

Benefits and Use of Network Reconnaissance Tool Spiderfoot












Russell Dranch
11/15/2020
Rochester Institute of Technology

*All images, graphs and charts were obtained from the Spiderfoot site

Spiderfoot is an OSINT automation tool designed for asset discovery, attack surface monitoring, security assessments, and threat intelligence. Written in Python 3, Spiderfoot integrates an extremely wide range of data sources and leverages various discovery-based modules in order to conduct data analysis.

While the command line version of the tool is [extremely well documented](#), the Web interface (or HX) version [lacks some](#) and the goal of this paper will be to explain and walkthrough the steps when conducting a scan.

The most important thing to know before starting to use an unfamiliar tool is where your limitations stand. With that being said, Spiderfoot breaks down its tool into tiers. The tiers can be shown here (prices by month):

Features	Hobby Free	Standard €59	Professional €179	Enterprise €599
Scans per month 	3	10	50	200
User accounts 	1	2	5	Unlimited
Scan duration limit 	4 hours	24 hours	72 hours	120 hours
Targets per scan 	1	256	1024	4096
Data retention 	90 days	12 months	12 months	24 months
Correlations 	✓	✓	✓	✓
Support 	✗	✓	✓	✓
Investigations 	✗	✓	✓	✓
Attack Surface Monitoring 	✗	✓	✓	✓
SpiderFoot HX API 	✗	✓	✓	✓
Screenshotting 	✗	✗	✓	✓

- Scans per month are counted by anything run under the “scans” tab or any scan run by a monitor you set up. Investigations are *not* counted towards your monthly quota.
 - ◆ Monitors usually run on intervals of daily, weekly, or monthly. Changes between each scan are tracked and alerted.
- User accounts are how many users can be included in your web scan at a time.
- Scan duration limit is the amount of time a scan can be run before being aborted.
- Targets are, as the name implies, how many targets can be scanned at one time.
 - ◆ A target is an ip address, subnet, ASN, email address, name, domain name, etc.
 - All of the above are considered one target except for a subnet which is considered the amount of targets in that subnet.
 - I.e subnet 1.1.1.0/24 is 255 targets
- Data retention is how long data is stored on the Spiderfoot servers before being removed.
- Correlations match and compare data found to other sources in order to identify interesting information.
- Support includes customer support from Spiderfoot staff.
- Investigations are non-automated, step-by-step manual scans you perform that allow you the ability to take control of the scan.
- Attack Surface Monitoring detects when information about your target becomes available.
- Spiderfoot HX API can automatically trigger scans and fetch data.
- Screenshotting automatically captures information about interesting information.
- ***Modules might have their own individual price per API - not all are free nor do all come with the Spiderfoot package(s).**

Scans

The screenshot shows the Spiderfoot HX Scans interface. At the top, there is a search bar labeled 'Search Bar' and a status filter labeled 'Status Filter'. Below these is a table of scans. The table has columns: Name, Target, Started, Finished, User, Status, Elements, Correlations, and Action. A single scan is listed with the name 'Example', target 'example.com', started at '2020-11-07 18:52:54', finished 'Not yet', user 'rdranch275...', and status 'RUNNING'. The 'Elements' column shows '1830' and 'Correlations' shows '-'. Above the table, there are buttons for 'New Scan' (a plus icon), 'Re-run/Stop Selected' (a circular arrow icon), 'Compare Selected' (a double arrow icon), 'Refresh/Export' (a refresh icon), and 'Delete' (a trash icon). The text 'Total Scans: 1' is at the bottom left.

All scans can be found under the “scan” section of Spiderfoot HX. The scan being conducted was under the “Hobby” package so only the results and information from the “Hobby” package will be discussed.

To create a scan, you will need to press the “new scan” button located on the right. From there, Spiderfoot will present you with various data and options that you will need to sift through and select to your needs.

New Scan

Scan Name & Targets

Help

Import

Iteration

Modules

Options

Run Scan Now


Save as Scan Profile...

Apply Scan Profile...

Hobby Subscription

Scans this month (limit)	1 (3)
Individual scan duration limit	4 hours
Maximum targets per scan	1
Scan data retention	90 days

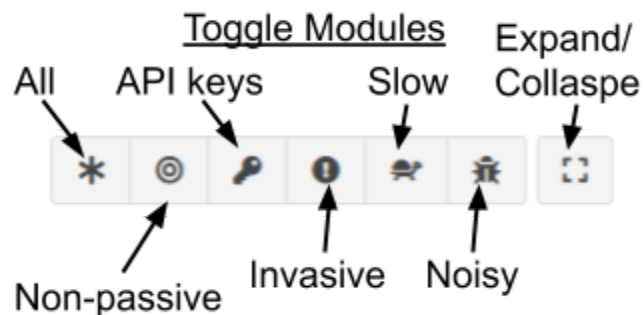
API Keys



You don't have any API keys configured.
Consider [adding some](#) for wider reach.

- Scan Name & Targets
 - This field allows you to identify a “target” (as defined earlier) and provide the scan a meaningful name that will allow you or other researchers to identify what is being scanned.
 - It is possible to import information from a single column list or Hunchly.
- Iteration
 - Iteration gives the user the option to select:
 - Entity Iteration
 - Entity iteration will attempt to identify relevant or similar information discovered during the scanning process. Without this option, applied modules will only be tested against the target.
 - Affiliate Iteration
 - Affiliate iteration will attempt to identify information about affiliates during the scanning process.
 - Co-host Iteration
 - Co-host iteration will attempt to identify information about co-hosts during the scanning process.
 - Human name Iteration
 - Human name iteration will attempt to identify information about names during the scanning process.
- Modules

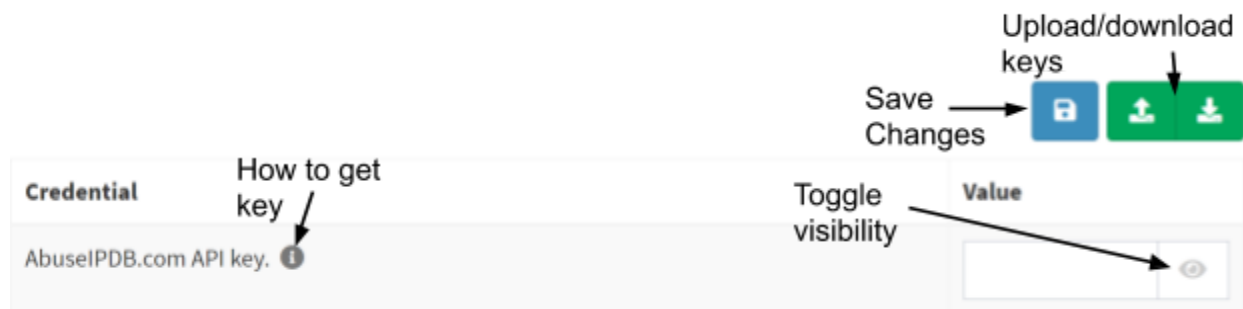
- Modules can be broken down further into subsections:
 - Content Analysis
 - Crawling and Scanning
 - DNS
 - Leaks, Dumps, and Breaches
 - Passive DNS
 - Public Registries
 - Real World
 - Reputation Systems
 - Search Engines
 - Secondary Networks
 - Social Media
- Modules can be searched for in the “search modules” bar at the top
- Toggles are available to filter out unwanted or unnecessary modules quickly



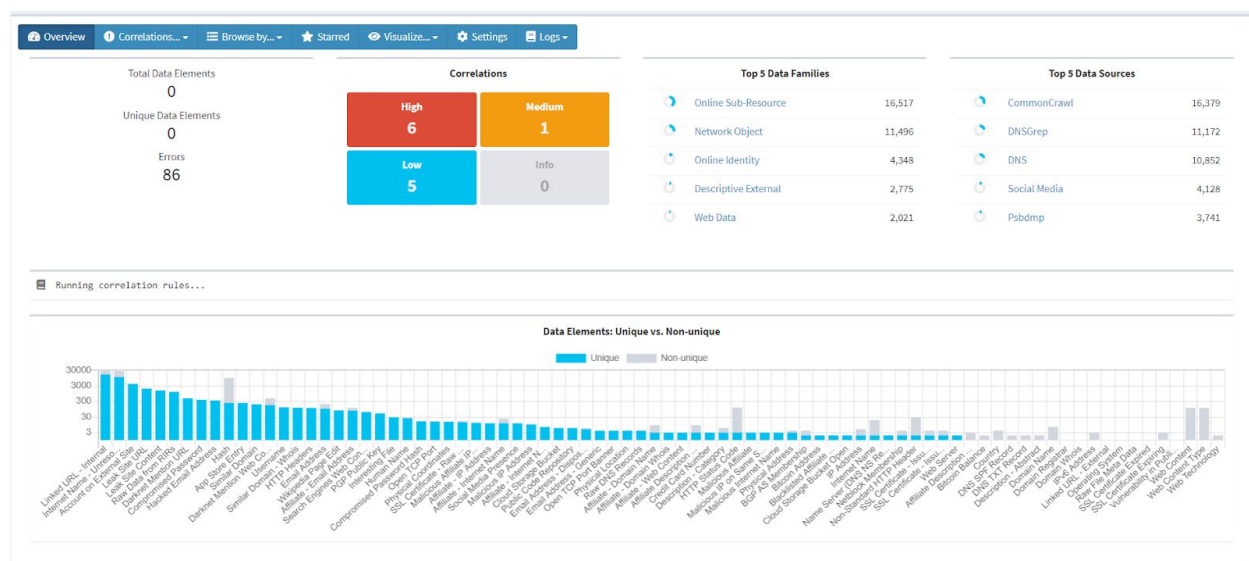
- Similar symbols can be seen above the module to determine what it does. The following module is slow, noisy, invasive, and non-passive. All invasive modules are non-passive but not all non-passive modules are invasive.



- Options
 - Correlation
 - Rules to correlate interesting data.
 - Screenshotting
 - Automatically captures information about interesting information.
 - Email notification on completion
 - Emails the user(s) once the scan has been completed or if an issue occurs that halts the scan.
- Hobby Subscription
 - The scan limits outlined in your plan subscription.
- API Keys



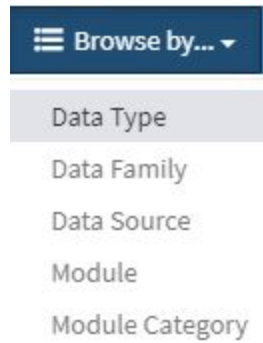
Spiderfoot breaks down the information into the above sections. Each section displays various information.



*photo taken while scan was in progress/aborted

Above is the *Overview* of the Spiderfoot scan tab. Information is broken down into many parts to better graphically and visually identify key information. The number of unique and total elements found are represented in numerical form on the left, top list form on the right, and graphically below. The graph in the lower portion of the picture is broken down further (apart from unique vs. non-unique) by modules. Every module that did not require a key was used here.

The *Browse By* tab is broken up into the following subsections. While there is too much information to go through just here, I will be focusing my attention on the Module tab as this option has some of the most interesting data.



A list of the tested modules can be found here:

- ARIN
- Account Finder
- Ahmia
- Apple iTunes
- Azure Blob Finder
- BGPView
- Certificate Transparency
- Country Name Extractor
- Crobat API
- DNS Brute-forcer
- DNS Look-aside
- DNS Raw Records
- DNS Resolver
- DNSGrep
- DuckDuckGo
- E-Mail Address Extractor
- Flickr
- Github
- Maltiverse
- Mnemonic PassiveDNS
- Open Bug Bounty
- Open Passive DNS Database
- Port Scanner- TCP
- RIPE
- SSL Certificate Analyzer
- Scylla
- Similar Domain Finder
- SpiderFoot UI
- ThreatCrowd
- Tool - Nmap
- Trumail
- Watchguard
- Web Spider
- Whois
- grep.app

The modules tested here can provide interesting information and are all common in the security industry. What makes Spiderfoot so powerful is its ability to handle this wide-range testing completely automatically. Spiderfoot breaks down data into three categories so you can sort through information that was determined to be “important”, first.

The module ARIN is shown below. Spiderfoot has broken down the information to allow you to see risky data, unique data, total data, and the percent of the total that this module makes up [so far]. Similar to the main page, you are able to export and search through the data here with the tabs located on the top right (refer to earlier picture).

Module	Risky	Unique	Total	% of Data
ARIN	0	170	207	2%

Correlation	Data Elements	Criticality	Created
Cloud storage bucket for internal use widely open: example-files.s3-external-1.amazonaws.com: 20 files found. ①	1	HIGH	2020-11-07 23:44:47

The *Correlations* tab is broken down into different ratings consisting of: high, medium, low, and info. Correlations match and compare data found to other sources in order to identify interesting information. Each correlation has a piece of associated information including the number of elements found, how serious the information is, and when the correlation was created. When you click on the correlation, more information is available.



By clicking on any of direct children, correlations, distance from target, or the details tab (blue button), the following popup appears which shows the above information in a more detailed format.

Details:

Shows the data by:

- Type
- Starred
- Module
- Raw data
- Data family
- Risky
- Source

idontwannasignup@example.com:ha4793FjcMy [8tracks.com] ☆

Details

Relationships

Direct Children (0)

Instances (1)

Discovery Path

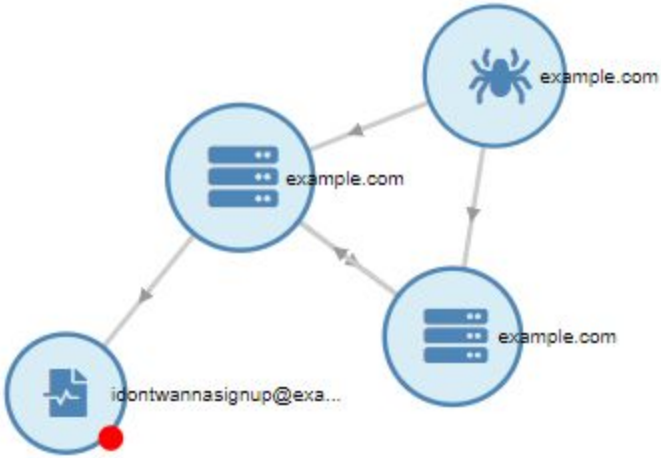
Correlations (0)

Annotation

Data Type	Compromised Password	Data Family	Status Description
Starred	☆ No	Risky Data Type?	⚠ Yes
Source Module(s)	Scylla	Data Source(s)	Scylla
Raw Data	idontwannasignup@example.com:ha4793FjcMy [8tracks.com]		

Relationships:

Shows the connections of the data between various elements. Where the data came from and where it's going in a graphical format is shown here.



Discovery path:

Shows the path taken to get to the data point. You are able to click on any of the data points here and see where they came from in more detail (similar to the details tab of this data).



The *Settings* tab provides information regarding (almost) every aspect of the scan. All things from meta information to modules options are here for the user to view and/or

modify. This information is useful for researchers who might want to configure their search to be more exclusive.

Meta Information	
Name:	Example
Internal ID:	a33d0b4fc4beabf134649dcc7b82c475d6de2747700182837804999179eef30
Target(s):	example.com
Started:	2020-11-07 18:52:55
Completed:	Not yet
Status:	RUNNING
User:	
Scanner ID:	
Global Settings	
Option	Value
Iterate Affiliates	Enabled
Slack webhook URL for scan completion notifications.	
User-Agent string to use for HTTP requests or supply a URL to load the list from there.	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:62.0) Gecko/20100101 Firefox/82.0
Correlations	Enabled
Additional e-mail on completion	
List of usernames that if found as usernames or as part of e-mail addresses, should be treated differently to non-generics.	abuse,admin,billing,compliance,devnull,dns,ftp,hostmaster,noc,ispfeedback,ispupport,list-request,list,maildaemon,marketing,noc,no-reply,noreply,null,peering,peering-notify,peering-request,phish,phishing,postmaster,privacy,registrar,registry,root,routing-registry,rr,sales,security,spam,support,sysadmin,tech,undisclosed-recipients,unsubscribe,usenet,uucp,webmaster,www
Modules enabled for the scan.	sfp__stor_db sfp_abusech sfp_accounts sfp_adblock sfp_ahmia sfp_allenvaultiprep sfp_apple_itunes