📖 TheHarvester – Full Manual (Kali Linux)

🔹 Overview

TheHarvester is a **Python-based reconnaissance tool** used for **information gathering**.
It collects:

- Emails
- Subdomains
- Hosts
- IPs
- Usernames

from **public sources** (search engines, PGP, social networks, certificates, etc.).
It is widely used in the **OSINT (Open Source Intelligence)** phase of penetration testing.

🔹 Basic Syntax

```
theHarvester -d <domain> -l <limit> -b <source> [options]
```

🔹 Key Parameters

Option	Description	Example
-d	Target domain name	-d tesla.com
-l	Limit results	-l 500
-s	Start from result number (pagination)	-s 200
-b	Source (search engine / service)	-b google
-f	Save results to file (HTML/XML)	-f tesla_report
-v	Verbose output	-v
-h	Help menu	-h

◈ Supported Sources (-b)

Source	Description
google	Google search
bing	Bing search
yahoo	Yahoo search
duckduckgo	DuckDuckGo search
baidu	Baidu search
crtsh	Certificate Transparency logs
linkedin	LinkedIn users (requires API)
pgp	PGP key servers
virustotal	Uses VirusTotal’s passive DNS
hunter	Email hunter API (needs key)
intelx	IntelligenceX search (needs key)
anubis	Subdomain data
all	Use all available sources

◈ Basic Examples

Collect from Bing

```
theHarvester -d facebook.com -l 300 -b bing
```

Use all sources

```
theHarvester -d facebook.com -l 300 -b all
```

Save to HTML report

```
theHarvester -d facebook.com -l 300 -b bing -f fb_report
```

Verbose mode

```
theHarvester -d tesla.com -l 200 -b yahoo -v
```

◈ Advanced Techniques

1. Save in XML & parse for automation

```
theHarvester -d target.com -l 500 -b all -f target_data.xml
```

- Useful for automation & integration with scripts.
- Can be imported into Burp Suite / custom parsers.

---

2. Use API Keys for More Results

Some sources require API keys.
Edit the config file:

```
nano /etc/theHarvester/api-keys.yaml
```

Example (Hunter.io):

```
hunter:

  key: "YOUR_HUNTER_API_KEY"

virustotal:

  key: "YOUR_VT_API_KEY"

intelx:

  key: "YOUR_INTELX_API_KEY"
```

Run with API-based sources:

```
theHarvester -d target.com -l 500 -b hunter
```

---

3. Combine with Amass for deeper subdomain enumeration

```
amass enum -d target.com -o amass_output.txt

theHarvester -d target.com -l 500 -b all -f harvester_output.xml

cat amass_output.txt harvester_output.xml | sort -u > final_subdomains.txt
```

✅ Gives a more complete picture of the target domain.

---

4. Pivot with Maltego (Visual Recon)

- Import TheHarvester results into **Maltego**.
- Graphically visualize relationships between emails, domains, IPs, and organizations.

---

5. Use in Red Team Engagements

- Save HTML/XML reports for documentation.
- Export emails and run **password spray / phishing simulation**.
- Extract subdomains and feed into **Nmap**:

```
for i in $(cat final_subdomains.txt); do nmap -sV $i; done
```

---

6. Stealth Reconnaissance

- Instead of -b all, choose **single providers** to avoid detection.
- Use proxychains/Tor to mask origin:

```
proxychains theHarvester -d target.com -l 200 -b google
```

---

7. Chaining with Other Tools

Tool	**Integration**
Nmap	Scan harvested hosts/subdomains for open ports.
Metasploit	Use discovered emails for **spear-phishing** modules.
Sublist3r	Validate & cross-check subdomains.
theHarvester + Shodan	Use IPs for deeper host intel.

� Output

- **Terminal** → Emails, subdomains, IPs, hosts.
- **HTML/XML Reports** → Saved for documentation.
- Can be parsed by other tools for automation.

💠 Pro Tips

- Always start with -b all, then refine with specific sources.
- Add API keys for better results from **Hunter.io, VirusTotal, IntelX, etc.**
- Save outputs in XML/HTML for further processing.
- Use TheHarvester in **early recon**, then move to **Amass, Shodan, Nmap** for deeper scans.
- For stealth, use **Tor + proxychains**.

🔥 With this manual, you now have both **basic & advanced techniques** for TheHarvester.

🕵️ Reconnaissance Workflow (Kali Linux)

```
Target Domain / IP
   |
   ├──▶ TheHarvester
   |      ├– Collect emails
   |      ├– Collect subdomains
   |      ├– Collect hostnames
   |      └─ Collect IPs
   |
   ├──▶ Subdomain Enumeration
   |      ├– Amass
   |      ├– Sublist3r
   |      └─ crt.sh (certificate transparency)
   |
   ├──▶ Data Validation
   |      ├– Remove duplicates
   |      ├– Compare TheHarvester + Amass
   |      └─ Final list of subdomains/hosts
```

```
|
├──▶ Scanning
|     ├– Nmap
|     |     ├– Port scan
|     |     ├– Service version scan
|     |     └─ Script scan (vuln, http-enum, etc.)
|     ├– Masscan (fast wide scan)
|     └─ Rustscan (optimized scanning)
|
├──▶ Intelligence Gathering
|     ├– Shodan (host/IP intelligence)
|     ├– Censys (certs & IP data)
|     ├– VirusTotal (passive DNS, malware check)
|     └─ Hunter.io / IntelX (email intelligence)
|
├──▶ Pivoting
|     ├– Export results to Maltego (graph visualization)
|     ├– Feed hosts into Metasploit (exploit modules)
|     └─ Use emails for password spraying / phishing simulation
|
└──▶ Documentation
      ├– Save TheHarvester reports (HTML/XML)
      ├– Export Nmap scans
      └─ Create recon report for pentest
```

⚡ Quick Workflow Commands

1. **Gather Initial Data**

```
theHarvester -d target.com -l 500 -b all -f target_report
```

2. **Subdomain Enumeration**

```
amass enum -d target.com -o amass.txt
```

```
sublist3r -d target.com -o sublist3r.txt
```

3. **Merge + Deduplicate**

```
cat target_report.xml amass.txt sublist3r.txt | sort -u > hosts.txt
```

4. **Scan Hosts**

```
nmap -sV -iL hosts.txt -oN nmap_scan.txt
```

5. **Shodan Lookup**

```
shodan host <IP>
```

6. **Visualize with Maltego**

- Import emails/domains → Build relationship graph

💠 Pro Tips

- Use **TheHarvester first** → gets you starting emails, subdomains, and IPs.
- Always **validate with Amass/Sublist3r** for deeper enumeration.
- Use **Shodan + Censys** to get extra hidden intel about discovered IPs.
- Save everything (-f) → makes reporting easier.
- For stealth, combine with **proxychains + Tor**.