# Information Gathering – Foot printing and Reconnaissance

## 1. Introduction to Information Gathering

This is the **first and most critical phase** of security assessment (or an attack). The goal is to gather as much information as possible about a target system, network, or organization without yet launching an attack. The more information we have the more effective and targeted our later steps will be.

## 2. Key Concepts & Their Relationship

There are several methods. It's important to understand how they relate to each other. **Foot printing and Reconnaissance** are often used interchangeably as the overall phase name, while **Scanning** and **Enumeration** are most specific, technical sub-steps that follow.

- **Foot printing/Reconnaissance (The "What"):** The broad process of collecting publicly available information to create a profile of the target. It's about understanding the target's digital foot print.
- **Scanning (The "Where"):** Using technical tools to discover live systems, open ports, and running services on the target's network. You are moving from "What exists" to "Where it is".
- **Enumeration (The "How"):** Actively querying the systems and services discovered during scanning to extract more detailed information, such as user names, group names, shares, and other network resources.

## 3. Active vs Passive Techniques

| Feature | Passive Reconnaissance | Active Reconnaissance |
|---|---|---|
| **Definition** | Gathering information **without directly interacting** with the target. Using third-party sources. | Gathering Information by **directly interacting** with the target's system. |
| **Interaction** | No packets are sent to the target | Packets are sent to the target |

| | | |
|---|---|---|
| **Stealth Level** | **Very Stealthy.** The target is unaware they are being probed. | **Less Stealthy.** The interaction can be logged by firewalls and IDS/IPS |
| **Speed** | Slower, as you rely on cached or archived data | Faster, as you get real-time, accurate data |
| **Risk of Detection** | **Very Low.** | **High.** |
| **Examples** | Searching Google, viewing social media checking WHOIS records. | Ping sweeps, pot scanning (nmap), DNS queries directly to targets server. |

From a **security defender's point of view,** Passive Reconnaissance is a bigger concern in the initial stages because it's almost impossible to detect or prevent. An attacker can gather huge amount of data without you ever knowing.

## 4. OSINT (Open-Source Intelligence)

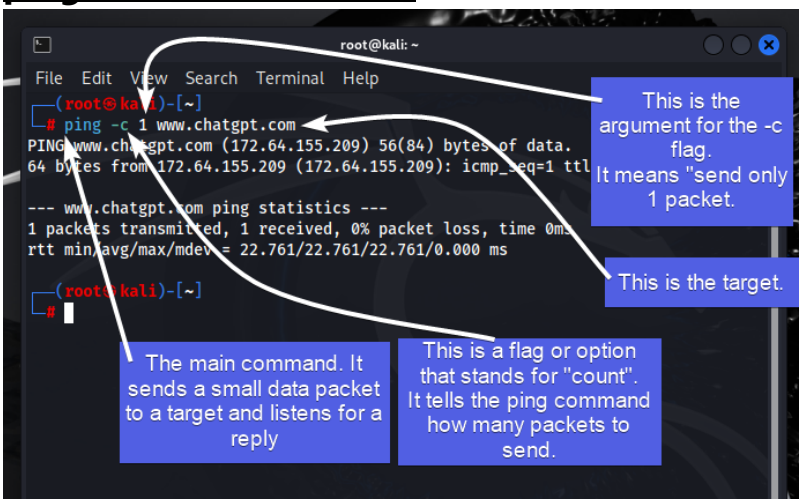
This is the cornerstone of the foot printing phase. It means collecting information from publicly available sources. The information was posted by the target or about the target, often without realizing its value to an attacker.

## 5. Practical Steps & Commands for Foot printing

**Step 1:** Select a Target (e.g., a website)

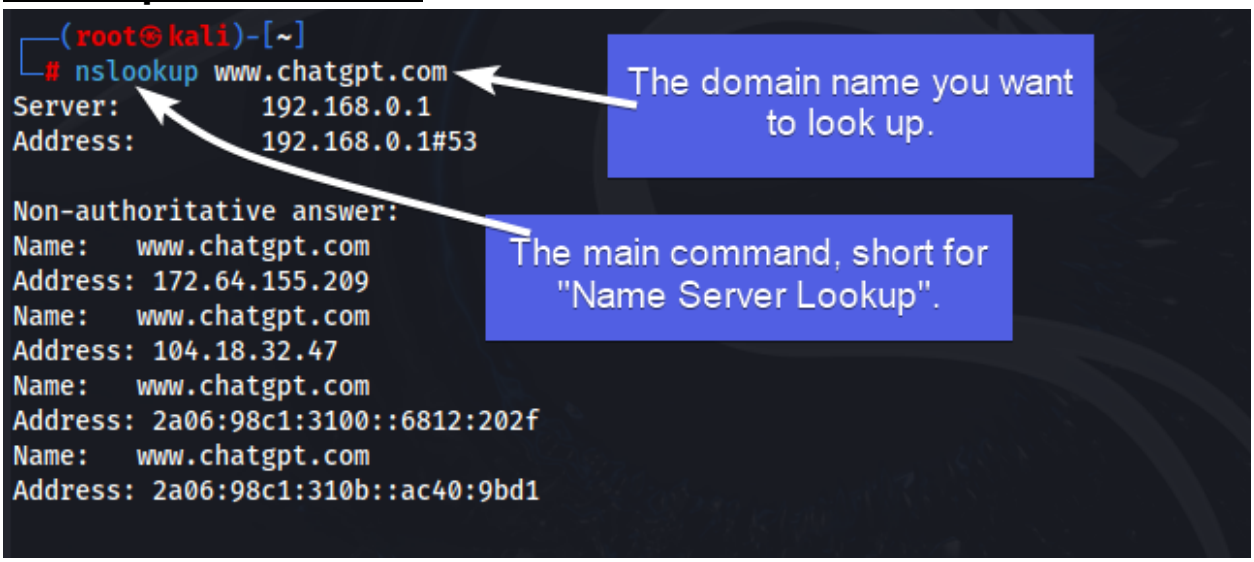
**Step 2:** Discover IP Address and Network Information.

- **ping -c 1 [website.com]**



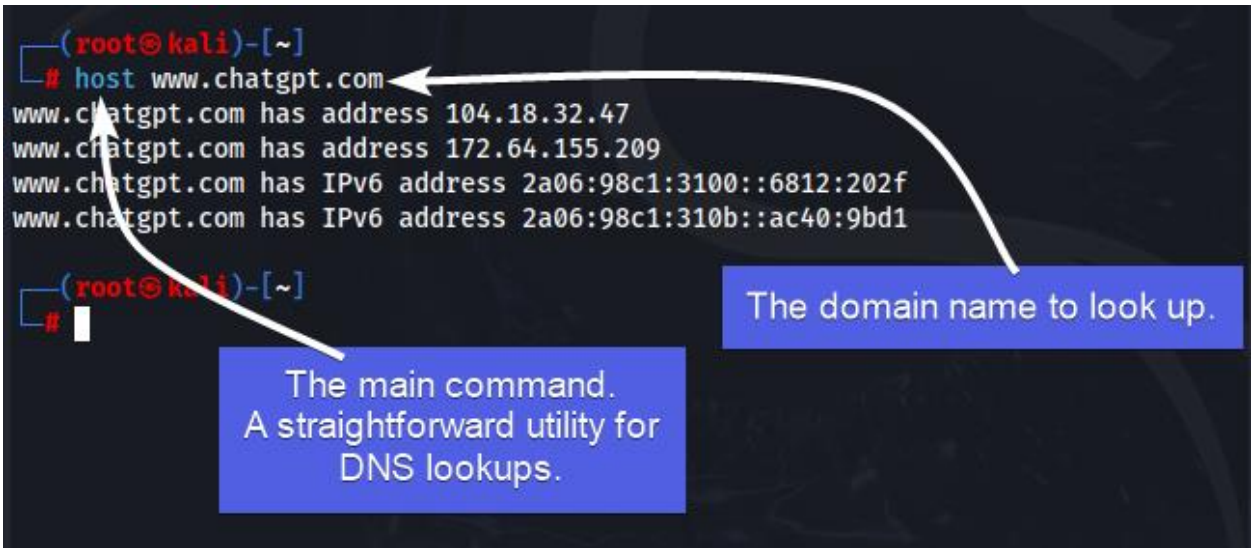
- **What it does:** Sends a single ICMP packet to the host. The reply shows the IP address and round trip time.
- **Note:** This is an **Active** Technique. Many modern networks block ICMP, so it may not always work
- **ICMP:** ICMP (Internet Control Message Protocol) is a network protocol used by devices to send error messages and operational information, like when a host is unreachable or to test connectivity with **ping**.

- **nslookup [website.com]**



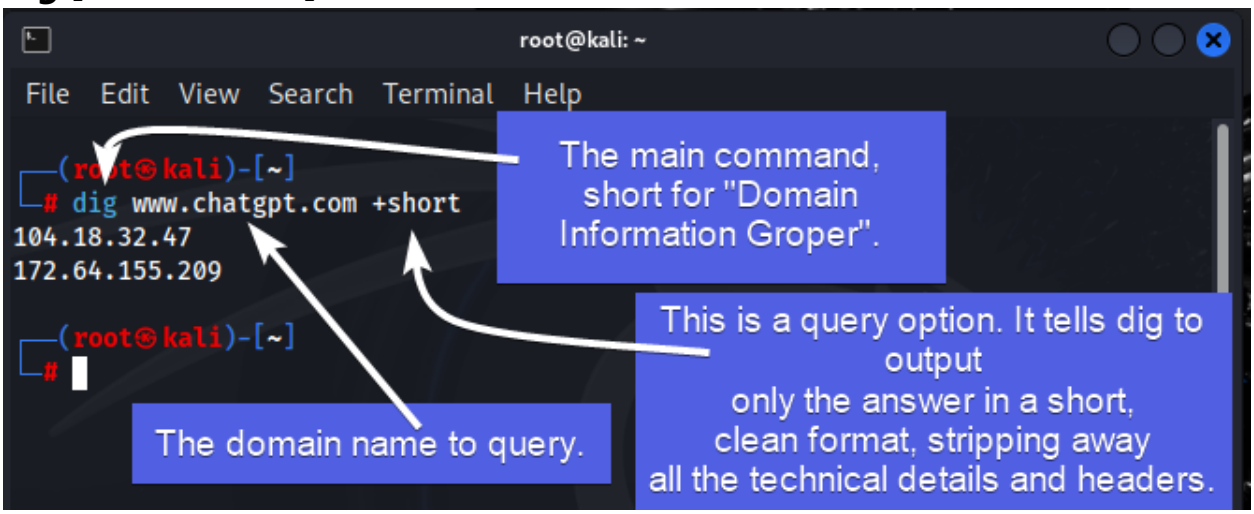

- **What it does:** A standard tool for querying the Domain Name System (DNS) to get the IP address and other DNS records.

- **host [website.com]**



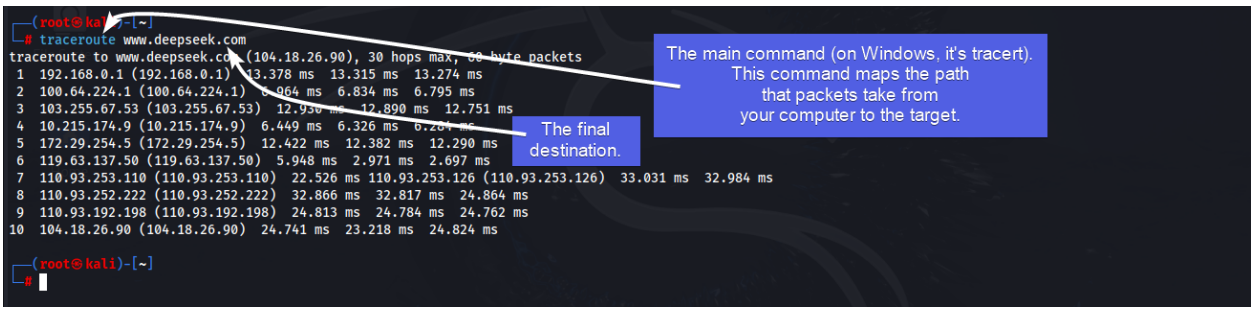

- **What it does:** A simpler, Linux-based alternative to nslookup for performing DNS lookups.

- **dig [website.com] +short**



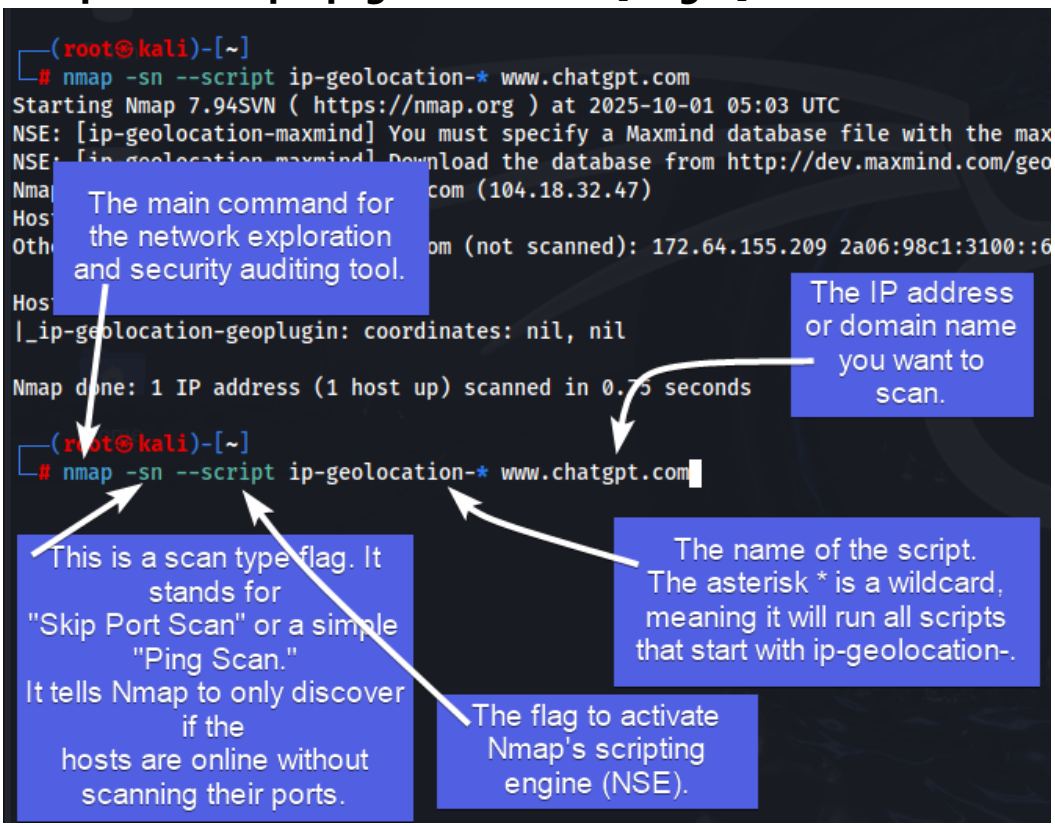

  - **What it does:** A powerful DNS lookup tool, preferred by many professionals. The +short option gives a clean, abbreviated output (just the IP).
  - **Note:** Without +short, dig provides a very verbose output with sections like HEADER, QUESTION, ANSWER, and AUTHORITY, which is useful for debugging but overwhelming for a quick lookup.

- **traceroute [website.com] (or tracert on Windows)**




  - **What it does:** Shows the path (the "hops") that packets take from your machine to the target.
  - **The more hops there are:** Each hop is a router. More hops generally mean a longer path, which can lead to higher latency (lag).

- **nmap -sn --script ip-geolocation-* [target]**




  - **What it does:** Check if this target is online, and if it is, run all your geolocation scripts to try and find its physical location on a map.
- **nmap -sn --script traceroute-geolocation [target]**
  - **What it does:** Similar to the above but attempts to geolocate every hop in the traceroute path.

## 6. Website Registration Details (WHOIS)

This is a critical passive reconnaissance step. A WHOIS query tells you who registered the domain name.

# Information Gathering - Passive Reconnaissance Techniques

## 1. Active vs. Passive Commands: A Critical Clarification

**Correction and Clarification:** While we used command-line tools yesterday, the technique (Active or Passive) is defined by how and from where we use them, not just by the tool itself.

- **Active Reconnaissance (Direct):** When you run ping, nslookup, or traceroute directly from your terminal, you are performing active reconnaissance. Your IP address sends packets directly to the target, creating a log in their firewall and revealing your presence.
- **Passive Reconnaissance (Indirect):** When you use a third-party website (like CentralOps.net) to run these same commands, you are performing passive reconnaissance.

**Analogy:**

- **Active:** You personally go to a store and ask the manager questions (they see you).
- **Passive:** You send a friend to ask the questions (the store sees your friend, not you).

## 2. Passive Reconnaissance Tools & Techniques

### CentralOps.net

- **Concept:** A free online toolkit that acts as your "proxy" for information gathering.
- **How it works:**
  You enter a target (e.g., example.com) on the CentralOps website.
  Your IP → CentralOps.net → Target Website
  The target only sees the request coming from CentralOps.net's IP address, not yours.
- **Use Case:** Performing DNS lookups, ping scans, and traceroutes without revealing your own IP address to the target.

### Archive.org (The Wayback Machine)

- **Purpose:** To view historical versions of a website.
- **Why it's useful for Pentesters:**

    - Find old, exposed files (e.g., robots.txt, admin login pages) that are no longer on the live site.
    - Discover technical information the company may have accidentally published in the past (server types, software versions, employee names).
    - Understand the evolution of the company's web infrastructure, which can reveal legacy systems that might be vulnerable.
- **Command Equivalent:** This is a purely passive technique with no direct command-line equivalent.

## Netcraft.com

This is a powerful suite of tools for passive reconnaissance.

### A. Site Report (sitereport.netcraft.com):

- **What it provides:**
    - **Hosting History:** Which company hosts the website and where the server is located.
    - **DNS History:** How the domain's DNS records have changed over time.
    - **Technographic Profile:** The underlying technologies powering the site (e.g., WordPress, Apache, nginx, specific JavaScript libraries).
    - **Security Assessment:** Netcraft's own risk rating for the site.

### B. DNS Search (searchdns.netcraft.com):

- **Purpose:** To find other domains and subdomains owned by the same organization that don't appear in regular Google searches.
- **Why it's Critical:** Attackers often find vulnerabilities in smaller, less-secure "shadow IT" systems or development/staging subdomains (e.g., dev.company.com, test.company.com, admin.company.com) and use them to pivot into the main network.
- **How it works:** It searches Netcraft's extensive database of web data for domains hosted on the same network block or registered under the same organization.

# Advanced Reconnaissance Techniques

## 1. Website Mirroring with wget:

### Command:

wget --mirror --convert-links --adjust-extension --no-parent --page-requisites corvit.com

### Breakdown and Explanation:

- **wget:** The main command - "Web Get" - a powerful tool for downloading content from the web.
- **--mirror:**
  - **What it does:** Turns on options suitable for mirroring an entire website.
  - **Effect:** Recursively downloads the entire website structure.
- **--convert-links:**
  - **What it does:** After downloading, converts the links in the HTML files to work locally for offline viewing.
  - **Why it's important:** Makes the mirrored site actually usable on your local machine.
- **--adjust-extension:**
  - **What it does:** Adds proper file extensions (like .html) to files if they're missing.
  - **Example:** A file called about would become about.html.
- **--no-parent:**
  - **What it does:** Prevents wget from ascending to the parent directory when downloading recursively.
  - **Why it's important:** Keeps the download focused only on the target domain and doesn't wander to unrelated sites.
- **--page-requisites:**
  - **What it does:** Downloads all necessary files to display the page properly (images, CSS, JavaScript).
  - **Without this:** You'd get only the HTML, but the site would look broken.
- **corvit.com:** The target website to mirror.

## Practical Use Case:

**This creates a perfect offline copy of the entire website for:**

- Offline analysis of source code
- Finding hidden content not easily visible through browsing
- Preserving evidence of the current site state
- Testing without network connectivity

## 2. Email and Domain Intelligence with theHarvester:

### Command:

theHarvester -d corvit.com -b yahoo,bing

### Breakdown and Explanation:

- **theHarvester:** A powerful OSINT tool specifically designed for gathering emails, subdomains, hosts, and employee names.
- **-d corvit.com:**
  - **-d:** Stands for "domain"
  - **corvit.com:** The target domain to investigate
- **-b yahoo,bing:**
  - **-b:** Stands for "bdata source" (or "backend")
  - **yahoo,bing:** Specifies which search engines to use for gathering information

### What it gathers:

- Email addresses associated with the domain
- Subdomains
- Hosts/IP addresses
- Employee names (from public sources)

### Professional Use:

Perfect for building target profiles during penetration testing or red team exercises.

## 3. Web Application Firewall Detection with wafw00f:

### Command:

wafw00f corvit.com

### Breakdown and Explanation:

- **wafw00f:** "Web Application Firewall Finder" - a tool that detects and identifies Web Application Firewalls.
- **corvit.com:** The target website to check for WAF protection.

### How it works:

- Sends specially crafted HTTP requests to the target
- Analyzes the responses for WAF fingerprints
- Identifies specific WAF products like:
  - Cloudflare
  - Akamai
  - Imperva
  - ModSecurity
  - AWS WAF

### Why it matters:

- Knowing the WAF helps tailor attack vectors
- Different WAFs have different bypass techniques
- Critical for planning web application penetration tests

## 4. IP Geolocation & Network Tools

### A. www.nirsoft.net/countryip

**What it is:** A website by NirSoft that provides IP address ranges organized by country.

**Key Features:**

- **Country IP Blocks:** Shows all IP ranges assigned to specific countries
- **Useful For:**
  - Geolocation analysis
  - Identifying where a company's infrastructure is hosted
  - Understanding the geographic footprint of a target

**Format:** Provides IP ranges in CIDR notation (e.g., 192.168.0.0/16)

### B. networksdb.io

**What it is:** An online network database and IP intelligence platform.

**Key Capabilities:**

- **IP to Company Mapping:** Find what organization owns an IP range
- **Network Reconnaissance:** Discover related networks and domains
- **ASN Lookup:** Autonomous System Number information
- **Reverse DNS Lookup:** Find domains hosted on the same IP
- **BGP Routing Information:** Network path analysis

**Professional Use:** Essential for mapping out a target's entire network infrastructure during external penetration tests.

| Tool | Primary Purpose | Reconnaissance Type |
|---|---|---|
| wget | Website mirroring & offline analysis | Passive |
| theHarvester | Email, subdomain & employee discovery | Passive |
| wafw00f | Web Application Firewall detection | Active |

| Tool | Primary Purpose | Reconnaissance Type |
|---|---|---|
| NirSoft CountryIP | IP range geolocation data | Passive |
| NetworksDB | Network infrastructure mapping | Passive |