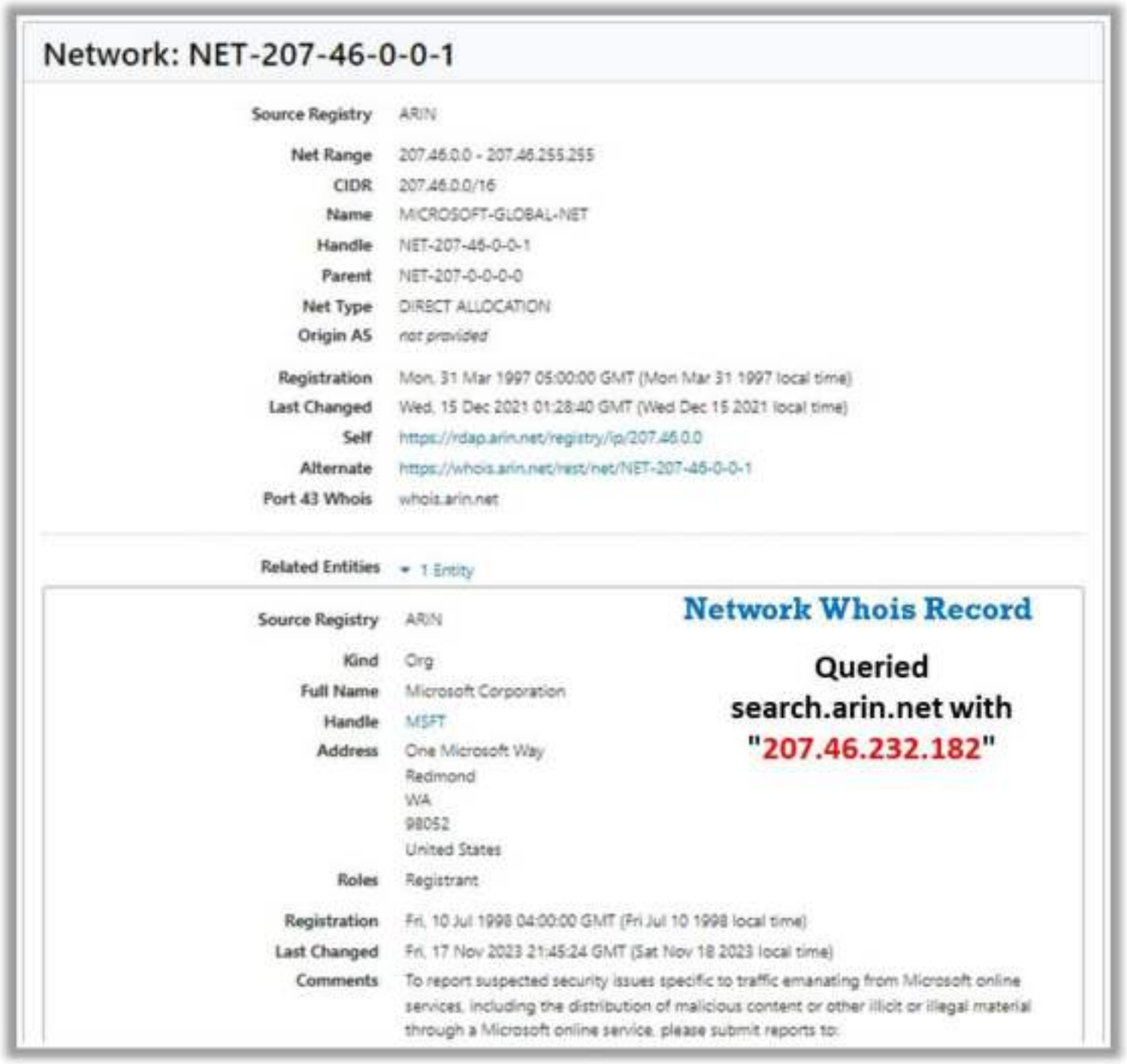


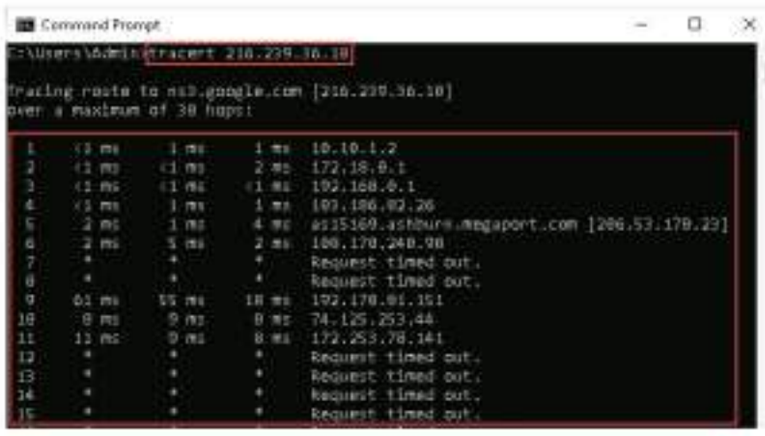
Figure 2.74: Screenshot showing result of ARIN Whois database search result

Attackers typically use more than one tool to obtain network information, as a single tool cannot provide all the required information.

Traceroute

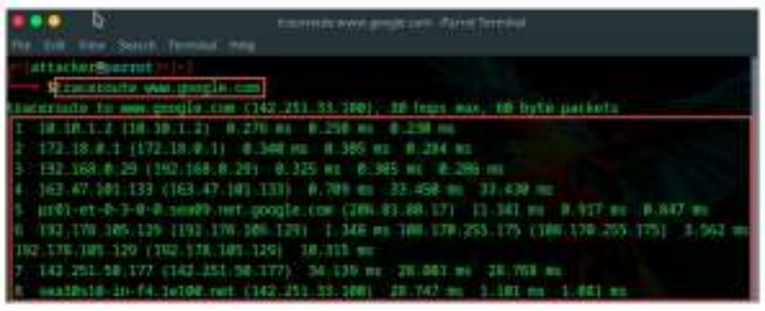
Traceroute programs work on the concept of **ICMP protocol** and **use the TTL field in the header of ICMP packets** to discover the routers on the path to a target host

ICMP Traceroute



TCP Traceroute

UDP Traceroute



Traceroute

Finding the route of the target host on the network is necessary to test against man-in-the-middle attacks and other related attacks. Most operating systems come with a Traceroute utility to perform this task. It traces the path or route through which the target host packets travel in the network.

Traceroute uses the ICMP protocol and Time to Live (TTL) field of the IP header to find the path of the target host in the network.

The Traceroute utility can detail the path through which IP packets travel between two systems. The utility can trace the number of routers the packets travel through, the round-trip time (duration in transiting between two routers), and, if the routers have DNS entries, the names of the routers and their network affiliation. It can also trace geographic locations. It works by exploiting a feature of the Internet Protocol called TTL. The TTL field indicates the maximum number of routers a packet may traverse. Each router that handles a packet decrements the TTL count field in the ICMP header by one. When the count reaches zero, the router discards the packet and transmits an ICMP error message to the originator of the packet.

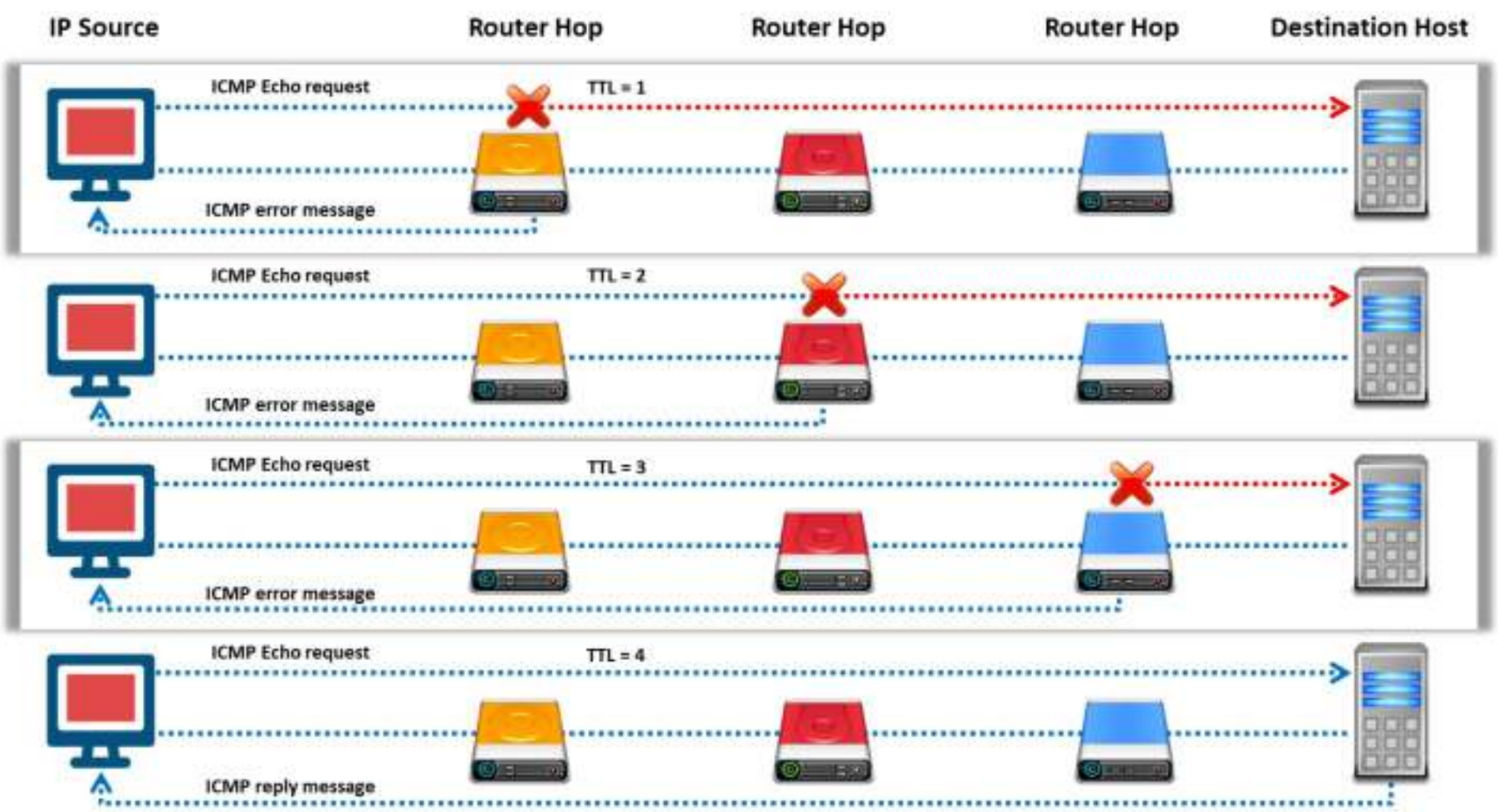


Figure 2.75: Illustration of Traceroute

The utility records the IP address and DNS name of the router and sends out another packet with a TTL value of two. This packet makes it through the first router and then times-out at the next router in the path. This second router also sends an error message back to the originating host. Traceroute continues to do this and records the IP address and name of each router until a packet finally reaches the target host or until it decides that the host is unreachable. In the process, it records the time taken for each packet to make a round trip to each router. Finally, when it reaches the destination, the normal ICMP ping response will be sent back to the sender. The utility helps to reveal the IP addresses of the intermediate hops in the route to the target host from the source.

ICMP Traceroute

Windows operating system by default uses ICMP traceroute. Go to the command prompt and type the **tracert** command along with the destination IP address or domain name as follows:

```
C:\>tracert 216.239.36.10
```

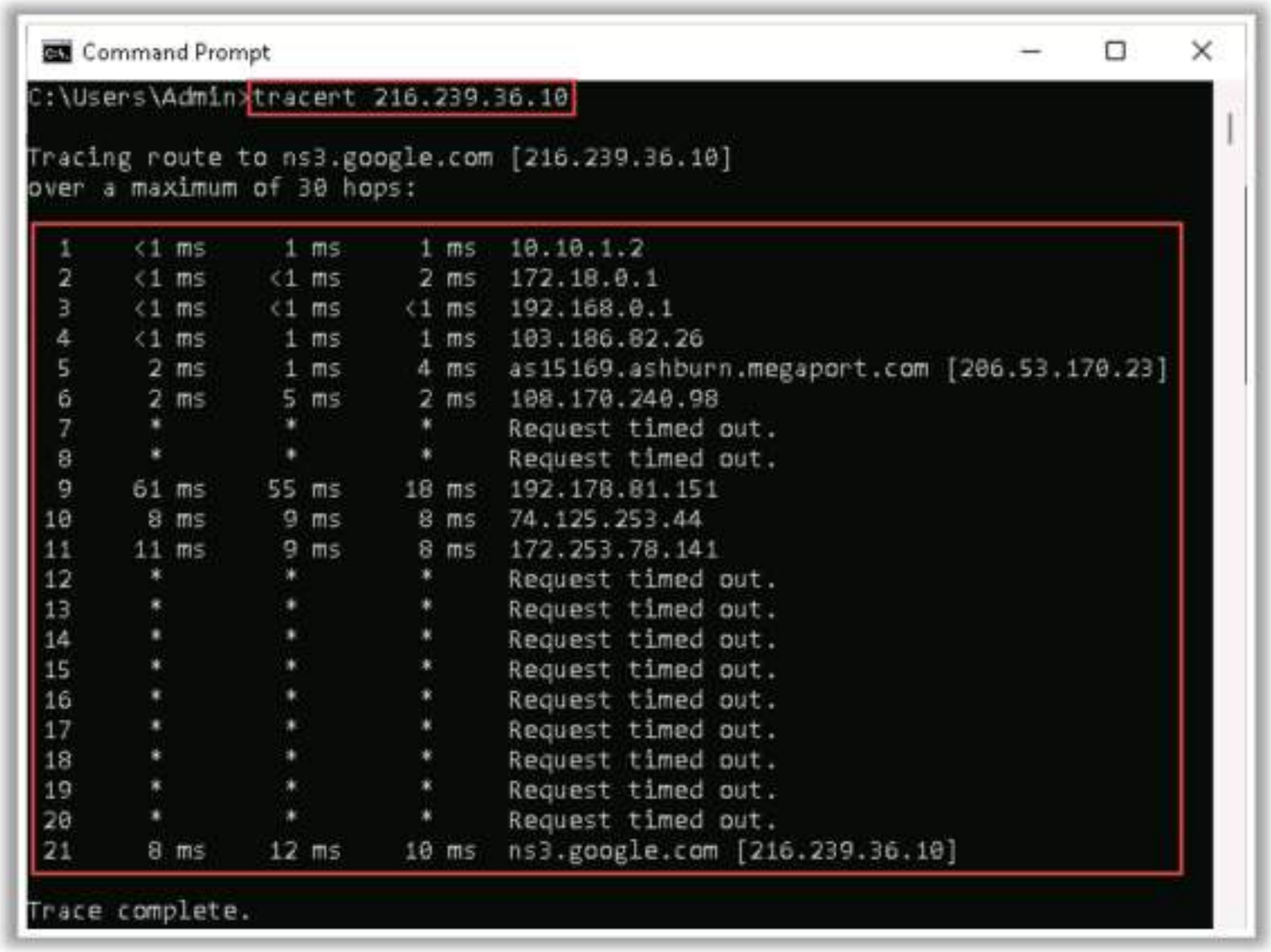


Figure 2.76: Screenshot showing the output of tracert

TCP Traceroute

Many devices in any network are generally configured to block ICMP traceroute messages. In this scenario, an attacker uses TCP or UDP traceroute, which is also known as Layer 4 traceroute. Go to the terminal in Linux operating system and type the **tcptraceroute** command along with the destination IP address or domain name as follows:

```
sudo tcptraceroute www.google.com
```

Figure 2.77: Screenshot showing the output of TCP Traceroute

UDP Traceroute

Like Windows, Linux also has a built-in traceroute utility, but it uses the UDP protocol for tracing the route to the destination. Go to the terminal in the Linux operating system and type the **traceroute** command along with the destination IP address or domain name as follows:

```
traceroute www.google.com
```

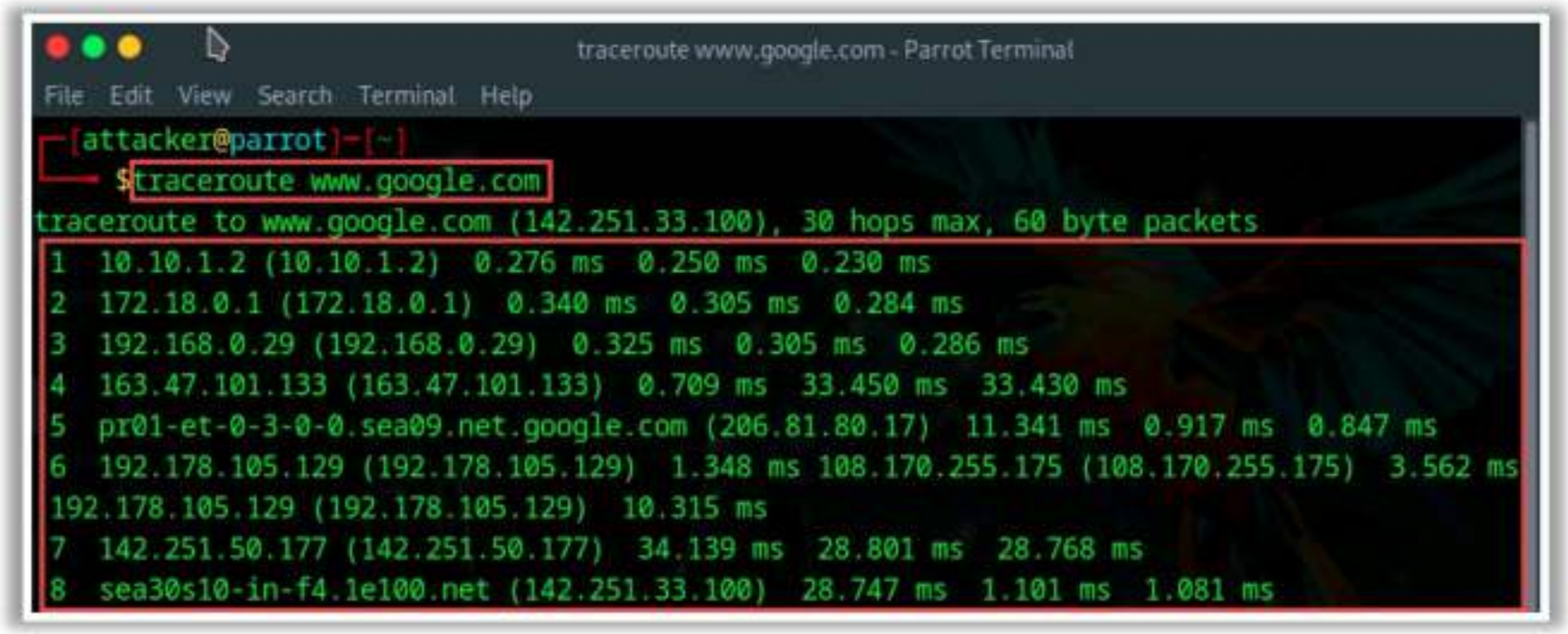
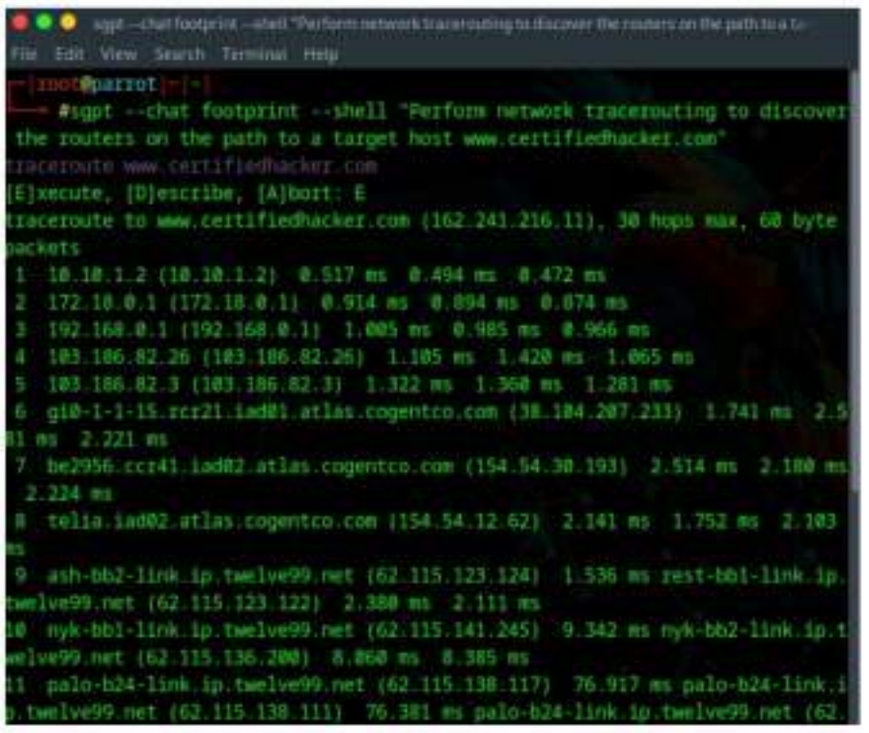


Figure 2.78: Screenshot showing the output of UDP Traceroute

Traceroute with AI

- An attacker can also leverage **AI-powered ChatGPT** or other generative AI technology to perform this task by using an appropriate prompt such as

"Perform network tracerouting to discover the routers on the path to a target host www.certifiedhacker.com"



Traceroute with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting operations. With the aid of AI, attackers can effortlessly perform tracerouting on a target.

For example,

Attackers can use ChatGPT to execute this task by using an appropriate prompt such as:

"Perform network tracerouting to discover the routers on the path to a target host www.certifiedhacker.com"

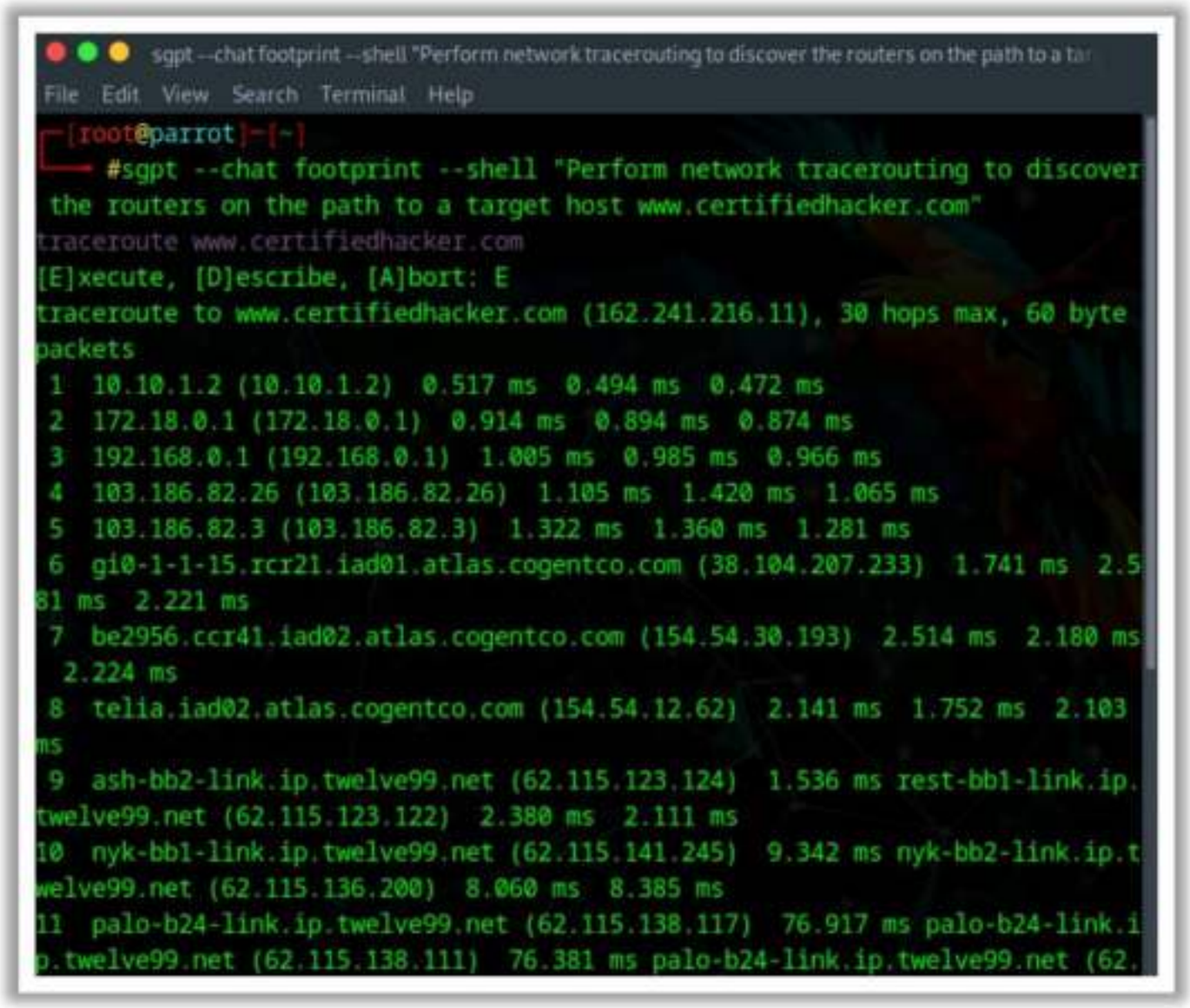


Figure 2.79: Prompt and Output for performing Traceroute with AI

The following shell command is designed to perform network tracerouting using the "traceroute" tool to discover the routers on the path to the www.certifiedhacker.com host:

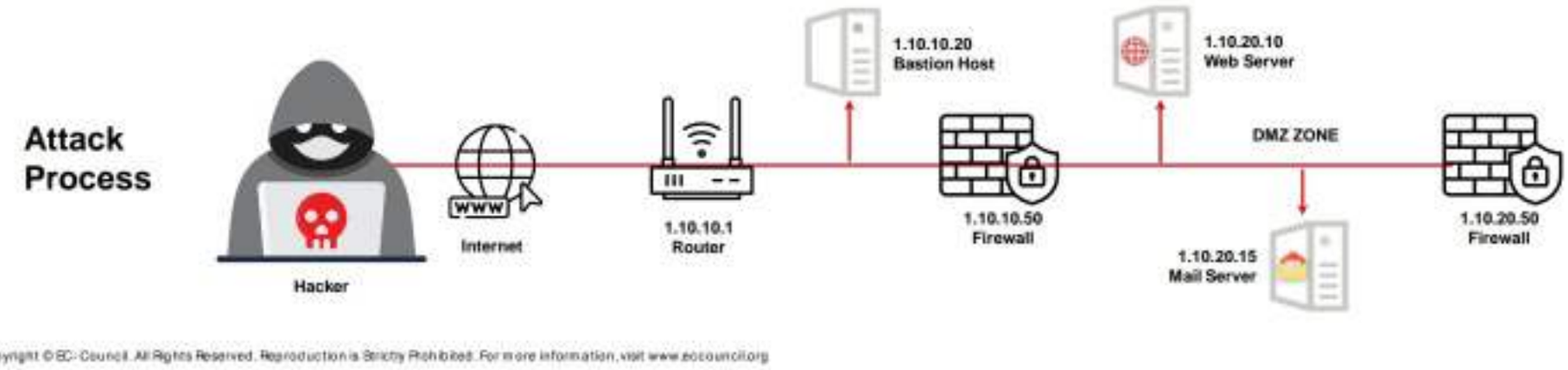
```
traceroute www.certifiedhacker.com
```

Explanation of the command:

- **`traceroute www.certifiedhacker.com`**: Initiates the traceroute tool to discover the routers on the path to the www.certifiedhacker.com host by sending packets to the destination with increasing time to live (TTL) values and analyzing the responses received from intermediate routers.

Traceroute Analysis

- Attackers execute traceroute to find the **IP addresses of intermediate devices** such as routers and firewalls present between a source and its destination
- For example, after running several **traceroutes**, an attacker might obtain the following information:
  - traceroute 1.10.10.20, second to last hop is 1.10.10.1
  - traceroute 1.10.20.10, third to last hop is 1.10.10.1
  - traceroute 1.10.20.10, second to last hop is 1.10.10.50
  - traceroute 1.10.20.15, third to last hop is 1.10.10.1
  - traceroute 1.10.20.15, second to last hop is 1.10.10.50
- By compiling this information, attackers can identify the intermediate devices or hosts in the path to the target network



Traceroute Analysis

We have discussed how the traceroute utility helps find the IP addresses of intermediate devices such as routers and firewalls present between a source and its destination. After running several traceroutes, an attacker can find the location of a hop in the target network. Consider the following traceroute results:

- `traceroute 1.10.10.20, second to last hop is 1.10.10.1`
- `traceroute 1.10.20.10, third to last hop is 1.10.10.1`
- `traceroute 1.10.20.10, second to last hop is 1.10.10.50`
- `traceroute 1.10.20.15, third to last hop is 1.10.10.1`
- `traceroute 1.10.20.15, second to last hop is 1.10.10.50`

By analyzing these results, an attacker can identify the intermediate devices or hosts in the path to reach the target network, as shown in the figure.

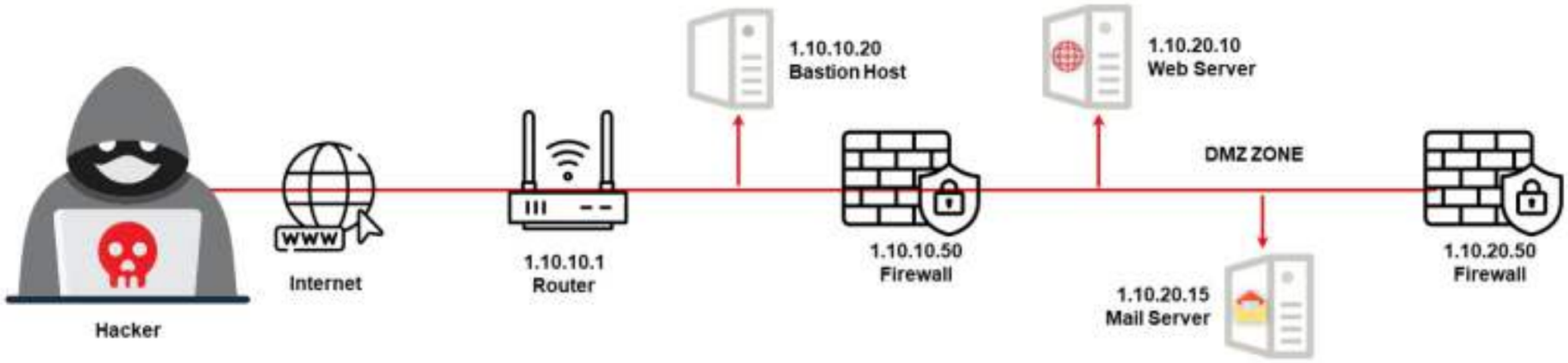


Figure 2.80: Traceroute Analysis

Traceroute Tools

Traceroute tools such as NetScanTools Pro, PingPlotter, Traceroute NG, and tracert are useful for extracting information about the geographical location of routers, servers, and IP devices in a network. Such tools help us to trace, identify, and monitor the network activity on a world map. Some of the features of these tools are as follows:

- Hop-by-hop traceroutes
- Reverse tracing
- Historical analysis
- Packet loss reporting
- Reverse DNS
- Ping plotting
- Port probing
- Detect network problems
- Performance metrics analysis
- Network performance monitoring
- **NetScanTools Pro**

  Source: *https://www.netscantools.com*

  Attackers can use NetScanTools Pro to trace the route packets traverse from their machine to the target device on a local area network or across the Internet. The tool offers ICMP, UDP, or TCP traceroute methods and allows attackers to identify the intermediate devices along the route. Also, it helps attackers locate the country assigned to each IPv4 address in each hop.

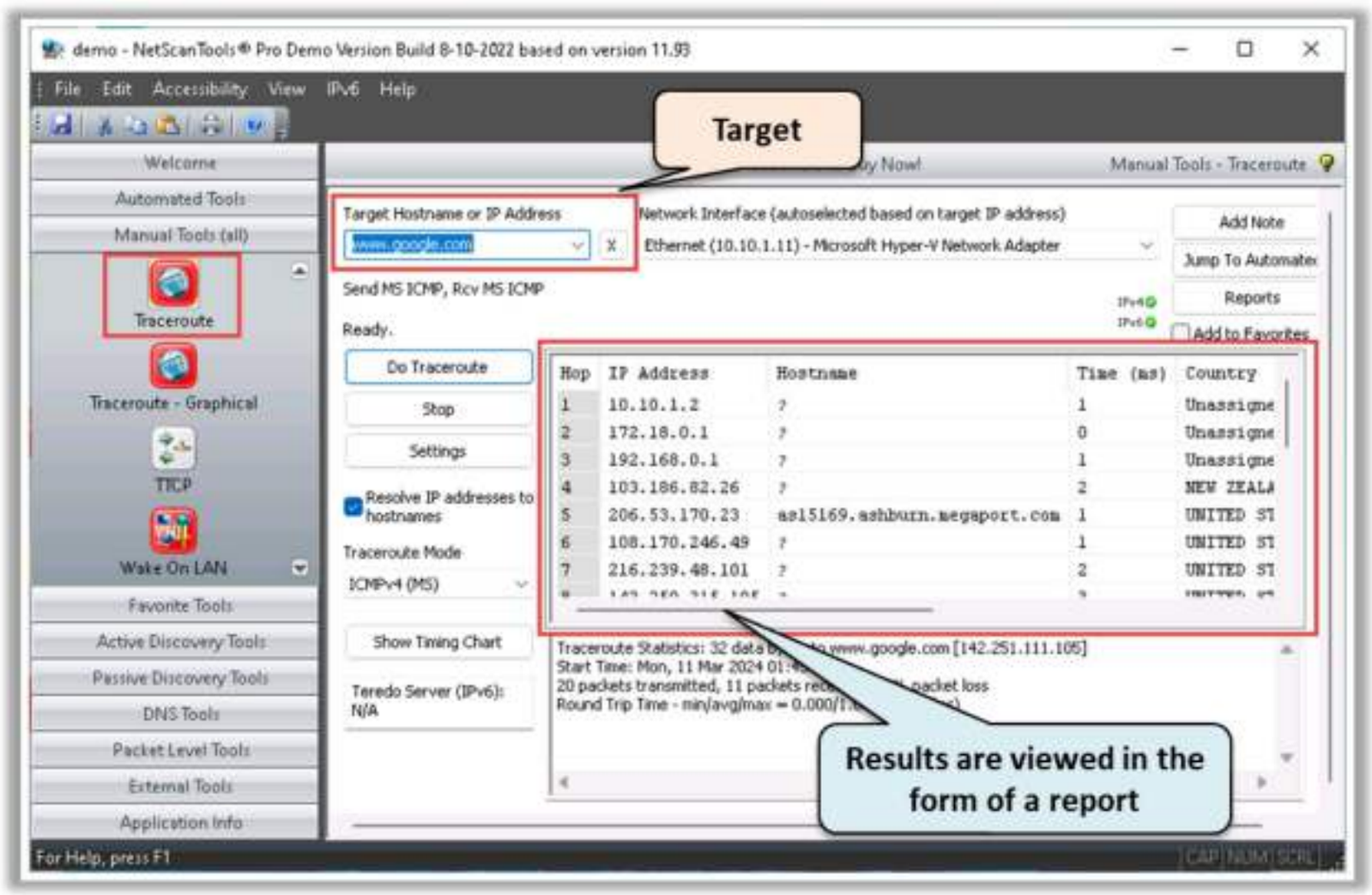


Figure 2.81: Screenshot of NetScanTools Pro

- **PingPlotter**

  Source: *https://www.pingplotter.com*

  PingPlotter allows attackers to collect traceroute data for target hosts using ICMP, UDP, and TCP packets. It automatically discovers the network hops and tracks latency and packet loss over time. Using this tool, attackers can visualize the traceroute data in readable graphs. This tool aids attackers in identifying bandwidth bottlenecks, WiFi interference, or hardware faults on the target network.

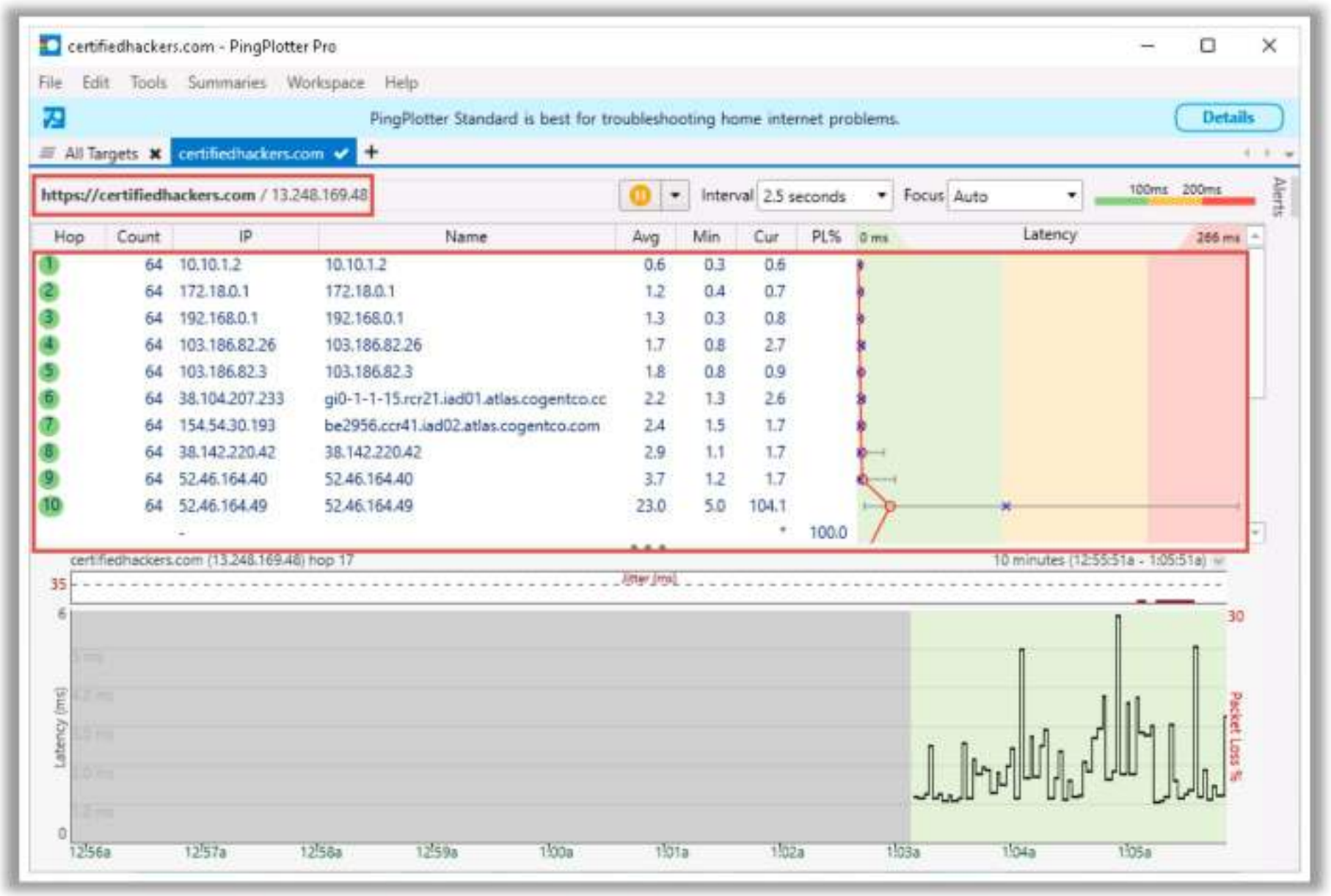
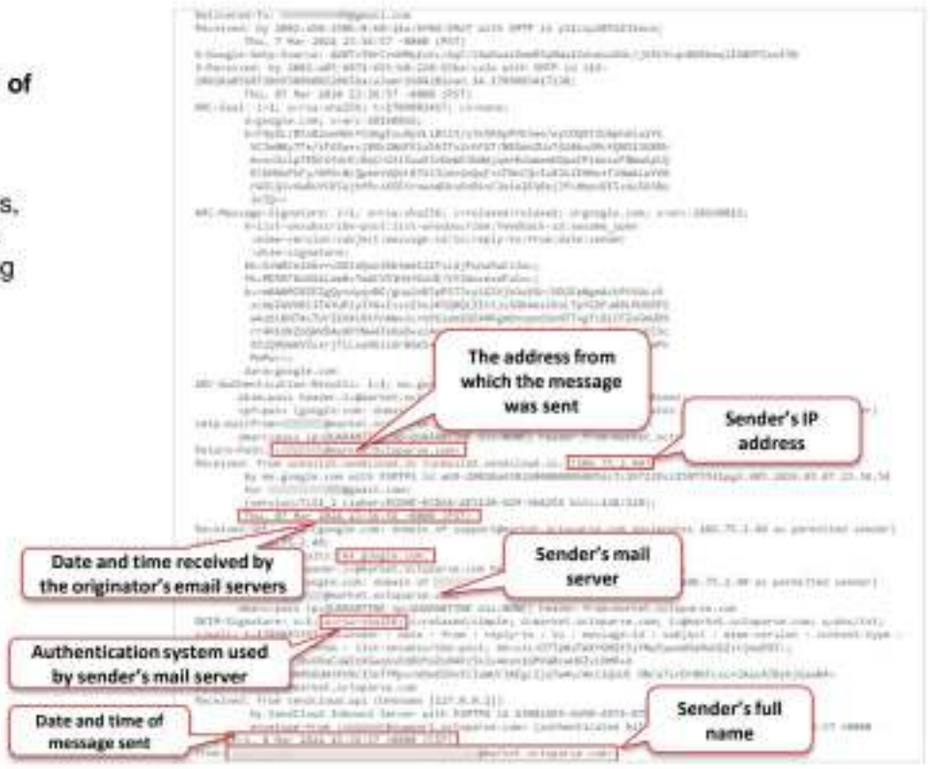


Figure 2.82: Screenshot of PingPlotter

EC-Council CEH™

Tracking Email Communications

- Email tracking is used to **monitor the delivery of emails** to an intended recipient
- Attackers track emails to **gather information about a target recipient**, such as IP addresses, geolocation, browser and OS details, to build a hacking strategy and perform social engineering and other attacks



Collecting Information from Email Header

Tracking Email Communications

Email tracking monitors the email messages of a particular user. This kind of tracking is possible through digitally time-stamped records that reveal the time and date when the target receives and opens a specific email. Email tracking tools allow an attacker to collect information such as IP addresses, mail servers, and service providers involved in sending the email. Attackers can use this information to build a hacking strategy and to perform social engineering and other attacks. Examples of email tracking tools include IP2LOCATION's Email Header Tracer, MxToolbox, DNS Checker Email Header Analyzer, and Social Catfish.

Information about the victim gathered using email tracking tools includes:

- **Recipient's System IP address**: Allows tracking of the recipient's IP address
- **Geolocation**: Estimates and displays the location of the recipient on the map and may even calculate the distance from the attacker's location
- **Email Received and Read**: Notifies the attacker when the email is received and read by the recipient
- **Read Duration**: The time spent by the recipient in reading the email sent by the sender
- **Proxy Detection**: Provides information about the type of server used by the recipient
- **Links**: Checks whether the links sent to the recipient through email have been checked
- **Operating System and Browser information**: Reveals information about the operating system and the browser used by the recipient. The attacker can use this information to find loopholes in that version of the operating system and browser to launch further attacks

- **Forward Email**: Determines whether the email sent to the user is forwarded to another person
- **Device Type**: Provides information about the type of device used to open and read the email, e.g., desktop computer, mobile device, or laptop
- **Path Travelled**: Tracks the path through which the email traveled via email transfer agents from source to destination system

Collecting Information from Email Header

An email header contains the details of the sender, routing information, addressing scheme, date, subject, and recipient. Email headers also help attackers to trace the routing path taken by an email before it is delivered to the recipient. Each email header is a useful source of information for an attacker to launch attacks against the target. The process of viewing the email header varies with different email programs.

Commonly used email programs:

- eM Client
- Mailbird
- Hiri
- Mozilla Thunderbird
- Spike
- Claws Mail
- SmarterMail Webmail
- Outlook
- Apple Mail
- ProtonMail
- AOL Mail
- Tuta

The email header contains the following information:

- Sender's mail server
- Date and time of receipt by the originator's email servers
- Authentication system used by the sender's mail server
- Data and time of sending the message
- A unique number assigned by mx.google.com to identify the message
- Sender's full name
- Sender's IP address and address from which the message was sent

The attacker can trace and collect all this information by performing a detailed analysis of the complete email header.

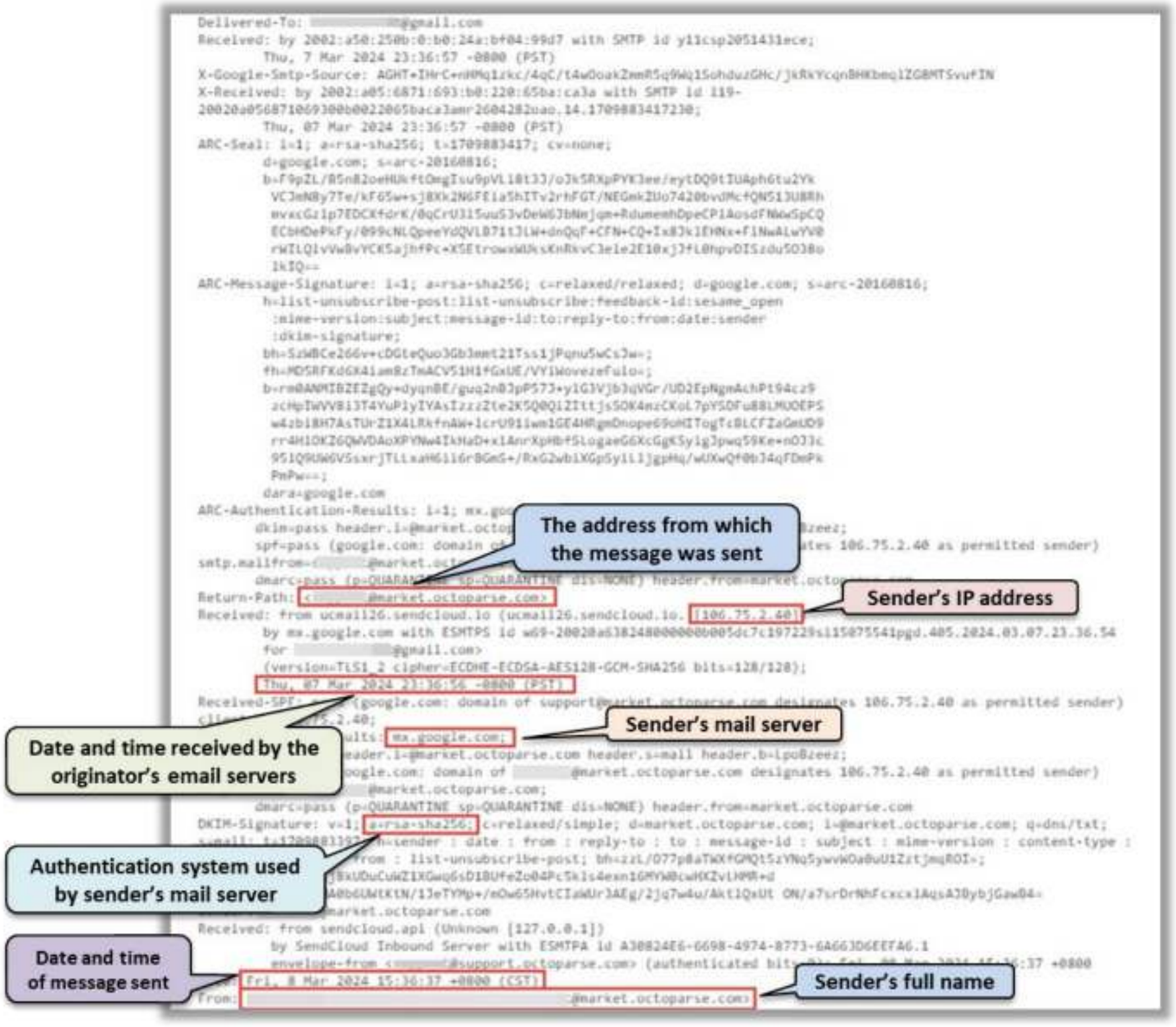


Figure 2.83: Screenshot showing detailed analysis of the email header

Email Tracking Tools

Email tracking tools such as IP2LOCATION's Email Header Tracer, MxToolbox, eMailTrackerPro, Holehe, DNS Checker Email Header Analyzer, and Social Catfish allow an attacker to track an email and extract information such as sender identity, mail server, sender's IP address, location, and so on. Attackers use the extracted information to track the email path from the attacker's location to the target mail server using IP addresses in the email header.

- **eMailTrackerPro**

  Source: *http://www.emailtrackerpro.com*

  As shown in the screenshot, attackers use eMailTrackerPro to analyze email headers and extract information such as the sender's geographical location, IP address, and so on. It allows an attacker to review the traces later by saving past traces.

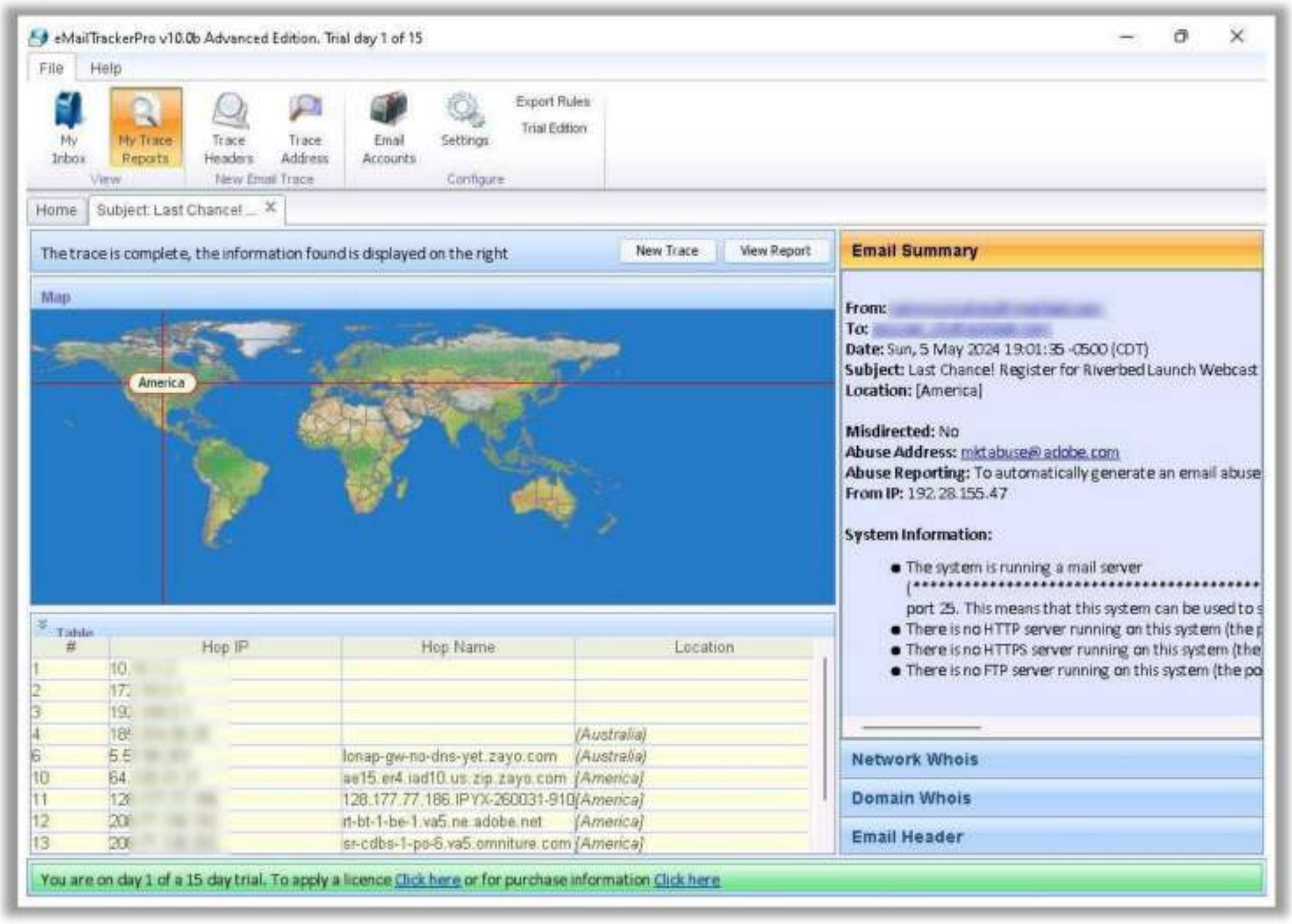


Figure 2.84: Screenshot of eMailTrackerPro

- **IP2LOCATION's Email Header Tracer**

  Source: *https://www.ip2location.com*

  IP2LOCATION's Email Header Tracer is an open-source service that attackers can use to analyze and trace email paths using the email header. It enables attackers to trace back the target location and the mail servers through which the email is passing, utilizing the IP addresses found in the email header.

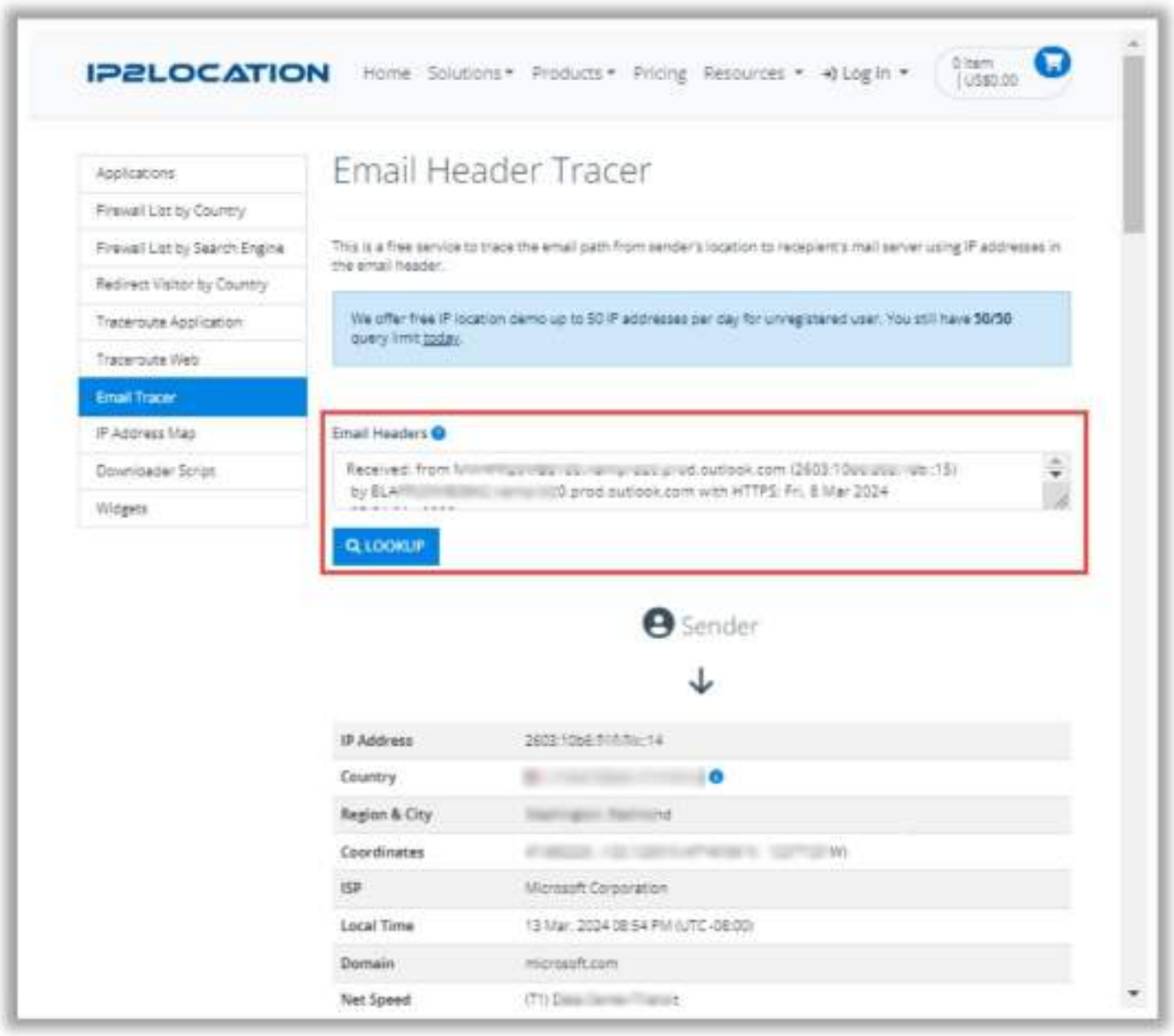


Figure 2.85: Screenshot of IP2LOCATION's Email Header Tracer

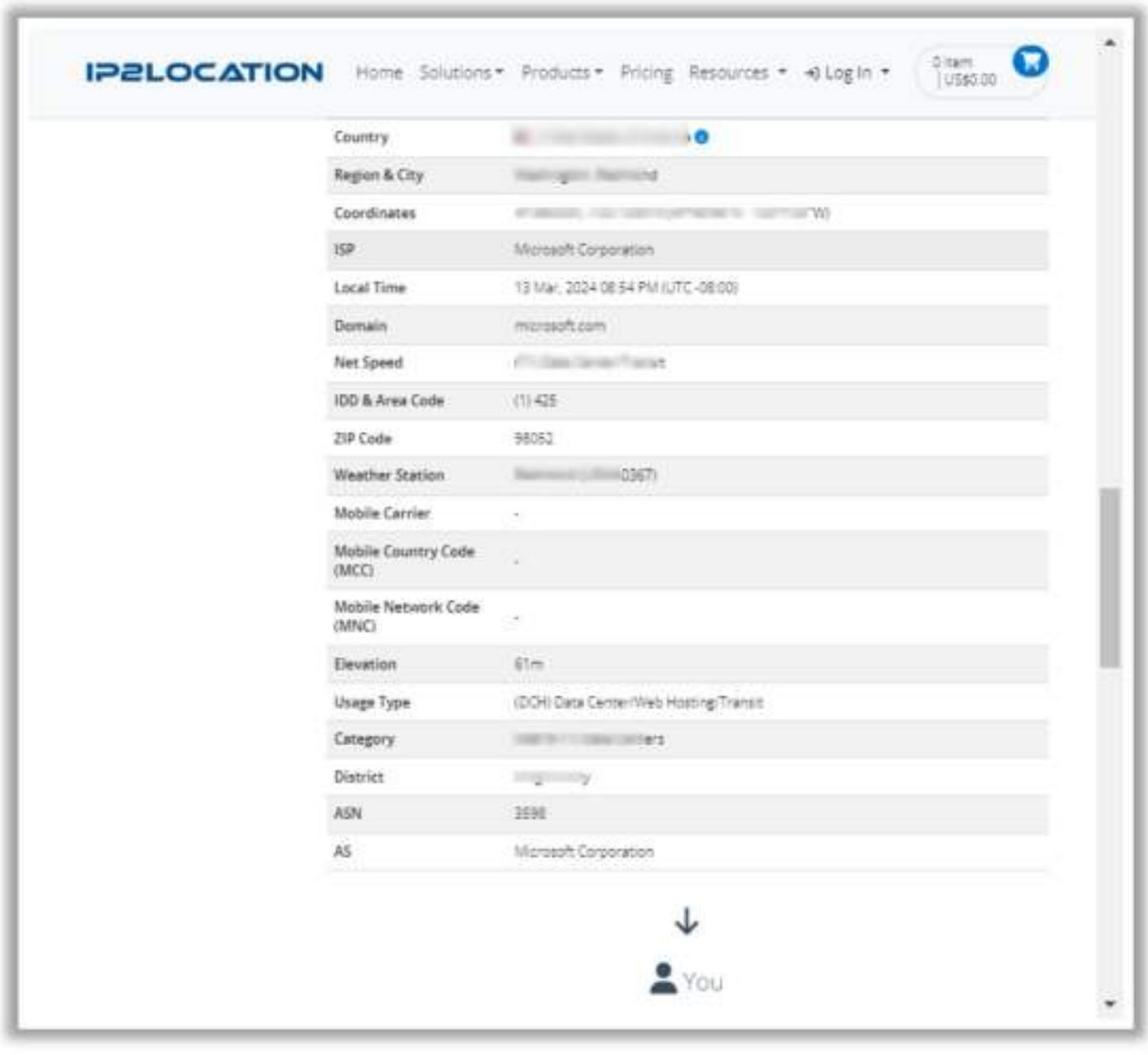


Figure 2.86: Screenshot of IP2LOCATION's Email Header Tracer

Objective 08

Demonstrate Footprinting through Social Engineering

Active Reconnaissance

Footprinting through Social Engineering

- Social engineering is an art of exploiting human behaviour to **extract confidential information**
- Social engineers depend on the fact that **people are unaware** of their valuable information and are careless about protecting it

Social engineers attempt to gather

- Credit card details and social security number
- Usernames and passwords
- Security products in use
- Operating systems and software versions
- Network layout information
- IP addresses and names of servers

Social engineering techniques include

- Eavesdropping
- Shoulder surfing
- Dumpster diving
- Impersonation

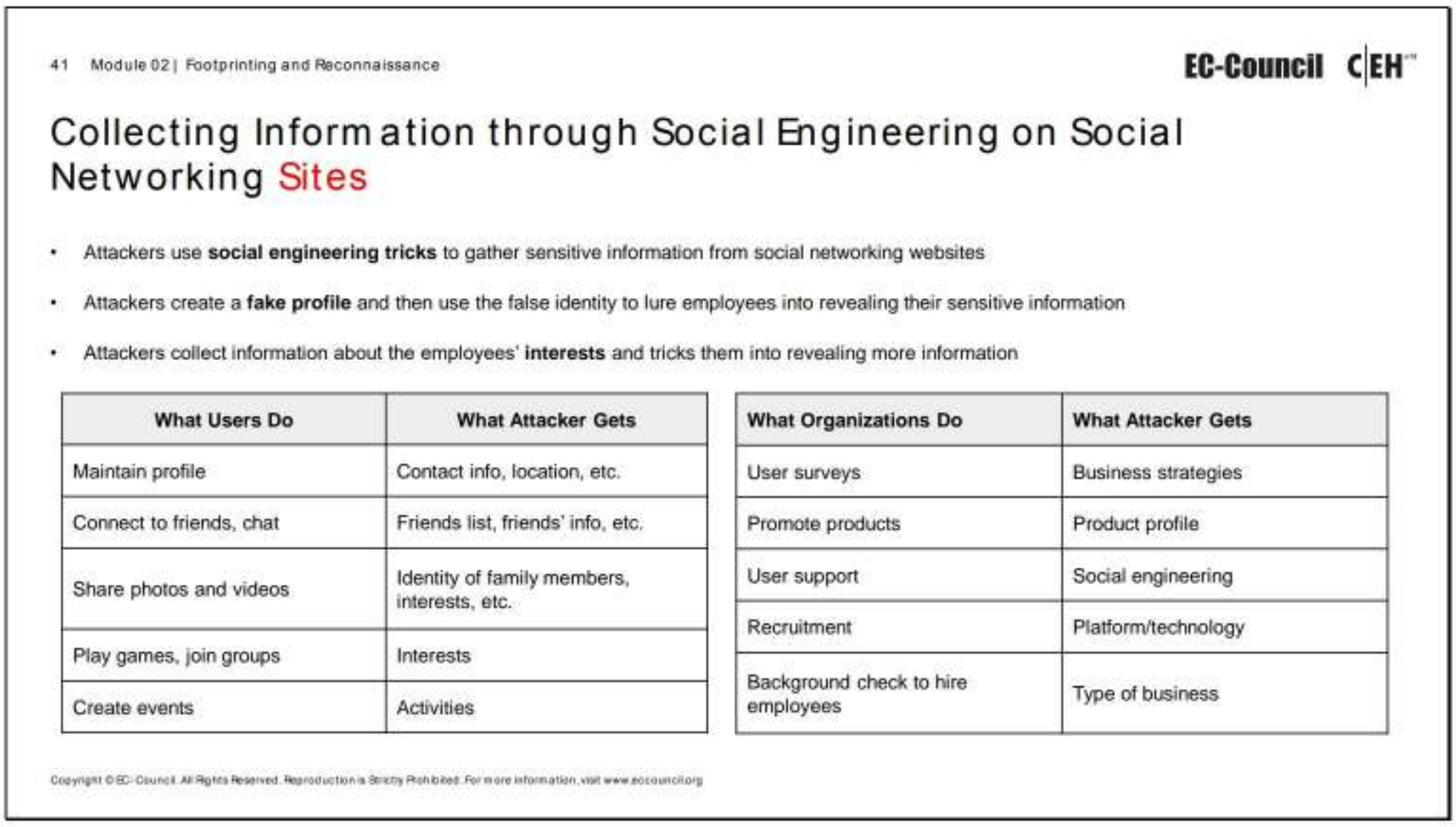
Footprinting through Social Engineering

So far, we have discussed the different techniques for gathering information using online resources or tools. Now, we will discuss footprinting through social engineering, i.e., the art of obtaining information from people by exploiting their weaknesses. This section covers the concept as well as the techniques used to gather information through social engineering.

Social engineering is a non-technical process in which an attacker misleads a person into providing confidential information inadvertently. In other words, the target is unaware of the fact that someone is stealing confidential information. The attacker takes advantage of the gullible nature of people and their willingness to provide confidential information.

To perform social engineering, an attacker first needs to gain the confidence of an authorized user and then mislead that user into revealing confidential information. The goal of social engineering is to obtain the required confidential information and then use that information for malicious purposes such as gaining unauthorized access to the system, identity theft, industrial espionage, network intrusion, fraud, and so on. The information obtained through social engineering may include credit card details, social security numbers, usernames and passwords, other personal information, security products in use, OS and software versions, IP addresses, names of servers, network layout information, and so on.

Social engineering can be performed in many ways, such as eavesdropping, shoulder surfing, dumpster diving, impersonation, tailgating, third-party authorization, piggybacking, reverse social engineering, and so on.



Collecting Information through Social Engineering on Social Networking Sites

Social networking sites are online services, platforms, or other sites that allow people to connect and to build interpersonal relations. The use of social networking sites is increasing rapidly. Examples of such sites include LinkedIn, Facebook, Instagram, Twitter, Pinterest, YouTube, and so on. Each social networking site has its own purpose and features. One site may connect friends, family and so on, while another helps users to share professional profiles. Social networking sites are open to everyone. Attackers may take advantage of this feature to gather sensitive information from users either by browsing through users' public profiles or by creating a fake profile to pose as a genuine user. On social networking sites, people may post personal information such as date of birth, educational information, employment background, spouse's names, and so on. Organizations often post information such as potential partners, websites, and upcoming news about the company.

For an attacker, social networking sites can be valuable sources of information about the target person or organization. The attacker can only gather the information that is posted by individuals. There are no barriers for attackers to access the public pages of accounts created on social networking sites. To obtain more information about the target, attackers may create fake accounts and use social engineering techniques to lure the victim into revealing more information. For example, the attacker can send a friend request to the target person from a fake account; if the victim accepts the request, then the attacker can access even the restricted pages of the target person on that website.

Information Available on Social Networking Sites

So far, we have discussed *how* an attacker can collect information from social networking sites. Now, we will discuss *what* information an attacker can get from social networking sites.

People usually maintain profiles on social networking sites to provide basic information about themselves and to help create and maintain connections with others. A profile generally contains personal information such as a person's name, contact information (cell phone number, email address), friends' information, information about family members, interests, and activities. People usually connect with friends and chat with them. Attackers can gather sensitive information through these chats. Social networking sites also allow people to share photos and videos. If users fail to set the appropriate privacy settings for their albums, then attackers can see the pictures and videos shared by them. Users may join groups to play games or to share their views and interests. Attackers can collect information about the victim's interests by tracking his or her groups and can then mislead the victim into revealing more information. Users may create events to notify other users about upcoming occasions, from which attackers will come to know about the user's activities.

The activities of users on social networking sites and the respective information that an attacker can collect is summarized in the following table.

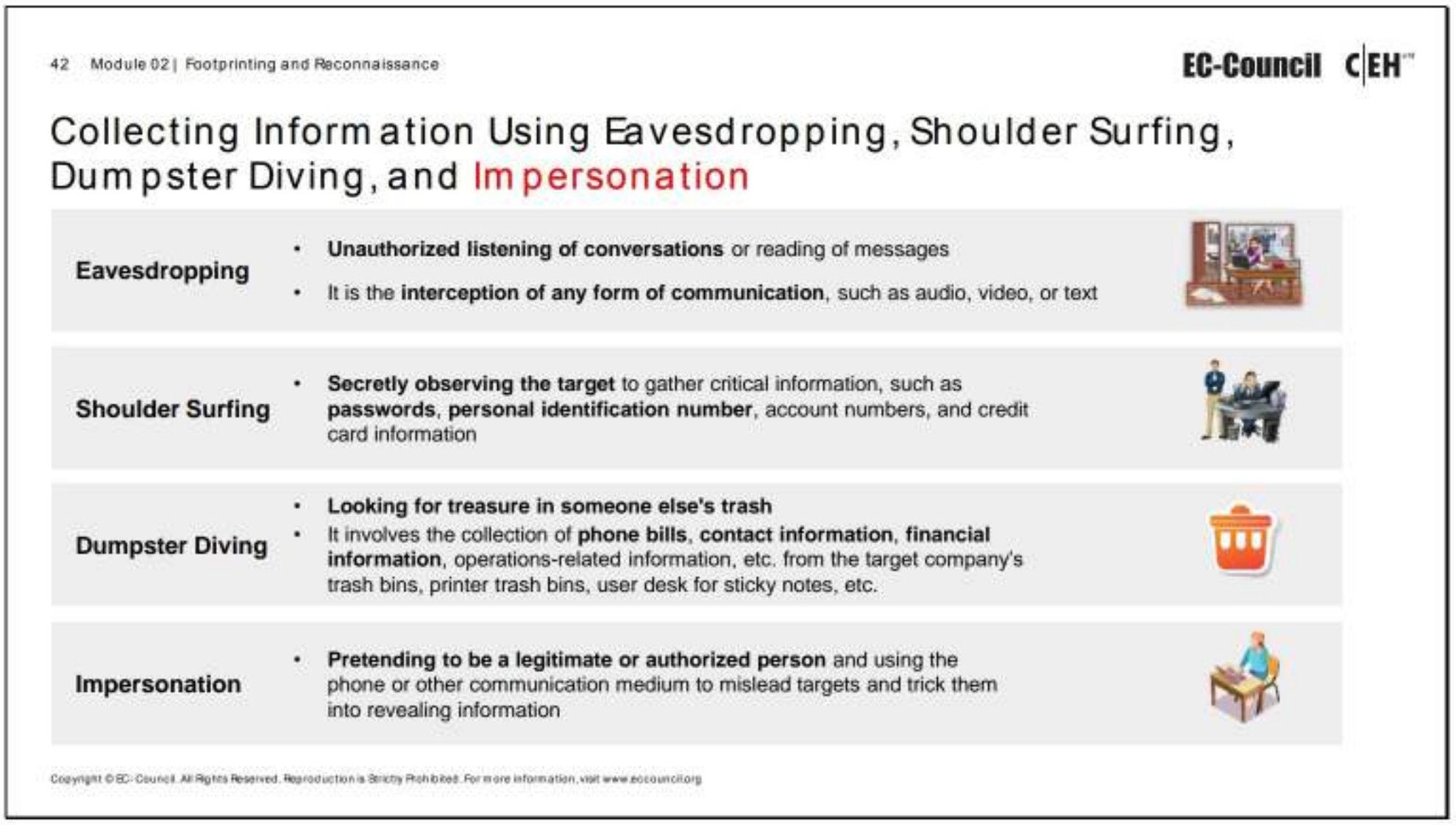
What Users Do	What Attacker Gets
Maintain profile	Contact info, location, and related information
Connect to friends, chat	Friends list, friends' info, and related information
Share photos and videos	Identity of family members, interests, and related information
Play games, join groups	Interests
Create events	Activities

Table 2.5: Activities of users on the social networking sites and the respective information

Like individuals, organizations also use social networking sites to connect with people, promote their products, and gather feedback about their products and services. The activities of an organization on social networking sites and the respective information that an attacker can collect are summarized in the table below.

What Organizations Do	What Attacker Gets
User surveys	Business strategies
Promote products	Product profile
User support	Social engineering
Recruitment	Platform/technology information
Background check to hire employees	Type of business

Table 2.6: Activities of the organization on the social networking sites and the respective information



Collecting Information Using Eavesdropping, Shoulder Surfing, Dumpster Diving, and Impersonation

Eavesdropping, shoulder surfing, dumpster diving, and impersonation are social engineering techniques widely used to collect information from people.

- **Eavesdropping**

  Eavesdropping is the act of intercepting communication in any form, such as audio, video, or text, without the consent of the communicating parties. It also includes reading confidential messages from communication media such as instant messaging or fax transmissions. The attacker can gain information by tapping phone conversations or intercepting audio, video, or written communications.

- **Shoulder Surfing**

  Shoulder surfing is a technique whereby attackers secretly observe the target to gain critical information. In the shoulder surfing technique, an attacker stands behind the victim and secretly observes the victim's activities on the computer, such as keystrokes while entering usernames, passwords, and so on. The technique is effective in gaining passwords, personal identification numbers, security codes, account numbers, credit card information, and similar data. Attackers can easily perform shoulder surfing in a crowded place, as it is relatively easy to stand behind and watch the victim without his or her knowledge.

- **Dumpster Diving**

  This uncouth technique, also known as trashing, involves the attacker rummaging for information in garbage bins. The attacker may gain vital information such as phone bills, contact information, financial information, operations-related information, printouts of

source codes, printouts of sensitive information, and so on from the target company's trash bins, printer waste bins, sticky notes at users' desks, and so on. The attacker may also gather account information from ATM trash bins. The information can help the attacker to commit attacks.

- **Impersonation**

  Impersonation is a technique whereby an attacker pretends to be a legitimate or authorized person. Attackers perform impersonation attacks personally or use phones or other communication media to mislead targets and trick them into revealing information. The attacker might impersonate a courier/delivery person, janitor, businessman, client, technician, or he/she may pretend to be a visitor. Using this technique, an attacker gathers sensitive information by scanning terminals for passwords, searching important documents on desks, rummaging bins, and so on. The attacker may even try to overhear confidential conversations and "shoulder surf" to obtain sensitive information.

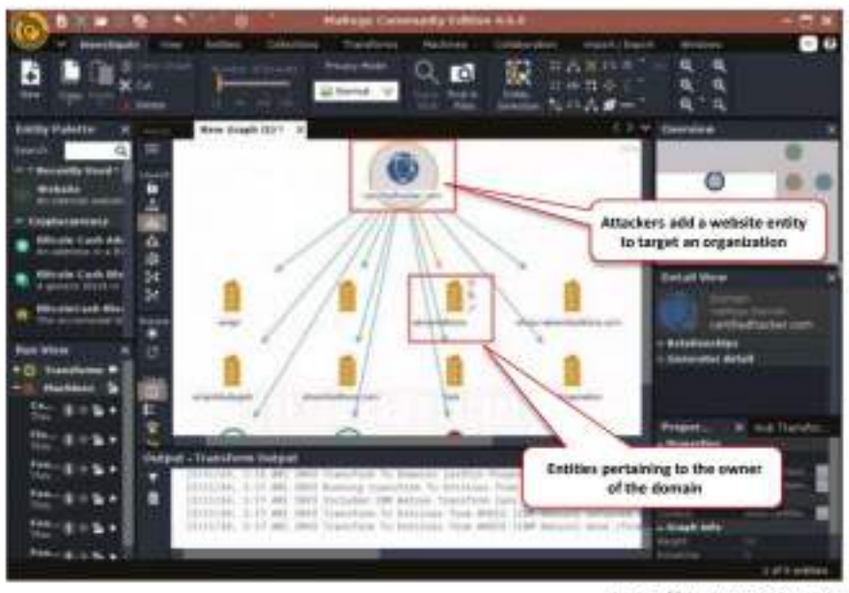
Objective 09

Automate Footprinting Tasks using Advanced Tools and AI

Active Reconnaissance

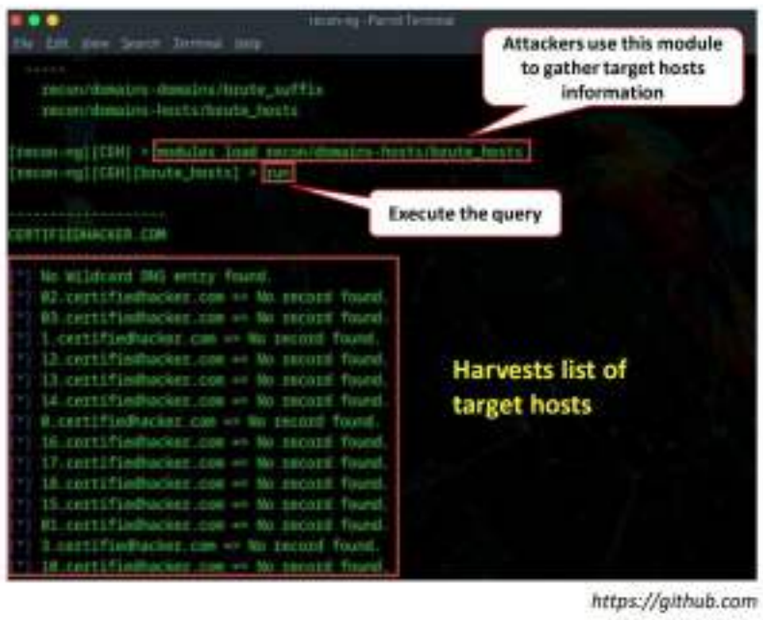
Footprinting Tools: Maltego and Recon-ng

Maltego | Maltego can be used to determine the **relationships and real world links** between people, groups of people, organizations, websites, Internet infrastructure, documents, etc.



https://www.maltego.com

Recon-ng | Recon-ng is a **Web Reconnaissance framework** with independent modules and database interaction, which provides an environment in which open source, web-based reconnaissance can be conducted



https://github.com

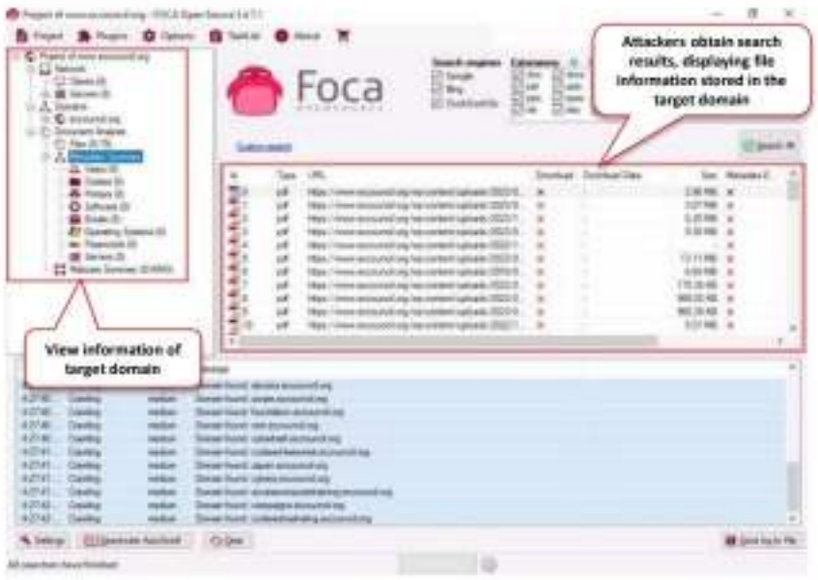
Footprinting Tools: FOCA and subfinder

FOCA

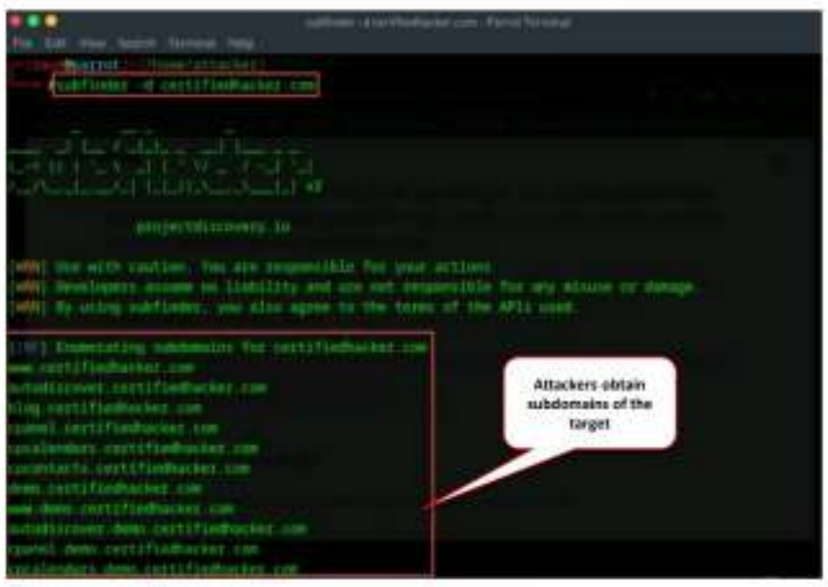
FOCA (Fingerprinting Organizations with Collected Archives) is a tool used mainly to find metadata and hidden information in the documents it scans

subfinder

subfinder is a **subdomain discovery** tool that helps attackers find valid subdomains for websites



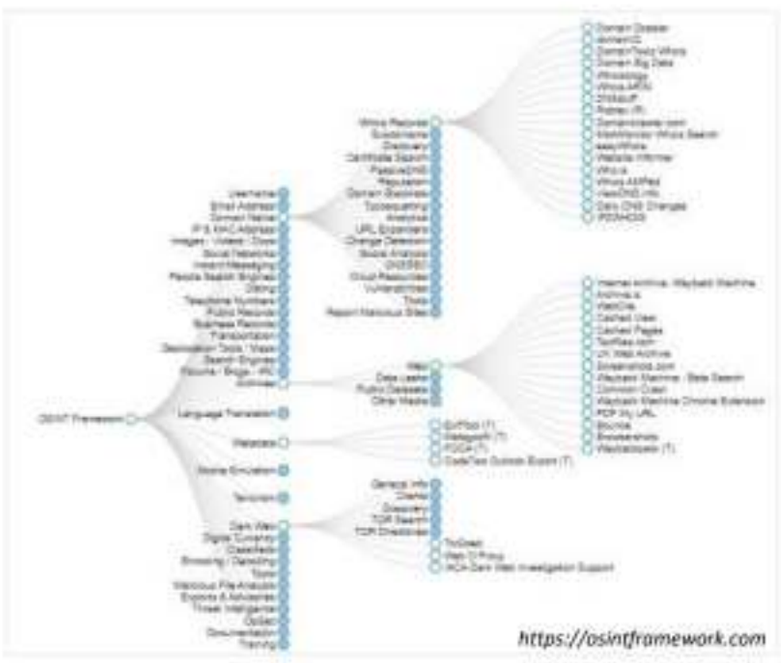
https://www.elevenpaths.com



https://github.com

Footprinting Tools: OSINT Framework

- OSINT Framework is an **open source intelligence gathering framework** that is focused on gathering information from free tools or resources
- It provides a simple web interface that lists various OSINT tools arranged by categories and is shown as **OSINT tree structure** on the web interface
- Tools listed includes the following indicators:
  - (T) - Indicates a link to a tool that must be installed and run locally
  - (D) - Google Dork
  - (R) - Requires registration
  - (M) - Indicates a URL that contains the search term and the URL itself must be edited manually



https://osintframework.com

Other Tools:

Sudomy	BillCipher	whatweb	Raccoon	Recon-Dog	Web Check	OSINT.SH
https://github.com	https://github.com	https://github.com	https://github.com	https://github.com	https://web-check.xyz	https://osint.sh

Automate Footprinting Tasks using Advanced Tools and AI

Many organizations offer tools that facilitate information gathering. This section describes the tools used to obtain information from various sources.

Footprinting tools are used to collect basic information about target systems to exploit them. Information collected by footprinting tools includes the target's IP location, routing information,

business information, address, phone number, social security number, details about the source of an email and a file, DNS information, and domain information.

- **Maltego**

  Source: *https://www.maltego.com*

  Maltego is an automated tool that can be used to determine the relationships and real-world links between people, groups of people, organizations, websites, Internet infrastructure, documents, etc.

  Attackers can use different entities available in the tool to obtain information such as email addresses, a list of phone numbers, and a target's Internet infrastructure (domains, DNS names, Netblocks, IP addresses information).

  As shown in the screenshot, attackers add a **`Website entity`**, rename it with the target's domain, and obtain the email addresses and phone numbers associated with the target.

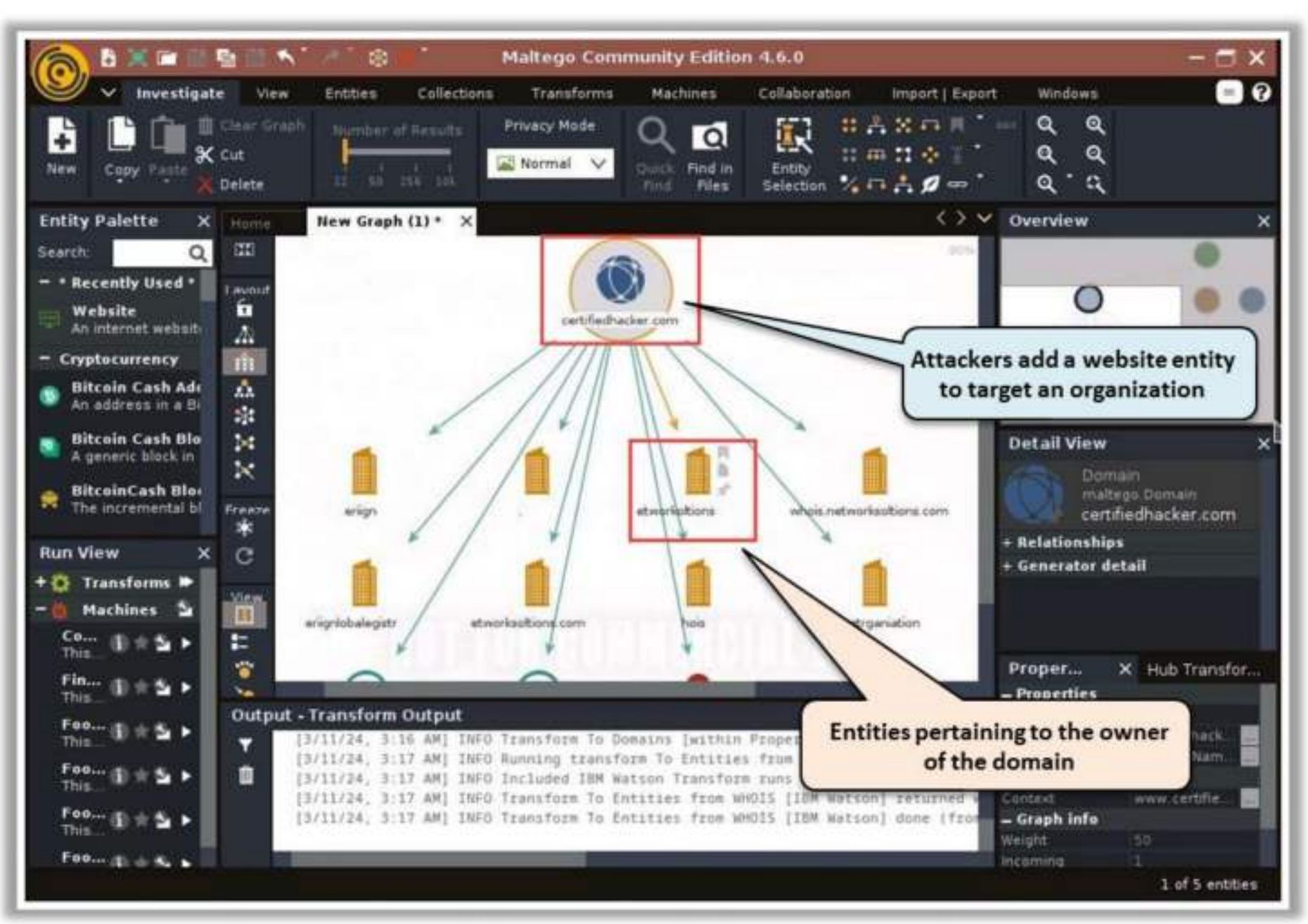


Figure 2.87: Screenshot of Maltego

- **Recon-ng**

  Source: *https://github.com*

  Recon-ng is a web reconnaissance framework with independent modules for database interaction that provides an environment in which open-source web-based reconnaissance can be conducted.

As shown in the screenshot, attackers use the module **recon/domains-hosts/brute_hosts** to extract a list of hosts associated with the target URL.

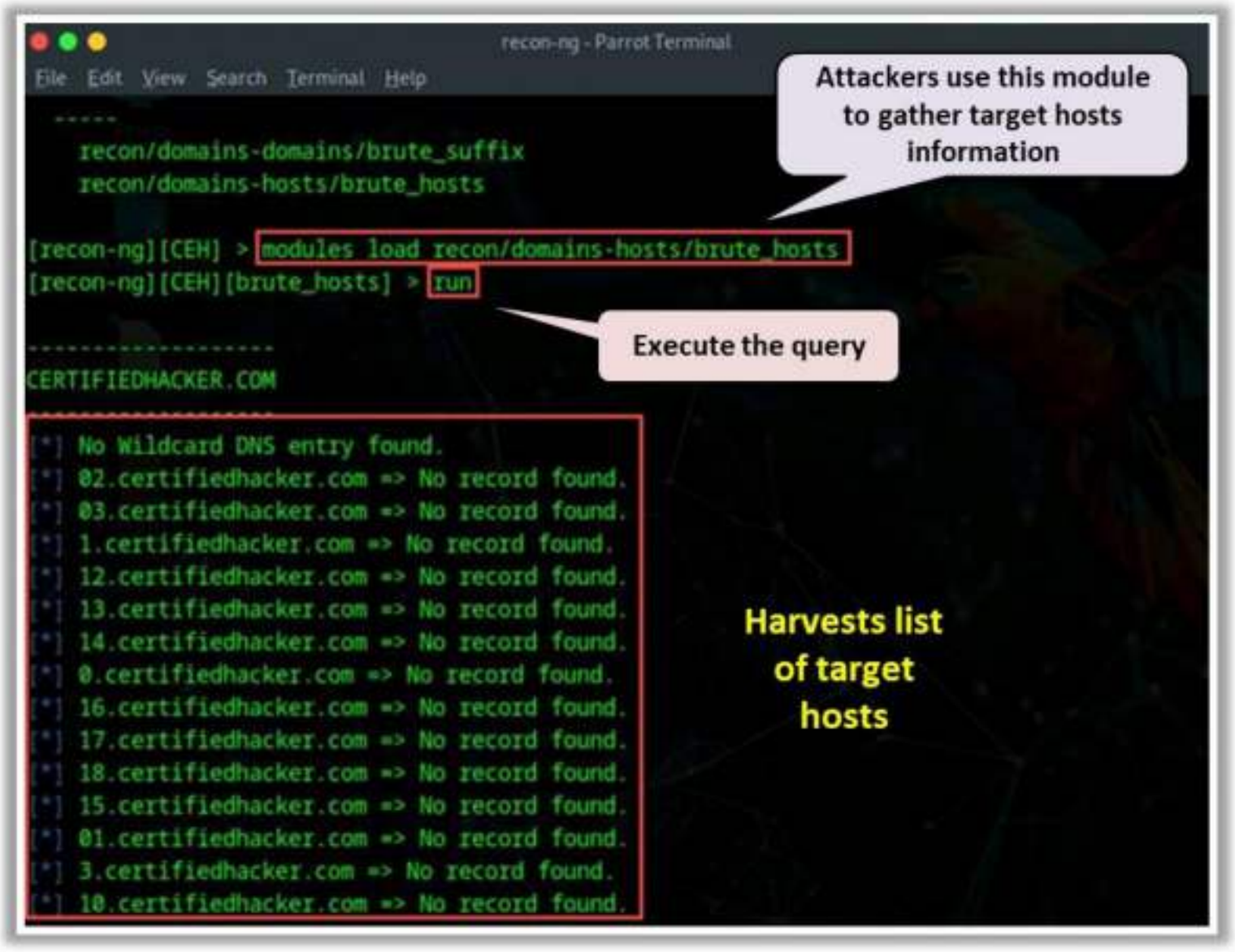


Figure 2.88: Screenshot of recon-ng

- **FOCA**

  Source: *https://www.elevenpaths.com*

  Fingerprinting Organizations with Collected Archives (FOCA) is a tool used mainly to find metadata and hidden information in the documents that its scans. FOCA is capable of scanning and analyzing a wide variety of documents, with the most common ones being Microsoft Office, Open Office, or PDF files.

  Features:

  - **Web Search** - Searches for hosts and domain names through URLs associated with the main domain. Each link is analyzed to extract information from its new host and domain names.
  - **DNS Search** - Checks each domain to ascertain the host names configured in NS, MX, and SPF servers to discover the new host and domain names.
  - **IP Resolution** - Resolves each host name by comparison with the DNS to obtain the IP address associated with this server name. To perform this task accurately, the tool performs analysis against the organization's internal DNS.
  - **PTR Scanning** - Finds more servers in the same segment of a determined address; IP FOCA executes a PTR log scan.

- **Bing IP** - Launches FOCA, which is a search process for new domain names associated with that IP address for each IP address discovered.
- **Common Names** - Perform dictionary attacks against the DNS.

As shown in the screenshot, attackers search the target domain and obtain the file information stored in it. The extracted files can be viewed on the web browser. Further, the attackers can view additional information such as network domains, roles, vulnerabilities, and metadata of the target domain.

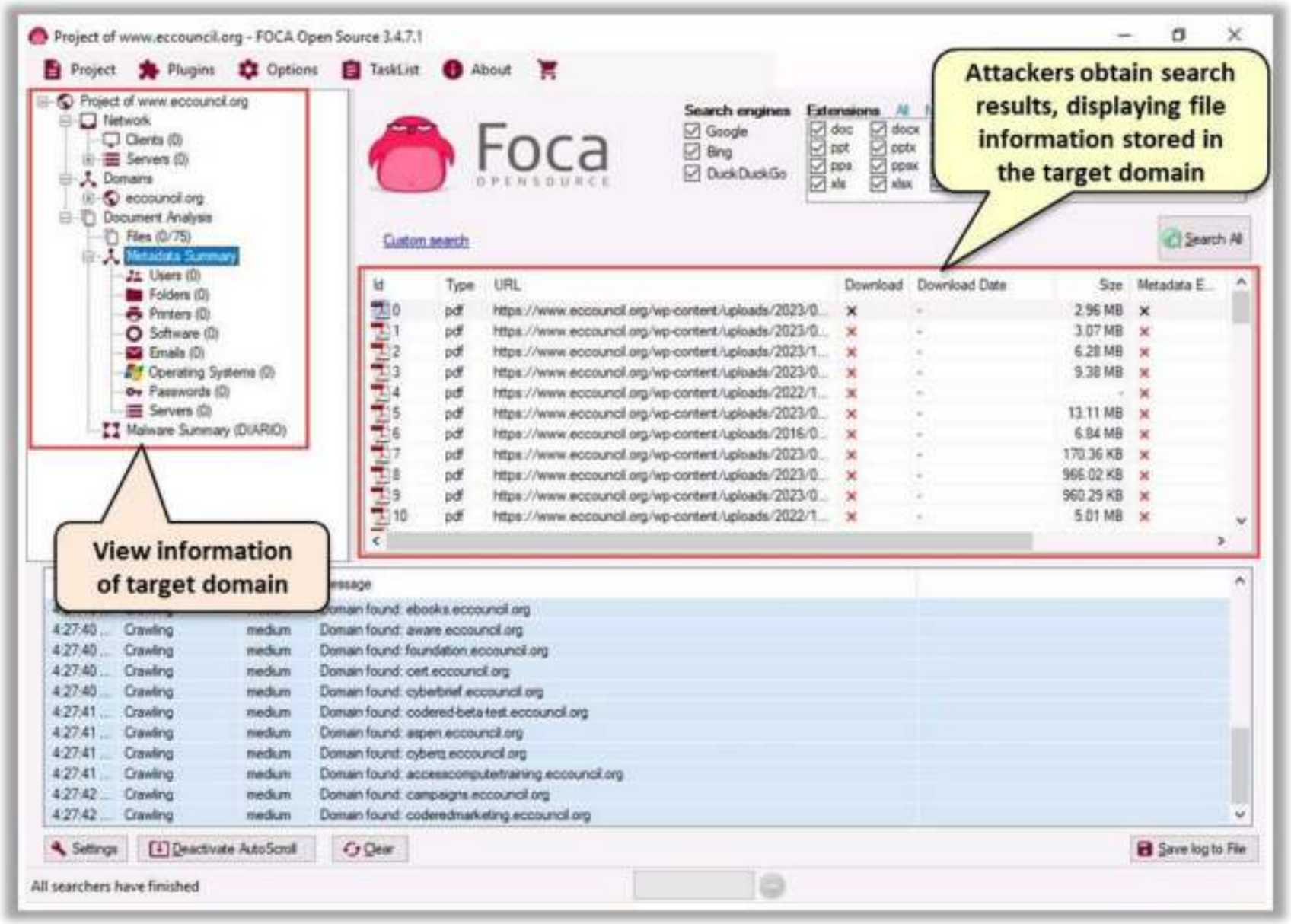


Figure 2.89: Screenshot of FOCA

- **subfinder**

Source: *https://github.com*

subfinder is a subdomain discovery tool that helps attackers find valid subdomains for websites using passive online sources. It supports multiple output formats (JSON, file, stdout).

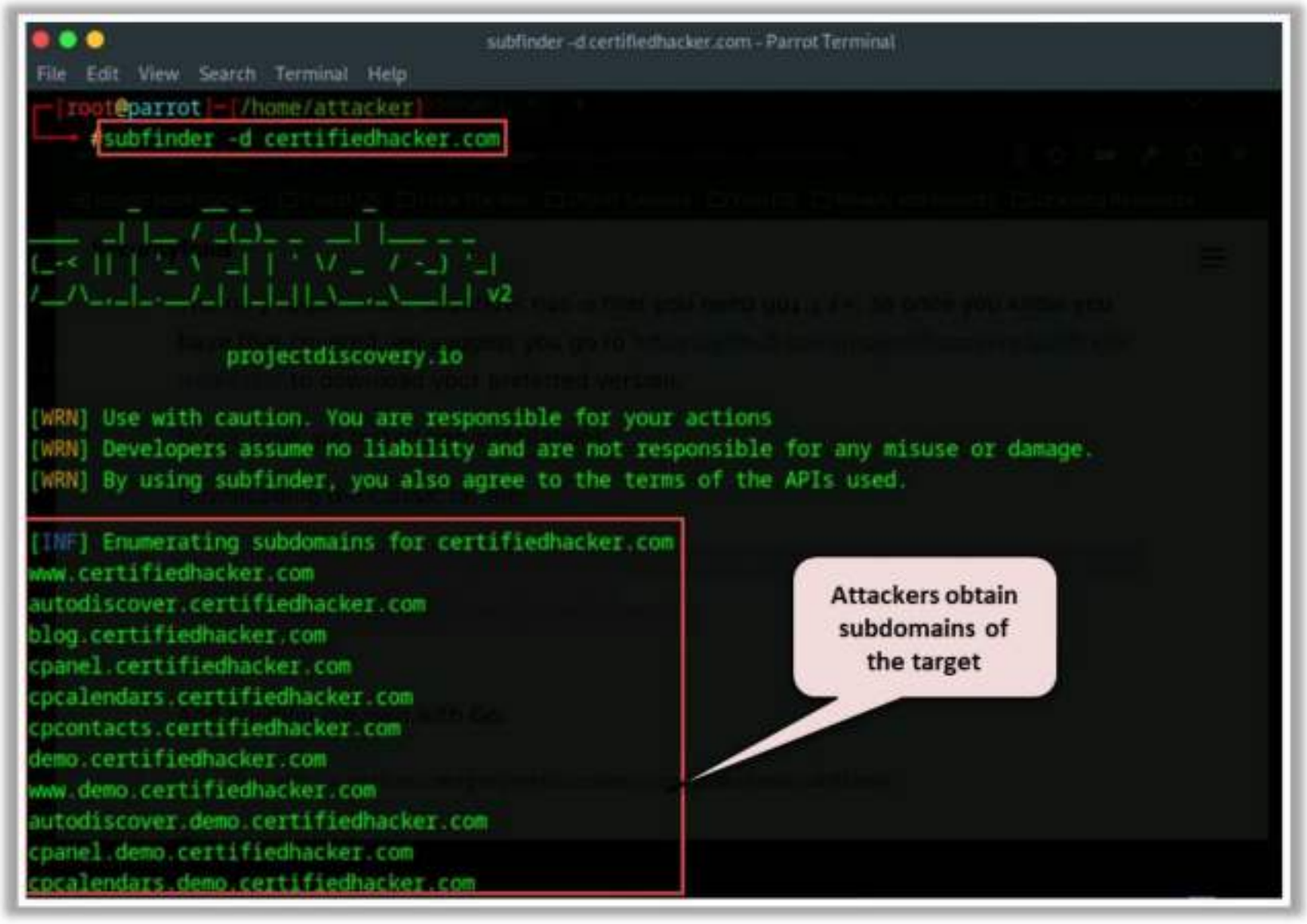


Figure 2.90: Screenshot of subfinder

- **OSINT Framework**

  Source: *https://osintframework.com*

  OSINT Framework is an open source intelligence gathering framework that helps security professionals in performing automated footprinting and reconnaissance, OSINT research, and intelligence gathering. It is focused on gathering information from free tools or resources. This framework includes a simple web interface that lists various OSINT tools arranged by category, and it is shown as an OSINT tree structure on the web interface.

  As shown in the screenshot, the tools listed include the following indicators:

  - (T) - Indicates a link to a tool that must be installed and run locally
  - (D) - Google dork
  - (R) - Requires registration
  - (M) - Indicates a URL that contains the search term and the URL itself must be edited manually

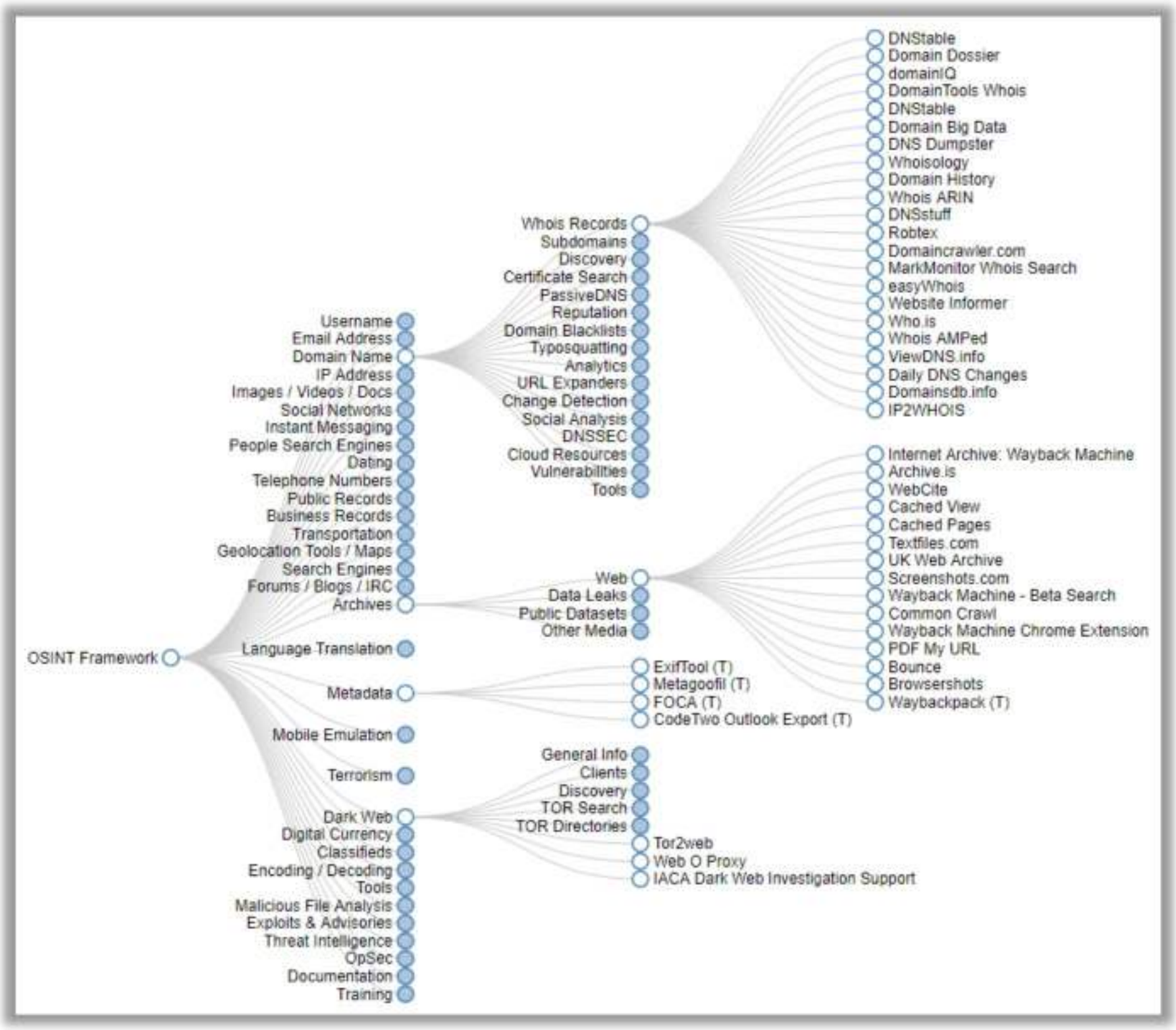


Figure 2.91: Screenshot of OSINT Framework

- **Recon-Dog**

  Source: *https://www.github.com*

  Recon-Dog is an all-in-one tool for all basic information gathering needs. It uses APIs to collect information about the target system.

  Features:

  - **Censys**: Uses censys.io to gather a massive amount of information about an IP address.
  - **NS lookup**: Performs name server lookup
  - **Port scan**: Scans most common TCP ports
  - **Detect CMS**: Can detect 400+ content management systems
  - **Whois lookup**: Performs a Whois lookup
  - **Detect honeypot**: Uses shodan.io to check if the target is a honeypot

- **Find subdomains**: Uses findsubdomains.com to find subdomains
- **Reverse IP lookup**: Performs a reverse IP lookup to find domains associated with an IP address
- **Detect technologies**: Uses wappalyzer.com to detect 1000+ technologies
- **All**: Runs all utilities against the target

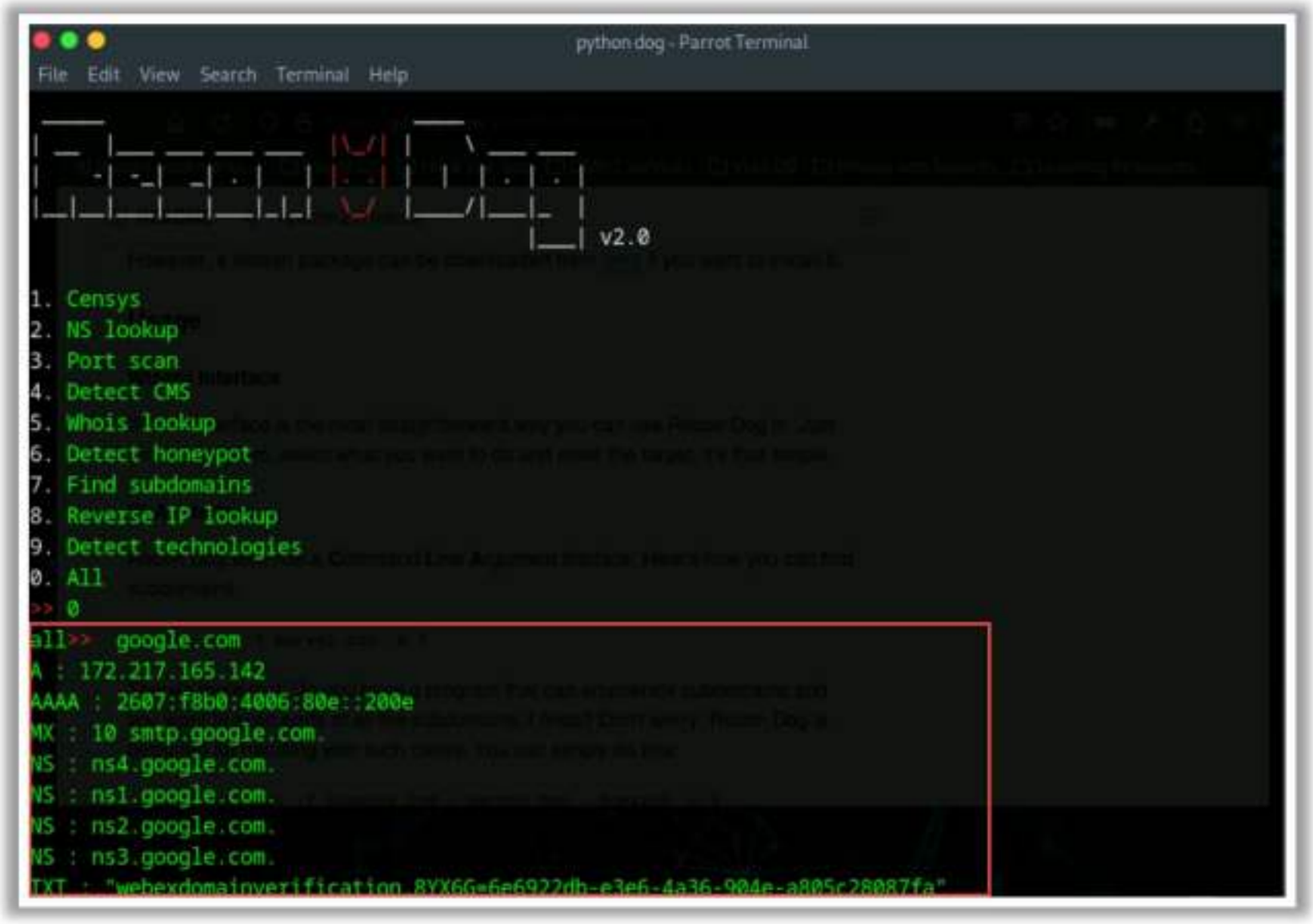


Figure 2.92: Screenshot of Recon-Dog

- **BillCipher**

  Source: *https://www.github.com*

  BillCipher is an information gathering tool for a website or IP address. It can work on any operating system that supports Python 2, Python 3, and Ruby. This tool includes various options such as DNS lookup, Whois lookup, port scanning, zone transfer, host finder, and reverse IP lookup, which help to gather critical information.

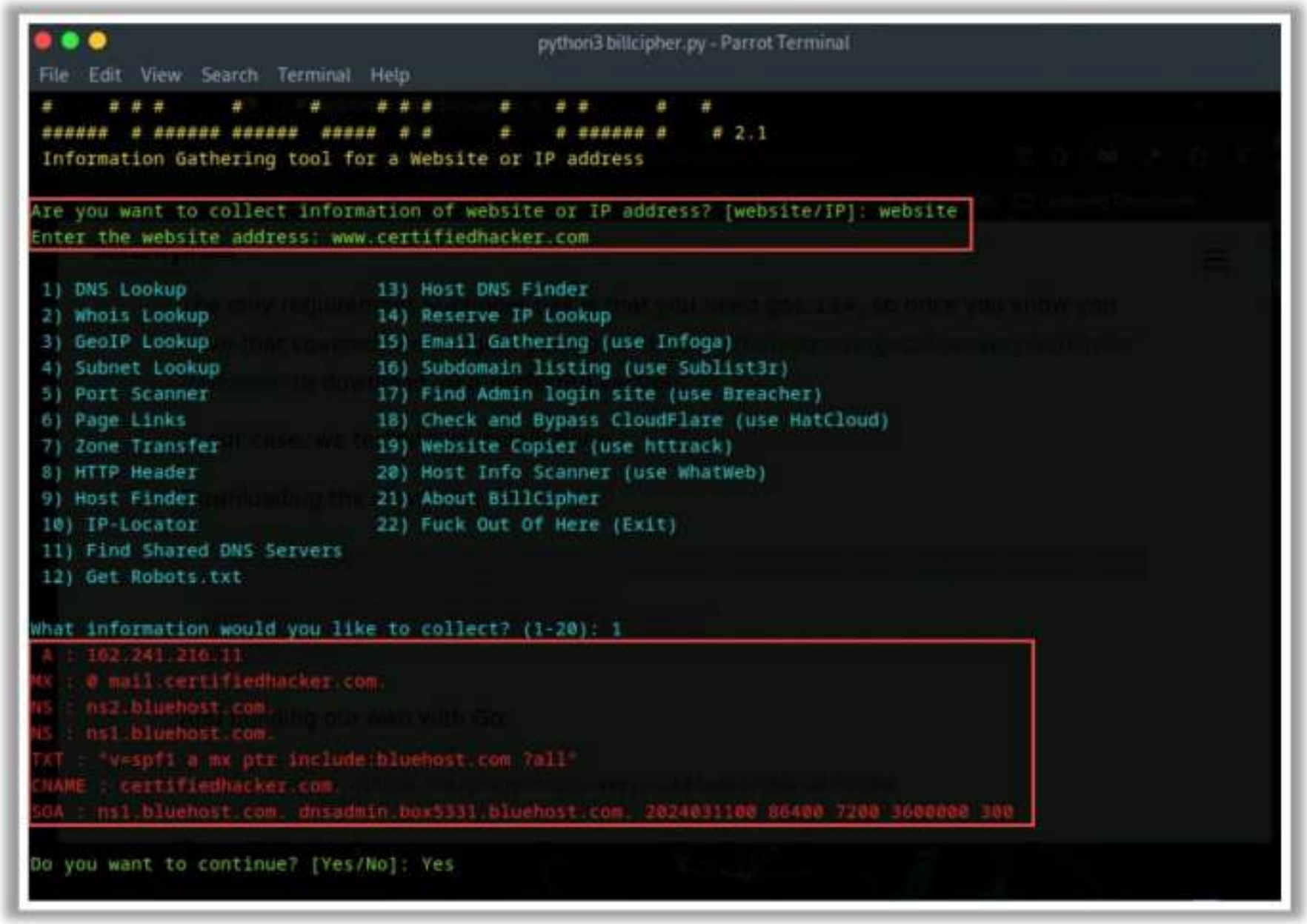
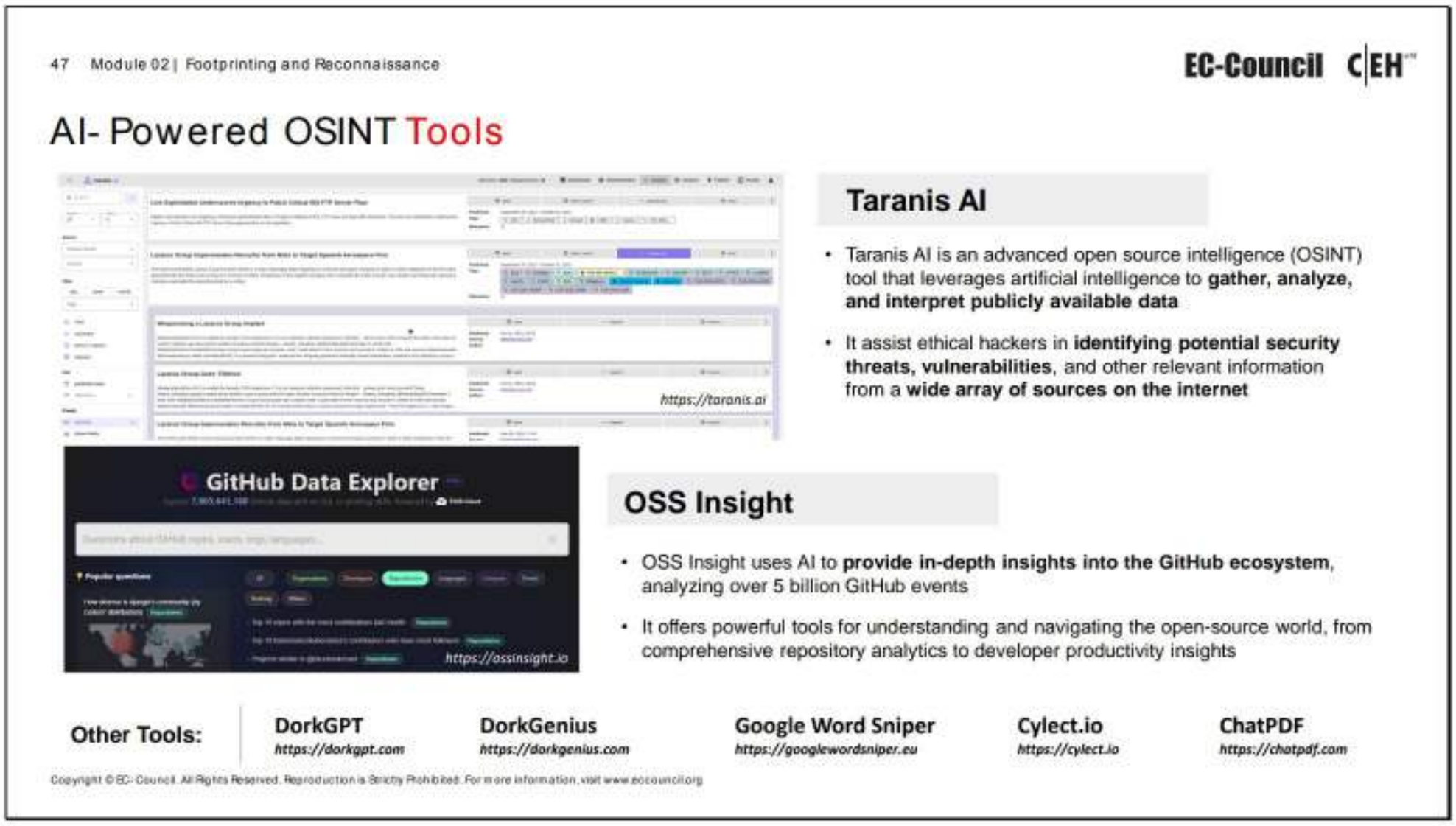


Figure 2.93: Screenshot of BillCipher

Some additional footprinting tools are listed below:

- Sudomy (*https://github.com*)
- theHarvester (*https://www.edge-security.com*)
- whatweb (*https://github.com*)
- Raccoon (*https://github.com*)
- Orb (*https://github.com*)
- Web Check (*https://web-check.xyz*)
- OSINT.SH (*https://osint.sh*)



AI-Powered OSINT Tools

AI has revolutionized open-source intelligence (OSINT) by significantly enhancing investigative capabilities through advanced data collection, analysis, and prediction. AI automates data processing, extracts relevant insights, delivers actionable intelligence more efficiently than traditional methods, and enhances the OSINT tools.

AI-powered tools offer numerous advantages for OSINT. The following are some key use cases in which AI can significantly benefit OSINT researchers.

- **Web Scraping**: AI techniques utilize online data from sources such as social media, blogs, forums, and deep web databases. This data enables the tracking of entities over time or the monitoring of public behavior. Machine-learning models can automate the extraction of specific information such as social media comments and replies.
- **Pattern Recognition**: Machine learning (ML) techniques can identify entities within large datasets and analyze files to identify the relationships between different entities. These entities include names, company details, addresses, emails, phone numbers, and relevant data.
- **Content Summarization**: NLP algorithms can summarize large volumes of data. OSINT gatherers can utilize this capability to extract pertinent information from extensive datasets. For example, an AI summarization tool can extract company names from a set of PDF files spanning hundreds of pages.
- **Sentiment Analysis**: AI technology can interpret human emotions through text analysis, which is particularly useful for understanding public sentiment. OSINT researchers can use AI to assess the emotional state of users based on social media posts and comments or to predict consumer behavior based on reviews.

- **Image Recognition**: Computer vision, a subset of AI, can analyze digital media files such as images and videos. In OSINT investigations, computer vision can assist in:
  - **Face Recognition**: Identifying and tracking individuals across different media.
  - **Metadata Analysis**: Extracting metadata from digital files.
  - **Reverse Image Search**: Enhancing reverse image search capabilities and detecting deepfake images.
- **AI Detection**: AI can also identify content generated by other AI tools, which is crucial for detecting malicious activities facilitated by AI.

Benefits of Integrating AI in OSINT:

- **Improved Efficiency**: AI enhances OSINT efficiency by automating tasks such as web scraping and data extraction, which accelerates the data collection and analysis processes. This allows investigators to focus on higher-level analysis and decision-making, ultimately expediting investigations and delivering timely insights.
- **Greater Scope**: AI expands the scope of OSINT by analyzing vast data from the surface web, deep web, and dark web, ensuring comprehensive intelligence coverage. Its ability to process large datasets and identify connections enables investigators to uncover hidden patterns and relationships that are difficult to detect manually.
- **Enhanced Visibility**: AI enhances the visibility of intelligence data by connecting billions of seemingly unrelated data points to coherent networks of information, enabling investigators to quickly identify suspicious activities and establish connections between threat actors or events. AI-powered tools present these networks as user-friendly graphical interfaces, making it simpler for investigators to recognize and comprehend complex relationships and trends within the data.
- **Increased Investigator Safety**: AI enhances the safety of investigators by enabling anonymized and automated investigations. This reduces the risk of exposing an investigator's identity or compromising sensitive information. AI tools can conduct thorough investigations without the need for direct human involvement in potentially dangerous environments such as the dark web.

AI-Powered OSINT Tool: Taranis AI

Source: https://taranis.ai

Taranis AI is an advanced OSINT tool uses AI to enhance information gathering and situational analyses. It uses NLP and AI to improve the quality of data received from data sources, such as websites, to gather unstructured news articles. Analysts then transform these AI-enhanced articles into organized reports that are used as the basis for deliverables such as PDF files that are eventually published.

Features of Taranis AI:

- **Advanced OSINT Capabilities:** Taranis AI searches multiple data sources, including websites, to collect unstructured news articles and provides a comprehensive and enriched intelligence feed.
- **AI-Enhanced Analysis:** Taranis enhances the collected articles, ensuring higher content quality and relevance using AI and NLP.
- **Multi-Format Output:** Taranis AI creates a variety of end products, including structured reports and PDF files, tailored to meet specific informational needs and requirements.
- **Seamless Publishing:** The platform enables easy publication of finalized intelligence products, ensuring timely dissemination of critical information to stakeholders.

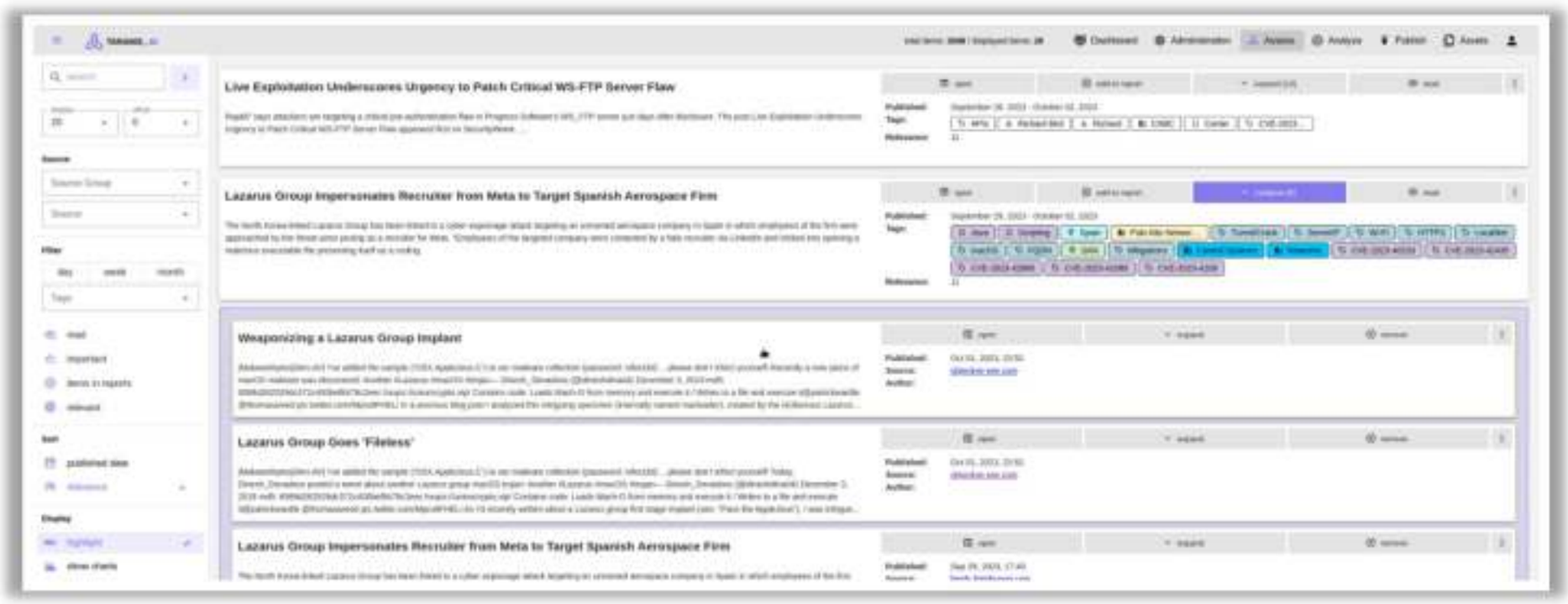


Figure 2.94: Screenshot of Taranis AI

AI-Powered OSINT Tool: OSS Insight

Source: *https://ossinsight.io*

OSS Insight leverages AI to delve deep into the GitHub ecosystem by analyzing an extensive dataset of over five billion GitHub events. This capability enables it to offer comprehensive insights and tools to enhance the understanding and navigation of the open-source world. From detailed repository analytics encompassing metrics such as stars, forks, and commits to insights into developer productivity and collaboration patterns, OSS Insight is equipped with powerful resources for informed decision-making and strategic planning in open-source software development.

Key Features of OSS Insight

- **GPT-Powered Data Exploration**: It enables users to query GitHub data using natural languages, generate SQL queries, and present the results visually. It helps ethical hackers gather intelligence on repositories, developer activities, and trends without requiring advanced SQL skills, thereby enhancing investigative efficiency.
- **Technical Fields Analytics**: The platform curates GitHub collections in technical domains and provides analytics in areas such as web frameworks, AI, and Web3. Ethical hackers

can use this data to identify emerging technologies, potential vulnerabilities, and understand specialized security landscapes.

- **Developer Analytics**: The tool monitors developer productivity metrics such as commits, pull requests, and code contributions while also analyzing collaboration behaviors and engagement levels. Ethical hackers can use this to assess contributor activity and reliability, identify potential weaknesses in security practices, and identify high-risk or low-activity contributors.
- **Repository Analytics**: This tool assesses GitHub repository metrics, such as popularity (stars and forks), update frequency, and community engagement. It provides historical trends and comparative insights that aid in benchmarking and strategic decisions. Ethical hackers can analyze the health and security postures of repositories, track how issues and pull requests are handled, and identify repositories that are vulnerable to attacks owing to low engagement or outdated code.
- **Compare Projects**: This enables easy comparison of metrics from different GitHub projects, including activity levels, contributor demographics, issue-handling efficiency, and technical metrics. This helps ethical hackers identify projects with better security practices, active communities, or potential vulnerabilities for strategic targeting and assessment.

Importance of OSS Insight in OSINT for Ethical Hacking

- **Comprehensive Data Analysis**: By analyzing over five billion GitHub events, OSS Insight provides a wealth of data that ethical hackers can use to gather intelligence on software vulnerabilities, popular frameworks, and emerging trends. These data can inform vulnerability assessments and aid in the identification of potential attack vectors.
- **Real-time and Historical Data**: The integration of real-time data updates and historical data from the GHArchive ensures that ethical hackers have access to up-to-date information on ongoing developments and past incidents within the GitHub ecosystem.
- **AI-Powered Querying**: The AI-powered GitHub Data Explorer simplifies the querying of complex datasets using natural language, making it easier for ethical hackers to extract specific information relevant to their investigations.

Figure 2.95: Screenshot of OSS Insight

Additional AI-Powered OSINT Tools

AI OSINT tools that leverage artificial intelligence to enhance the efficiency and accuracy of open-source intelligence gathering are as follows:

- **DorkGPT**

  Source: https://dorkgpt.com

  DorkGPT is an AI-powered tool designed to assist Google Dorking, a technique used to find information that is not easily accessible through regular search queries. It leverages the capabilities of GPT (Generative Pre-trained Transformer) models to generate and refine search queries, helping users uncover sensitive information, hidden pages, and other data that may be relevant to cybersecurity, ethical hacking, or research purposes.

- **DorkGenius**

  Source: https://dorkgenius.com

  DorkGenius is an AI-powered tool that automates Google Dorking and helps users generate advanced search queries to find specific information on the internet. It is useful for uncovering hidden files, directories, sensitive information, and security vulnerabilities, particularly in the case of ethical hackers.

- **Google Word Sniper**

  Source: https://googlewordsniper.eu

  Google Word Sniper helps to refine search queries for more effective Google results. It identifies targeted keywords and phrases, making it easier to find specific information, hidden content, and niche data. This tool is valuable for researchers, marketers, and cybersecurity professionals, as it enhances their ability to uncover valuable buried information in search results.

- **Cylect.io**

  Source: https://cylect.io

  Cylect.io is an advanced AI-powered OSINT tool that integrates multiple databases into a user-friendly interface, providing a vast collection of resources for ethical hackers and enabling efficient and confident OSINT investigations. Developed to address the inefficiencies of traditional search engines, Cylect.io simplifies the search process and enhances the speed and accuracy of data collection in investigative contexts.

- **ChatPDF**

  Source: https://chatpdf.com

  ChatPDF is an OSINT tool that leverages AI to analyze and extract information from PDF documents through a conversational interface. Users can upload PDF files and interact with the tool to quickly retrieve specific data, summaries, and insights, making it a valuable resource for ethical hacking.

- **Bardeen.ai**

  Source: https://www.bardeen.ai

  Bardeen.ai is an automation tool that can be used for OSINT by enabling users to streamline and automate data collection and analysis processes from various online sources. This enhances the speed and accuracy of OSINT activities, making them useful assets for cybersecurity professionals, researchers, and investigators.

- **DarkGPT**

  Source: https://github.com/luijait/DarkGPT

  DarkGPT is an AI assistant that uses GPT-4-200K to query leaked databases, aiding in efficient and targeted searches within compromised data sources. This enables users to extract vital information and insights, enhancing the OSINT capabilities of cybersecurity analysts and researchers.

- **PenLink Cobwebs**

  Source: https://cobwebs.com

  PenLink Cobwebs is an advanced AI-powered OSINT tool that specializes in gathering and analyzing data from various online sources. It offers comprehensive capabilities for collecting, processing, and visualizing information to support cybersecurity investigations.

- **Explore AI**

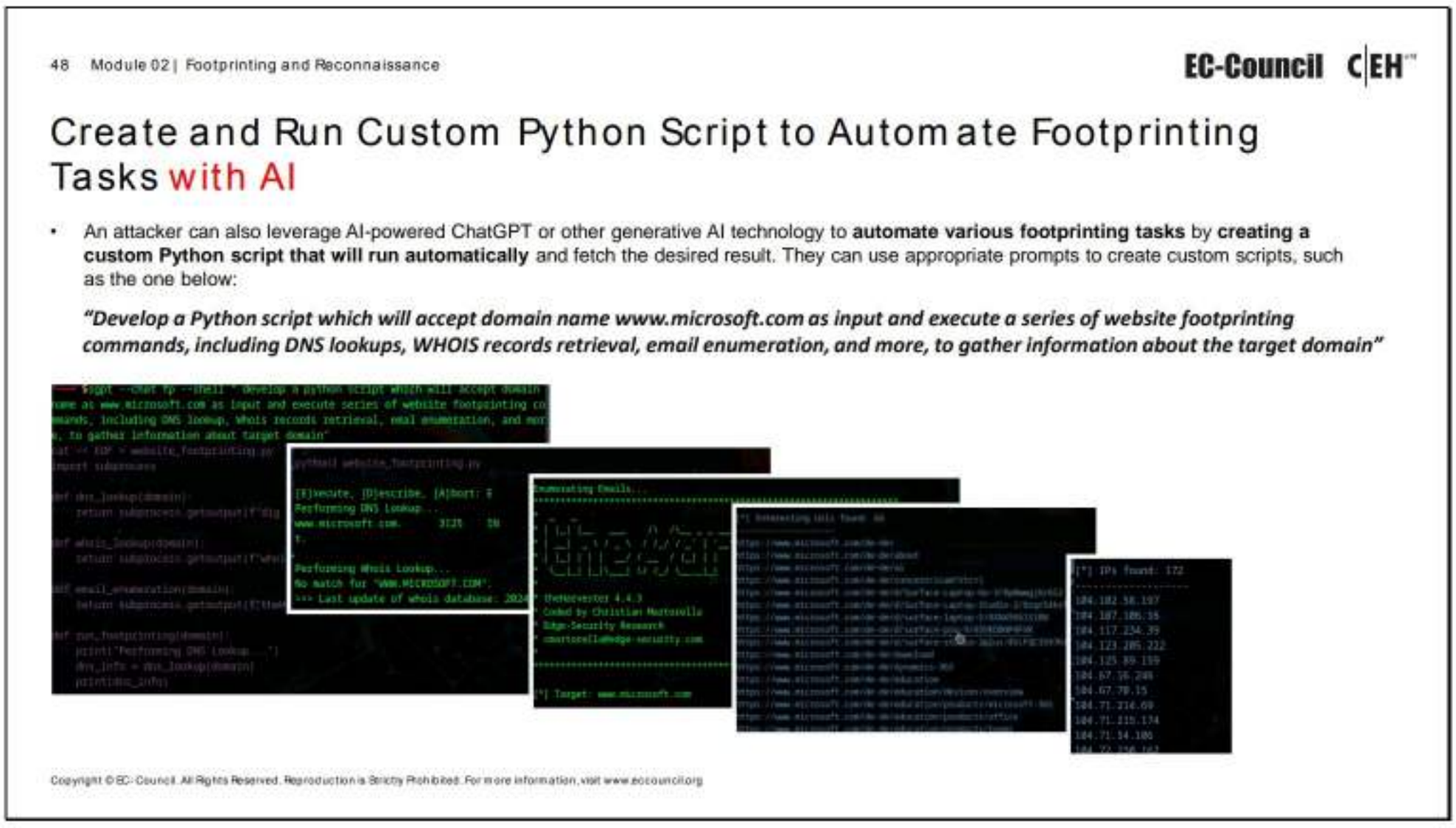
  Source: https://exploreai.vercel.app

  Explore AI is an AI-powered YouTube search engine that uses artificial intelligence to search for and extract information from YouTube videos, making it easier to access information for ethical hacking purposes.

- **AnyPicker**

  Source: https://app.anypicker.com

  AnyPicker is a powerful visual web scraper and AI OSINT tool designed to extract data from websites without requiring coding skills. This tool supports scraping multiple pages simultaneously and provides a real-time preview of the extraction results, offering flexibility and efficiency in web data collection.



Create and Run Custom Python Script to Automate Footprinting Tasks with AI

Attackers can leverage AI-powered technologies to enhance and automate their footprinting tasks. With the aid of AI, attackers can effortlessly create and run custom footprinting scripts and acquire valuable insights about their targets. By developing custom scripts empowered by artificial intelligence (AI), attackers can efficiently execute a series of website footprinting commands to gather information about a target domain for cybersecurity assessments.

For example,

Attackers can use ChatGPT to guide the development of such a script by using an appropriate prompt such as:

"Develop a Python script which will accept the domain name www.microsoft.com as input and execute a series of website footprinting commands, including DNS lookups, WHOIS records retrieval, email enumeration, and more, to gather information about the target domain."

```
└──╼ $sgpt --chat fp --shell " develop a python script which will accept domain
name as www.microsoft.com as input and execute series of website footprinting co
mmands, including DNS lookup, Whois records retrieval, emal enumeration, and mor
e, to gather information about target domain"
cat << EOF > website_footprinting.py
import subprocess

def dns_lookup(domain):
    return subprocess.getoutput(f"dig {domain} ANY +noall +answer")

def whois_lookup(domain):
    return subprocess.getoutput(f"whois {domain}")

def email_enumeration(domain):
    return subprocess.getoutput(f"theHarvester -d {domain} -b all -l 100")

def run_footprinting(domain):
    print("Performing DNS Lookup...")
    dns_info = dns_lookup(domain)
    print(dns_info)
```

Figure 2.96: Prompt and Output for performing Traceroute with AI – Screenshot 1

```
    print("\\nPerforming Whois Lookup...")
    whois_info = whois_lookup(domain)
    print(whois_info)

    print("\\nEnumerating Emails...")
    emails = email_enumeration(domain)
    print(emails)

domain = 'www.microsoft.com'
run_footprinting(domain)

EOF

python3 website_footprinting.py
```

Figure 2.97: Prompt and Output for performing Traceroute with AI – Screenshot 2

The following shell script is designed to perform website footprinting including DNS lookups, WHOIS records retrieval, email enumeration, and more, to gather information about the target domain: www.microsoft.com

```
cat <<< EOF > website_footprinting.py
import subprocess


def dns_lookup(domain):
    return subprocess.getoutput(f"dig {domain} ANY +noall +answer")
```

```
def whois_lookup(domain):
    return subprocess.getoutput(f"whois {domain}")

def email_enumeration(domain):
    return subprocess.getoutput(f"theHarvester -d {domain} -b all -l 100")

def run_footprinting(domain):
    print("Performing DNS Lookup...")
    dns_info = dns_lookup(domain)
    print(dns_info)

    print("\nPerforming Whois Lookup...")
    whois_info = whois_lookup(domain)
    print(whois_info)

    print("\nEnumerating Emails...")
    emails = email_enumeration(domain)
    print(emails)

domain = 'www.microsoft.com'
run_footprinting(domain)

EOF

python3 website_footprinting.py
```

Explanation:

This Python script defines four functions: dns_lookup, whois_lookup, email_enumeration, and run_footprinting.

- **`dns_lookup(domain)`**: Performs a DNS lookup for the specified domain using the dig command.
- **`whois_lookup(domain)`**: Retrieves WHOIS records for the specified domain using the whois command.
- **`email_enumeration(domain)`**: Enumerates emails associated with the specified domain using theHarvester tool.
- **`run_footprinting(domain)`**: Executes a series of website footprinting commands (DNS lookup, WHOIS lookup, email enumeration) for the given domain and prints the results.

You can run this script using Python3 to perform website footprinting on the specified domain (www.microsoft.com).

```
python3 website_footprinting.py

[E]xecute, [D]escribe, [A]bort: E
Performing DNS Lookup...
www.microsoft.com.      3125    IN      CNAME   www.microsoft.com-c-3.edgekey.n
t.

Performing Whois Lookup...
No match for "WWW.MICROSOFT.COM".
>>> Last update of whois database: 2024-03-13T13:25:50Z <<<
```

Figure 2.98: Output file for script with AI – Screenshot 1

```
Enumerating Emails...
********************************************************************************
*                                                                              *
*  _   _                                                 _                     *
* | |_| |__   ___    /\  /\__ _ _ __ _____  _____  ___| |_ ___ _ __          *
* | __|  _ \ / _ \  / /_/ / _` | '__\ V / _ \/ __|/ _ \ __/ _ \ '__|         *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \  __/ ||  __/ |           *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\___|\__\___|_|           *
*                                                                              *
* theHarvester 4.4.3                                                           *
* Coded by Christian Martorella                                                *
* Edge-Security Research                                                       *
* cmartorella@edge-security.com                                                *
*                                                                              *
********************************************************************************

[*] Target: www.microsoft.com
```

Figure 2.99: Output file for script with AI – Screenshot 2

```
[*] Interesting Urls found: 64
------------------
https://www.microsoft.com/de-de/
https://www.microsoft.com/de-de/about
https://www.microsoft.com/de-de/ai
https://www.microsoft.com/de-de/concern/scam?rtc=1
https://www.microsoft.com/de-de/d/Surface-Laptop-Go-3/8p0wwgj6c6l2
https://www.microsoft.com/de-de/d/Surface-Laptop-Studio-2/8rqr54krf1dz
https://www.microsoft.com/de-de/d/surface-laptop-5/8XN49V61S1BN
https://www.microsoft.com/de-de/d/surface-pro-9/93VKD8NP4FVK
https://www.microsoft.com/de-de/d/surface-studio-2plus/8VLFQC3597K4
https://www.microsoft.com/de-de/download
https://www.microsoft.com/de-de/dynamics-365
https://www.microsoft.com/de-de/education
https://www.microsoft.com/de-de/education/devices/overview
https://www.microsoft.com/de-de/education/products/microsoft-365
https://www.microsoft.com/de-de/education/products/office
https://www.microsoft.com/de-de/education/products/teams
```

Figure 2.100: Output file for script with AI – Screenshot 3

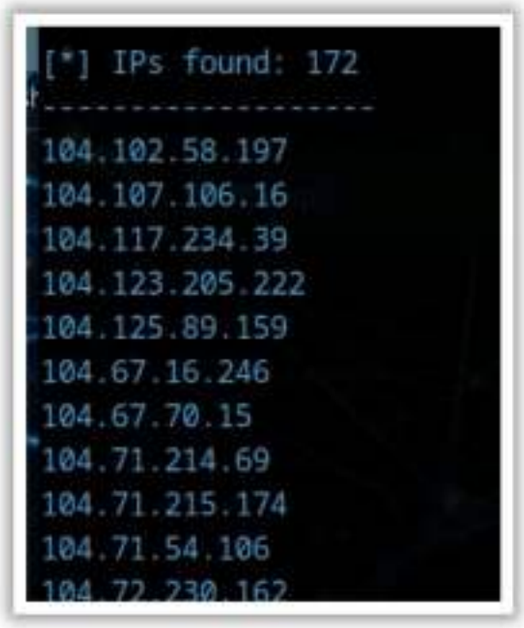


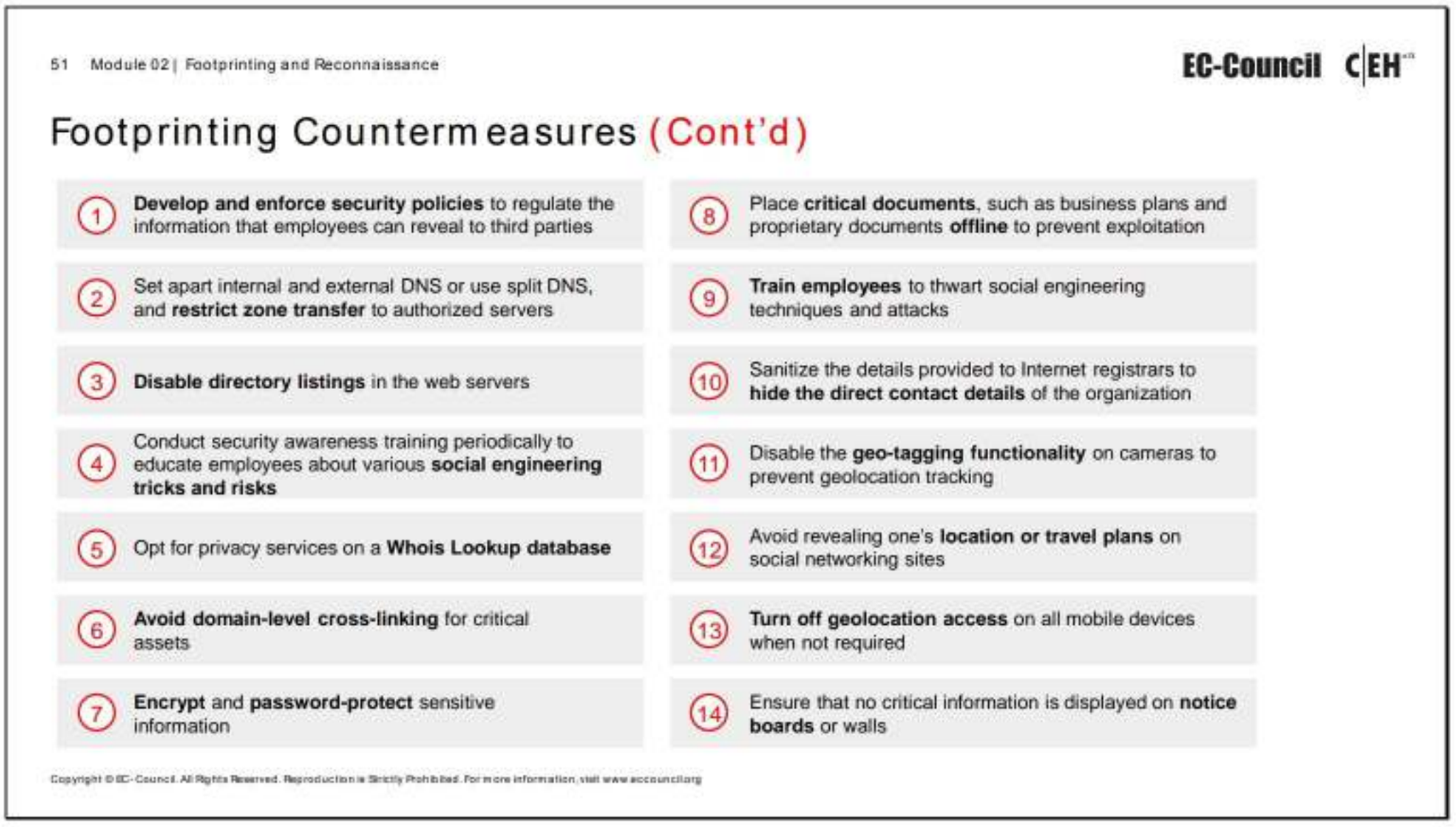
Figure 2.101: Output file for script with AI – Screenshot 4

Objective 10

Explain Footprinting Countermeasures

Footprinting Countermeasures

- **Restrict the employees' access to social networking sites from the organization's network**
- **Configure web servers to avoid information leakage**
- **Educate employees to use pseudonyms on blogs, groups, and forums**
- **Do not reveal critical information in press releases, annual reports, product catalogues, etc.**
- **Limit the amount of information published on a website or the Internet**
- **Use footprinting techniques to discover and remove any sensitive information that is publicly available**
- **Prevent search engines from caching a web page and use anonymous registration services**



Footprinting Countermeasures

Thus far, we have discussed the importance of footprinting, various methods to perform footprinting, and tools that help in its execution. Now, we discuss footprinting countermeasures, i.e., the measures or actions taken to prevent or offset information disclosure.

Some of the footprinting countermeasures are as follows:

- Restrict the employees' access to social networking sites from the organization's network.
- Configure web servers to avoid information leakage.
- Educate employees to use pseudonyms on blogs, groups, and forums.
- Do not reveal critical information in press releases, annual reports, product catalogs, etc.
- Limit the amount of information published on a website or the Internet.
- Use footprinting techniques to discover and remove any sensitive information that is publicly available.
- Prevent search engines from caching a web page and use anonymous registration services.
- Develop and enforce security policies such as information security and password policies to regulate the information that employees can reveal to third parties.
- Implement multi-factor authentication mechanisms to enhance the security of the organization's systems and resources.
- Set apart internal and external DNS or use split DNS, and restrict zone transfer to authorized servers.

- Disable directory listings in the web servers.
- Conduct security awareness training periodically to educate employees about various social engineering tricks and risks.
- Opt for privacy services on a Whois lookup database.
- Avoid domain-level cross-linking for critical assets.
- Encrypt and password-protect sensitive information.
- Implement captchas and rate limiting on public-facing services to prevent automated tools from collecting information at a rapid pace.
- Do not enable protocols that are not required.
- Always use TCP/IP and IPsec filters for defense in depth.
- Configure Internet Information Services (IIS) to avoid information disclosure through banner grabbing.
- Hide the IP address and related information by implementing a VPN or keeping the server behind a secure proxy.
- Request archive.org to delete the history of the website from the archive database.
- Keep the domain name profile private.
- Place critical documents such as business plans and proprietary documents offline to prevent exploitation.
- Train employees to thwart social engineering techniques and attacks.
- Sanitize the details provided to the Internet registrars to hide the direct contact details of the organization.
- Disable the geo-tagging functionality on cameras to prevent geolocation tracking.
- Avoid revealing one's location or travel plans on social networking sites.
- Turn off geolocation access on all mobile devices when not required.
- Ensure that no critical information, such as strategic plans, product information, or sales projections, is displayed on notice boards or walls.
- Disable or delete the accounts of employees who left the organization.
- Configure mail servers to ignore mails from anonymous individuals.
- Deploy honeypots or honeynets within the network to attract and detect attackers that can divert potential footprinters away from critical systems.

Module Summary

- In this module, we have discussed the following:
  - Footprinting concepts and the objectives of footprinting
  - Various footprinting techniques, such as footprinting through search engines, footprinting through Internet search services, and footprinting through social networking sites
  - Whois, DNS, and email footprinting
  - Network footprinting and footprinting through social engineering
  - Some important footprinting tools
  - How organizations can defend against footprinting and reconnaissance activities
- In the next module, we will discuss in detail how attackers, ethical hackers, and pen testers perform network scanning to collect information about a target of evaluation before an attack or audit

Module Summary

This module presented footprinting concepts along with the objectives of footprinting. It provided a detailed explanation of the various techniques used for footprinting through search engines. Further, it described footprinting through Internet research services and social networking sites. It also explained Whois and DNS footprinting in detail. Moreover, it described network footprinting along with traceroute analysis. In addition, it discussed email footprinting techniques. It also explained footprinting through social engineering. Finally, it presented an overview of important footprinting tools. The module ended with a detailed discussion of how organizations can defend themselves against footprinting and reconnaissance activities.

In the next module, we will discuss in detail how attackers as well as ethical hackers and pen testers perform network scanning to collect information about a target for evaluation before an attack or audit.

This page is intentionally left blank.