Reconnaissance, Info Gathering & OSINT

Overview: **Information gathering (reconnaissance)** is the initial step in penetration testing. It involves collecting public information about the target to identify vulnerabilities. It's divided into **passive** and **active** information gathering.

Passive Information Gathering

Passive information gathering collects data without direct interaction, minimizing detection risk. **Open-Source Intelligence (OSINT)** utilizes publicly available sources like news, blogs, and social media. Techniques include **Web Scraping, Google Dorking, and social media profiling**. Tools used include: **host, nslookup, dig, whois, knockpy, netdiscover, traceroute, whatweb, theHarvester, sherlock, wfw00f, Google Dorking, OSINT framework.**

Active Information Gathering

Active information gathering (scanning) involves direct interaction with the target network to gather data. It is more detectable, as it often leaves traces. Requires written permission of the system owner. **nmap** is a common tool.

Host

host is a utility that performs **DNS lookups** to convert names to IP addresses and vice versa. Can show A records (IPv4), AAAA records (IPv6), MX records (mail servers), and NS records (name servers).

Nslookup

nslookup (Name Server Lookup) is used for **name to IP address mapping**. It retrieves specific DNS records like A, AAAA, MX, NS, and TXT. Can show authoritative or non-authoritative answers. Performs **Reverse DNS (rDNS) lookup** which finds domain name associated with IP.

Dig

dig (Domain Information Groper) is a DNS lookup utility providing detailed output. Includes: * Query response header (opcode, status, ID, flags). * Question section (domain name, query type, record type). * Answer section (domain name, TTL, query class, query type, IP address). * Query statistics (query time, server used, date/time, response size).

Whois

Whois retrieves **domain registration information** from whois databases. This includes the **registry domain ID, registrar WHOIS server, registrar URL, creation/updated/expiration dates, registrar details, domain status, and name servers**. Provides info about the **IP range, organization details, and abuse contacts** when used with an IP address. Commands: `bash $ sudo apt-get install whois $ whois google.com $ whois 8.8.8.8` Online web services for domain information: * https://whois.domaintools.com/ * https://centralops.net/co/ * https://ipinfo.io/

Knockpy

Knockpy is a tool for **subdomain enumeration**. It identifies subdomains associated with a target domain. Commands: `bash $ git clone https://github.com/guelfoweb/knock.git $ cd knock $ pip3 install -r requirements.txt $ which knockpy $ knockpy --version $ knockpy -h $ knockpy -d pu.edu.pk --recon --bruteforce --threads 50`

Netdiscover

Netdiscover is an **active/passive network discovery tool** using **ARP** to identify hosts in a LAN. Commands: `bash $ sudo apt-get install netdiscover $ man netdiscover`

Active Scanning

Performs active scanning by sending ARP requests. Command: `bash $ sudo netdiscover -r 10.0.2.0/8`

Passive Scanning

Listens to network traffic to detect devices without sending requests. Command: `bash $ sudo netdiscover -p -r 10.0.2.0/8`

TraceRoute

Traceroute traces the path that packets take from your device to a remote server. Displays the list of routers/gateways (hops). Output includes the **hop number, IP of the router, and round-trip-time (RTT)**. The `-I` option specifies ICMP echo requests. Command: `bash $ traceroute -I arifbutt.me`

Whatweb

Whatweb identifies web technologies used on the target website. It can identify **HTTP Headers, web server, CMS, frameworks & libraries, plugins and extensions**. Command: `bash $ sudo apt-get install whatweb $ whatweb -v pucit.edu.pk`

Whatweb - Aggressive Scan

Whatweb can perform aggressive scans on IP ranges using the `-a` option. Level of aggression controls the trade-off between speed/stealth and reliability. The `--no-errors` option can suppress errors for non-existent addresses. Command: `bash $ whatweb -v -a 3 10.0.2.1-10.0.2.254`

TheHarvester

TheHarvester is a command-line utility for **OSINT** gathering. It collects data like domain names, IP addresses, and email addresses from public sources. Key features: * **Email Address Gathering** * **Subdomain Enumeration** * **IP Address and Hostname Discovery** Commands: `bash $ sudo apt-get install theharvester $ theharvester –help $ theHarvester -d pucit.edu.pk -l 100 -b yahoo` The `-d` option specifies the domain, `-l` the limit of search results and `-b` specifies the source(s).

Sherlock

Sherlock is a command-line tool for finding usernames across social media platforms. Useful for OSINT investigations. Commands: `bash $ sudo apt update $ sudo apt install python3 python3-pip git $ git clone https://github.com/sherlock-project/sherlock.git $ cd sherlock $ sudo pip3 install -r requirements.txt $ sherlock -h $ sherlock --version $ sherlock <username>`

Wafw00f

Wafw00f identifies **Web Application Firewalls (WAFs)**. It helps in detecting the presence and type of WAF protecting a web application. Purposes: * **Identify WAFs** * **Analyze WAF Types** * **Inform Security Testing** * **Reconnaissance** Commands: `bash $ sudo apt update $ sudo apt install python3 python3-pip git $ git clone https://github.com/EnableSecurity/wafw00f.git $ cd wafw00f $ sudo pip3 install . $ man wafw00f $ wafw00f <target_url> $ wafw00f -i <urls.txt> # To check multiple URLs, specify them in text file`

Google Hacking/Dorking

Google Dorking uses advanced search operators to find information not readily available. Operators: * `filetype:` Searches for specific file types (e.g., `filetype:pdf "Advanced Network Security"`). * `inurl:` Finds words within the URL (e.g., `inurl:admin.php`). * `intitle:` Searches for terms in the title of a webpage (e.g., `intitle:"index of"`). * `link:` Finds pages that link to a specific URL (e.g., `link:arifbutt.me`). * `site:` Searches within a specific site (e.g., `site:pucit.edu.pk inurl:admin`). * `intext:` Searches for text within the content of a webpage (e.g., `site:daraz.pk intext:admin`).

OSINT Framework

Open-Source Intelligence (OSINT) involves gathering and analyzing publicly available information. It's applied in fields like cybersecurity and law enforcement. The **OSINT Framework** (https://osintframework.com/) is a collection of free OSINT tools and resources organized for efficient investigation.

OSINT Framework Categories:

- **Search Engines:** Web search tools, metadata extraction.
- **People Search:** Finding information about individuals.
- **Usernames and social media:** Tracking social media profiles and activity, username enumeration.
- **Email Addresses:** Finding and verifying email addresses, searching for breaches.
- **Domain and IP Information:** Gathering website, domain, IP, and DNS data.
- **Public Records:** Accessing government and organizational databases.
- **Geolocation:** Extracting geolocation data from media.
- **Malware and Threat Intelligence:** Analyzing malware, IP blacklists, and threat actors.
- **Metadata and File Analysis:** Extracting metadata from files.
- **Dark Web Tools:** Navigating/searching the dark web.