Guide for Beginners theHarvester Tool

Souleiman Guedi

Content

- What is theHarvester
- Installing theHarvester
- Understanding the Syntax and Basic Options
- Using APIs for Enhanced Results
- Recommendations

What is TheHarvester?

TheHarvester is a powerful and widely used OSINT (Open Source Intelligence) tool that helps Ethical Hackers, Penetration Testers, and cybersecurity professionals collect information related to target domains and organizations from various search engines, databases, and other publicly available services. The primary function of theHarvester is to gather critical information about a target domain, such as:

- Subdomains: Alternative domain addresses linked to the target.
- Emails: Employee or organizational email addresses. One of the most valuable pieces of information you can collect during a reconnaissance phase is a list of email addresses. These emails can later be used for social engineering attacks or identifying potential weak points in security configurations.
- IP Addresses: TheHarvester can map domain names to associated IP addresses. These addresses can be used for further network scanning and vulnerability analysis.
- Hostnames: Additional domain names or services linked to the target.

This information is primarily used during the reconnaissance phase of penetration testing or when conducting security assessments. It gathers information without directly interacting with the target system, minimizing detection.

Installing theHarvester

The initial step involves installing **theHarvester**, a widely used reconnaissance tool. While it comes pre-installed on **Kali Linux**, it can also be manually installed on other Linux distributions or on **Windows** systems. This flexibility allows security professionals to integrate TheHarvester into a variety of operating environments based on their specific needs.

1. **Kali Linux:** TheHarvester comes pre-installed on most penetration testing distributions like Kali Linux.

You can verify the installation by using:

Understanding the Syntax and Basic Options

theHarvester offers a versatile set of command-line options that allow users to customize and refine their reconnaissance activities. Understanding its syntax and commonly used parameters is essential for effective information gathering. Below are some of the key options and their usage:

Common options:
Here's a breakdown of the most commonly used options in TheHarvester:

- **-d <domain>:** Specifies the domain to search.
- **-b <source>:** Defines the data source (e.g., yahoo, bing, shodan). You can specify multiple sources separated by commas.
- **-l <limit>:** Limit the number of results fetched from the data sources.
- **-f <filename>:** Save the output into a file (in HTML format).
- **-n:** Perform DNS enumeration using search results.
- **-t:** Perform DNS TLD expansion.
- **-s <start>:** Start with a specific result number (useful when you want to skip initial results).
- **-v:** Enable verbose mode for more detailed output.

  Basic Usage Example: To search for a subdomain (e.g., tesla.com) using bing as the search engine:

theHarverster

SG-Learning

```
┌──(root㉿SG)-[/home/sg-learning]
└─# theHarvester -d tesla.com -b bing

Read proxies.yaml from /root/.theHarvester/proxies.yaml
*******************************************************************
*  _   _                                                          *
* | |_| |__   ___    ^  /\__ _ _ ____   _____ ___| |_ ___ _ __   *
* | __|  _ \ / _ \  / /_/ / _` | '__\ \ / / _ \ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* theHarvester 4.6.0                                              *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*                                                                 *
*******************************************************************

[*] Target: tesla.com

Created default api-keys.yaml at /root/.theHarvester/api-keys.yaml
        Searching 0 results.
[*] Searching Bing.
```

This command gathers subdomain data using bing and displays the results.

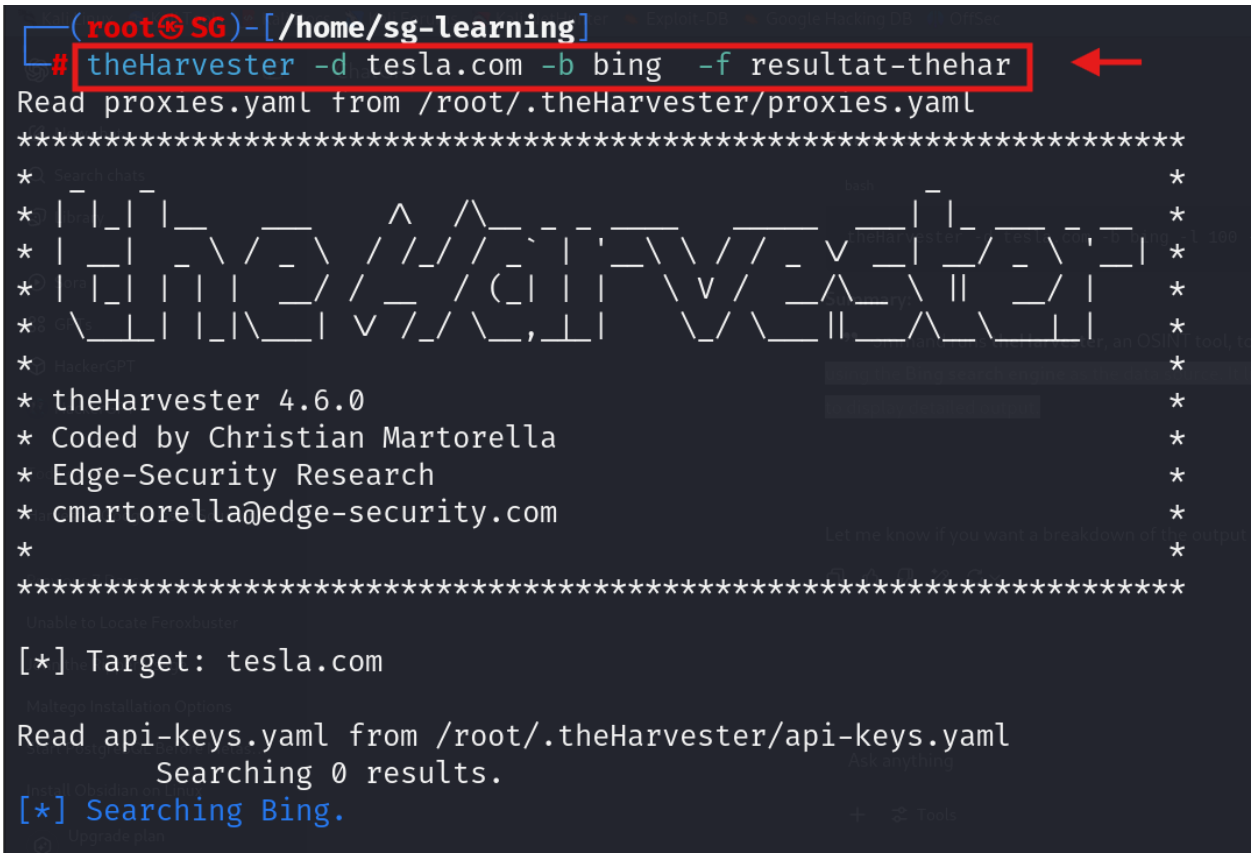
using the Bing search engine as the data source. It limits results to 100 entries and uses verbose mode (-v) to display detailed output.

```
┌──(root㉿SG)-[/home/sg-learning]
└─# theHarvester -d tesla.com -b bing  -l 100 -v
Read proxies.yaml from /root/.theHarvester/proxies.yaml
*******************************************************************
*  _   _                                                          *
* | |_| |__   ___    ^  /\__ _ _ ____   _____ ___| |_ ___ _ __   *
* | __|  _ \ / _ \  / /_/ / _` | '__\ \ / / _ \ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* theHarvester 4.6.0                                              *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*                                                                 *
*******************************************************************

[*] Target: tesla.com

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

You can save your results for future reference by using the -f option followed by the filename.



Using APIs for Enhanced Results

To improve the accuracy and depth of your reconnaissance, **theHarvester** supports integration with various third-party services through API keys—such as **Hunter.io**. Leveraging these APIs can significantly enhance the quality and quantity of the collected data.

Tips for Better Usage

To maximize the effectiveness of **theHarvester**, consider the following practical tips:

- **Use a VPN or Proxy**
  Conduct reconnaissance through a VPN or proxy service to maintain anonymity and avoid IP-based rate limiting from search engines or APIs.
- **Combine Multiple Data Sources**
  Use multiple search engines or APIs (-b option) in a single query to gather more comprehensive data. Each source may return unique results.

- **Limit and Filter Results**
  Apply options like `-l` (limit) and `-f` (save to file) to manage output size and organize findings efficiently for later analysis.
- **Integrate into Automated Workflows**
  Incorporate theHarvester into automated reconnaissance or red team scripts to streamline data collection processes.
- **Respect Target Scope**
  Ensure that all targets are within your authorized testing scope to avoid legal or ethical violations.

Practical Recommendations

To fully leverage theHarvester in real-world scenarios, keep the following recommendations in mind:

- Stay Updated
  Regularly update the tool to benefit from the latest bug fixes, data source support, and feature enhancements.
- Use API Keys for Advanced Results
  Configure and use API keys (e.g., for Hunter.io, Bing, or Shodan) to unlock deeper intelligence and bypass limitations of free sources.
- Validate Collected Data
  Cross-reference harvested information with other tools (e.g., Maltego, Recon-ng) to verify accuracy and expand your dataset.
- Document Findings
  Store and organize your output in structured formats (e.g., CSV, JSON, or Markdown) for reporting or further exploitation.
- Incorporate into OSINT Training
  TheHarvester is an excellent educational tool. Use it in labs or cybersecurity training environments to teach reconnaissance techniques.