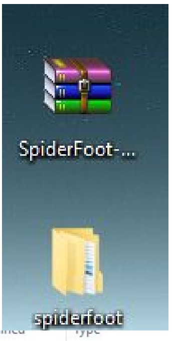
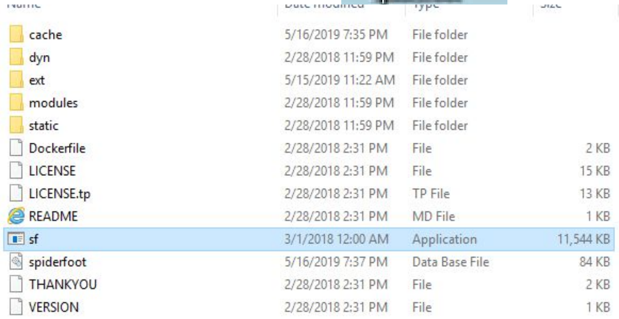
Alec Miller
Spiderfoot Project

SpiderFoot

When first downloading Spiderfoot a zip file will be downloaded. With this zip file, I extracted the contents of it into a new folder, which I named “spiderfoot” for easy memorization.

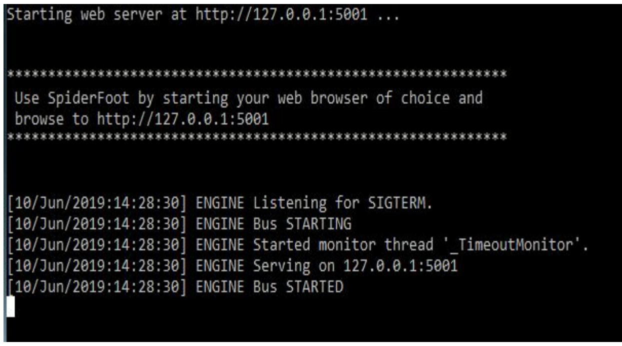


When opening the new folder with all the extracted files from the .zip there will be lots of random folders and files. In the mess of all these files there is one application called “sf” which is short for “Spiderfoot”. Opening this application starts running the Spiderfoot program.



Name	Date modified	Type	Size
cache	5/16/2019 7:35 PM	File folder	
dyn	2/28/2018 11:59 PM	File folder	
ext	5/15/2019 11:22 AM	File folder	
modules	2/28/2018 11:59 PM	File folder	
static	2/28/2018 11:59 PM	File folder	
Dockerfile	2/28/2018 2:31 PM	File	2 KB
LICENSE	2/28/2018 2:31 PM	File	15 KB
LICENSE.tp	2/28/2018 2:31 PM	TP File	13 KB
README	2/28/2018 2:31 PM	MD File	1 KB
sf	3/1/2018 12:00 AM	Application	11,544 KB
spiderfoot	5/16/2019 7:37 PM	Data Base File	84 KB
THANKYOU	2/28/2018 2:31 PM	File	2 KB
VERSION	2/28/2018 2:31 PM	File	1 KB

When first opening Spiderfoot, the user is welcomed by a command prompt. Unlike a normal command prompt, this one doesn’t allow for typing of any sort, which can be confusing at first glance. Yet there is hope to end this confusion! Between the asterisks in the command prompt there is a URL.



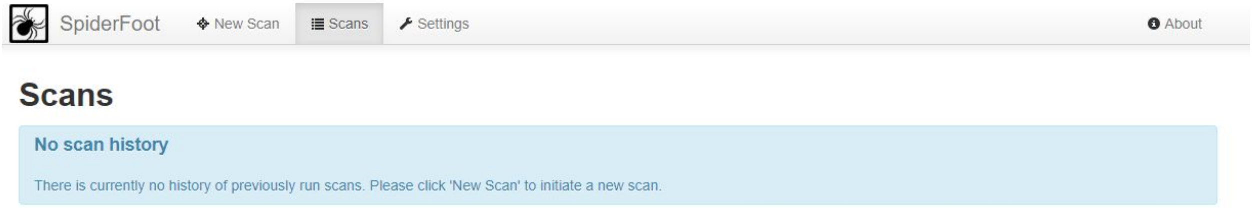
```
Starting web server at http://127.0.0.1:5001 ...


*****************************************************************
 Use SpiderFoot by starting your web browser of choice and
 browse to http://127.0.0.1:5001
*****************************************************************



[10/Jun/2019:14:28:30] ENGINE Listening for SIGTERM.
[10/Jun/2019:14:28:30] ENGINE Bus STARTING
[10/Jun/2019:14:28:30] ENGINE Started monitor thread '_TimeoutMonitor'.
[10/Jun/2019:14:28:30] ENGINE Serving on 127.0.0.1:5001
[10/Jun/2019:14:28:30] ENGINE Bus STARTED
```

With this newly found URL (everything from “http” to the last number following it, in this case it’s “5001”), the user can copy and paste that information into a web browser to begin using Spiderfoot to its fullest potential. After copy and pasting the URL into a preferred web browser the website should look like:

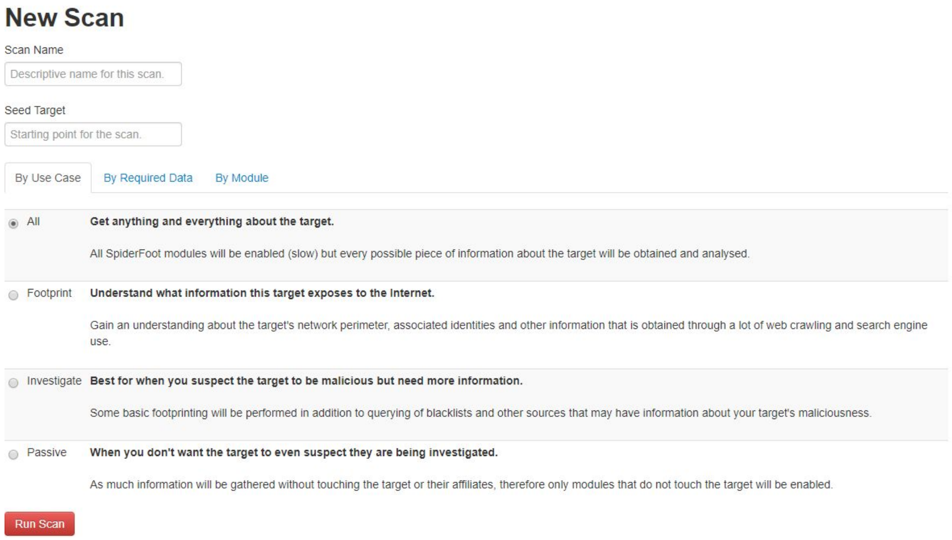


Scans

No scan history

There is currently no history of previously run scans. Please click 'New Scan' to initiate a new scan.

There are no previous scans in the history, since one hasn’t been run yet. This can all be changed when going into the “New Scan” tab such as the one shown below.



New Scan

Scan Name

Descriptive name for this scan.

Seed Target

Starting point for the scan.

By Use Case | By Required Data | By Module

● All — **Get anything and everything about the target.**

All SpiderFoot modules will be enabled (slow) but every possible piece of information about the target will be obtained and analysed.

○ Footprint — **Understand what information this target exposes to the Internet.**

Gain an understanding about the target's network perimeter, associated identities and other information that is obtained through a lot of web crawling and search engine use.

○ Investigate — **Best for when you suspect the target to be malicious but need more information.**

Some basic footprinting will be performed in addition to querying of blacklists and other sources that may have information about your target's maliciousness.

○ Passive — **When you don't want the target to even suspect they are being investigated.**

As much information will be gathered without touching the target or their affiliates, therefore only modules that do not touch the target will be enabled.

Run Scan

There are four types of scans that are available: the first type is All, the second type of scan is Footprint which correlates to information the target has added to the internet whether it be their identities, their network parameters, or other types of information, the third scan is called Investigate correlating to whether the user suspects the target to be malicious and needs more information, the last scan is called Passive which allows the user to gather information about their target without touching the target or their accomplices.

When beginning a new scan the first initial blank to fill in is the name of the scan, this can be named anything under the sun just as long as the box is marked, usually with an important marker of what is being targeted.

Scan Name

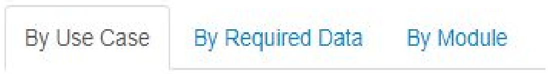
Descriptive name for this scan.

The next box below the name is the Seed Target. There’s a multitude of variations that can be searched, such as: a Domain name such as Google.com or other websites, an IP address, a Hostname/Subdomain which include websites such as www.microsoft.com, a Subnet, and finally an email address.

Seed Target

Starting point for the scan.

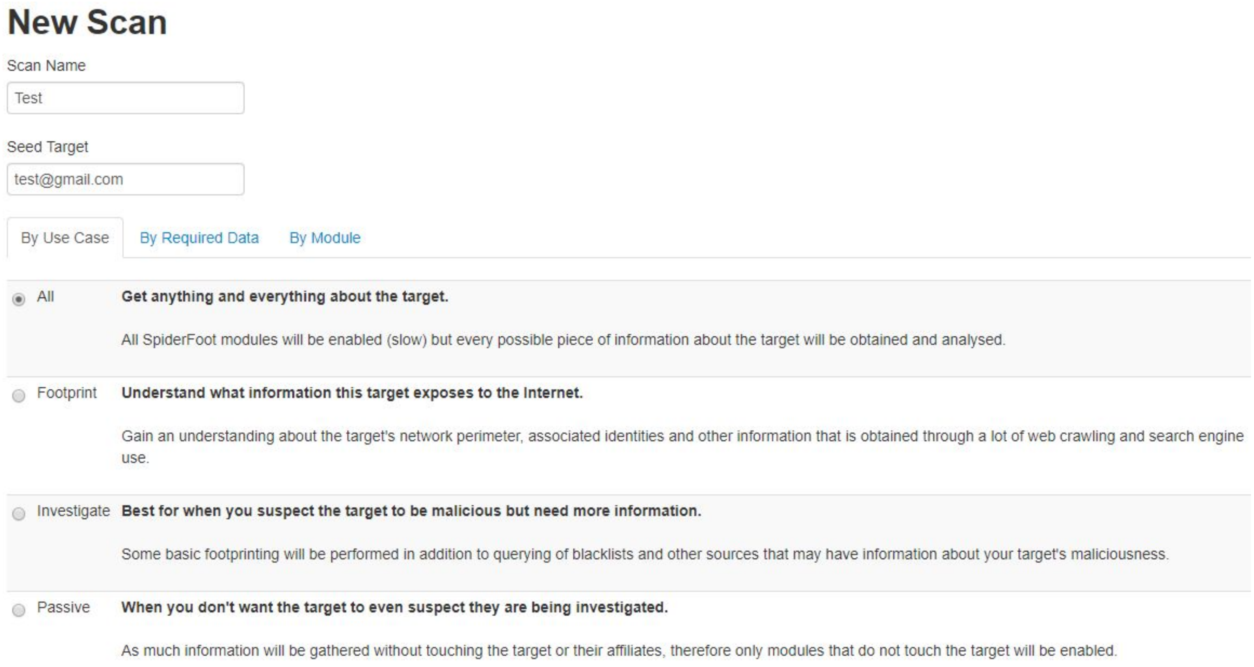
The final piece of the puzzle to complete is the type of scan. As previously mentioned before, there are four types of scans: Footprint, Investigate, Passive, and All. Before the scan there are two different tabs called “By Required Data” and “By



By Use Case | By Required Data | By Module

Module" which allows the user to select or deselect certain websites or hosts on which the scan will run. There are a large number of modules and required data types that Spiderfoot searches through. All the data that is searched between each tab can be seen using the provided links: By Required Data and By Module.

Using all this information and data, we're going to test how Spiderfoot functions and what the scans reveal about the target. For this experiment, the target will be "test@gmail.com" and will be scanned using "All" since the goal is to gather as much intel about the target as possible. When all the information has been filled out, the scan is ready to start.



New Scan

Scan Name: Test

Seed Target: test@gmail.com

● All — **Get anything and everything about the target.**

All SpiderFoot modules will be enabled (slow) but every possible piece of information about the target will be obtained and analysed.

○ Footprint — **Understand what information this target exposes to the Internet.**

Gain an understanding about the target's network perimeter, associated identities and other information that is obtained through a lot of web crawling and search engine use.

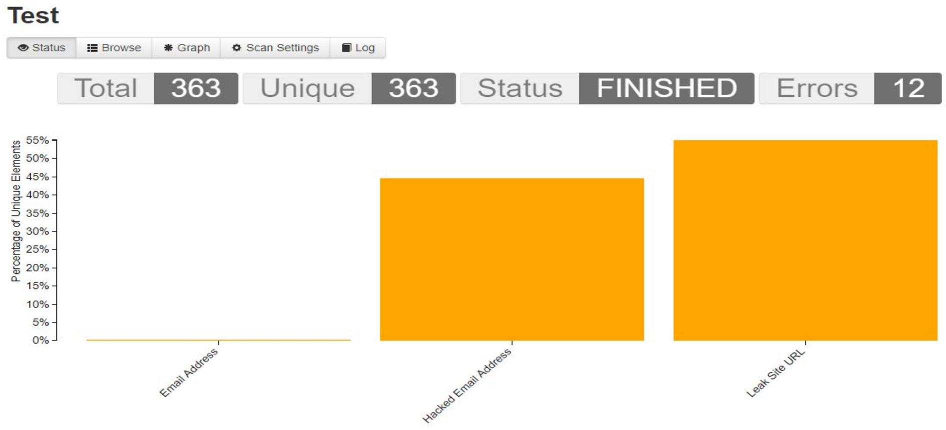
○ Investigate — **Best for when you suspect the target to be malicious but need more information.**

Some basic footprinting will be performed in addition to querying of blacklists and other sources that may have information about your target's maliciousness.

○ Passive — **When you don't want the target to even suspect they are being investigated.**

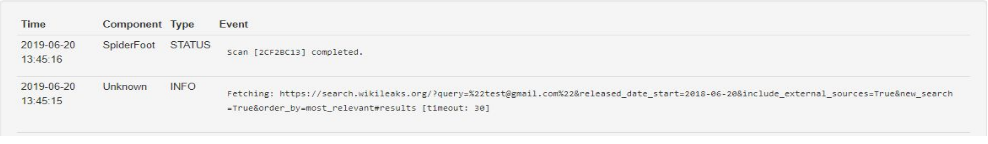
As much information will be gathered without touching the target or their affiliates, therefore only modules that do not touch the target will be enabled.

Spiderfoot will run through all the selected sources and output what has/has not been found. If there hasn't been any credentials set for certain areas Spiderfoot will show these as ERRORS. To know when the scan is completely finished the "Status" bar at the top will display "FINISHED" to the right of it.



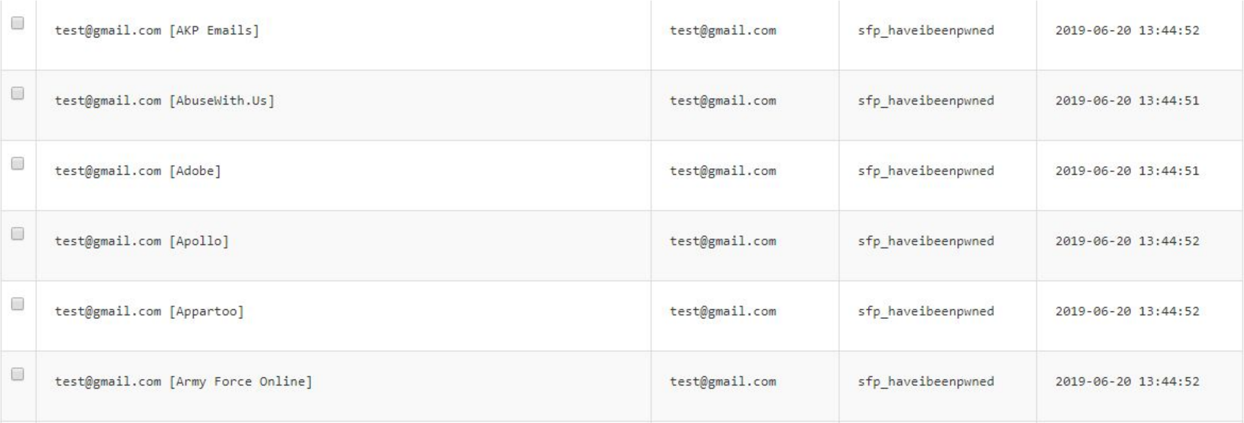
Test

Total 363 | Unique 363 | Status FINISHED | Errors 12



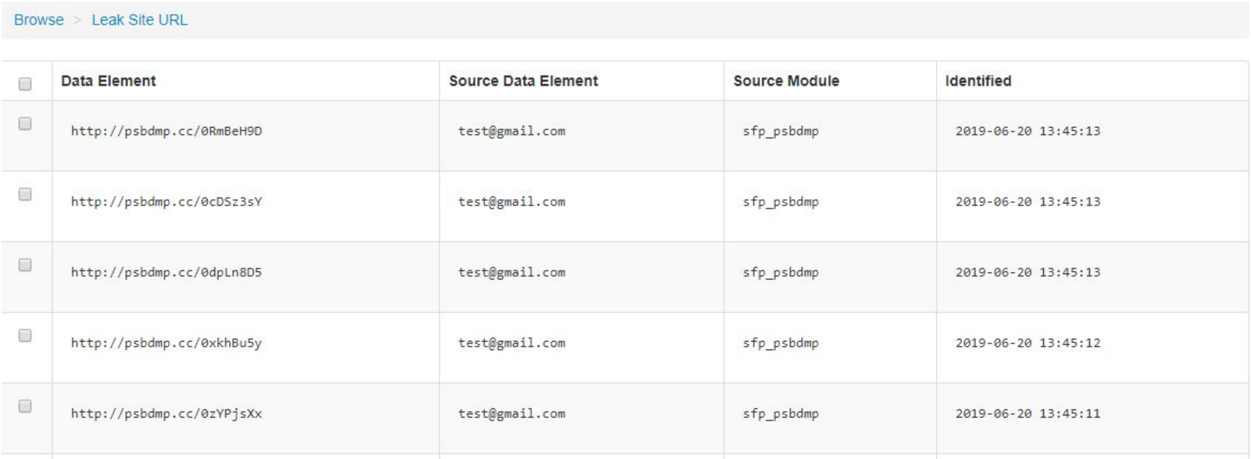
Time	Component	Type	Event
2019-06-20 13:45:16	SpiderFoot	STATUS	`Scan [2CF2BC13] completed.`
2019-06-20 13:45:15	Unknown	INFO	`Fetching: https://search.wikileaks.org/?query=%22test@gmail.com%22&released_date_start=2018-06-20&include_external_sources=True&new_search=True&order_by=most_relevant#results [timeout: 30]`

The orange bars represent what has been found regarding the target, for this test case there is only one email address, since began the scan using an email address, there are 162 unique elements regarding Hacked Email Address and 200 unique elements corresponding to Leak Site URL. The orange bars are able to be clicked to show what data has been uncovered about the target. When selecting Hacked Email Address, this shows the websites which have an account created under that email address. To further this, when selecting Leak Site URL it displays links regarding the target on different websites, for this test many pastebin links appeared about the target.



□	test@gmail.com [AKP Emails]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:52
□	test@gmail.com [AbuseWith.Us]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:51
□	test@gmail.com [Adobe]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:51
□	test@gmail.com [Apollo]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:52
□	test@gmail.com [Appartoo]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:52
□	test@gmail.com [Army Force Online]	test@gmail.com	sfp_haveibeenpwned	2019-06-20 13:44:52

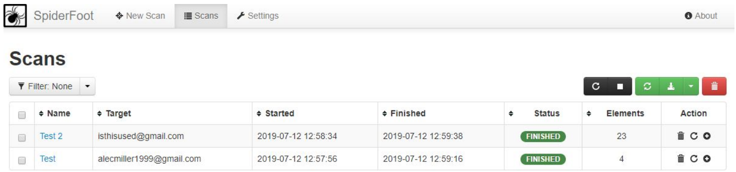
Hacked Email Address shown above.



□	Data Element	Source Data Element	Source Module	Identified
□	http://psbdmp.cc/0RmBeH9D	test@gmail.com	sfp_psbdmp	2019-06-20 13:45:13
□	http://psbdmp.cc/0cDSz3sY	test@gmail.com	sfp_psbdmp	2019-06-20 13:45:13
□	http://psbdmp.cc/0dpLn8D5	test@gmail.com	sfp_psbdmp	2019-06-20 13:45:13
□	http://psbdmp.cc/0xkhBu5y	test@gmail.com	sfp_psbdmp	2019-06-20 13:45:12
□	http://psbdmp.cc/0zYPjsXx	test@gmail.com	sfp_psbdmp	2019-06-20 13:45:11

Leak Site URL shown above.

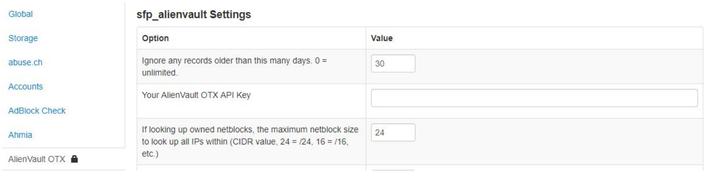
Every scan that is completed is saved under the “Scans” tab at the top of the website. Each scan can be selected and viewed, displaying all the information gathered on the target during the scan. Scans can be deleted by selecting on them and using the



□	Name	Target	Started	Finished	Status	Elements	Action
□	Test 2	isthisused@gmail.com	2019-07-12 12:58:34	2019-07-12 12:59:38	FINISHED	23	
□	Test	alecmiller1999@gmail.com	2019-07-12 12:57:56	2019-07-12 12:59:16	FINISHED	4	

red trash bin above the action column. If information has been gathered that the user wants to save, there's an export option to the left of the trash bin to save any scans for later.

Spiderfoot as a whole is an amazing OSINT resource, but in the settings tab, there are more in depth options which could allow the user to have deeper searches on their target. Each selection with a lock next to the name has an option to add an API Key. These keys will allow the program to access more sources to gather intel from! These API keys that are gathered from various websites are very useful in regards to OSINT. Every key that is acquired should be kept safe and known in case of future programs requiring certain keys.



sfp_alienvault Settings

Option	Value
Ignore any records older than this many days. 0 = unlimited.	30
Your AlienVault OTX API Key	
If looking up owned netblocks, the maximum netblock size to look up all IPs within (CIDR value, 24 = /24, 16 = /16, etc.)	24

Spiderfoot is a great program to begin using when starting out with OSINT. It's very easy to navigate through the program and gathering information is very straightforward. One of the great things about Spiderfoot is the ease of using the program while also having the ability to go deeper into the settings, add some API keys, maybe change some other settings and as a result have more in depth searches because of it. Spiderfoot is a very dynamic program, great for both beginners and experts of OSINT.