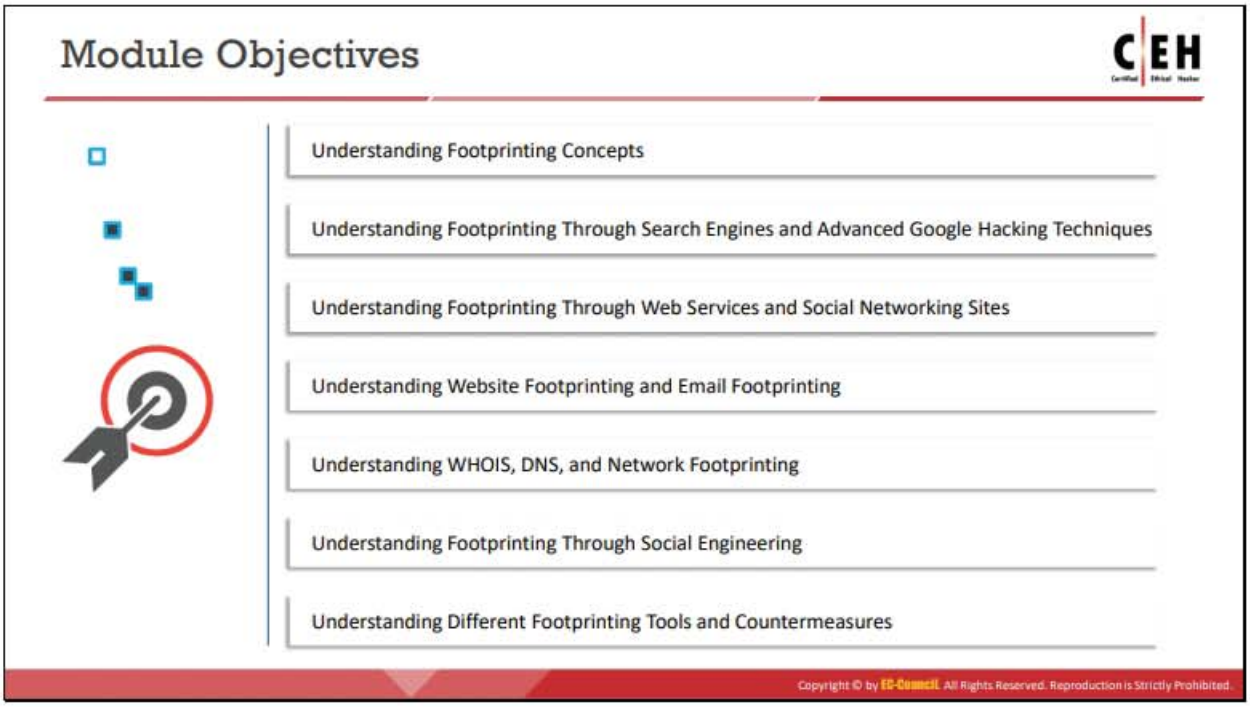
Module 02:
Footprinting and Reconnaissance



Module 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.

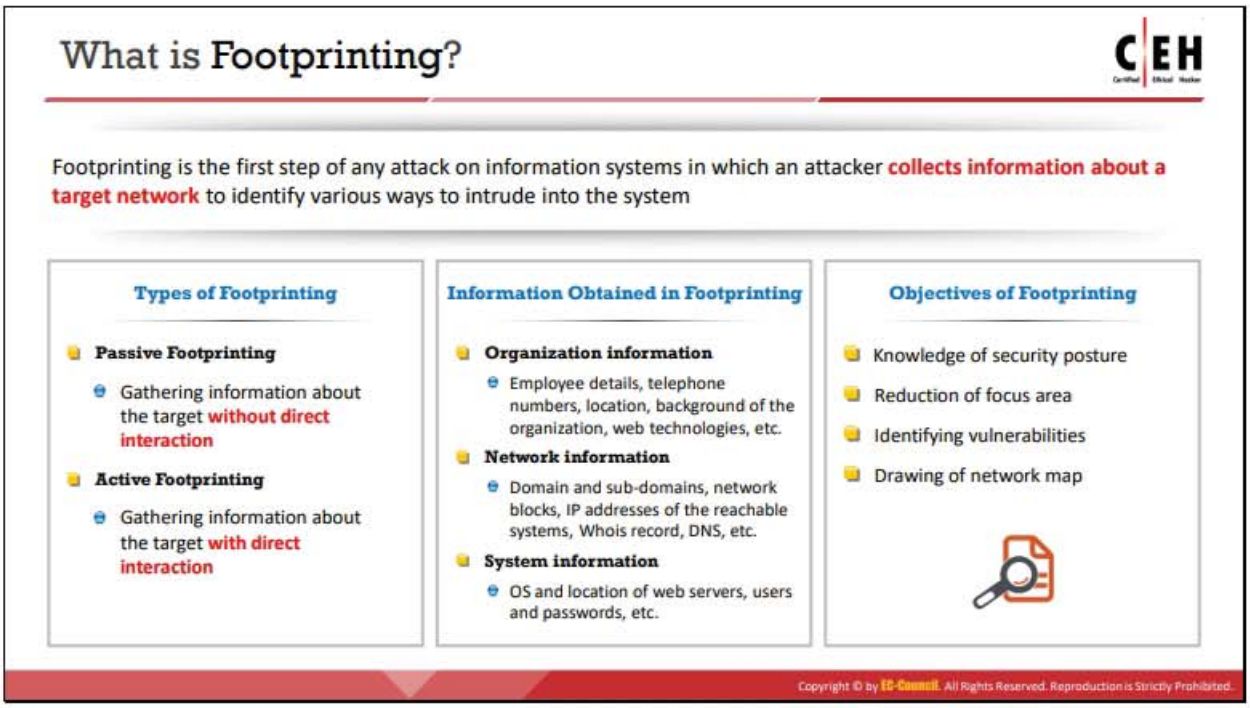
At the end of this module, you will be able to:

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

Footprinting Concepts

Ethical hacking is legal in nature and conducted to evaluate the security of a target organization's IT infrastructure with their consent. Footprinting, where an attacker tries to gather information about a target, is the first step in ethical hacking. 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.



What is Footprinting?

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

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.

Passive footprinting techniques include:

- Finding information through search engines
- Finding the Top-level Domains (TLDs) and sub-domains of a target through web services
- Collecting location information on the target through web services
- Performing people search using social networking sites and people search services
- Gathering financial information about the target through financial services
- Gathering infrastructure details of the target organization through job sites
- Collecting information through deep and dark web footprinting
- Determining the operating systems in use by the target organization
- Performing competitive intelligence
- Monitoring the target using alert services
- Gathering information using groups, forums, blogs, and NNTP Usenet newsgroups
- Collecting information through social engineering on social networking sites
- Extracting information about the target using Internet archives
- Gathering information using business profile sites
- Monitoring website traffic of the target
- Tracking the online reputation of the target

- **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.

Active footprinting techniques include:

- Querying published name servers of the target
- Searching for digital files
- Extracting website links and gathering wordlists from the target website
- Extracting metadata of published documents and files
- Gathering website information using web spidering and mirroring tools
- Gathering information through email tracking

- Harvesting email lists
- Performing Whois lookup
- Extracting DNS information
- Performing traceroute analysis
- Performing social engineering

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**: Such 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 need to gather information about the target organization's network. They then use such information to locate 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; this plays a vital role in the hacking process.

Footprinting helps to

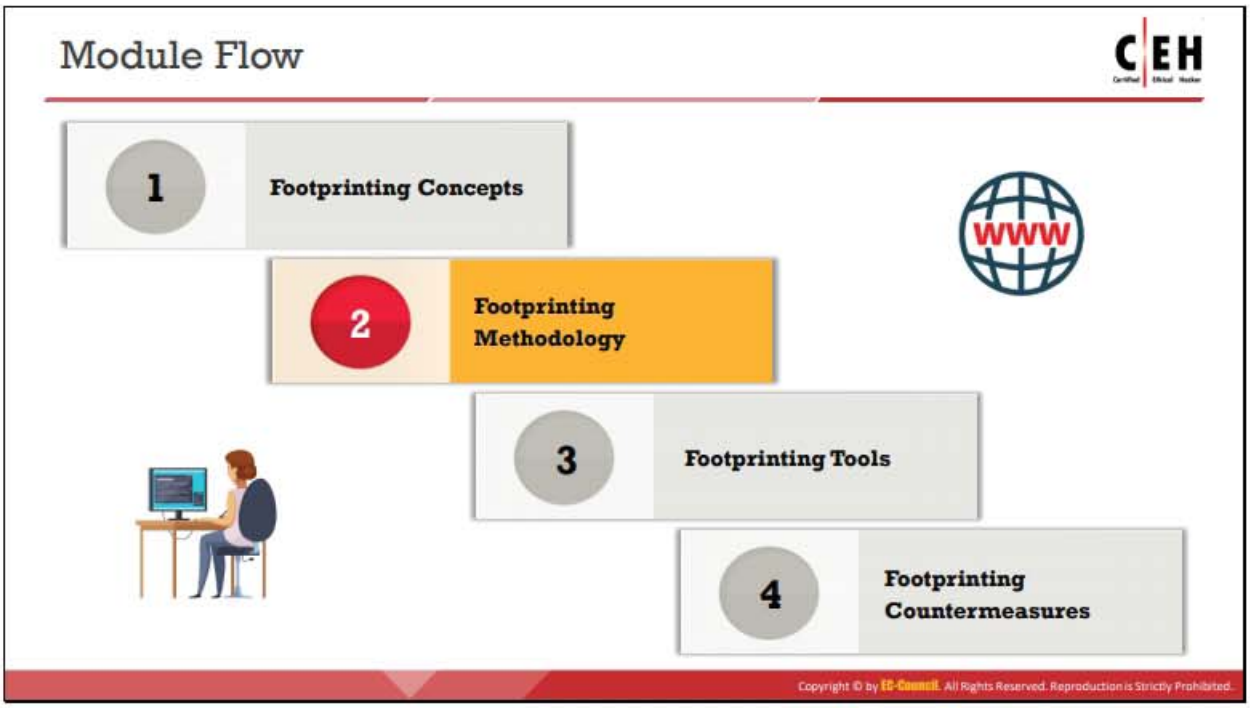
- **Know Security Posture**: Performing footprinting on the target organization gives the complete profile of the organization's security posture. Hackers can then analyze the report to identify loopholes in the security posture of the target organization and build a hacking plan accordingly.
- **Reduce Focus Area**: 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, as well as many other details pertaining to its security posture.
- **Identify Vulnerabilities**: A detailed footprint provides maximum information about the target organization. It allows the attacker to identify vulnerabilities in the target systems to select appropriate exploits. Attackers can build their own information database about the security weaknesses of the target organization. Such a database can then help in identifying the weakest link in the organization's security perimeter.
- **Draw Network Map**: Combining footprinting techniques with tools such as Tracert allows the attacker to create diagrammatic representations of the target organization's network presence. Specficially, it allows attackers to draw a map or outline of the target organization's network infrastructure to know about the actual environment that they are going to break into. A network map will depict the attacker's understanding of the target's Internet footprint. These network diagrams can guide the attacker in performing an attack.

Footprinting Threats

Attackers perform footprinting as the first step of any attack on information systems. In this phase, attackers attempt to collect valuable system-level information such as account details, operating system and other software versions, server names, database schema details, and so on, which will be useful in the hacking process.

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 secure 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. This section discusses the common techniques used to collect information about the target organization from different sources.

Footprinting techniques:

- Footprinting through search engines
- Footprinting through web services
- Footprinting through social networking sites
- Website footprinting
- Email footprinting
- Whois footprinting
- DNS footprinting
- Network footprinting
- Footprinting through social engineering

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.

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.

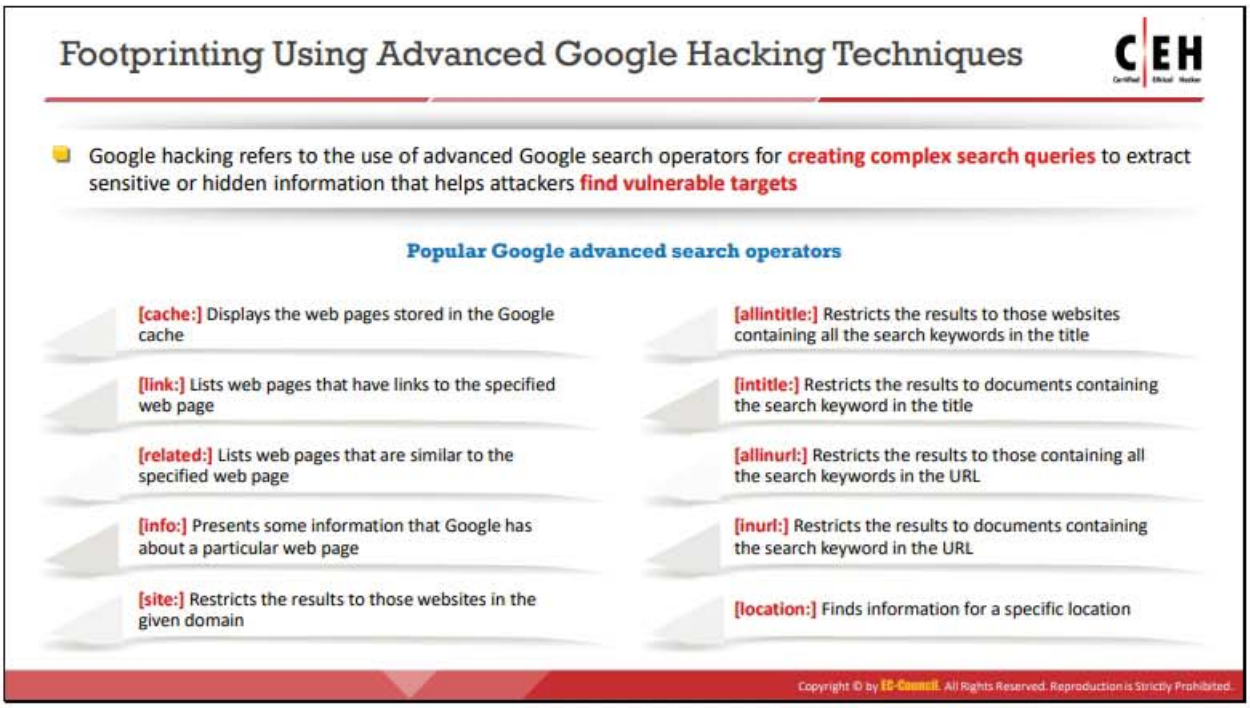
For example, consider an organization, perhaps Microsoft. Type **Microsoft** in the **Search** box of a search engine and press **Enter**; this will display the results containing information about Microsoft. Browsing the results often provides critical information such as physical location, contact 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, 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.



Footprinting Using Advanced Google Hacking Techniques

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: *http://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."

- **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.

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 uses Google operators that help locate specific strings of text within the search results. Thus, the attacker can not only detect websites and web servers that are vulnerable to exploitation but also locate private, sensitive information about others, such as credit card numbers, social security numbers, passwords, and so on. Once a vulnerable site is identified, attackers try 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.

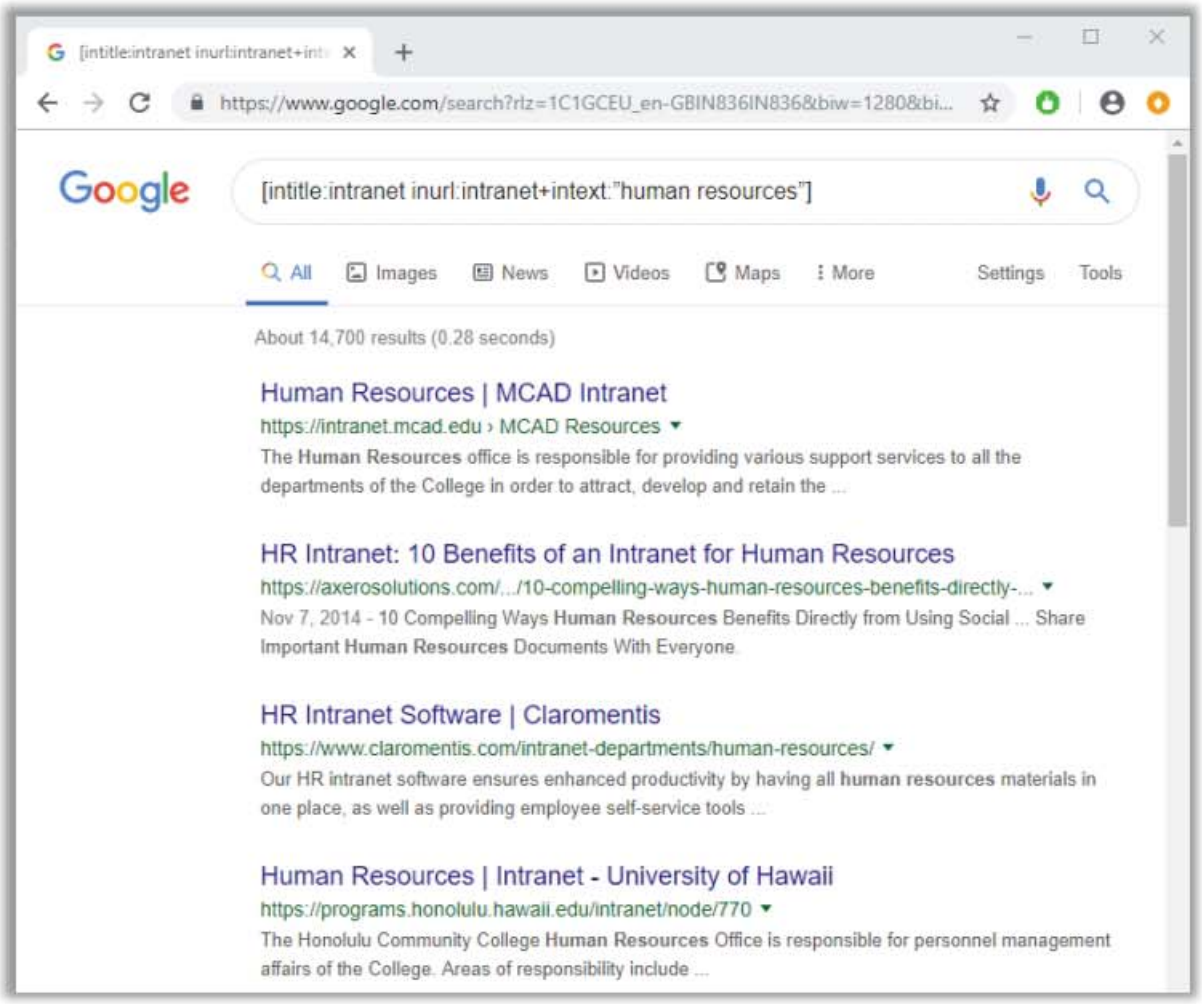
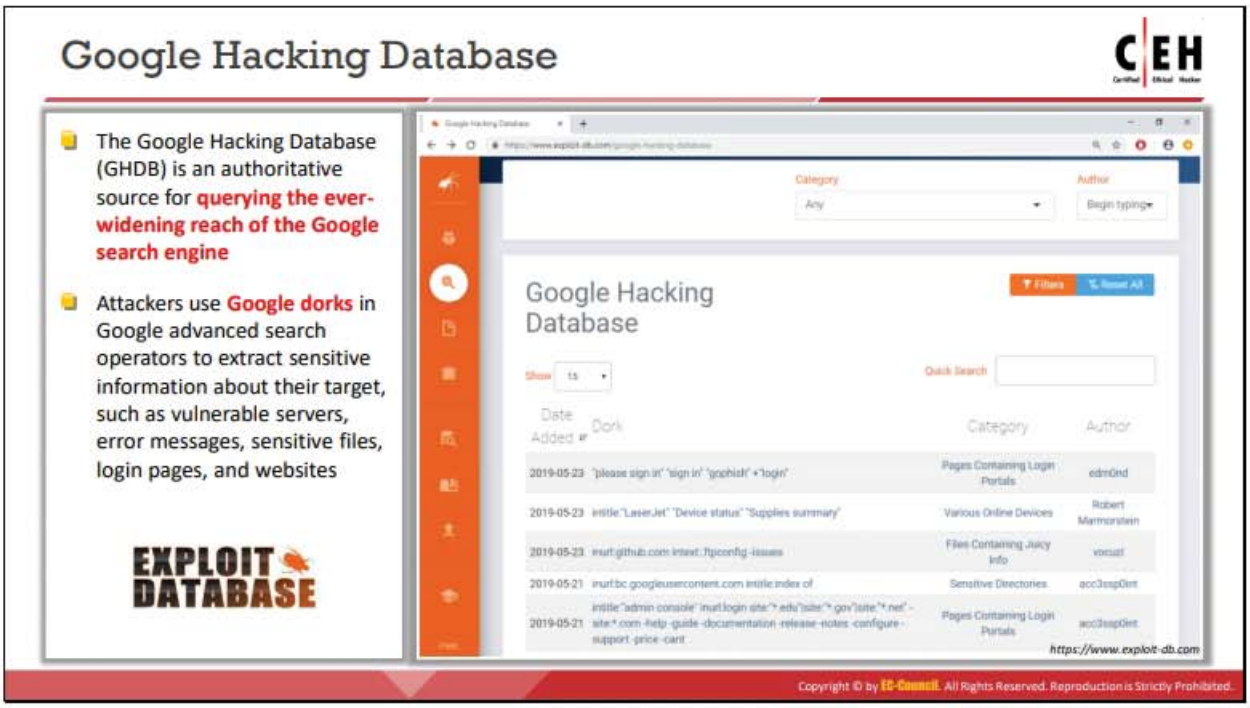


Figure 2.1: Search engine results for given Google Advance Operator syntax



Google Hacking Database

Source: *https://www.exploit-db.com*

The Google Hacking Database (GHDB) is an authoritative source for querying the ever-widening scope of the Google search engine. In the GHDB, you will find search terms for files containing usernames, vulnerable servers, and even files containing passwords. The Exploit Database is a Common Vulnerabilities and Exposures (CVE) compliant archive of public exploits and corresponding vulnerable software, developed for use by penetration testers and vulnerability researchers.

Using GHDB dorks, attackers can rapidly identify all the publicly available exploits and vulnerabilities of the target organization's IT infrastructure. Attackers use Google dorks in Google advanced search operators to extract sensitive information about the target, such as vulnerable servers, error messages, sensitive files, login pages, and websites.

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

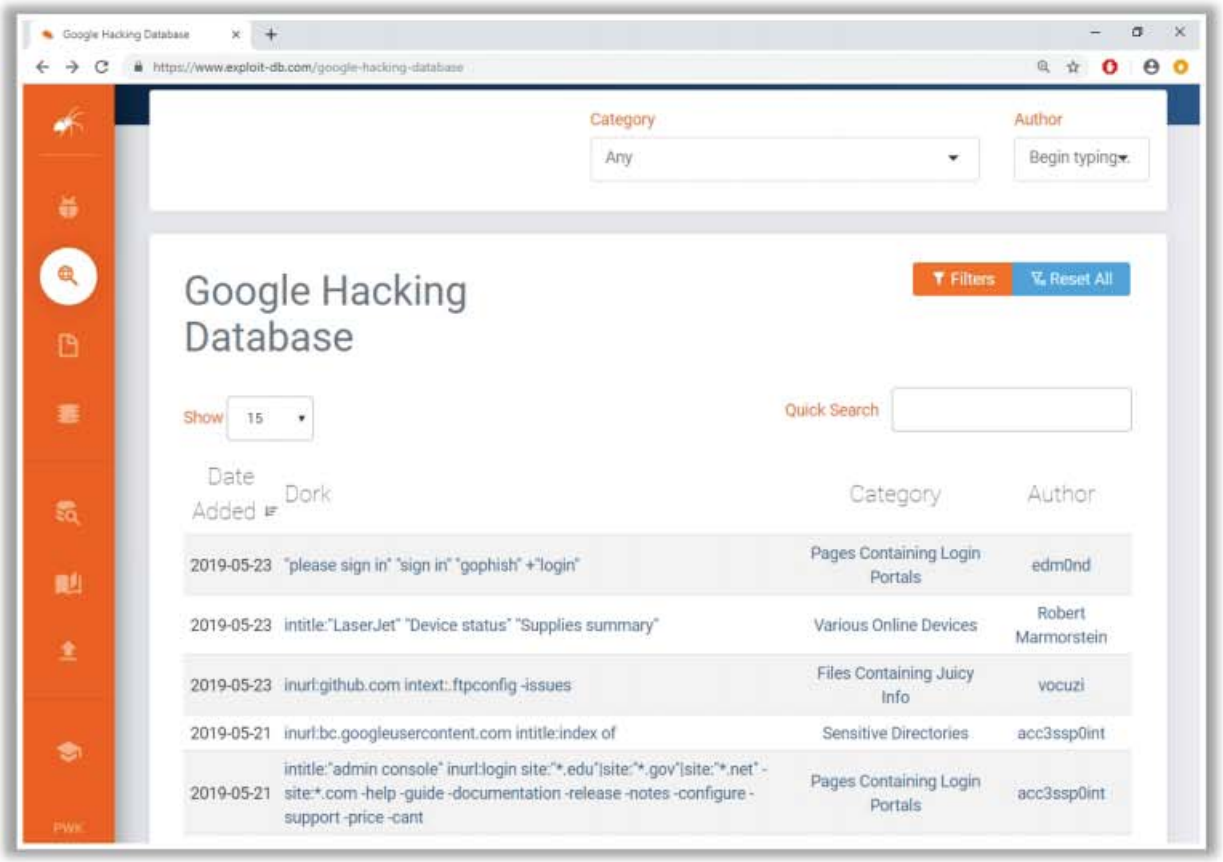
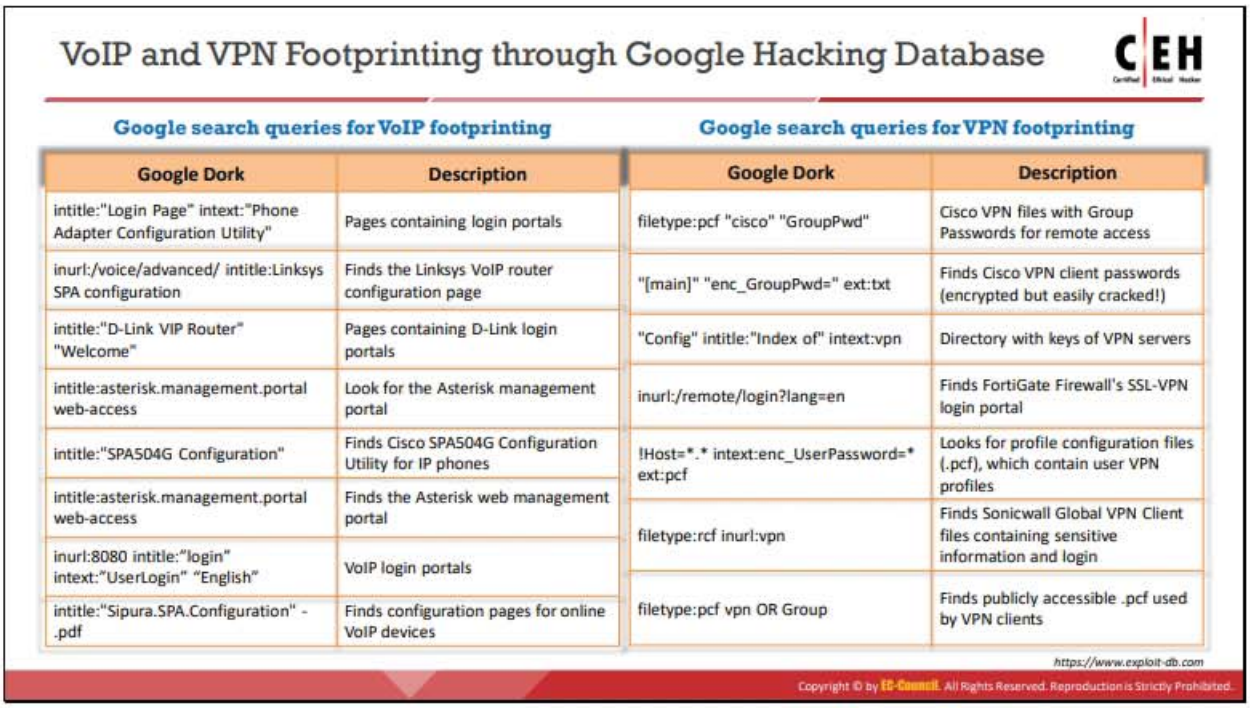


Figure 2.2: Google Hacking Database screenshot



VoIP and VPN Footprinting through Google Hacking Database

Google hacking involves the implementation of advanced operators in the Google search engine to match specific strings of text within the search result. These advanced operators help refine searches to expose sensitive information, vulnerabilities, and passwords. You can use these Google hacking operators or Google dorks for footprinting VoIP and VPN networks. Thus, you can extract information such as pages containing login portals, VoIP login portals, directories with keys of VPN servers, and so on.

The following tables summarize some of the Google hacking operators or Google dorks to obtain specific information related to VoIP and VPN footprinting, respectively.

Google search queries for VoIP footprinting

Google Dork	Description
intitle:"Login Page" intext:"Phone Adapter Configuration Utility"	Pages containing login portals
inurl:/voice/advanced/ intitle:Linksys SPA configuration	Finds the Linksys VoIP router configuration page
intitle:"D-Link VoIP Router" "Welcome"	Pages containing D-Link login portals
intitle:asterisk.management.portal web-access	Looks for the Asterisk management portal
intitle:"SPA504G Configuration"	Finds Cisco SPA504G Configuration Utility for IP phones

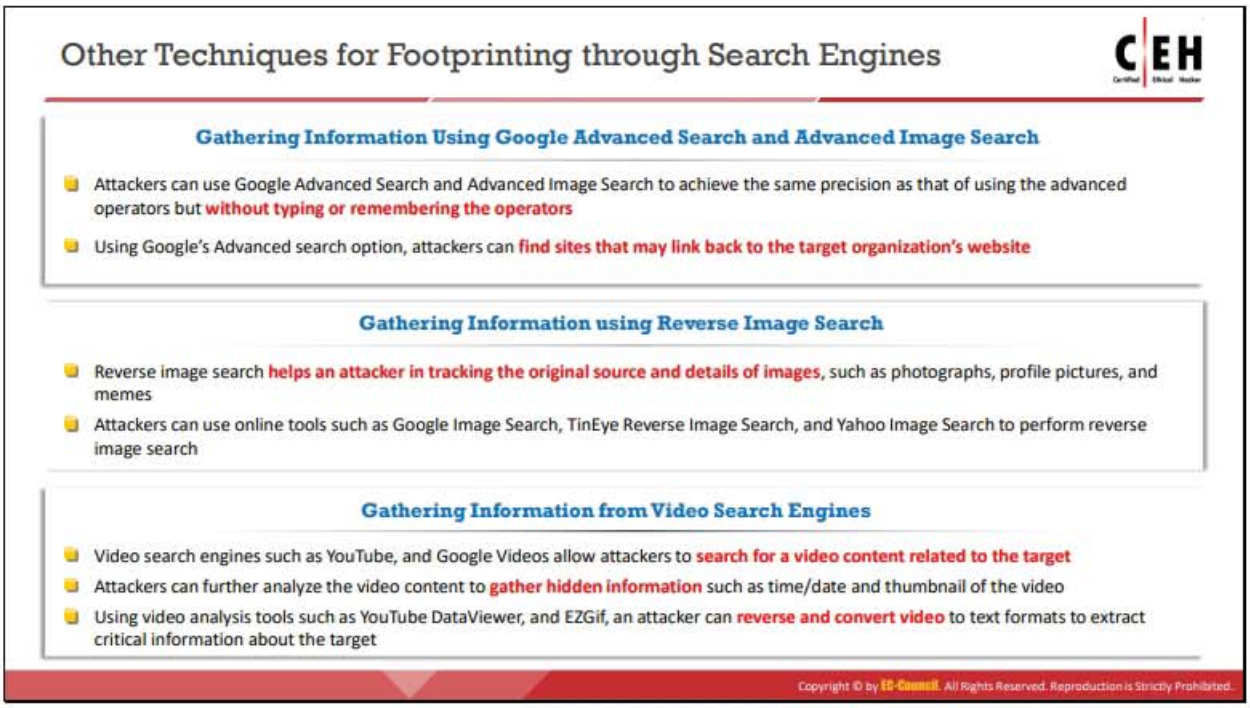
intitle:"Sipura.SPA.Configuration" -.pdf	Finds configuration pages for online VoIP devices
intitle:asterisk.management.portal web-access	Finds the Asterisk web management portal
inurl:8080 intitle:"login" intext:"UserLogin" "English"	VoIP login portals

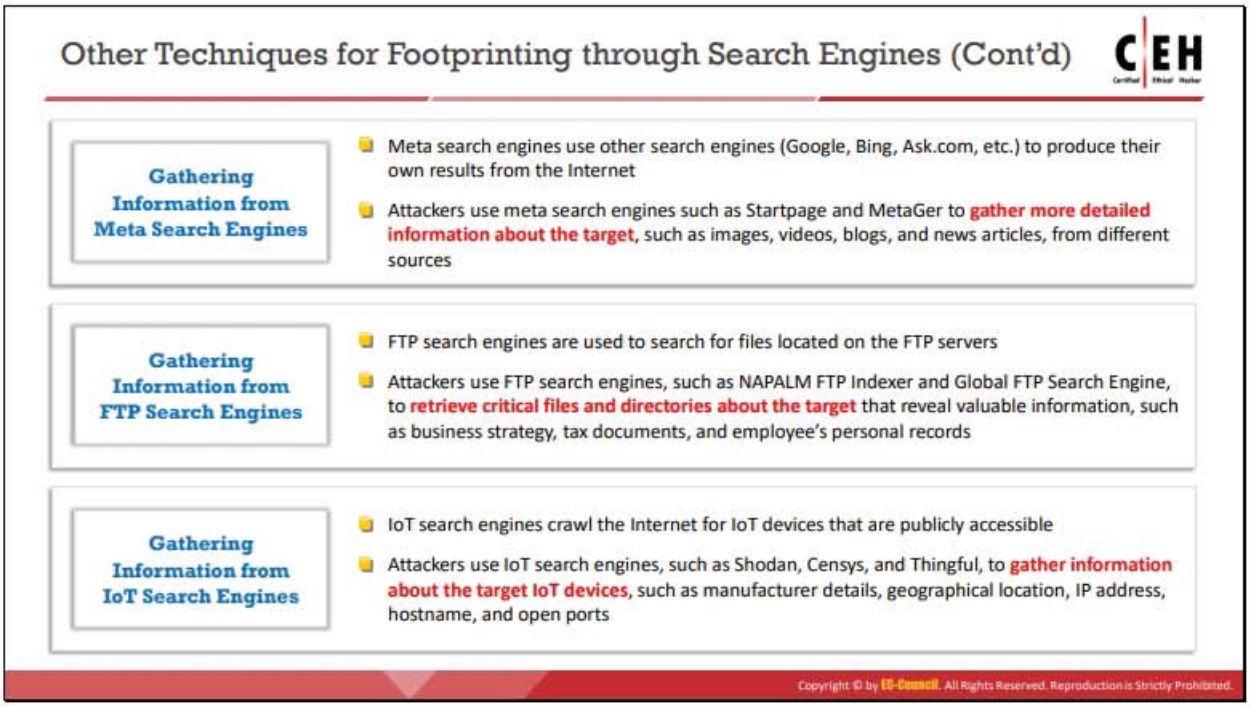
Table 2.1: Google search queries for VoIP footprinting

Google search queries for VPN footprinting

Google Dork	Description
filetype:pcf "cisco" "GroupPwd"	Cisco VPN files with Group Passwords for remote access
"[main]" "enc_GroupPwd=" ext:txt	Finds Cisco VPN client passwords (encrypted but easily cracked)
"Config" intitle:"Index of" intext:vpn	Directory with keys of VPN servers
inurl:/remote/login?lang=en	Finds FortiGate Firewall's SSL-VPN login portal
!Host=*.* intext:enc_UserPassword=* ext:pcf	Looks for profile configuration files (.pcf), which contain user VPN profiles
filetype:rcf inurl:vpn	Finds Sonicwall Global VPN Client files containing sensitive information and login
filetype:pcf vpn OR Group	Finds publicly accessible .pcf used by VPN clients
vpnssl	Retrieves login portals containing vpnssl companies' access
intitle:"SSL VPN Service" + intext:"Your system administrator provided the following information to help understand and remedy the security conditions:"	Finds Cisco asa login web pages

Table 2.2: Google search queries for VPN footprinting





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).

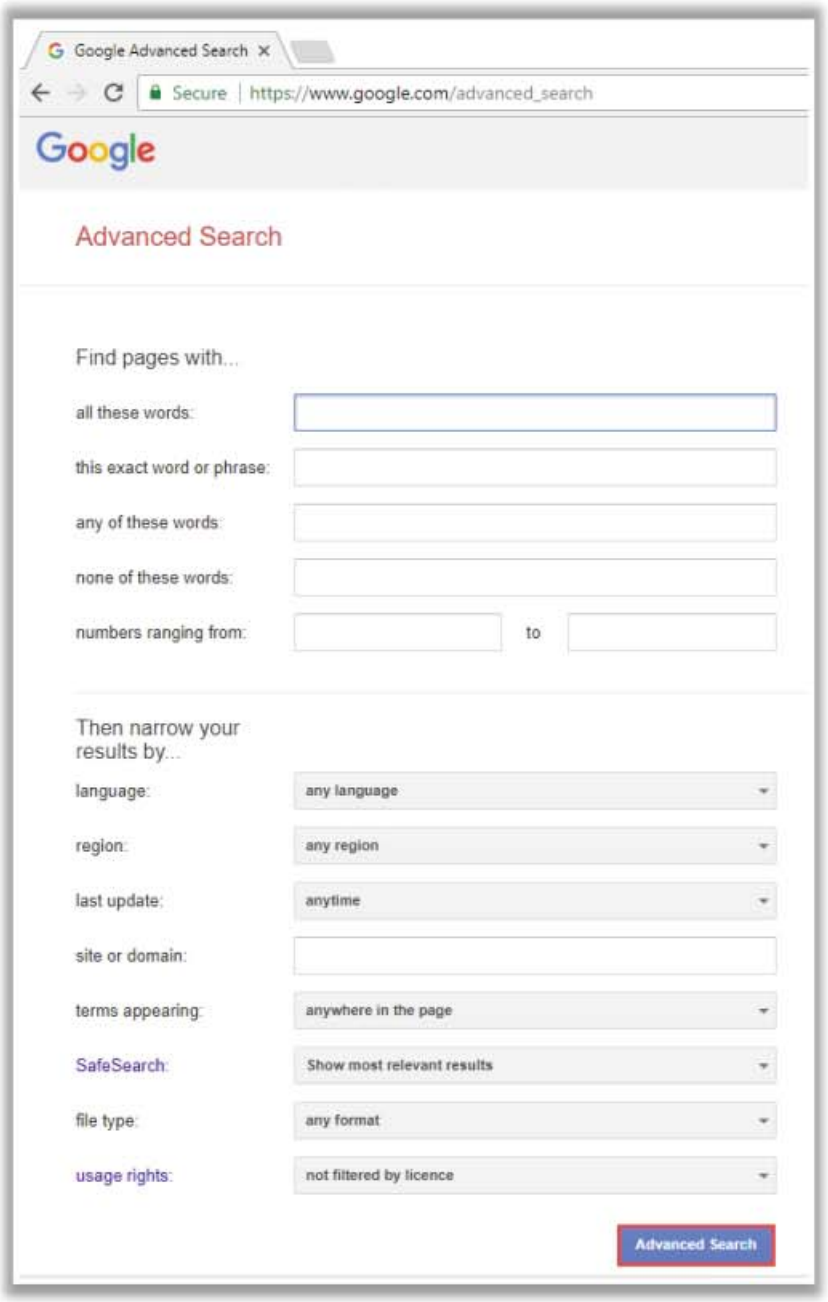


Figure 2.3: Google Advance Search

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.

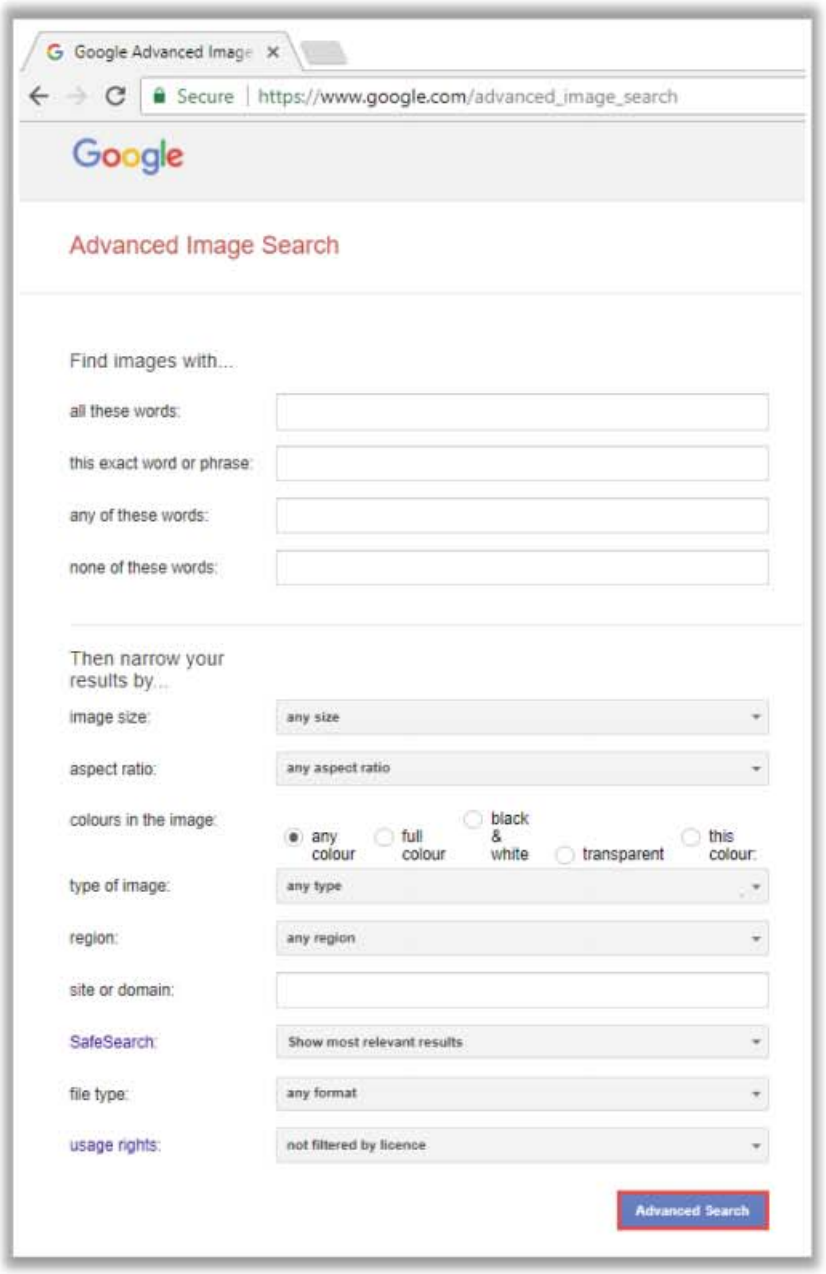


Figure 2.4: Google Advance Image Search

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, and Bing Image Search to perform a reverse image search.

Figure 2.5: 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.

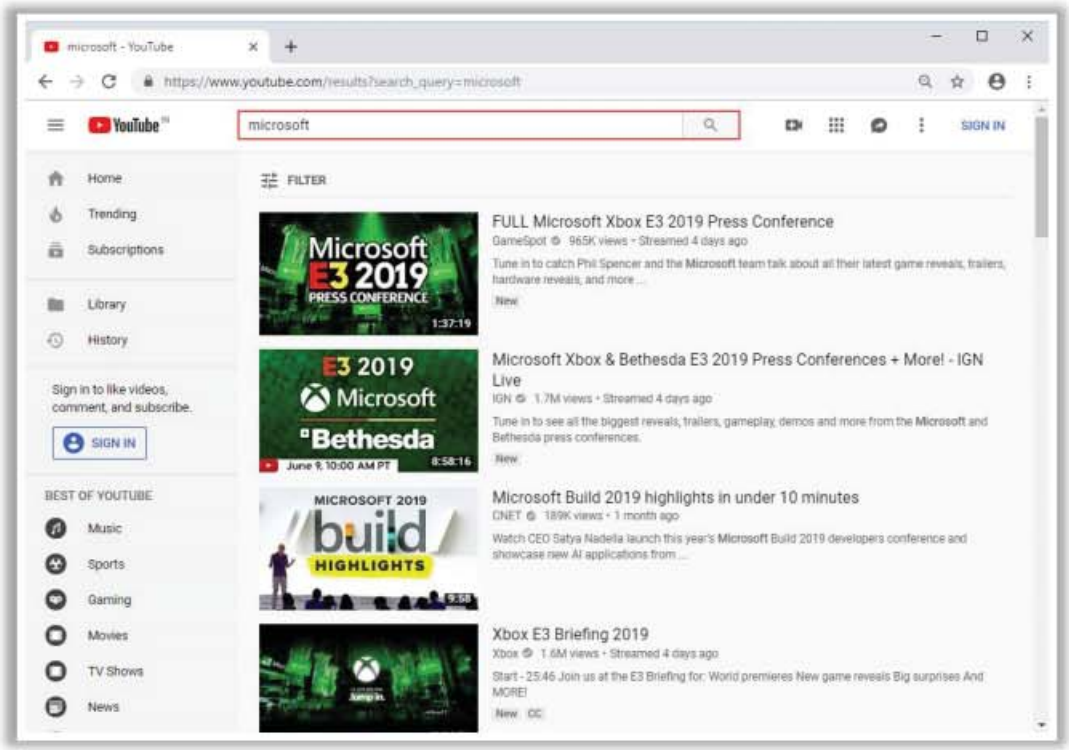


Figure 2.6: Screenshot of YouTube showing search results for Microsoft

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 DataViewer, 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 2.7: Screenshot of YouTube DataViewer 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, 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.

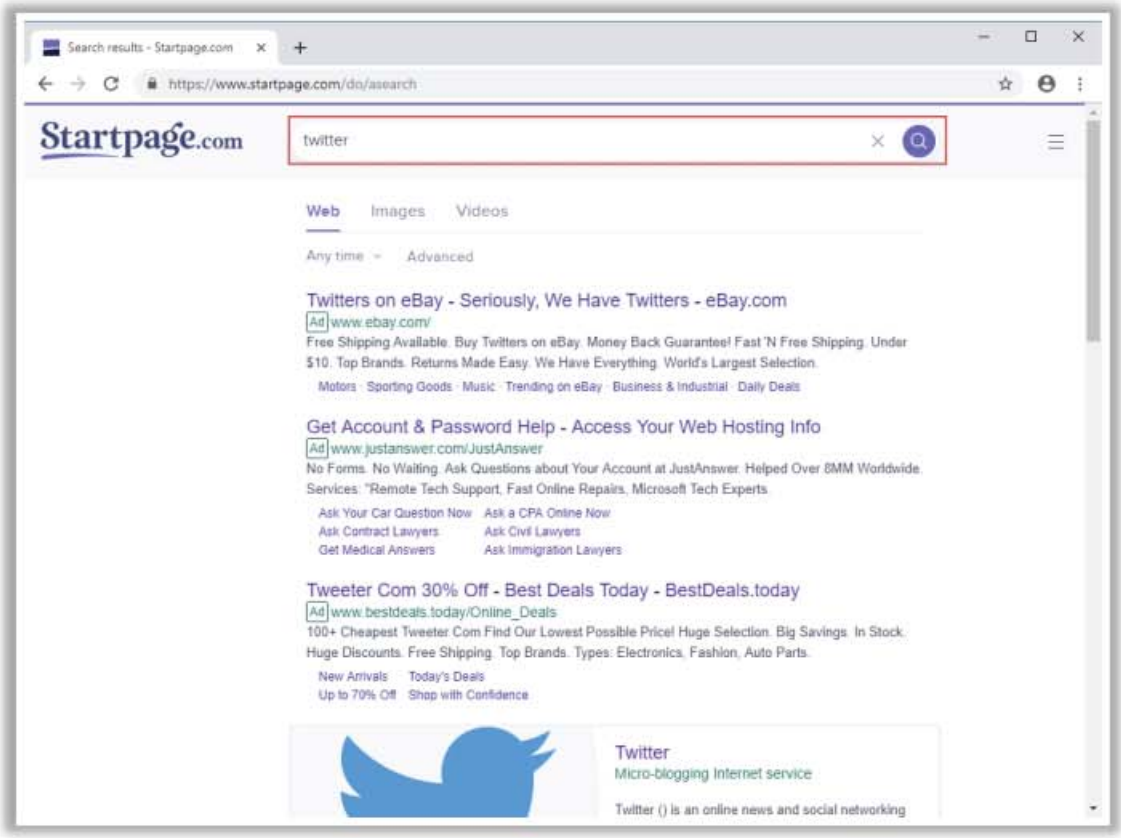


Figure 2.8: Screenshot of Meta Search Engine StartPage.com showing search results for Twitter

- **Gathering Information from FTP Search Engines**

FTP search engines are used to search for files located on FTP servers that contain valuable information about the target organization. Many industries, institutions, companies, and universities use FTP servers to store large file archives and other software that are shared among their employees. A special client such as FileZilla (*https://filezilla-project.org*) can be used to access the 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 through web browsers directly.

Using FTP search engines such as NAPALM FTP Indexer, Global FTP Search Engine, and FreewareWeb FTP File Search, attackers can search for critical files and directories

containing valuable information such as business strategies, tax documents, employee's personal records, financial records, licensed software, and other confidential information.

Listed below are some of the important advanced Google search queries to find FTP servers:

Google Dork	Description
inurl:github.com intext:.ftpconfig -issues	Returns SFTP/FTP server credentials on Github
type:mil inurl:ftp ext:pdf | ps	Returns sensitive directories on FTP
intext:pure-ftpd.conf intitle:index of	Returns servers exposing pure-ftpd configuration files
intitle:"Index Of" intext:sftp-config.json	Extracts list of FTP/SFTP passwords from sublime text
inurl:"ftp://www." "Index of /"	Displays various online FTP servers
inurl:~/ftp://193 filetype:(php | txt | html | asp | xml | cnf | sh) ~'/html'	Returns a list of FTP servers by IP address, mostly Windows NT servers with guest login capabilities

Table 2.3: Google search queries to find FTP servers

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

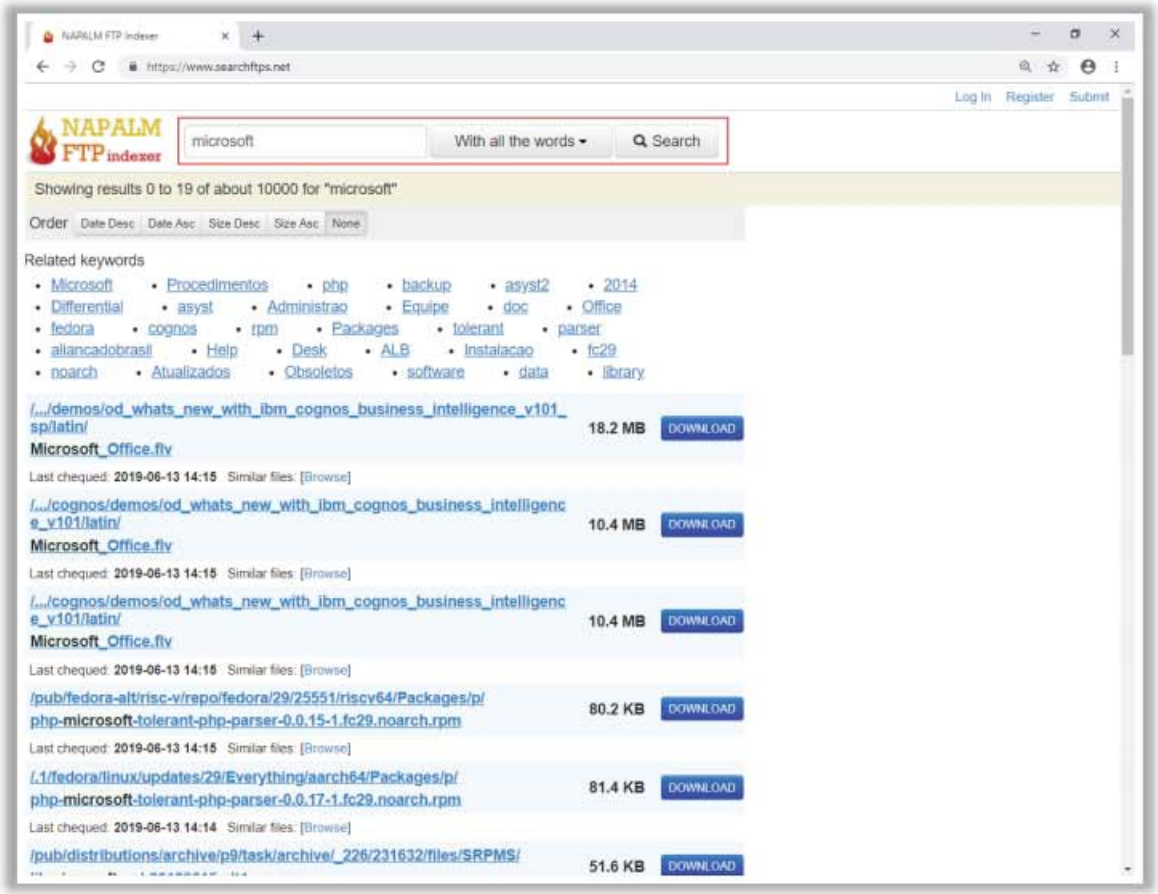


Figure 2.9: Screenshot of FTP Search Engine NAPALM FTP Indexer showing search results for Microsoft

- **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 Thingful, 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.

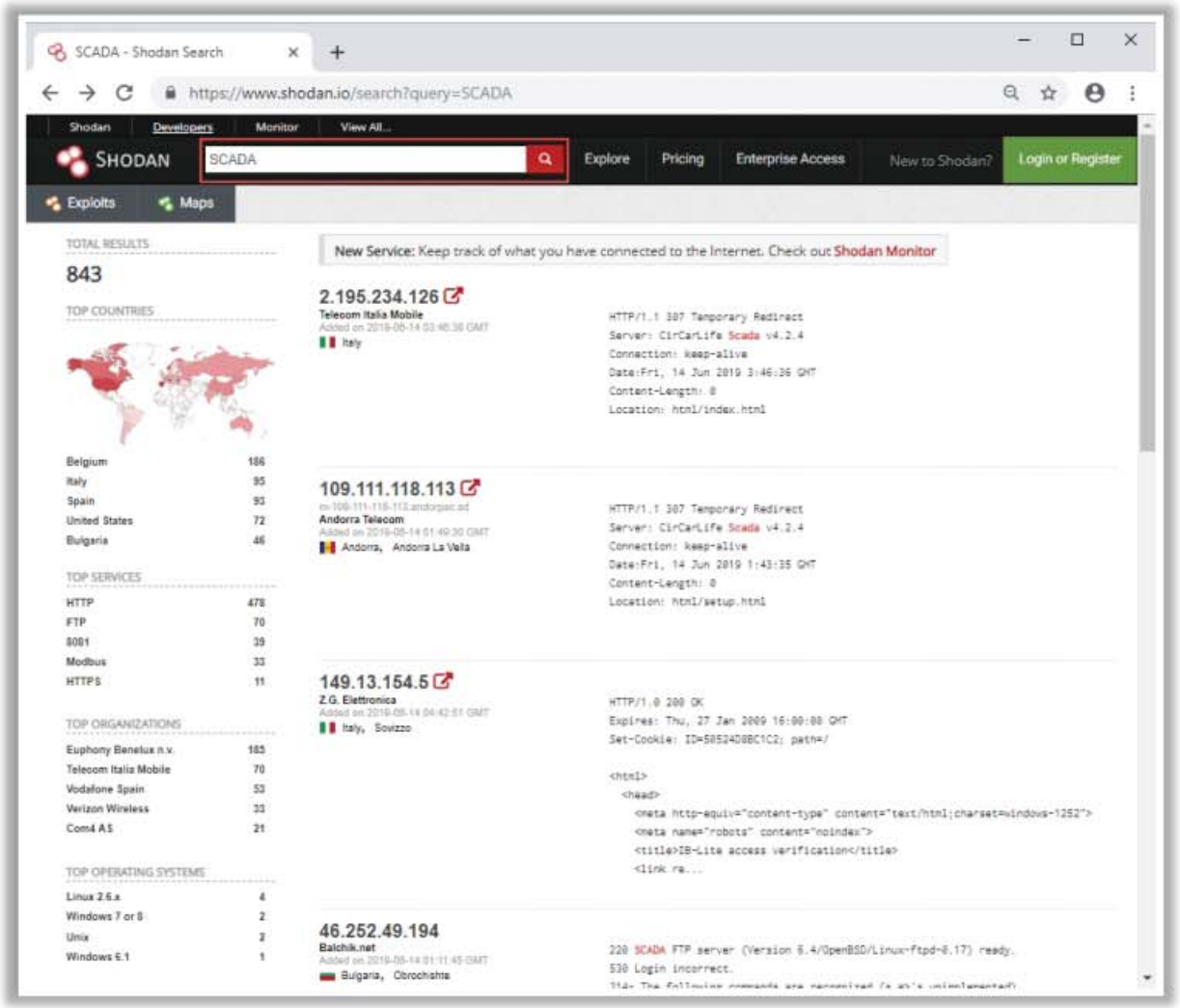


Figure 2.10: Screenshot of Shodan showing search results for SCADA devices

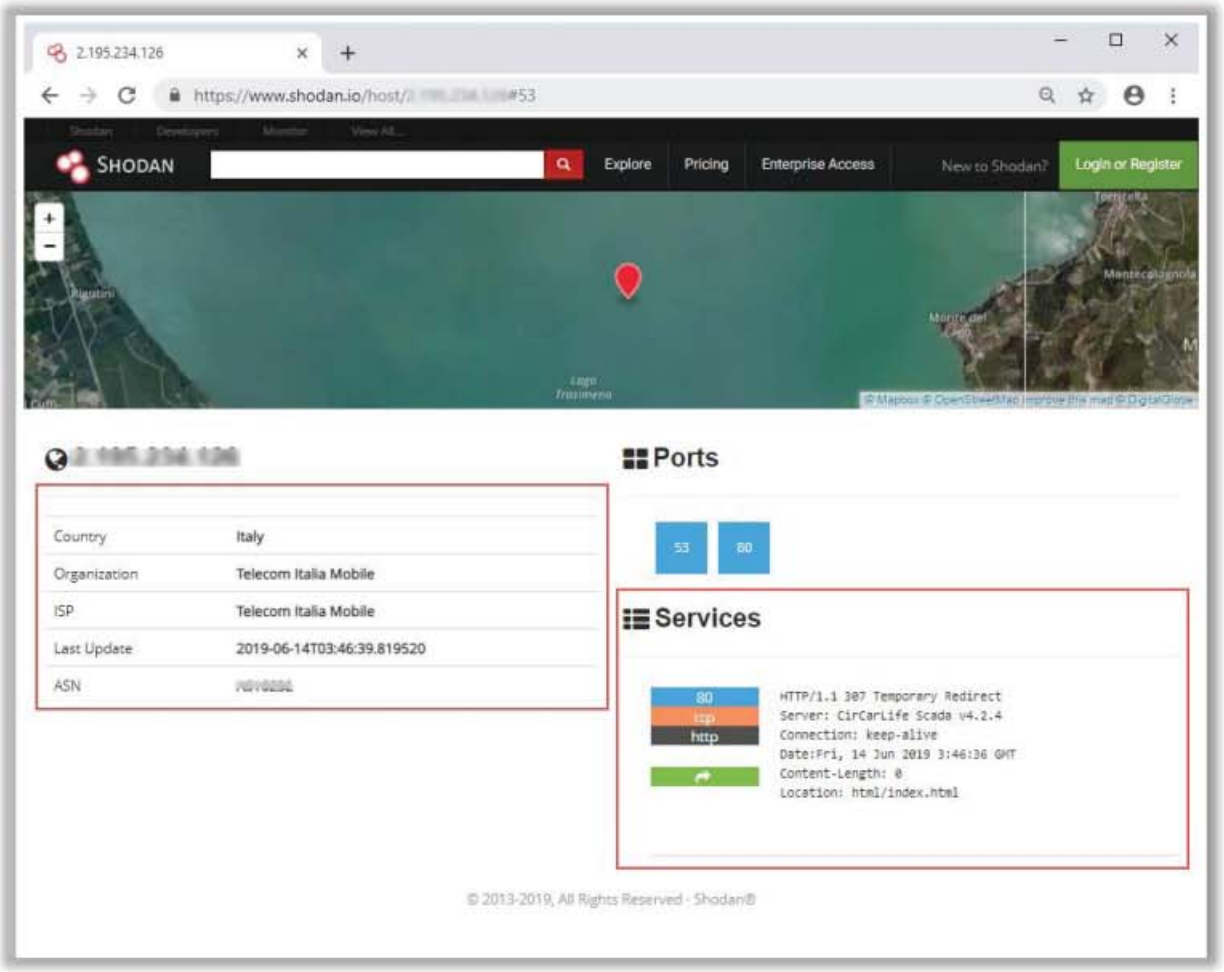
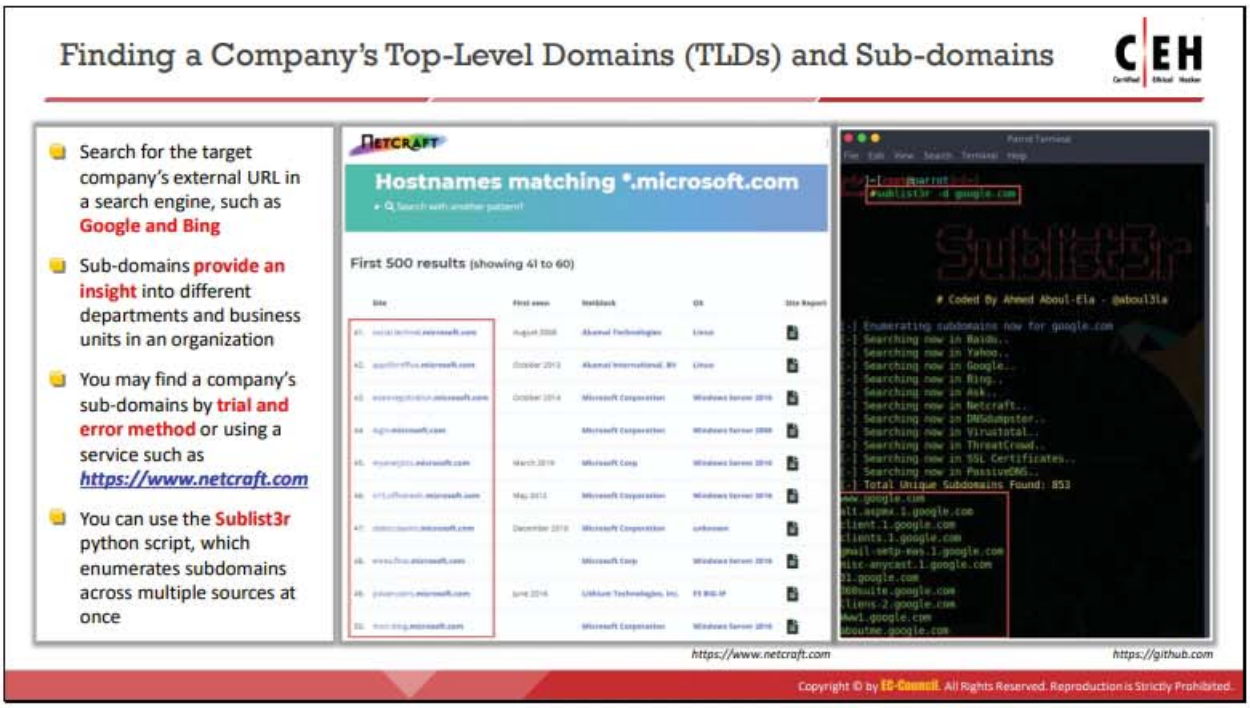


Figure 2.11: Screenshot of Shodan showing open ports and services of a SCADA system



Footprinting through Web Services

Web services such as people search services can provide sensitive information about the target. Internet archives may also provide sensitive information that has been removed from the World Wide Web (WWW). Social networking sites, people search services, alerting services, financial services, and job sites provide information about a target such as infrastructure details, physical location, and employee details. Moreover, groups, forums, and blogs can help attackers in gathering sensitive information about a target, such as public network information, system information, and personal information. 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 social networking sites and people search services, gathering information from job sites, financial services, third-party data repositories, performing deep and dark web footprinting, determining the operating system, VOIP and VPN footprinting through Shodan, 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.

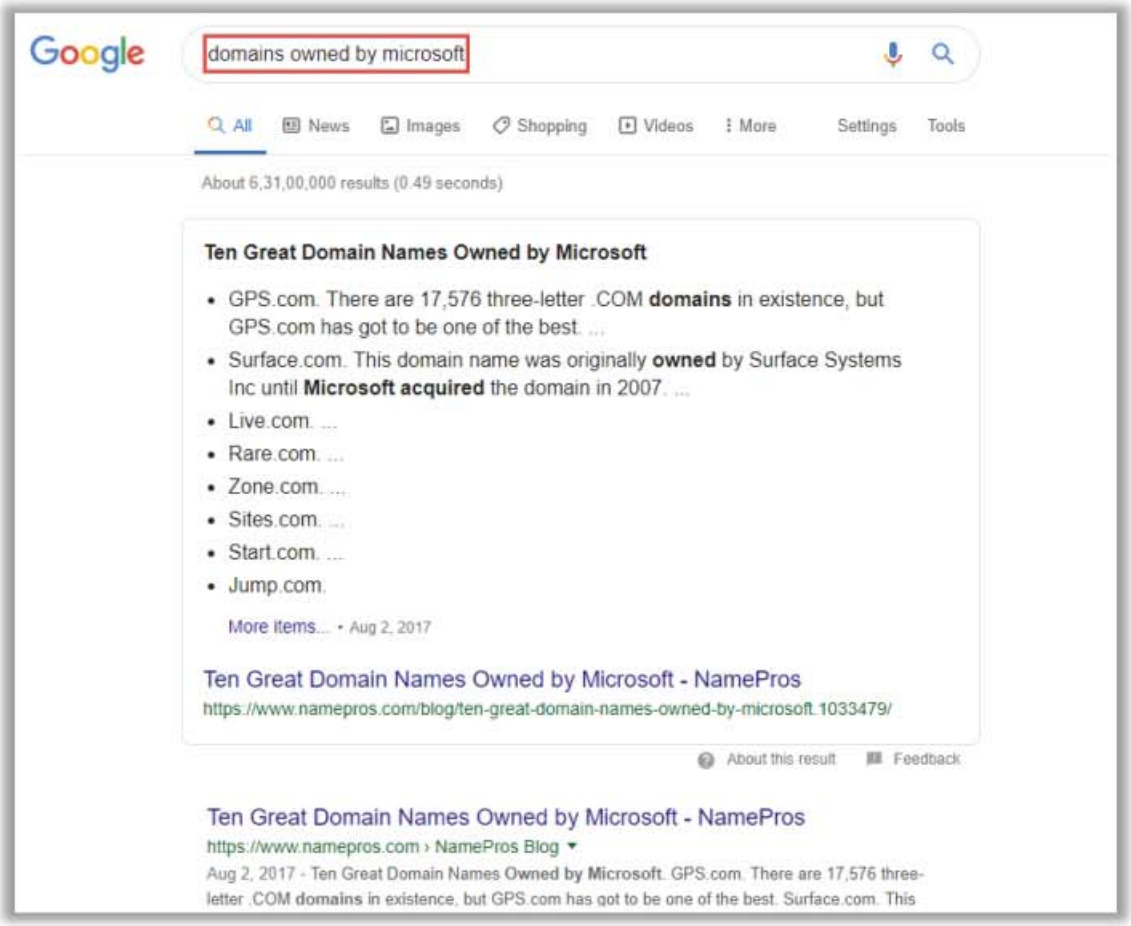


Figure 2.12: Google search engine showing results for given syntax

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
```

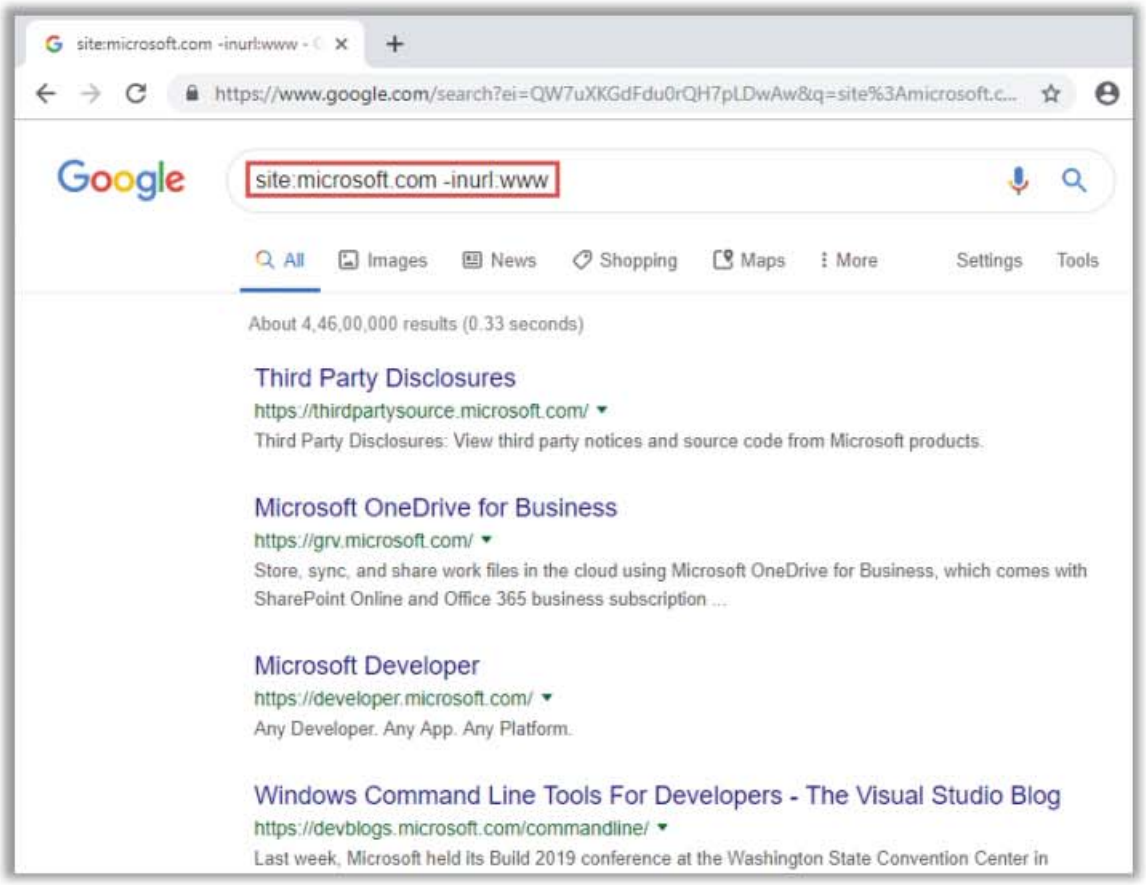


Figure 2.13: Finding sub-domains using Google Advanced Search Operator

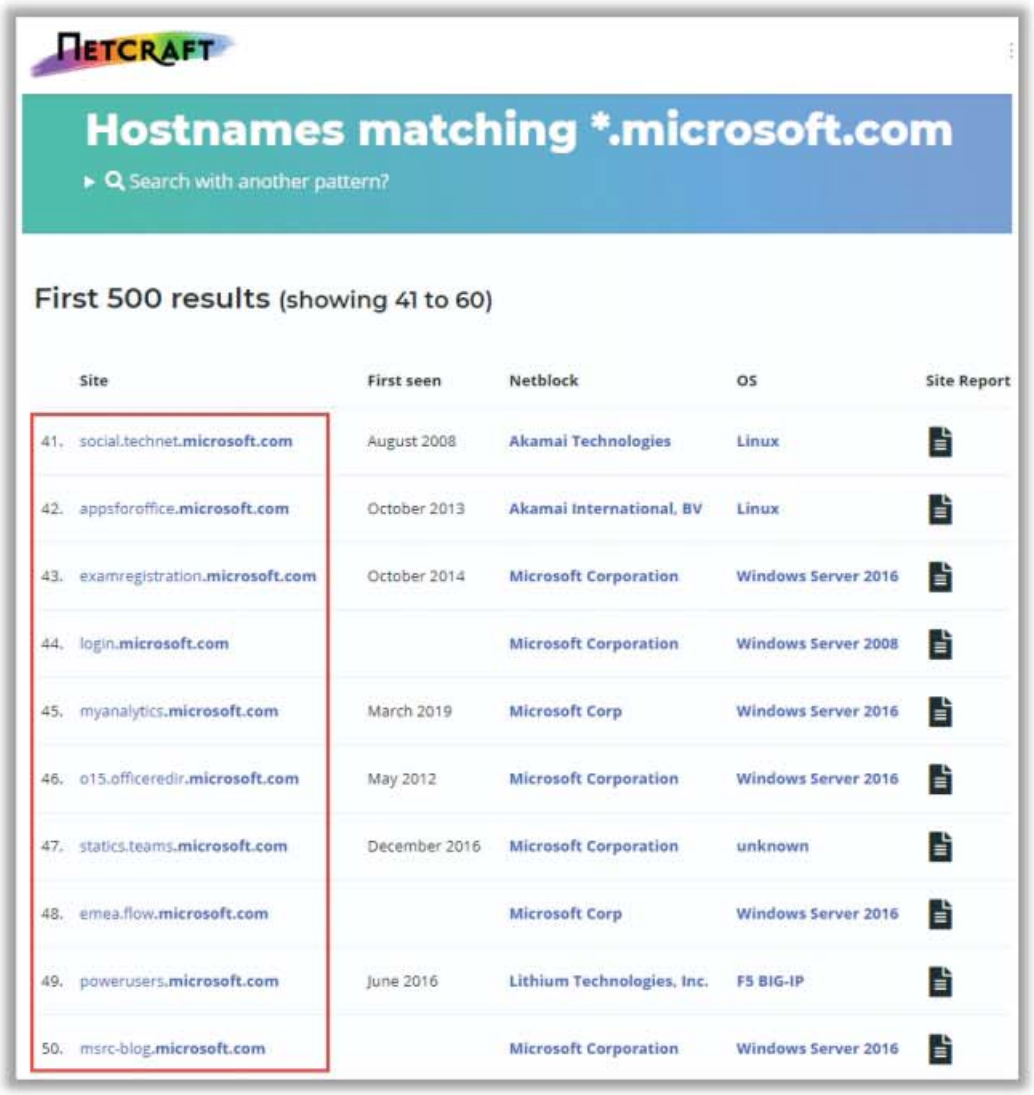
Tools to Search Company's Sub-domains

- **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.



NETCRAFT

Hostnames matching *.microsoft.com

▸ Search with another pattern?

First 500 results (showing 41 to 60)

	Site	First seen	Netblock	OS	Site Report
41.	social.technet.microsoft.com	August 2008	Akamai Technologies	Linux	
42.	appsforoffice.microsoft.com	October 2013	Akamai International, BV	Linux	
43.	examregistration.microsoft.com	October 2014	Microsoft Corporation	Windows Server 2016	
44.	login.microsoft.com		Microsoft Corporation	Windows Server 2008	
45.	myanalytics.microsoft.com	March 2019	Microsoft Corp	Windows Server 2016	
46.	o15.officeredir.microsoft.com	May 2012	Microsoft Corporation	Windows Server 2016	
47.	statics.teams.microsoft.com	December 2016	Microsoft Corporation	unknown	
48.	emea.flow.microsoft.com		Microsoft Corp	Windows Server 2016	
49.	powerusers.microsoft.com	June 2016	Lithium Technologies, Inc.	F5 BIG-IP	
50.	msrc-blog.microsoft.com		Microsoft Corporation	Windows Server 2016	

Figure 2.14: Screenshot of Netcraft displaying sub-domains of microsoft.com

- **Sublist3r**

  Source: *https://github.com*

  Sublist3r is a Python script designed to enumerate the subdomains of websites using OSINT. It enables you to enumerate subdomains across multiple sources at once. Further, it helps penetration testers and bug hunters in collecting and gathering subdomains for the domain they are targeting. It enumerates subdomains using many search engines such as Google, Yahoo, Bing, Baidu, and Ask. It also enumerates subdomains using Netcraft, VirusTotal, ThreatCrowd, DNSdumpster, and ReverseDNS.

  Syntax:

  ```
  sublist3r [-d DOMAIN] [-b BRUTEFORCE] [-p PORTS] [-v VERBOSE][-t
  THREADS] [-e ENGINES] [-o OUTPUT]
  ```

Short Form	Long Form	Description
-d	--domain	Domain name to enumerate subdomains of
-b	--bruteforce	Enable the subbrute bruteforce module
-p	--ports	Scan the found subdomains against specific TCP ports
-v	--verbose	Enable the verbose mode and display results in real time
-t	--threads	Number of threads to use for subbrute bruteforce
-e	--engines	Specify a comma-separated list of search engines
-o	--output	Save the results to a text file
-h	--help	Show the help message and exit

Table 2.4: Sublist3r options with description

Examples 1:

As shown in the screenshot, Sublist3r helps attackers in enumerating the subdomains of a target company from multiple sources at the same time.

Figure 2.15: Screenshot of Sublist3r displaying sub-domains of google.com

Examples 2:

Sublist3r also helps attackers in enumerating the subdomains of a target company with a specific port open.

As shown in the screenshot, attackers search for subdomains of google.com (-d google.com) using the Bing search engine (-e Bing) with port 80 (-p 80) open.

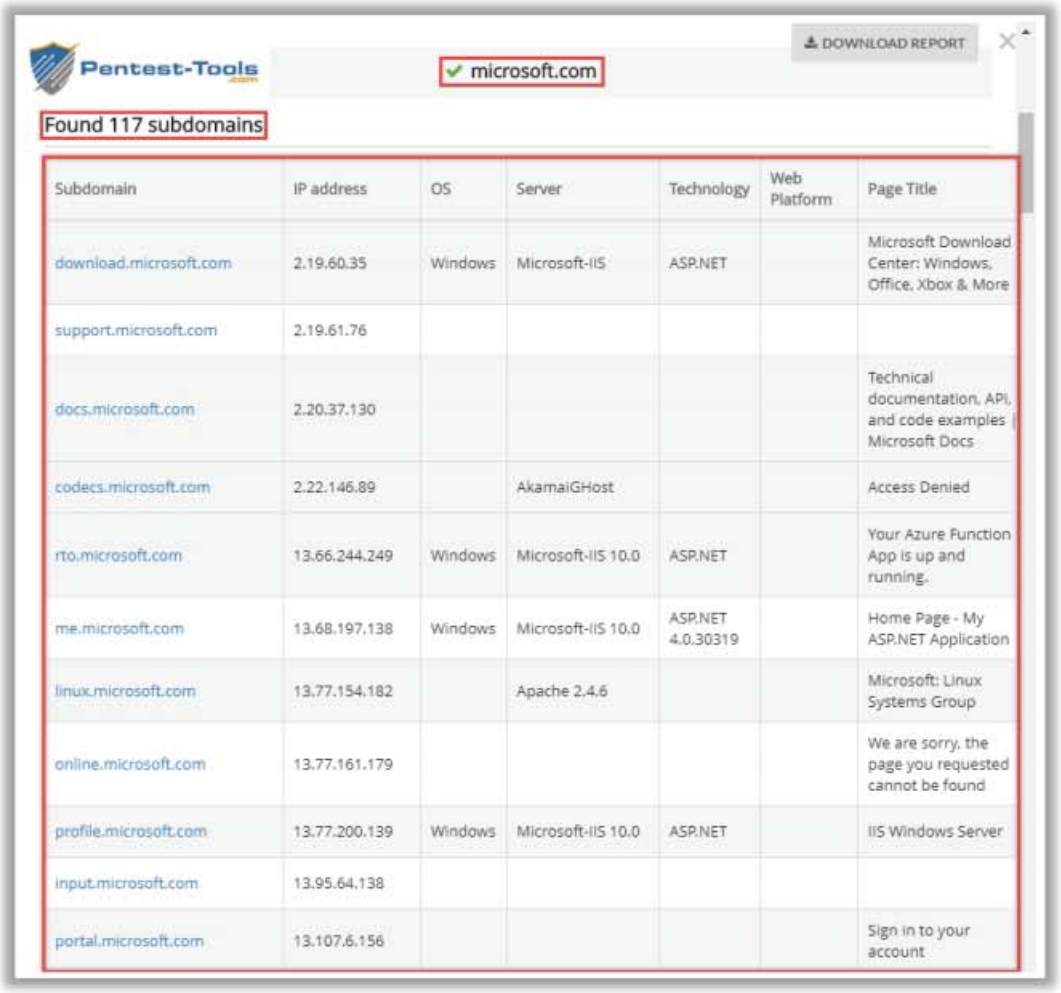
Figure 2.16: Screenshot of Sublist3r displaying sub-domains of google.com with port 80 open

- **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.



Pentest-Tools .com

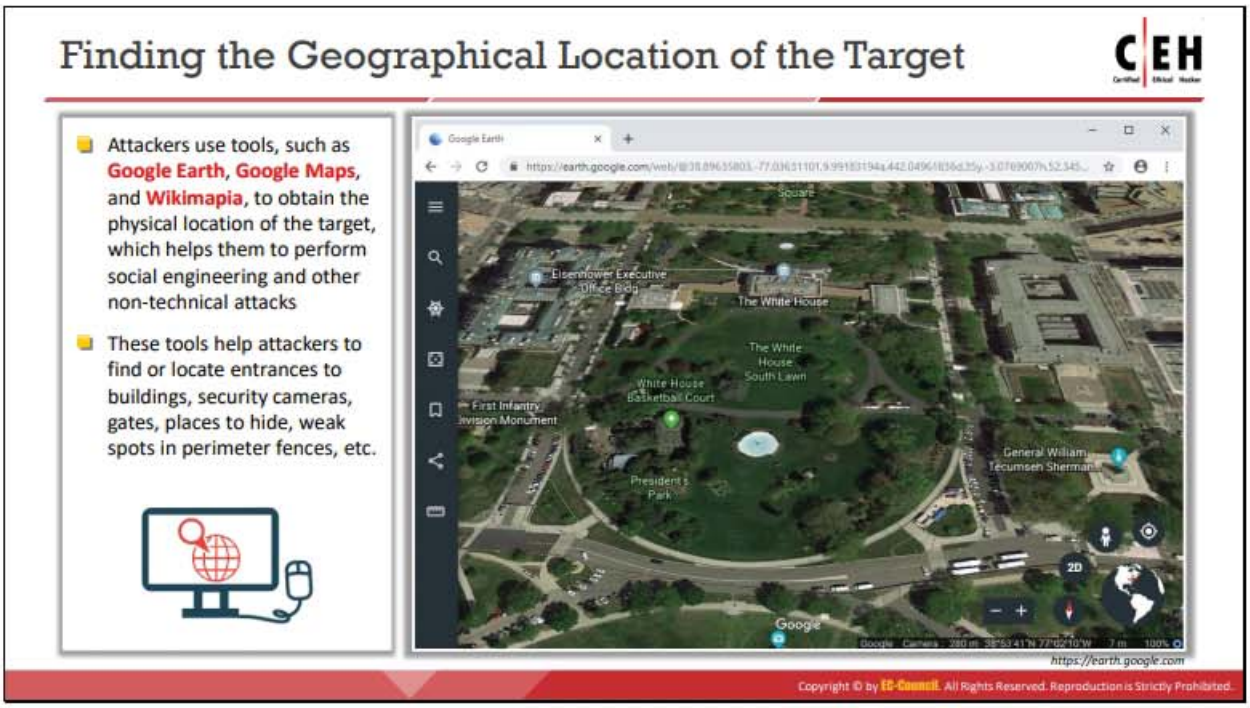
✔ microsoft.com

DOWNLOAD REPORT

Found 117 subdomains

Subdomain	IP address	OS	Server	Technology	Web Platform	Page Title
download.microsoft.com	2.19.60.35	Windows	Microsoft-IIS	ASP.NET		Microsoft Download Center: Windows, Office, Xbox & More
support.microsoft.com	2.19.61.76					
docs.microsoft.com	2.20.37.130					Technical documentation, API, and code examples	Microsoft Docs
codecs.microsoft.com	2.22.146.89		AkamaiGHost			Access Denied
rto.microsoft.com	13.66.244.249	Windows	Microsoft-IIS 10.0	ASP.NET		Your Azure Function App is up and running.
me.microsoft.com	13.68.197.138	Windows	Microsoft-IIS 10.0	ASP.NET 4.0.30319		Home Page - My ASP.NET Application
linux.microsoft.com	13.77.154.182		Apache 2.4.6			Microsoft: Linux Systems Group
online.microsoft.com	13.77.161.179					We are sorry, the page you requested cannot be found
profile.microsoft.com	13.77.200.139	Windows	Microsoft-IIS 10.0	ASP.NET		IIS Windows Server
input.microsoft.com	13.95.64.138					
portal.microsoft.com	13.107.6.156					Sign in to your account

Figure 2.17: Screenshot of Pentest-Tools displaying sub-domains of microsoft.com



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 can obtain this information using footprinting. In addition to the physical location, a hacker can also acquire information such as surrounding public Wi-Fi hotspots that may offer a way to break into the target organization's network.

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. Once the attackers discern the location of the target, they can obtain detailed satellite images of the location using various sources available on the Internet such as Google Earth and Google Maps. The attackers can use this information to gain unauthorized access to buildings, wired and wireless networks, and systems.

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. Tools such as Google Maps even locate entrances of the building, security cameras, and gates. These tools provide interactive maps, outline maps, satellite imagery, and information on how to interact with and create one's own maps. Google Maps, Yahoo Maps, and other tools provide driving directions, traffic conditions, landmarks, and detailed address and contact information.

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** (*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.

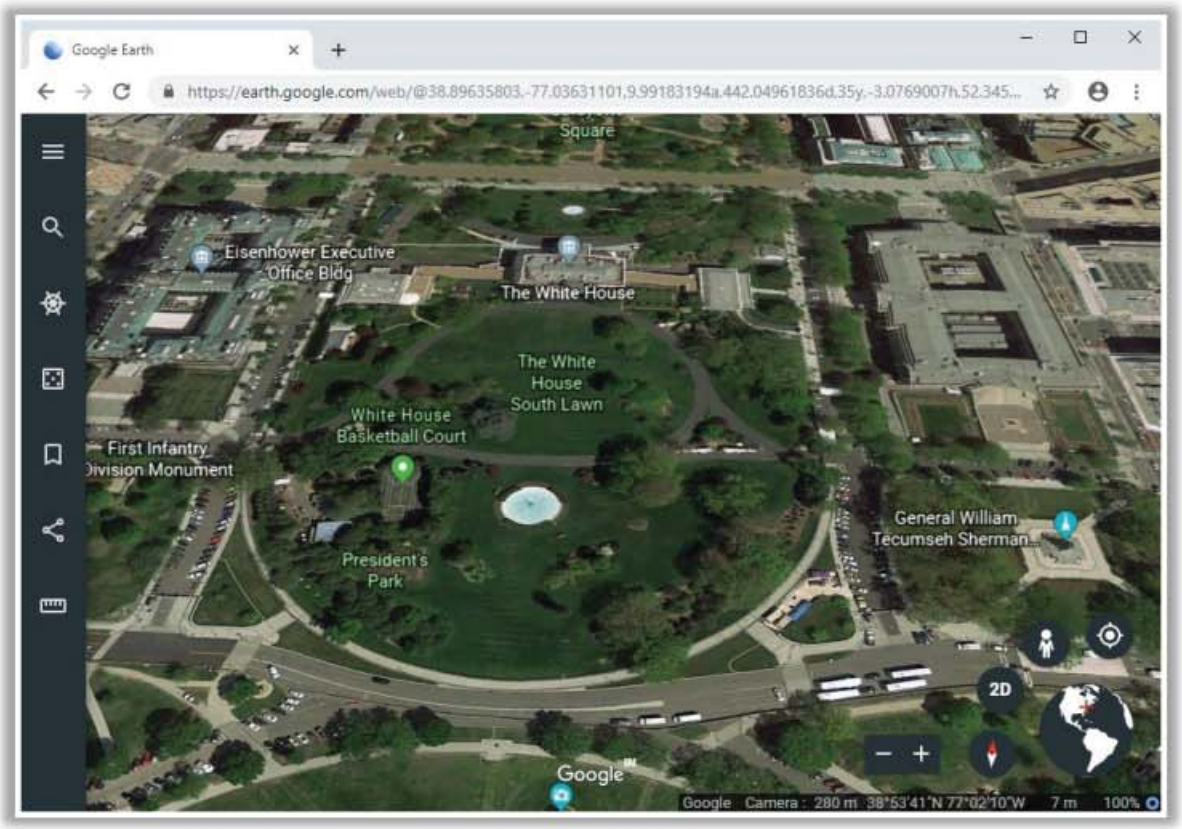


Figure 2.18: Screenshot of Google Earth

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 that users provide in their profiles. They help to directly or indirectly relate people to each other through various fields such as common interests, work location, 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.

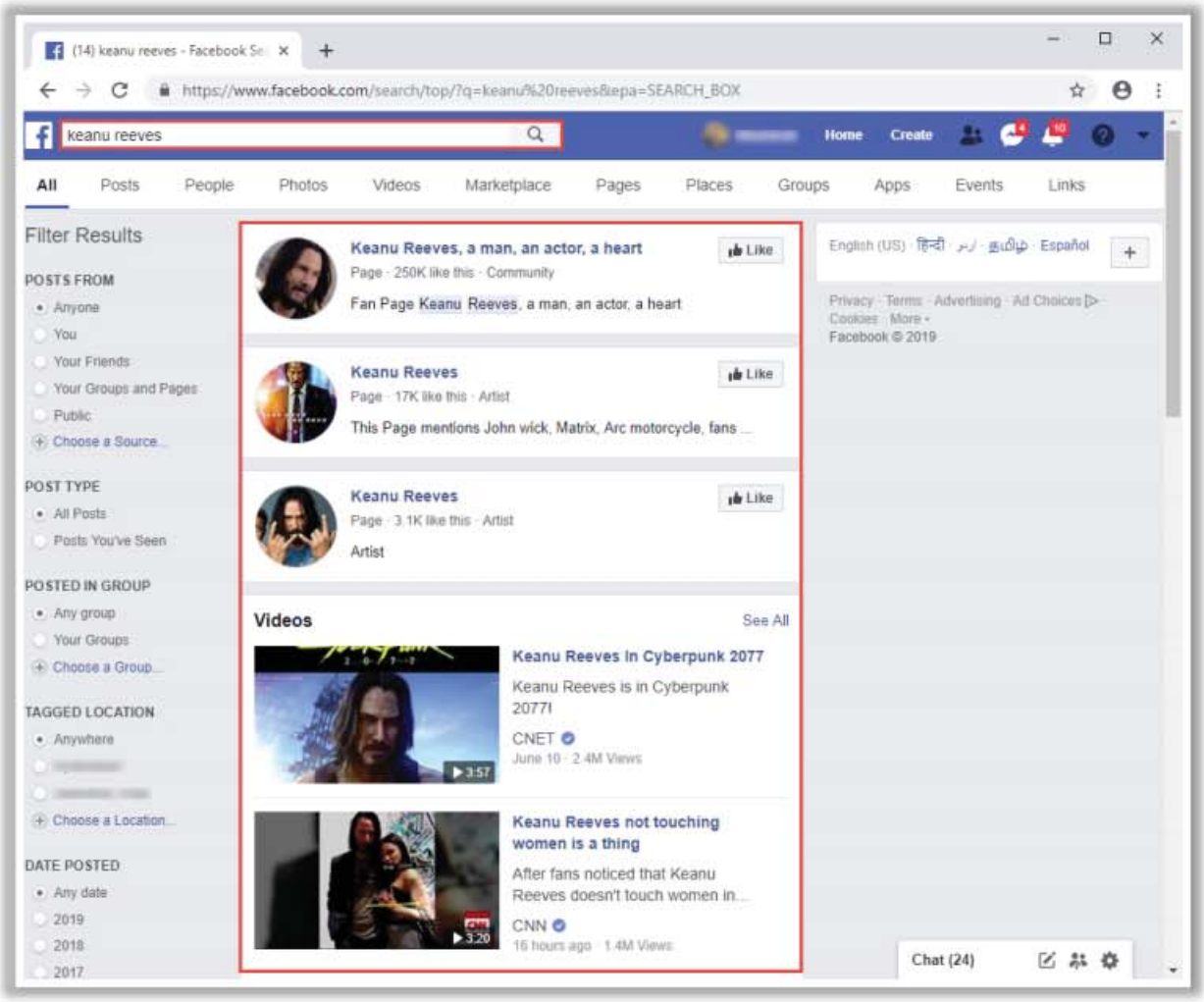


Figure 2.19: Screenshot of Facebook showing search results

People Search on 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 pipl, Intelius, BeenVerified, Whitepages, 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. Examples of such people search services include Intelius, pipl, and AnyWho.

- **People search service - Intelius**

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

  Attackers can use the Intelius people search online service to search for people belonging to the target organization. Using this service, attackers obtain information such as phone numbers, address history, age, date of birth, relatives, previous work history, educational background, and so on.

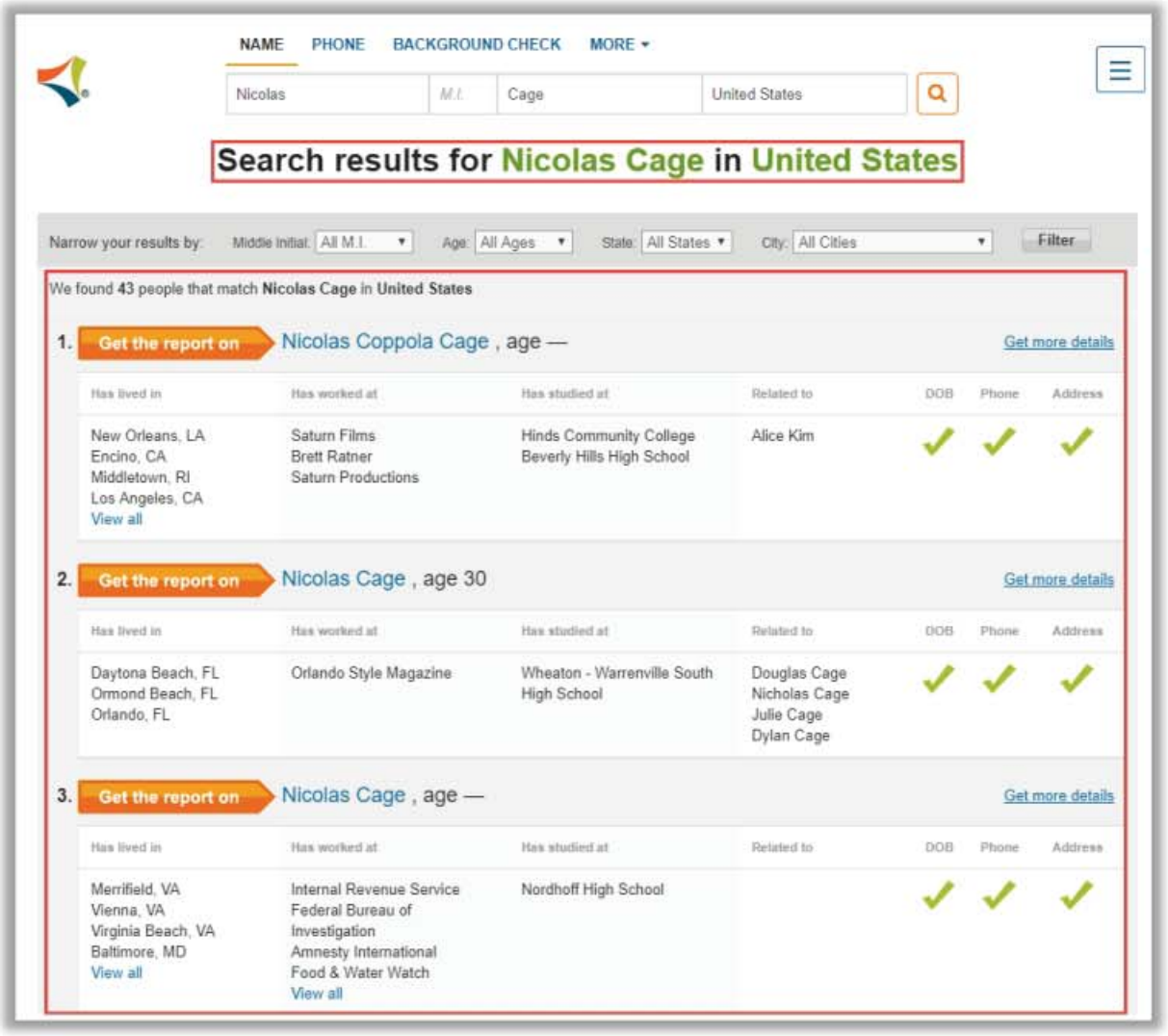


Figure 2.20: Screenshot of Intelius People Search

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: *http://www.edge-security.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.

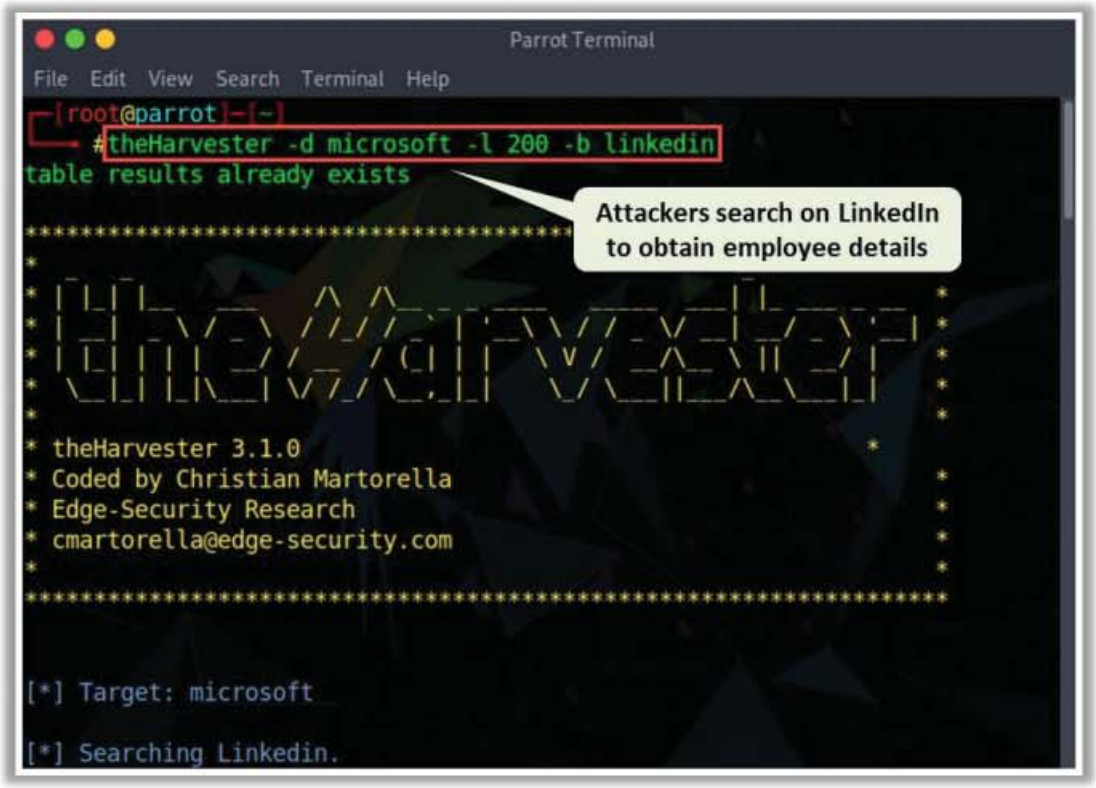


Figure 2.21: Screenshot showing theHarvester command to enumerate users on LinkedIn

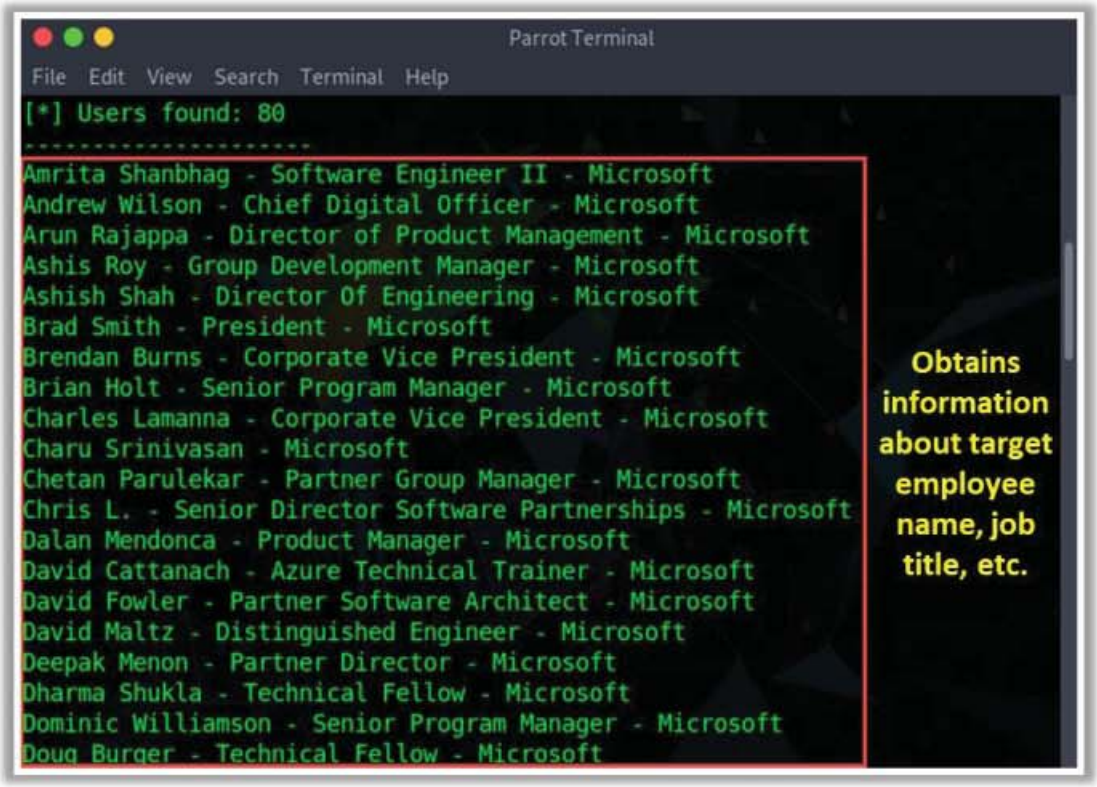
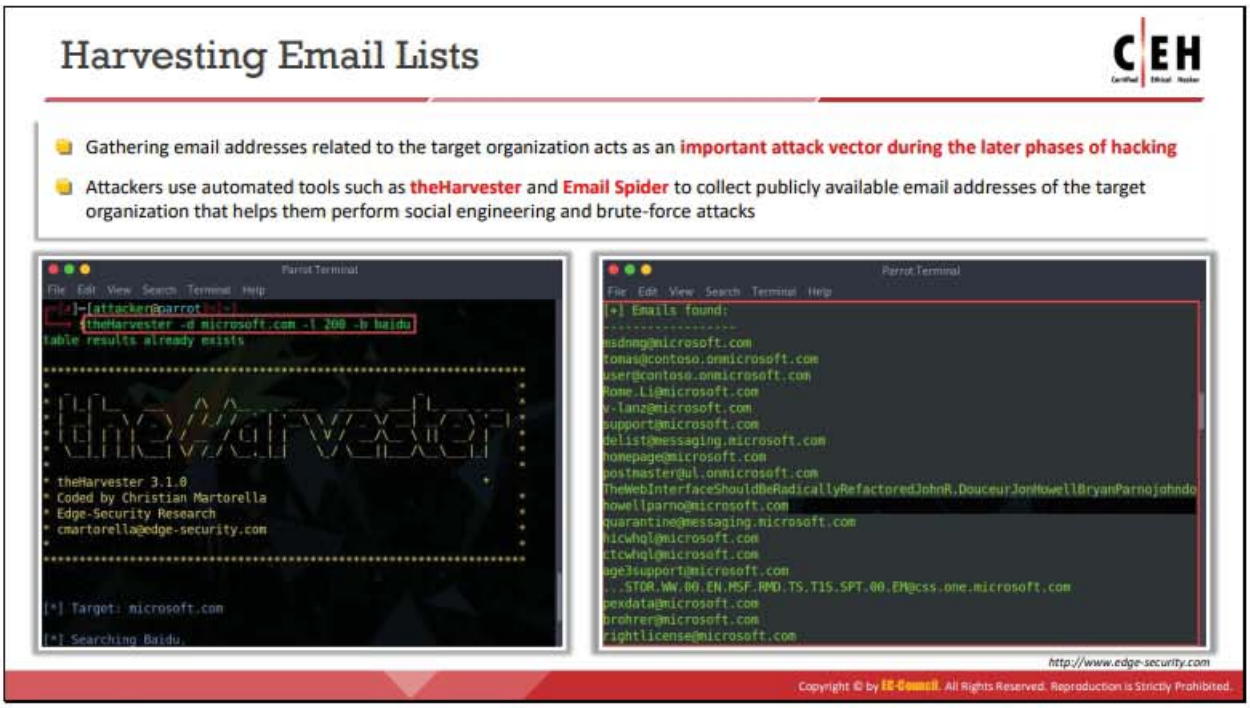


Figure 2.22: Screenshot showing theHarvester search results from LinkedIn



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 Baidu. Attackers use these email lists and usernames to perform social engineering and brute force attacks on the target organization.

- **theHarvester**

  Source: *http://www.edge-security.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.

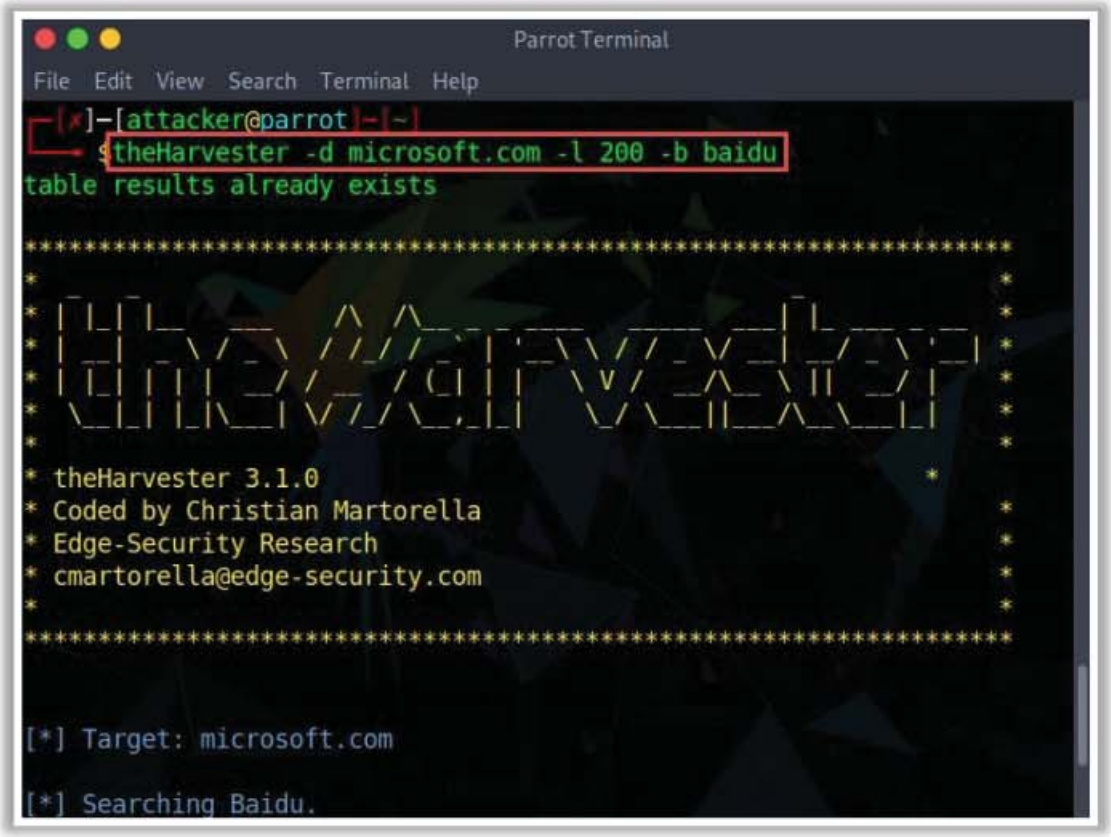


Figure 2.23: Screenshot showing theHarvester command to extract email addresses

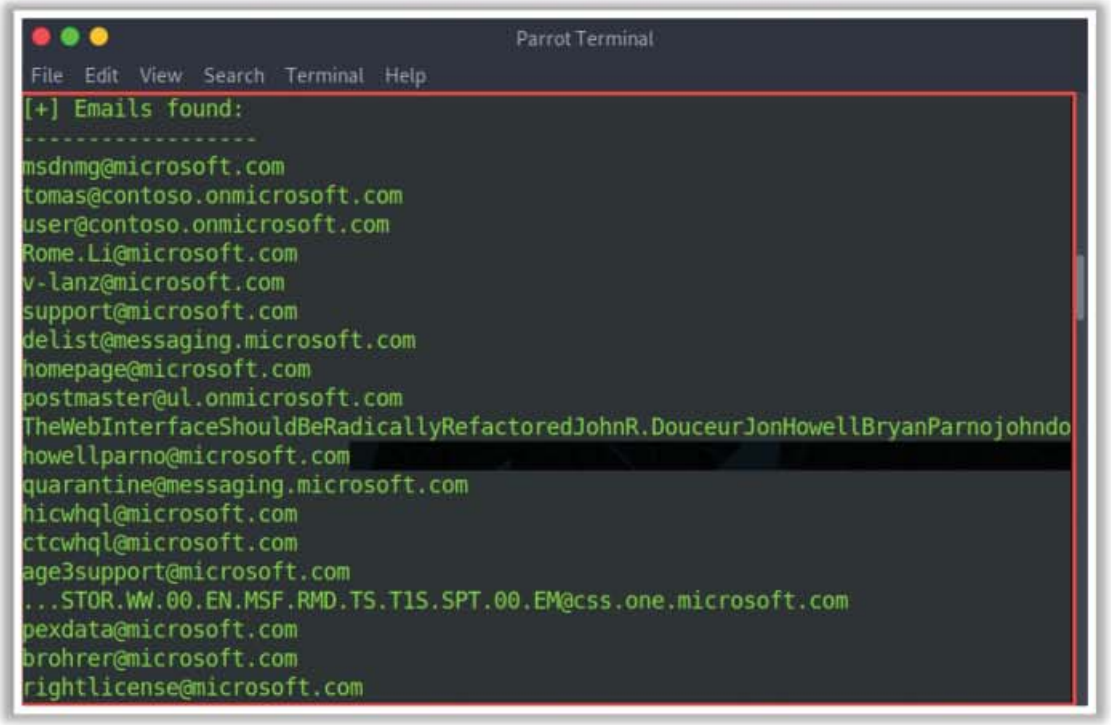
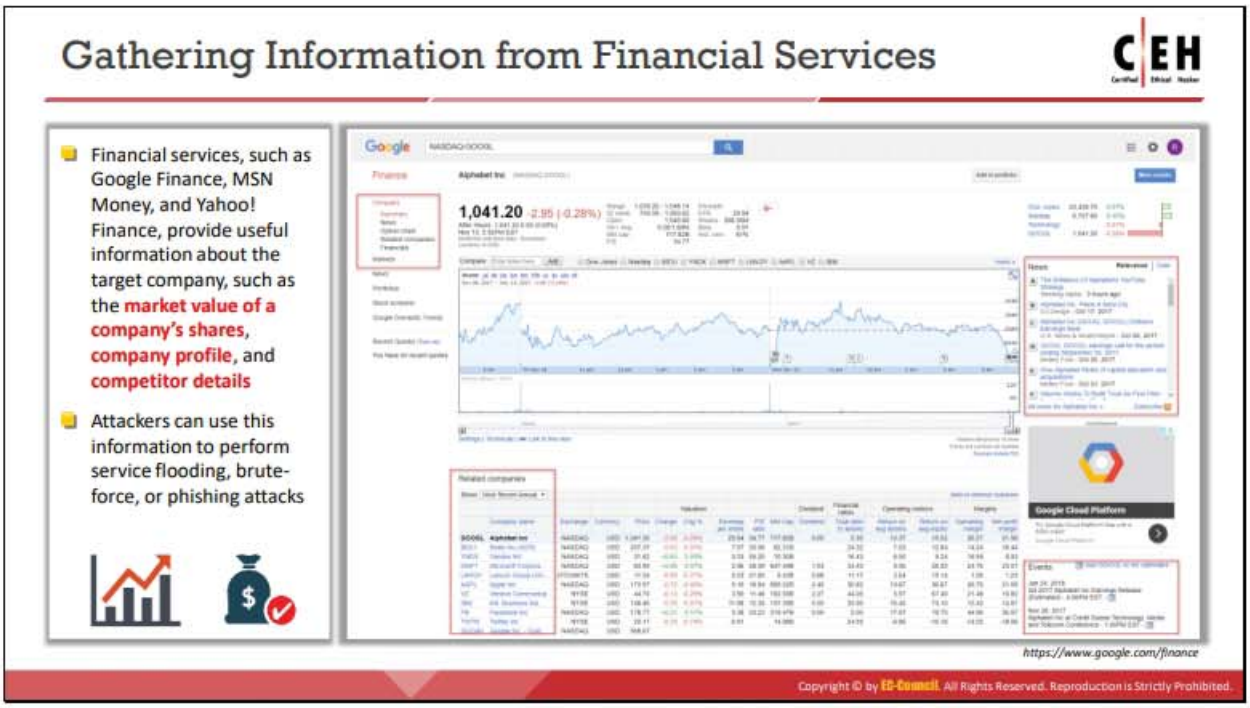


Figure 2.24: Screenshot showing the email list extracted by theHarvester



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.

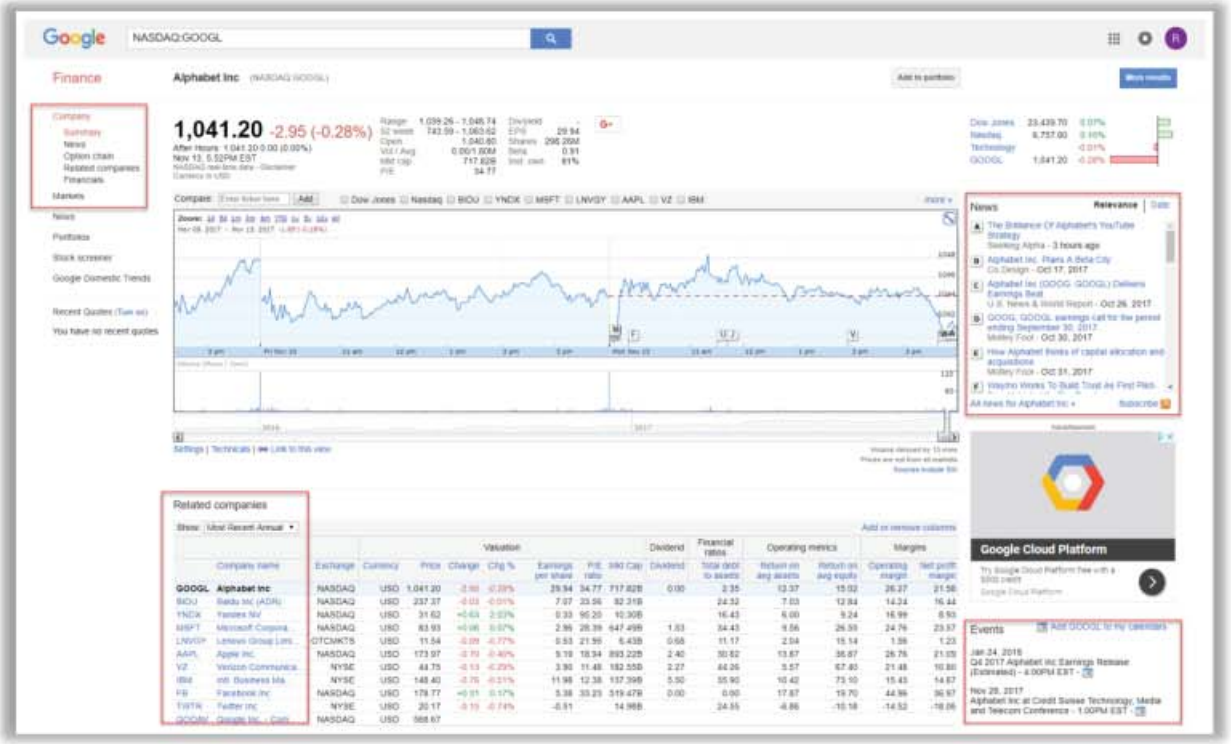
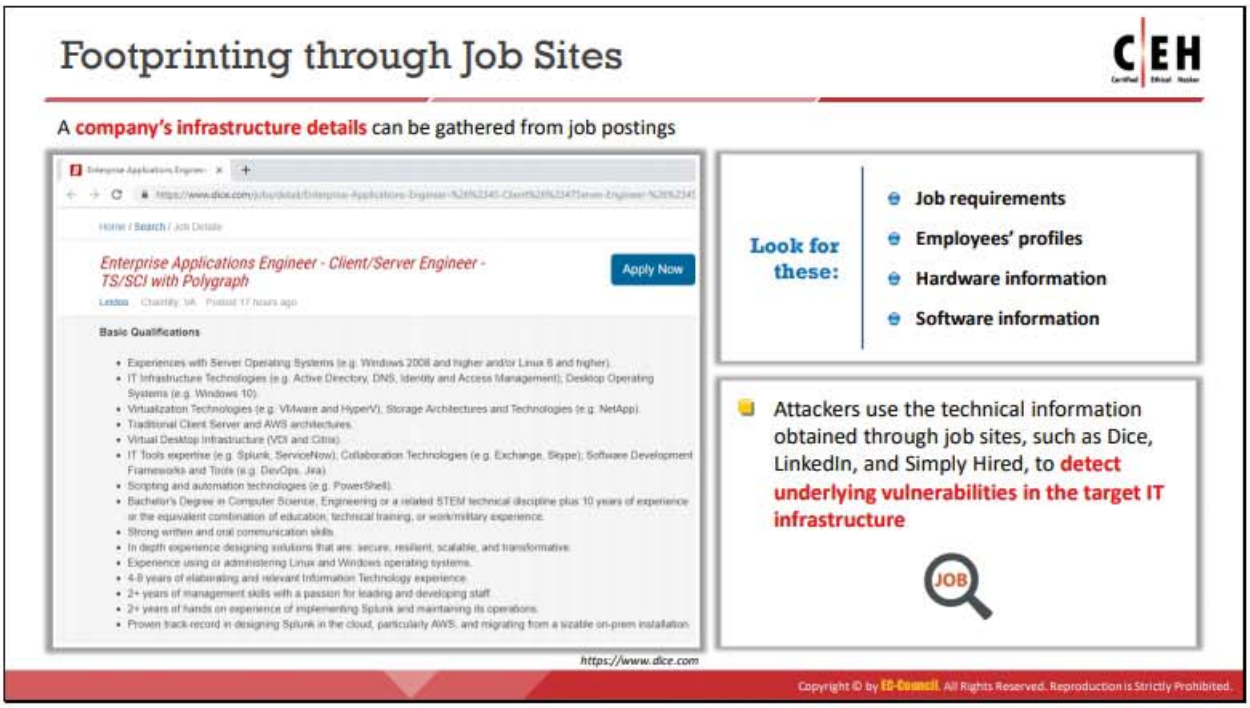


Figure 2.25: Screenshot of Google Finance Service



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, 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, and Simply Hired to detect underlying vulnerabilities in the target IT infrastructure.

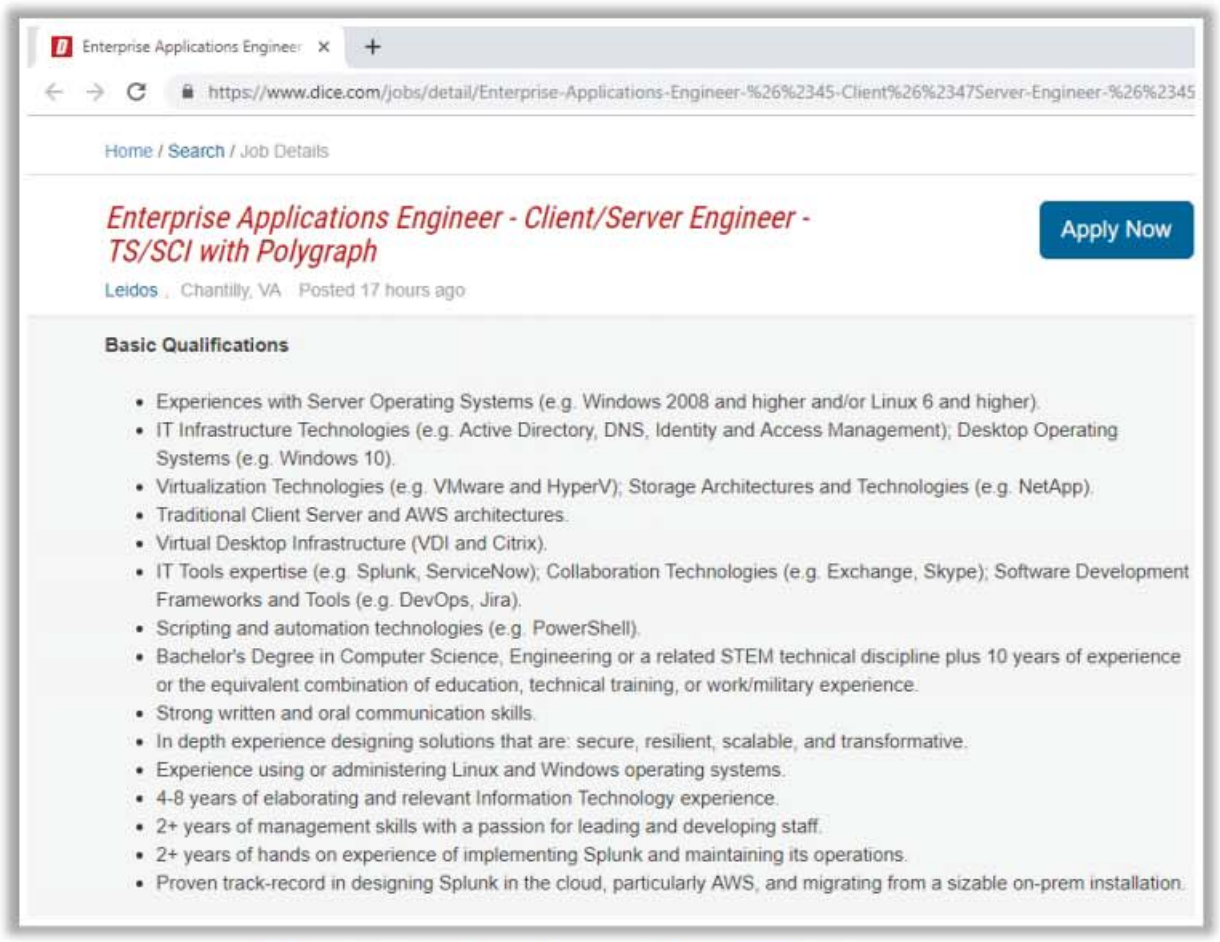


Figure 2.26: Screenshot of job posting showing valuable information

Deep and 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 deep and 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 deep and 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 deep web.

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

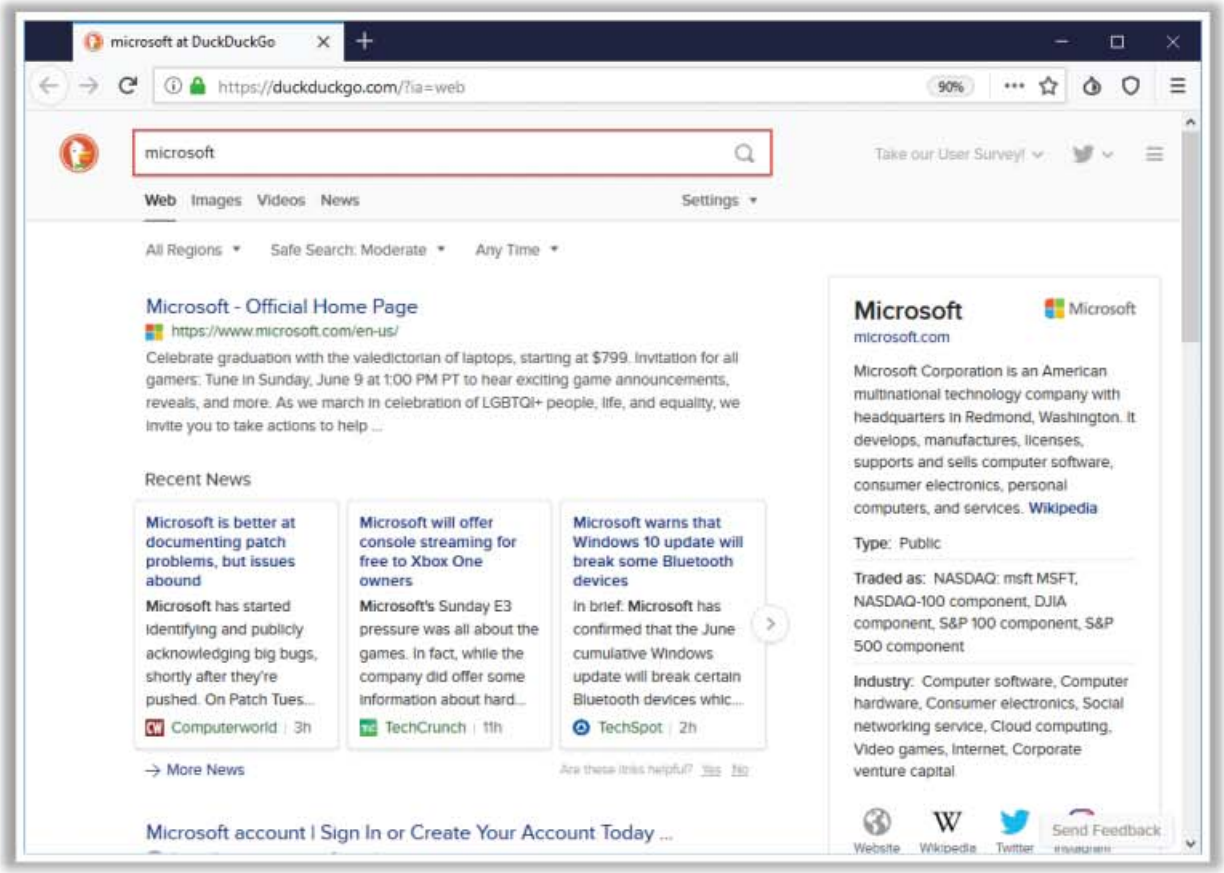
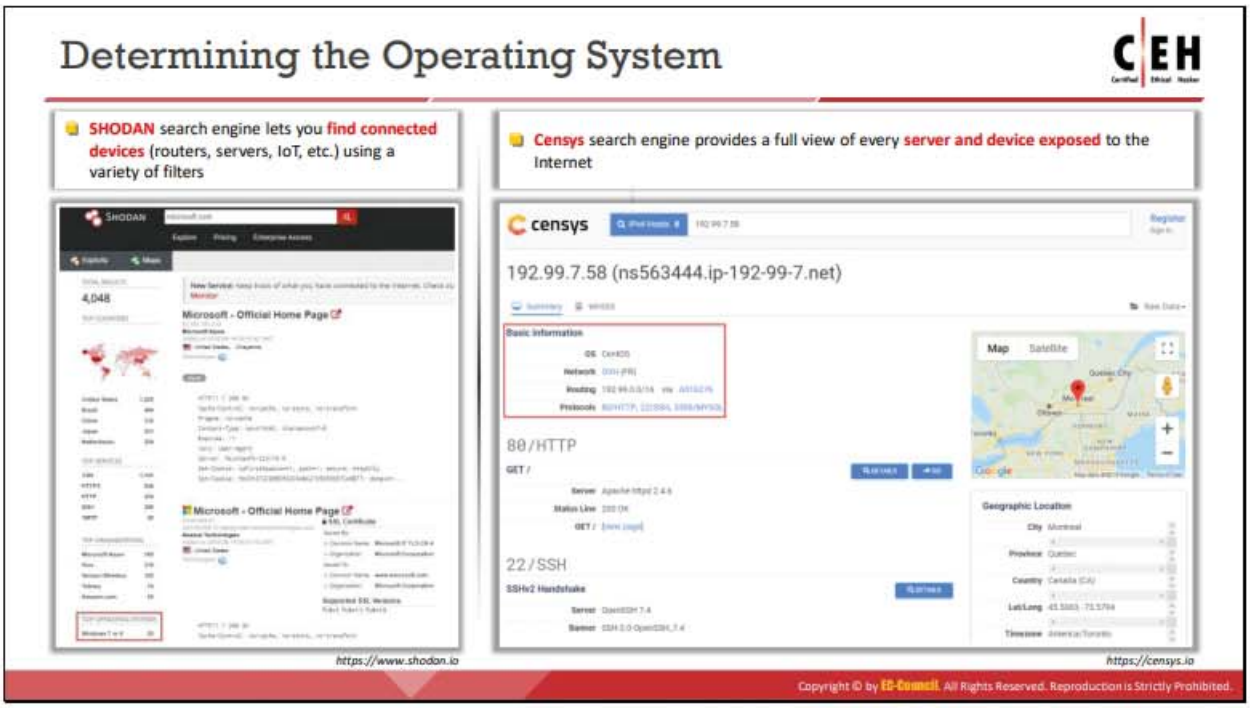


Figure 2.27: Screenshot of Tor Browser



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*** in the browser and type the domain name of the target network 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.

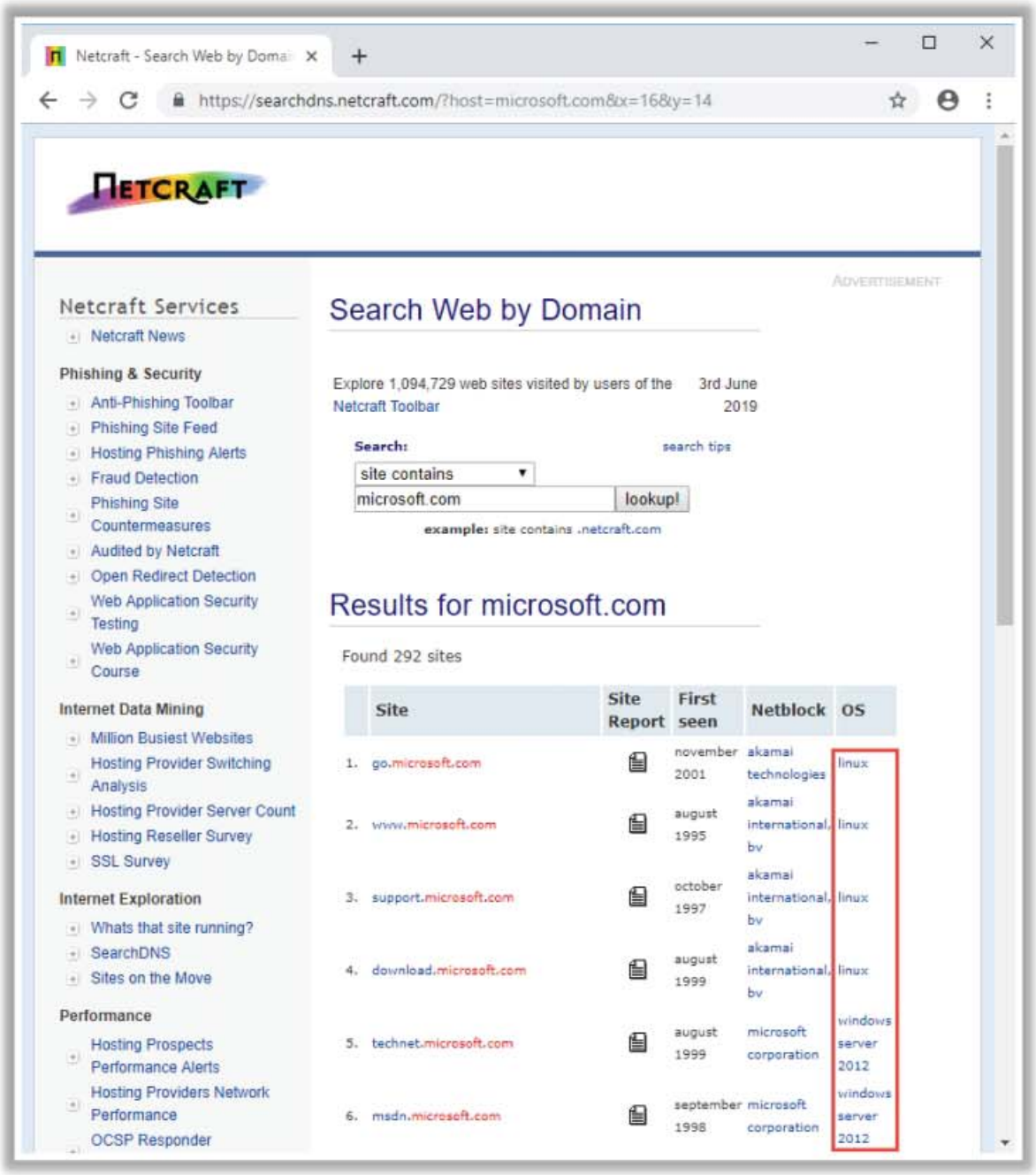


Figure 2.28: Screenshot of Netcraft showing target operating system

- **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.

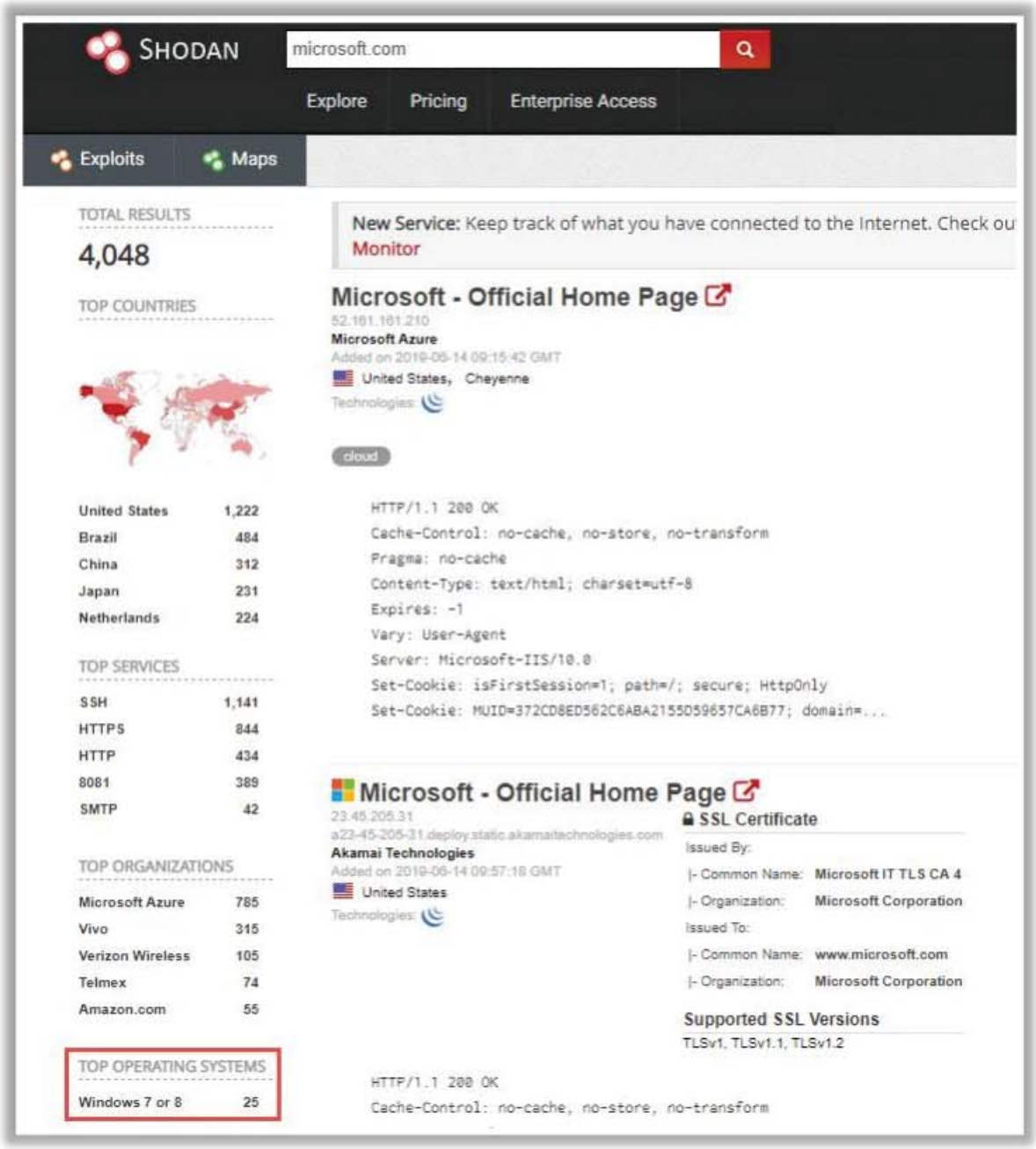


Figure 2.29: Screenshot of SHODAN Search Engine showing target operating system

- **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.

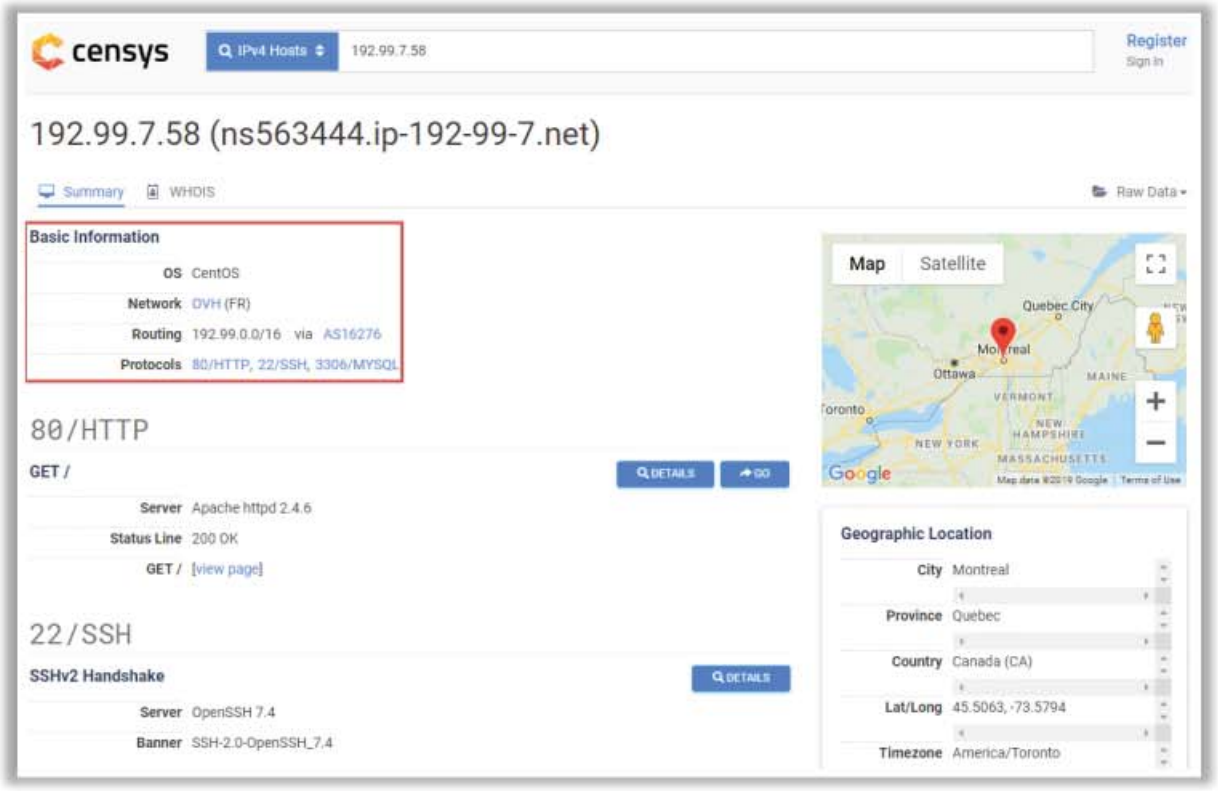
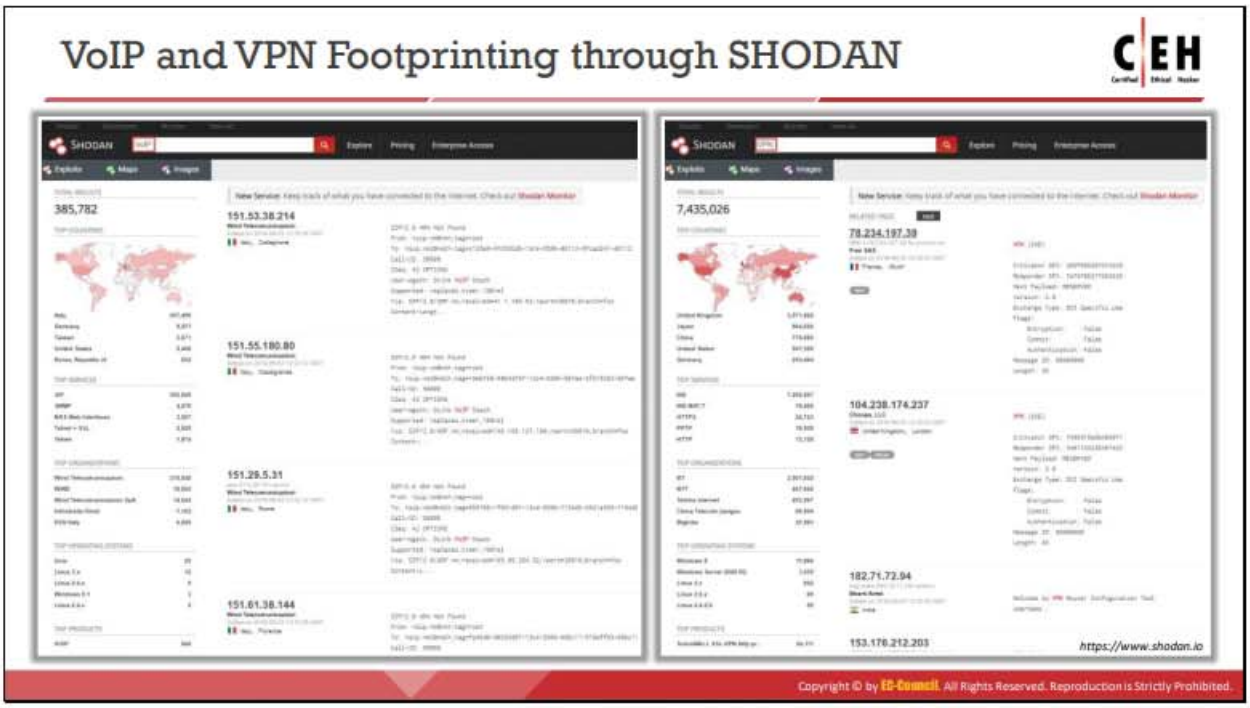


Figure 2.30: Screenshot of Censys Search Engine showing target operating system



VoIP and VPN Footprinting through SHODAN

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:

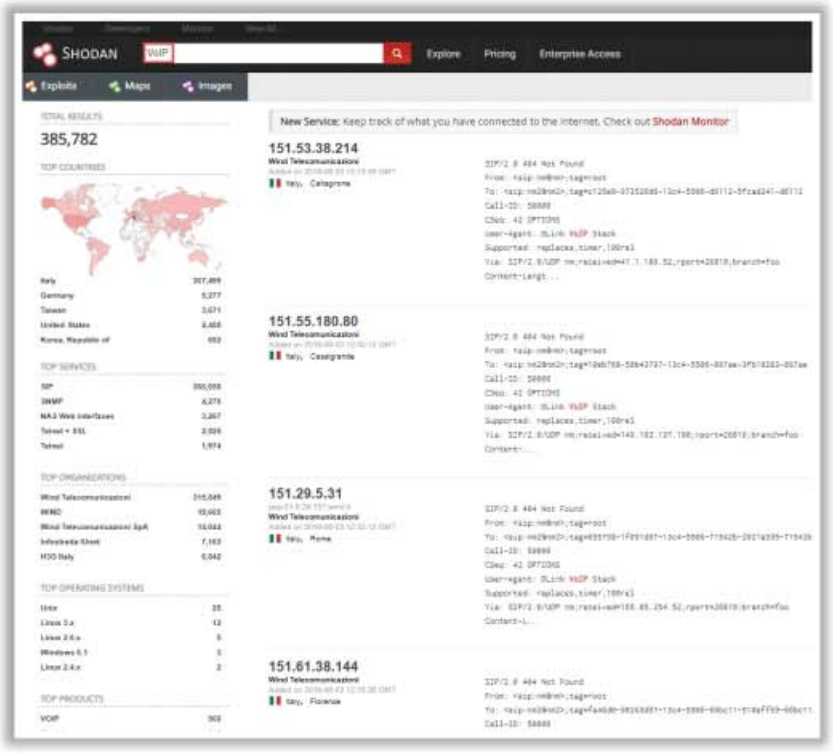


Figure 2.31: Screenshot of SHODAN search engine showing VoIP results

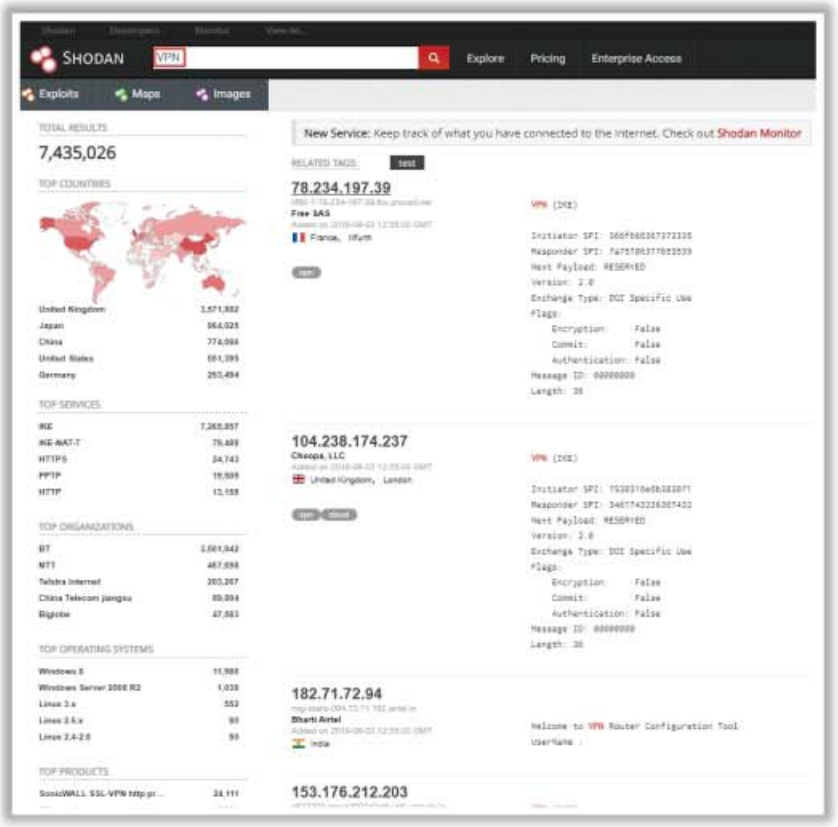
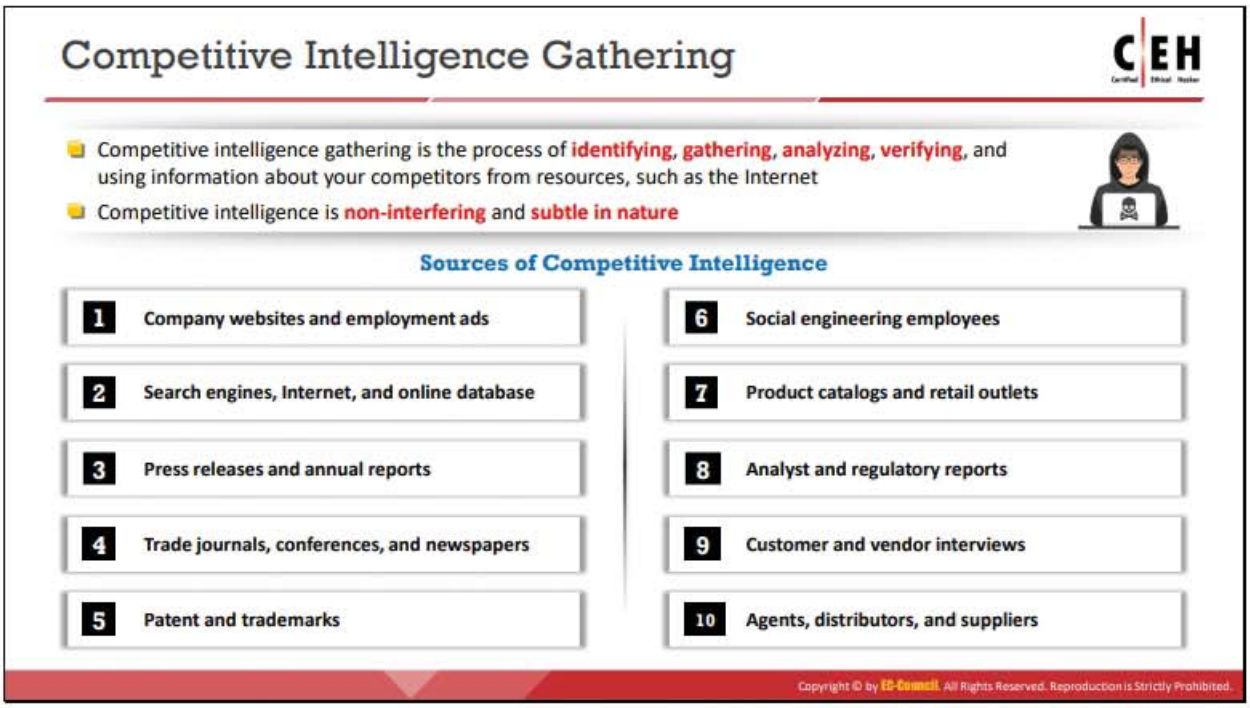
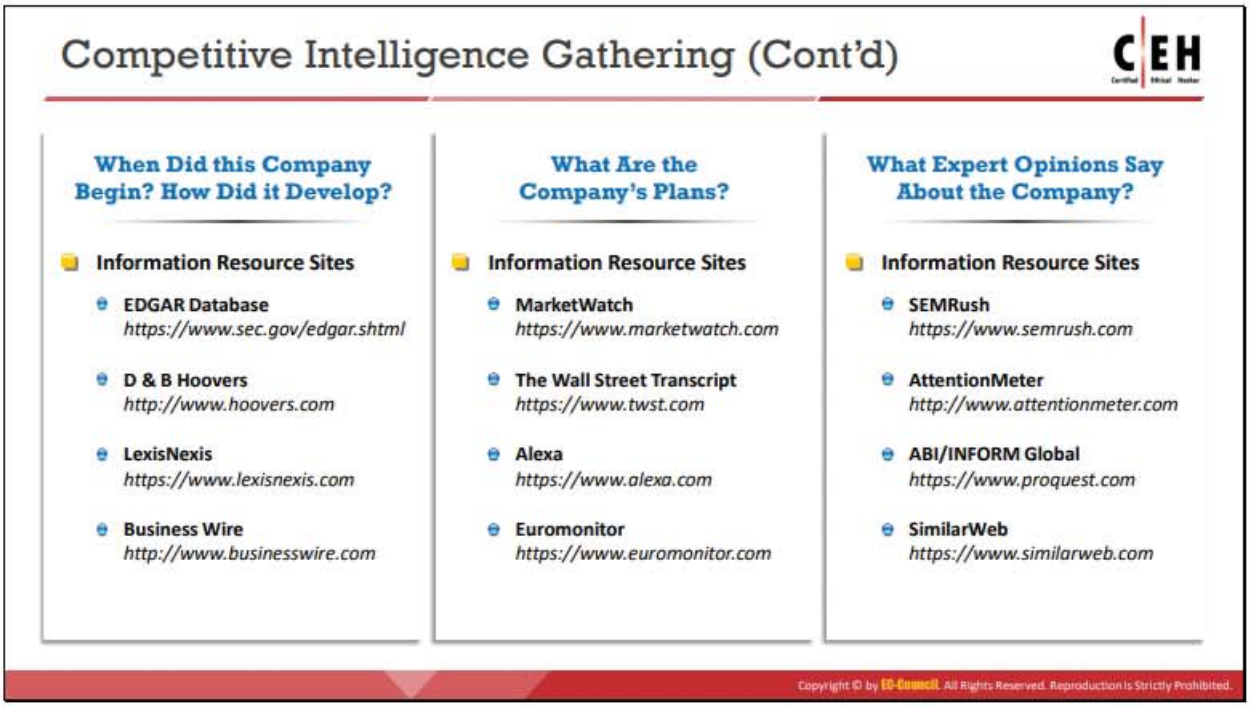


Figure 2.32: Screenshot of SHODAN search engine showing VPN results





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., Hoover's
  - Online job postings

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.shtml*

  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: *http://www.hoovers.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.

- **Alexa**

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

  Alexa is a great tool to dig deep into the analytics of other companies. It allows users to

  - Discover influencer outreach opportunities by uncovering sites that link to their competitors using Competitor Backlink Checker.
  - Benchmark and track their company's performance relative to their competitors using Competitive Intelligence Tools.

- **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

- **SEC Info**

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

  SEC Info offers the U.S. Securities and Exchange Commission (SEC) EDGAR database service on the web, with many links added to SEC documents. It allows searches by name, industry, business, SIC code, area code, accession number, file number, CIK, topic, ZIP code, and so on.

- **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.

- **AttentionMeter**

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

  AttentionMeter is a tool for comparing websites (traffic) by using Alexa, Compete, and Technorati. It gives a snapshot of traffic data as well as graphs from Alexa, Compete, and Technorati for the specified websites.

- **ABI/INFORM Global**

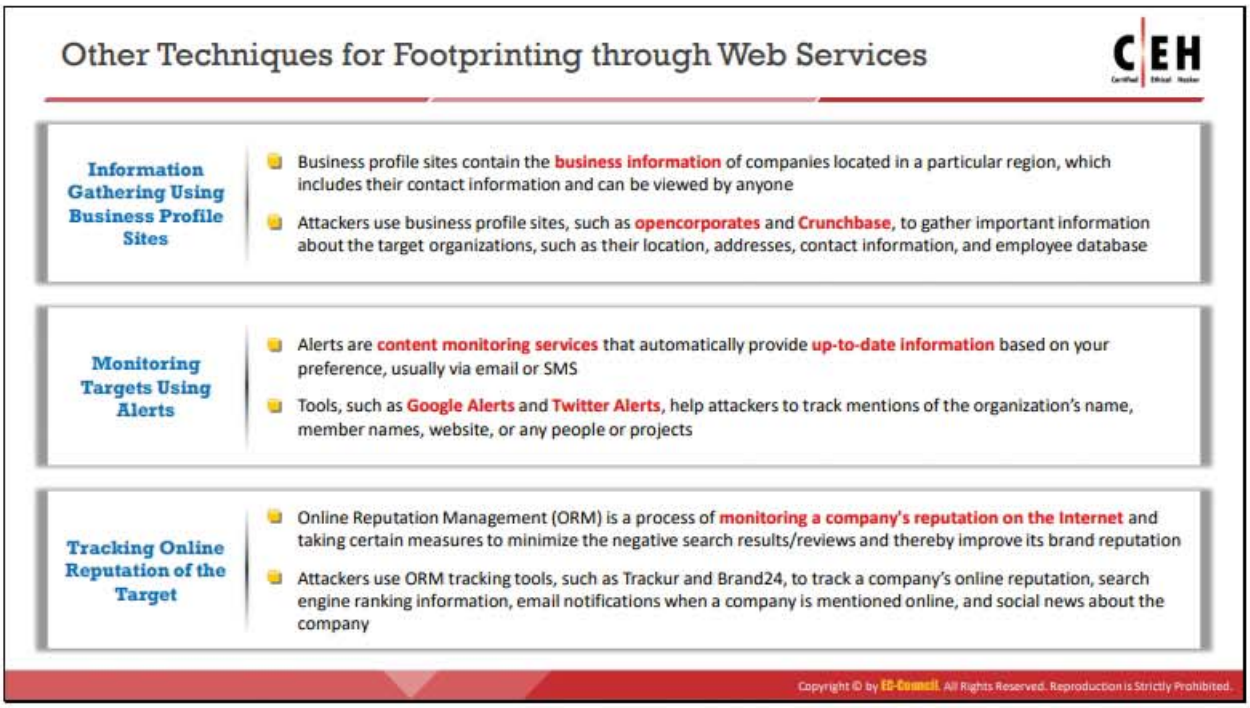
  Source: *https://www.proquest.com*

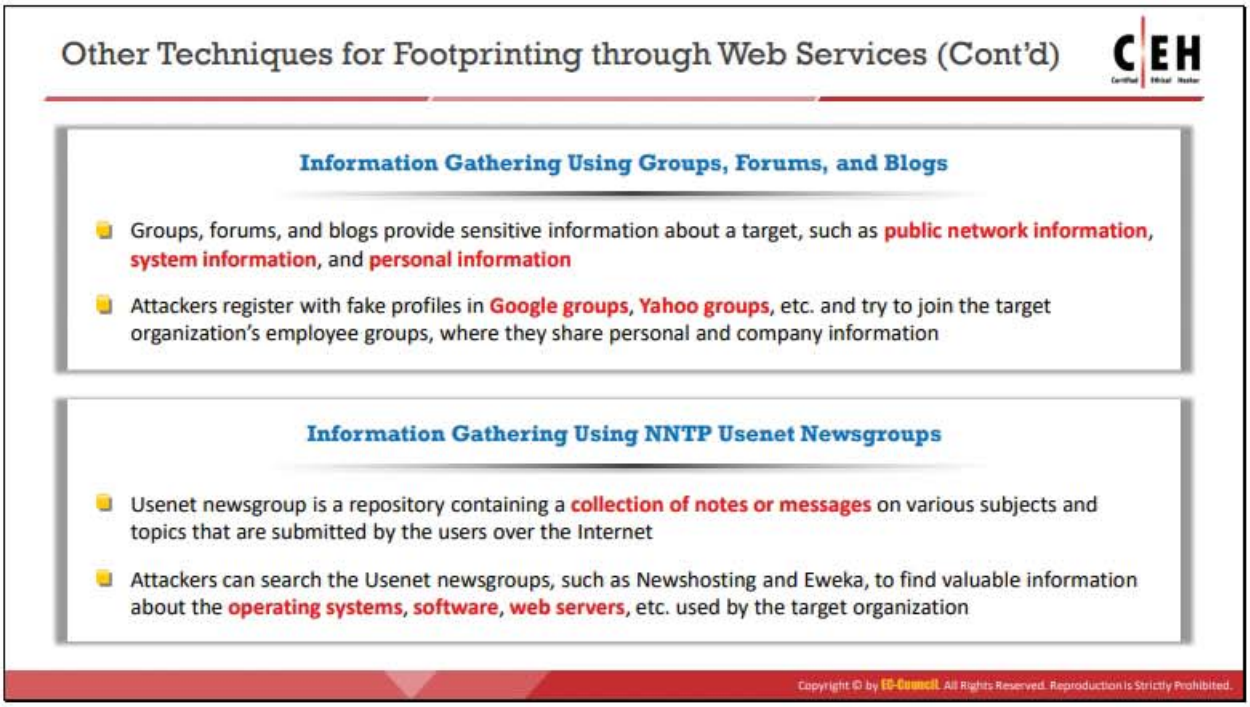
  ABI/INFORM Global is a business database. ABI/INFORM Global offers the latest business and financial information for researchers. With ABI/INFORM Global, users can determine business conditions, management techniques, business trends, management practice and theory, corporate strategy and tactics, and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





Other Techniques for Footprinting through Web Services

- **Information Gathering Using 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.

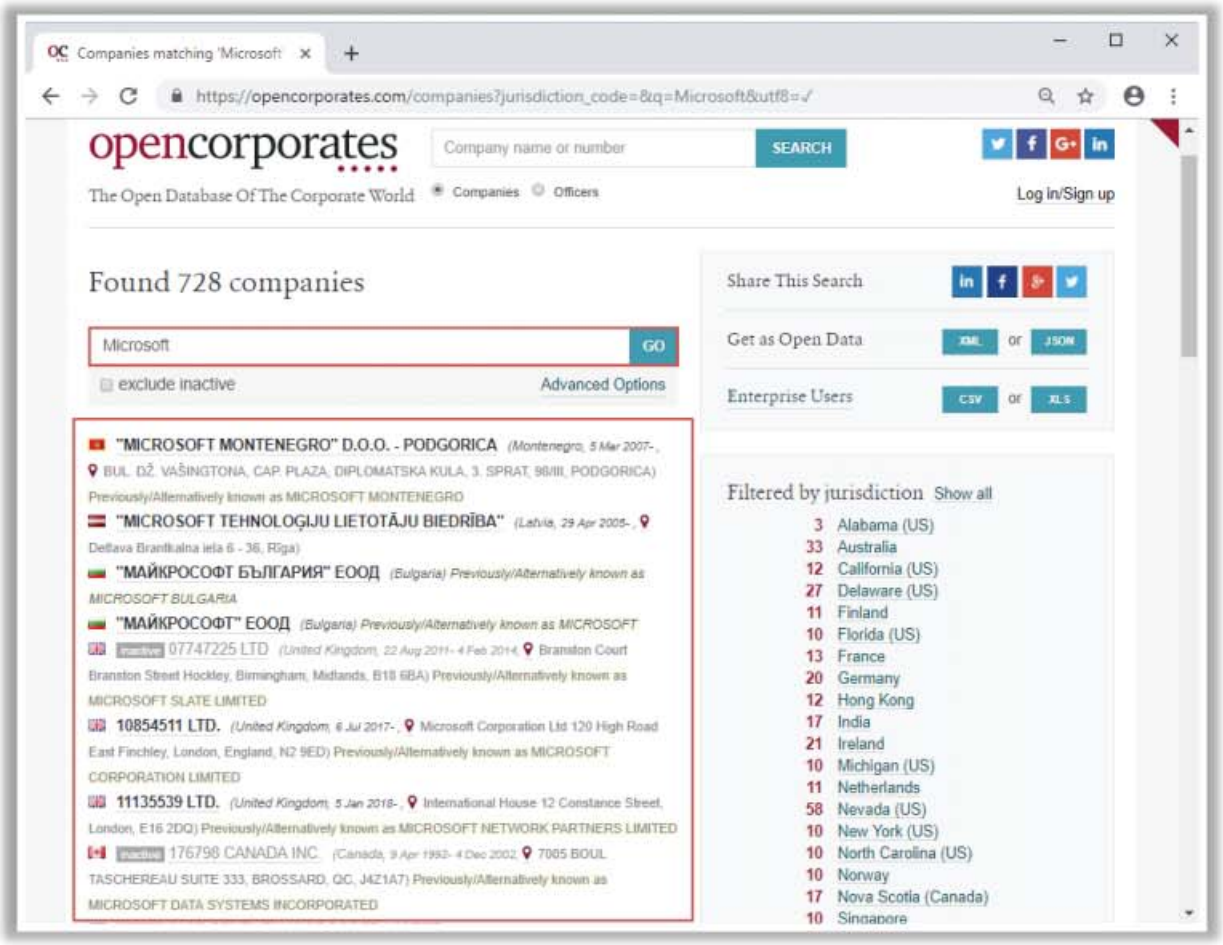


Figure 2.33: Screenshot of opencorporates showing search results of Microsoft

- **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. Alerts are sent via email or SMS notifications.

  Tools such as Google Alerts, Twitter 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.

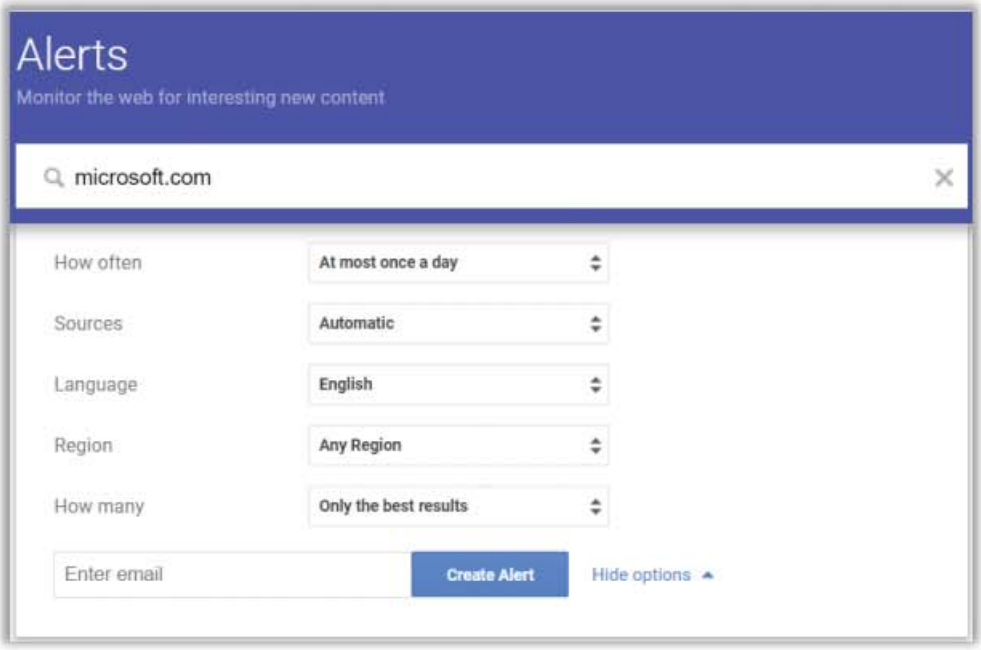


Figure 2.34: Screenshot of Google Alert

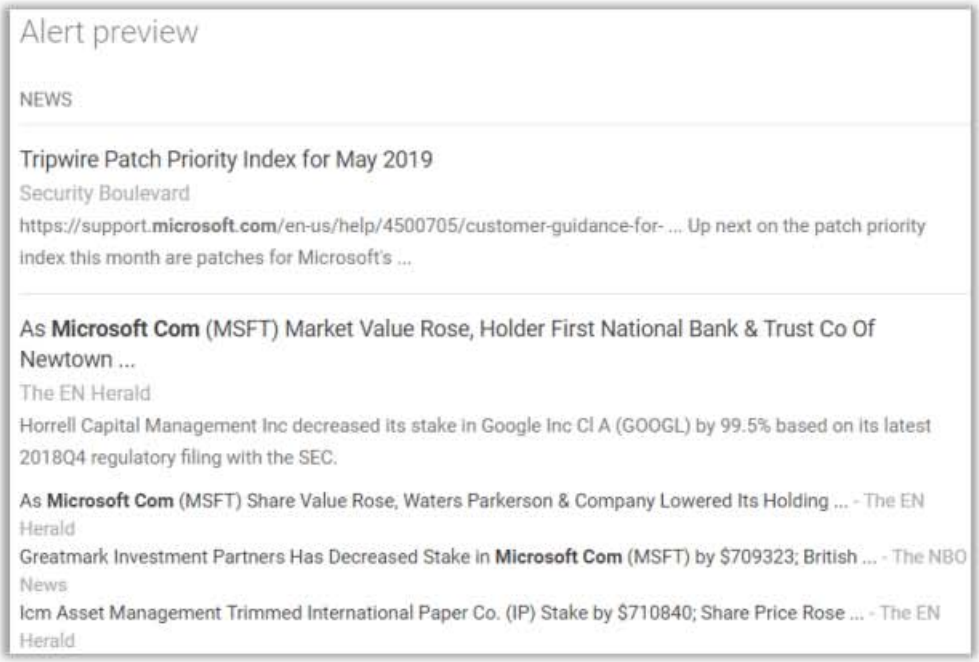


Figure 2.35: Screenshot of Google Alert Preview

- **Tracking 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.

  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.

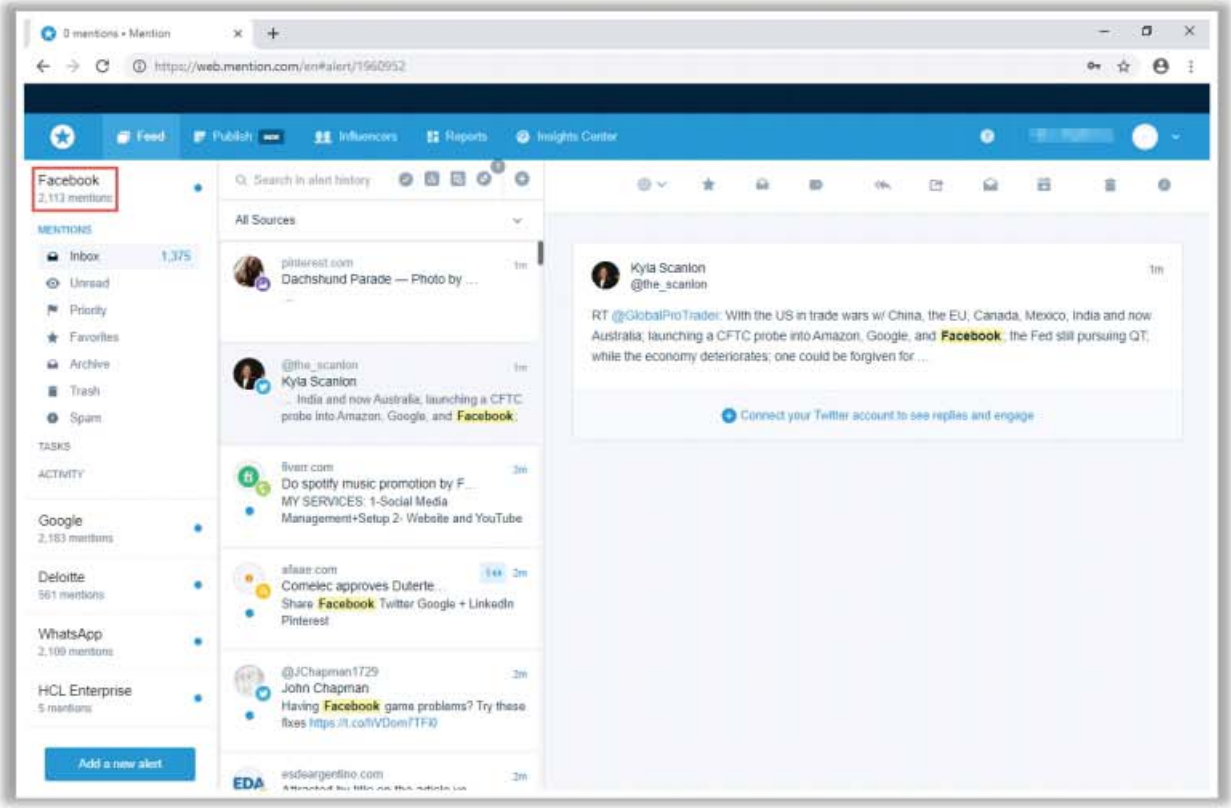


Figure 2.36: Screenshot of Mention

- **Information Gathering Using 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 see 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, Yahoo 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

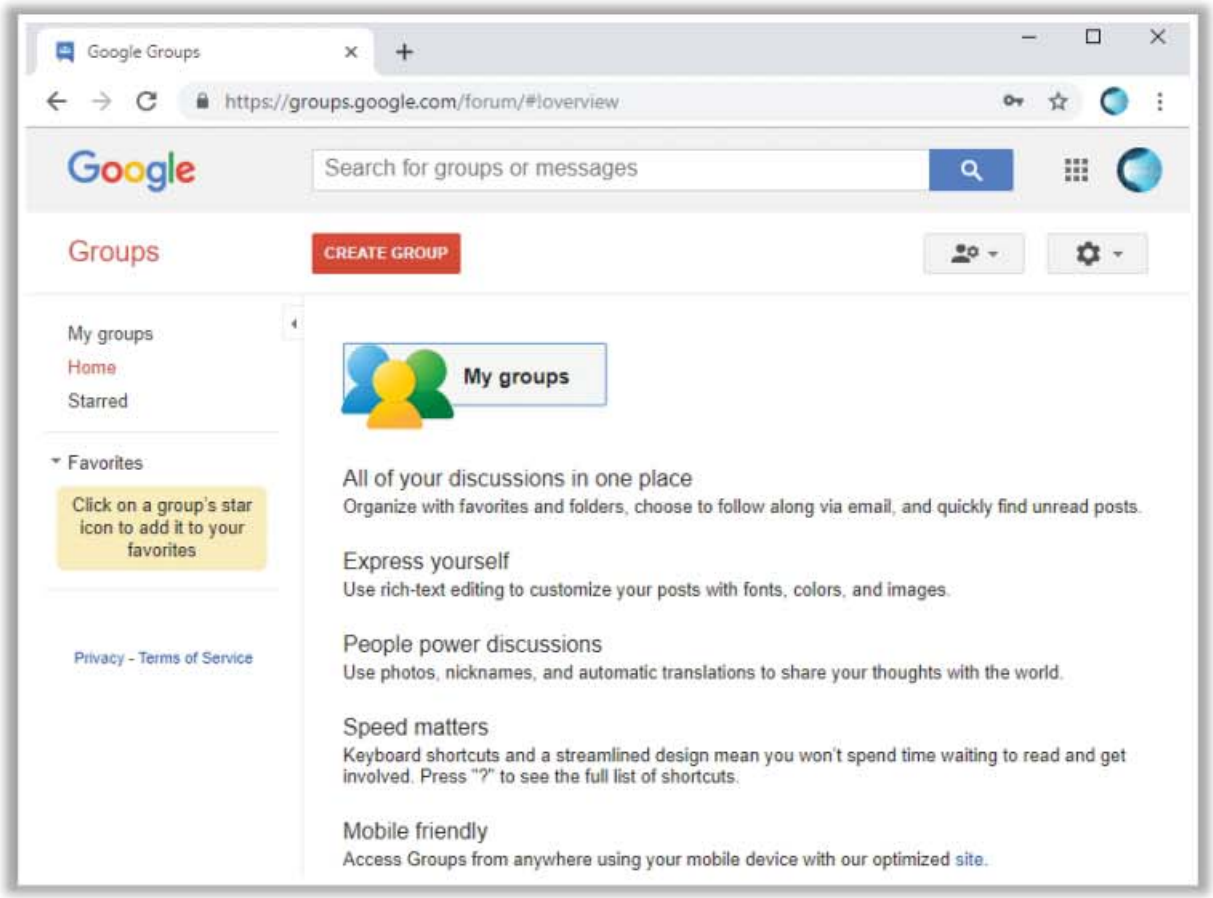


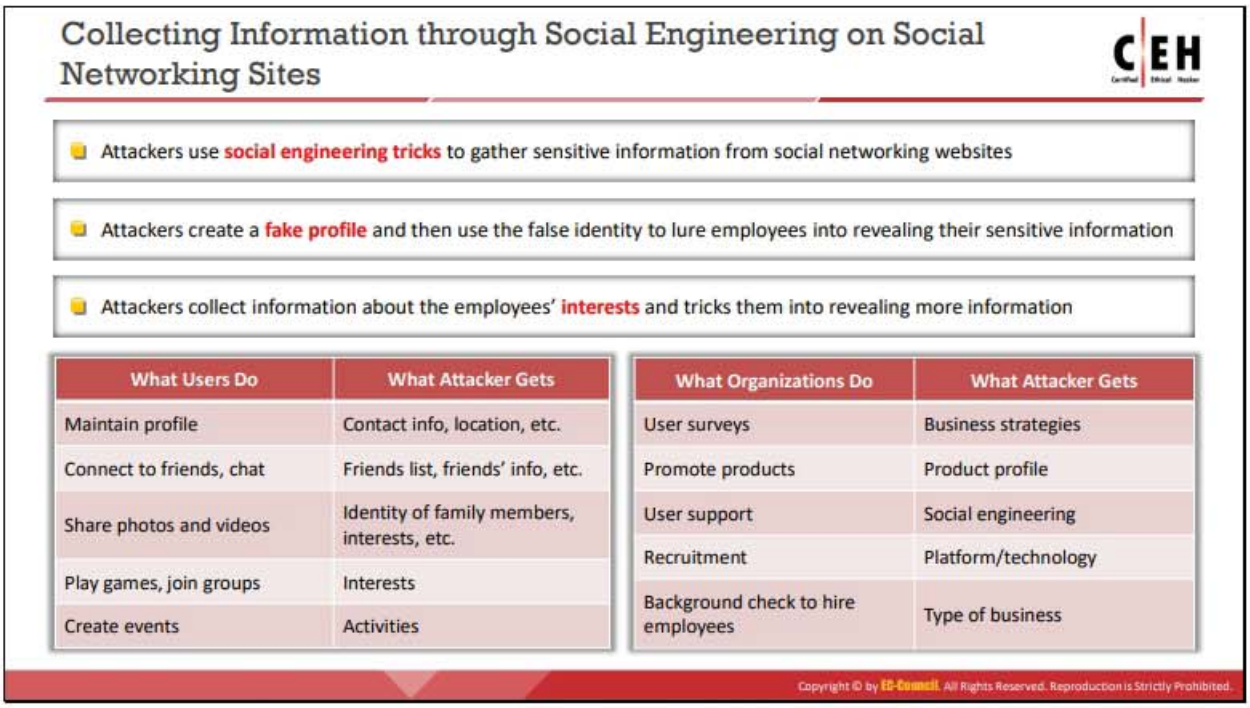
Figure 2.37: Screenshot of Google Groups

- **Information Gathering Using NNTP Usenet Newsgroups**

  Usenet newsgroup is a repository containing a collection of notes or messages on various subjects and topics that are submitted by the users over the Internet. Network News Transfer Protocol (NNTP) is used to relay Usenet news articles from the discussions over the newsgroup. Usenet newsgroups can be a useful source of valuable information about the target. People seek help by posting questions and asking for a solution on Usenet newsgroups. Many professionals use the newsgroups to resolve their technical issues by posting questions on Usenet. To obtain solutions for these issues, sometimes they post more detailed information about the target than needed. Attackers can search Usenet newsgroups or mailing lists such as Newshosting, Eweka, and Supernews to find valuable information about the operating systems, software, web servers, etc., used by the target organization.

For example, from the screenshot given below, you can understand that the target organization is using a Red Hat Linux 6.2 machine that is running Apache web server 1.3.23. This information helps attackers in performing web server and web application attacks.

Figure 2.38: Screenshot of sample USENET newsgroup posting



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 using social engineering and how it can be obtained. It aims to familiarize you with locating information from social media sites using various online services and resources.

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

General Resources for Locating Information from Social Media Sites

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.

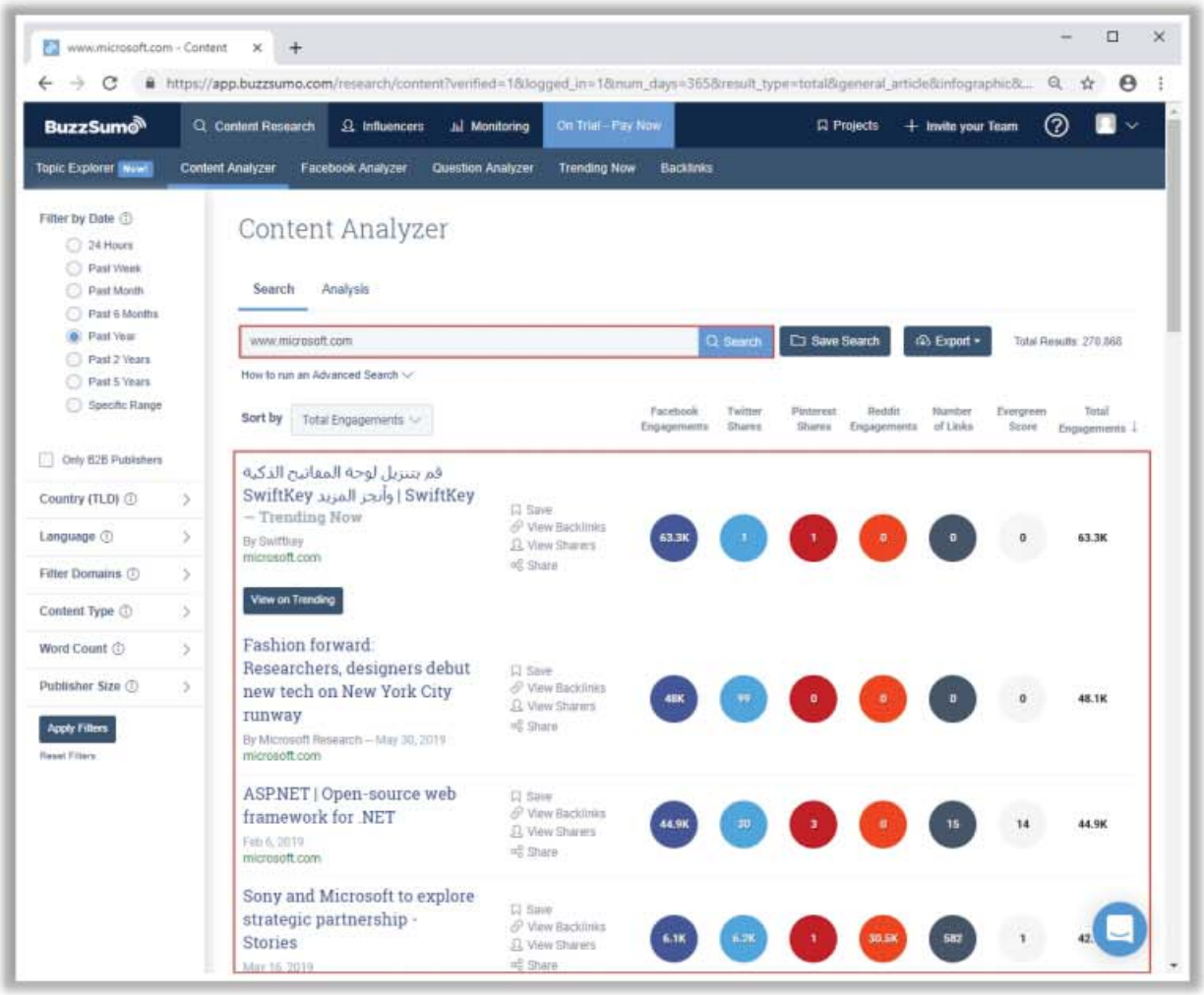


Figure 2.39: Screenshot of BuzzSumo showing the shared content

Conducting Location Search on Social Media Sites

Conducting location search on social media sites such as Twitter, Instagram, and Facebook helps attackers to detect the geolocation of the target. This information further helps attackers to perform various social engineering and non-technical attacks. Many online tools such as Followerwonk, Hootsuite, and Sysomos are available to search for both geotagged and non-geotagged information on social media sites. Attackers search social media sites using these online tools using keywords, usernames, date, time, and so on.

- **Followerwonk**

  Source: *https://followerwonk.com*

  Followerwonk helps you explore and grow your social graph: Dig deeper into Twitter analytics: Who are your followers? Where are they located? When do they tweet?

  As shown in the screenshot, attackers use Followerwonk to track the geolocation of the target Twitter users.

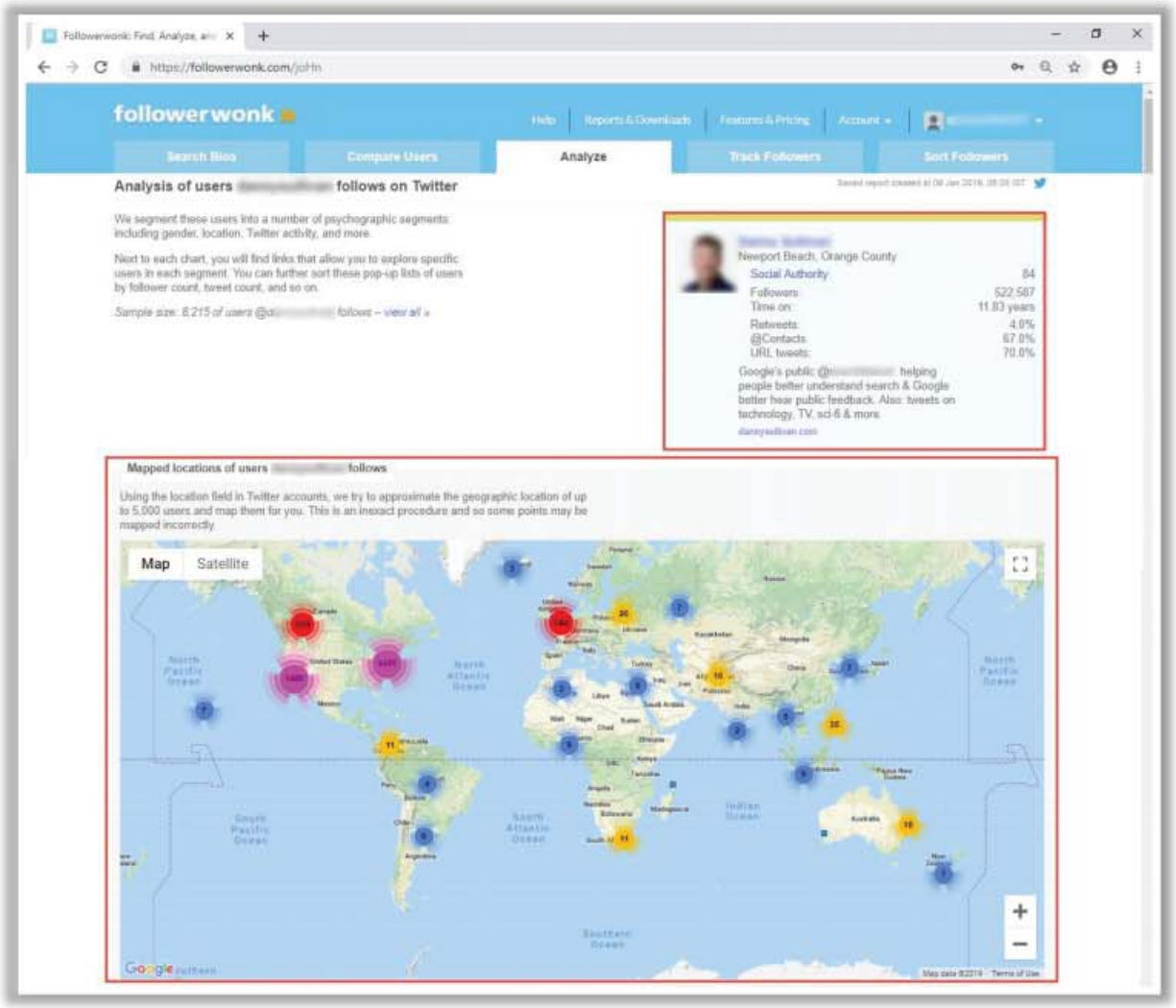


Figure 2.40: Screenshot of Followerwonk showing the geolocation

Tools for Footprinting through Social Networking Sites

Attackers use various tools such Sherlock, Social Searcher, and UserRecon to footprint social networking sites such as Twitter, Instagram, Facebook, and Pinterest to gather sensitive information about the target such as DOB, educational qualification, employment status, name of the 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 to locate the target user on various social networking sites along with the complete URL.

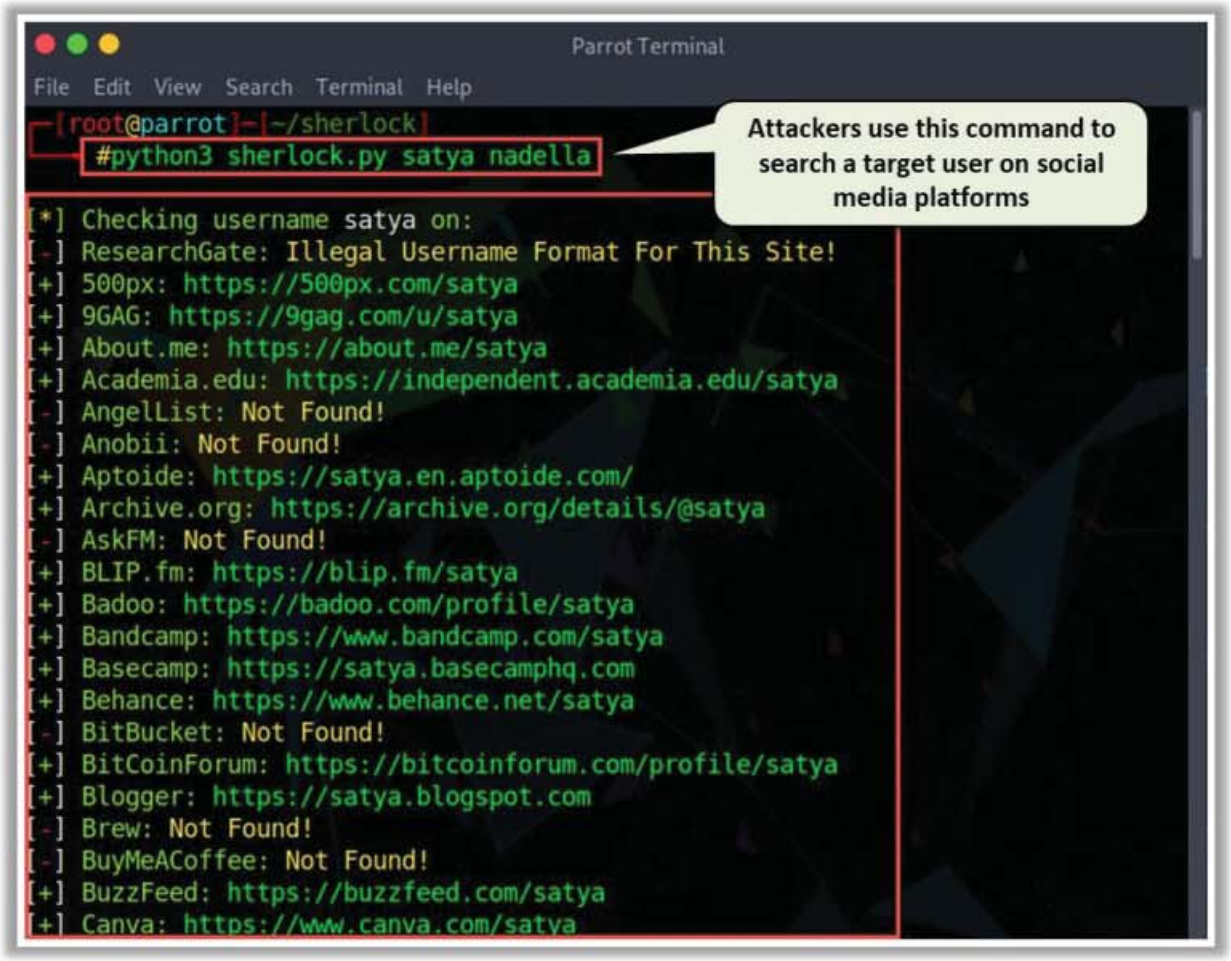


Figure 2.41: Screenshot showing the result of Sherlock tool

- **Social Searcher**

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

  Social Searcher allows attackers to search for content in 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.

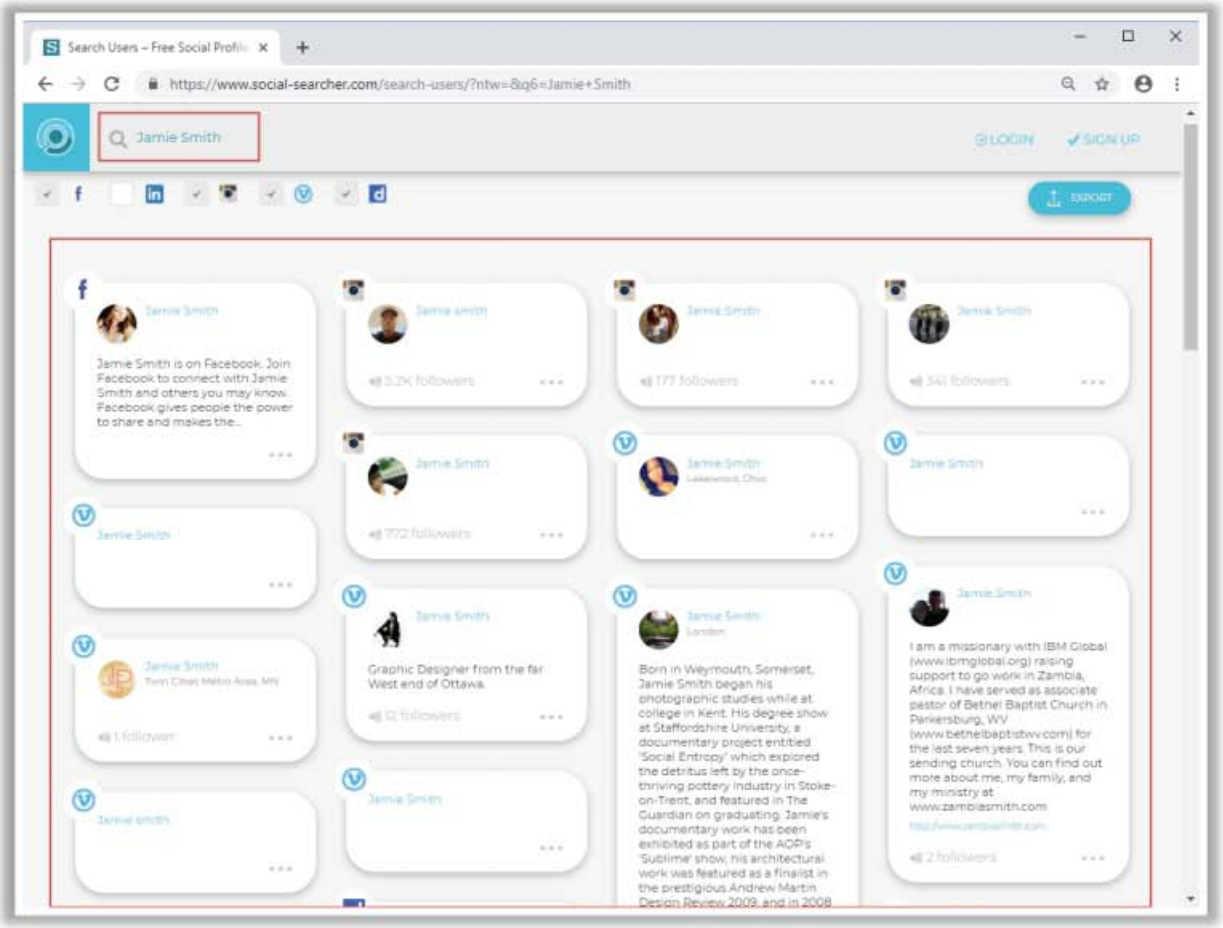
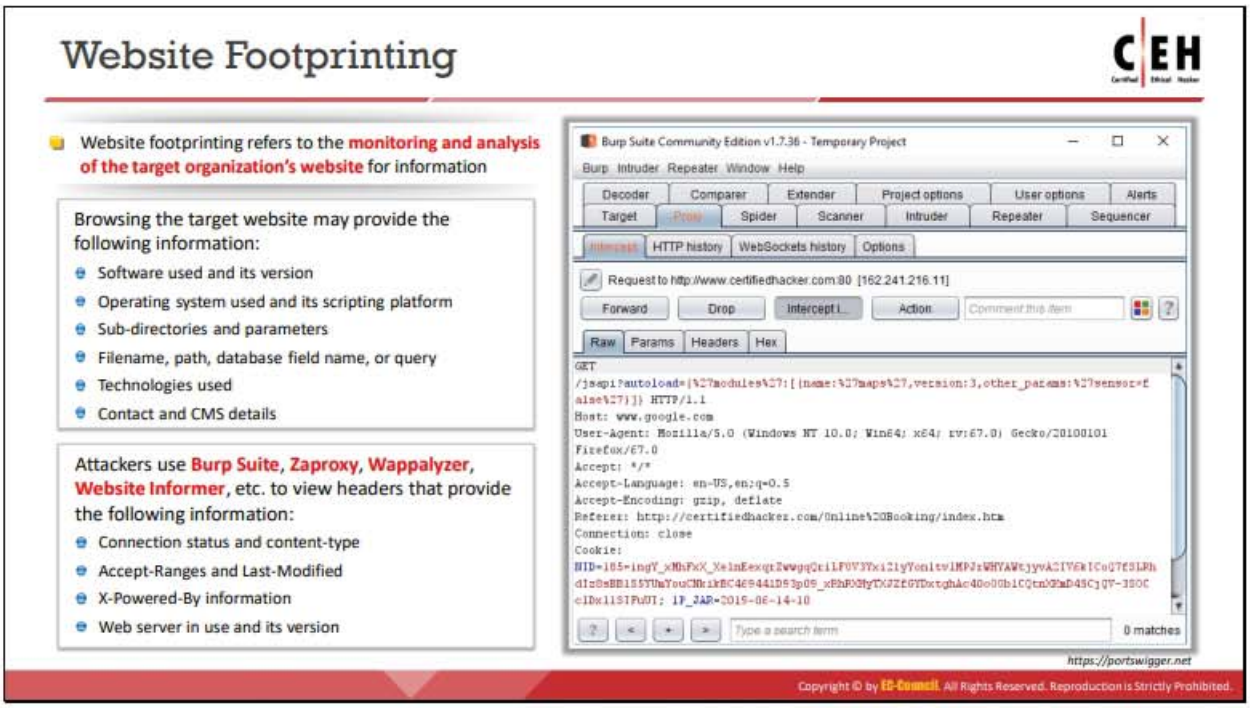
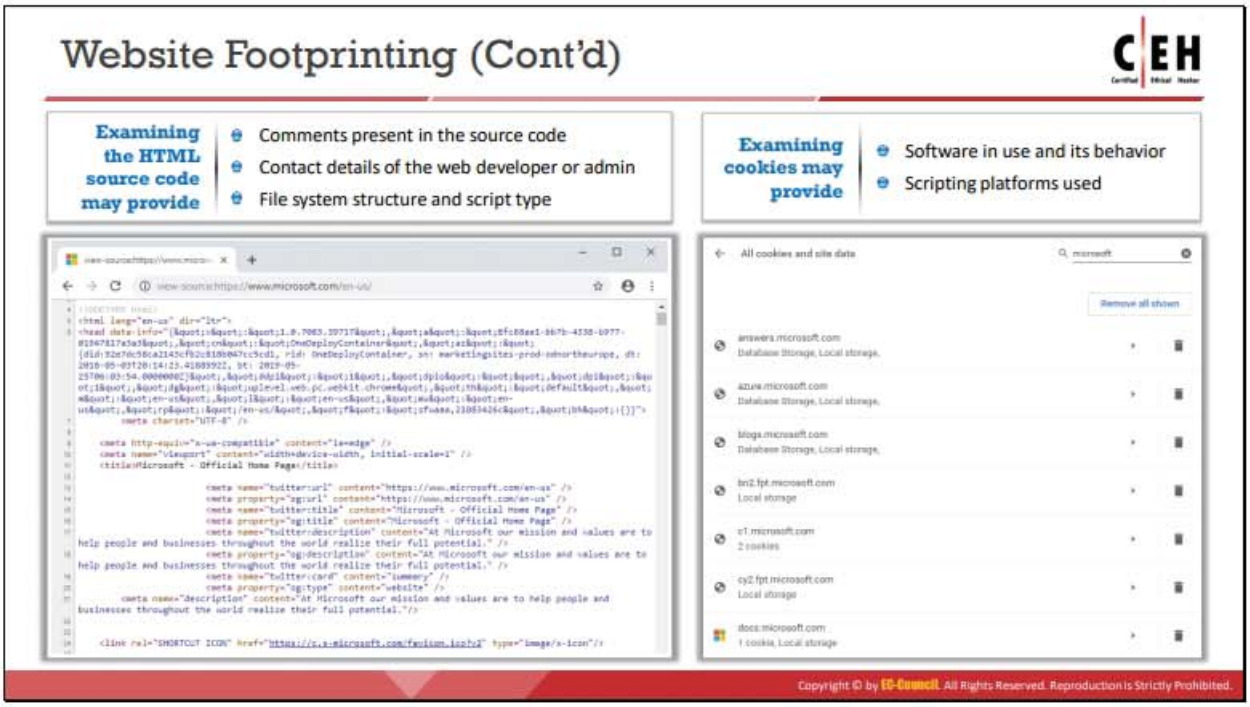


Figure 2.42: Screenshot of Social Searcher showing user content on social networks





Website Footprinting

So far, we have discussed footprinting through search engines, web services, and social networking sites. Hereafter, we will discuss website footprinting. An organization's website is the first place to get sensitive information such as names and contact details of the leaders of the organization, upcoming project details, and so on.

This section covers the website footprinting concept, mirroring websites, extracting website information and links, gathering wordlists, extracting metadata of public documents, and monitoring web updates and website traffic.

Website footprinting refers to monitoring and analyzing a target organization's website for information. An attacker can build a detailed map of a website's structure and architecture without triggering the IDS or arousing the suspicion of any system administrator. Attackers use sophisticated footprinting tools or the basic tools that come with the operating system, such as Telnet, or a browser.

The Netcraft tool can gather website information such as IP address, registered name and address of the domain owner, domain name, host of the site, and OS details. However, the tool may not give all these details for every site. In such cases, the attacker can browse the target website.

Browsing the target website will typically provide the following information:

- **Software used and its version**: An attacker can easily find the software and version in use on an off-the-shelf software-based website.
- **Operating system used**: Usually, the operating system in use can also be determined.
- **Sub-directories and parameters**: Searches can reveal the sub-directories and parameters by making a note of the URLs while browsing the target website.
- **Filename, path, database field name, or query**: The attacker will often carefully analyze anything after a query that looks like a filename, path, database field name, or query to check whether it offers opportunities for SQL injection.
- **Scripting platform**: With the help of script filename extensions such as .php, .asp, or .jsp, one can easily determine the scripting platform that the target website is using.
- **Technologies Used**: By inspecting the URLs of the target website, one can easily determine the technologies (.NET, J2EE, PHP, etc.) used to build that website.
- **Contact details and CMS details**: The contact pages usually offer details such as names, phone numbers, email addresses, and locations of admin or support personnel. An attacker can use these details to perform a social engineering attack. CMS software allows URL rewriting to disguise the script filename extensions if the attacker is willing to devote additional effort toward determining the scripting platform.

Attackers use Burp Suite, Zaproxy, WhatWeb, BuiltWith, Wappalyzer, and Website Informer to view headers that provide:

- Connection status and content type
- Accept-Ranges and Last-Modified information
- X-Powered-By information
- Web server in use and its version

Burp Suite

Source: *https://portswigger.net*

Burp Suite is an integrated platform for performing security testing of web applications. Its various tools work together to support the entire testing process, from initial mapping and analysis of an application's attack surface to finding and exploiting security vulnerabilities.

Burp Proxy allows attackers to intercept all requests and responses between the browser and the target web application and obtain information such as web server used, its version, and web-application-related vulnerabilities.

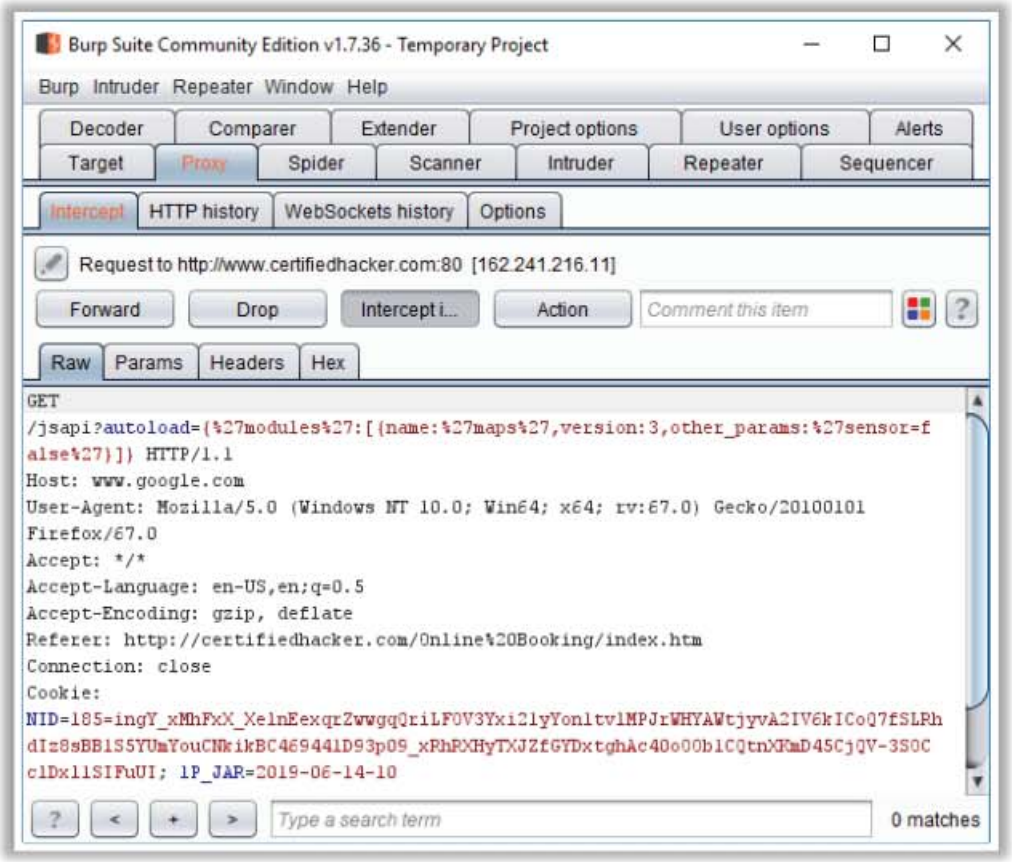


Figure 2.43: Screenshot of Burp Suite

Website footprinting can be performed by examining HTML source code and cookies.

- **Examining the HTML source code**

  Attackers can gather sensitive information by examining the HTML source code and following the comments that are inserted manually or those that the CMS system creates. The comments may provide clues as to what is running in the background. They may even provide contact details of the web developer or administrator.

  Observe all the links and image tags to map the file system structure. This will reveal the existence of hidden directories and files. Enter fake data to determine how the script works. It is sometimes possible to edit the source code.

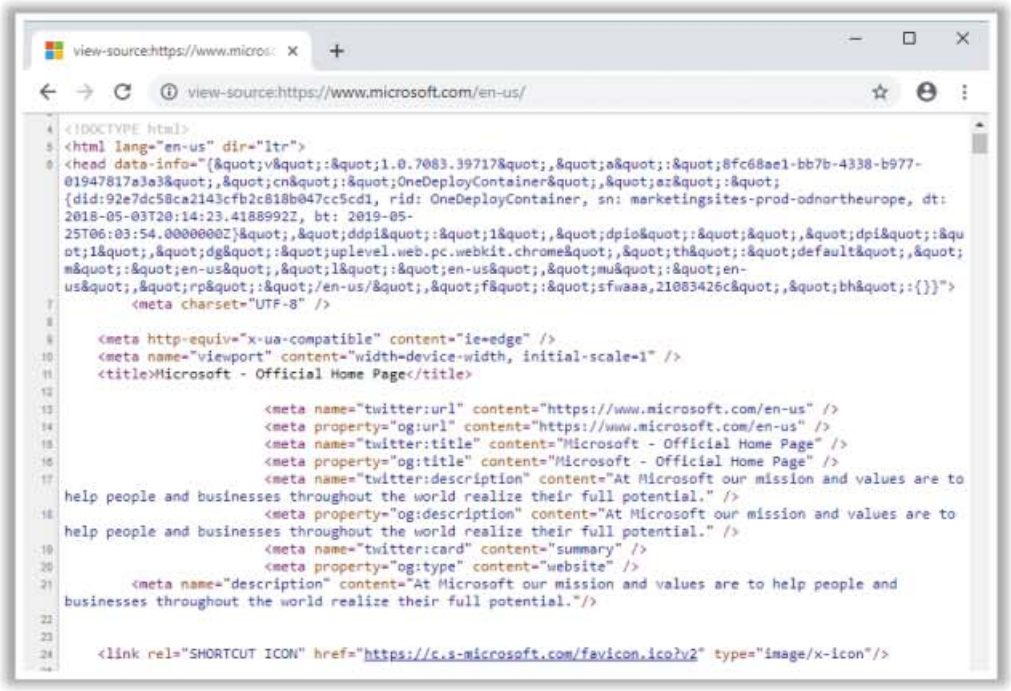


Figure 2.44: Screenshot showing HTML source code

- **Examining Cookies**

  To determine the software running and its behavior, one can examine cookies set by the server. Identify the scripting platforms by observing sessions and other supporting cookies. The information about cookie name, value, and domain size can also be extracted.

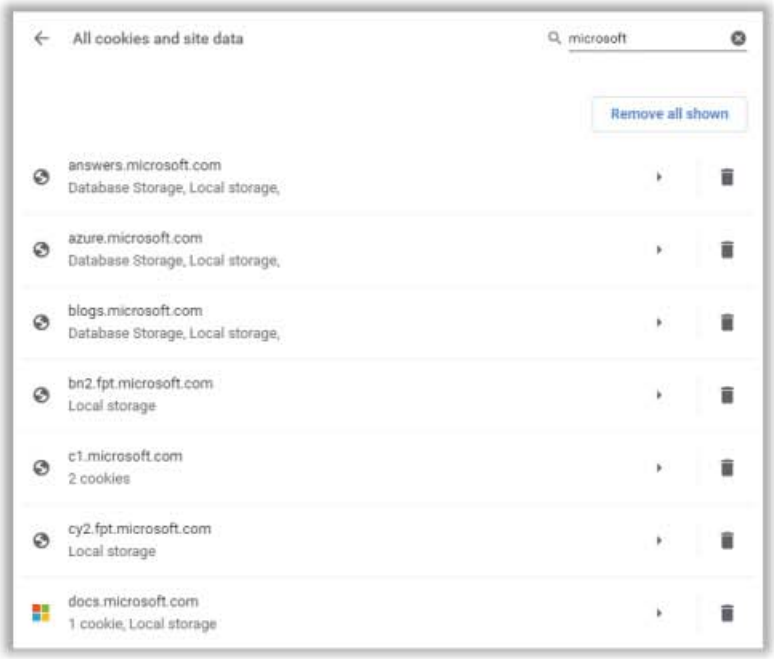
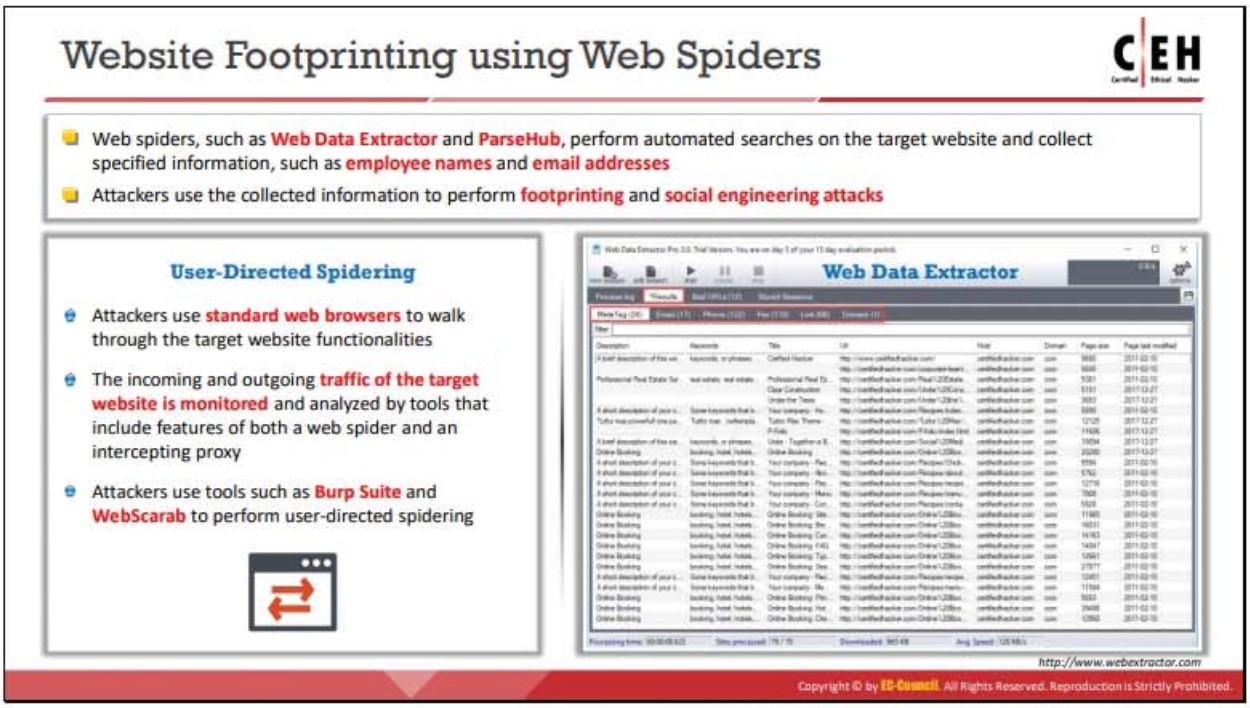


Figure 2.45: Screenshot showing cookies



Website Footprinting using Web Spiders

A web spider (also known as web crawler or web robot) is a program or automated script that browses websites in a methodical manner to collect specific information such as employee names and email addresses. Attackers then use the collected information to perform footprinting and social engineering attacks. Web spidering fails if the target website has the robots.txt file in its root directory with a listing of directories to prevent crawling.

Attackers can uncover all the files and web pages on the target website by simply feeding the web spider with a URL. Then, the web spider sends hundreds of requests to the target website and analyzes the HTML code of all the received responses for identifying additional links. If any new links are found, then the spider adds them to the target list and starts spidering and analyzing the newly discovered links. This method helps attackers to not only detect exploitable web-attack surfaces but also to find all the directories, web pages, and files that make up the target website.

User-Directed Spidering

Attackers, in some cases, use a more sophisticated technique for spidering the target website instead of using automated tools. They use standard web browsers to walk through the target website in an attempt to navigate through all the functionalities provided by the web application. While performing this task, the resulting incoming and outgoing traffic of the website is monitored and analyzed by the tools that include features of both a web spider and an intercepting proxy. Further, these tools create a map of the web application consisting of all the URLs visited by the browser. It also analyzes the responses of the application and updates the map with the discovered content and its functionalities. Attackers use tools such as Burp Suite and WebScarab to perform user-directed spidering.

Web spidering tools such as Web Data Extractor, ParseHub, and SpiderFoot can collect sensitive information from the target website.

- **Web Data Extractor**

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

  Web Data Extractor automatically extracts specific information from web pages. It extracts targeted contact data (email, phone, and fax) from the website, extracts the URL and meta tags (title, description, keyword) for website promotion, searches directory creation, performs web research, and so on.

  As shown in the screenshot, attackers use Web Data Extractor to automatically gather critical information such as lists of meta tags, e-mail addresses, and phone and fax numbers from the target website.

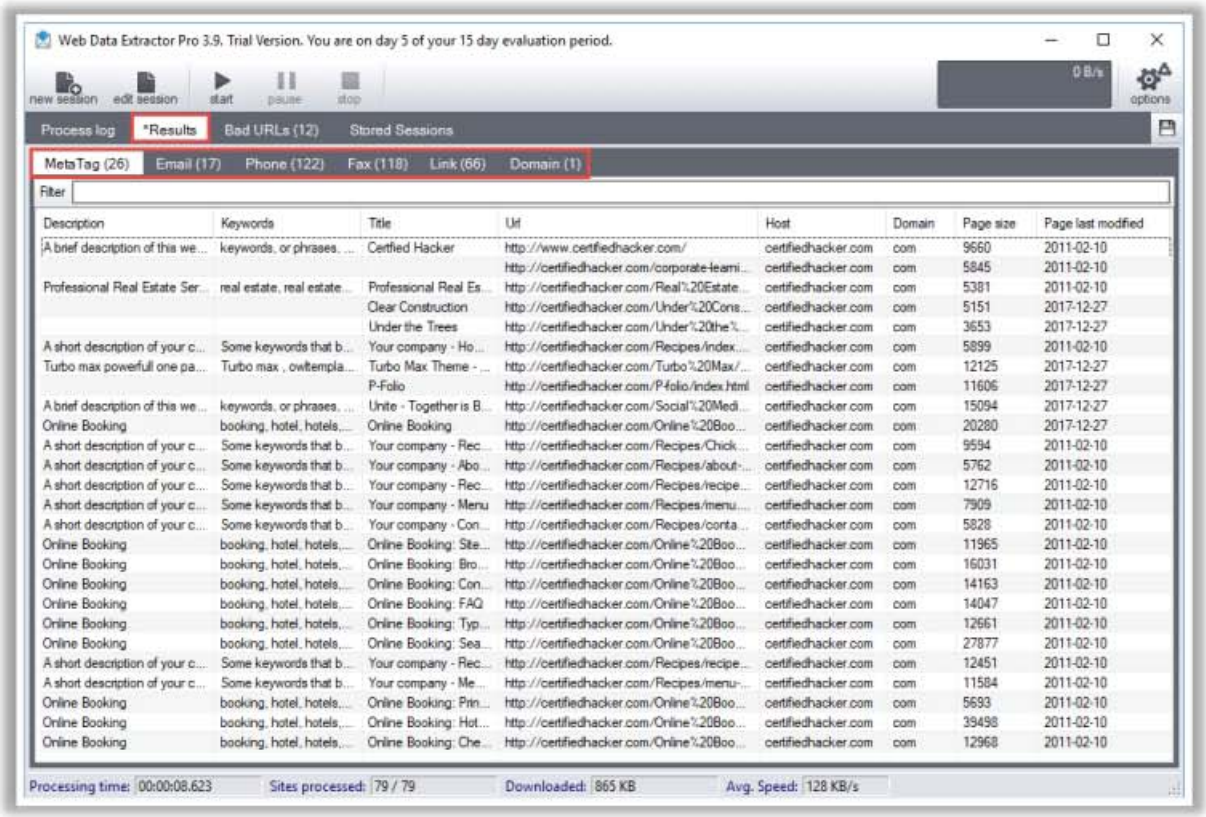
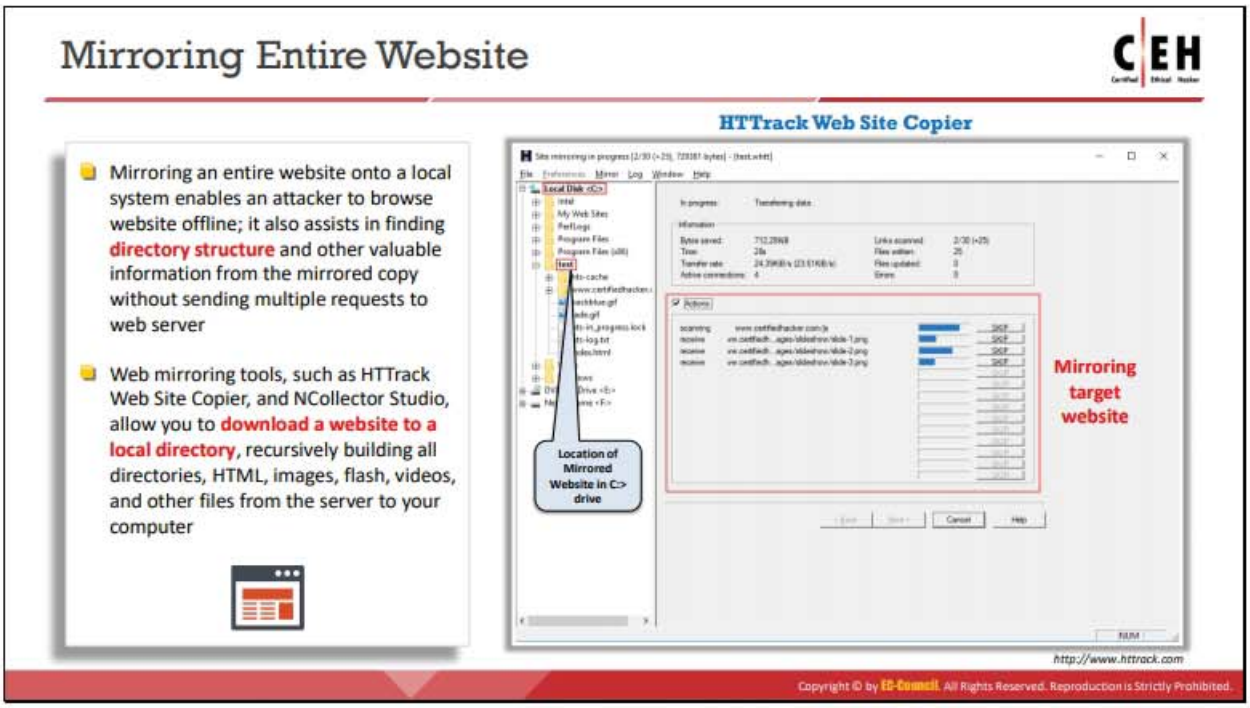


Figure 2.46: Screenshot of Web Data Extractor



Mirroring Entire Website

Website mirroring is the process of creating a replica or clone of the original website. Users can duplicate websites using mirroring tools such as HTTrack Web Site Copier and NCollector Studio. These tools download a website to a local directory and recursively build all the directories including HTML, images, flash, videos, and other files from the webserver on another computer.

Website mirroring has the following benefits:

- It is helpful for offline site browsing
- It enables an attacker to spend more time in viewing and analyzing the website for vulnerabilities and loopholes
- It helps in finding the directory structure and other valuable information from the mirrored copy without multiple requests to the webserver

Attackers can use this information to perform various web application attacks on the target organization's website.

Website Mirroring Tool: HTTrack Web Site Copier

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

HTTrack is an offline browser utility. It downloads a website from the Internet to a local directory and recursively builds all the directories including HTML, images, and other files from the web server on another computer.

As shown in the screenshot, attackers use HTTrack to mirror the entire website of the target organization, store it in the local system drive, and browse the local website to identify possible exploits and vulnerabilities.

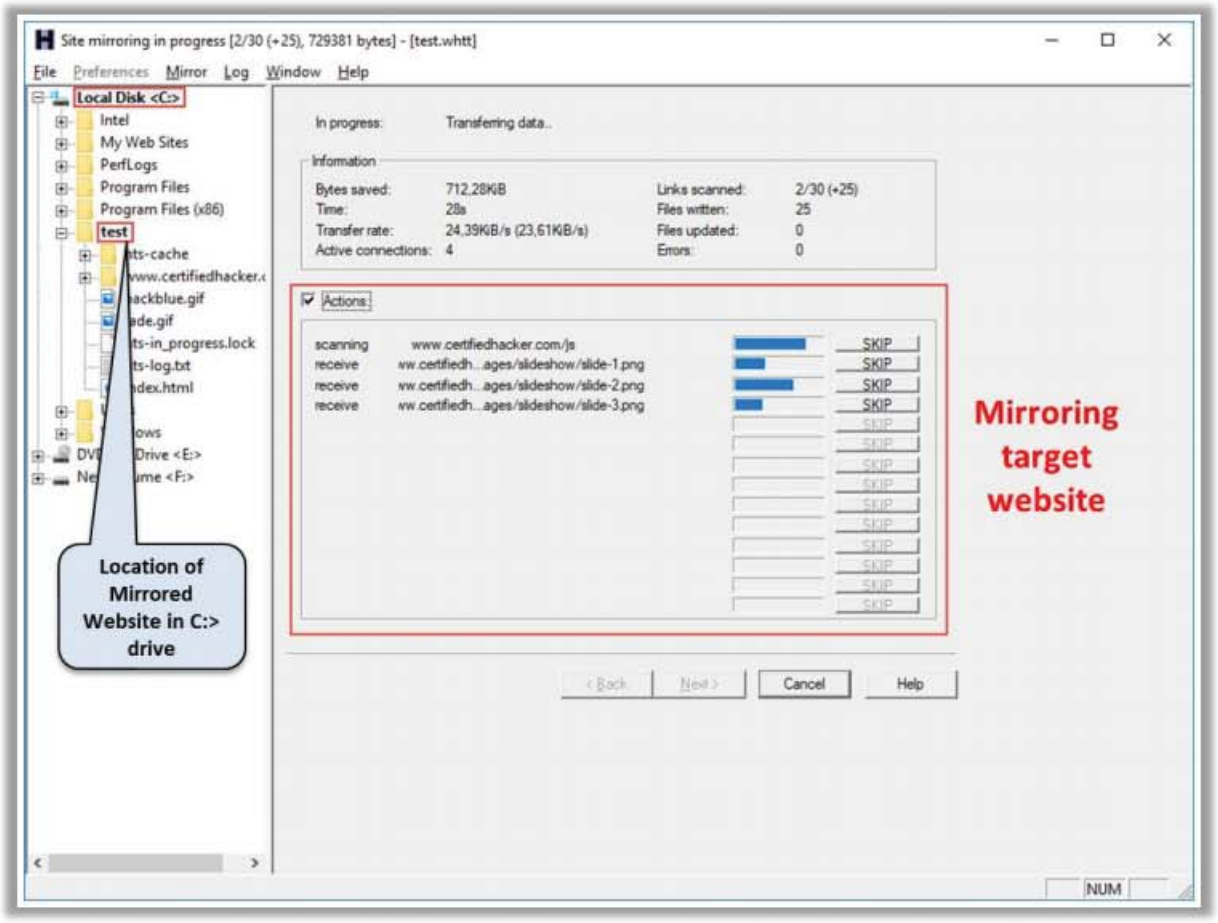
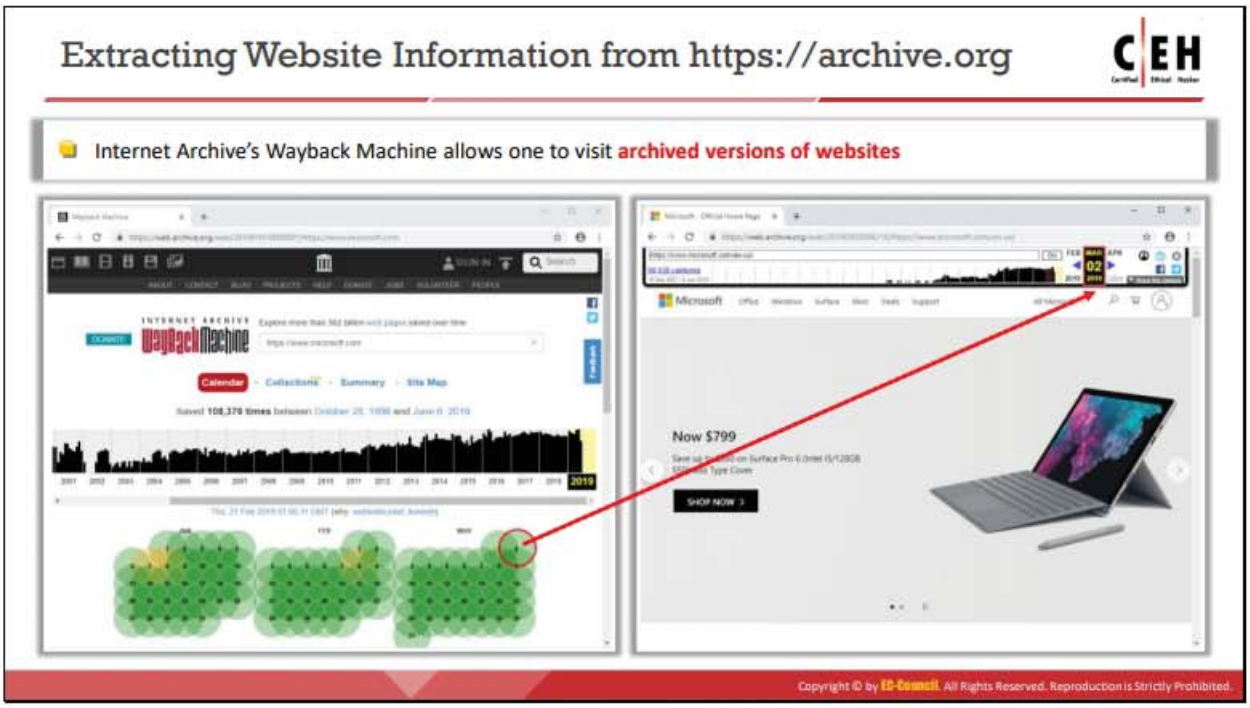


Figure 2.47: Screenshot of HTTrack Web Site Copier



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.

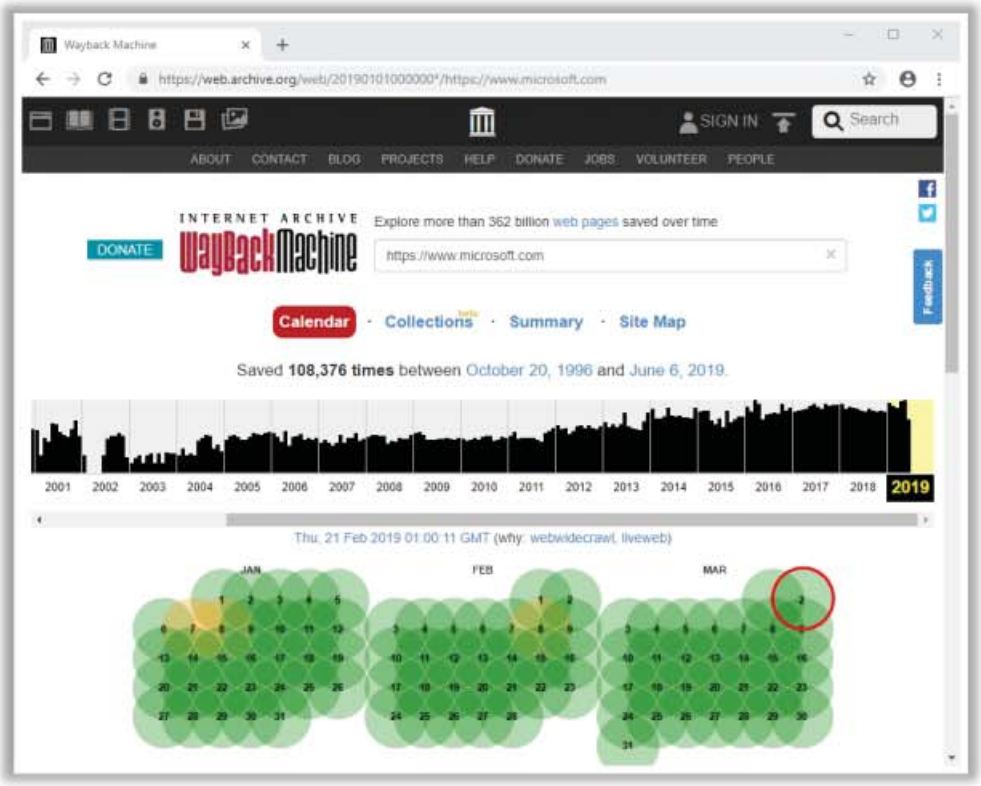


Figure 2.48: Screenshot of Archive showing archived versions of microsoft.com

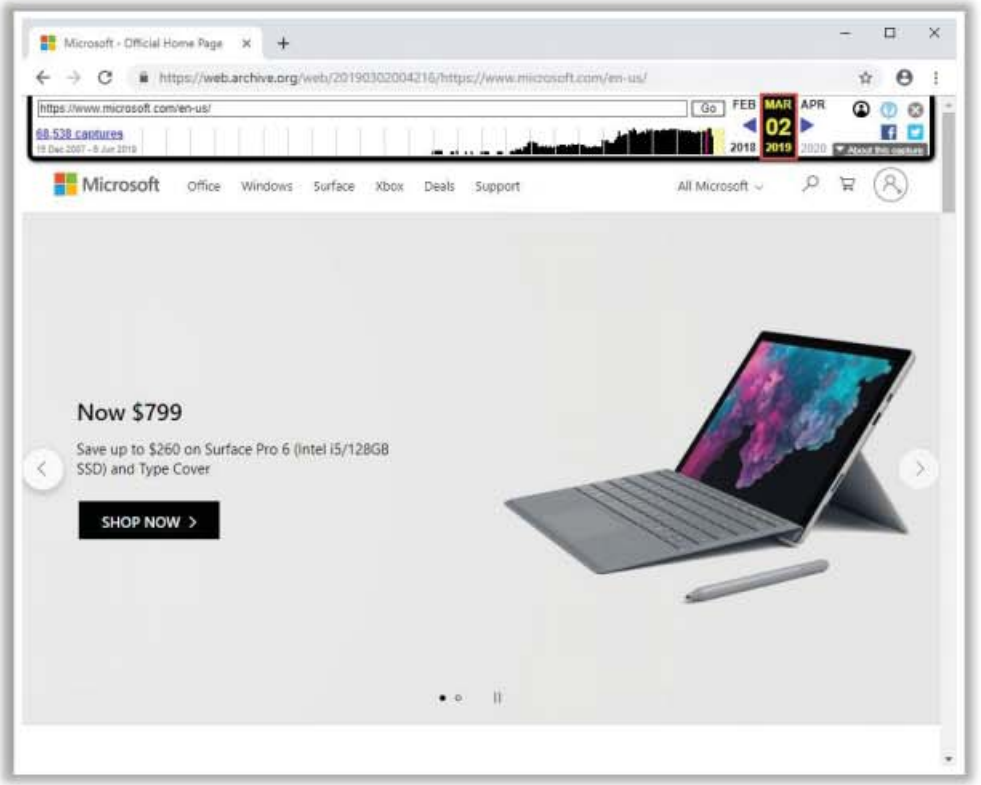
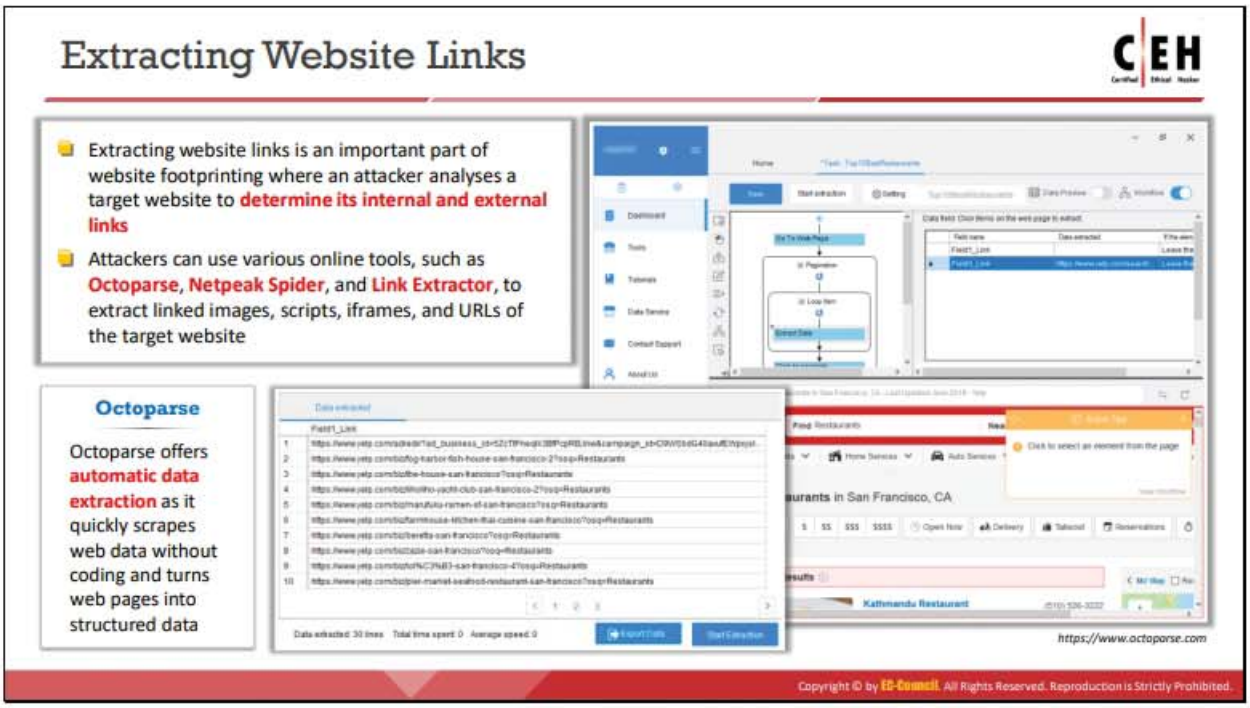


Figure 2.49: Screenshot of Archive showing archived web pages of microsoft.com



Extracting Website Links

Extracting website links is an important part of website footprinting, where an attacker analyzes a target website to determine its internal and external links. Using the gathered information, an attacker can find out the applications, web technologies, and other related websites that are linked to the target website. Further, dumping the obtained links can reveal important connections and extract URLs of other resources such as JavaScript and CSS files. This information helps attackers to identify vulnerabilities in the target website and find ways to exploit the web application.

Attackers can use various online tools or services such as Octoparse, Netpeak Spider, and Link Extractor to extract linked images, scripts, iframes, URLs, etc., of the target website. Using these tools, an attacker can also extract backlinks to a target website, which can provide important and useful information about the target to perform further exploitation.

- **Octoparse**

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

  Octoparse offers automatic data extraction, as it quickly scrapes web data without coding and turns web pages into structured data. As shown in the screenshot, attackers use Octoparse to capture information from webpages, such as text, links, image URLs, or html code.

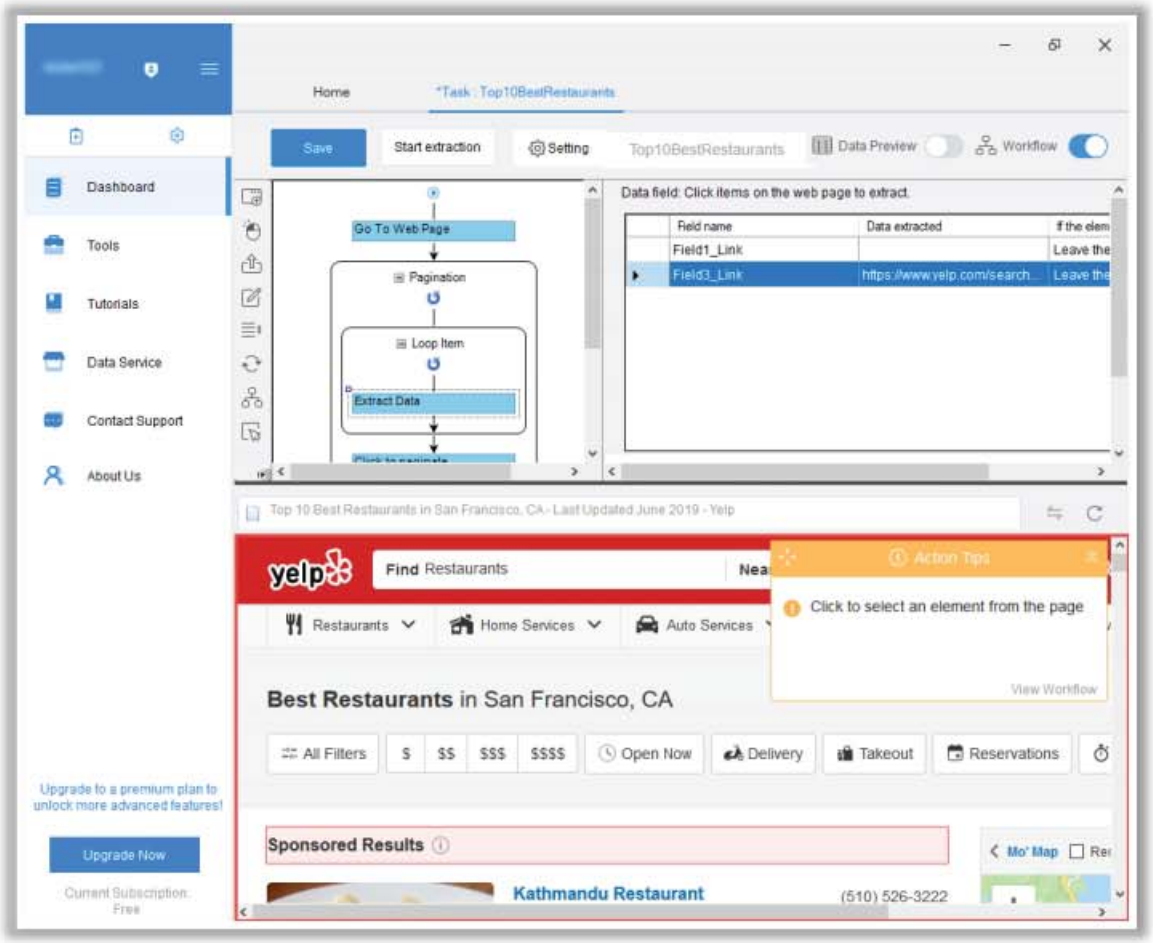


Figure 2.50: Screenshot of Octoparse

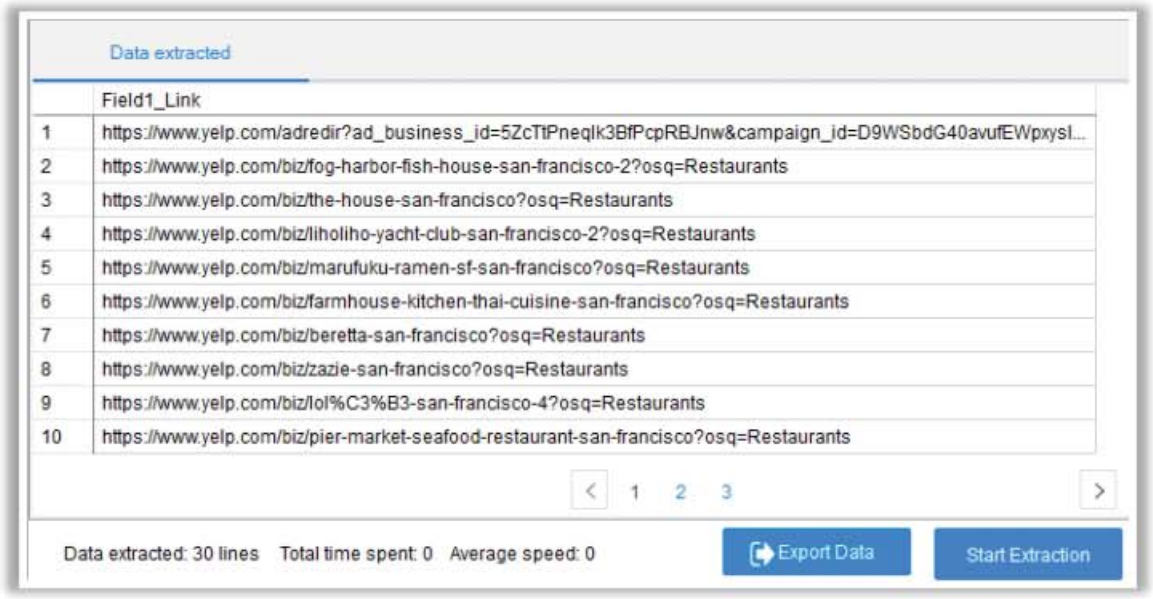


Figure 2.51: Screenshot showing output of Octoparse

Gathering Wordlist from the Target Website

The words available on the target website may reveal critical information that helps attackers to perform further exploitation. Attackers gather a list of email addresses related to the target organization using various search engines, social networking sites, web spidering tools, etc. After obtaining these email addresses, an attacker can gather a list of words available on the target website. This information helps the attacker to perform brute-force attacks on the target organization. An attacker uses the CeWL tool to gather a list of words from the target website and perform a brute-force attack on the email addresses gathered earlier.

To run the CeWL tool, issue the following commands:

- **`ruby cewl.rb --help`**

  This command displays various options that a user can use to obtain a list of words from the target website.

- **`cewl www.certifiedhacker.com`**

  This command returns a list of unique words present in the target URL.

- **`cewl --email www.certifiedhacker.com`**

  In this case, the target website is **`www.certifiedhacker.com`**, and the **`'--email'`** option is used to fetch a list of words and email addresses from the target website.

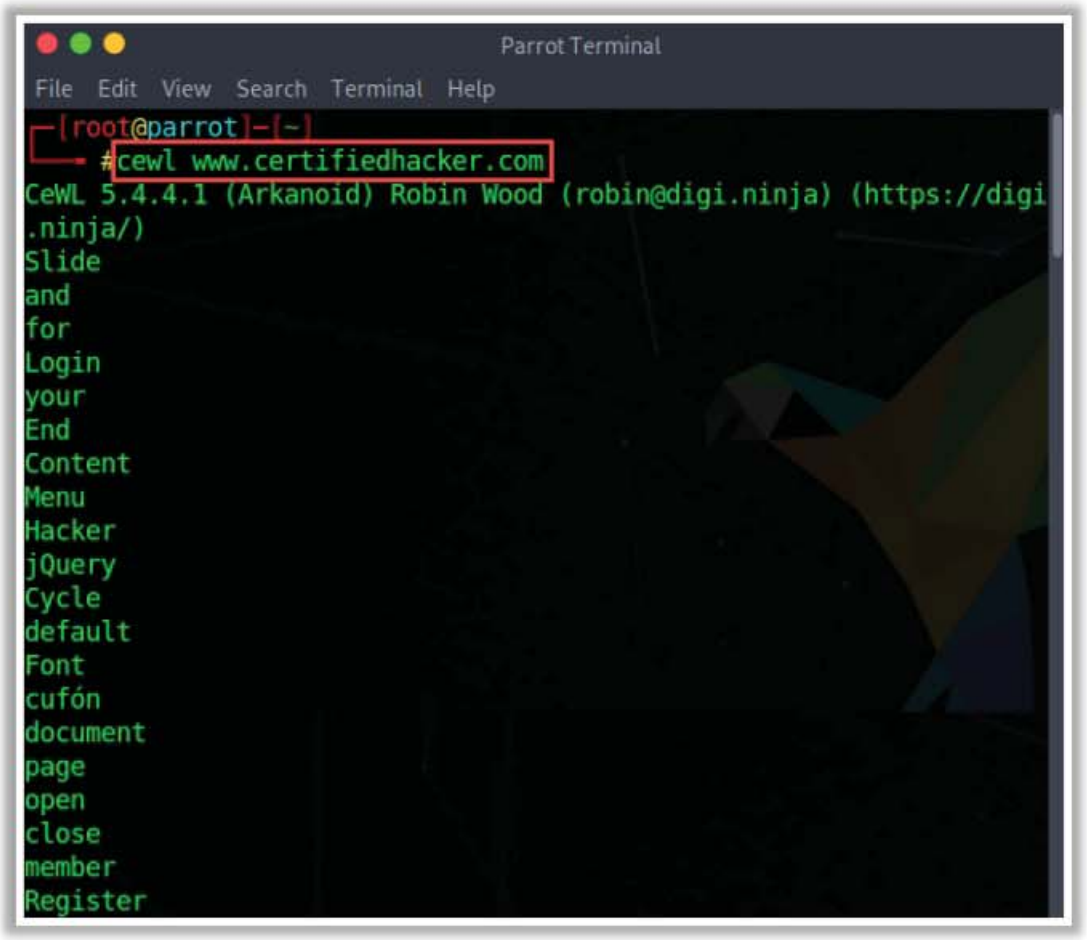
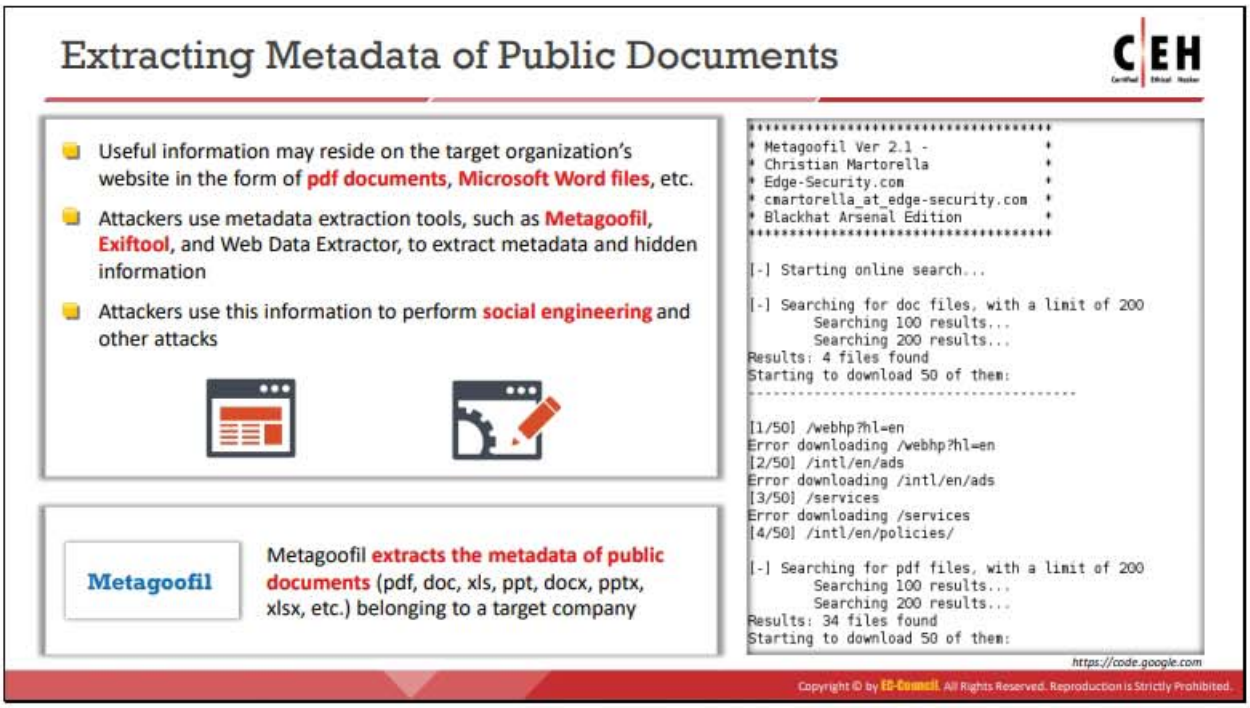


Figure 2.52: Screenshot showing results obtained from CeWL tool



Extracting Metadata of Public Documents

Useful information may reside on the target organization's website in the form of pdf documents, Microsoft Word files, and other files in various formats. Attackers extract valuable data, including metadata and hidden information from such documents. The data mainly contains hidden information about the public documents that can be analyzed to extract information such as the title of the page, description, keywords, creation/modification date and time of the content, and usernames and e-mail addresses of employees of the target organization.

An attacker can misuse this information to perform malicious activities against the target organization by brute-forcing authentication using the usernames and e-mail addresses of employees, or perform social engineering to send malware, which can infect the target system.

Metadata Extraction Tools

Metadata extraction tools such as Metagoofil, Exiftool, and Web Data Extractor automatically extract critical information that includes the usernames of clients, operating systems (exploits are OS-specific), email addresses (possibly for social engineering), list of software (version and type) used, list of servers, document date creation/modification, and authors of the website.

- **Metagoofil**

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

  Metagoofil extracts metadata of public documents (pdf, doc, xls, ppt, docx, pptx, and xlsx) belonging to a target company. It performs a Google search to identify and download the documents to the local disk and then extracts the metadata with different libraries such as Hachoir, PdfMiner, and others.

As shown in the screenshot, Metagoofil generates a report with usernames, software versions, and servers or machine names, which helps attackers in the information gathering phase.

```
************************************
* Metagoofil Ver 2.1 -             *
* Christian Martorella             *
* Edge-Security.com                *
* cmartorella_at_edge-security.com  *
* Blackhat Arsenal Edition         *
************************************

[-] Starting online search...

[-] Searching for doc files, with a limit of 200
        Searching 100 results...
        Searching 200 results...
Results: 4 files found
Starting to download 50 of them:
----------------------------------------

[1/50] /webhp?hl=en
Error downloading /webhp?hl=en
[2/50] /intl/en/ads
Error downloading /intl/en/ads
[3/50] /services
Error downloading /services
[4/50] /intl/en/policies/

[-] Searching for pdf files, with a limit of 200
        Searching 100 results...
        Searching 200 results...
Results: 34 files found
Starting to download 50 of them:
```

Figure 2.53: Screenshot of Metagoofil

Other Techniques for Website Footprinting

- **Monitoring Web Pages for Updates and Changes**

  Attackers monitor the target website to detect web updates and changes. Monitoring the target website helps attackers to access and identify changes in the login pages, extract password-protected pages, track changes in the software version and driver updates, extract and store images on the modified web pages, and so on. Attackers analyze the gathered information to detect underlying vulnerabilities in the target website, and based on these vulnerabilities, they perform exploitation of the target web application.

  Web Updates Monitoring Tools

  Web updates monitoring tools are capable of detecting any changes or updates on a particular website, and they can send notifications or alerts to interested users through email or SMS.

  - **WebSite-Watcher**

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

    WebSite-Watcher helps to track websites for updates and automatic changes. When an update or change occurs, WebSite-Watcher automatically detects and saves the last two versions onto your disk.

    As shown in the screenshot, attackers use WebSite-Watcher to extract the older and newer versions of web pages related to the target website.

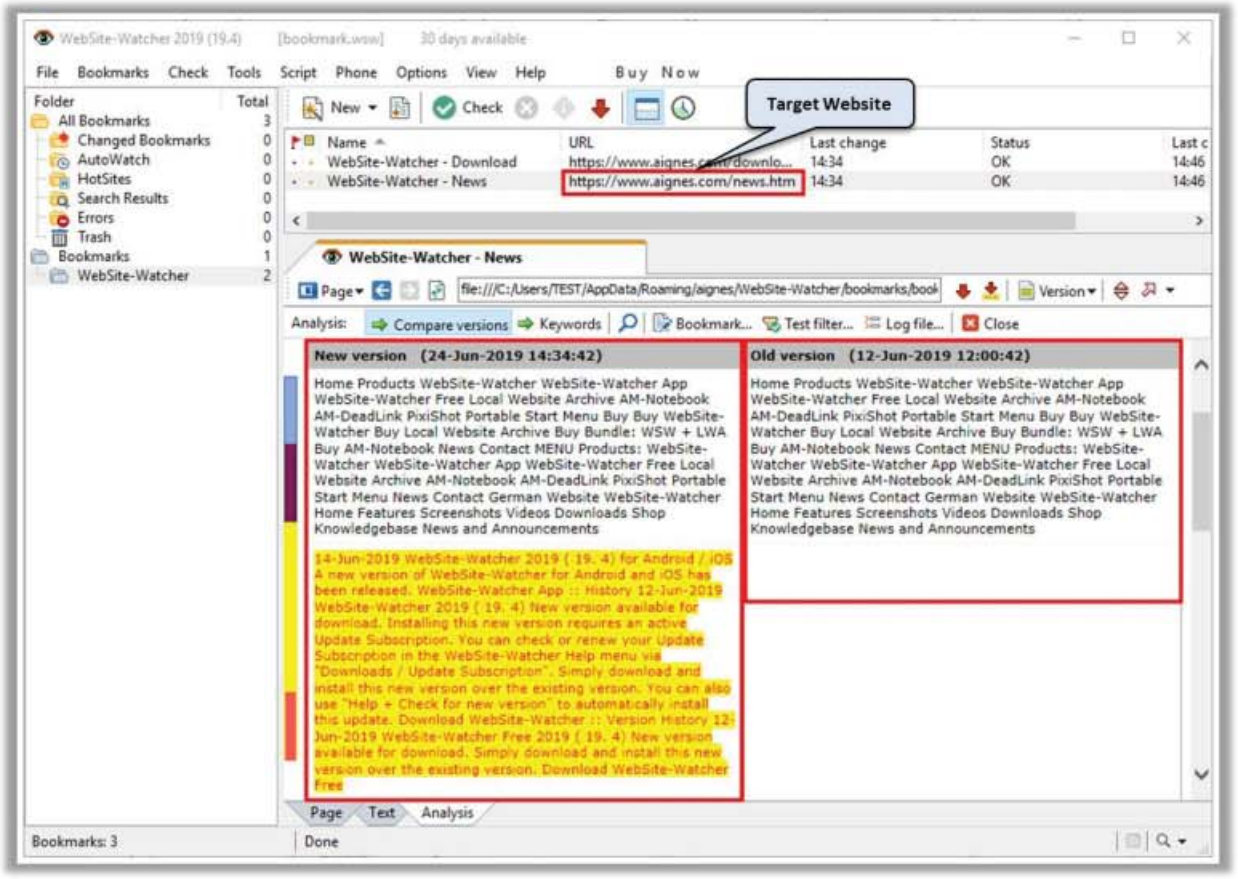


Figure 2.54: Screenshot of WebSite-Watcher

- **Searching for Contact Information, Email Addresses, and Telephone Numbers from Company Website**

  Attackers can search the target company's website to gather crucial information about the company. Generally, organizations use websites to inform the public about what they do, what type of services or products they provide, how to contact them, etc. Attackers can exploit this information to launch further attacks on the target company.

  For example, attackers can search for the following information on the company's website:

  - Company contact names, phone numbers, and email addresses
  - Company locations and branches
  - Partner Information
  - News
  - Links to other sites
  - Product, project, or service data

- **Searching for Web Pages Posting Patterns and Revision Numbers**

  Copyright is a protecting mechanism provided by the law of a country, which grants the creator of an original work exclusive rights for its use and distribution. To restrict third parties from accessing their data freely, most organizations ensure that there is a copyright notice on every single piece of their published work.

  A typical copyright notice contains the following information:

  - The Copyright Symbol
  - The Year of Creation
  - The Name of the Author
  - A Rights Statement

  An attacker can search for copyright notices on the web and use these details to perform a deep analysis of the target organization. Further, attackers can search and note down the revision number of a product. The revision number is a unique string that acts as an identifier for the revision of a given document, and it can be found within the documents of the company.

  Attackers can also search for the document numbers that are assigned to the documents after revision, which can be searched from the Internet and recorded to launch further attacks on the target.

- **Monitoring Website Traffic of Target Company**

  Attackers can monitor a target company's website traffic using tools such as Web-Stat, Alexa, and Monitis to collect valuable information. These tools help to collect information about the target's customer base, which help attackers to disguise themselves as customers and launch social engineering attacks on the target.

  The information collected includes:

  - **Total visitors**: Tools such as Clicky (*https://clicky.com*) find the total number of visitors browsing the target website.
  - **Page views**: Tools such as Opentracker (*https://www.opentracker.net*) monitor the total number of pages viewed by the users along with the timestamps and the status of the user on a particular web page (whether the webpage is still active or closed).
  - **Bounce rate**: Tools such as Google Analytics (*https://analytics.google.com*) measure the bounce rate of the target company's website.
  - **Live visitors map**: Tools such as Web-Stat (*https://www.web-stat.com*) track the geographical location of the users visiting the company's website.
  - **Site ranking**: Tools such as Alexa (*https://www.alexa.com*) track a company's rank on the web.
  - **Audience geography**: Tools such as Alexa track a company's customer locations on the globe.

- **Track Visitors and monitor sales**: Tools such as goingup! (*https://goingup.com*) track visitors, monitor sales, and show conversation rates with the company's website.

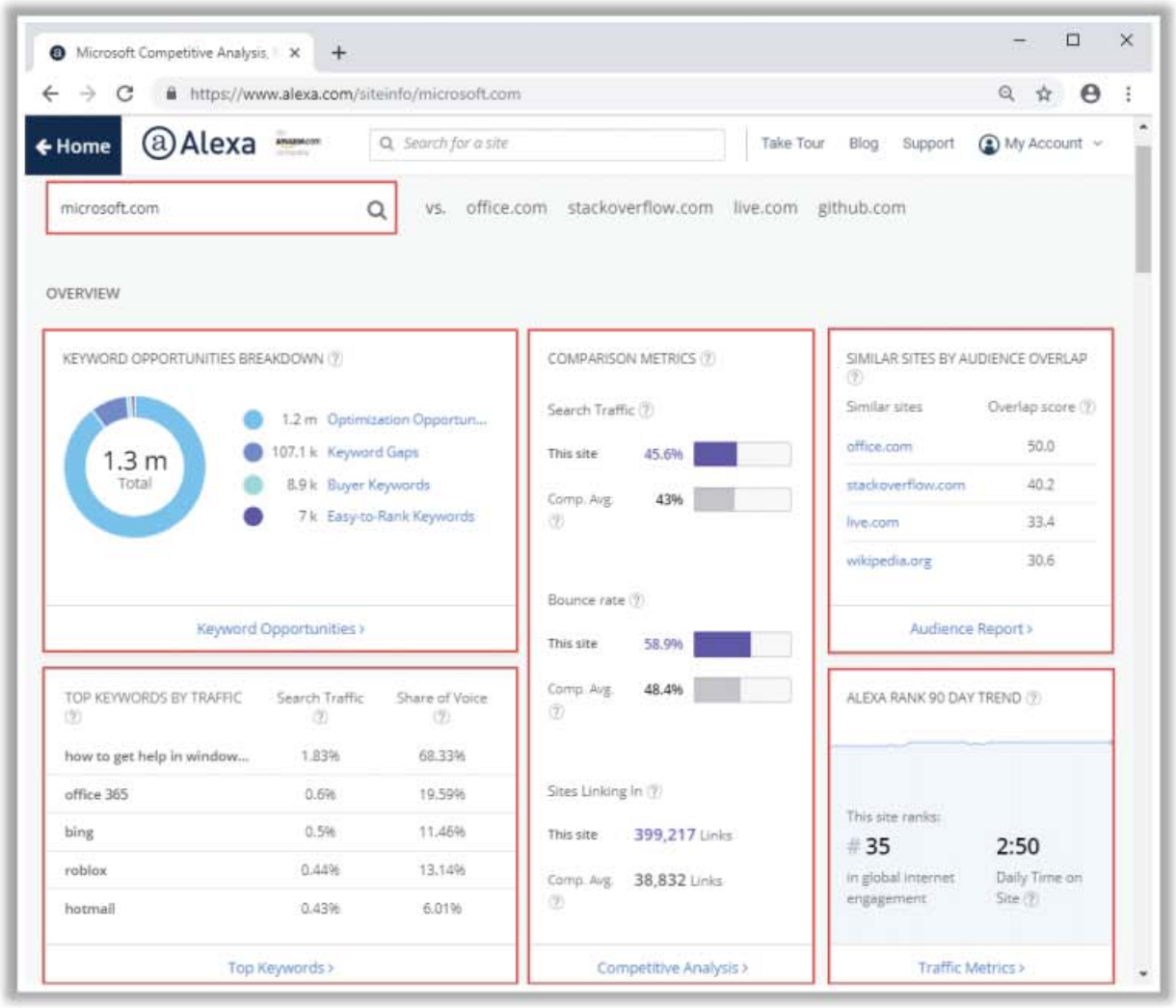
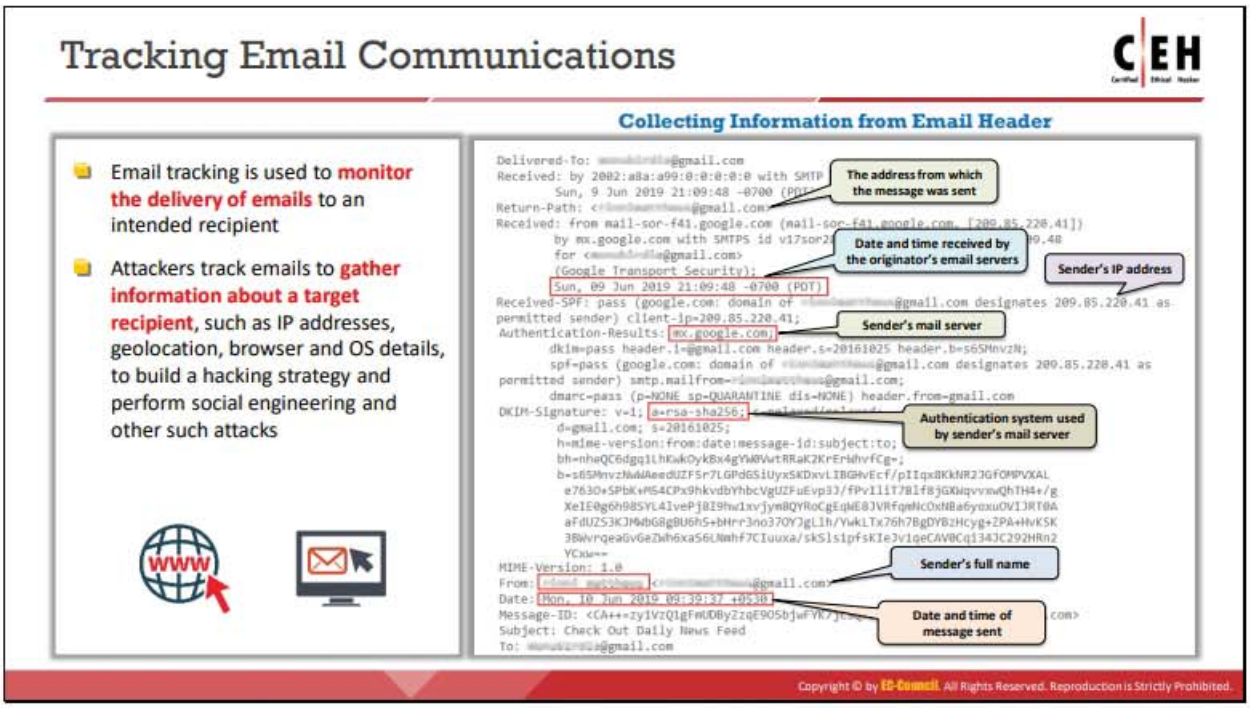


Figure 2.55: Screenshot of Alexa



Email Footprinting

So far, we have discussed footprinting through search engines, footprinting using Google, footprinting through social networking sites, and website footprinting. Now, we will discuss email footprinting. This section describes how to track email communications, how to collect information from email headers, and email tracking tools.

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 eMailTrackerPro, Infoga, and Mailtrack.

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 Lite
- Hiri
- Mozilla Thunderbird
- Spike
- Claws Mail
- SmarterMail Webmail
- Outlook

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.

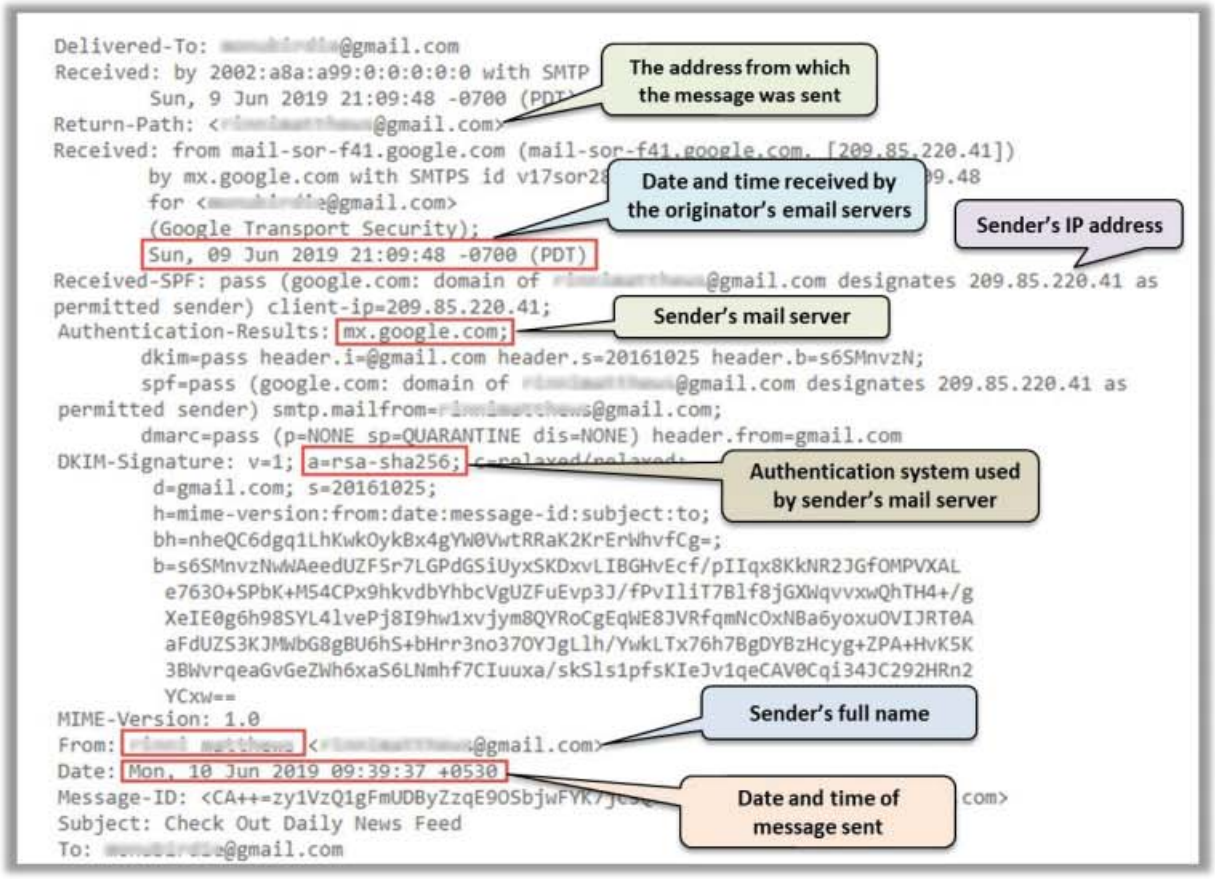


Figure 2.56: Screenshot showing detailed analysis of the email header

Email Tracking Tools

Email tracking tools allow an attacker to track an email and extract information such as sender identity, mail server, sender's IP address, location, and so on. These tools send notifications automatically when the recipients open the mail and provide status information about whether the email was successfully delivered or not. Attackers use the extracted information to attack the target organization's systems by sending malicious emails.

- **Infoga**

  Source: *https://github.com*

  Infoga is a tool used for gathering email account information (IP, hostname, country, etc.) from different public sources (search engines, pgp key servers, and Shodan), and it checks if an email was leaked using the haveibeenpwned.com API. For example, the command

  ```
  python infoga.py --domain microsoft.com --source all --breach -v
  2 --report ../microsoft.txt
  ```

  will retrieve all the publicly available email addresses related to the domain microsoft.com along with email account information.

  ```
  python infoga.py --info m4ll0k@protonmail.com --breach -v 3 --
  report ../m4ll0k.txt
  ```

  The above command will retrieve email account information for a specified email address.

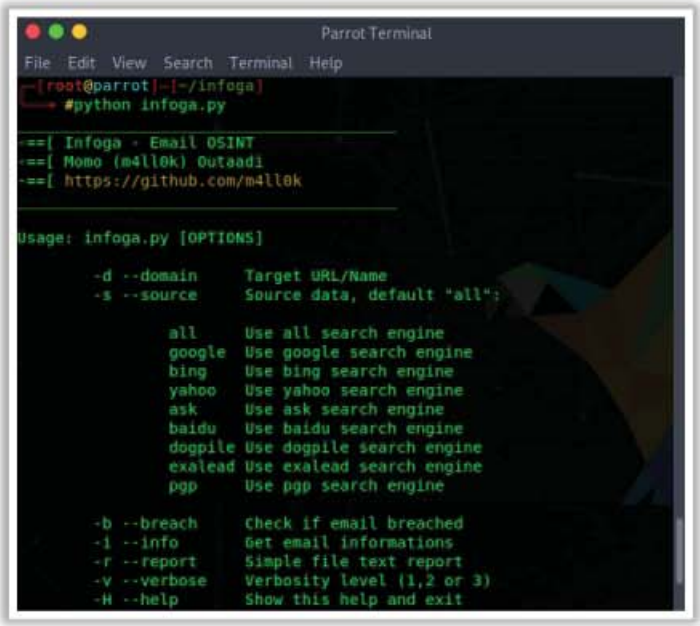


Figure 2.57: Screenshot showing various options of Infoga

- **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.

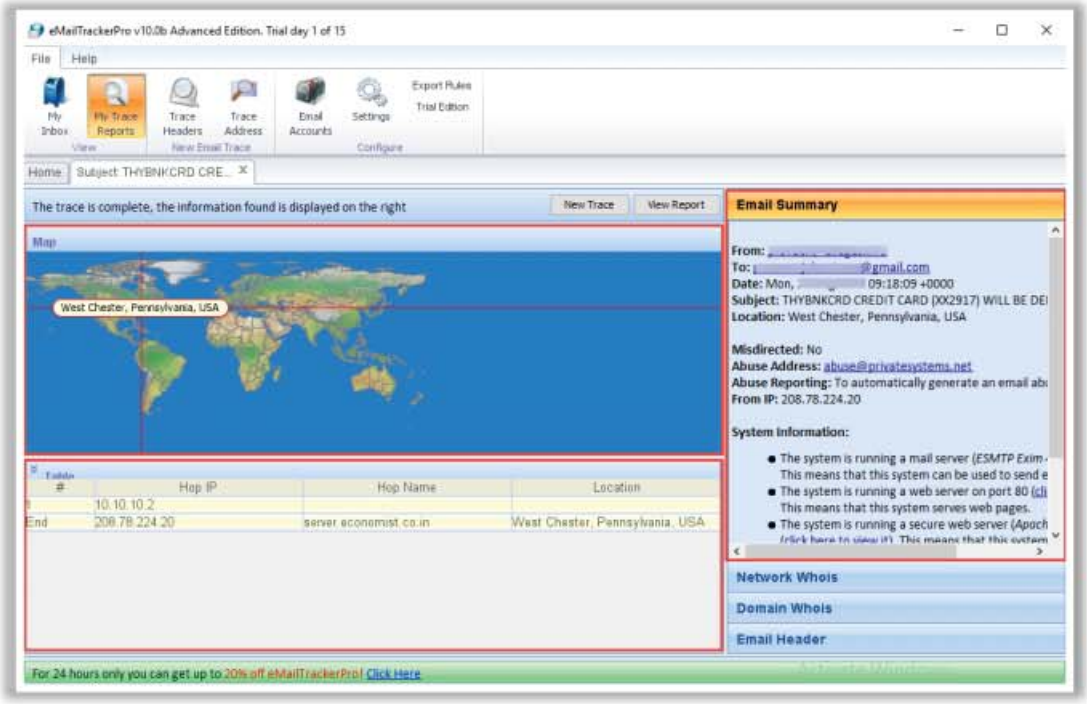
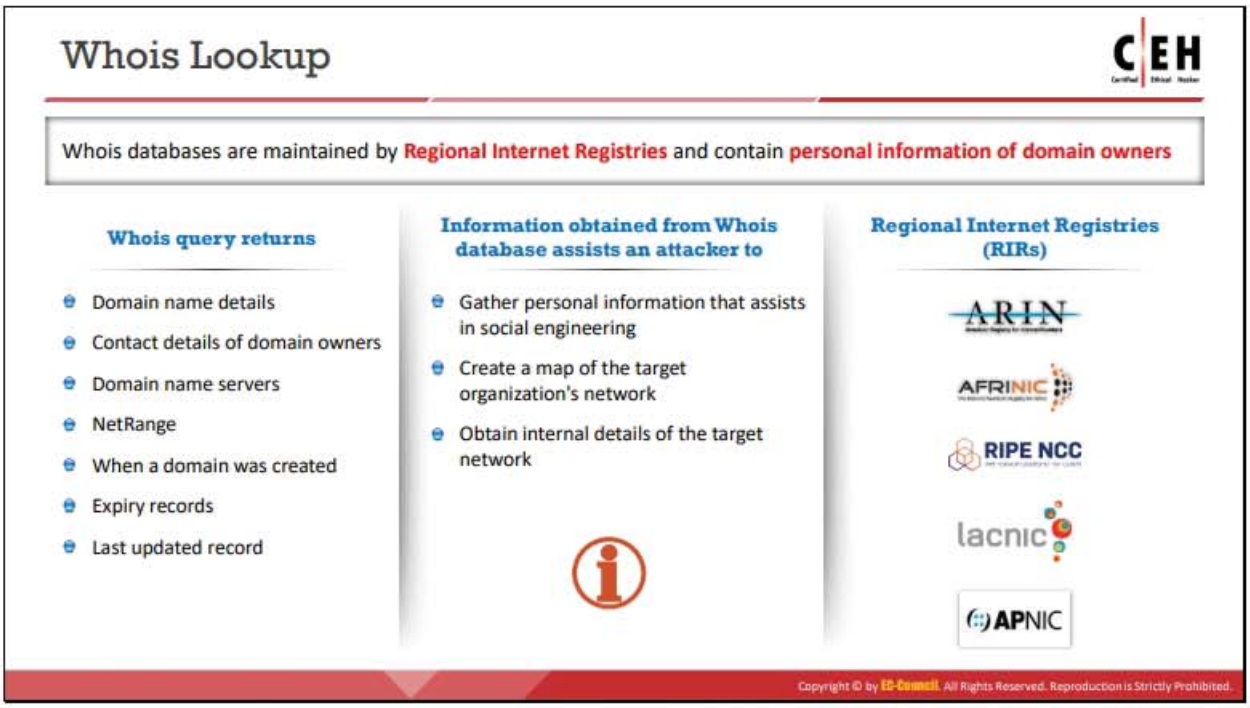
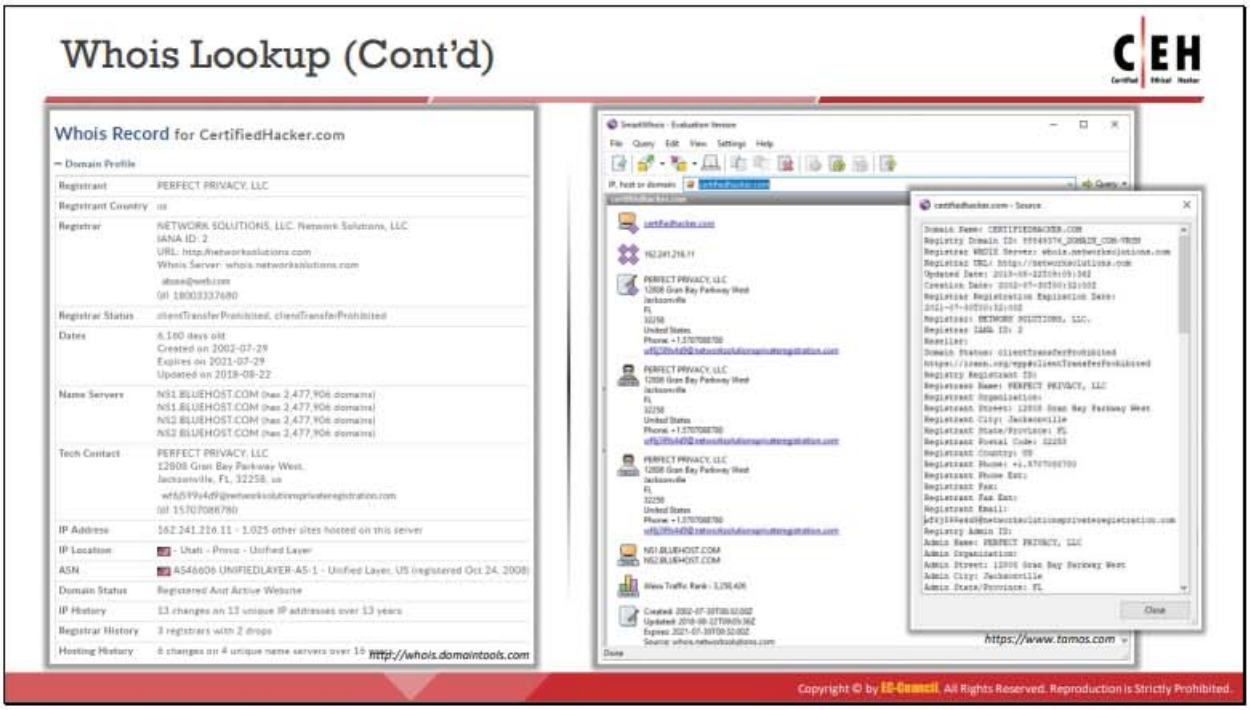


Figure 2.58: Screenshot of eMailTrackerPro





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 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).

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

- **Thick Whois** - Stores the complete Whois information from all the registrars for a particular set of data.
- **Thin Whois** - 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.

Whois query returns the following information:

- Domain name details
- Contact details of the domain owner
- Domain name servers
- NetRange
- When a domain has been created
- Expiry records
- Records last updated

An attacker queries a Whois database server to obtain information about the target domain name, contact details of its owner, expiry date, creation date, and so on, 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 then obtain internal details of the network.

Regional Internet Registries (RIRs)

The RIRs include:

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

Whois Lookup Result

Whois services such as *http://whois.domaintools.com* or *https://www.tamos.com* can help to perform Whois lookups. The screenshot shows the result analysis of a Whois lookup obtained with the two above-mentioned Whois services. The services perform Whois lookup by entering the target's domain or IP address. The domaintools.com service provides Whois information such as registrant information, email, administrative contact information, creation and expiry date, and a list of domain servers. SmartWhois, available at *http://www.tamos.com*, gives 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 also helps in finding the owner of the domain, the owner's contact information, the owner of the IP address block, registered date of the domain, and so on. It supports Internationalized Domain Names (IDNs), which means 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. Network Solutions, LLC IANA ID: 2 URL: http://networksolutions.com Whois Server: whois.networksolutions.com abuse@web.com (p) 18003337680
Registrar Status	clientTransferProhibited, clientTransferProhibited
Dates	6,160 days old Created on 2002-07-29 Expires on 2021-07-29 Updated on 2018-08-22
Name Servers	NS1.BLUEHOST.COM (has 2,477,906 domains) NS1.BLUEHOST.COM (has 2,477,906 domains) NS2.BLUEHOST.COM (has 2,477,906 domains) NS2.BLUEHOST.COM (has 2,477,906 domains)
Tech Contact	PERFECT PRIVACY, LLC 12808 Gran Bay Parkway West, Jacksonville, FL, 32258, us wf6j599s4d9@networksolutionsprivateregistration.com (p) 15707088780
IP Address	162.241.216.11 - 1,025 other sites hosted on this server
IP Location	- Utah - Provo - Unified Layer
ASN	AS46606 UNIFIEDLAYER-AS-1 - Unified Layer, US (registered Oct 24, 2008)
Domain Status	Registered And Active Website
IP History	13 changes on 13 unique IP addresses over 13 years
Registrar History	3 registrars with 2 drops
Hosting History	6 changes on 4 unique name servers over 16 years

Figure 2.59: Screenshot of Whois

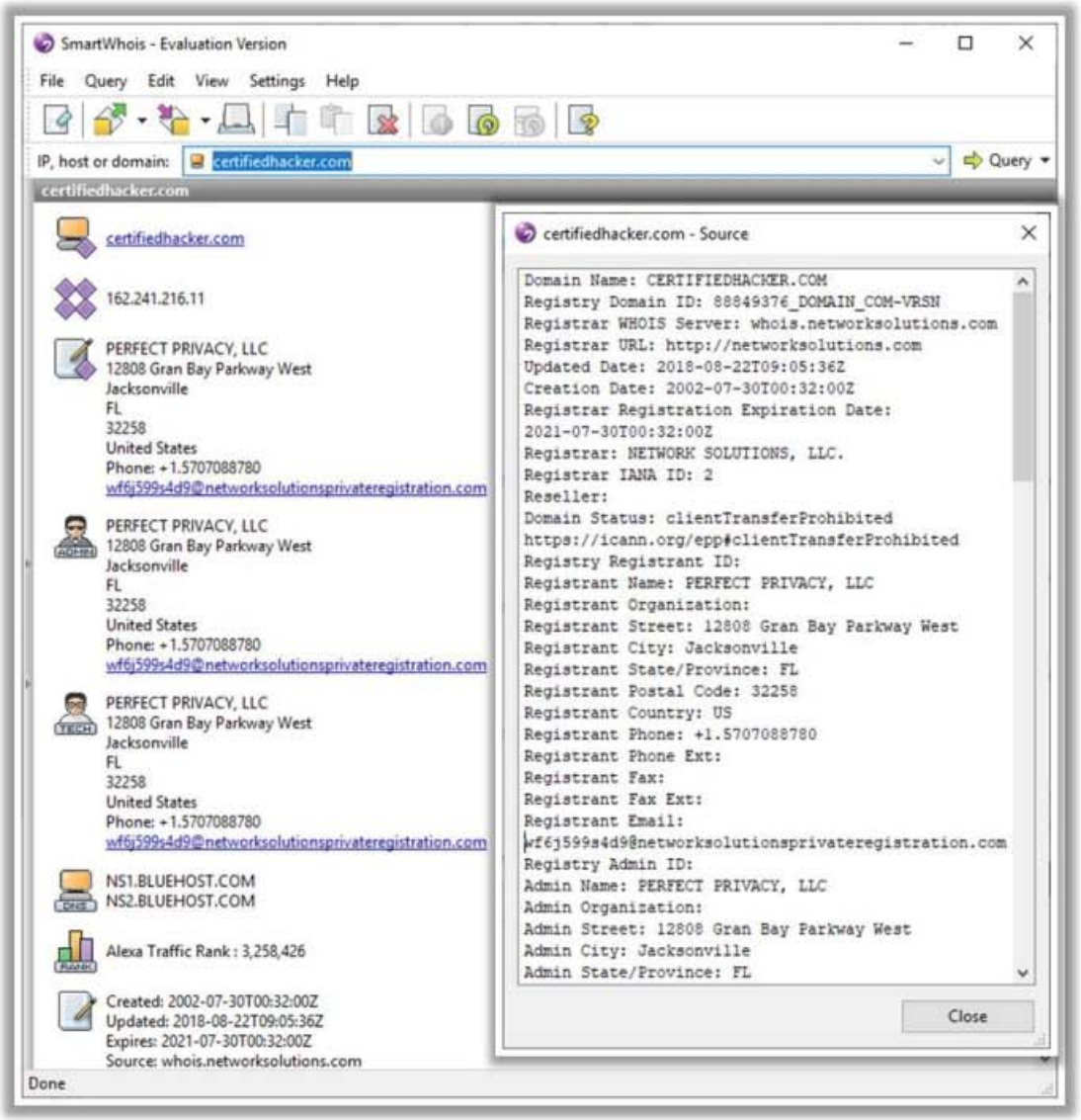
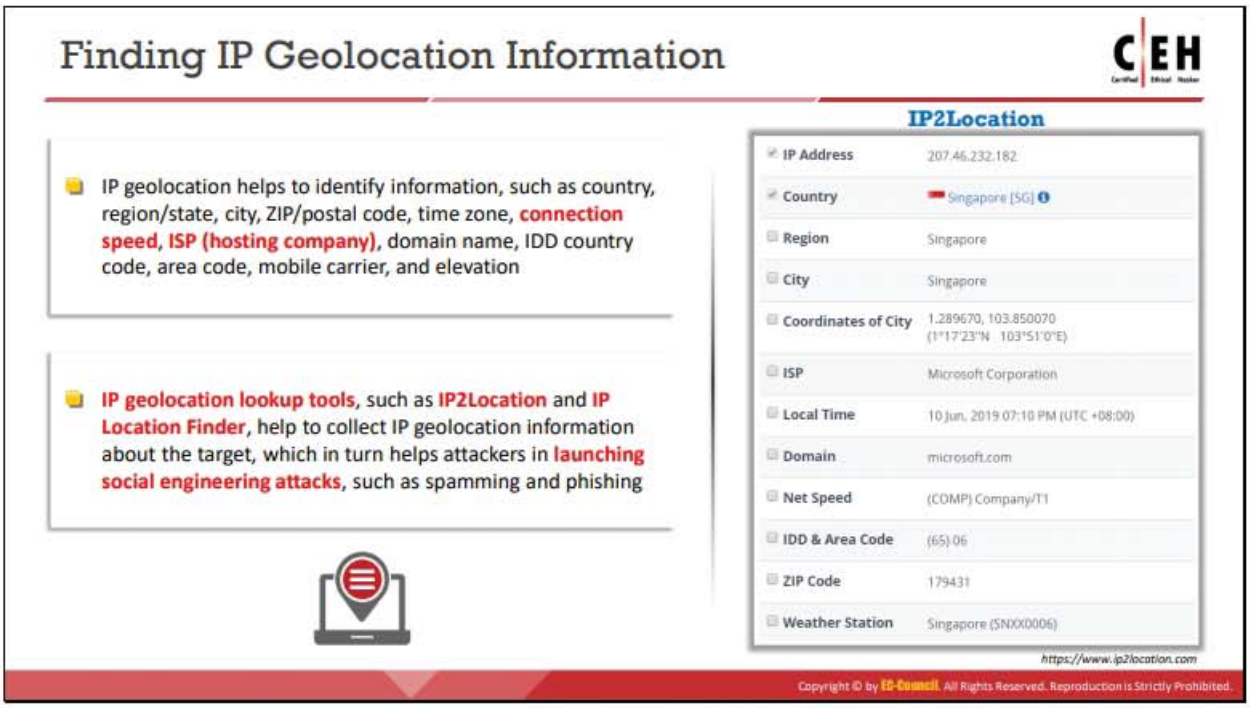


Figure 2.60: Screenshot of SmartWhois

Attackers use Whois lookup tools such as Batch IP Converter, WhoIs Analyzer Pro, and ActiveWhois to extract information such as IP addresses, hostnames or domain names, registrant information, and DNS records, including the country, city, state, phone and fax numbers, network service providers, administrators, and technical support information, for any IP address or domain name.



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 using 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.

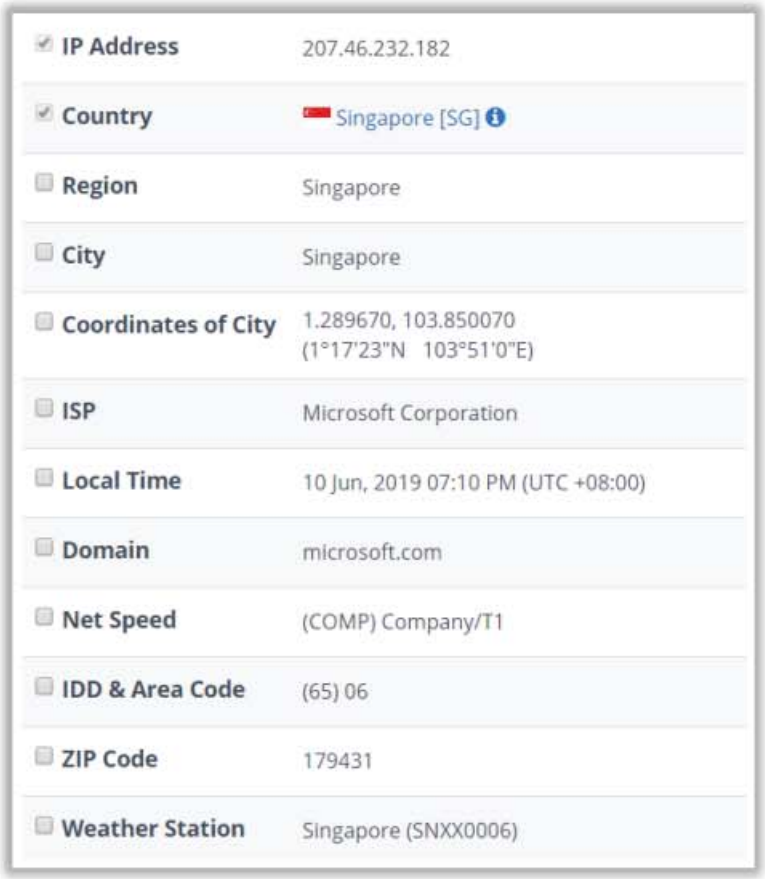
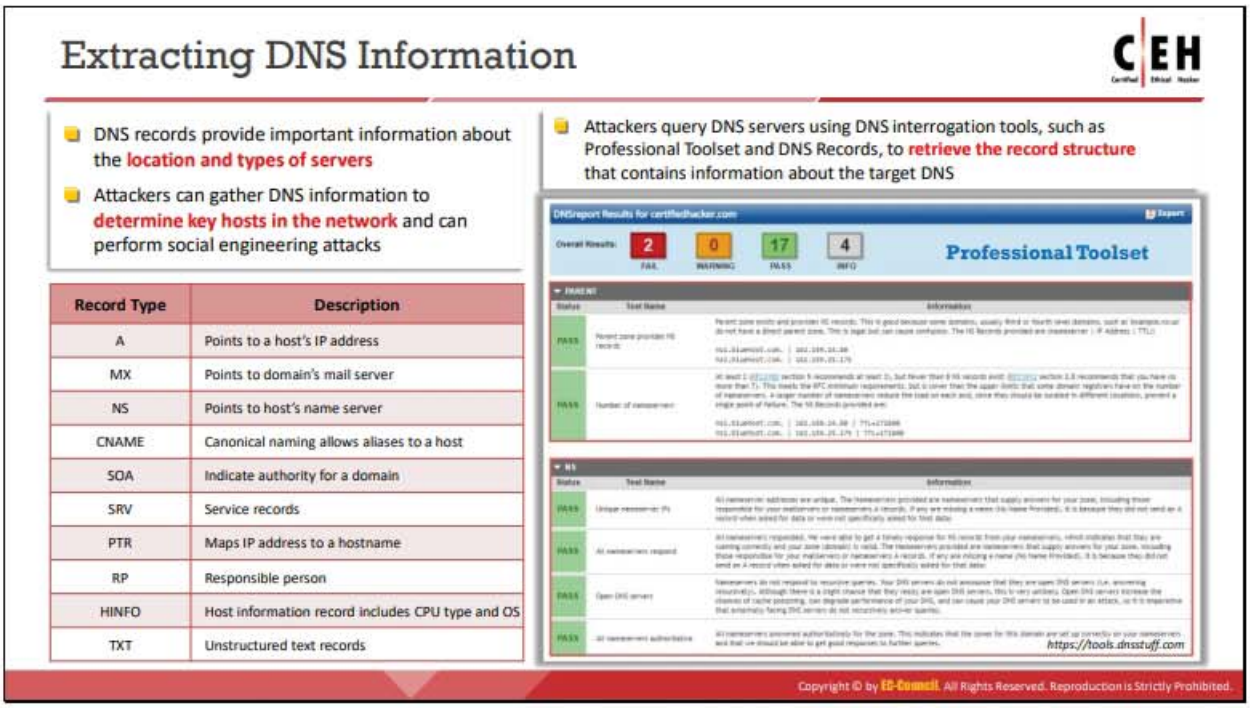


Figure 2.61: Screenshot of IP2Location



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

DNS Footprinting

After collecting Whois records about the target, the next phase in the footprinting methodology is DNS footprinting. Attackers perform DNS footprinting to gather information about DNS servers, DNS records, and types of servers used by the target organization. This information helps attackers to identify the hosts connected in the target network and perform further exploitation on the target organization.

This section describes how to extract DNS information, perform the reverse DNS lookup, and collect information from DNS zone transfers, as well as DNS interrogation tools.

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
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 2.7: DNS records and their description

DNS interrogation tools such as Professional Toolset (*https://tools.dnsstuff.com)* and DNS Records (*https://network-tools.com*) enable the user to perform DNS footprinting. DNSstuff (Professional Toolset) extracts DNS information about IP addresses, mail server extensions, DNS lookups, Whois lookups, and so on. It can extract a range of IP addresses using an IP routing lookup. If the target network allows unknown, unauthorized users to transfer DNS zone data, then it is easy for an attacker to obtain the information about DNS with the help of the DNS interrogation tool.

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

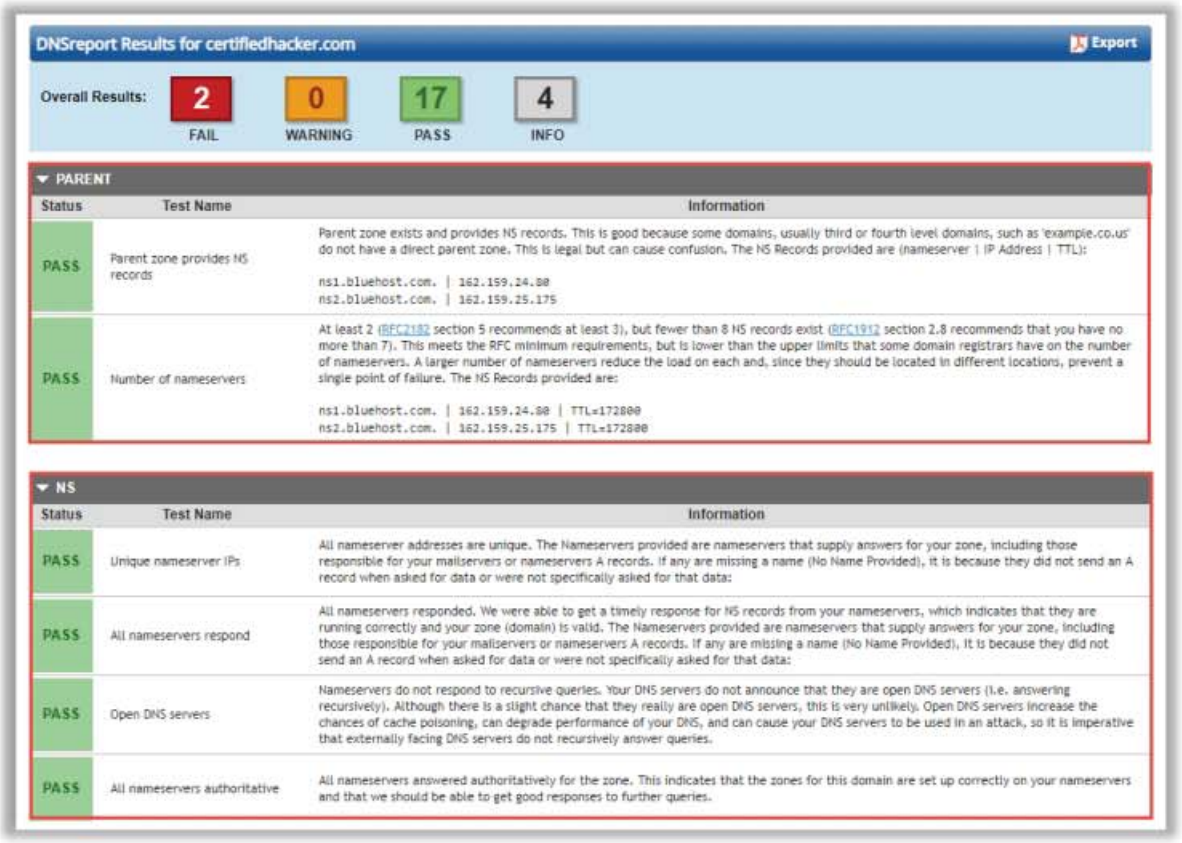
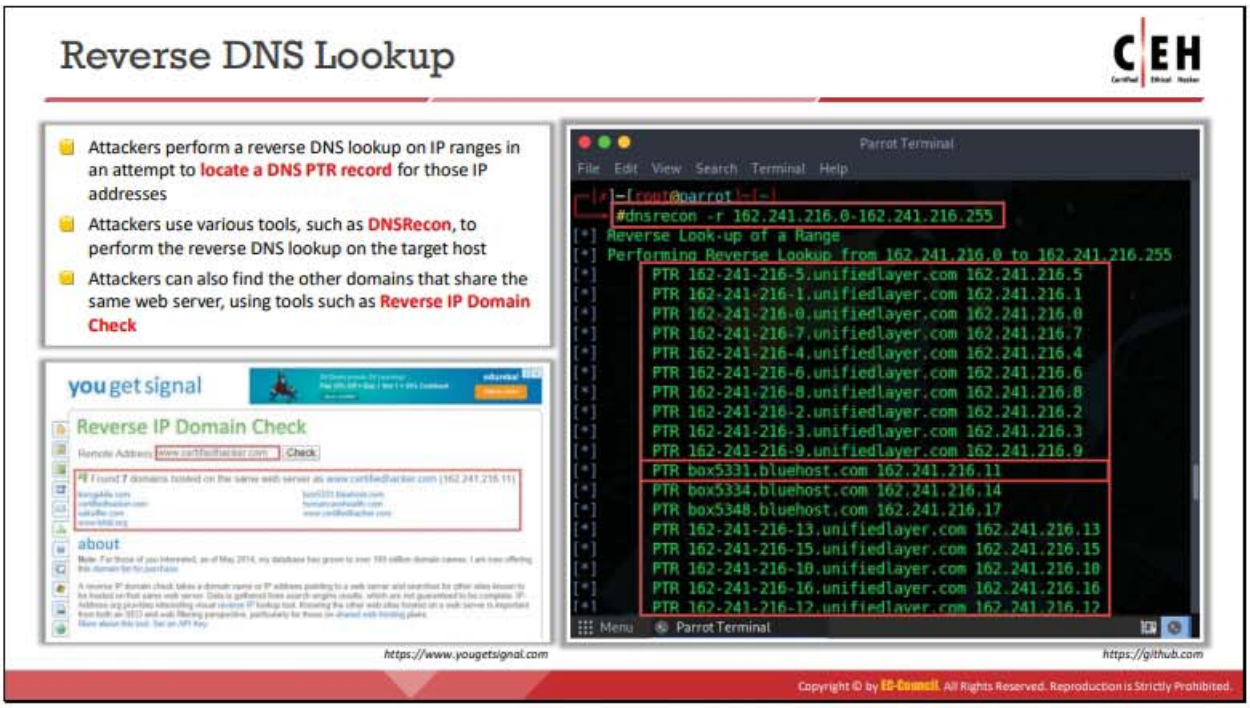


Figure 2.62: Screenshot of Professional Toolset

Attackers also use DNS lookup tools such as DNSdumpster.com, Bluto, and Domain Dossier to retrieve DNS records for a specified domain or hostname. These tools retrieve information such as domains and IP addresses, domain Whois records, DNS records, and network Whois records.



Reverse DNS Lookup

DNS lookup is used for finding the IP addresses for a given domain name, and the reverse DNS operation is performed to obtain the domain name of a given IP address. When you are looking for a domain and type the domain name in the browser, the DNS converts that 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 by 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 and Reverse IP Domain Check for performing the reverse DNS lookup on the target host. When we get 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-last) for a reverse lookup by brute force.

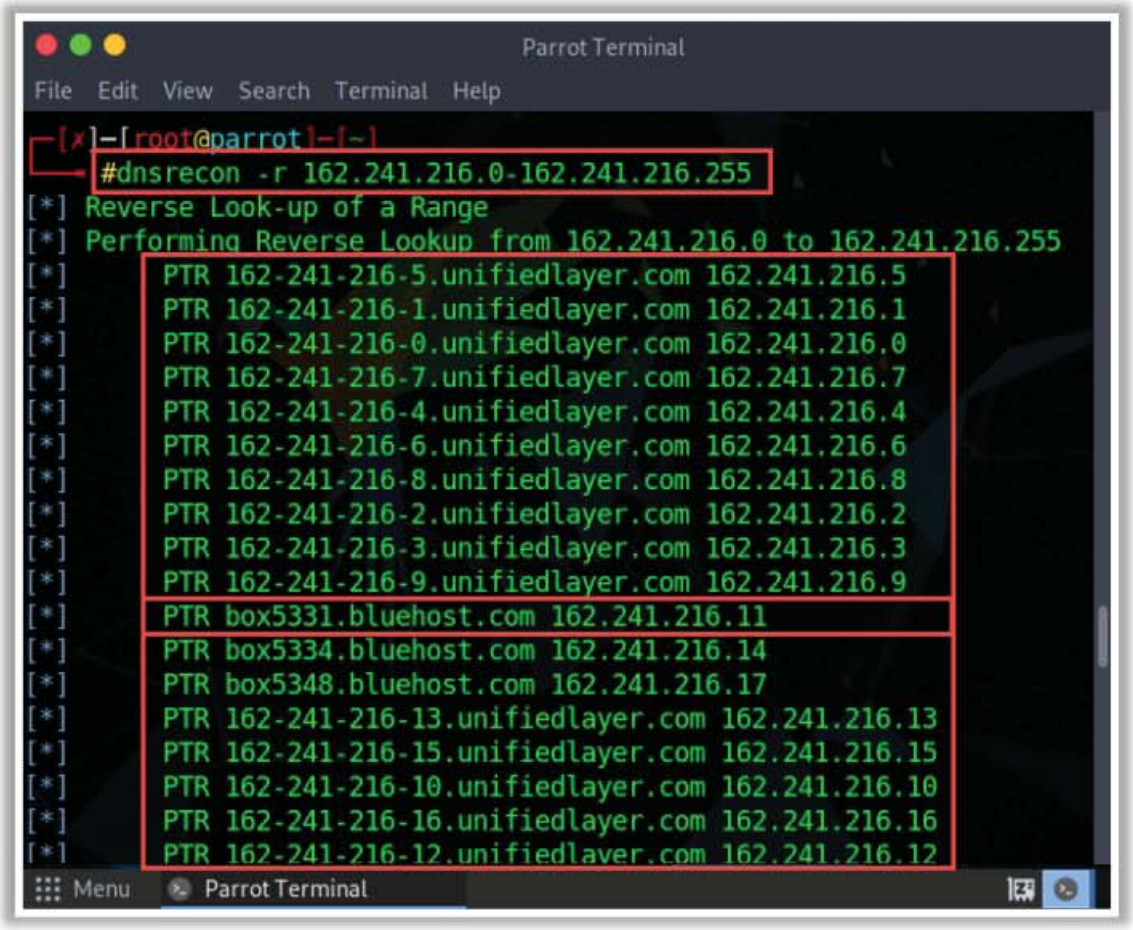


Figure 2.63: Screenshot of DNSRecon showing reverse DNS lookup information

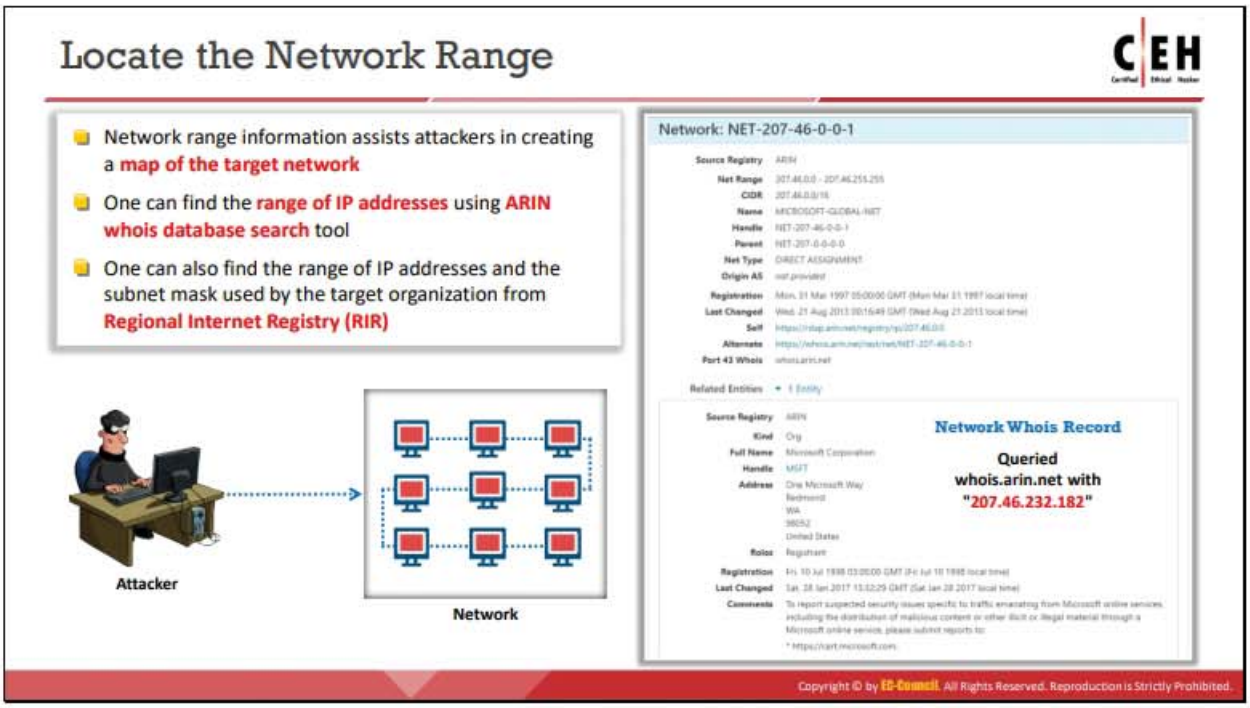
Attackers also find the other domains that share the same web server using tools such as Reverse IP Domain Check. These tools list the possible domains that are hosted on the same web server.

- **Reverse IP Domain Check**

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

  As shown in the screenshot, a reverse IP domain check takes a domain name or IP address pointing to a web server and searches for other sites known to be hosted on the same web server.

Figure 2.64: Screenshot of Reverse IP Domain Check



Network Footprinting

The next step after retrieving the DNS information is gathering network-related information. We will now discuss network footprinting, a method of gathering the footprint of the target organization's network. This section describes how to locate the network range, traceroute analysis, and traceroute tools.

Locate the Network Range

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 they do to perform network footprinting. The answers to these questions provide information about the internal structure of the target network.

After gathering the information, an attacker can proceed to find the network range of a target system. Detailed information is available from the appropriate regional registry database regarding IP allocation and the nature of the allocation. An attacker can also determine the subnet mask of the domain and trace the route between the system and the target system. Traceroute tools that are widely used include Path Analyzer Pro and VisualRoute.

Obtaining private IP addresses can be useful to attackers. The Internet Assigned Numbers Authority (IANA) has reserved the following three blocks of the 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, the attacker can get information about how the network is structured and which machines in the networks are alive. Using the network range also helps to identify the network topology, access control device, and OS used in the target network. To find the network range of the target network, one needs to enter the server IP address (that was 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 in the **SEARCH Whois** text box. This gives the network range of the target network. Improperly set up DNS servers offer attackers a good chance of obtaining a list of internal machines on the server. In addition, sometimes, if an attacker traces a route to a machine, it is possible to obtain the internal IP address of the gateway, which can be useful.

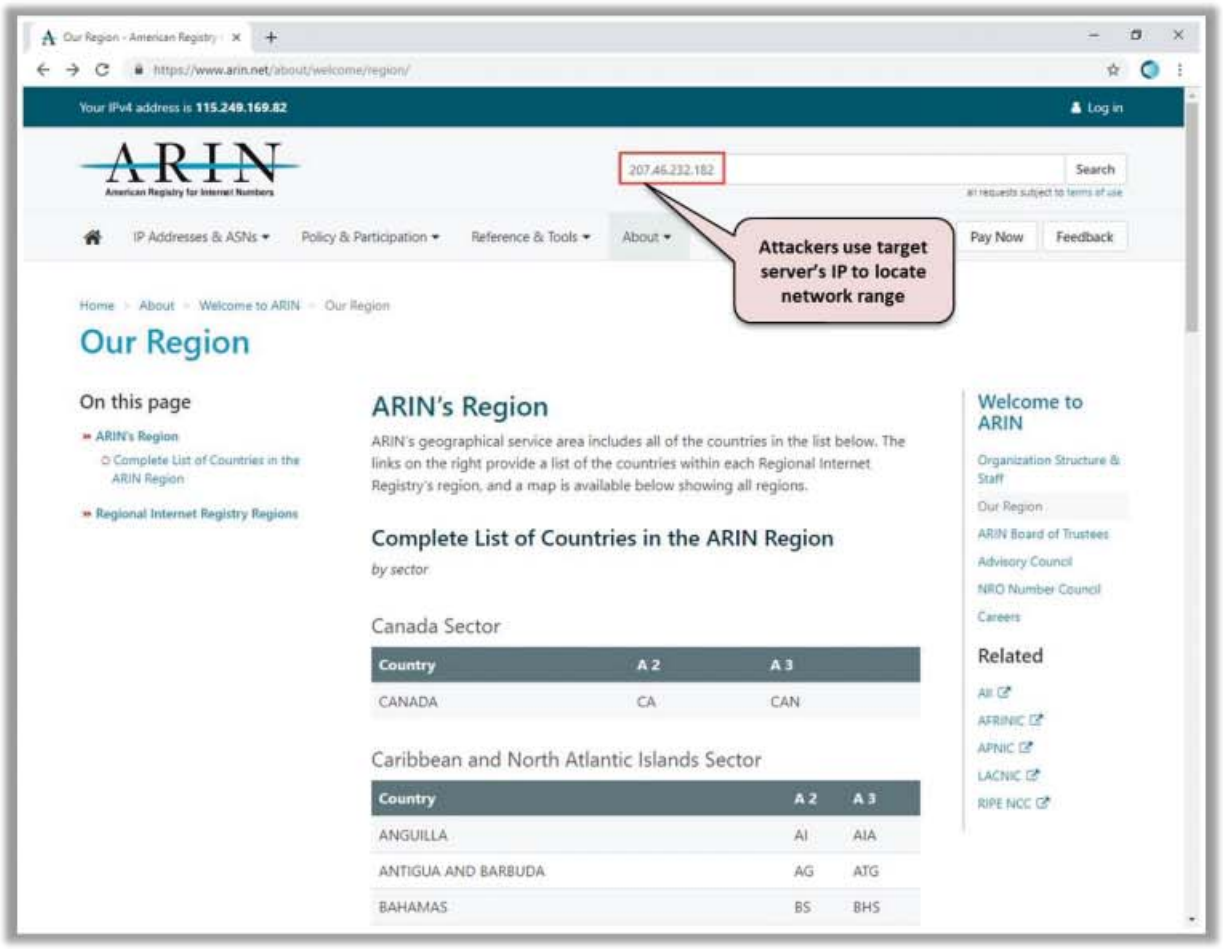


Figure 2.65: Screenshot of ARIN's Region

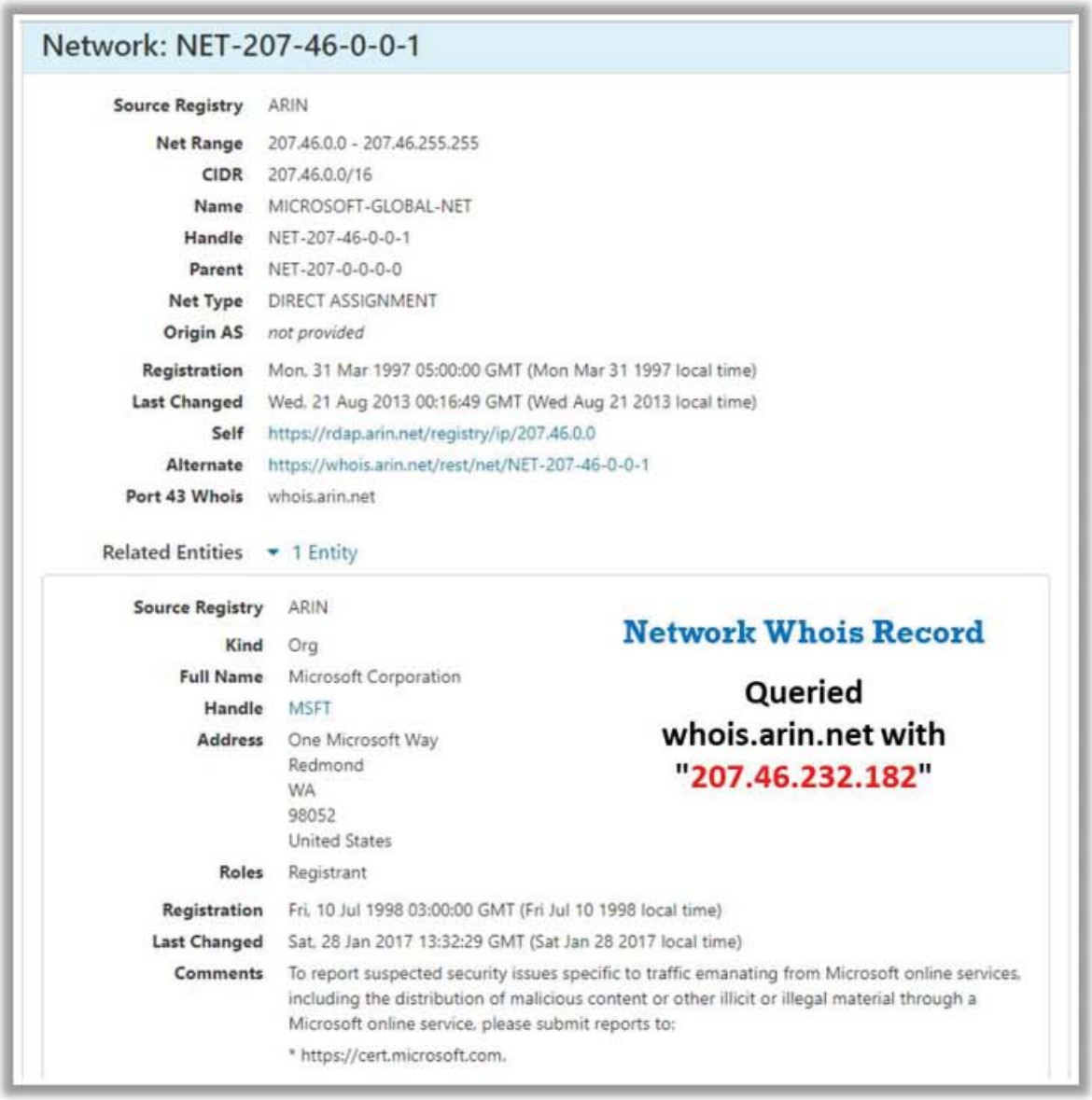
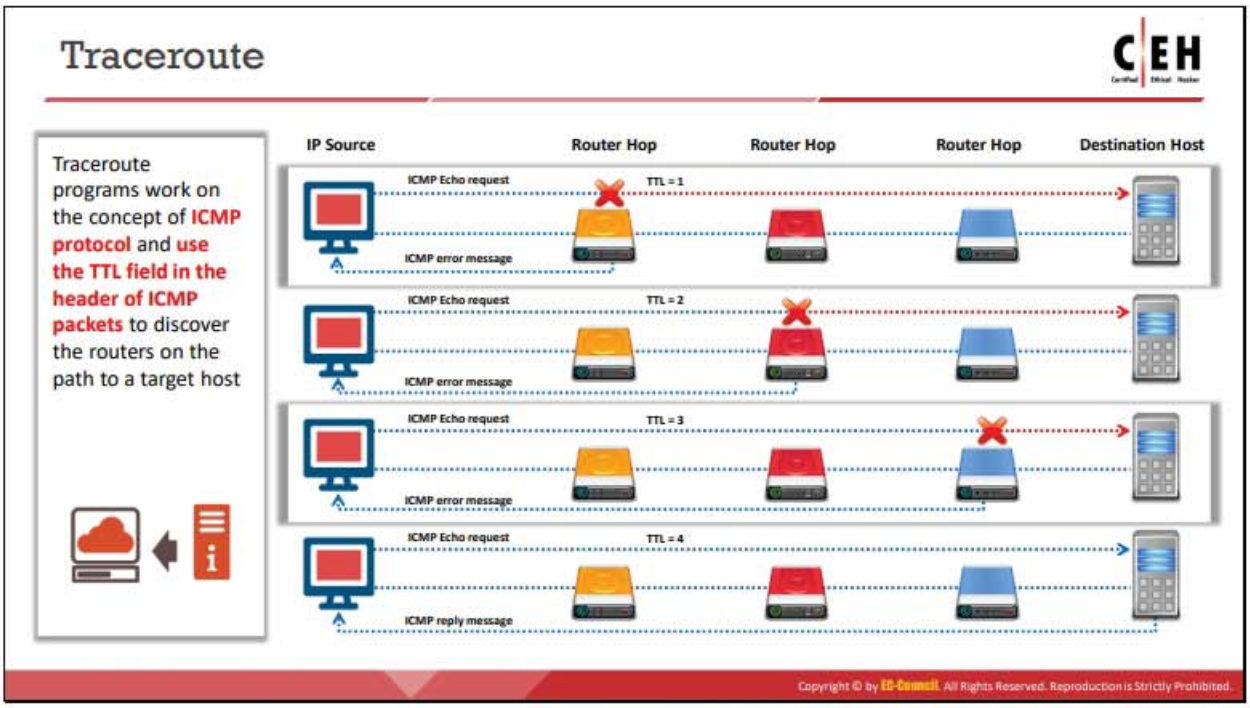
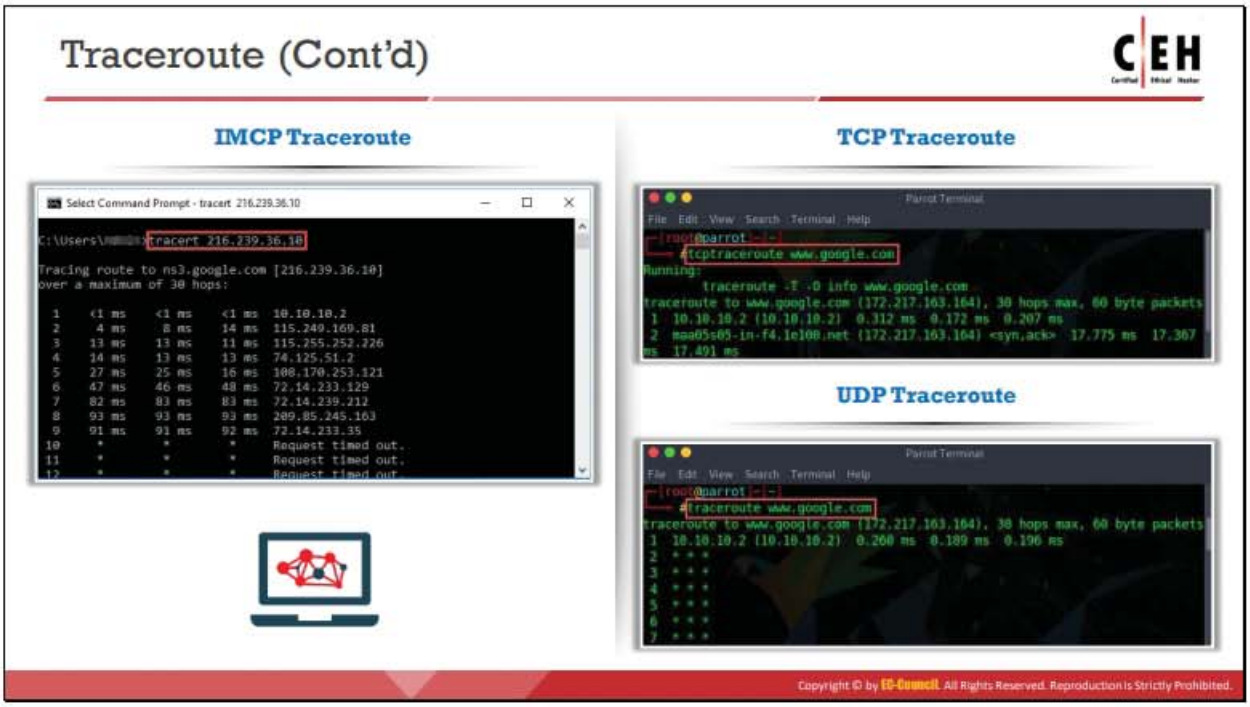


Figure 2.66: Screenshot showing result of ARIN Whois database search result

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





Traceroute

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 concept 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.

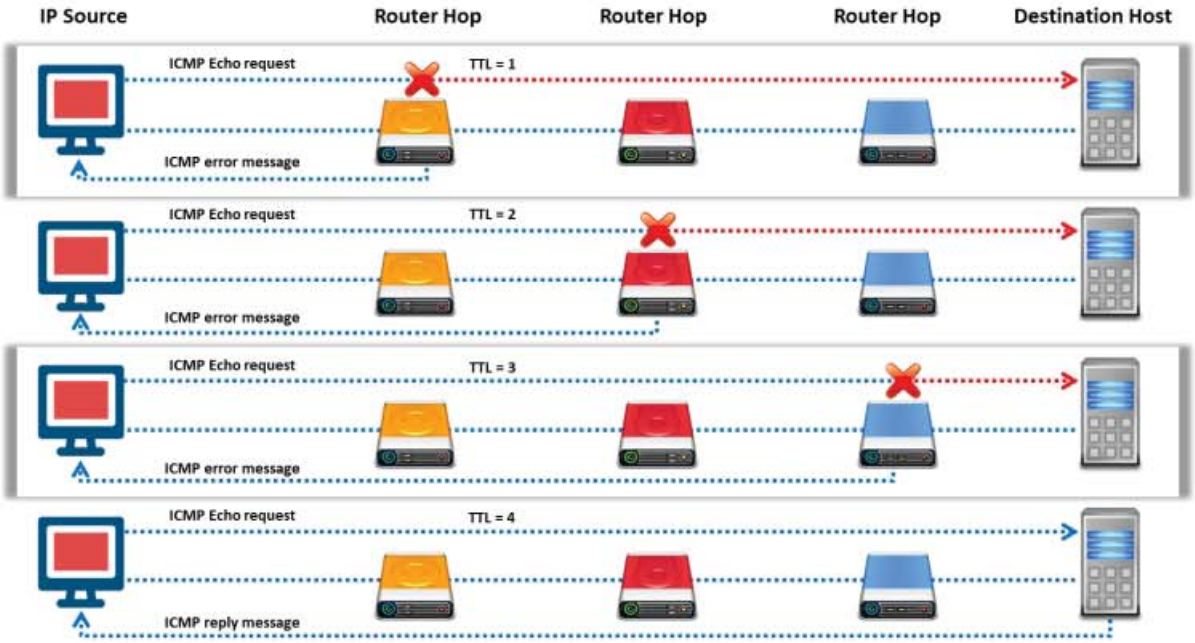


Figure 2.67: Illustration of Traceroute

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

Tracing route to ns3.google.com [216.239.36.10] over a maximum of 30
hops:

  1   <1 ms     <1 ms     <1 ms   10.10.10.2
  2   20 ms      4 ms      5 ms   1.6.15.234
  3   21 ms     19 ms     21 ms   100.66.8.23
  4   20 ms     19 ms     19 ms   100.68.8.23
  5   23 ms     41 ms     20 ms   72.14.210.200
  6   21 ms     21 ms     23 ms   108.170.248.163
  7   68 ms     67 ms     67 ms   209.85.242.115
  8  102 ms    102 ms    102 ms   209.85.247.194
  9  100 ms    106 ms    122 ms   72.14.239.175
 10  114 ms    119 ms    114 ms   209.85.244.31
 11  114 ms    112 ms    112 ms   209.85.247.118
 12  114 ms    118 ms    115 ms   74.125.253.85
 13  111 ms    112 ms    113 ms   ns3.google.com [216.239.36.10]

 Trace complete.
```

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:

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

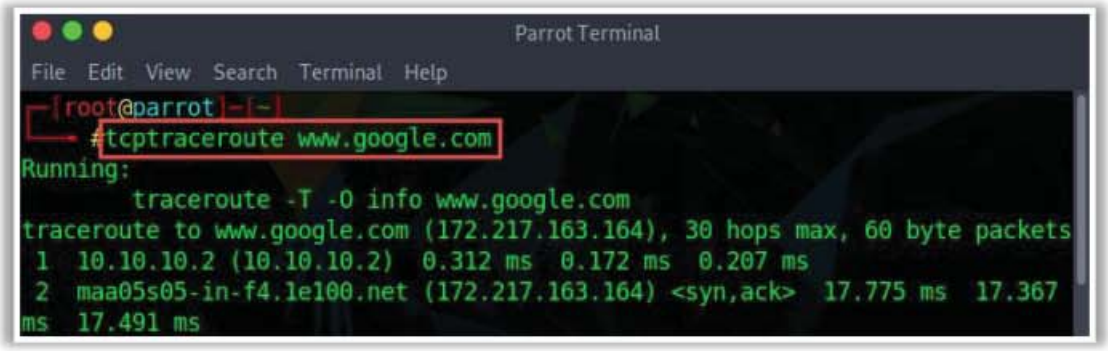


Figure 2.68: Screenshot showing the output of TCP Traceroute

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
```

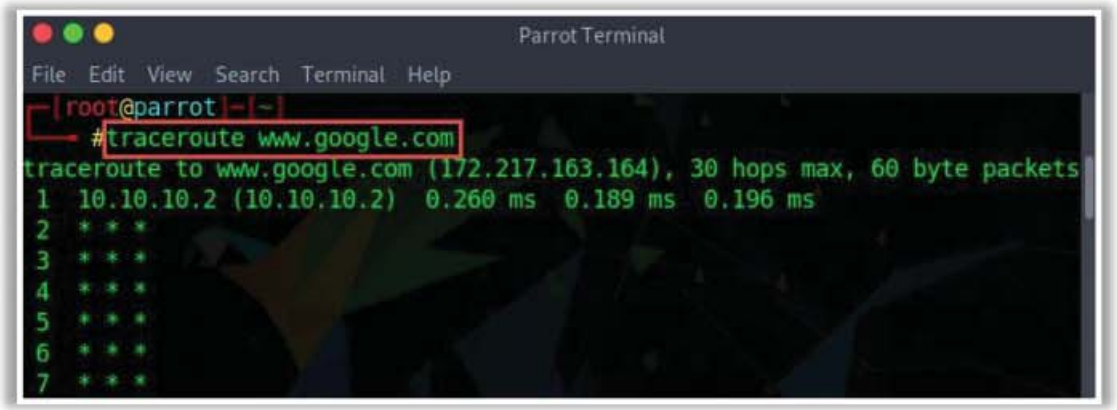
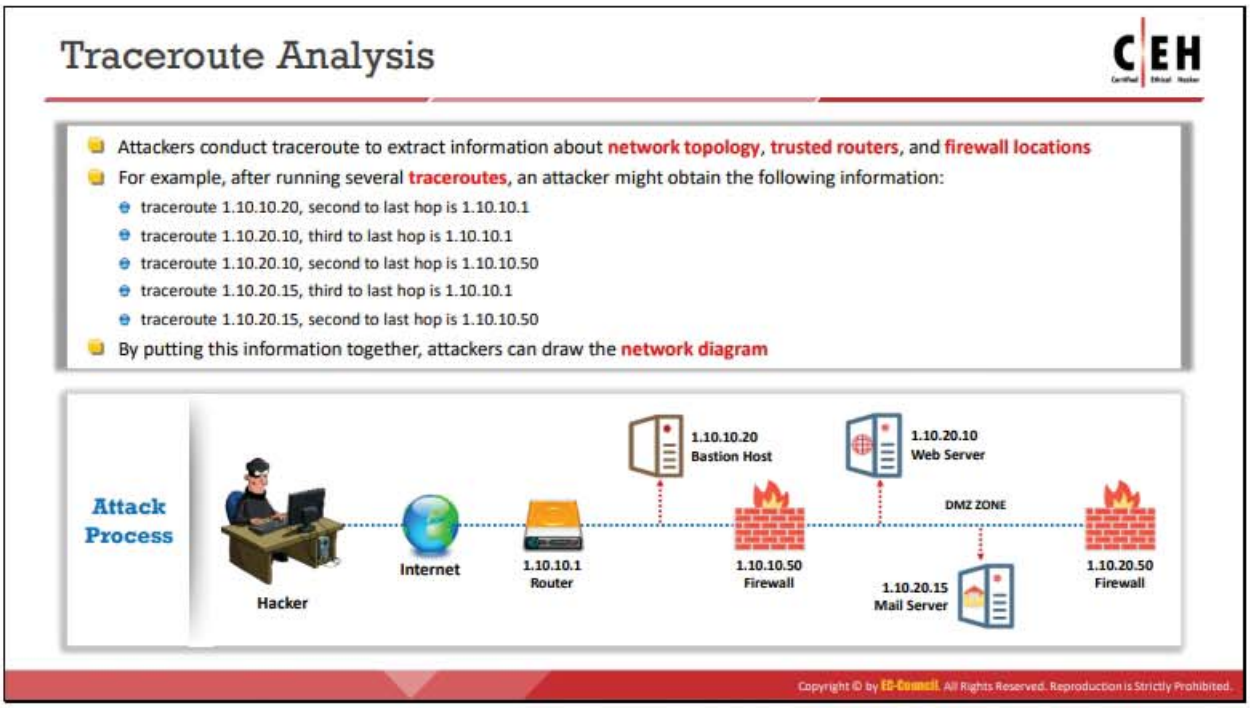


Figure 2.69: Screenshot showing the output of UDP Traceroute



Traceroute Analysis

We have seen how the Traceroute utility helps to 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 will be able to find the location of a hop in the target network. Consider the following traceroute results obtained:

- `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 draw the network topology diagram of the target network, as shown below.

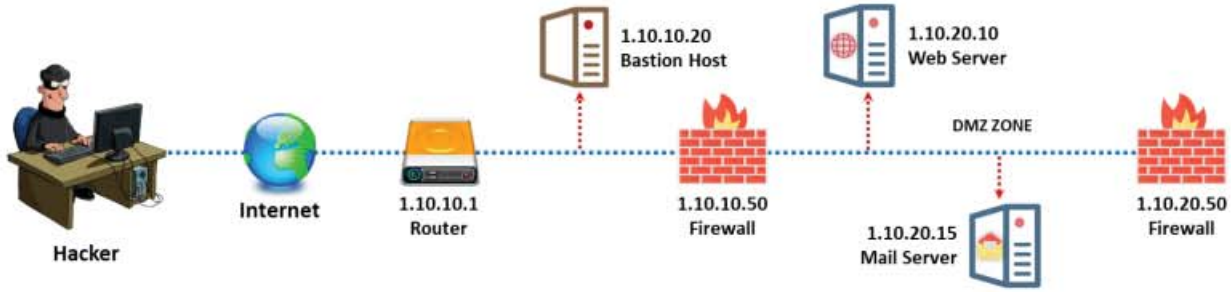
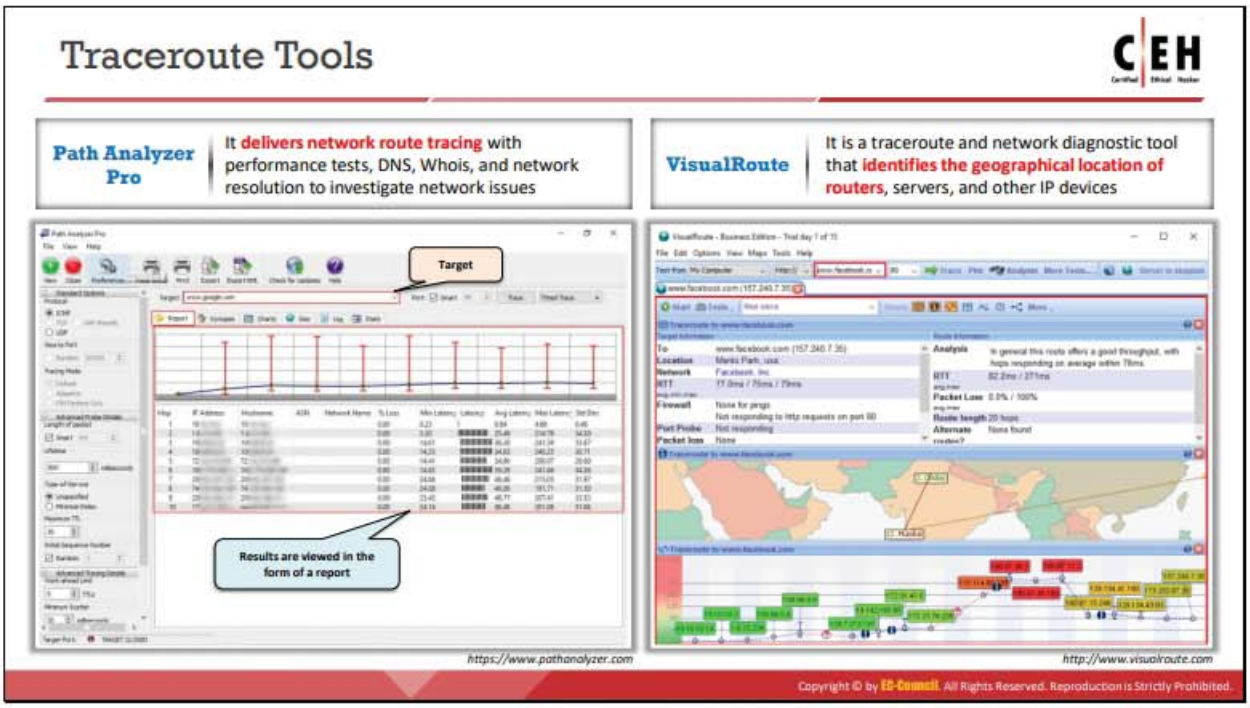


Figure 2.70: Traceroute Analysis



Traceroute Tools

Traceroute tools such as Path Analyzer Pro, VisualRoute, Traceroute NG, and PingPlotter 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
- **Path Analyzer Pro**

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

  Path Analyzer Pro performs network route tracing with performance tests, DNS, Whois, and network resolution to investigate network issues.

  Attackers use Path Analyzer Pro to identify the route from the source to destination target systems graphically. As shown in the screenshot, this tool helps attackers to gather information such as the hop number, its IP address, hostname, ASN, network name, percentage loss, latency, average latency, and standard deviation for each hop in the path.

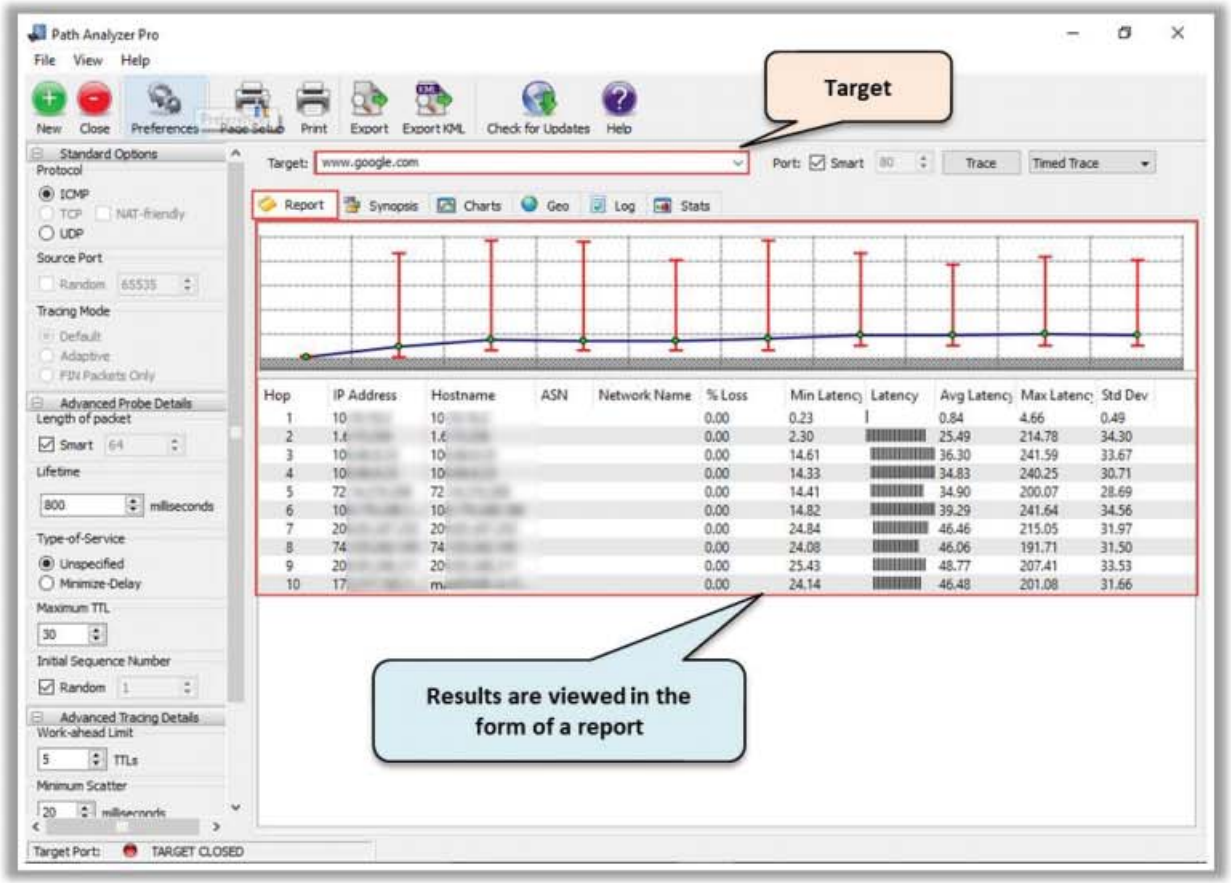


Figure 2.71: Screenshot of Path Analyzer Pro

- **VisualRoute**

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

  VisualRoute is a traceroute and network diagnostic tool. Attackers use VisualRoute to identify the geographical location of routers, servers, and other IP devices in the target network.

  This tool helps attackers in tracking the path between the source and destination systems and obtaining the results in a graphical format. As shown in the screenshot, using VisualRoute tool enables attackers to gather information such as hop number, IP address, node name, and geographical location of each hop in the route.

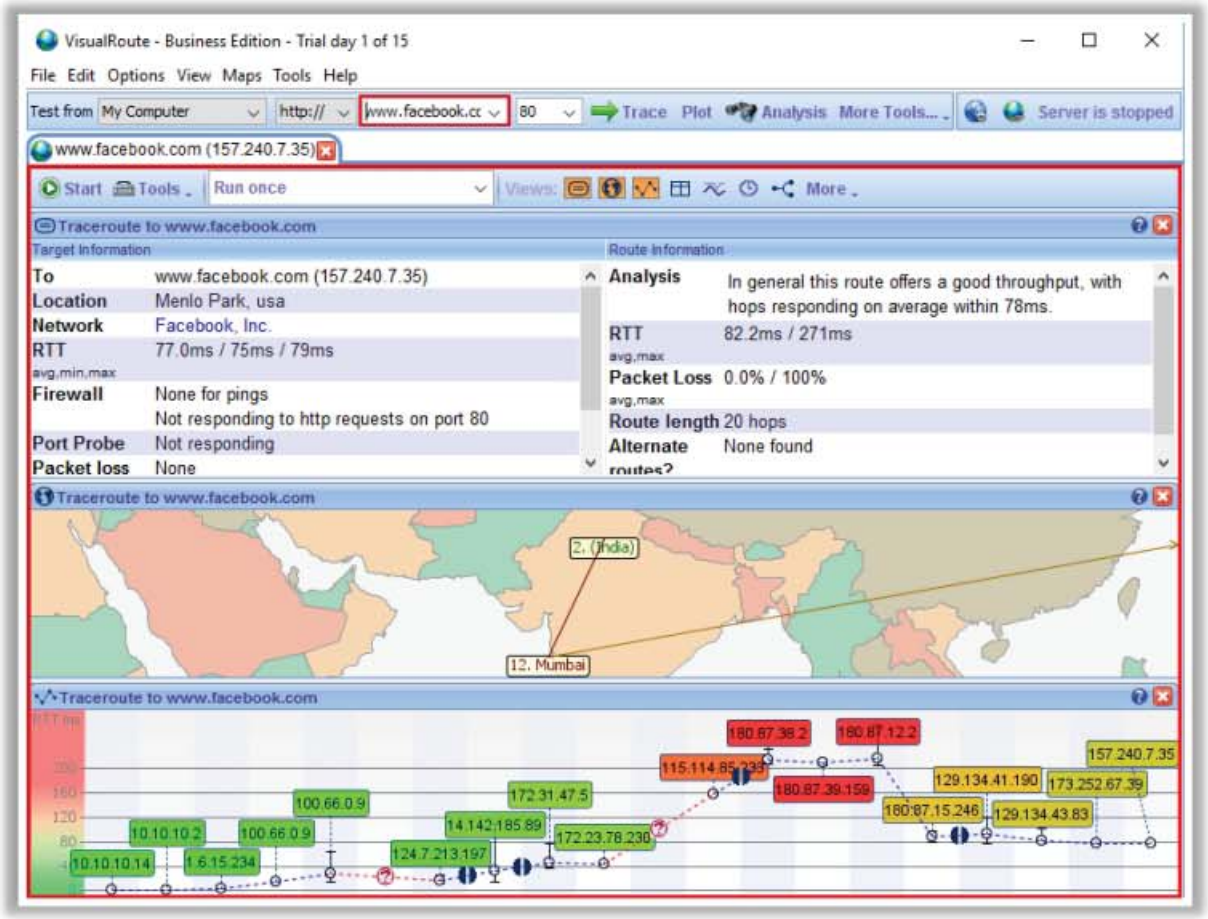
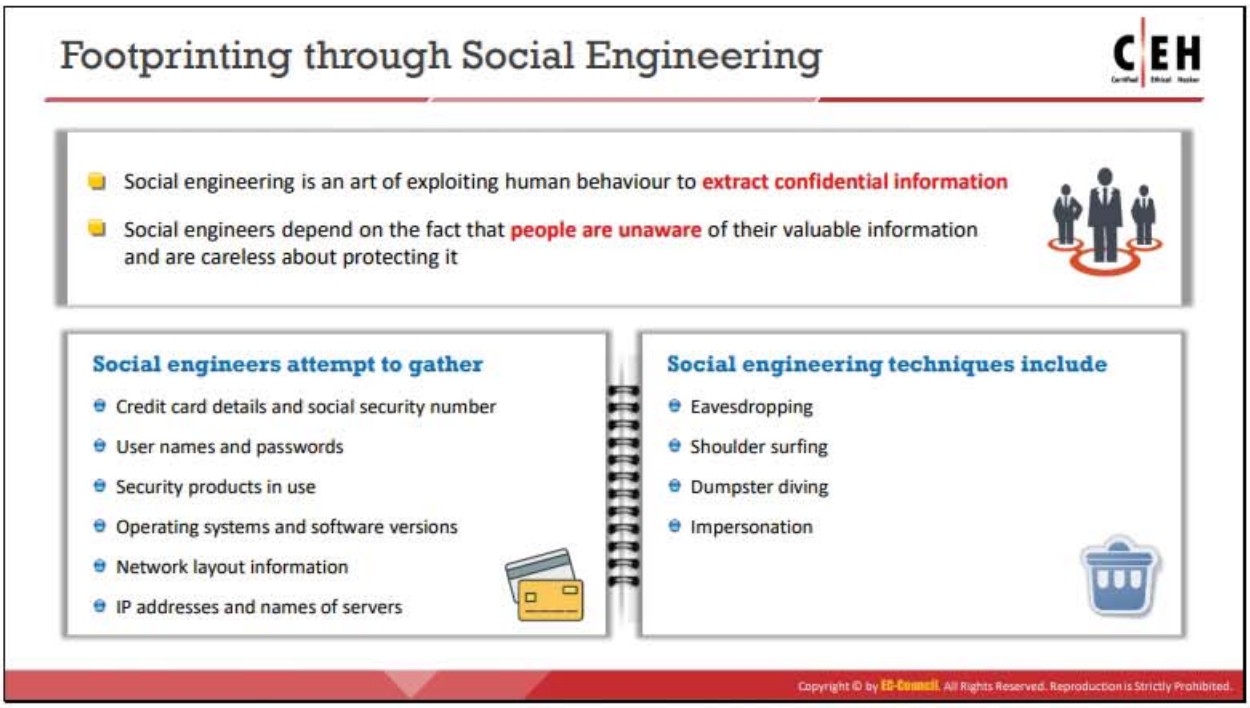


Figure 2.72: Screenshot of VisualRoute



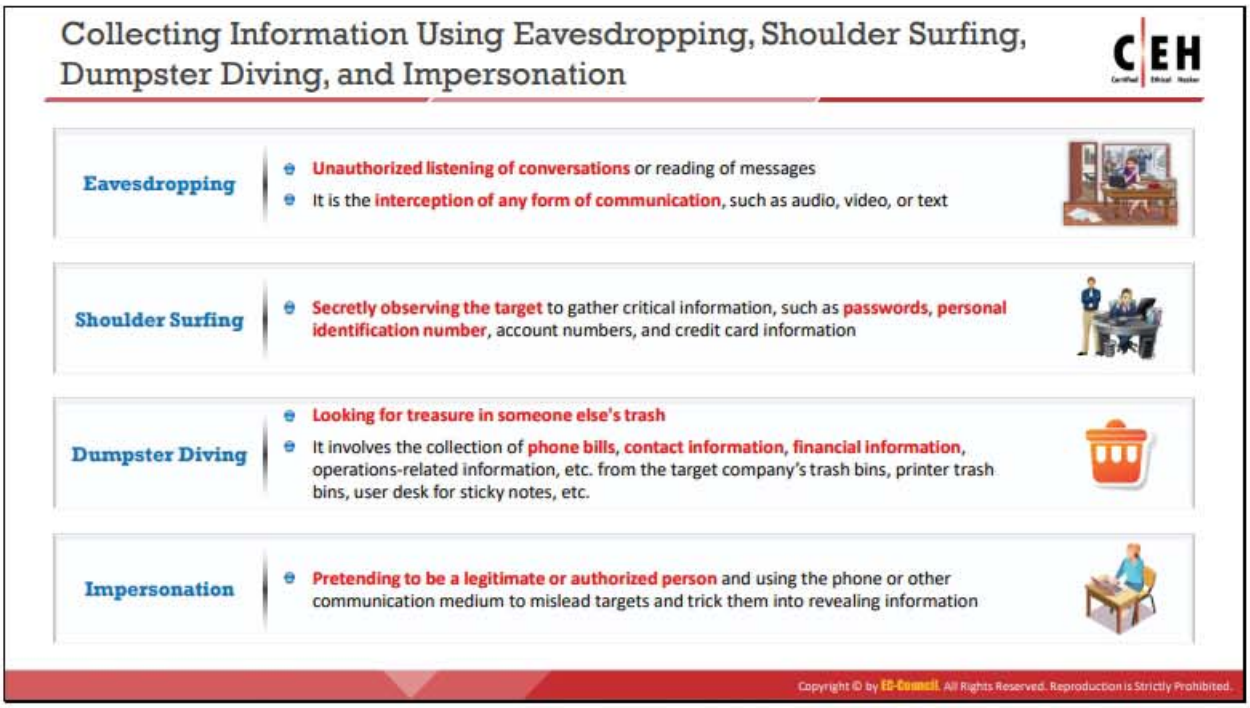
Footprinting through Social Engineering

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 Using Eavesdropping, Shoulder Surfing, Dumpster Diving, and Impersonation

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

- **Eavesdropping**

  Eavesdropping is the act of secretly listening to the conversations of people over a phone or video conference without their consent. It also includes reading confidential messages from communication media, such as instant messaging or fax transmissions. It is the act of intercepting communication in any form such as audio, video, or text without the consent of the communicating parties. The attacker gains information by tapping phone conversations or intercepting audio, video, or written communication.

- **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**

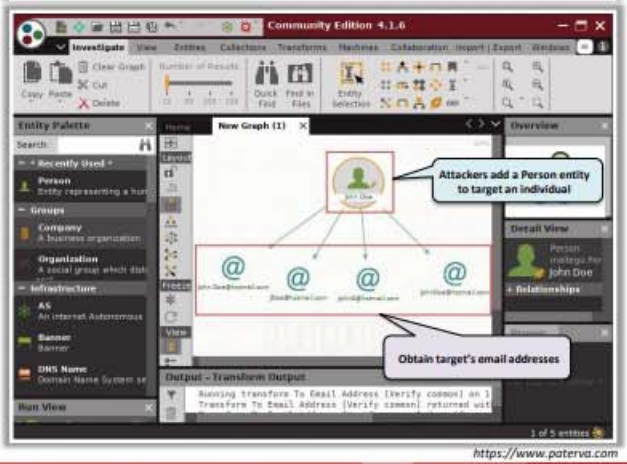
  Impersonation is a technique whereby an attacker pretends to be a legitimate or authorized person. Attackers perform impersonation attacks personally or use phones or other communication media to mislead targets and trick them into revealing information. The attacker might impersonate a courier/delivery 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.

Module Flow

1. **Footprinting Concepts**
2. **Footprinting Methodology**
3. **Footprinting Tools**
4. **Footprinting Countermeasures**

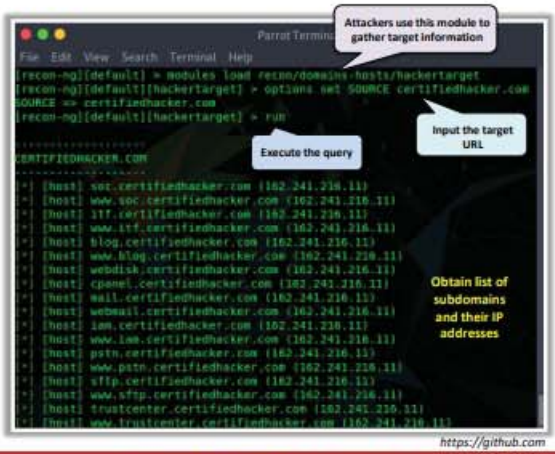
Footprinting Tools: Maltego and Recon-ng

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.paterva.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

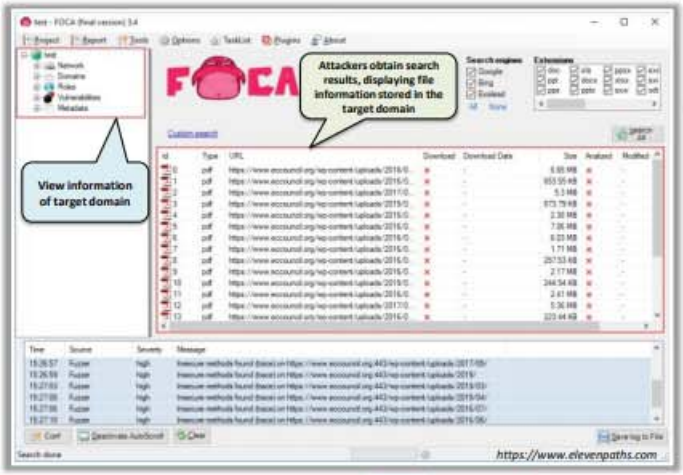


https://github.com

Footprinting Tools: FOCA and OSRFramework

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

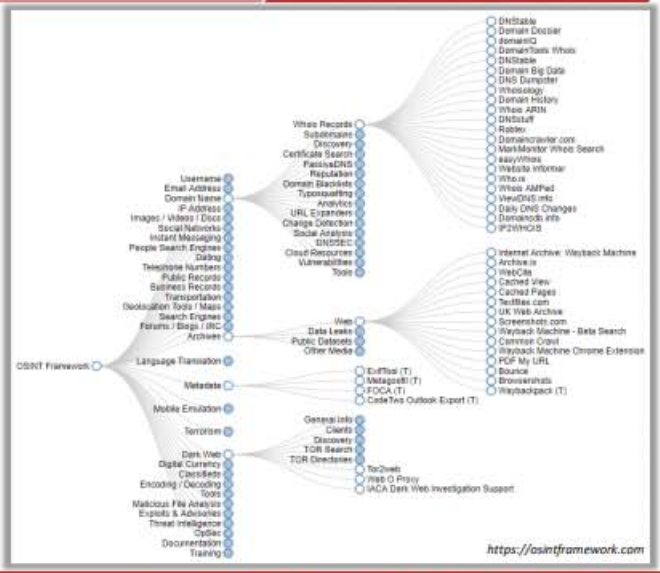
OSRFramework includes applications related to username checking, DNS lookups, information leaks research, deep web search, regular expressions extraction, etc.

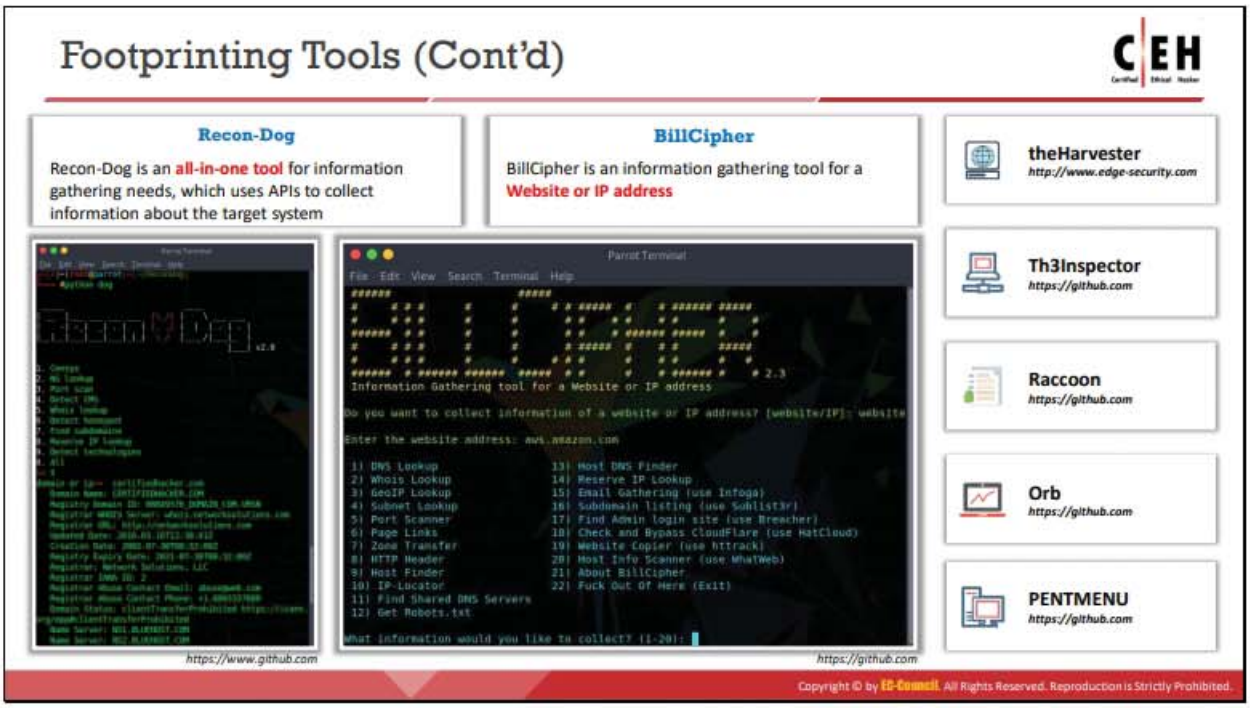


Footprinting Tools: OSINT Framework

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





Footprinting Tools

Various tools help attackers in footprinting. Many organizations offer tools that make information gathering an easy task. This section describes tools intended for obtaining information from various sources.

Footprinting tools are used to collect basic information about target systems to exploit them. Information collected by the footprinting tools includes the target's IP location information, routing information, business information, address, phone number and social security number, details about a source of an email and a file, DNS information, domain information, and so on.

- **Maltego**

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

  Maltego is a program 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 **`Person entity`**, rename it with the target's name, and obtain the email addresses associated with the target.

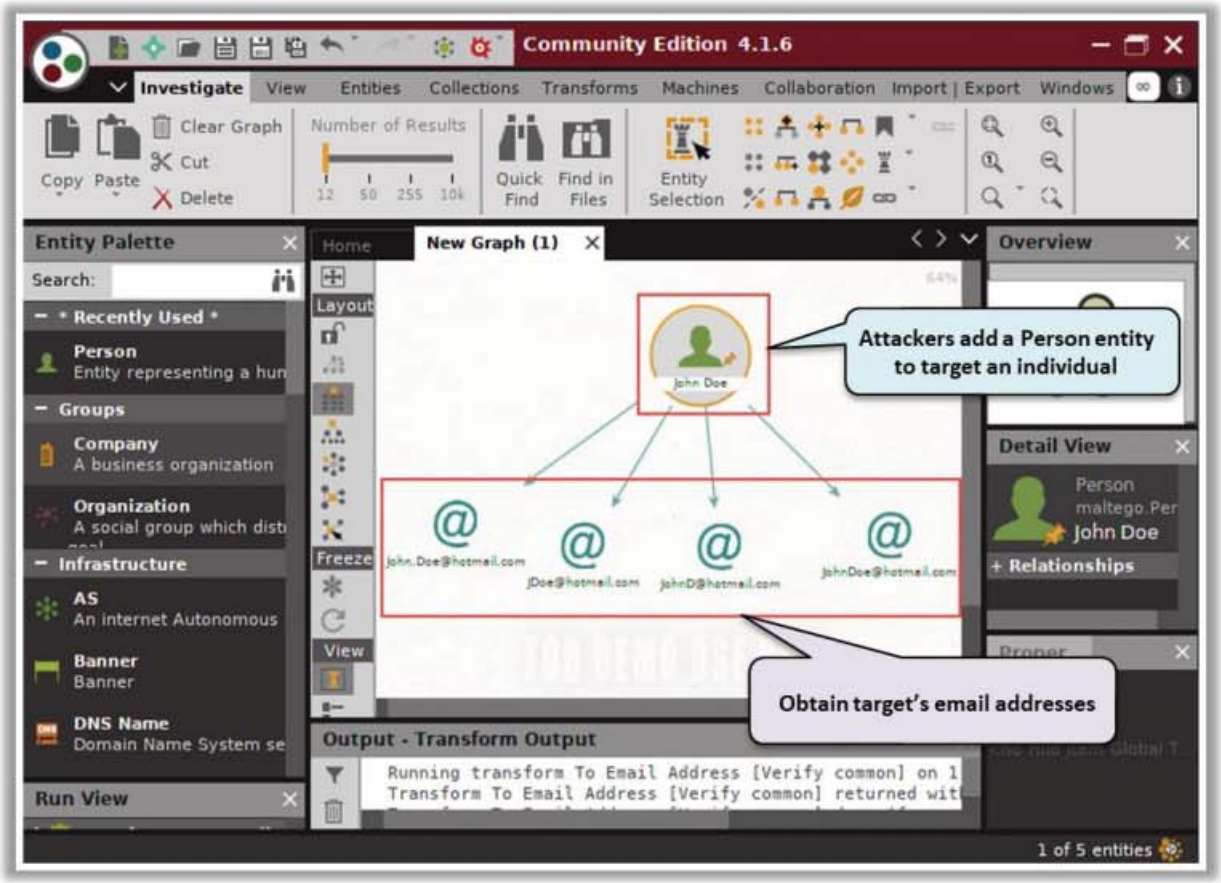


Figure 2.73: Screenshot of Maltego

- **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/hackertarget`** to extract a list of subdomains and IP addresses associated with the target URL.

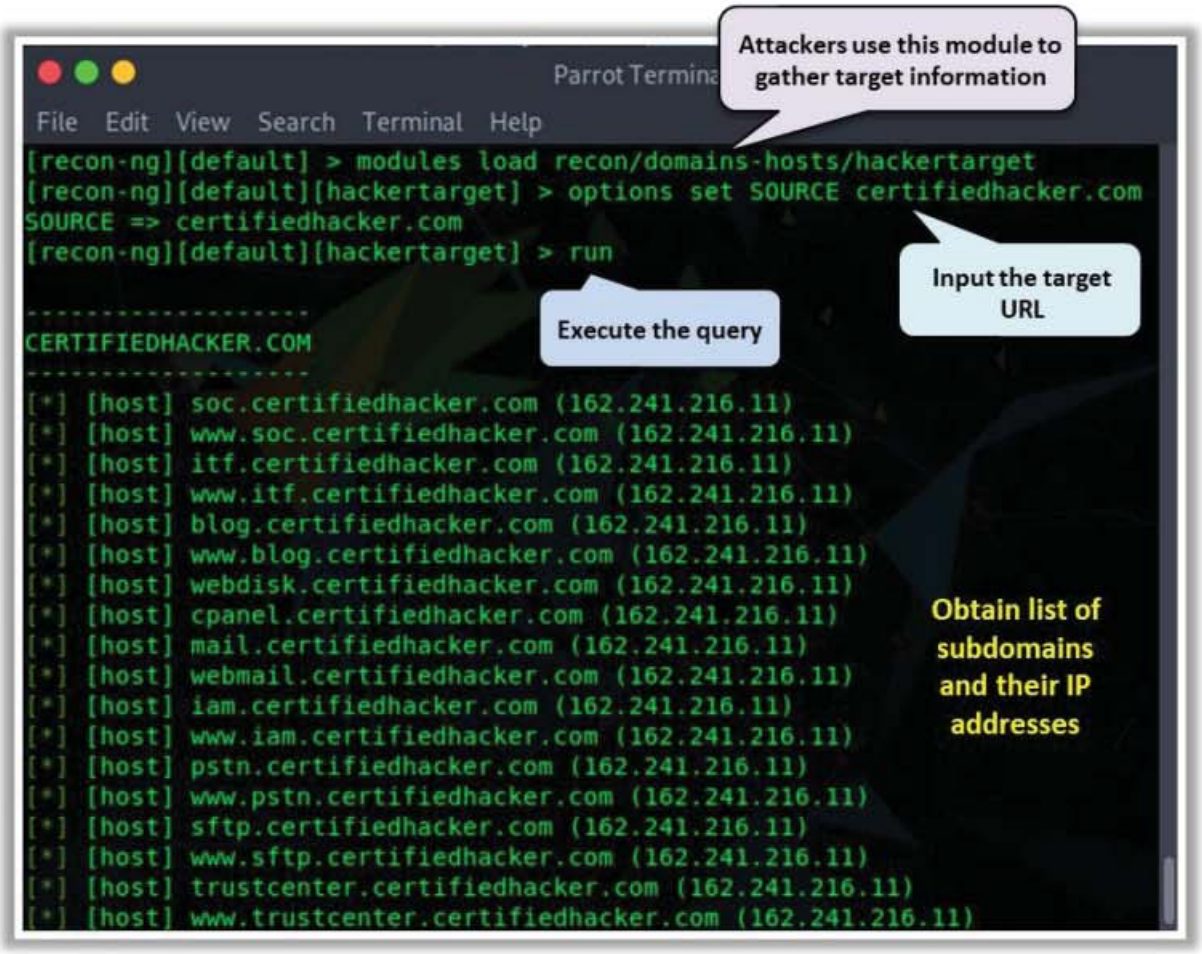


Figure 2.74: Screenshot of recon-ng

- **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.

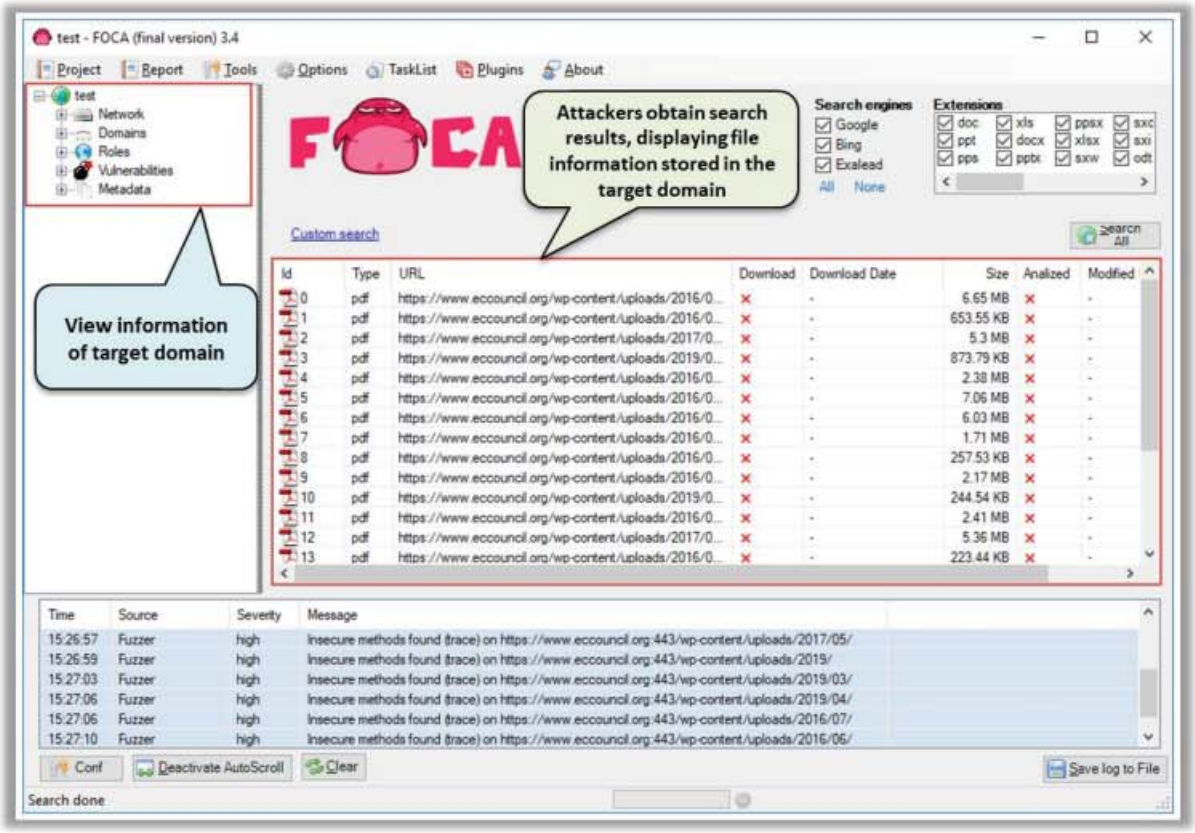


Figure 2.75: Screenshot of FOCA

- **OSRFramework**

  Source: *https://github.com*

  OSRFramework includes applications related to username checking, DNS lookups, information leaks research, deep web search, and regular expression extraction.

  The tools included in the OSRFramework package that attackers can use to gather information on the target are listed below:

  - **usufy.py** - Checks for a user profile on up to 290 different platforms
  - **mailfy.py** - Check for the existence of a given email
  - **searchfy.py** - Performs a query on the platforms in OSRFramework
  - **domainfy.py** - Checks for the existence of domains
  - **phonefy.py** - Checks for the existence of a given series of phones
  - **entify.py** - Uses regular expressions to extract entities

  As shown in the screenshot, attackers use the following command to search for a target user on social media platforms,

  usufy.py -n Mark Zuckerberg -p twitter facebook youtube

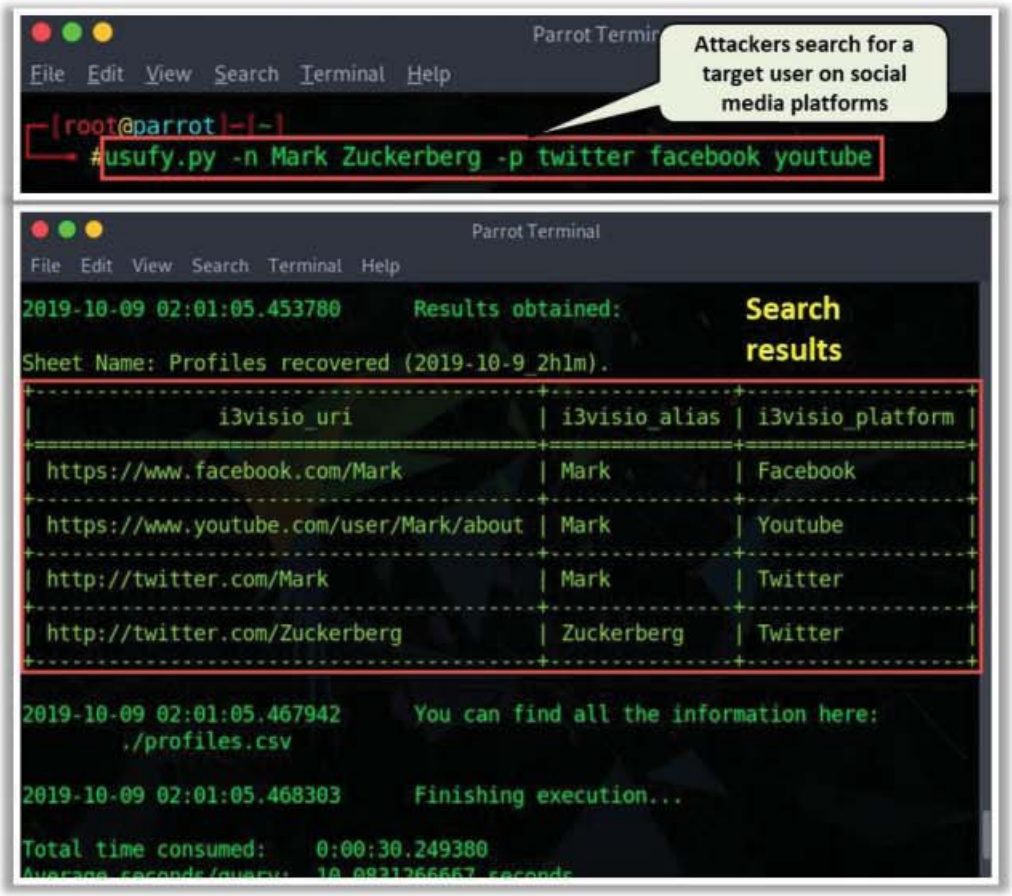


Figure 2.76: Screenshot of OSRFramework

- **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

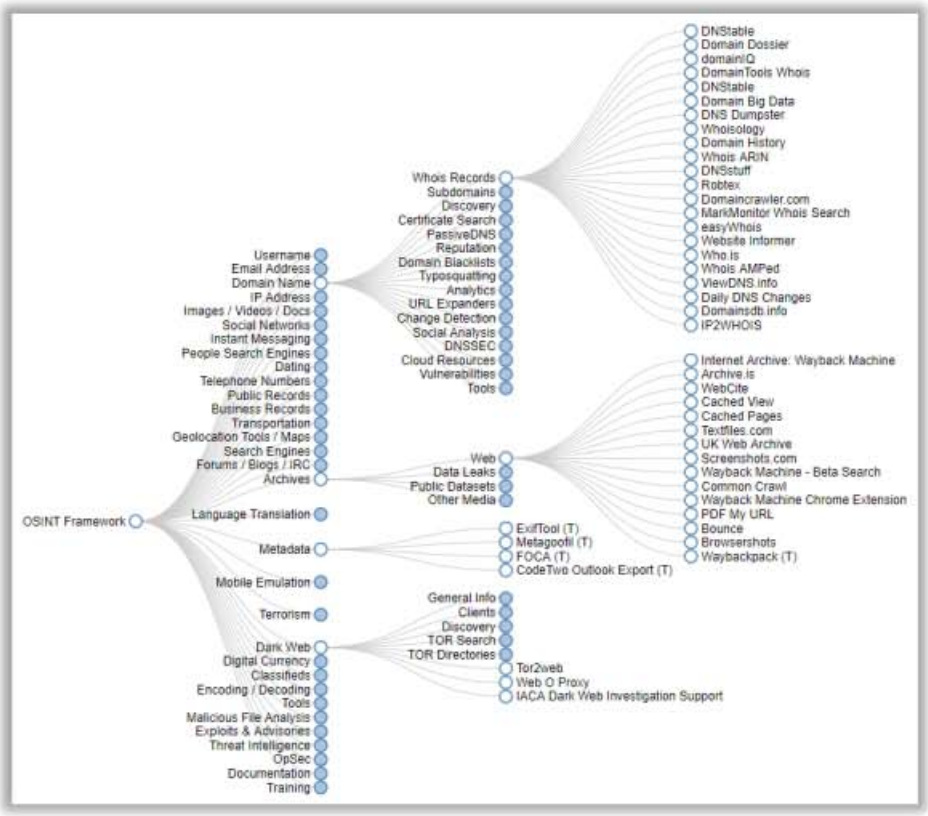


Figure 2.77: Screenshot of OSINT Framework

- **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 2.78: 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.

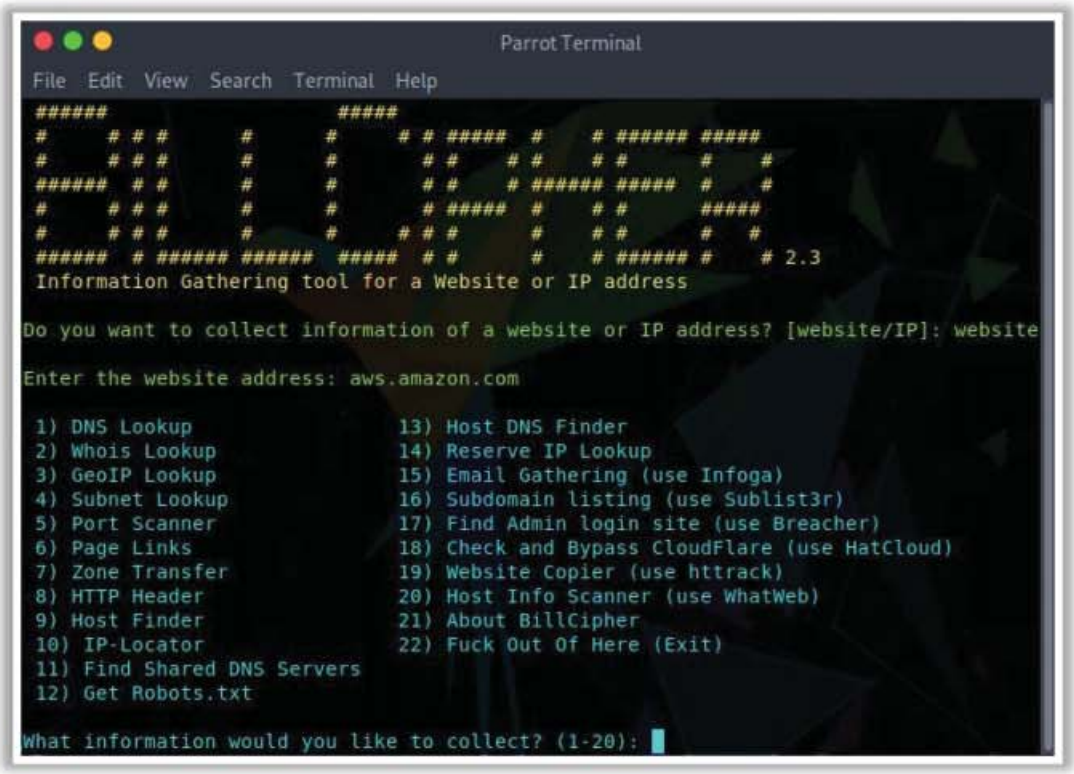


Figure 2.79: Screenshot of BillCipher

Some additional footprinting tools are listed below:

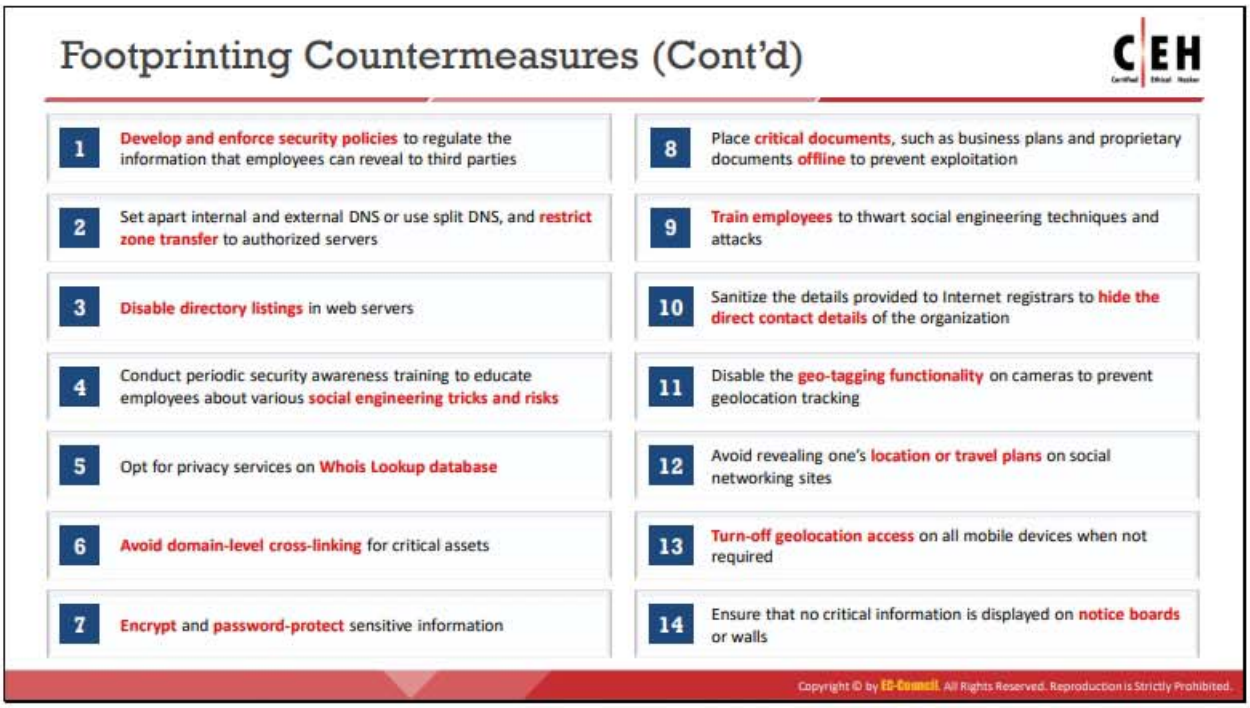
- theHarvester (*http://www.edge-security.com*)
- Th3Inspector (*https://github.com*)
- Raccoon (*https://github.com*)
- Orb (*https://github.com*)
- PENTMENU (*https://github.com*)

Module Flow

1. Footprinting Concepts
2. Footprinting Methodology
3. Footprinting Tools
4. 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 the website/Internet
- Use footprinting techniques to discover and remove any sensitive information publicly available
- Prevent search engines from caching a web page and use anonymous registration services



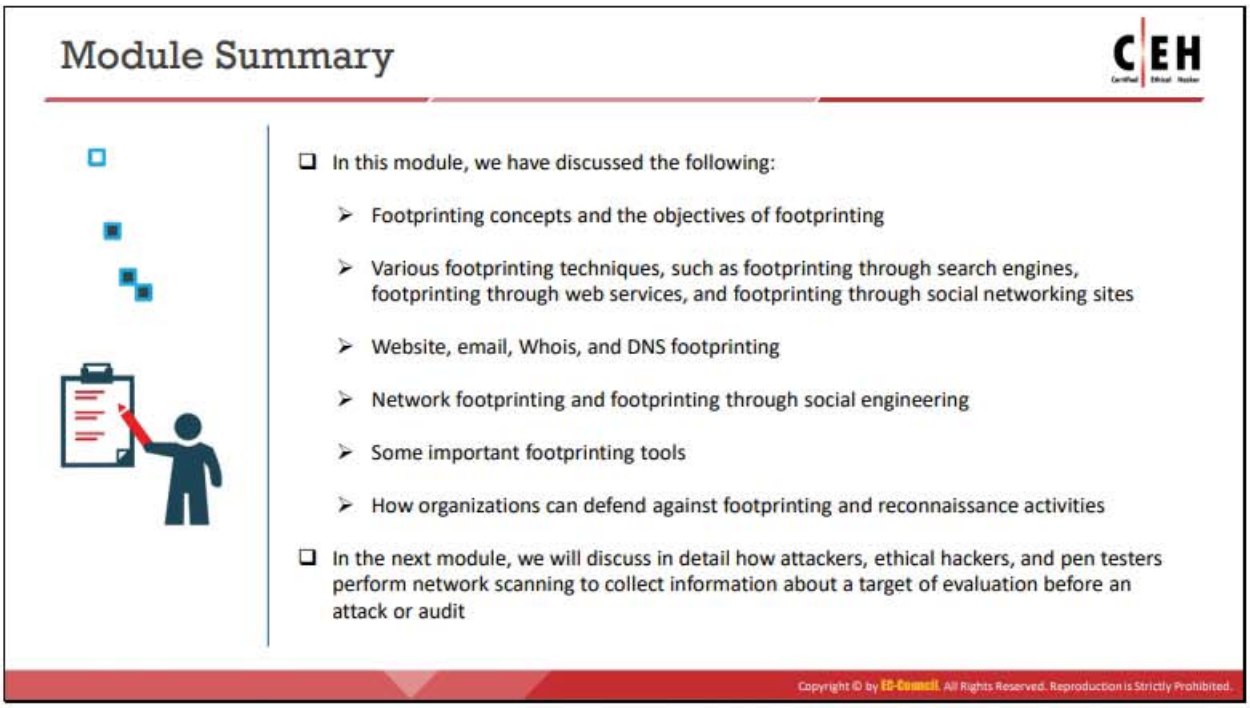
Footprinting Countermeasures

So far, we have discussed the importance of footprinting, various ways to perform the task, and the tools that help in its execution. Now, we will 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, and so on.
- Limit the amount of information that you are publishing on the website/Internet
- Use footprinting techniques to discover and remove any sensitive information publicly available
- Prevent search engines from caching a web page and use anonymous registration services
- Develop and enforce security policies such as information security policy, password policy, and so on, to regulate the information that employees can reveal to third parties
- 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 Whois lookup database
- Avoid domain-level cross-linking for critical assets
- Encrypt and password-protect sensitive information
- Do not enable protocols that are not required
- Always use TCP/IP and IPSec filters for defense in depth
- Configure IIS to avoid information disclosure through banner grabbing
- Hide the IP address and the related information by implementing 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, and sales projections is displayed on notice boards or walls



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 web services and social networking sites. In addition, it discussed website and email footprinting techniques. It also explained Whois and DNS footprinting in detail. Moreover, it described network footprinting along with traceroute analysis. 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.

EC-Council

EC-COUNCIL OFFICIAL CURRICULA