**Footprinting**

An essential aspect of footprinting is identifying the level of risk associated with the organization’s publicly accessible information. 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.

Types

* Active Footprinting
* passive Footprinting

The major objectives of footprinting include collecting the network information, system information, and organizational information of the target.

**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.

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

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

**Footprinting helps to:**

* Know security posture
* Reduce Focus Area
* Identify Vulnerabilities
* Draw Network Map

**🡺 Footprinting through Search Engines**

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

**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. 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.

Syntax: **operator: search\_term**

**Some popular Google advanced search operators include**

* **site:** [games site: www.certifiedhacker.com] query gives information on games from the certifiedhacker site.
* **allinurl:** [allinurl: google career] query returns only pages containing the words “google” and “career” in the URL.
* **inurl:** [inurl: copy site:www.google.com] query returns only Google pages in which the URL has the word “copy.”
* **allintitle:** [allintitle: detect malware] query returns only pages containing the words “detect” and “malware” in the title.
* **intitle:** [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:** [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:** [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:** [cache:www.eff.org] will show Google’s cached version of the Electronic Frontier Foundation home page.
* **link:** [link:www.googleguide.com] finds pages that point to Google Guide’s home page.

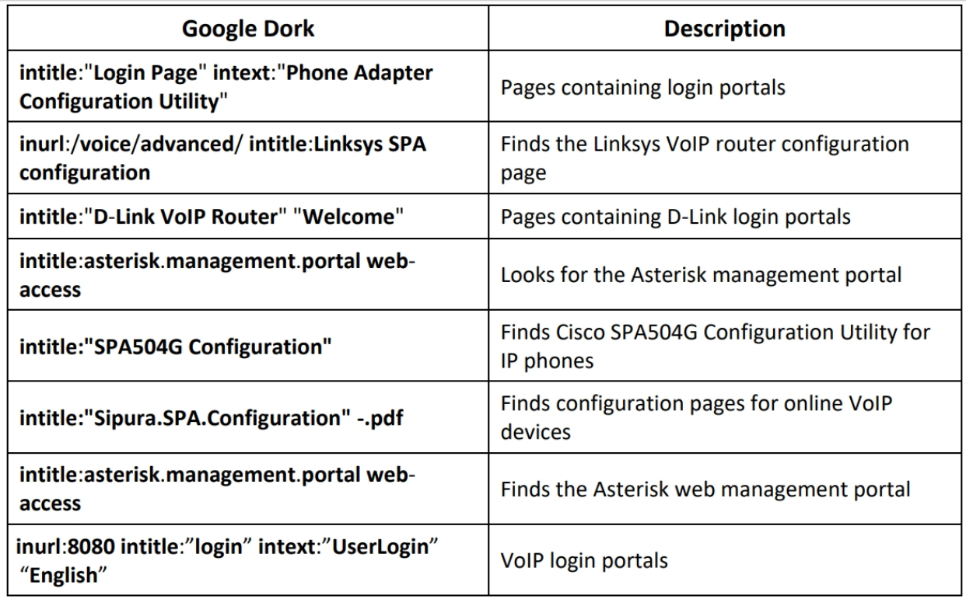
Note: “you cannot combine a link: search with a regular keyword search.”Also note that when you combine link: with another advanced operator, Google may not return all the pages that match.

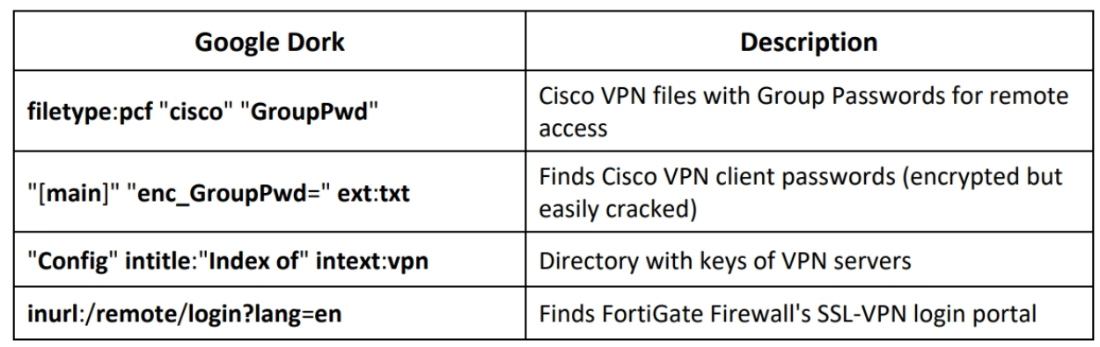
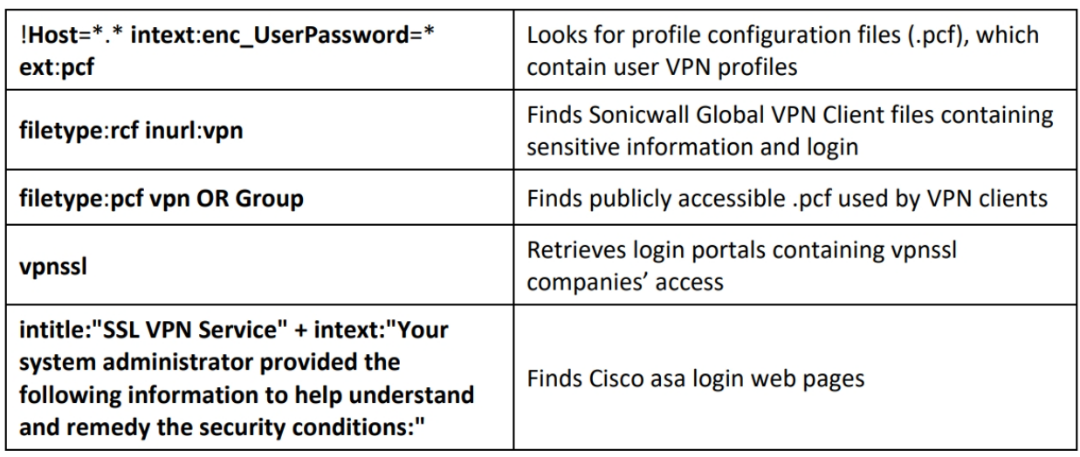
* **related:** [related:www.microsoft.com] provides the Google search engine results page with websites similar to microsoft.com.
* **info:** [info:gothotel.com] provides information about the national hotel directory GotHotel.com home page.
* **location:** [location: 4 seasons restaurant] will give you results based on the term “4 seasons restaurant.”
* **Filetype:** [jasmine:jpg] will provide jpg files based on jasmine.

**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.

**Google Hacking Database (GHDB**) is an authoritative source for querying the ever-widening scope of the Google search engine. 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 website.

**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.

**Google search queries for VoIP footprinting**

**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
* Gathering Information from Video Search Engine
* Gathering Information from Meta Search Engines
* Gathering Information from FTP Search Engines
* Gathering Information from IoT Search Engines

**🡺 Footprinting through Web Services**

**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.

**Tools to Search Company’s Sub-domains**

* **Netcraft:** Netcraft provides Internet security services, including anti-fraud and anti-phishing services, application testing, and PCI scanning.
* **Sublist3r:** 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.

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

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

**Tools for Finding the Geographical Location**

* Google Earth
* Google Map
* Wikimapia

**People Search on Social Networking Sites**

* Instagram
* Facebook
* LinkedIn
* Twitter

**People Search on People Search Services**

* Intelius: Attackers can use the Intelius people search online service to search for people belonging to the target organization.

**Gathering Information from LinkedIn**

**theHarvester:** It 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.

**Syntax: theHarvester -d microsoft -l 200 -b linkedin**

-d: domain or company name to search, -l: number of results, and -b” data source as LinkedIn.

**Harvesting Email Lists**

Attackers use theHarvester tool to extract email addresses related to the target domain.

**Syntax: theharvester -d microsoft.com -l 200 -b baidu**

-d: domain or company name to search, -l: number of results, and -b tells theHarvester to extract the results from the Baidu search engine; alternatively, you can use Google, Bing, etc.

**Gathering Information from Financial Services**

* Google Finance
* MSN Money
* Yahoo Finance
* Investing.com

**Deep and Dark Web Footprinting**

* Tor Browser

**Determining the Operating System**

The technique of obtaining information about the target network operating system is called OS fingerprinting.

* **Netcraft:** Attackers use the Netcraft tool to identify all the sites associated with the target domain along with the operating system running at each site.
* **Shodan:** It is a computer search engine that searches the Internet for connected devices It helps attackers to keep track of all the devices on the target network that are directly accessible from the Internet.
* **Censys:** It 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.

**VoIP and VPN Footprinting through SHODAN**

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.

**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.

**Sources of Competitive Intelligence**

* **Direct Approach:** trade shows, social engineering of employees and customers
* **Indirect Approach:** Company websites and employment ads, Support threads and reviews, Search engines, Internet, and online database, social media postings, Press releases and annual reports

**Information Resource Sites Information**

* **EDGAR:** 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.
* D&B Hoovers: 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: 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 D&B Hoovers.

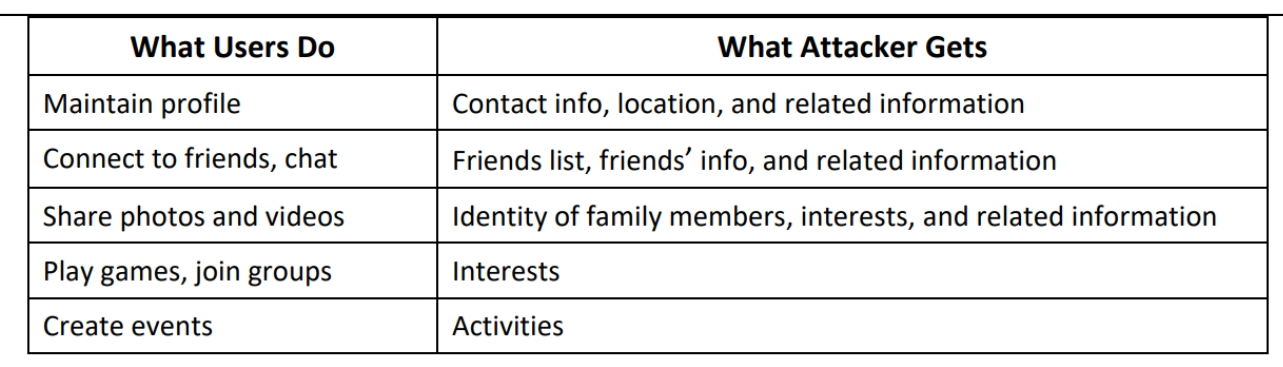
**Other Techniques for Footprinting through Web Services**

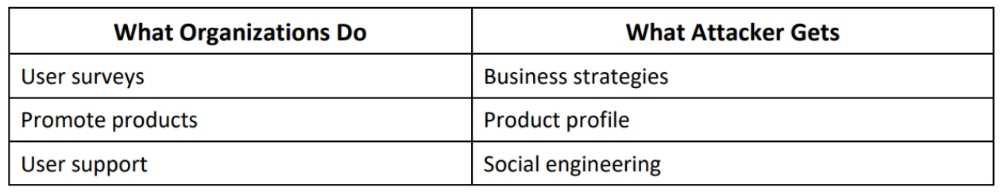
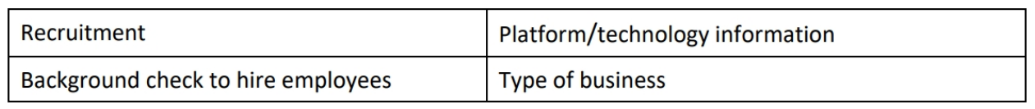
* Information Gathering Using Business Profile Sites
* Monitoring Targets Using Alerts
* Tracking Online Reputation of the Target: Online Reputation Management (ORM) is a process of monitoring displays when someone searches for your company’s reputation on the Internet.
* Information Gathering Using Groups, Forums, and Blogs
* Information Gathering Using Network News Transfer Protocol (NNTP) Usenet Newsgroups

🡺 **Footprinting through Social Networking Sites**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.

**Collecting Information through Social Engineering on Social Networking Sites**

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.

**Information Available on Social Networking Sites**



**General Resources for Locating Information from Social Media Sites**

* BuzzSumo: It is an advanced social search engine finds the most shared content for a topic, author, or domain.
* Google Trends
* Hashatit
* Ubersuggest

**Conducting Location Search on Social Media Sites**

* Followerwonk: It 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?
* Hootsuite
* Sysomos

**Tools for Footprinting through Social Networking Sites**

* Sherlock: 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.
* Social Searcher: Social Searcher allows attackers to search for content in social networks in real time and provides deep analytics data.
* UserRecon

**🡺 Website Footprinting**

Website footprinting refers to monitoring and analyzing a target organization’s website for information. 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. 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.

**Browsing the target website will typically provide the following information:**

* Software used and its version
* Operating system used
* Sub-directories and parameters
* Filename, path, database field name, or query
* Scripting Platform
* Technologies used
* Contact details and CMS details

**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

**Website footprinting can be performed by:**

* **Examining The HTML Source Code:** 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.
* **Examining Cookies:** Identify the scripting platforms by observing sessions and other supporting cookies. The information about cookie name, value, and domain size can also be extracted.

**Website Footprinting using Web Spiders**

A web Spider is a program or automated script that browses websites in a methodical manner to collect specific information. 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.

**User Directed Spidering**

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.

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

**Mirroring Entire Website**

Website mirroring is the process of creating a replica or clone of the original website.

**HTTrack:** It 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.

**Extracting Website Links**

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.

**Gathering Wordlist from the Target Website**

To run the CeWL tool, issue the following commands:

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

command displays options that user can use to obtain 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.

**Extracting Metadata of Public Documents**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.

**Other Techniques for Website Footprinting**

* Monitoring Web Pages for Updates and Changes
* Searching for Contact Information, Email Addresses, and Telephone Numbers from Company Website
* Searching for Web Pages Posting Patterns and Revision Numbers
* Monitoring Website Traffic of Target Company

**🡺 Email Footprinting**

**Tracking Email Communication**

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.

**Information about the victim gathered using email tracking tools includes:**

* Recipient's System IP address
* Email Received and Read
* Read Duration
* Proxy Detection: Provides information about the type of server used by the recipient
* Links
* Operating System and Browser information
* Forward Email
* Device Type
* Path Travelled

**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.

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
* ender’s IP address and address from which the message was sent

**Email Tracking Tools**

* **eMailTrackerPro**
* **Infoga:** 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.
* **python infoga.py --domain microsoft.com --source all --breach -v 2 --report ../microsoft.txt**

command 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**

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

**🡺 Whois Footprinting**

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.

**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

**The RIRs include:**

* ARIN (American Registry for Internet Numbers)
* AFRINIC (African Network Information Center)
* APNIC (Asia Pacific Network Information Center)
* RIPE (Réseaux IP Européens Network Coordination Centre)
* LACNIC (Latin American and Caribbean Network Information Center)

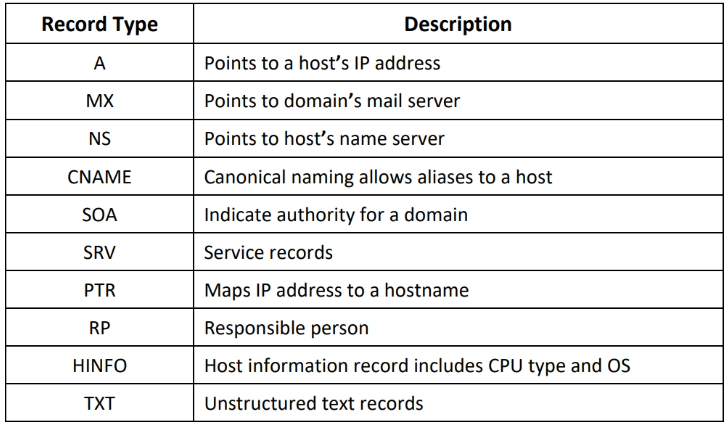
**Finding IP Geolocation Information**

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.

**🡺 DNS Footprinting**

**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:**

**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.

* **Reverse IP Domain Check**
* **DNSRecon**

**dnsrecon -r 162.241.216.0-162.241.216.255**

-r option specifies the range of IP addresses (first-last) for a reverse lookup by brute force.

**🡺 Network Footprinting**

**Locate the Network Range**

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.

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).

**Traceroute**

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, 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.

**ICMP Traceroute**

Windows operating system by default uses ICMP traceroute.

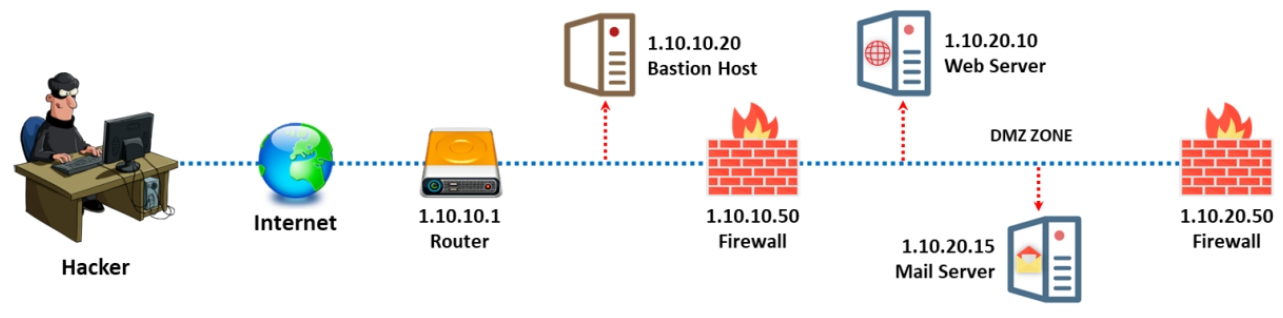
**tracert Ipaddress**

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.

**TCP Traceroutetcptraceroute www.google.com**

**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. **traceroute www.google.com**

**Traceroute Analysis**



**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.

**🡺 Footprinting through Social Engineering**

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.

**Collecting Information Using**

* **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.
* **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.
* **Dumpster diving** 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 from the trash
* **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.

**🡺 Footprinting Tools**

* **Maltego**: It 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.
* Recon-ng: It 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. Attackers use the module recon/domains-hosts/hackertarget to extract a list of subdomains and IP addresses associated with the target URL.
* FOCA: Fingerprinting Organizations with Collected Archives (FOCA) is a tool used mainly to find metadata and hidden information in the documents that its scans.

Features: Web Search, DNS Search, IP Resolution, PTR Scanning, Bing IP, Common Names

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

Attackers use the following command to search for a target user on social media platforms, **usufy.py -n Mark Zuckerberg -p twitter facebook youtube**

* OSINT Framework: It is an open source intelligence gathering framework that helps securityprofessionals in performing automated footprinting and reconnaissance, OSINT research, and intelligence gathering.

Features: Censys, NS Lookup, Port Scan, Detect CMS, Whois lookup, Detect honeypot, Find Subdomains, Reverse IP lookup, Detect technologies

* BillCipher: It 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.