Information Gathering

Information gathering, also referred to as reconnaissance, is the **initial** and one of the most critical phases in hacking and penetration testing. During this phase, an attacker collects as much data as possible about a target system, organization, or individual to identify potential attack vectors or vulnerabilities.

This phase lays the groundwork for all subsequent activities in a penetration test. It helps in understanding the target's infrastructure, identifying weak points, and planning a strategy for testing. Proper information gathering can mean the difference between a successful penetration test and a failed one.

Types of Information Gathering

1. Passive Information Gathering

Definition: Involves collecting information without directly interacting with the target system.

Examples:

Searching publicly available data (e.g., Google, social media, forums).

Examining DNS records or IP address blocks.

Scraping leaked data repositories.

Tools:

WHOIS: Retrieves domain registration information and other details about a target domain or IP address.

```
whois [options] <domain_name> or <IP_address>
```

nslookup: querying the Domain Name System (DNS) to obtain domain name or IP address mapping information.

```
nslookup [options] [hostname | IP_address]
```

dig(Domain Information Groper): versatile DNS query tool used to gather detailed information about DNS records.

```
dig [@server] [domain] [type] [options]
```

Shodan: a powerful search engine designed to discover internet-connected devices, including servers, routers, IoT devices, webcams, and industrial control systems.

Google Dorking: a technique that uses advanced search operators in Google to find specific information that is not easily accessible through conventional searches.

Common Google Dorking Examples

- Finding Login Pages

  ```
  inurl:login
  ```

  Returns pages with "login" in the URL, potentially exposing login portals.

- Discovering Exposed Files

  ```
  filetype:pdf site:example.com
  ```

  Searches for PDF files on the specified domain.

- Identifying Exposed Sensitive Directories

  ```
  intitle:"index of" "backup"
  ```

  Returns directories labeled "index of" containing "backup," often indicating publicly accessible directories.

- Exposed Databases

  ```
  filetype:sql site:example.com
  ```

  Searches for SQL database files on a specific site.

2. Active Information Gathering

Definition: Involves interacting with the target system to extract more detailed information.

Examples:

Scanning open ports.

Conducting banner grabbing to identify services and versions.

Testing endpoints of APIs or web applications.

Tools:

Nmap(Network Mapper): a versatile network scanning utility used to discover hosts, services, and vulnerabilities in a network.

```
nmap [Scan Type(s)] [Options] <Target>
```

Metasploit: a powerful penetration testing framework used for developing and executing exploit code against a target machine.

Netcat(nc): a versatile networking tool used for tasks such as port scanning, transferring files, and setting up reverse or bind shells.

```
nc [options] [hostname] [port]
```

Burp Suite: a comprehensive toolset for web application security testing. It includes features for intercepting traffic, scanning for vulnerabilities, and testing manually.

FAQ’s

1. Why Information gathering is the first step?
   Information gathering is the first step because it helps ethical hackers understand the target's infrastructure, identify potential vulnerabilities, and plan a strategic approach for testing. Without this phase, the testing process would lack focus and effectiveness.

2. What is the difference between **reconnaissance** and **footprinting**?
   - **Reconnaissance** is a broader term that refers to the entire information gathering process, including both passive and active techniques used to gather data about the target.
   - **Footprinting** is a specific part of reconnaissance, focused on gathering detailed information about the target’s network infrastructure, such as IP addresses, DNS records, and domain names. Footprinting is typically more about creating a map of the target’s online presence.