HO# 2.2: Reconnaissance, Info Gathering & OSINT

Overview (Phase 1- Reconnaissance and Information Gathering)

The Information gathering phase (reconnaissance) is the initial step in the penetration testing lifecycle. This phase involves collecting as much public information as possible about the organization, systems, networks, applications, and employees to identify potential vulnerabilities and formulate a strategy for further testing. Information gathering can be divided into two main categories:

Passive Information Gathering (Reconnaissance): Involves collecting data about a target without direct interaction, reducing detection risk. It relies on Open-Source Intelligence (OSINT) from public sources like news, blogs, and social media. Techniques include Web Scraping, Google Dorking, and social media profiling. Tools used include netdiscover, traceroute, host, nslookup, dig, whois, whatweb, theHarvester, sherlock, knockpy, wfw00f, and OSINT framework.

Active Information Gathering (Scanning): Involves direct interaction with the target to discover open ports, services, and vulnerabilities, often leaving traces on the system. It's typically used in the second phase of penetration testing. Tools like nmap are commonly used. Active scanning requires written permission from the system owner.

Host: A DNS lookup utility used to convert domain names to IP addresses and vice versa. To install DNS utilities on Linux**, use:** $ sudo apt-get install dnsutils

Exmaple: $ host arifbutt.me

Nslookup: A versatile tool for DNS lookups, used to map domain names to IP addresses and retrieve specific DNS records like A, AAAA, MX, NS, and TXT.

Example usage: $ nslookup arifbutt.me

Reverse DNS Lookup: Resolves an IP address to its domain name.

Example: $ nslookup 68.65.120.238

Dig: A DNS lookup tool (Domain Information Groper) similar to nslookup but provides more detailed and structured output. It's commonly used to query DNS servers and troubleshoot DNS issues. **Example usage:** $ dig google.com

Whois: A command used to retrieve domain registration information from whois databases, including registrar details, registration and expiration dates, name servers, and contact information for domain owners. This is useful for checking the availability of domain names.

Example for domain lookup: $ whois google.com

Note: Students should visit the following online web services to get information about a specific domain at their own time:

https://whois.domaintools.com/

https://centralops.net/co/

https://ipinfo.io/

Using Whois with IP Addresses: When an IP address is passed, it provides details about the organization managing that IP, including the IP range, organization information, and abuse contacts. **Example for IP lookup**: $ whois 8.8.8.8

Knockpy: An open-source tool for subdomain enumeration that identifies subdomains associated with a target domain by sending requests and collecting responses. It's useful for penetration testing and OSINT activities.

Example for subdomain enumeration: $ knockpy -d pu.edu.pk --recon --bruteforce --threads 50

Netdiscover: An active/passive network discovery tool that uses Address Resolution Protocol (ARP) to identify hosts in a local area network (LAN). It can perform both active and passive scanning.

Active Scanning: By default, Netdiscover performs active scanning by sending ARP requests to every IP in a specified range. Use the `-r` option to define the range.

Example: $ sudo netdiscover -r 10.0.2.0/8

Passive Scanning: To avoid detection, use the `-p` option to listen to network traffic without sending requests. It outputs results when network activity occurs.

Example: $ sudo netdiscover -p -r 10.0.2.0/8

Traceroute: A command used to trace the path that packets take from your device to a remote server (domain or IP address). It lists the routers or gateways (hops) that packets pass through, showing the round-trip time (RTT) for each hop.

To use ICMP echo requests instead of the default UDP packets, use the `-I` option.

Example: $ traceroute -I arifbutt.me

Whatweb: A tool used to identify web technologies on a target website, including server software, content management systems (CMS), frameworks, libraries, and plugins.

Usage: To analyze a website and view detailed information: $ whatweb -v pucit.edu.pk

Output Information:

- **HTTP Headers**: Displays server type and metadata.
- **Web Server Information**: Identifies server software (e.g., Apache, Nginx).
- **CMS Detection**: Detects CMS like WordPress or Joomla.
- **Frameworks and Libraries**: Identifies frameworks and JavaScript libraries.
- **Plugins and Extensions**: Lists detected plugins/extensions.

Aggressive Scan on IP Range: Use the `-a` option to set the aggression level when scanning an IP range with Whatweb. To ignore errors from non-existent addresses, use the `--no-errors` option. **Example:** $ whatweb -v -a 3 10.0.2.1-10.0.2.254

TheHarvester: A command-line utility for gathering open-source intelligence (OSINT) about targets, including domain names, IP addresses, and email addresses. It is commonly used in the information-gathering phase of penetration tests and security audits.

Key Features:

- **Email Address Gathering**: Collects email addresses associated with a domain.
- **Subdomain Enumeration**: Identifies subdomains of a target domain using various search engines.
- **IP Address and Hostname Discovery**: Retrieves related IP addresses and hostnames.

Example: Specify the domain with `-d`, limit search results with `-l`, and choose data sources with `-b`.

```
$ theHarvester -d pucit.edu.pk -l 100 -b yahoo,baidu
```

Sherlock: A command-line tool designed to search for usernames across various social media platforms and websites. It's widely used in OSINT (Open-Source Intelligence) investigations to find profiles associated with a specific username, aiding in social engineering or information gathering.
Example: To search for a specific username: $ sherlock arifutt.me

Wafw00f: A tool for identifying and analyzing Web Application Firewalls (WAFs) to assess the security posture of web applications. It helps security professionals with:

- **Identifying WAFs**: Detects if a web application is protected by a WAF and identifies the specific WAF in use.
- **Analyzing WAF Types**: Distinguishes between various WAF vendors and types, aiding in understanding their capabilities and limitations.
- **Informing Security Testing**: Assists in crafting appropriate test cases during penetration testing by knowing the WAF's presence and potential bypass techniques.
- **Reconnaissance**: Enhances strategic planning by understanding the security landscape, including WAF presence.

Example for Multiple URLs: To check multiple URLs, list them in a text file: $ wafw00f -i <urls.txt>

Google Hacking/Dorking is a technique using advanced search operators to uncover hidden information on the internet. Often employed by security professionals, it helps identify vulnerabilities in systems. Here are some useful operators:

- **filetype:** Searches for specific file types (e.g., PDFs).
  - Example: `filetype:pdf "Advanced Network Security"`
- **inurl:** Finds specific words in the URL.
  - Example: `inurl:admin.php`
- **intitle:** Searches for terms in the webpage title.
  - Example: `intitle:"index of"`
- **link:** Finds pages linking to a specific URL.
  - Example: `link:arifbutt.me`
- **site:** Searches within a specific site.
  - Example: `site:pucit.edu.pk inurl:admin`
- **intext:** Searches for specific text within the content of a webpage.
  - Example: `site:daraz.pk intext:admin`

These techniques can reveal sensitive information like usernames, passwords, and other private data.

OSINT Framework

Open-Source Intelligence (OSINT) is the process of collecting and analyzing publicly available information from sources like websites, social media, and public records. It is used across various fields, including cybersecurity and journalism, to gather insights about individuals or organizations.

The **OSINT Framework** (https://osintframework.com/) is a comprehensive collection of free tools and resources for conducting OSINT investigations. It categorizes tools for efficient access to information related to:

- **Search Engines:** Tools for web searches and metadata extraction.
- **People Search:** Resources for finding individual information (names, emails, phone numbers).
- **Usernames and Social Media:** Tools for tracking social media profiles and activity.
- **Email Addresses:** Tools to find and verify email addresses.
- **Domain and IP Information:** Resources for gathering website and DNS details.
- **Public Records:** Access to government databases (court records, property records).
- **Geolocation:** Tools for extracting location data from images.
- **Malware and Threat Intelligence:** Tools for analyzing malware and threat actors.
- **Metadata and File Analysis:** Extracting metadata from files for investigations.
- **Dark Web Tools:** Resources for navigating the dark web.

The framework is especially useful for security professionals, researchers, and investigators seeking information from public sources.