

theHarvester

theHarvester is another tool like sublist3r which is developed using Python. This tool can be used by penetration testers for gathering information of emails, sub-domains, hosts, employee names, open ports, and banners from different public sources like search engines, PGP key servers, and SHODAN computer database. This tool can be used in passive reconnaissance and by anyone who needs to know what an attacker can see about the organization.

theHarvester (purposely spelt with a lower-case 't' at the beginning) is a commandline-based tool made by the team at **Edge-Security**. It is a Python-based tool meant to be used in the initial stages of an investigation by leveraging open source Intelligence (OSINT) to help determine a company's external threat landscape on the internet.

The tool was originally designed to be used in the early stages of a penetration test or red team engagement. However, the passive reconnaissance abilities of theHarvester also make it suitable for blue or purple teams, depending on the situation.

TheHarvester can retrieve various types of information, including:

- **Email Addresses:** It can search for email addresses associated with a domain from search engines, social media platforms, and other public sources.
- **Domain Names:** It can enumerate subdomains associated with a target domain.
- **Hostnames:** It can discover hostnames associated with a target domain or IP address.
- **Open Ports:** It can identify open ports on target systems.
- **Virtual Hosts:** It can discover virtual hosts associated with web servers.
- **DNS Information:** It can gather DNS information such as DNS server names, mail exchanger (MX) records, and name server (NS) records.
- **Network Information:** It can retrieve information about the network infrastructure of a target organization.

Steps to install and run:-

1. theHarvester tool is pre-installed in Linux distribution. You can run it by just typing theHarvester inside the terminal.

```

root@kali: /home/kali
File Actions Edit View Help

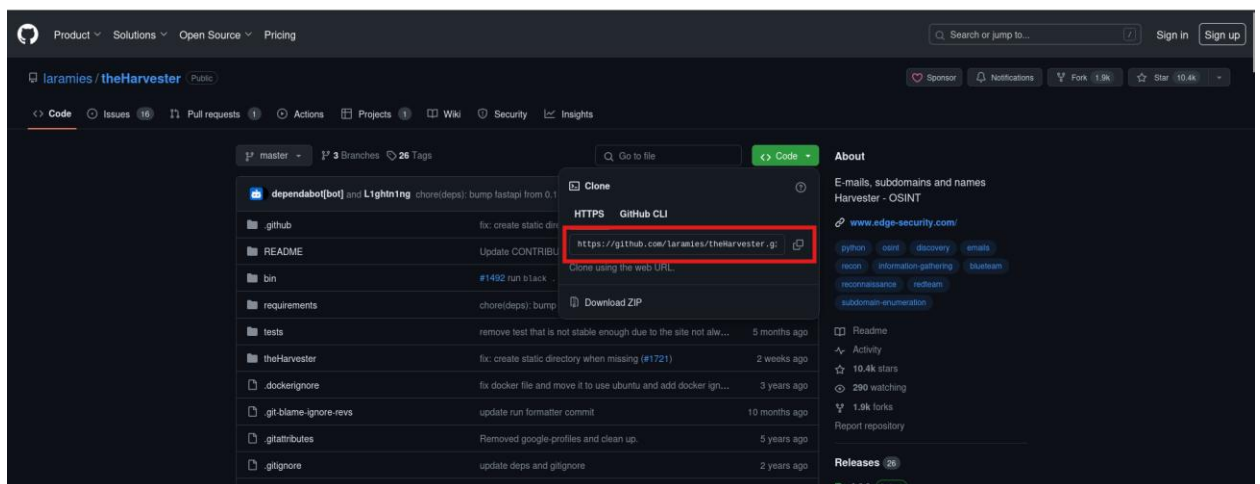
root@kali) [/home/kali]
# theHarvester --help
Created default proxies.yaml at /root/.theHarvester/proxies.yaml
*****
*
* theHarvester
*
* theHarvester 4.5.1
* Coded by Christian Martorella
* Edge-Security Research
* cmartorella@edge-security.com
*
*****
usage: theHarvester [-h] -d DOMAIN [-l LIMIT] [-S START] [-p] [-s] [--screenshot SCREENSHOT] [-v] [-e DNS_SERVER] [-t]
                  [-r [DNS_RESOLVE]] [-n] [-c] [-f FILENAME] [-b SOURCE]

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

options:
-h, --help            show this help message and exit
-d DOMAIN, --domain DOMAIN
                      Company name or domain to search.
-l LIMIT, --limit LIMIT
                      Limit the number of search results, default=500.
-S START, --start START
                      Start with result number X, default=0.
-p, --proxies          Use proxies for requests, enter proxies in proxies.yaml.
-s, --shodan           Use Shodan to query discovered hosts.
--screenshot SCREENSHOT
                      Take screenshots of resolved domains specify output directory: --screenshot output_directory
-v, --virtual-host     Verify host name via DNS resolution and search for virtual hosts.

```

2. If its not installed then, you can install it through its Github repository.



3. Just use git clone command into your terminal with the copied url of the tool.

```
(root@kali)-[/home/kali]
# git clone https://github.com/laramies/theHarvester.git
Cloning into 'theHarvester' ...
remote: Enumerating objects: 14518, done.
remote: Counting objects: 100% (1666/1666), done.
remote: Compressing objects: 100% (261/261), done.
remote: Total 14518 (delta 1488), reused 1531 (delta 1405), pack-reused 12852
Receiving objects: 100% (14518/14518), 7.70 MiB | 9.71 MiB/s, done.
Resolving deltas: 100% (9144/9144), done.
```

4. Now after cloning it move towards its directory where you have clone it.

```
(root@kali)-[/home/kali]
# cd theHarvester

(root@kali)-[/home/kali/theHarvester]
# ls
bin                Dockerfile          README              requirements         restfulHarvest.py   theHarvester        theHarvester-logo.webp
docker-compose.yml pyproject.toml      README.md           requirements.txt     tests               theHarvester-logo.png theHarvester.py
```

5. Now install the requirements.txt file using pip command.

```
(root@kali)-[/home/kali/theHarvester]
# pip install -r requirements.txt
Collecting aiodns==3.2.0 (from -r requirements/base.txt (line 1))
  Downloading aiodns-3.2.0-py3-none-any.whl.metadata (4.0 kB)
Requirement already satisfied: aiofiles==23.2.1 in /usr/lib/python3/dist-packages (from -r requirements/base.txt (line 2)) (23.2.1)
Collecting aiohttp==3.9.5 (from -r requirements/base.txt (line 3))
  Downloading aiohttp-3.9.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.5 kB)
Collecting aiomultiprocess==0.9.1 (from -r requirements/base.txt (line 4))
  Downloading aiomultiprocess-0.9.1-py3-none-any.whl.metadata (4.8 kB)
Collecting aiosqlite==0.20.0 (from -r requirements/base.txt (line 5))
  Downloading aiosqlite-0.20.0-py3-none-any.whl.metadata (4.3 kB)
Requirement already satisfied: beautifulsoup4==4.12.3 in /usr/lib/python3/dist-packages (from -r requirements/base.txt (line 6)) (4.12.3)
Collecting censys==2.2.12 (from -r requirements/base.txt (line 7))
  Downloading censys-2.2.12-py3-none-any.whl.metadata (7.0 kB)
Collecting certifi==2024.2.2 (from -r requirements/base.txt (line 8))
  Downloading certifi-2024.2.2-py3-none-any.whl.metadata (2.2 kB)
Collecting dnspython==2.6.1 (from -r requirements/base.txt (line 9))
  Downloading dnspython-2.6.1-py3-none-any.whl.metadata (5.8 kB)
Collecting fastapi==0.111.0 (from -r requirements/base.txt (line 10))
  Downloading fastapi-0.111.0-py3-none-any.whl.metadata (25 kB)
Collecting lxml==5.2.1 (from -r requirements/base.txt (line 11))
  Downloading lxml-5.2.1-cp311-cp311-manylinux_2_28_x86_64.whl.metadata (3.4 kB)
Collecting netaddr==1.2.1 (from -r requirements/base.txt (line 12))
  Downloading netaddr-1.2.1-py3-none-any.whl.metadata (5.0 kB)
Requirement already satisfied: ujson==5.9.0 in /usr/lib/python3/dist-packages (from -r requirements/base.txt (line 13)) (5.9.0)
Collecting playwright==1.43.0 (from -r requirements/base.txt (line 14))
  Downloading playwright-1.43.0-py3-none-manylinux1_x86_64.whl.metadata (3.5 kB)
Requirement already satisfied: PyYAML==6.0.1 in /usr/lib/python3/dist-packages (from -r requirements/base.txt (line 15)) (6.0.1)
Collecting python-dateutil==2.9.0.post0 (from -r requirements/base.txt (line 16))
```

Now it is ready to use. theHarvester is available with various switches. I am going to demonstrate few of them.

1. **-d** is used to define the target domain while **-b** is used to define the source of the search. Here I have searched for Microsoft.com in all the available search field by using the all parameter in the -b.

```
(root@kali)~/home/kali/theHarvester
# theHarvester -d microsoft.com -b all
Read proxies.yaml from /root/.theHarvester/proxies.yaml
*****
* theHarvester                               *
* theHarvester 4.5.1                         *
* Coded by Christian Martorella              *
* Edge-Security Research                     *
* cmartorella@edge-security.com              *
*****

[*] Target: microsoft.com
Created default api-keys.yaml at /root/.theHarvester/api-keys.yaml
[!] Missing API key for bevigil.
Read api-keys.yaml from /root/.theHarvester/api-keys.yaml
[!] Missing API key for binaryedge.
Read api-keys.yaml from /root/.theHarvester/api-keys.yaml
Read api-keys.yaml from /root/.theHarvester/api-keys.yaml
Read api-keys.yaml from /root/.theHarvester/api-keys.yaml
```

2. Now I just want to search under the scanurl search engine so I have define it inside -b.

```
(root@kali)~/home/kali/theHarvester
# theHarvester -d kali.org -b urlscan
Read proxies.yaml from /root/.theHarvester/proxies.yaml
*****
* theHarvester                               *
* theHarvester 4.5.1                         *
* Coded by Christian Martorella              *
* Edge-Security Research                     *
* cmartorella@edge-security.com              *
*****

[*] Target: kali.org
[*] Searching Urlscan.
[*] ASNs found: 13
AS13335
AS135822
AS14618
AS16819
AS16276
AS1835
AS19858
AS24540
AS47610
AS54113
AS377
AS68733
AS8472

[*] Interesting Urls found: 21
http://archive.kali.org/
```

Here are few more commands for the particular purposes:

- **Search for Email Addresses:** theHarvester -d example.com -l 100 -b google
- **Enumerate Subdomains:** theHarvester -d example.com -l 100 -b Baidu
- **Discover Hostnames:** theHarvester -d example.com -l 100 -b bing
- **Identify Open Ports:** theHarvester -d example.com -b shodan
- **Discover Virtual Hosts:** theHarvester -d example.com -l 100 -b bing -f output.txt
- **Gather DNS Information:** theHarvester -d example.com -b all
- **Retrieve Network Information:** theHarvester -d example.com -l 100 -b netcraft