

Prepared by Dr. Nermin Hamza

CYB740

Penetration test and Ethical hacking

Lecture 02



Reconnaissance and Information gathering

Lecture 02

CYB740

Penetration test and Ethical hacking

Prepared by
Dr. Nermin Hamza



Outlines

- Reconnaissance
- Introduction: Information gathering
- Information gathering categories
- Sources of Information Gathering
- Information gathering techniques
- Information gathering tools



Remember: Phases of Penetration Testing

- Reconnaissance and Information Gathering
- Network Enumeration and Scanning
- Vulnerability Testing and Exploitation
- Reporting



Reconnaissance

- **Reconnaissance**, also known as information gathering, is arguably the most important of the four phases we will discuss. The more time you spend collecting information on your target, the more likely you are to be successful in the later phases.



Introduction: Information gathering

- **Information gathering** is the first phase of hacking.
- In this phase, we gather as much information as possible regarding the target's online presence, which in turn reveal useful information about the target itself.
- It is the first step or the beginning stage of Ethical Hacking, where the penetration testers or hackers (both black hat or white hat) performed this stage; this is a necessary and crucial step to be performed.



Introduction :Information gathering

- The required information will depend on whether we are doing a network pen-test or a web application pen-test.
- The more the information gathered about the target, the more the probability to obtain relevant results.
- Information gathering is not just a phase of security testing; it is an **art that every penetration-tester** (pen-tester) and hacker should master for a better experience in penetration testing.



Introduction :Information gathering

- There are various tools, techniques, and websites, including public sources such as Whois, nslookup that can help hackers gather information.
- This step is necessary because you may need any information (such as his pet name, best friend's name, age, or phone number to perform password guessing attack or other kinds of attacks) while performing attacks on any target.



Information gathering categories

all information gathering techniques can be classified into two main categories:

1. Active information gathering
2. Passive information gathering



Active Information Gathering

Active Information Gathering

- In active information gathering, we would directly engage with the target, for example, gathering information about what ports are open on a particular target, what services they are running, and what operating system they are using.



Active Information Gathering

Active Information Gathering

- However, the techniques involving active information gathering would be very noisy at the other end.
- As they are easily detected by IDS, IPS, and firewalls and generate a log of their presence, and hence are not recommended sometimes



Passive Information Gathering

Passive Information Gathering

- In passive information gathering, we do not directly engage with the target. Instead, we use search engines, social media, and other websites to gather information about the target. This method is recommended, since it does not generate any log of presence on the target system.



Passive Information Gathering

Passive Information Gathering

- A common example would be to use LinkedIn, Facebook, and other social networks to gather information about the employees and their interests. This would be very useful when we perform phishing, keylogging, browser exploitation, and other client side attacks on the employees.



Sources of Information Gathering

There are many sources of information; the most important ones are as follows:

- Social media
- Search engines
- Press releases
- Job sites



Social Media

- With all the Social Media platforms our lives become less and less private.
- Some of us try going incognito but most of our family members don't as a result this act of ours ends up not helping as much as it should.
- if you fail to contact me, just try someone close to me instead. This benefits hackers very much.



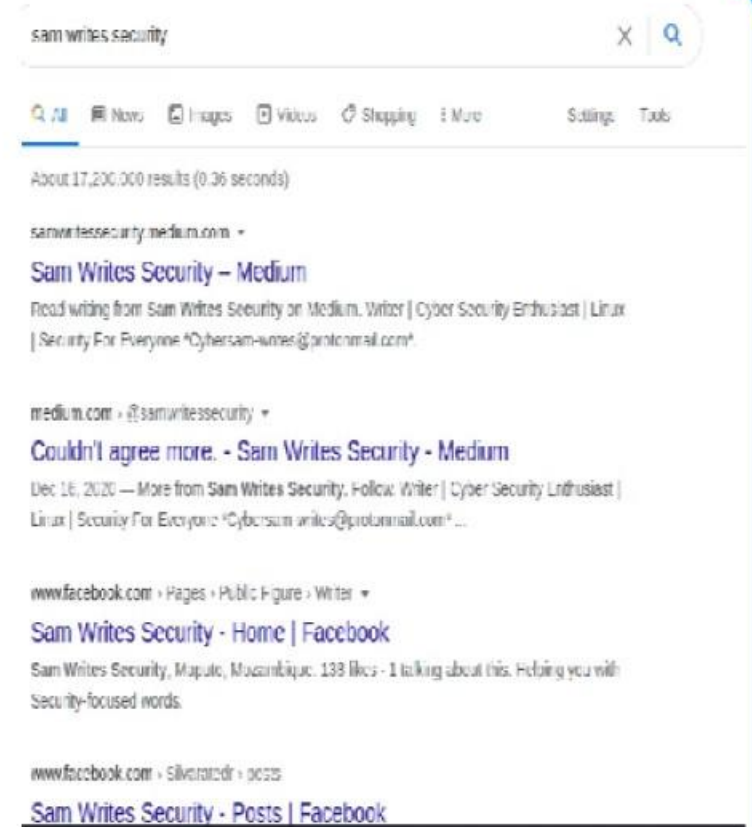
Social Media

- You might use a pen name for most of your writing, a fake name on all your Social Media profiles and email addresses but your partner may be doing the exact opposite, like sharing your full name on a post when it's your birthday, sharing your wedding photos, your son or daughter's birth month, and a few more things we don't take into consideration which should be too personal to go online but isn't, at least not anymore.



Search engines

- Compared to Social Media platforms search engines get more results because instead of a single profile they get you everything it can find on your search, not to mention that the Social Media profiles also get indexed on engines.
- So, a simple search is how you should start.



Press Release

- A press release is a written document prepared for the media ("the press") that announces something newsworthy.



Job sites

- Job Site means a specific worksite that deals specifically with employment or careers.
- Many employment websites are designed to allow employers to post job requirements for a position to be filled and are commonly known as job boards.
- Other employment sites offer employer reviews, career and job-search advice, and describe different job descriptions or employers.



Passive Information gathering techniques



Information gathering techniques

- Google Directives
- DNS Server
- E-MAIL SERVERS



GOOGLE DIRECTIVES:

- These directives are **keywords** that enable us to more accurately extract information from the Google Index.



Information gathering using :Google Directives

To properly use a Google directive, you need three things:

1. The name of the directive you want to use
2. A colon
3. The term you want to use in the directive.

To utilize the “site:” directive, we need to enter the following into a Google search box:

site:domain term(s) to search

Example:

site:dsu.edu pat engebretson



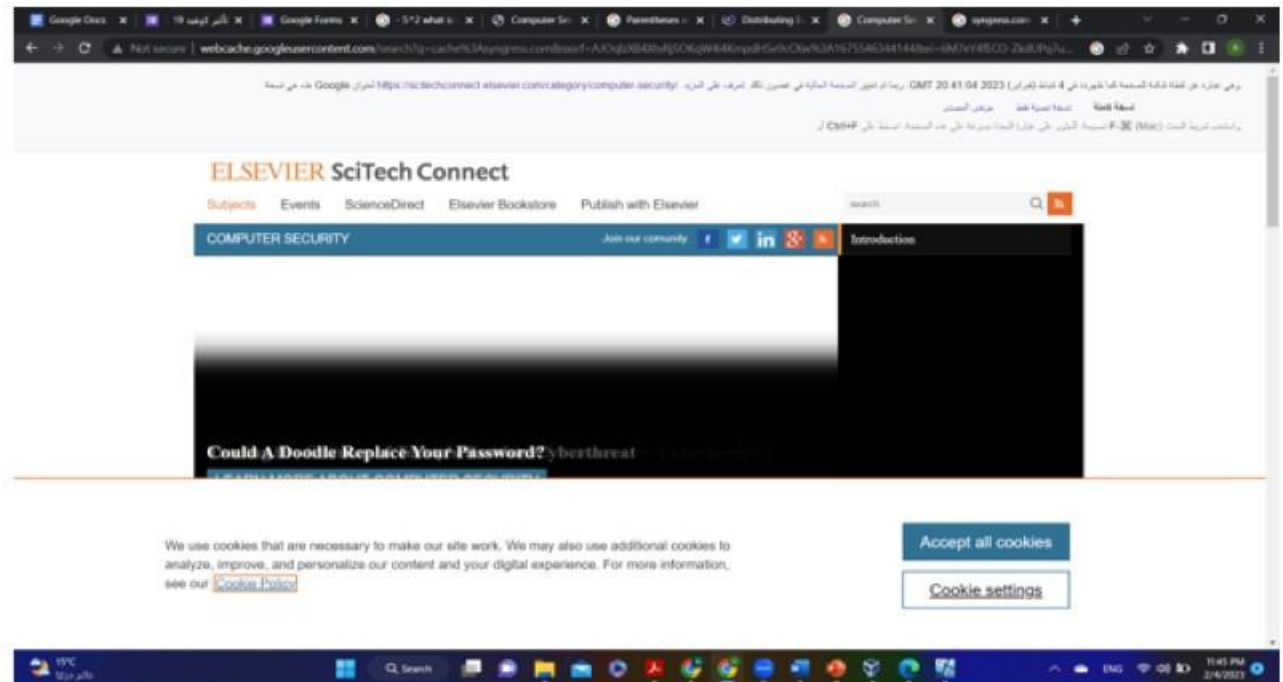
Chapter

02

Information gathering using :Google Directives

- We can use the cache: directive to limit our search results and show only information pulled directly from the Google cache.
- The following search will provide us with the cached version of the Syngress homepage:

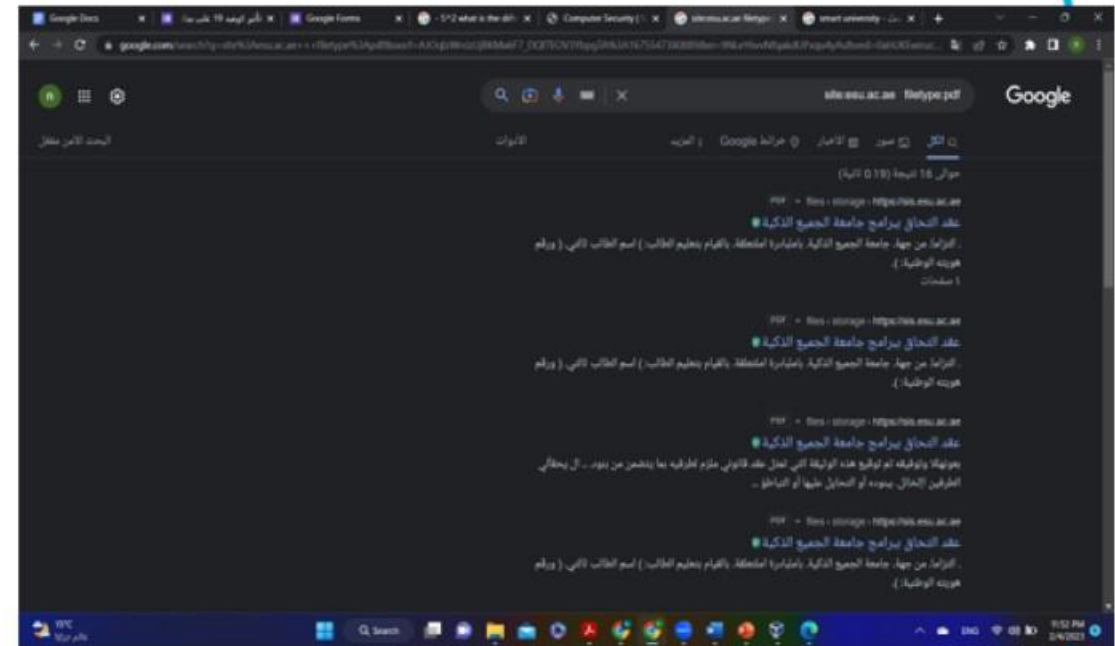
cache:syngress.com



Information gathering using :Google Directives

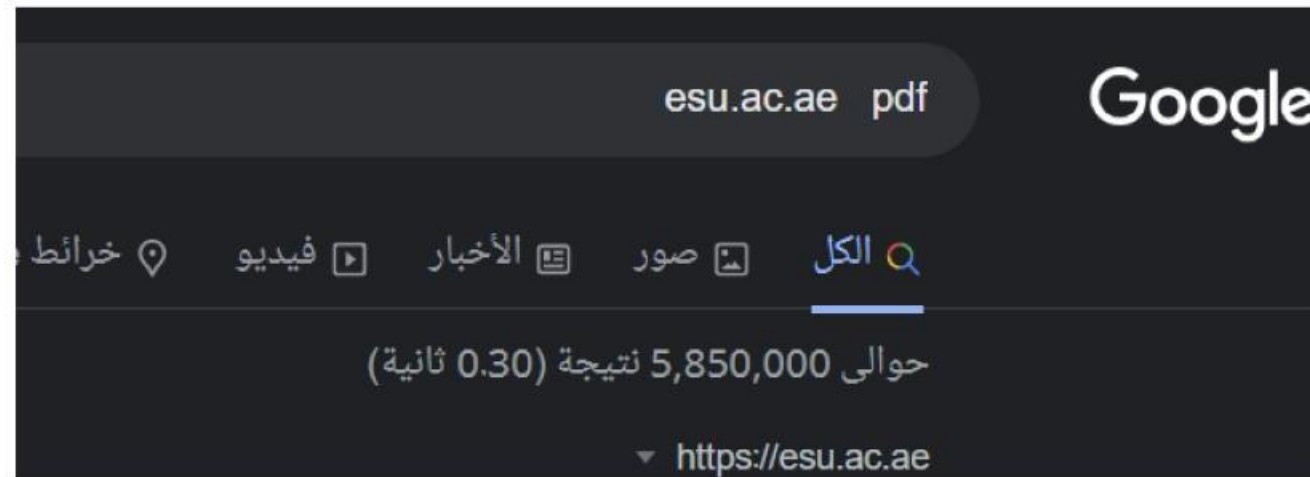
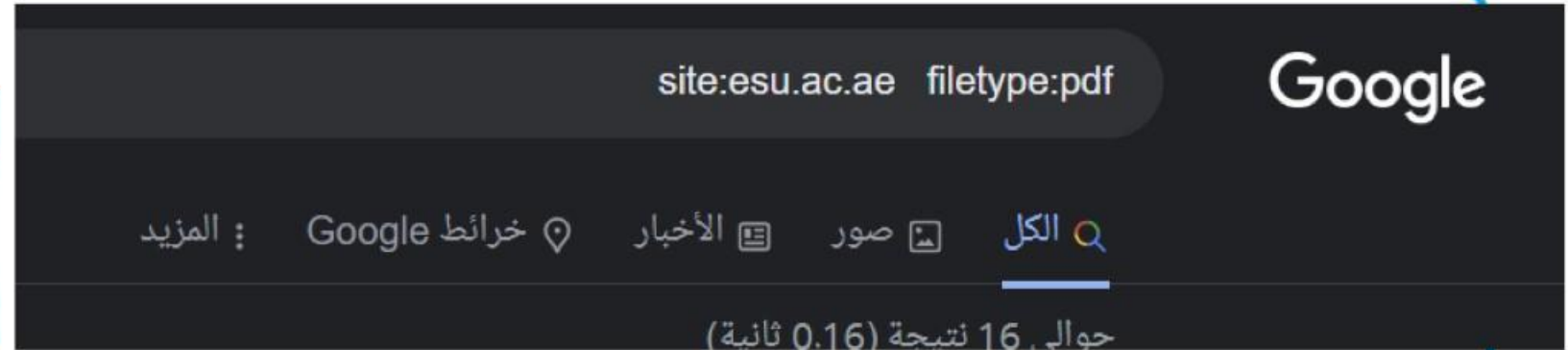
- Another directive we will cover here is “filetype:”. We can utilize “filetype:” to search for specific file extensions.
- This is extremely useful for finding specific types of files on your target’s website.
- For example, to return only hits that contain PDF documents, you would issue the following command:
- filetype:pdf

site:esu.ac.ae filetype:pdf



Information gathering using :Google Directives

What is the difference?



Information gathering using :Google Directives

Chapter

02

Google Search Terms

Operator	Description
Filetype	This operator directs Google to search only within the test of a particular type of file. Example: filetype:xls
Inurl	This operator directs Google to search only within the specified URL of a document. Example: inurl:search-text
Link	The link operator directs Google to search within hyperlinks for a specific term. Example link: www.domain.com
Intitle	The intitle operator directs Google to search for a term within the title of a document. Example intitle: "Index of...etc"



Chapter

02

Information gathering using DNS servers

- A DNS server is a computer server that contains a database of public IP addresses and their associated hostnames, and in most cases serves to resolve, or translate, those names to IP addresses as requested.
- DNS servers run special software and communicate with each other using special protocols.
- The DNS server sits in the space between humans and computers to help facilitate their communication.
- It's easier to remember a domain or hostname like lifewire.com than it is to remember the site's IP address numbers 151.101.2.114. So when you access a website, like Lifewire, all you have to type is the URL <https://www.lifewire.com>.



- DNS servers are an excellent target for hackers and penetration testers. They usually contain information that is considered highly valuable to attackers.
- DNS is a core component of both our local networks and the Internet. Among other things, DNS is responsible for the process of translating domain names to IP addresses.
- As humans, it is much easier for us to remember “google.com” rather than `http://74.125.95.105`. However, machines prefer the reverse.
- DNS serves as the middle man to perform this translation process.



EXTRACTING INFORMATION FROM DNS

- As penetration testers, it is important to focus on the DNS servers that belong to our target. The reason is simple?
- In order for DNS to function properly, it needs to be aware of both the IP address and the corresponding domain name of each computer on its network.



EXTRACTING INFORMATION FROM DNS

- The primary tool to query DNS servers is **nslookup**.
- Nslookup provides machine name and address information.

```
C:\Users\Nermin>nslookup www.google.com
Server:  dns.google
Address:  8.8.8.8

Non-authoritative answer:
Name:     www.google.com
Addresses: 2a00:1450:4002:403::2004
          142.250.180.164
```



Information gathering from E-MAIL SERVERS

- E-mail servers can provide a wealth of information for hackers and penetration testers. In many ways, e-mail is like a revolving door to your target's organization.
- we can often gather significant pieces of information by interacting directly with the e-mail sever.
- send an e-mail to the organization with an empty .bat file or a nonmalicious .exe file like calc.exe. (the goal is to send a message to the target e-mail server inside the organization in the hope of having the e-mail server inspect, and then reject the message.)



- Once the rejected message is returned back to us, we can attempt to extract information about the target e-mail server.
- In many cases, the body of the message will include a pre-canned write-up explaining.
- This message often indicates the specific vendor and version of antivirus that was used to scan the e-mail. As an attacker, this is a great piece of information to have.



Active Information gathering techniques



Information gathering techniques

- Website Copier
- IP Addresses
- Port Scanning
- FTP Testing – (File Server Probing)



- we begin Step 1 by closely reviewing the target's website.
- In some cases, it maybe helpful to use a tool called **HTTrack** to make a page-by-page copy of the website.
- HTTrack is a free utility that creates an identical, offline copy of the target website.
- Utilizing a website-copying tool like HTTrack allows us to explore and thoroughly mine the website "offline" without having to spend additional time traipsing around on the company's web server.



Information Gathering using IP Addresses

- Ping
- You would need to know the IP address of the webserver in order to trace the exact location

```
C:\Users\Nermin>ping www.google.com

Pinging www.google.com [142.250.180.132] with 32 bytes of data:
Reply from 142.250.180.132: bytes=32 time=1012ms TTL=117
Reply from 142.250.180.132: bytes=32 time=451ms TTL=117
Reply from 142.250.180.132: bytes=32 time=715ms TTL=117
Reply from 142.250.180.132: bytes=32 time=738ms TTL=117

Ping statistics for 142.250.180.132:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 451ms, Maximum = 1012ms, Average = 729ms
```



Information Gathering using IP Addresses

- After determining the webserver's IP, we can use some online tools to track the exact location of the webserver.

- http://www.ip-adress.com/ip_tracer/yourip
- <https://www.ipaddress.com/>

<https://www.ipaddress.com/ipv4/142.250.180.132>

 **IPAddress.com**
The Best IP Address Tools

What Is My IP:
41.69.206.161 EG

142.250.180.132 us

The public IP address **142.250.180.132** is located in *United States*. It is a address belongs to **ASN 15169** which is delegated to *GOOGLE*.

Please have a look at the tables below for full details about **142.250.180** approximate IP location for any public IP address.

142.250.180.132 IP Address Location	
Reverse IP (PTR)	mil04s43-in-f4.1e100.net
ASN	15169 (GOOGLE)
ISP / Organization	Google Servers



- When we move to the IP Addresses area, we find extensive information regarding the IPs, including Summary by Regional Registrar, Stats by IP subnet size, as well as the full list of IPs filtered by number of IPs blocks, IP Count, Unique User Agents, RIR, Hostnames and number of hosted domain names.



Port Scanning

Port scanning is a technique used in cybersecurity to **identify open ports** on a target device or server.

Tool	Description
Nmap	Most popular tool for port scanning
Netcat	For manual connection tests



FTP Testing – (File Server Probing)

FTP testing checks if an FTP server is:

- Allowing **anonymous access**
- Exposing **sensitive files**
- Running outdated or **vulnerable software**
- Using **plain-text credentials**



Chapter

02

Information gathering Tools



Information gathering tools

- **Whois**
- **Traceroute**
- **Netcraft**
- **theHarvester**
- **Nmap**



Information gathering tools

Chapter

02

Tool	Type	Classification	Reason
WHOIS	Domain lookup	Passive	Queries public WHOIS databases without contacting the target directly.
Traceroute	Network mapping	Active	Sends packets to the target and intermediate routers—interacts directly with the network.
Netcraft	Web intelligence	Passive	Gathers public data (e.g., hosting provider, OS, technologies) from third-party sources, not the target itself.



Information gathering tools

Tool	Type	What It Finds	Main Use
theHarvester	Passive	Emails, subdomains from public sources	OSINT, email phishing
Nmap	Active	Open ports, services, versions, OS	Network scanning, footprinting



WHOIS: The WHOIS command is a great source of data for fetching domain- and IP-related information, including admin names, telephones, addresses, country, DNS servers, etc.



Information Gathering : Whois

Chapter

02

Whois :

- very simple but effective means for collecting additional information about our target.
- Whois holds a huge database that contains information regarding almost every website that is on the web, most common information are “who owns the website” and “the e-mail of the owner,” which can be used to perform social engineering attacks.



Information Gathering : Whois

```
(kali@kali)-[~]  
$ whois www.google.com  
No match for "WWW.GOOGLE.COM".  
>>> Last update of whois database: 2023-02-06T15:03:30Z <<<  
  
NOTICE: The expiration date displayed in this record is the date the  
registrar's sponsorship of the domain name registration in the registry is  
currently set to expire. This date does not necessarily reflect the expiration  
date of the domain name registrant's agreement with the sponsoring  
registrar. Users may consult the sponsoring registrar's Whois database to  
view the registrar's reported date of expiration for this registration.  
  
TERMS OF USE: You are not authorized to access or query our Whois  
database through the use of electronic processes that are high-volume and  
automated except as reasonably necessary to register domain names or  
modify existing registrations; the Data in VeriSign Global Registry
```



Traceroute: As one of the most popular network tools used to track the path of networks packets between one IP address to another, it's a powerful recon tool that will let you gain critical network information about IP addresses and networking routes.



Traceroute

- Traceroute is a very popular utility available in both Windows and Linux. It is used for network orientation.
- By network orientation I don't mean scanning a host for open ports or scanning for services running on a port.
- It means to figure out how the network topology, firewalls, load balancers, and control points, etc. are implemented on the network.



Information Gathering :Traceroute

```
C:\Users\Nermin>tracert www.google.com
```

```
Tracing route to www.google.com [142.250.180.164]  
over a maximum of 30 hops:
```

1	1 ms	1 ms	1 ms	192.168.1.1
2	29 ms	27 ms	27 ms	41.69.0.2
3	32 ms	28 ms	33 ms	172.31.248.141
4	31 ms	28 ms	29 ms	host-81.10.86.137.tedata.net [81.10.86.137]
5	*	*	*	Request timed out.
6	71 ms	71 ms	70 ms	be3154.ccr32.mrs02.atlas.cogentco.com [154.54.76.217]
7	79 ms	83 ms	79 ms	be2865.rcr21.mil01.atlas.cogentco.com [154.54.74.178]
8	90 ms	89 ms	90 ms	ix-ae-12-0.thar1.wi3-milan.as6453.net [195.219.166.196]
9	73 ms	250 ms	72 ms	72.14.213.156
10	74 ms	75 ms	72 ms	108.170.245.81
11	72 ms	71 ms	72 ms	142.250.211.23
12	73 ms	72 ms	73 ms	mil04s44-in-f4.1e100.net [142.250.180.164]

```
Trace complete.
```



NETCRAFT

- Another great source of information is Netcraft. You can visit their site at <http://news.netcraft.com>.
- Start by searching for your target in the “What’s that site Running?”
- Netcraft contains a huge online database with useful information on websites and can be used for passive information gathering against the target.



Information Gathering :Netcraft

Site report for https://esu.ac.ae

► 🔍 Look up another site?

Share:     

Background

Site title	جامعة الامم المتحدة	Date first seen	December 2021
Site rank	181727	Netcraft Risk Rating 	1/10 
Description	Not Present	Primary language	Arabic

Network

Site	https://esu.ac.ae 	Domain	esu.ac.ae
------	---	--------	-----------



Information Gathering :theHarvester

The package contains a tool for gathering subdomain names, e-mail addresses, virtual hosts, open ports/ banners, and employee names from different public sources (search engines, pgp key servers).

<https://www.kali.org/tools/theharvester/>

```
usage: theHarvester [-h] -d DOMAIN [-l LIMIT] [-S START] [-p] [-s]
                  [--screenshot SCREENSHOT] [-v] [-e DNS_SERVER] [-t]
                  [-r [DNS_RESOLVE]] [-n] [-c] [-f FILENAME] [-w WORDLIST]
                  [-a] [-q] [-b SOURCE]
```

theHarvester is used to gather open source intelligence (OSINT) on a company or domain.



Information Gathering : theHarvester

```
theHarvester -d example.com -b google -l 100
```

Now let me walk you through this like a teacher:

- theHarvester is the tool we're running.
- -d example.com tells theHarvester that our **target domain** is example.com.
- -b google means we want to **search Google** as our source of public data.
- -l 100 limits the search to **the first 100 results**.

It works by searching public sources like:

- Google
- Bing
- LinkedIn
- Twitter
- Hunter.io
- DNSDumpster
- Shodan
- ...and more.



Information Gathering : theHarvester

[-] Emails found:
admin@example.com
careers@example.com

[-] Hosts found:
mail.example.com
vpn.example.com
intranet.example.com



```
(kali@kali)-[~]
$ theHarvester -d fifa.com -b bing -l 100
```

```
* Coded by Christian Martorella
* Edge-Security Research
* cmartorella@edge-security.com
*
*****
usage: theHarvester [-h] -d DOMAIN [-l LIMIT] [-s START] [-p] [-s]
                  [--screenshot SCREENSHOT] [-v] [-e DNS_SERVER] [-t]
                  [-r [DNS_RESOLVE]] [-n] [-c] [-f FILENAME] [-b SOURCE]
theHarvester: error: unrecognized arguments: -k YOUR_API_KEY

(kali@kali)-[~]
$ theHarvester -d fifa.com -b bing -l 100
Read proxies.yaml from /home/kali/.theHarvester/proxies.yaml
*****
*
* theHarvester
*
* theHarvester 4.6.0
* Coded by Christian Martorella
* Edge-Security Research
* cmartorella@edge-security.com
*
*****

[*] Target: fifa.com

Read api-keys.yaml from /home/kali/.theHarvester/api-keys.yaml
Searching 0 results.
[*] Searching Bing.

[*] No IPs found.

[*] No emails found.

[*] Hosts found: 5
-----
ar.fifa.com
digitalhub.fifa.com
inside.fifa.com
play.fifa.com
store.fifa.com
```



Information Gathering : Nmap

What is Nmap?

Nmap is short for Network Mapper. It is an open-source Linux command-line tool that is used to scan IP addresses and ports in a network and to detect installed applications.

Nmap allows network admins to find which devices are running on their network, discover open ports and services, and detect vulnerabilities.



Information Gathering : Nmap

```
nmap 192.168.1.1
```

PORT	STATE	SERVICE
------	-------	---------

22/tcp	open	ssh
--------	------	-----

80/tcp	open	http
--------	------	------

443/tcp	open	https
---------	------	-------





Thank You