ZROC102

Module1: Information Gathering Techniques

Passive Information Gathering Overview

- What is Passive Reconnaissance?
  - Trying to collect the information about the target without directly accessing the target. This involves collecting information from search engines, social media, public websites etc.

- Identifying our targets: https://www.megacorpone.com
  - Whois Enumeration
    - Whois is a TCP service, tool, and a type of database that can provide information about a domain name, such as the name server and registrar. This information is often public since registrars charge a fee for private registration.
  - Google Hacking

WHOIS Enumeration

Queries Internet Registries for domain registration information, including

- Domain ownership
- Addresses
- Phone numbers
- Locations
- Etc.

"Passive" footprinting tool – does not always query the target

- Many websites to run WHOIS queries
- Internic.net (Demo)
- Domaintools.com
- Whois.net
- Netcraft.com

Google Hacking

- Google hacking is creating complex search queries to find information in a laser-focused manner.
  - When used maliciously, can be harnessed to locate exploits, vulnerable targets, sensitive information, and conduct “anonymous” footprinting.
  - When used in a friendly manner, it can locate vulnerabilities within your own networks, find sensitive information online, and increase your awareness of threats.

Why is Google Hacking Important?

Google indexes publicly accessible websites, even those administrators believe not to be accessible to the public.

Google caches copies of pages it has accessed, so even if the sensitive information is removed, it is still available via the Google cache.

Safeguarding tip: Think about what you post online – and before you do, assume anything you put online is

- 1) freely available to the public.

2) can't be removed - EVER.

Google Search Operators

- A basic Google search query is comprised of keywords and operators.
  - A keyword can be anything of interest you'd like to search for.
  - Operators are special words which control the results shown.

- Many operators can be applied to other operators.
  - For example: -intitle: "Invalid Results" will not show pages with "invalid results" in the title

Google Search Operators

AND	Using AND between keywords is implied in Google queries – searches will return pages which contain all the words you specify.
OR	Using OR between keywords will return pages that contain either keyword you specify. For example: cat OR dog
NOT	functionally the same as the minus sign. For example: cat NOT dog, or cat -dog
*** (asterisk)**	Wildcard, but on a word level, not character level. For example: see * run, or go * go

Cont’d

“” (quotes)	Using quotes around keywords will return pages with exactly those words, in the order you specify Use with caution – you might miss some results which are related, but do not explicitly contain your query. For example, searching for “Alexander Bell” will return pages with “Alexander Bell”, but not say, “Alexander G. Bell”.
+ (plus sign)	Using a plus sign before a keyword will return pages that contain this keyword exactly Disables Google’s use of synonyms for this keyword
- (minus sign)	Using a minus sign before a keyword will return pages that DO NOT contain this keyword.

Search Examples

- “football” +ronaldo -messi – returns pages with football and ronaldo, that do not have the word messi
- “football” ronaldo OR messi – returns pages with football and either ronaldo or messi.
- ice * cone – returns pages which contain phrases like “football cone”, or “ice snow cone”.
- “ice * cone” -cream – returns pages like “ice snow cone” or “ice lance, cone of cold” that do not have the word “cream” in them.

Advanced Search Operators

intitle / allintitle	Returns pages with keywords in the title of the HTML document – Remember, the title usually appears at the top of the browser window. For example: intitle:liveapplet
inurl / allinurl	Returns pages with keywords in the URL, or address of the page For example: inurl:passwords
filetype	Returns pages with files matching the specified extension For example: filetype:xls +bank

Cont’d

allintext	Returns pages with keywords in the actual text of the page For example: allintext:passwords
site	Returns pages matching your query from a specific website For example: site:msdn.microsoft.com
link	Returns pages with links to the specified domain For example: link:www.us.gov
inanchor	Returns pages with keywords in the link text For example: inanchor:”United States Government”

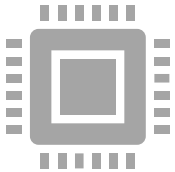
Google cache

- Google caches pages, with the result being that information posted online can effectively never be removed or deleted.
- This cached copy is available, even if the original server is disconnected from the Internet, or all its pages wiped clean.
- Reviewing cached pages may yield sensitive information that was found by site administrators and removed. AND, these searches are “anonymous” to the attacker – no traffic is sent to the target site

Finding Vulnerabilities and Targets

CGI – Common Gateway Interface scripts are commonly seen on the web – some of these files may contain vulnerabilities.

Examples userreg.cgi, mailview.cgi, maillist.cgi, cphost.dll, findvserver.asp, SQLQHit.asp



Using Google hacking techniques, you can easily locate servers with CGI files on them.

Examples inurl:/cgi-bin/userreg.cgi fileext:cgi

A CGI scanner can locate these files and determine if they are vulnerable. E.g. Nikto

Login Portals

- Main login pages can be good sources of information for social engineering attacks.
- Some login pages have instructions for users to get help if they have forgotten their password or otherwise are unable to login.
- Run a Google query looking for login pages
  - intitle:Login site:yourdomain.com
  - login OR logon

Finding Email Addresses

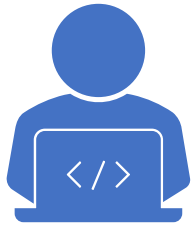
- Looking for an email address to send a phishing message to? Google has also got you covered.
- If you don’t care WHO you send phishing messages to, and just want a target inside the network, try searching for just the domain, perhaps limited to files that typically store such information.
  - domain.com filetype:xls

Google Hacking Resources

Google API	Application Programming Interface to programmatically access the Google search engine.
GHDB	Google hacking database, from the father of Google Hacking, contains lots of predefined queries for sensitive information. Now hosted at hackersforcharity.org.
SiteDigger	Scans Google's cache for sensitive information
Gooscan	Automates queries to Google to find vulnerabilities on web pages; built for Google search appliances, but can be run against Google in violation of its Terms of Service

Open-Source Code

open-source projects and online code repositories, such as GitHub,153 GitLab,154 and SourceForge are perfect places to perform passive information gathering.



Code stored online can provide a glimpse into the programming languages and frameworks used by an organization. (Former SolarWinds CEO blames intern for "solarwinds123" password leak – CNNPolitics)

In some rare occasions, developers have even accidentally committed sensitive data and credentials to public repos.

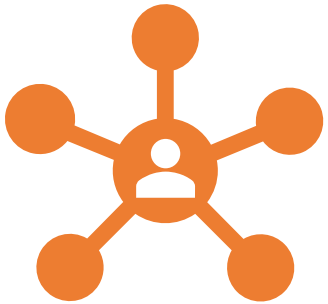
Github

- On GitHub, we will be able to search a user's or organization's repos, but we need an account if we want to search across all public repos.
- Let's search that megarcorpone's repos for interesting information.
  - We can use "filename:users" to search for any files with the word "users" in the name.
- This manual approach will work best on small repos. For larger repos, we can use several tools to help automate some of the searching, such as *Gitrob* and *Gitleaks.*

Shodan

- What is shodan?
  - Shodan is a search engine that crawls devices connected to the Internet including but not limited to the World Wide Web. This includes the servers that run websites but also devices like routers and IoT devices.
- Google and other search engines look for web server content, while Shodan searches for Internet-connected devices, interacts with them, and displays information about them.
- Typical shodan search query
  - hostname:megacorpone.com
  - hostname:megacorpone.com port:"22"

Pastebin



Pastebin is a website for storing and sharing text. The site doesn’t require an account for basic usage.

since Pastebin is a public service, we can use it to search for sensitive information.

User Information Gathering

- In addition to gathering information about our target organization's resources, we can also gather information about the organization's employees.
- Our purpose for gathering this information is to compile user or password lists, build pretexting for social engineering, augment phishing campaigns or client-side attacks.
- Email Harvesting
  - $ theharvester –d megacorpone.com –b google

Data breaches

- Malicious hackers often dump breached credentials on Pastebin or other less reputable websites. These password dumps can be extremely valuable for generating wordlists. (RockYou2021: Largest Ever Password Compilation Leaked | CyberNews)

Active Information Gathering

Unlike passive information gathering, here we will explore techniques that involve direct interaction with target services.

we will look at some of the more common active information gathering techniques in this module including port scanning and DNS, SMB, NFS, SMTP, and SNMP enumeration

DNS Enumeration

- The Domain Name System (DNS) is one of the most critical systems on the Internet and is a distributed database responsible for translating user-friendly domain names into IP addresses.
- Each domain can use different types of DNS records. Some of the most common types of DNS records include:
  - • NS - Nameserver records contain the name of the authoritative servers hosting the DNS records for a domain.
  - A - Also known as a host record, the “a record” contains the IP address of a hostname (such as www.megacorpone.com).
  - • MX - Mail Exchange records contain the names of the servers responsible for handling email for the domain. A domain can contain multiple MX records.
  - • PTR - Pointer Records are used in reverse lookup zones and are used to find the records associated with an IP address.
  - • CNAME - Canonical Name Records are used to create aliases for other host records.
  - • TXT - Text records can contain any arbitrary data and can be used for various purposes, such as domain ownership verification

DNS Enumeration

- we'll use the host command to find the IP address of www.megacorpone.com
  - $ host www.megacorpone.com
- By default, the host command looks for an A record, but we can also query other fields, such as MX or TXT records. To do this, we can use the -t option to specify the type of record we are looking for
  - $ host -t mx megacorpone.com
  - $host –t txt megacorpone.com

DNS Zone Transfer

- A zone transfer is basically a database replication between related DNS servers in which the zone file is copied from a master DNS server to a slave server.
- The zone file contains a list of all the DNS names configured for that zone. Zone transfers should only be allowed to authorized slave DNS servers but many administrators misconfigure their DNS servers, and in these cases, anyone asking for a copy of the DNS server zone will usually receive one.
- This is equivalent to handing a hacker the corporate network layout on a silver platter. All the names, addresses, and functionality of the servers can be exposed to prying eyes.

Automated DNS Enumeration

DNSRecon

- $ dnsrecon -d megacorpone.com -t axfr

DNSenum

- $ dnsenum zonetransfer.me

Port Scanning

- Port scanning is the process of inspecting TCP or UDP ports on a remote machine with the intention of detecting what services are running on the target and what potential attack vectors may exist.
  - $ nc -nvv -w 1 -z 10.11.1.220 3388-3390 (TCP)
  - $ nc -nv -u -z -w 1 10.11.1.115 160-162 (UDP)
- Port scanning with NMAP
  - Nmap is one of the most popular, versatile, and robust port scanners available. It has been actively developed for over a decade and has numerous features beyond port scanning.
    - Stealth / SYN scanning: SYN scanning is a TCP port scanning method that involves sending SYN packets to various ports on a target machine without completing a TCP handshake. If a TCP port is open, a SYN-ACK should be sent back from the target machine, informing us that the port is open.
      - $ nmap –sS IP

Port Scanning

TCP Connect Scan

- Unlike Syn scan, it waits for the 3 way handshake to be completed
- Slower but more accurate scan.
  - $ nmap –sT IP

UDP Scanning

- $ nmap –sU IP

Network Sweeping

- Host discovery, similar to netdiscover
  - $ nmap –sn IP range

OS fingerprinting

- $ nmap –O IP

Port Scanning

- Banner Grabbing / Service Enumeration.
  - We can also identify services running on specific ports by inspecting service banners (-sV) and running various OS and service enumeration scripts (–A) against the target:
    - $ nmap –sV –sT –A IP
- Massscan
  - arguably the fastest port scanner; it can scan the entire Internet in about 6 minutes, transmitting an astounding 10 million packets per second!
    - $ sudo masscan –p80 IPrange

Hunting for Subdomains

- Search for publicly available internal subdomains / URLs.
- Misconfigurations, or someone's “good idea” of making an internal resource available to traveling members, may unintentionally open up an internal resource to the outside world and thus the attacker.
- Finding such a foothold may potentially allow access/insight into the organization’s internal architecture