theHarvester

Bernd Storck

Grundlagen / Einführung

`theHarvester` ist ein Open-Source-Sicherheitstool, das zur Informationsbeschaffung (Reconnaissance) verwendet wird. Es wird hauptsächlich von Sicherheitsexperten und Penetrationstestern eingesetzt, um Daten über ein bestimmtes Ziel, in der Regel eine Domain oder IP-Adresse, zu sammeln. Hier sind die grundlegenden Aspekte von `theHarvester`:

Funktionen von `theHarvester`

1. Suchmaschinen-Abfrage:
   - Das Tool kann Suchmaschinen wie Bing, Yahoo und andere nutzen, um Informationen über E-Mail-Adressen, Subdomains, Hosts und mehr zu sammeln.
2. Social Media-Scraping:
   - `theHarvester` kann Informationen von verschiedenen sozialen Netzwerken wie LinkedIn und Twitter abrufen, um Netzwerkinformationen über das Ziel zu erhalten.
3. DNS-Abfragen:
   - Es kann DNS-Abfragen durchführen, um Subdomains und IP-Adressen zu identifizieren.
4. Web-Server-Integrationen:
   - Das Tool kann auch Informationen von verschiedenen Web-Servern abrufen, die mit dem Ziel in Verbindung stehen.

Installation

`theHarvester` kann auf verschiedenen Betriebssystemen installiert werden, oft in einer Python-Umgebung. Um es zu installieren, kannst du folgendes tun:

```
git clone https://github.com/laramies/theHarvester.git
cd theHarvester
pip install -r requirements.txt
```

Verwendung

Ein einfaches Beispiel für die Verwendung von `theHarvester`:

`python3 theHarvester.py -d <Ziel-Domain> -b <Suchmaschine>`

- `-d <Ziel-Domain>`: gibt die Domain an, die du untersuchen möchtest.
- `-b <Suchmaschine>`: wählt die Suchmaschine aus, die für die Abfragen verwendet werden soll (z.B. `google`, `bing`, etc.).

Beispiel

Wenn du Informationen über die Domain `example.com` mit Google sammeln möchtest, würdest du den folgenden Befehl verwenden:

`python3 theHarvester.py -d example.com -b google`

Wichtige Hinweise

- Nutzungsrichtlinien: Stell sicher, dass du beim Sammeln von Informationen die rechtlichen und ethischen Richtlinien einhältst. Verwende `theHarvester` nur auf Zielsystemen, für die du autorisiert bist.
- Feedback und Verbesserungen: Da `theHarvester` ein Open-Source-Projekt ist, ist die Community aktiv und nimmt Feedback entgegen, um das Tool ständig zu verbessern.

Installationsprobleme

Die Version von `theHarvester` in dne Repositorien der eigenen Linux-Distribution ist nicht unbedingt aktuell. Die offizielle Version von `theHarvester` wird auf GitHub gepflegt.

2. Installationsmethode

Die empfohlene Methode zur Installation von `theHarvester` ist das Klonen des GitHub-Repositories und die direkte Ausführung mit Python.

3. Ausführbare Datei nicht im PATH

Wenn Sie `pyenv` verwenden, wird die ausführbare Datei möglicherweise in einem Verzeichnis installiert, das nicht in Ihrem PATH enthalten ist. In Ihrem Fall sollte die ausführbare Datei in `/home/bernds/.pyenv/versions/3.11.6/bin/` liegen, aber dieses Verzeichnis ist möglicherweise nicht in Ihrem PATH enthalten.

Lösungsvorschläge:

1. Installieren Sie die offizielle Version:

   ```
   git clone https://github.com/laramies/theHarvester.git
   cd theHarvester
   pip install -r requirements.txt
   ```

2. Überprüfen Sie den PATH: Fügen Sie das Verzeichnis, in dem `theHarvester` installiert ist, zu Ihrem PATH hinzu. Ergänzen Sie Ihre `.bashrc` Datei um folgende Zeile:

   ```
   export PATH="$HOME/.pyenv/versions/3.11.6/bin:$PATH"
   ```

   Führen Sie danach `source ~/.bashrc` aus, um die Änderungen zu aktivieren.

3. Lokalisieren Sie die ausführbare Datei: Verwenden Sie den `find`-Befehl, um nach der ausführbaren Datei zu suchen:

   ```
   find ~/.pyenv -name "theHarvester*"
   ```

4. Überprüfen Sie die pyenv-Umgebung: Stellen Sie sicher, dass die richtige Python-Version aktiviert ist:

   ```
   pyenv versions
   pyenv global 3.11.6
   ```

5. Ausführen von theHarvester: Nach der Installation aus dem GitHub-Repository können Sie `theHarvester` direkt mit Python ausführen:

   ```
   python3 theHarvester.py -h
   ```

6. Überprüfen Sie die Abhängigkeiten: Stellen Sie sicher, dass alle erforderlichen Abhängigkeiten installiert sind. Dies wird normalerweise durch die Installation der `requirements.txt` aus dem GitHub-Repository erledigt.

Wenn Sie nach diesen Schritten noch Probleme haben, überprüfen Sie die Ausgabe von `which theHarvester` und `pyenv which theHarvester`, um zu sehen, wo das System die ausführbare Datei erwartet.

Praxis / Benutzung

Die Hilfeseite von theHarvester

```
*******************************************************************
*  _   _                                             _            *
* | |_| |__   ___    /\  /\__ _ _ ____   _____  ___| |_ ___ _ __  *
* | __|  _ \ / _ \  / /_/ / _` | '__\ \ / / _ \/ __| __/ _ \ '__| *
* | |_| | | |  __/ / __  / (_| | |   \ V /  __/\__ \ ||  __/ |    *
*  \__|_| |_|\___| \/ /_/ \__,_|_|    \_/ \___||___/\__\___|_|    *
*                                                                 *
* theHarvester 4.7.0                                              *
* Coded by Christian Martorella                                   *
* Edge-Security Research                                          *
* cmartorella@edge-security.com                                   *
*                                                                 *
*******************************************************************
usage: theHarvester.py [-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.
  -e DNS_SERVER, --dns-server DNS_SERVER
                        DNS server to use for lookup.
  -t, --take-over       Check for takeovers.
  -r [DNS_RESOLVE], --dns-resolve [DNS_RESOLVE]
                        Perform DNS resolution on subdomains with a resolver
                        list or passed in resolvers, default False.
  -n, --dns-lookup      Enable DNS server lookup, default False.
  -c, --dns-brute       Perform a DNS brute force on the domain.
  -f FILENAME, --filename FILENAME
                        Save the results to an XML and JSON file.
  -b SOURCE, --source SOURCE
                        anubis, baidu, bevigil, binaryedge, bing, bingapi,
                        brave, bufferoverun, censys, certspotter, criminalip,
                        crtsh, duckduckgo, fullhunt, github-code,
                        hackertarget, hunter, hunterhow, intelx, netlas,
                        onyphe, otx, pentesttools, projectdiscovery, rapiddns,
                        rocketreach, securityTrails, sitedossier,
                        subdomaincenter, subdomainfinderc99, threatminer,
                        tomba, urlscan, virustotal, yahoo, zoomeye
```

Erläuterung einzelner Optionen

```
-p, --proxies         Use proxies for requests, enter proxies in proxies.yaml.
-s, --shodan          Use Shodan to query discovered hosts.
-t, --take-over       Check for takeovers.
-r [DNS_RESOLVE], --dns-resolve [DNS_RESOLVE]
                      Perform DNS resolution on subdomains with a resolver list or passed in
                      resolvers, default False.
-n, --dns-lookup      Enable DNS server lookup, default False.
-c, --dns-brute       Perform a DNS brute force on the domain.
```

1. `-p, --proxies`
   Diese Option ermöglicht die Nutzung von Proxys, um Anfragen zu verschleiern. Besonders nützlich, wenn du häufige Abfragen durchführst und vermeiden möchtest, dass eine Quelle deine IP blockiert.
   - Proxys werden in einer Datei namens `proxies.yaml` angegeben.
   - Beispielsweise kannst du über diesen Ansatz geografische Einschränkungen umgehen oder deine Identität schützen.
2. `-s, --shodan`
   Mit dieser Option werden Daten von Shodan, einer Suchmaschine für verbundene Geräte (wie Server, Router etc.), abgefragt.
   - Du kannst Informationen zu Hosts wie offene Ports, Betriebssysteme und mehr sammeln.
   - Voraussetzung: ein Shodan-API-Schlüssel, den du unter Shodan.io generieren kannst.
3. `-t, --take-over`
   Diese Funktion prüft, ob potenzielle Subdomain-Takeover-Schwachstellen existieren.
   - Ein Subdomain-Takeover tritt auf, wenn eine Subdomain auf einen nicht mehr genutzten Hosting-Dienst verweist, der von Angreifern übernommen werden kann.
   - Besonders nützlich bei Sicherheitsanalysen!
4. `-r [DNS_RESOLVE], --dns-resolve [DNS_RESOLVE]`
   Diese Option führt DNS-Auflösungen für Subdomains durch.
   - Du kannst eine eigene Resolver-Liste verwenden (falls angegeben) oder die Standard-DNS-Server verwenden.
   - Eignet sich, um IP-Adressen hinter Subdomains zu identifizieren und mögliche Dienste aufzudecken.
5. `-n, --dns-lookup`
   Aktiviert die DNS-Lookup-Funktion.
   - DNS-Lookups ermöglichen das Abfragen der DNS-Einträge einer Domain, um Details wie IP-Adressen oder MX-Einträge (für E-Mail-Server) zu extrahieren.
   - Perfekt, um grundlegende Informationen über eine Domain zu erhalten.
6. `-c, --dns-brute`
   Hierbei wird ein Brute-Force-Angriff auf DNS durchgeführt, um Subdomains zu entdecken.
   - Dazu wird eine vordefinierte Liste von möglichen Subdomain-Namen ausprobiert (z. B. `www`, `mail`, `test`).
   - Kann versteckte Subdomains aufdecken, die nicht in anderen Quellen auftauchen.

Quellen, die Harvester nutzen kann

Quelle	Fokus & Besonderheiten	Eignung für theHarvester
anubis	Domain- und E-Mail-OSINT; grundlegende Informationsquelle	Für erste, grobe Informationen zu Domains und zugehörigen E-Mail-Adressen
baidu	Chinesische Suchmaschine; liefert hauptsächlich Ergebnisse aus dem chinesischsprachigen Raum	Gut, wenn du Daten zu chinesischen Domains sammeln möchtest
bevigil	Aggregierte Domain- und Netzwerkdaten; experimentell	Ergänzt andere Quellen, kann zusätzliche Hinweise liefern

Quelle	Fokus & Besonderheiten	Eignung für theHarvester
binaryedge	Scan-Daten und Host-Informationen; liefert technische Details (oft API-gebunden)	Sehr nützlich für tiefergehende technische Recherchen, wenn API-Zugang besteht
bing	Etablierte Suchmaschine; erfasst allgemeine OSINT-Daten wie Subdomains, E-Mails, Hosts	Zuverlässige, breit gefächerte Quelle für passive Daten
bingapi	API-basierter Zugriff auf Bing-Daten; strukturiert und für Automatisierung geeignet	Ideal, wenn du automatisierte Abfragen mit Bing durchführen möchtest (API-Schlüssel erforderlich)
brave	Privacy-orientierte Suchmaschine; alternative allgemeine Suchergebnisse	Ergänzt andere Suchmaschinen; liefert oftmals etwas andere Ergebnisse
bufferoverun	DNS-basierte Subdomain-Erkennung; spezialisiert auf DNS-Daten	Sehr effektiv zur Erfassung von Subdomains
censys	Umfassende Internet-Scan-Daten; technisch orientiert (häufig API-gebunden)	Hervorragend für detaillierte Sicherheits- und Host-Informationen (API-Zugang oft nötig)
certspotter	Sammelt Daten aus Zertifikatstransparenz-Logs (CT-Logs)	Exzellent für das Aufspüren von Subdomains und teils auch E-Mail-Daten
criminalip	Informationen zu IP-Adressen und deren Reputation	Kann ergänzend genutzt werden, Ergebnisse variieren aber oft in der Tiefe
crtsh	Nutzt CT-Logs zur Sammelung von Zertifikatsdaten	Sehr beliebt und oft sehr effektiv für Subdomain-Erkennung
duckduckgo	Datenschutzfreundliche allgemeine Suchmaschine	Alternative zu anderen Suchmaschinen, liefert allgemeine OSINT-Daten, aber oft weniger spezialisierte Ergebnisse
fullhunt	Plattform für Bug-Bounty- und Sicherheitsforscher; aggregiert diverse OSINT-Daten	Vielseitig einsetzbar für ein breites Spektrum an OSINT-Informationen
github-code	Durchsucht öffentliche GitHub-Repositories auf Hinweise (z. B. sensible Daten oder Fehlkonfigurationen)	Gut, um Leaks oder versehentlich veröffentlichte Konfigurationen und Kontaktdaten aufzuspüren
hackertarget	Aggregierte OSINT-Daten aus diversen Quellen	Liefert oft gute erste Einblicke und breit gefächerte Daten
hunter	Spezialisiert auf die Suche nach Unternehmens-E-Mail-Adressen	Sehr gut geeignet, wenn du gezielt nach Kontakt-E-Mails recherchieren möchtest
hunterhow	Ähnlich wie Hunter; unterstützt die E-Mail-Suche	Ergänzt hunter, wenn die Standardquelle einmal nicht ausreicht
intelx	Breite Aggregation von OSINT-Daten; holt Informationen aus vielen Quellen (häufig API-gebunden)	Bietet ein breites Spektrum an Informationen, eignet sich aber vor allem für automatisierte Workflows (API nötig)
netlas	Liefert Scan-Daten und technische Informationen zu Hosts	Gut für detaillierte technische Recherchen
onyphe	Französische OSINT-Plattform, oft regional fokussiert	Besonders nützlich bei Recherchen im französischsprachigen Raum oder in spezielleren Regionen
otx	AlienVault OTX bietet globale Bedrohungsdaten, inkl. Domain- und Host-Suchen	Ideal, um Sicherheitswarnungen und Angriffsindikatoren zu identifizieren
pentesttools	Sammlung diverser OSINT-Tools	Eignet sich als ergänzende Quelle, um erste Recherchen zu unterstützen
projectdiscovery	Anbieter moderner Sicherheits- und OSINT-Tools; liefert oft aktuelle und hochwertige Daten	Sehr wertvoll für präzise und aktuelle Sicherheitsrecherchen

Quelle	Fokus & Besonderheiten	Eignung für theHarvester
rapiddns	Fokussiert auf DNS-basierte Subdomain-Erkennung; sammelt speziell DNS-Daten	Sehr effektiv und zielgerichtet, wenn es ausschließlich um Subdomain-Ermittlung geht geht
rocketreach	Sucht nach Kontaktdaten (E-Mail, Telefon), oft für Unternehmensprofile genutzt	Gut für die Suche nach Kontaktdaten, jedoch oft in der freien Nutzung limitiert
securityTrails	Bietet umfangreiche, historische und aktuelle DNS-, Domain- und IP-Daten; sehr ausführlich (häufig API-gebunden)	Exzellent für tiefgehende und langfristige Untersuchungen, ideal für detaillierte OSINT-Analysen (API erforderlich)
sitedossier	Analysiert Webseiten- und Domain-Daten umfassend	Nützlich, um ein ganzheitliches Profil einer Zielseite zu erstellen
subdomaincenter	Spezialisierte Subdomain-Erkennung	Kann gezielt ergänzend eingesetzt werden, wenn es um das Auffinden von Subdomains geht
subdomainfinder99	Nutzt mehrere Quellen, um Subdomains zu ermitteln	Ergänzt die anderen Quellen und schließt Lücken in der Subdomain-Erfassung
threatminer	Aggregiert Daten zu Bedrohungen, Domains, IPs und E-Mails	Hilfreich bei der Sicherheitsanalyse und zur Erkennung von Zusammenhängen zwischen verschiedenen OSINT-Daten
tomba	Fokussiert auf die Suche nach E-Mail-Adressen, ähnlich zu Hunter	Sehr gut geeignet, um gezielt Unternehmenskontakte zu extrahieren
urlscan	Führt Webseitenscans durch (inkl. Screenshots) und analysiert Domains	Nützlich zur visuellen und technischen Überprüfung von Webseiten und als ergänzende Informationsquelle
virustotal	Aggregiert Informationen aus Datei-, URL- und Domain-Scans; vor allem für Malware-Analysen	Hilfreich, um zu prüfen, ob eine Domain mit schädlichen Aktivitäten in Verbindung gebracht wird
yahoo	Klassische Suchmaschine	Kann ergänzende OSINT-Daten liefern, ist aber weniger spezialisiert im Vergleich zu anderen Suchanbietern
zoomeye	Ähnlich zu Shodan; liefert detaillierte technische Informationen zu Hosts, offenen Ports und Diensten (API oft erforderlich)	Sehr nützlich für technisch orientierte Recherchen und zum Auffinden von verbundenen Sicherheitslücken

---

Hinweise:

- API-Zugang: Einige Quellen (z. B. binaryedge, censys, intelx, securityTrails, zoomeye) erfordern möglicherweise API-Schlüssel oder haben Nutzungslimits.
- Ergänzende Nutzung: In der Regel ist es sinnvoll, mehrere Quellen zu kombinieren, um ein umfassenderes Bild zu erhalten.
- Zielabhängigkeit: Die Eignung einzelner Quellen kann stark von der Zieldomain und dem geografischen/inhaltlichen Kontext abhängen.