© SANS Institute 2019

487.4
Networks, Government, and Business

© SANS Institute 2019

Copyright © 2019 Micah Hoffman. All rights reserved to Micah Hoffman and/or SANS Institute.

PLEASE READ THE TERMS AND CONDITIONS OF THIS COURSEWARE LICENSE AGREEMENT ("CLA") CAREFULLY BEFORE USING ANY OF THE COURSEWARE ASSOCIATED WITH THE SANS COURSE. THIS IS A LEGAL AND ENFORCEABLE CONTRACT BETWEEN YOU (THE “USER”) AND SANS INSTITUTE FOR THE COURSEWARE. YOU AGREE THAT THIS AGREEMENT IS ENFORCEABLE LIKE ANY WRITTEN NEGOTIATED AGREEMENT SIGNED BY YOU.

With the CLA, SANS Institute hereby grants User a personal, non-exclusive license to use the Courseware subject to the terms of this agreement. Courseware includes all printed materials, including course books and lab workbooks, as well as any digital or other media, virtual machines, and/or data sets distributed by SANS Institute to User for use in the SANS class associated with the Courseware. User agrees that the CLA is the complete and exclusive statement of agreement between SANS Institute and you and that this CLA supersedes any oral or written proposal, agreement or other communication relating to the subject matter of this CLA.

BY ACCEPTING THIS COURSEWARE, YOU AGREE TO BE BOUND BY THE TERMS OF THIS CLA. BY ACCEPTING THIS SOFTWARE, YOU AGREE THAT ANY BREACH OF THE TERMS OF THIS CLA MAY CAUSE IRREPARABLE HARM AND SIGNIFICANT INJURY TO SANS INSTITUTE, AND THAT SANS INSTITUTE MAY ENFORCE THESE PROVISIONS BY INJUNCTION (WITHOUT THE NECESSITY OF POSTING BOND) SPECIFIC PERFORMANCE, OR OTHER EQUITABLE RELIEF.

If you do not agree, you may return the Courseware to SANS Institute for a full refund, if applicable.

User may not copy, reproduce, re-publish, distribute, display, modify or create derivative works based upon all or any portion of the Courseware, in any medium whether printed, electronic or otherwise, for any purpose, without the express prior written consent of SANS Institute. Additionally, User may not sell, rent, lease, trade, or otherwise transfer the Courseware in any way, shape, or form without the express written consent of SANS Institute.

If any provision of this CLA is declared unenforceable in any jurisdiction, then such provision shall be deemed to be severable from this CLA and shall not affect the remainder thereof. An amendment or addendum to this CLA may accompany this Courseware.

SANS acknowledges that any and all software and/or tools, graphics, images, tables, charts or graphs presented in this Courseware are the sole property of their respective trademark/registered/copyright owners, including:

AirDrop, AirPort, AirPort Time Capsule, Apple, Apple Remote Desktop, Apple TV, App Nap, Back to My Mac, Boot Camp, Cocoa, FaceTime, FileVault, Finder, FireWire, FireWire logo, iCal, iChat, iLife, iMac, iMessage, iPad, iPad Air, iPad Mini, iPhone, iPhoto, iPod, iPod classic, iPod shuffle, iPod nano, iPod touch, iTunes, iTunes logo, iWork, Keychain, Keynote, Mac, Mac Logo, MacBook, MacBook Air, MacBook Pro, Macintosh, Mac OS, Mac Pro, Numbers, OS X, Pages, Passbook, Retina, Safari, Siri, Spaces, Spotlight, There’s an app for that, Time Capsule, Time Machine, Touch ID, Xcode, Xserve, App Store, and iCloud are registered trademarks of Apple Inc.

PMP and PMBOK are registered marks of PMI.

SOF-ELK® is a registered trademark of Lewes Technology Consulting, LLC. Used with permission.

SIFT® is a registered trademark of Harbingers, LLC. Used with permission.

Governing Law: This Agreement shall be governed by the laws of the State of Maryland, USA.

© SANS Institute 2019

SEC487.4 | Open Source Intelligence (OSINT) Gathering and Analysis

SANS

Networks, Government, and Business

© 2019 Micah Hoffman | All Rights Reserved | Version E02_01

Welcome to SEC487 Open Source Intelligence (OSINT) Gathering and Analysis!

© SANS Institute 2019

TABLE OF CONTENTS	PAGE
Whois	3
IP Addresses	24
DNS	38
Findings Online Devices	61
Wireless Networks	78
Recon Tool Suites and Frameworks	98
Government Data	137
Researching Companies	168

487.4 Table of Contents

This table is a reference for you to quickly move to certain topics in this 487.4 book.

©2019 Micah Hoffman

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. **Whois**
2. IP Addresses
3. DNS
4. Findings Online Devices
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. Government Data
8. Researching Companies

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

The Whois Protocol

When domain names are recorded with ICANN,[1] data about who is registering it is submitted

This data can have: names, addresses, phone numbers, email addresses, and other servers

When registering, users can make their data private to hide it from the public

Data can be false, old and/or plainly incorrect

Five regional registrars compile whois records. Can you name all five?

The Whois Protocol

When a new domain is registered with a domain registrar, certain personal (or business) information is requested so that the registrar (GoDaddy, Enom, etc.) can tie the domain to an actual person or entity. People usually submit full names, physical addresses, phone numbers, email addresses, and DNS (Domain Name Service) servers to these registrars. The data is sent to Regional Internet Registries (RIRs) found at https://sec487.info/n9 and can be looked up using a variety of tools and web sites. It is public unless the person or company registering the domain allows the registrar to mask the public-facing data with their own. So, instead of theprincessbride.test domain reported as being owned by Prince Humperdinck, it would show the registrar's data (GoDaddy, Enom, etc.) with a unique address for this customer. This privacy feature is usually an extra step and perhaps an extra cost and not the default when new domains are purchased. Because of this, whois data can be very valuable for an OSINT researcher.

One thing to note is that, once a domain is registered to a person or organization, there is no requirement to update the data in the whois registry when the data should change. For instance, theprincessbride.test domain might be owned by The Prince Humperdinck Corporation and originally registered by Vizzini. Then, years later, Vizzini moves to another company and Inigo Montoya becomes the point of contact for the domain. The information should be updated in the registry but frequently is not, so if you perform a whois on that domain, you may get Vizzini's data and not the current point of contact.

Reference: [1] https://sec487.info/fe

©2019 Micah Hoffman

© SANS Institute 2019

Command-line Whois Tool Example

Command-line tool shows less data than it used to

Still see important domain dates (1) and name servers (2)

May see additional data depending on the domain and privacy

```
student@sec487 (10:06:25) :~$ whois sec487.info
Domain Name: SEC487.INFO
Registry Domain ID: D503300000016954742-LRMS
Registrar WHOIS Server: whois.godaddy.com
Registrar URL: http://www.godaddy.com
Updated Date: 2017-11-15T23:58:55Z
Creation Date: 2016-07-24T18:37:13Z
Registry Expiry Date: 2020-07-24T18:37:13Z
Registrar Registration Expiration Date:
Registrar: GoDaddy.com, LLC
Registrar IANA ID: 146
Registrar Abuse Contact Email: abuse@godaddy.com
Registrar Abuse Contact Phone: +1.4806242505
Reseller:
Domain Status: clientDeleteProhibited https://icann.org/epp#clientDeleteProhibited
Domain Status: clientRenewProhibited https://icann.org/epp#clientRenewProhibited
Domain Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited
Domain Status: clientUpdateProhibited https://icann.org/epp#clientUpdateProhibited
Registrant Organization: Domains By Proxy, LLC
Registrant State/Province: Arizona
Registrant Country: US
Name Server: ASA.NS.CLOUDFLARE.COM
Name Server: BEAU.NS.CLOUDFLARE.COM
DNSSEC: signedDelegation
URL of the ICANN Whois Inaccuracy Complaint Form is https://www.icann.org/wicf/
>>> Last update of WHOIS database: 2019-09-26T14:13:10Z <<<
```

1

2

SANS Open Source Intelligence (OSINT) Gathering and Analysis 5

Command-line Whois Tool Example

As mentioned previously, the command-line tool `whois` on some operating systems can still be used to retrieve domain registration data from the whois system. You will have varying degrees of success using this tool. The amount of data returned depends on the privacy settings the registering person set and those of the registrar.

Even with the privacy settings enabled, we can still retrieve some information such as the dates when the domain was created, when it was updated, and when its registration expires (arrow 1). Additionally, you may see the name servers for the domain listed (arrow 2).

© SANS Institute 2019

DomainTools Whois

Websites such as DomainTools.com have whois searches that combine additional data about the target domain into an informative web page

We have important dates, name server information, and registrar used

Registrant Country	us
Registrar	GoDaddy.com, LLC IANA ID: 146 URL: http://www.godaddy.com Whois Server: whois.godaddy.com abuse@godaddy.com (p) 14806242505
Registrar Status	clientDeleteProhibited, clientRenewProhibited, clientTra clientUpdateProhibited
Dates	1,160 days old Created on 2016-07-24 Expires on 2020-07-24 Updated on 2017-11-15
Name Servers	ASA.NS.CLOUDFLARE.COM (has 17,041,736 domains) BEAU.NS.CLOUDFLARE.COM (has 17,041,736 domains)
Tech Contact	–
IP Address	104.27.162.135 - 513 other sites hosted on this server
IP Location	- California - San Francisco - Cloudflare Inc.
ASN	AS13335 CLOUDFLARENET - Cloudflare, Inc., US (re
Domain Status	Registered And Active Website
IP History	5 changes on 5 unique IP addresses over 3 years
Hosting History	1 change on 2 unique name servers over 3 years

SANS Open Source Intelligence (OSINT) Gathering and Analysis 6

DomainTools Whois

The DomainTools.com web site, along with several others that we will see in the course, allows users to search for a domain and retrieve whois and other domain-relevant data. The above image shows some of the data returned from a search of sec487.info in the site (https://sec487.info/i8). These web interfaces to command-line tools, like whois, can make it much easier to collect data, as the analyst only needs access to a web browser to gather this data.

The content returned may contain important information for your OSINT case. In the slide, we noted five interesting points:

1. The registrar that currently holds the domain registration.
2. Dates when the domain was created, modified, and when it expires.
3. What the DNS NS records are for the domain.
4. Is there a web site on the domain?
5. IP address history that would show if this domain name has been hosted on other networks.

Image from https://sec487.info/i8, September 27, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Whois and GDPR

macOS and Linux systems have a command-line "whois" client that used to retrieve detailed data like:

- Domain name
- Registrant name
- Registrant address
- Registrant phone
- Registrant email

When GDPR[1] began to be enforced in May 2018, most of the whois systems in the world began filtering their content

They retained detailed records for domains, but most would not show it to the public anymore

The command-line "whois" command is less useful now

Whois and GDPR

macOS and Linux systems have a command-line tool named "whois" that queries the whois registries to find detailed data about domains. Well, that is what it used to do before the European GDPR[1] rules went into enforcement in May 2018. We used to retrieve detailed data from whois registrars about who registered a domain, phone numbers, names, addresses, email addresses, as well as data about when the domain's registration was submitted and when it expires. Now, even though the registrars still have that information, most do not provide it to the public due to fears of divulging European Union resident data and violating the GDPR privacy rules.

We still can gain access to some of that data, as you will see shortly, but the command-line whois tool is less useful to us now since the results that it retrieves will most likely be redacted or masked. The OSINT Curious Project has a blog post about this topic at https://sec487.info/sz.

Reference [1] https://sec487.info/na.

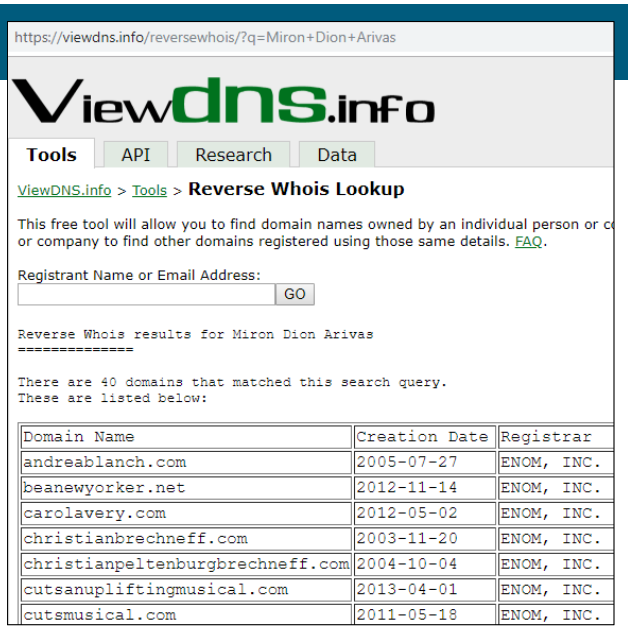
© SANS Institute 2019

Whois Pivot on Name

Even though we cannot see some data in the whois systems, we can search for people, phone numbers, and emails

A good example is using the name "Miron Dion Arivas"

We can perform a reverse whois query on that name and see what is returned



SANS Open Source Intelligence (OSINT) Gathering and Analysis 8

Whois Pivot on Name

With the whois data being filtered or redacted, we cannot see what personal information is being held in those databases. We can submit names, phone numbers, email addresses, and other information to whois sites and have them search for that data in the whois records they hold. Remember that, although they will not show us the personal data, they still have it. We submit some information, they search for corresponding domains, and then they show us what domains were associated with the data we submitted.

To see this in action, we can use the name "Miron Dion Arivas" and perform a reverse whois search on a site like viewdns.info (https://sec487.info/8r). Entering that name in the field and submitting it yields 40 domains where the name appears as a registrant. Look at the names of the domains. We not only can see who his previous customers were (Andrea Blanch, Carol Avery, Christian Brechneff) but we also see when the domains were registered (2005, 2012, and 2003, respectively). Scrolling down the results page, we also see that Mr. Arivas is somehow associated with the domain "mandypatinkin.org".

jmjcarpentry.com	2013-04-17
joemasteroff.com	2011-09-04
johnbdlaw.com	2007-11-21
kathryngrody.com	2004-10-28
larrymantello.com	2006-04-21
leszuke.com	2014-08-06
lindapellegrini.com	2018-03-28
loissasson.com	2014-12-05
lyndabenglis.com	2004-12-13
mandypatinkin.org	2003-03-26
nancyhwolf.com	2007-11-02
naomimargolis.com	2005-09-26
petragroupnewyork.com	2003-04-18
richtercapital.com	2012-03-27
robertmichaelburns.com	2010-09-21
solarisentertainment.com	1999-10-08
thebaselinstallation.com	2007-08-11
thegreekhousesifnos.com	2013-04-18
timalovejoy.com	2006-10-19
tracymartin.net	2003-07-21
watersources.us	2004-05-26

Images from https://sec487.info/8r, January 4, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Whois and Organizations

Sometimes your target is a certain company/group

Finding information about its users is helpful for social engineering and other attacks

What if we used the whois system to find information about people at an organization?

Whois and Organizations

Sometimes our investigations seek to uncover information about companies, and sometimes we want information about people. Now that we understand that the whois system may contain "people" data for certain companies, we can extend our use of whois. What if we could find people from a certain domain (and by extension a company or organization) that registered certain domains that were not for their company? Let's find out!

© SANS Institute 2019

Reverse Whois (1)

Each entry in the whois databases have at least one email address in the admin, technical, or registrant contact fields

If we search for buttercup@sec487.info in the whois system, we will find all records that have that email address tied to a domain

We can perform reverse lookups to find domains that have certain information in the whois records

We can record these domains for later analysis

Reverse Whois (1)

We already discussed that each whois entry has one to three points of contact (admin, technical, and registrant) and that many of the domains registered do not use the "private masking" techniques discussed earlier. When we perform a whois search, we request information about a domain. When we perform a reverse whois search, we request information that may be found in the responses that are returned.

If your OSINT subject uses the email address buttercup@sec487.info, you can perform a reverse whois, and information for any record where the contact email is buttercup@sec487.info will show in the results. We can record this data for later analysis.

©2019 Micah Hoffman

© SANS Institute 2019

Reverse Whois (2)

What if you wanted to see all the domains where people with an @sec487.info email address appear in the whois system?

Certain web applications allow us to search *@sec487.info and will return all domains where that data is recorded

Some people use their "work" email addresses to register "personal" domains

Using this process, we can find personal domains that employees/members of an organization are connected to

Why is this important?

Reverse Whois (2)

Reverse whois queries, using certain tools, can be asked to return all the domains where anyone with an email address ending in a specified domain is found. For instance, instead of searching for buttercup@sec487.info, we could query whois for all @sec487.info email addresses that appear in whois records. Using the "glob" symbol (*) we would make the query for *@sec487.info and then review the results.

Let us get into the minds of some of our targets. Many people use their work email addresses to register personal domains. So instead of using buttercup@personalemaildomain.test, Princess Buttercup may use her work email (buttercup@sec487.info) to register her personal domains. Using this technique, we can rapidly discover a variety of the interests of the staff of certain organizations. Want to know some of the interests of people with @dhs.gov (the United States Department of Homeland Security)? Perform a reverse whois lookup of *@dhs.gov and then perform a regular whois lookup on those results.

© SANS Institute 2019

ViewDNS.info

ViewDNS.info has an excellent web application for performing rWhois (reverse Whois) lookups

It also has an API that can be purchased for automated queries

Here is a search with *@dhs.gov as our target

ViewDNS.info

Tools | API | Research | Data

ViewDNS.info > Tools > **Reverse Whois Lookup**

This free tool will allow you to find domain names owned by an individual person or company. or company to find other domains registered using those same details. FAQ.

Registrant Name or Email Address:

GO

Reverse Whois results for *@dhs.gov
==========

There are 24 domains that matched this search query.
These are listed below:

Domain Name	Creation Date	Registrar
2beststyle.co	2013-07-10	GODADDY.COM, INC.
asistxgulfcoast.org	2010-03-08	DOMAIN.COM, LLC
bethlehemcog.org	2015-11-17	TUCOWS INC.
blainemwr.com	2011-02-02	ENOM, INC.
bostonjrhuskies.com	2016-01-14	GODADDY.COM, LLC
bostonjuniorhuskies.com	2016-01-27	GODADDY.COM, LLC
crossfitpistoleros.com	2014-10-16	GODADDY.COM, LLC
dejavuvintage.biz	2017-03-15	TUCOWS DOMAINS INC.
feaep.org	2007-12-28	TUCOWS INC.
fletc.us	2002-04-18	NEUSTAR RESERVE ACCOUNT A

SANS Open Source Intelligence (OSINT) Gathering and Analysis 12

ViewDNS.info

Let us use this reverse whois technique to discover some of the sites that people with dhs.gov email addresses have registered but that might not be DHS-owned domains. We can use the free viewdns.info web site's reverse whois lookup tool (https://sec487.info/4c) for this purpose. The key thing about ViewDNS's tool is that it allows the "*" character to act like "all characters," so it matches any email address at the domain we specify. The web site will show the first 500 results domains. If there are more results that could be returned, you will need to purchase an account with ViewDNS.info for a monthly fee (https://viewdns.info/api/).

When the reverse whois is performed using "*@dhs.gov", we get 24 domains, when they were registered, and the registrar where they were registered. All of this can be helpful, especially if the company or organization uses a single vendor for its domains. Then we can easily pick out the domains that are not using that registrar and further investigate them.

Image from https://sec487.info/ff, January 3, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

ViewDNS.info Results

We now have the domains where the *@dhs.gov users have registered domains

There are some obvious non-work-related domains in the results, and also some surprises that may provide pivot points

bostonjuniorhuskies.com
crossfitpistoleros.com
heavensentsisters.org
horseplayobx.com
keelermotorsports.com
robertparsons.biz
stephiehartz.com
thecranes.guru
zarmishaminister.org

ViewDNS.info Results

Now that we have the domains where DHS users used their work emails to register domains, we need to sort through and evaluate each to determine if it is an official one or personal. The table above shows an excerpt from the results of our query. In this list you can easily pick out several domains (crossfitpistoleros.com and keelermotorsports.com) that look like they are most likely not official DHS web domains. These personal domains provide our investigations "pivot points" to perform more research and gather even more data.

© SANS Institute 2019

Keelermotorsports.com

This is most likely not an official DHS web site

Probably a side business of an employee

But we still don't know who this is, right? (Stand by, we will)

Keelermotorsports.com

The http://keelermotorsports.com/ site does not appear to be an official Department of Homeland Security (DHS) web site but, instead, the business of one of its employees. In the coming module, we will use historical whois data to show the email address of the person who registered this site.

Image from http://keelermotorsports.com/, January 3, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Going Directly to the Source - ARIN

We can query the regional registrars directly

Some have APIs that let us interact with them via web pages

ARIN's[1] and RIPE's services are easy to use

[1]American Registry for Internet Numbers

IP Address Block:

`https://whois.arin.net/rest/ip/8.8.0.0.json`

First and/or Last Names:

`https://whois.arin.net/rest/pocs;last=doe;first=j*`

Name of Organization:

`https://whois.arin.net/rest/orgs;name=sans*`

Name of the Domain of the POCs:

`https://whois.arin.net/rest/pocs;domain=dhs.gov`

Going Directly to the Source - ARIN

Several regional registrars have APIs that we can use to retrieve our OSINT data. In this slide, we show how to use the ARIN (American Registry for Internet Numbers) API to retrieve information about IP address blocks, people, organizations, and domains. We can make the below requests to the web sites via a web browser or through other tools such as curl or Python scripts.

IP Address Block:

`https://whois.arin.net/rest/ip/8.8.0.0.json (https://sec487.info/h0)`

First and/or Last Names:

`https://whois.arin.net/rest/pocs;last=doe;first=j* (https://sec487.info/h1)`

Name of Organization:

`https://whois.arin.net/rest/orgs;name=sans* (https://sec487.info/h2)`

Name of the Domain of the POCs:

`https://whois.arin.net/rest/pocs;domain=dhs.gov (https://sec487.info/h3)`

© SANS Institute 2019

Let's Look at the DHS Example

https://whois.arin.net/rest/pocs;domain=dhs.gov

Lopez, Stevie	LOPEZ187-ARIN
Mascitelli, Dave	DMA496-ARIN
McCauley, Terry	TMC155-ARIN
MCCORD, HAL	HEM3-ARIN
McGouirk, Jason	MCGOU13-ARIN
Meno, Tom	MENOT-ARIN
Meno, Tom	MENOT1-ARIN

Details

Let's Look at the DHS Example

When we visit the https://whois.arin.net/rest/pocs;domain=dhs.gov (https://sec487.info/h3) URL, we get a long page of all the points of contact (POCs) for that domain. Each account has a unique link on the right that will show details about the entry.

We will look at the details of the HEM3-ARIN record of Hal McCord in the coming slides.

©2019 Micah Hoffman

© SANS Institute 2019

Detailed View of ARIN Data

We have:

- Names
- Addresses
- Phones
- Emails

Enough for Social Engineering?

Name	MCCORD , HAL E
Handle	HEM3-ARIN
Company	Dept. of Homeland Security,
Street	1110 North Glebe Road
City	Arlington
State/Province	VA
Postal Code	22201
Country	US
Registration Date	2006-04-17
Last Updated	2017-08-03
Comments	
Phone	+1-850-452-6390 (Office)
Email	hal.mccord@us-cert.gov hal.mccord@hq.dhs.gov

Name	HELPDESK OIG
Handle	HELPD182-ARIN
Company	DHS-OIG DATA
Street	1120 VERMONT AVE NW
City	WASHINGTON
State/Province	DC
Postal Code	20005
Country	US
Registration Date	2012-12-04
Last Updated	2012-12-04
Comments	
Phone	+1-202-254-4301 (Office)
Email	oig.helpdesk@dhs.gov

SANS Open Source Intelligence (OSINT) Gathering and Analysis 17

Detailed View of ARIN Data

Pulling up the records for Hal McCord (HEM3-ARIN) and the OIG Helpdesk (HELPD182-ARIN), we can now see the wealth of data we can get from this API. We have names, addresses, phone numbers, and emails. Each of these are pivot points for further investigation. In addition to those, we also have dates that can help us understand when the data was submitted to ARIN and how recently some of it was updated.

Images from:

https://sec487.info/h4, March 28, 2018

https://sec487.info/h5, March 28, 2018

© SANS Institute 2019

We Can Retrieve JSON via the Command Line

```
$  curl -H "Accept: application/json"
"http://whois.arin.net/rest/pocs;domain=dhs.gov" | json_pp
...{
   "@name" : "Groves, Bob",
   "@handle" : "GROVE83-ARIN",
   "$" : "https://whois.arin.net/rest/poc/GROVE83-ARIN"
 },{
   "$" : "https://whois.arin.net/rest/poc/OHE6-ARIN",
   "@handle" : "OHE6-ARIN",
   "@name" : "HELPDESK, OIG"},...
```


We Can Retrieve JSON via the Command Line

Before we examine the details of the records, you should know that ARIN allows users to retrieve data in the JSON format which is perfect for post-processing with a scripting language. In the above example, we used the curl command-line tool to make a request to the ARIN API and ask for JSON data back. When data was returned, we sent it through the json_pp tool to make it look pretty and human-readable.

We can see that the data returned is similar to what we saw on the previous slide via the web interface.

©2019 Micah Hoffman

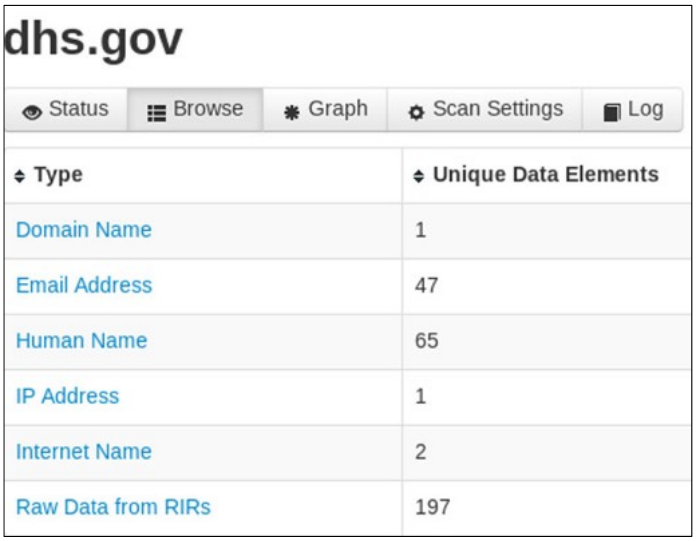
© SANS Institute 2019

SpiderFoot ARIN Module

Seed Domain: dhs.gov

Enable the following modules (at least):

- ARIN
- DNS resolver
- Email
- Name Extractor



Type	Unique Data Elements
Domain Name	1
Email Address	47
Human Name	65
IP Address	1
Internet Name	2
Raw Data from RIRs	197

SpiderFoot ARIN Module

Steve Micallef wrote a module for SpiderFoot to access the ARIN API and retrieve data. By enabling the ARIN, DNS resolver, email, and name extractor modules (at a minimum), we can retrieve all the data we received via manual methods, but this time via an automated tool. SpiderFoot categorizes the data it receives into the above groupings, which we can browse by clicking on them.

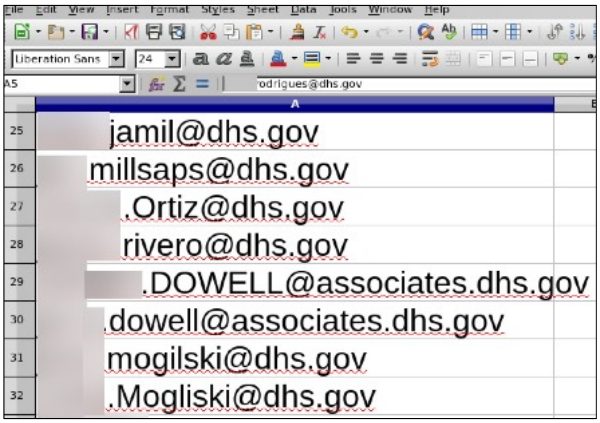
© SANS Institute 2019

SpiderFoot Emails (Browse and CSV Exported)

Browse > Email Address

Data Element	Source Data Element
iland@dhs.gov	{"poc":{"@xmlns":{"ns3":"http:\/\/www.arin.net\/rws\/rdns\/v1","$":"http:\/\/www.arin.net\/whois n.net\/public\/whoisinaccuracy\/index.xhtml","@t strationDate":{"$":"2008-04-01T08:09:07-04:00"}, N"},"city":{"$":"LAKEWOOD"},"companyName":{"$":" ":"USA"},"name":{"$":"United States"},"e164":{"$ N"},"lastName":{"$":"EILAND"},"emails":{"email" portUrl":"https:\/\/www.arin.net\/public\/whois .net\/whois_tou.html","limitExceeded":{"@limit" lidate the data for this POC, but has received r
ROGERS@DHS.GOV	{"poc":{"@xmlns":{"ns3":"http:\/\/www.arin.net\/rws\/rdns\/v1","$":"http:\/\/www.arin.net\/whois n.net\/public\/whoisinaccuracy\/index.xhtml","@t strationDate":{"$":"2013-01-29T16:06:42-05:00"},

Export to CSV



SpiderFoot Emails (Browse and CSV Exported)

Choosing the "Email Address" category in the web interface brings up the data on the left. We can see the "data element" is an email address and the "source data element" is a JSON block that the server sent to us. SpiderFoot allows users to export data into CSV (Comma Separated Values) files. On the right, we exported the email addresses and opened the CSV in Libre Office (although you could use a text editor or spreadsheet application). This list of email addresses is in a perfect format for other uses, such as running through the HaveIBeenPwned database or using in a phishing exercise.

©2019 Micah Hoffman

© SANS Institute 2019

Historical Whois and DNS Data

Sites like SecurityTrails show current and historical whois and DNS data about a target domain

Historical Data

A AAAA MX NS SOA TXT

IP Adresses	Organization	First Seen	Last Seen	Duration Seen
45.60.31.34 45.60.103.34	Incapsula Inc	2018-09-09 (1 year ago)	2019-09-25 (14 hours ago)	1 year
45.60.33.34 45.60.31.34	Incapsula Inc	2018-08-10 (1 year ago)	2018-09-09 (1 year ago)	1 month
45.60.33.34 45.60.31.34	Incapsula Inc	2017-09-18 (2 years ago)	2018-08-09 (1 year ago)	10 months
204.51.94.202	SANS INSTITUTE	2017-08-20 (2 years ago)	2017-09-18 (2 years ago)	29 days
45.60.34.77 45.60.32.77	Incapsula Inc	2017-08-14 (2 years ago)	2017-08-20 (2 years ago)	6 days
204.51.94.202	SANS INSTITUTE	2017-04-16 (2 years ago)	2017-08-14 (2 years ago)	4 months

SANS Open Source Intelligence (OSINT) Gathering and Analysis 21

Historical Whois and DNS Data

A site similar to DomainTools.com, shown earlier, is SecurityTrails.com. They aggregate a variety of DNS and whois data about domains and allow easy access to it via a web browser. One additional reason to use this site is the historical content that it keeps and shows when you search for a domain that it has been tracking.

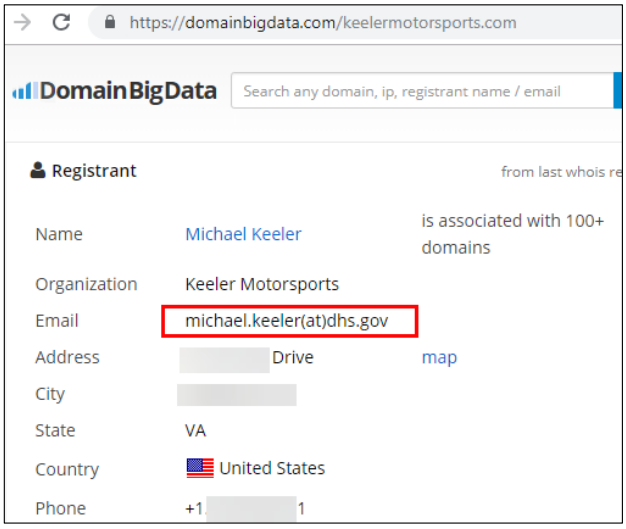
Image from https://sec487.info/i9, September 26, 2019.

© SANS Institute 2019

Remember keelermotorsports.com?

Some of the historical whois data is not filtered, redacted, or masked

Using domainbigdata.com, we can pull up the person that registered the domain we saw earlier



Remember keelermotorsports.com?

On the internet, we have many choices for places that hold data of interest. We need to be resilient and try many of them before admitting defeat. While the whois command-line tool and other whois web sites would not give us the old whois data from our target domain, domainbigdata.com does (https://sec487.info/nb).

Image from https://sec487.info/nb, January 3, 2019.

©2019 Micah Hoffman

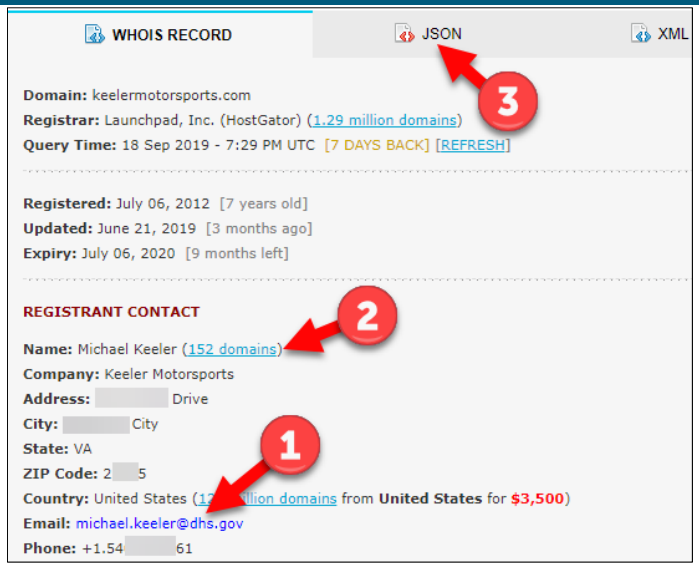
© SANS Institute 2019

Whoxy.com Whois Search

Whoxy.com is a free and paid whois domain search engine

Forward and reverse whois queries reveal live and historical information

Web and API interface



Whoxy.com Whois Search

The whoxy.com web site allows anonymous, registered, and paid users access to the information inside of the whois registrant system. They have both web and API interfaces to retrieve the data in human-readable, JSON (arrow 3), and XML formats.

Since May 2018, many sites began masking whois data due to GDPR privacy regulations. This prevented us from obtaining data that we would otherwise have been able to retrieve. The whoxy.com site does not apply additional masks to whois data for privacy and GDPR reasons. It not only shows us the data that is in whois now but also reveals historical whois content. Historical whois data can show true email addresses, phone numbers, and other piece of information tied to a domain.

We can see the whois registrant contact information for the keelermotorsports.com domain and see the person associated with this domain (arrow 1). This record shows that this registrant has over 150 other domains that they registered (arrow 2)!

Image from https://sec487.info/uc, September 25, 2019.

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. **IP Addresses**
3. DNS
4. Findings Online Devices
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. Government Data
8. Researching Companies

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

IPs and OSINT

IP addresses are important to collect for OSINT:

- Communications have a source and a destination IP
- They can (sometimes) be used to identify a device, network, company, or organization
- They have attributes and data associated with themselves
- It is useful to understand when they change
- Internet-facing IP addresses can be geolocated (sometimes)

IPs and OSINT

Since all networked devices on the internet need an IP address, collecting and analyzing this information may be important to your investigations. The originator of the network traffic is said to be the source, while the recipient is the destination. If my computer's web browser sends a request to Google, then my internet-facing IP address (for example, 96.232.133.18) would be the source and one of Google's IP addresses (such as 172.217.6.196) would be the destination. You can understand that collecting and analyzing the IP addresses that computers send traffic to, who owns them, and where they are might be of interest to an OSINT investigator.

There are other reasons why we may collect and use an IP address in our work. Among those are:

- When we find an IP address, it can be resolved to discover the IP address block owner, domains that may be registered to it, and what geographic location the IP address is registered to.
- Sometimes IP addresses have information about themselves that we can research on other sites.

© SANS Institute 2019

What Is an IP?

Internet Protocol (IP) addresses are how our devices communicate on the internet

Two versions of Internet Protocol:

- Version 4 (IPv4) - More common
  - Uses 4 "octets" in dotted decimal notation such as `87.111.235.2`
- Version 6 (IPv6) - Less common (but growing!)
  - Uses 8 groupings of hexadecimal such as `2001:98ae:1121::0988:2714`

What Is an IP?

When we wish to visit a physical location in the world, we usually will obtain its address. This address describes from the general (country) to the specific (street number) where the place we want to visit is. IP (Internet Protocol) addresses are similar for computers. IP addresses on the internet are similarly organized into a network portion and a specific system unique address. We won't go into private address spaces specified in RFC 1918 (https://sec487.info/48). Each device on the internet has at least one IP address so that it can communicate with other devices. IP version 4 (IPv4) is the most common protocol in use on the internet, and those addresses are organized into four octets and displayed in a dotted decimal notation, such as `87.111.235.2` or `10.20.30.40`. There are many tutorials and documents describing IPv4 addresses and networking. For additional information, https://sec487.info/46 has more details.

As the number of devices requiring IP addresses grew (computers, web sites, mobile devices, Internet of Things devices, etc.), it became apparent that the world was going to use up all of the internet-facing IPv4 addresses. The Internet Assigned Numbers Authority (IANA) created, socialized, and released the next version of the Internet Protocol, which you would expect to be IPv5, right? It is not (https://sec487.info/49). IPv6 (Internet Protocol version 6) was what the new standard was named. Many of the newer computers and mobile devices support IPv6 but may not be using it.

IPv6 addresses can be quite long and are up to eight groupings of four hexadecimal characters, such as `2001:98ae:1121:0000:0000:0000:0988:2714`.

IPv6 addresses can have four zeros in some of the addresses. When this happens, it is acceptable to collapse that section of the address and remove the zeros. So, an IPv6 address of `2001:98ae:1121::0988:2714` is equivalent to `2001:98ae:1121:0000:0000:0000:0988:2714`. For more details, visit https://sec487.info/47.

With the adoption of IPv6, more resources are beginning to focus on these newer addresses. Many of the tools we will discuss for IPv4-based research also have IPv6 components or also retrieve IPv6 data. With the majority of the internet still on the IPv4 version of the protocol, we will focus our efforts in that area.

Image from https://sec487.info/oh, January 5, 2019.

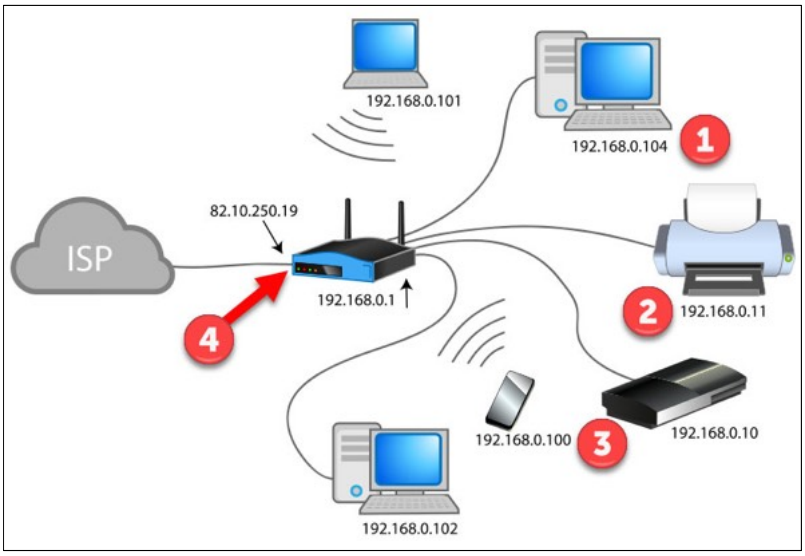
©2019 Micah Hoffman

© SANS Institute 2019

Simple Network Diagram

Each device on a network has at least 1 internal (not internet-facing) IP

The edge device, a router, gives its external IP address to all device traffic flowing through it



Simple Network Diagram

All the things that our devices do on the internet are tagged with an IP address. Let us examine where that IP address comes from. In a simple network, shown above, you have clients like printers, mobile devices, and laptops (1, 2, 3). Each of these devices gets an IP address that works on the local network inside the building, home, or office where those devices are located. Usually it is a router that gives the devices their internal IP addresses. We say "internal" because the router usually acts as a firewall and prevents internet traffic from directly accessing the devices inside the building, home, or office. Internal IP addresses are rarely internet-routable or able to be used directly on the internet.

When a device inside the network needs something from the internet, it sends its data to a gateway, which most of the time is a router. These routers connect two or more networks. In their most simple state, they can send data on the internal network and have a connection to an Internet Service Provider (ISP) to send data to the internet. It is their ISP-facing network connection that has the IP address that devices inside the network will use for their internet activities. In the example above, the router's external/internet-facing IP address is 82.10.250.19.

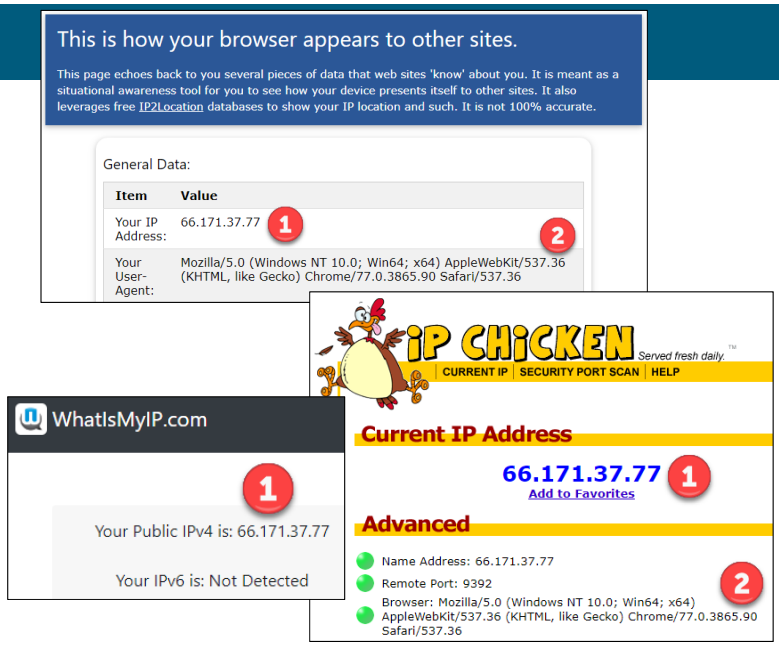
Image from https://sec487.info/yn, October 5, 2019.

© SANS Institute 2019

Viewing Your Internet IP

When your device communicates on the internet, your internet-facing IP address is sent to the remote system

You can use free web sites to discover what your IP address is



SANS Open Source Intelligence (OSINT) Gathering and Analysis 28

Viewing Your Internet IP

When our phones, computers, and other devices communicate on the internet, they give the remote system their IP addresses so that the remote system can send them responses. Sometimes this external IP is from a VPN or proxy, and sometimes it may be from the device you are using. We can use web sites to show us what IP address is being reported to web sites when we browser them. Three such sites are shown in the above slide: IPChicken.com, WhatIsMyIP.com, and whatu.info.

Some of these sites go beyond just showing your IP address (marked with 1's above) and will echo your device's current user agent string (marked with 2's above) or other data, such as the location of your IP address.

Images from https://www.whatismyip.com/, https://www.ipchicken.com/, and http://whatu.info, September 27, 2019.

©2019 Micah Hoffman

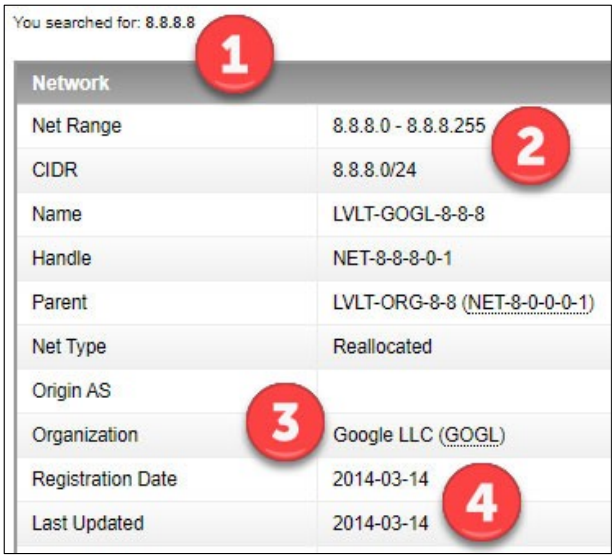
© SANS Institute 2019

IP Address Blocks

Companies buy blocks of IP addresses (IP blocks) for their internet-facing devices to use

When they buy them, the network ranges are noted along with the owner

We can look those up



IP Address Blocks

When companies have devices that they want to place onto the internet, they can rent, lease, or buy blocks of IP addresses from other companies. These registries are, for the most part, public, and we can search them from a variety of web sites. We searched the ARIN web site[1] for the IP address of 8.8.8.8 (1) and it showed the network (2), the owner (3), and when it was registered/updated (4).

Image from https://sec487.info/yp, October 5, 2019.

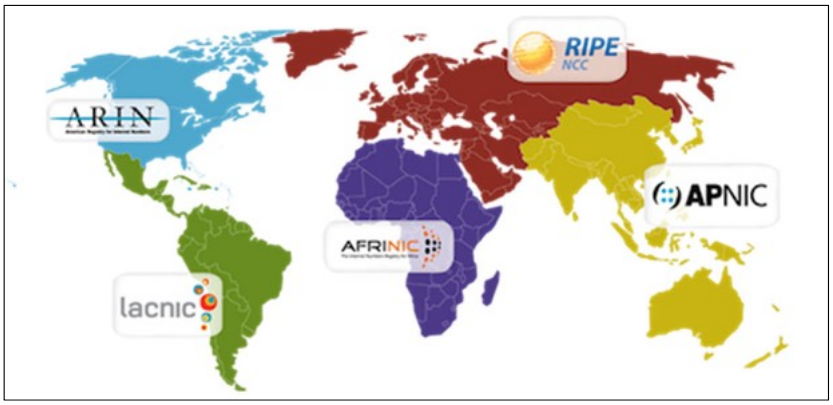
Reference: [1] https://sec487.info/yp

© SANS Institute 2019

RIRs

Regional Internet Registries (RIRs) were created to perform these registrations regionally

- LACNIC
- ARIN
- RIPE NCC
- APNIC
- AFRINIC



RIRs

Regional Internet Registrars (RIRs) were created in the 1990s to service the IP address and Autonomous System numbers (ASNs) for organizations in their regions of the world. The map above shows where the five international RIRs have their domains. We can look up IP address ownership in each of these registrars by visiting their sites.

- LACNIC - https://lacnic.net
- ARIN - https://arin.net
- RIPE NCC - https://ripe.net
- APNIC - https://apnic.net
- AFRINIC - https://afrinic.net/

Image from https://sec487.info/yq, October 5, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Searching Registrars for IPs

5.5.5.5 9.9.9.9

Enter an IP into any of the searches the RIRs have on their sites and you get who owns the address block associated with your target IP

```
inetnum:        5.4.0.0 - 5.7.255.255
netname:        DE-MEDIAWAYS-20120425
country:        DE
org:            ORG-TDG4-RIPE
admin-c:        MWH6-RIPE
tech-c:         MWH6-RIPE
status:         ALLOCATED PA
mnt-by:         RIPE-NCC-HM-MNT
mnt-by:         MDA-Z
mnt-lower:      MDA-Z
mnt-routes:     MDA-Z
created:        2012-04-25T06:13:17Z
last-modified:  2018-07-30T09:52:34Z
source:         RIPE

route:          5.4.0.0/14
descr:          Telefonica Germany GmbH & Co. OHG
remarks:        netname: DE-MEDIAWAYS
origin:         AS6805
mnt-by:         MDA-Z
created:        2018-08-08T09:03:25Z
last-modified:  2018-08-08T09:13:47Z
source:         RIPE
```

```
NetRange:       9.9.9.0 - 9.9.9.255
CIDR:           9.9.9.0/24
NetName:        CLEAN-97
NetHandle:      NET-9-9-9-0-1
Parent:         NET9 (NET-9-0-0-0-0)
NetType:        Direct Assignment
OriginAS:
Organization:   Quad9 (CLEAN-97)
RegDate:        2017-09-13
Updated:        2017-09-13
Ref:            https://rdap.arin.net/registry/ip

OrgName:        Quad9
OrgId:          CLEAN-97
Address:        1442 A Walnut Street, Suite 501
City:           Berkeley
StateProv:      CA
PostalCode:     94709
Country:        US
RegDate:        2017-09-07
Updated:        2018-12-04
```

SANS | Open Source Intelligence (OSINT) Gathering and Analysis 31

Searching Registrars for IPs

We can perform searches of the data in these RIRs by visiting one or more of their sites and looking for the search or whois pages. In the above slide, we searched for the IP address 5.5.5.5 in RIPE NCC (Europe). The RIR understands computer networking and shows the larger network that the 5.5.5.5 IP is a part of (1). Farther down that image (2), the site shows who owns the IP address block containing that target IP: Telefonica Germany GmbH & Co. OHG.

In the right image, we searched APNIC for the IP address 9.9.9.9 (3) to show that, although APNIC serves the Asia and Pacific region of the world, it still shows a United States address for Quad9 as owning the network containing our target IP.

Images from https://sec487.info/ys and https://sec487.info/yt, October 5, 2019.

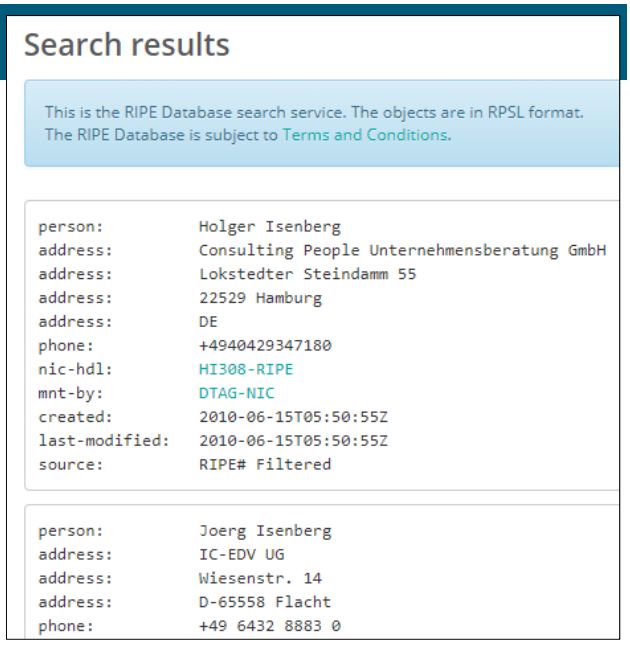
© SANS Institute 2019

Searching Registrars for Other Data

We can query RIRs for any information contained in IP block and domain registration

This includes names, addresses, phone numbers, company names, and emails

Here, we searched RIPE NCC for the last name "Isenberg"



Searching Registrars for Other Data

There is other information in the RIRs that we can retrieve. Any data submitted to the RIRs that is not protected by GDPR (and some that is!) can be found. Data such as first and last names, phone numbers, addresses, company names, and email addresses might be returned in search results. Above, we searched Europe's RIPE NCC RIR for the last name of "Isenberg" and found two records with details that may be interesting to our OSINT work.

Image from https://sec487.info/yu, October 5, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

IP Address Geolocation

Many companies have tied IP address blocks (subnets) to cities or addresses that they serve

It is an imperfect process…more of an educated guess in most cases

Since the internet is global, it helps to understand what regions traffic is from

For international sites, Content Delivery Networks (CDNs) can serve data from a server closer to your geographic location

This reduces global traffic and delivers faster results

Also allows computers to filter traffic to/from certain countries and regions

IP Address Geolocation

As humans, we live in a world with physical locations. Your office, your home, and your local restaurants most likely have street addresses so that you know how to get there and can share that information with others. While our cyber world is truly international, there is a need to try to map IP addresses to certain geographic regions. Recognize that this is an imperfect process in many cases and very accurate in others. Content Delivery Networks (CDNs) allow web-based content to be broadcast efficiently and rapidly to audiences around the world. If you live in Paris, France and wish to visit Google's web site, the CDN handling Google's web application traffic will look at the IP address of your computer or mobile device, note that you are in Paris, France, and have a server geographically close to your location respond to your searches. This is more efficient than your web request traversing the Atlantic Ocean, travelling across the United States to California, and being served by the Google servers in the data center near its headquarters.[1]

In the second case of very specific and accurate IP addresses is the situation of the home Internet Service Provider (ISP). Many ISPs know that the IP address given to your internet-facing router is for you, the customer who lives at your home address on file. ISPs may have specific network blocks that are for certain areas in which they have customers.

Another use for IP addresses tied to geographic locations is on the cyber defense side, where a company may wish to block traffic from certain parts of the world. [2] If your customer is concerned that a certain country may break into their web site, they can configure network devices to block network traffic that is most likely coming from that region.

Resources:

[1] CDNs - https://sec487.info/4g

[2] IP blocking - https://sec487.info/4h

© SANS Institute 2019

How to Find Geolocation from IP Address

Easiest method is to use a web application

There are MANY to choose from

The following have reliable results:

- whatismyipaddress.com
- www.infosniper.net
- www.ip2location.com

Most sites use commercial or open source databases on the backend to perform the queries

While this data may be roughly accurate, there is other data these sites divulge that could be relevant to your OSINT work

How to Find Geolocation from IP Address

To get the rough geolocation of an IP address, we usually visit one or more web applications. Each web application may have a different backend database that it is pulling the geolocation information from, so getting results from a variety of web sites can help narrow the actual location of the target. Finding a geolocation to IP web application is as simple as performing a query in your favorite search engine. The three sites noted above seem to provide valuable and fairly accurate information about location and other data about the IP address (such as carrier or ISP of the device).

©2019 Micah Hoffman

© SANS Institute 2019

IP2Location.com Results

IP Address	8.8.8.8
Country	United States [US]
Region	California
City	Mountain View
Coordinates of City	37.405992, -122.078515 (37°24'22"N 122°4'43"W)
ISP	Google LLC
Local Time	27 Sep, 2019 06:39 AM (UTC -07:00)
Domain	google.com
Net Speed	(COMP) Company/T1

IP Address	192.230.45.131
Country	South Africa [ZA]
Region	Gauteng
City	Johannesburg
Coordinates of City	-26.202270, 28.043630 (26°12'8"S 28°2'37"E)
ISP	Micfo LLC.
Local Time	27 Sep, 2019 03:48 PM (UTC +02:00)

Usage Type	(DCH) Data Center/Web Hosting/Transit
Anonymous Proxy	Yes
Proxy Type	(VPN) VPN Server

SANS Open Source Intelligence (OSINT) Gathering and Analysis 35

IP2Location Results

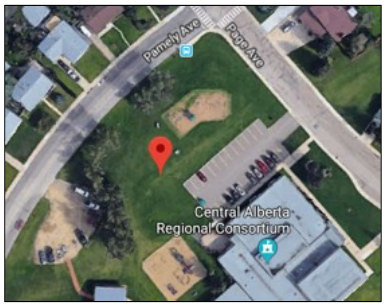
Using the https://sec487.info/4i site to look up Google's public DNS server at IP address 8.8.8.8, we can see the latitude and longitude (arrow 2) of the city where Google is located: Mountain View, California (arrow 1). Note that this is the latitude and longitude of the city of Mountain View, California and not the precise location of the server hosting the 8.8.8.8 IP address. This is a very common misconception. Additionally, this site reports other data, such as the local time for the city of registration (arrow 3) and what the network speed for the system might be (arrow 4).

If you choose to look up your own IP address, you might be surprised to discover that this site can tell if your IP address is from the Tor Darknet or using a VPN. Using a different IP address (right image above), 192.230.45.131, we can see (at arrow 6) that its ISP is Micfo LLC. (which gives us a business relationship to examine). If we scrolled down the ip2location.com web page for the South African IP address, we would see what is shown in the bottom right image (7). This IP address is tagged as coming from a data center and is flagged as an anonymous proxy and a VPN server. It is all of these things, as the ProtonVPN system uses this IP address for its VPN services.

Images from https://sec487.info/4i , September 27, 2019.

© SANS Institute 2019

Infosniper.net Results



Is this pin where the IP address is located?

Infosniper.net Results

The infosniper.net web site presents data a little differently and appears to draw from a different backend database. Choosing a different IP address at random, 24.66.15.87 (arrow 1), we see that it is owned by Shaw Communications Inc. (arrow 2), a cable internet provider (arrow 3) in the Red Deer Alberta, Canada (arrow 4) area (https://sec487.info/uk). Using that latitude and longitude (arrow 5) on the page to look up this location in the Google Maps application (https://sec487.info/uj) we can pull up the exact address that this IP address is tagged to (image on right above).

Is this accurate though? If we needed to serve a warrant or seize the computer at this IP address, would it be at this location? The answer is almost certainly, no.

©2019 Micah Hoffman

© SANS Institute 2019

IPs Aren't Always Where You Think

In some cases, such as with your home internet provider or a certain office building, the IP address is "close" to your location

In most cases it is a "guess"

MaxMind's "default" location is in the center of the United States, in Kansas, and was a farm

"Imagine having a constant stream of inexplicable and upsetting visitors to your house: cops looking for child pornographers; U.S. marshals looking for fugitives; responders rushing to the aid of suicidal people; sheriffs tracking down stolen vehicles. Incidents like this happened repeatedly"[1]

IPs Aren't Always Where You Think

While the IP address that you have at work or at home may be close to your actual location, it is mostly an approximation. Many people do not realize that fact. For many years, when an IP address was not mapped to a specific location, mapping companies such as MaxMind mapped those IPs to a default location. They picked a location in the center of the United States: a farm, in fact.

It turned out that many people did not understand that translating an IP address to its physical location on Earth was more art than science when they visited the above farm. Here is what happened. "The farm is a few hours' drive from the point that's the exact geographic center of the United States. Early last decade, a company called MaxMind was laying the foundation for what would become a location database for IP addresses. When their database doesn't know where a user's IP address is, it returns a latitude and longitude near the center of the country. They rounded the actual address, 39.8333333,-98.585522, to something shorter. Unfortunately, that nice and even address, 38.0000,-97.0000, is precisely in the farm's front lawn."[2]

"The following events appeared to originate at the residence and brought trespassers and/or law enforcement to the plaintiffs' home at all hours of the night and day: stolen cars, fraud related to tax returns and bitcoin, stolen credit cards, suicide calls, private investigators, stolen social media accounts, fund raising events, and numerous other events."[3]

Since the lawsuit, the MaxMind company has changed the default location of geolocated IPs to be a lake in Kansas and not the farm.[4]

[1] https://sec487.info/4k
[2] https://sec487.info/4j
[3] https://sec487.info/4l
[4] https://sec487.info/4m

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. **DNS**
4. Findings Online Devices
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. Government Data
8. Researching Companies

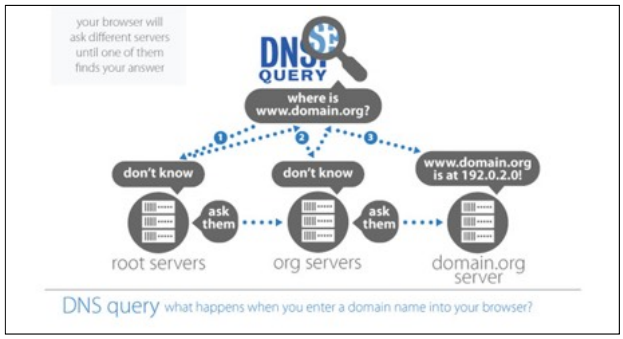
This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Domain Name Service (DNS)

Humans remember domains like sec487.info well but sometimes have challenges remembering IP addresses like 104.27.162.135 (IPv4) and 2400:cb00:2048:1::681b:a387 (IPv6)



DNS servers translate IP addresses to domain names and back again

Domain Name Service (DNS)

Humans remember words and web domains very well. It isn't hard to recall google.com or slashdot.org. Whereas people work in the semantic, word-filled world, computers process numbers more easily. For them, Internet Protocol (IP) addresses are important. The Domain Name Service (DNS) system was developed to simplify the complex computer networking world. DNS servers translate domain names to IP addresses and IP addresses into their corresponding domain names. A good example of this is the two domains shown above: google.com and slashdot.org. Their IP addresses are 172.217.7.142 and 216.34.181.45, respectively. You can understand that having to memorize IP addresses for every web site and email address domain would be incredibly challenging. DNS does that work for us.

While the intricacies of how DNS servers do their work is beyond the scope of this course, there are excellent resources on the internet that can explain the process (https://sec487.info/4d).

Image from https://sec487.info/4d, May 30, 2017.

© SANS Institute 2019

DNS Queries

We use DNS information to find

- The IP addresses of the servers of our targets
- Other domains that may be of interest in our work

It is common for multiple domain names to be tied to a single IP address, and with DNS queries we can find those other domains

What we will do is query DNS servers for certain information about domains and IP addresses

When you receive results from a DNS server (an IP address), perform another query against it to find additional records

DNS Queries

A DNS query occurs when you request information about an IP or domain name from a DNS server. Our computers do use the internet. We do this as analysts to retrieve interesting data for our OSINT work. What data is returned can vary depending upon how the DNS server is configured and what data has been stored in it. For our OSINT purposes, we collect all the information returned and, you guessed it, store and analyze it.

In the results of DNS server queries, we can find a variety of DNS record types that will be explored in upcoming slides. We will use a software client that may be an application on our computer or a web site to submit requests to DNS servers, asking for certain types of data. Sometimes we will ask to resolve a domain into an IP address and, other times, we will ask the DNS server for data about IP addresses and it will give us hostnames

©2019 Micah Hoffman

© SANS Institute 2019

DNS A, AAAA, and CNAME Records

DNS Record	Purpose	Example
A	IPv4 address to hostname	`hunch.ly -> 192.124.249.20`
AAAA	IPv6 address to hostname	`bing.com -> 2620:1ec:c11::200`
CNAME	Hostname to another hostname	`slack.sec487.info -> sec487alum.slack.com`

DNS A, AAAA, and CNAME Records

We begin our examination of DNS records by looking at some of the main records you might encounter in your DNS requests. In the table above, we show the different records for A, AAAA, and CNAMEs within DNS. But why are these important for OSINT?

A and AAAA records allow us to pivot from a hostname to an IP address (either IPv4 or IPv6, respectively). With this information, we can look up who owns the address block, check for the reputation of the IP, and other pivots to IP address searches.

A CNAME (canonical name) record ties one hostname to another. It is frequently used like a shortcut might be on your computer. Instead of copying a document from your Documents directory, you might make a shortcut or alias to it to refer to the original document. It is similar with CNAMEs. We look up the entry, such as slack.sec487.info, and the DNS server responds, telling our client or system to look up another hostname like sec487alum.slack.com. This allows us to link different sites. Also, it can reveal virtual server names; for example, if accounting.sec487.info and sales.sec487.info both resolve to the same IP address, they may both be served off the same web server or computer.

For a more complete listing of all the DNS record types, visit https://sec487.info/4e.

© SANS Institute 2019

DNS MX, NS, and TXT Records

DNS Record	Purpose	Example
NS	Name servers of the domain	`gab.com -> sue.ns.cloudflare.com`
MX	Email server(s) of the domain	`gab.com -> mail.protonmail.ch`
TXT	Any text, especially for validation purposes	`protonmail-verification=96ca5019e36fc3a6dc5c4399efabe68792e55cc7`

DNS MX, NS, and TXT Records

Let's move on to some other common DNS records you will encounter and what they can tell us about a domain. In this slide, we use the domain from the social media platform gab.com for our examples.

NS (name server) records are used when computers don't know how to resolve a hostname or IP address. Each domain has at least a primary NS record, and most have one or more secondary ones in case the primary name server is unavailable. With free DNS hosting providers offloading the burden of running DNS servers, it is more rare these days to find people or companies hosting their own DNS and, as such, the NS records pointing to IP addresses that they own. As shown above, the gab.com domain's NS record points to using Cloudflare's systems for their name resolution. What does this mean? If you wanted to shut down this domain, seize data about it and its owners, or understand its security posture, you now can look at the Cloudflare company.

MX (mail exchange) records are what systems use to determine where to send emails to for a given domain. When you send an email to a gab.com email address, your email system will look up the MX record for gab.com, which was mail.protonmail.com in September 2019. Now your email system knows where to send your email to get to its recipient. For OSINT purposes, the MX record can show business relationships and other systems that our target domains use for their business. In this case, gab.com uses Protonmail.com for its email.

TXT (text) records can contain any text data in them but are most commonly found containing verification and validation strings for internet services. The above example shows that the person who signed up to host gab.com email at Protonmail "proved" they owned the gab.com domain by inserting a TXT record in their DNS. Protonmail generated a unique string that the owner of gab.com inserted into a TXT record in the gab.com DNS. Then Protonmail can easily verify if this was done by performing a DNS request for a TXT record and viewing the string that was returned. TXT records need to be examined, as they can contain those business relationships that, if explored, may further our OSINT investigations.

©2019 Micah Hoffman

© SANS Institute 2019

Performing DNS Queries

From the command line:

- Windows
  - nslookup
  - Resolve-DnsName (PowerShell)
- Other operating systems
  - dig
  - host

Many web applications allow us to perform these queries too:

- Robtex.com
- ViewDNS.info
- dnsdumpster.com

Most recon tools and frameworks perform these queries as part of their enumeration processes

Performing DNS Queries

How do we actually perform DNS queries? It is quite simple. Let's use the sec487.info domain as our target.

From Windows computers:

1. Launch a command prompt and type: `nslookup sec487.info`
2. Or launch a PowerShell window and type: `Resolve-DnsName sec487.info`

On computers with macOS, Linux, or UNIX:

1. From a terminal window type: `host -a sec487.info`
2. From a terminal window type: `dig -t any sec487.info`

There are web applications that also allow these DNS queries. Some excellent sites for these searches are:

- Robtex.com - https://sec487.info/fi
- ViewDNS.info - https://sec487.info/fj
- DNSDumpster.com - https://sec487.info/nd

© SANS Institute 2019

NSLookup

Running nslookup is simple from a terminal or command line

The format of the command is "`nslookup DOMAIN`"

```
student@sec487 (09:41:56) : ~$ nslookup sans.org
Server:         8.8.8.8
Address:        8.8.8.8#53

Non-authoritative answer:
Name:   sans.org
Address: 45.60.103.34
Name:   sans.org
Address: 45.60.31.34
```


NSLookup

The easiest way to run a lookup of a domain from a terminal or command-line is using the nslookup command. Simply type: `nslookup DOMAIN` (where you replace DOMAIN with the domain you wish to resolve). The slide shows an example of this with the sans.org domain. The resolved addresses were 45.60.103.34 and 45.60.31.34.

Running the `nslookup` command without adding a domain on the command line makes the tool enter an interactive mode.

©2019 Micah Hoffman

© SANS Institute 2019

Dig DNS Command-line Tool

Standard and preferred command-line DNS query tool on macOS and Linux/UNIX

Request a single type of record or "all"

Specify DNS server to use with "@SERVER"

Common dig commands:

```
$ dig -t all sans.org
$ dig -t mx sans.org
$ dig @8.8.8.8 sans.org
```

Can add "`+noall +answer`" to the command for less-verbose output like:

```
$ dig +noall +answer sans.org
```


Dig DNS Command-line Tool

On macOS and Linux systems, the nslookup tool is deprecated (they do not want you to use it). Dig is the preferred DNS query command-line tool on those systems, and it is fairly simple to use. By passing the "`-t`" flag, you can request certain types of DNS records (such as MX, NS, and TXT). You can also send the DNS queries to certain DNS servers using the "`@DNSSERVER`" (where you replace the DNSSERVER with an IP address or domain name of the DNS server you wish to use).

The normal output of the dig command is verbose and contains so much irrelevant data it is sometimes hard to see the results. In these cases, we can tell dig to not show all the extra content and just show the results or answer using the "`dig +noall +answer DOMAIN`" format.

© SANS Institute 2019

Dig Example

Running a simple "dig sans.org" returns the output on the right with the data we want in the ANSWER SECTION

Robin Wood has a site for trying DNS zone transfers so you can see all the data DNS can hold (see notes)

```
student@sec487 (10:06:14) :~$ dig sans.org

; <<>> DiG 9.11.3-1ubuntu1.8-Ubuntu <<>> sans.org
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 18579
;; flags: qr rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 512
;; QUESTION SECTION:
;sans.org.                      IN      A

;; ANSWER SECTION:
sans.org.               9       IN      A       45.60.31.34
sans.org.               9       IN      A       45.60.103.34

;; Query time: 11 msec
;; SERVER: 8.8.8.8#53(8.8.8.8)
;; WHEN: Thu Sep 26 10:06:25 EDT 2019
;; MSG SIZE  rcvd: 69
```


Dig Example

Here we ran the dig command and asked it to retrieve information about the sans.org domain. Notice all the extra data we get back from the dig command? The only section we want is the "ANSWER SECTION" (inside the red box in the slide), which contains the IP addresses of the A records for the sans.org domain. However, dig sends more debugging information than we need.

While requesting individual records can be interesting, requesting all of the records that the DNS servers have about a domain can be a great shortcut to your reconnaissance. We can try to perform a DNS zone transfer on servers and see if they will divulge all the data about their domains. Unfortunately, most DNS servers are properly configured to not allow just any device to request this data. One server, owned and operated by Robin Wood (@digininja), allows us to try these zone transfers and examine the types of data that can be stored in a DNS server. His web site at https://sec487.info/sy explains how you can perform a zone transfer against his test domain. We show an image from the output of those commands below.

```
student@sec487 (13:37:23) : ~$ dig +noall +answer axfr @nsztm1.digi.ninja zonetrans
fer.me
zonetransfer.me.        7200    IN      SOA     nsztm1.digi.ninja. robin.digi.ninja
. 2017042001 172800 900 1209600 3600
zonetransfer.me.        300     IN      HINFO   "Casio fx-700G" "Windows XP"
zonetransfer.me.        301     IN      TXT     "google-site-verification=tyP28J7JA
UHA9fw2sHXMgcCC0I6XBmmoVi04VlMcwxA"
zonetransfer.me.        7200    IN      MX      0 ASPMX.L.GOOGLE.COM.
zonetransfer.me.        7200    IN      MX      10 ALT1.ASPMX.L.GOOGLE.COM.
zonetransfer.me.        7200    IN      MX      10 ALT2.ASPMX.L.GOOGLE.COM.
zonetransfer.me.        7200    IN      MX      20 ASPMX2.GOOGLEMAIL.COM.
zonetransfer.me.        7200    IN      MX      20 ASPMX3.GOOGLEMAIL.COM.
zonetransfer.me.        7200    IN      MX      20 ASPMX4.GOOGLEMAIL.COM.
zonetransfer.me.        7200    IN      MX      20 ASPMX5.GOOGLEMAIL.COM.
zonetransfer.me.        7200    IN      A       5.196.105.14
zonetransfer.me.        7200    IN      NS      nsztm1.digi.ninja.
zonetransfer.me.        7200    IN      NS      nsztm2.digi.ninja.
_sip._tcp.zonetransfer.me. 14000 IN     SRV     0 0 5060 www.zonetransfer.me.
14.105.196.5.IN-ADDR.ARPA.zonetransfer.me. 7200 IN PTR www.zonetransfer.me.
asfdbauthdns.zonetransfer.me. 7900 IN   AFSDB   1 asfdbbox.zonetransfer.me.
asfdbbox.zonetransfer.me. 7200  IN      A       127.0.0.1
asfdbvolume.zonetransfer.me. 7800 IN    AFSDB   1 asfdbbox.zonetransfer.me.
canberra-office.zonetransfer.me. 7200 IN A      202.14.81.230
cmdexec.zonetransfer.me. 300    IN      TXT     "; ls"
contact.zonetransfer.me. 2592000 IN     TXT     "Remember to call or email Pippa on
 +44 123 4567890 or pippa@zonetransfer.me when making DNS changes"
```


©2019 Micah Hoffman

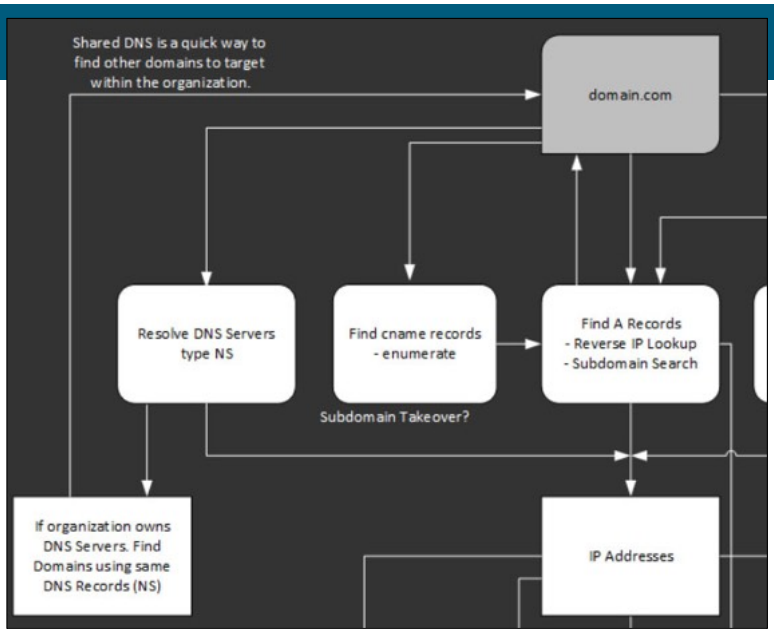
© SANS Institute 2019

DNSdumpster.com

Performs a variety of DNS lookups and analysis for you for free

Teaches why DNS is important and has sample work flow

Graphs connections between systems



DNSdumpster.com

The free https://dnsdumpster.com web site performs a variety of DNS functions for users. It will retrieve all manner of DNS records (MX, NS, TXT, A, etc.) and then group them in an easy-to-view format on the web page. Under each host are buttons that allow analysts to dive deeper into the data for that system. The site also has a feature to export all the hosts into an Excel XLSX file for archiving or post-processing.

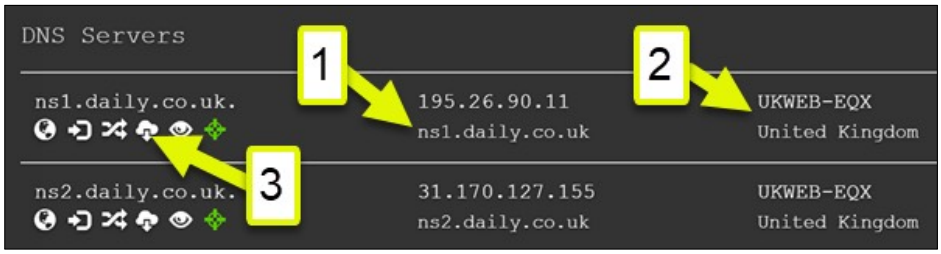
The site contains an excellent flow chart of how to perform DNS reconnaissance (https://sec487.info/ui), some of which is shown in the slide above.

Images from https://dnsdumpster.com/, September 27, 2019.

© SANS Institute 2019

DNSdumpster.com Example (1)

Domain:
leave.eu



OSINT data:

- NS records point to daily.co.uk (1)
- Network owned by UKWEB-EQX (2)

Both are additional sources for investigation

DNSdumpster.com Example (1)

Using the domain leave.eu, a domain with a web site dedicated to sharing content about leaving the European Union (EU), we will walk through why DNSdumpster.com and sites like it can advance our OSINT cases.

We searched for the leave.eu site on the DNSdumpster.com web site and saw that the DNS NS (name server) records pointed towards the ns1 and ns2.daily.co.uk servers (arrow 1). Those servers are hosted on the network owned by UKWEB-EQX company (arrow 2). These two points get added to our documentation as possible future points to research. Pressing the cloud icon (arrow 3) reveals any other hosts that DNSdumpster knows about that also use these name servers. Turns out that there are no other servers using these name servers. We note that and continue our work.

Images from https://dnsdumpster.com/, November 10, 2019.

©2019 Micah Hoffman

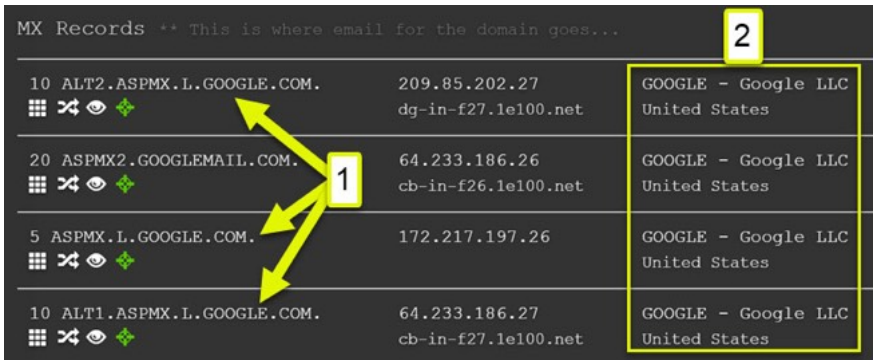
© SANS Institute 2019

DNSdumpster.com Example (2)

OSINT data:

- MX records point Google's servers (1)
- Network owned by Google Inc. (2)

So they may be using Google's G Suite for mail[1]



DNSdumpster.com Example (2)

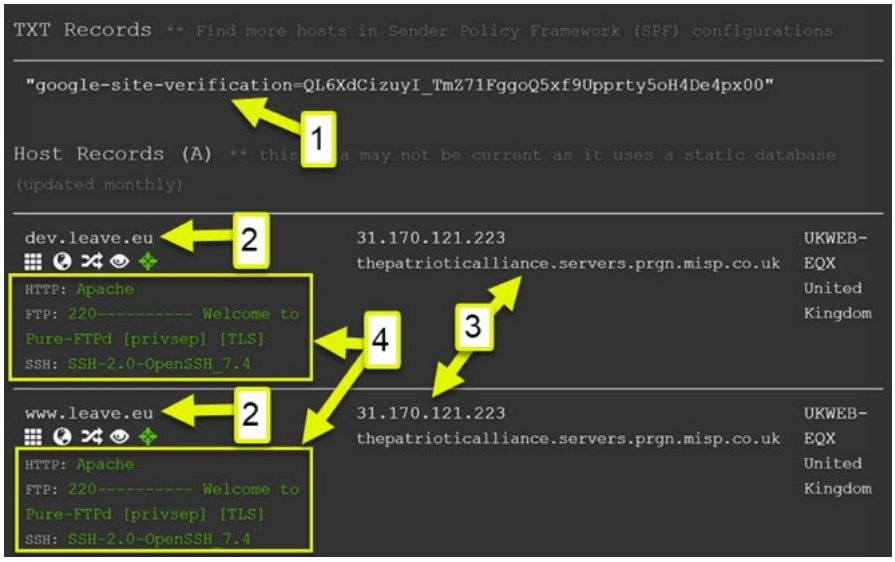
Remembering that MX records indicate what servers receive emails for a domain, we can see that leave.eu is sending its email traffic to Google's servers. This most likely means that the site is using Google's G Suite[1] online applications for its email and possibly for other activities, such as sharing files and calendars.

Images from https://dnsdumpster.com/, November 10, 2019.

Reference: [1] https://sec487.info/ul

© SANS Institute 2019

DNSdumpster.com Example (3)



DNSdumpster.com Example (3)

Continuing to scroll down the page at DNSdumpster.com, we see one TXT record (1). In this example, we see that the owners of the leave.eu domain proved to Google that they owned the leave.eu domain by inserting a TXT record with a "google-site-verification" entry. The entry value (beginning with QL6) should be noted and is unique to this domain and Google. You should not be able to find another domain that has the same google-site-verification value. Seeing it means that they most likely registered their domain with Google's services and supports that they are using Google products, perhaps the G Suite we mentioned earlier.

Finally, we have two DNS A records for dev.leave.eu and www.leave.eu (the arrows with the number 2). Looking to the right at what IP addresses they resolve to we can see they resolve to the same system, 31.170.121.223 (arrows at 3). This is common and further support for these subdomains being hosted on the same system comes from the services that both systems were found to be running in the boxes at number 4. The services are identical and show a web server, an FTP server, and an SSH server on the boxes. What this most likely means is that one of these systems is a CNAME for the other.

Images from https://dnsdumpster.com/, November 10, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Examining DNS Query Responses

We performed the queries; now what are we looking for in the results?

It really depends upon what your OSINT assessment goals are

We can find systems and user information from DNS record queries

Some possible next steps:

- Discover what services are running on the discovered domains/IPs
- Pivot and perform whois on domains found to get user information
- Collect the data for a social engineering attack
- Perform recursive DNS and whois queries

Examining DNS Query Responses

The DNS query results can be used for a variety of tasks in OSINT. We have seen how we can gather user, company, and other system information from these queries (and some other tools built into the web applications we might use). Depending upon what your OSINT goals are, there a several paths you could follow once you have this data:

- Use those IP addresses and domain names to find available services (such as web servers) that you wish to investigate
- Use the reverse whois techniques to discover what other domains users in your target organization might have established using their work email addresses
- Conduct a social engineering attack against a person or company
- Perform recursive DNS and whois queries against each IP and domain name that you discovered

© SANS Institute 2019

Common DNS Subdomain Enumeration Methods

Name	Explanation	Examples
Dictionary "Attack"	A tool uses a list of words to make many requests to a DNS server. The tool adds each word in the list, one by one, to a domain.	**dev**.sec487.info **admin**.sec487.info **www1**.sec487.info
Zone Transfer	A tool requests the entire "zone" from a DNS server. If the server is not secured, it may provide everything it stores about the given domain.	dig -t axfr sec487.info
Reverse or PTR Lookup	A tool requests 1 or more IP addresses in a range, looking for those with DNS entries.	dig -x 8.8.8.8

Common DNS Subdomain Enumeration Methods

We have several methods of extracting subdomains and IP addresses from these servers. Some of these techniques, such as a "dictionary attack," are called "attacks" but are merely rapid requests for information from a DNS server. They don't exploit the systems (aside from maybe using automation to make hundreds or thousands of requests for resolution each minute).

Above, we outlined three of the most common methods used to extract additional data from DNS servers. A few extra points about each are provided below.

- Dictionary attacks can use any list of words or strings to make subdomain names to be resolved. The tool being used to conduct the requests takes the first entry from the list, adds it to a domain that you provide, and then makes a DNS request for resolution. When it gets a reply, it then takes the next word from the list and adds it to the domain you gave it and makes another request. It continues to do this until the list, or dictionary, is exhausted.
- Zone transfers take advantage of misconfigured or less-hardened DNS servers. Zone transfers are valid methods of DNS servers staying synchronized. Primary and backup DNS servers must share their data so that if the primary crashes, the backup can continue to service the domain resolution requests. What we do is tell the DNS servers that they should synchronize their DNS data about a certain zone (domain) with our system and send us everything they have. Most DNS servers are configured to only provide this content to authorized, specific DNS servers (such as known backup servers). Some systems may allow our clients to conduct these types of DNS attacks.
- Reverse or PTR lookups are where our tools send one or more IP addresses to a DNS server and ask the server to tell us if any of the IP addresses have corresponding subdomains or hostnames. If we know that a business owns the addresses from 8.8.8.1 to 8.8.8.255, we can request each address (8.8.8.1, 8.8.8.2, 8.8.8.3, etc.) and see if there are entries for hostnames. This techniques is useful for identifying virtual servers that may resolve to the same IP address. Additionally, this technique can be effective when we know an IP address block that a company owns, as we can enumerate the entire block of addresses quickly.

Additional information about DNS enumeration can be found at https://sec487.info/xm.

©2019 Micah Hoffman

© SANS Institute 2019

DNSRecon DNS Enumeration

Carlos Perez created a free Python script

- Enumerates DNS names
- Performs reverse lookups
- Executes zone transfers
- Reverse PTR lookups
- Conducts dictionary attacks

```
student@sec487 (16:30:42) : /opt/tools/dnsrecon$ ./dnsrecon.py -t std -d sans.org
[*] Performing General Enumeration of Domain:
[-] DNSSEC is not configured for sans.org
[*]       SOA dns21a.sans.org 66.35.59.7
[*]       NS dns21b.sans.org 66.35.59.8
[*]       Bind Version for 66.35.59.8 9.3.6-P1-RedHat-9.3.6-25.P1.el5 11.12
[*]       NS dns31a.sans.org 204.51.94.7
[*]       Bind Version for 204.51.94.7 9.3.6-P1-RedHat-9.3.6-25.P1.el5_11.12
[*]       NS dns31b.sans.org 204.51.94.8
[*]       Bind Version for 204.51.94.8 9.3.6-P1-RedHat-9.3.6-25.P1.el5_11.12
[*]       NS dns21a.sans.org 66.35.59.7
[*]       Bind Version for 66.35.59.7 9.3.6-P1-RedHat-9.3.6-25.P1.el5_11.12
[*]       MX sans-org.mail.protection.outlook.com 216.32.181.42
[*]       MX sans-org.mail.protection.outlook.com 216.32.180.106
[*]       A sans.org 45.60.33.34
[*]       A sans.org 45.60.31.34
[*]       TXT sans.org MS=ms15381092
[*]       TXT sans.org status-page-domain-verification=2w964jh22625
[*]       TXT sans.org v=spf1 mx a:lists.sans.org a:mass1a.sans.org ip4:66.35.59.0
```


DNSRecon DNS Enumeration

Just looking to perform a variety of DNS lookups on targets? DNSRecon by Carlos Perez might be a great choice. This Python script can perform a wide variety of DNS queries and internet lookups to find DNS data. From zone transfers to brute-forcing domains in a dictionary attack (to find subdomains), DNSRecon can do it all. Download it from his GitHub repository at https://sec487.info/hp.

© SANS Institute 2019

Aquatone DNS Discovery

Ruby script by Michael Henriksen

Executes DNS queries and searches to find domains and IPs and can take pictures of web sites using "gather"

Uses APIs to query Shodan, Censys, and others

```
student@sec487 (16:40:13) : /opt/tools$ aquatone-discover --domain sans.org
                                                  __
  ____ ____ ___  _____ ____ ____ ___ _____ ____ ___
 / __ `/ __ `/ / / / __ `/ __/ __ \/ __ \/ __ \
/ /_/ / /_/ / /_/ / /_/ / /_/ /_/ / /_/ / /_/ /  __/
\__,_/\__,_/\__,_/\__,_/\__/\____/_/ /_/\___/
       /_/  discover v0.5.0 - by @michenriksen

Identifying nameservers for sans.org... Done
Using nameservers:

 - 66.35.59.8
 - 204.51.94.7
 - 204.51.94.8
 - 66.35.59.7

Checking for wildcard DNS... Done

Running collector: HackerTarget... Done (73 hosts)
Running collector: Wayback Machine... Done (33 hosts)
Running collector: Google Transparency Report... Done (22 hosts)
Running collector: DNSDB... Done (126 hosts)
```


Aquatone DNS Discovery

Michael Henriksen created the Ruby project called Aquatone. It runs via a terminal window and performs many of the same types of queries as DNSRecon. Aquatone also uses API keys (that you gather on your own and add to the application) and web queries to discover DNS names, subdomains, and IP addresses. The GitHub repository at https://sec487.info/hq has examples on how to run the program.

While aquatone-discover finds domain names and IP addresses, the aquatone-gather script can reach out to the sites that have web servers and collect HTTP responses and the web content being served, and it can take pictures of what the web pages look like when rendered in a browser.

©2019 Micah Hoffman

© SANS Institute 2019

Typosquatting and DNS

```
 __| |_.__  ___| |___      _(_)___| |_
/ _` | '_ \/ __| __\ \ /\ / / / __| __|
| (_| | | | \__ \ |_ \ V  V /| \__ \ |_
\__,_|_| |_|___/\__| \_/\_/ |_|___/\__| {20180623}
```

Ever accidentally visited https://goggle.com instead of google.com?

Typos happen, and attackers host malicious content on those domains

Use the Python dnstwist tool to find domain typosquatters

Homoglyph	sams.org	-
Homoglyph	ṣans.org	-
Homoglyph	sams.org	MX:mx.netidentity.com.cust.hostedemail.com
Homoglyph	sanṣ.org	-
Homoglyph	säns.org	-
Homoglyph	sans.org	-
Homoglyph	sáṇg.org	-
Homoglyph	saṛs.org	-

Replacement	sanx.org	-
Replacement	sany.org	MX:alt1.aspmx.l.google.com
Replacement	sand.org	-
Replacement	sane.org	-
Replacement	sahs.org	104.199.117.56 NS:dns047.b.register.com
Subdomain	s.ans.org	206.222.45.38
Subdomain	sa.ns.org	-
Subdomain	san.s.org	52.184.158.10
Transposition	asns.org	-
Transposition	snas.org	-
Transposition	sasn.org	-
Vowel-swap	sons.org	-
Vowel-swap	suns.org	-

SANS Open Source Intelligence (OSINT) Gathering and Analysis 55

Typosquatting and DNS

When we humans type, we make mistakes. We accidentally leave out a character or put the wrong letter into a domain in a URL. It happens. People who seek to exploit user browsers know this, too, so they use these common typographic errors in domains to register similar domains to a target's. When someone accidentally types https://goggle.com or https://yawho.com into their browser, the browser visits the attacker's domain instead of the normal site.

We can use the dnstwist Python script (https://sec487.info/ho) to enumerate many of the possible domains that someone may register to take advantage of domain typos. The script uses many rules to perform DNS lookups (and other tests) on a large amount of variations of a given domain name. It will bring back data about which are registered and to whom. It can also examine web content to see how similar it is to the original site (checking for phishing sites).

In the above image, we see the output from running the script against the sans.org domain.

Image from https://sec487.info/ho, March 31, 2018.

© SANS Institute 2019

dnstwister.report Web Site

Want a web site that does the same thing as dnstwist but MUCH faster?

That is the dnstwister.report site

sany.org	Replacement	192.124.249.168	analyse
s.ans.org	Subdomain	206.222.45.38	analyse
san.s.org	Subdomain	52.184.158.10	analyse
asns.org	Transposition	67.23.10.210	analyse
snas.org	Transposition	184.168.221.104	analyse
sasn.org	Transposition	184.168.221.104	analyse
suns.org	Vowel swap	52.52.34.204	analyse
sons.org	Vowel swap	69.172.201.153	analyse

dnstwister.report Web Site

If you like the dnstwist application but cannot run it from the command line using Python, visit the https://dnstwister.report web site and just enter your domain of interest in the search field. The site will do all the requests to look up the domains and then will provide you a clean report of what it found. In the above example for the sans.org domain (https://sec487.info/ne), we can see its output is similar to the dnstwist tool.

Image from https://sec487.info/ne, January 3, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

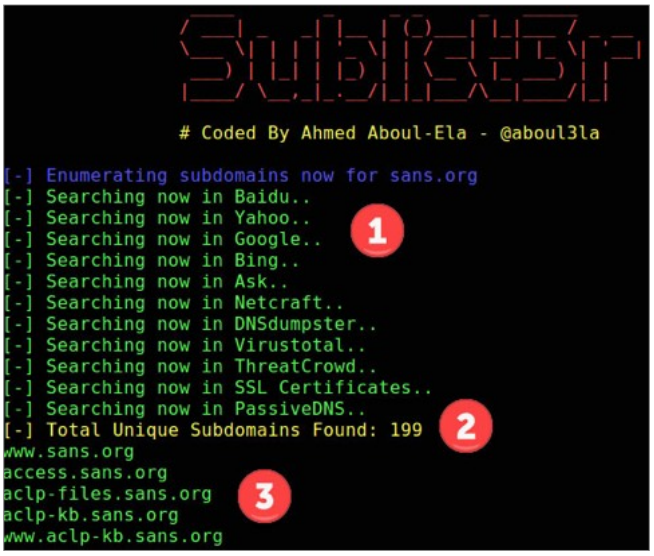
Sublist3r Subdomain Enumeration Tool

The Sublist3r Python tool retrieves domain information from web sites

It can also perform active attacks against systems like port scanning

Example command:

```
python sublist3r.py
-d sans.org
```



Sublist3r Subdomain Enumeration Tool

The Sublist3r Python tool can be found in the GitHub repository at https://sec487.info/nf and is a bit different from some of the other web sites and tools that we have seen. This tool can gather data about a domain from web sites like Google, Baidu, and DNSdumpster, but also can perform active interrogation of the domains it finds to find open TCP ports or perform DNS brute-force/dictionary queries against a DNS server. The tool runs with all sources enabled by default, or the analyst can choose only certain sources to be used for the recon.

To run the tool, type the following: `python sublist3r.py -d DOMAIN` (and replace DOMAIN with the domain you wish to discover data about). Once the tool starts, you will see it query a variety of sources (1 in the image above) and report the number of subdomains it found (2). Then it will output the subdomains it finds (3) to the screen or to a text file of your choosing.

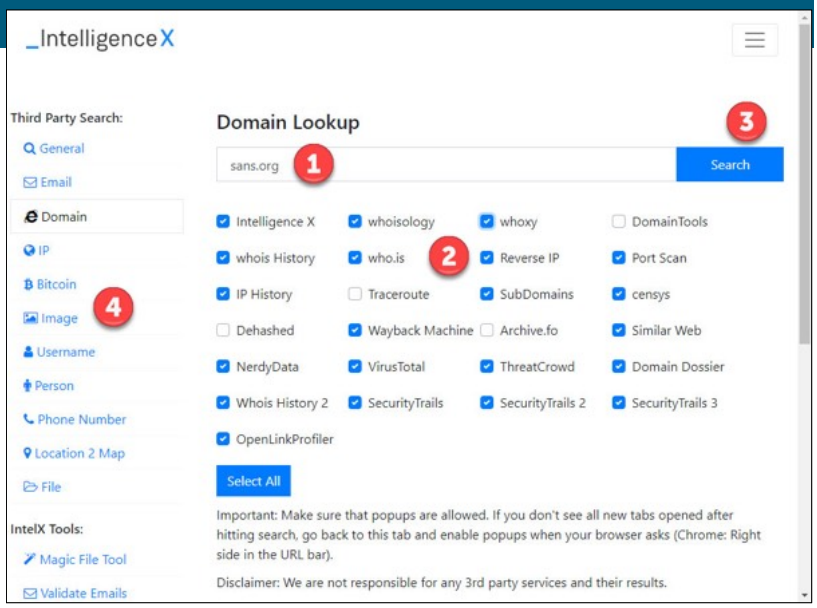
© SANS Institute 2019

Intelligence X (Intelx.io)

Fill out web form on the site

Performing searches causes new tabs to appear in your browser to make direct requests to other sites

Variety of "tools"



Intelligence X (Intelx.io)

For years, we have had web sites like Michael Bazzell's inteltechniques.com, where people could visit a "tools" web page, fill in some data, and then perform searches. Instead of his server making the requests and returning the data to your browser, his site would craft special URLs that embedded your search terms in the URL and then would open up each of those new URLs in a new tab of your local browser. That way, his server was not making the requests to the servers you wanted to search.

Earlier in 2019, the inteltechniques.com OSINT tools page was moved behind a pay wall to restrict access to it.[1] In the vacuum that was created by these tools not being there for OSINT professionals to use, sites like Intelligence X's intelx.io filled the gap. The tools section of this site (https://sec487.info/um) has a web form for domain name lookups (shown above). Enter the domain of your choice (1), select the third-party web sites to search (2), and click the search button (3). If you ensure that your browser allows new tabs to be created (popups) by this site, you will see additional tabs spawn for each site you selected to search.

We encourage you to be OSINT curious and click around on the intelx.io web site, as it has a large number of additional OSINT-related search forms that can help you in your work (4 above).

Image from https://sec487.info/um, September 27, 2019.

Reference: [1] https://sec487.info/un

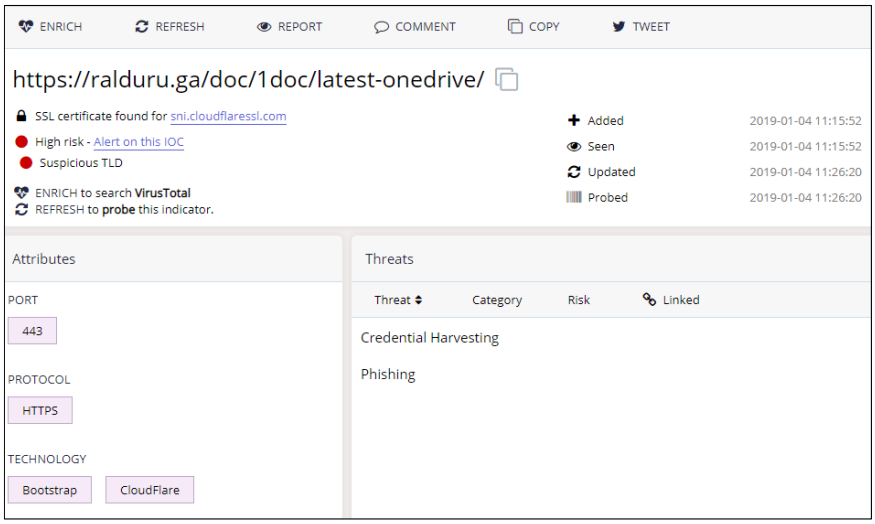
 ©2019 Micah Hoffman

© SANS Institute 2019

Pulsedive for Cyber Threat Intel and OSINT

Pulsedive aggregates data about IP addresses, domain names, malware, and IOCs

Anonymous and free account access



SANS | Open Source Intelligence (OSINT) Gathering and Analysis 59

Pulsedive for Cyber Threat Intel and OSINT

The pulsedive.com web site is a free, easy-to-use resource for searching for Indicators of Compromise (IOCs), researching IP addresses and domain names, and gaining a better understanding of emerging cyber threats. It can be accessed through the public and anonymous interface or, for increased access to site features, you can create a free account before performing your searches.

The site has a simple search field that can be used to submit data to be matched in its system. But this site goes further and allows you to comment, submit suspicious IOCs, and participate in the research. When you register for an account, you get an API key that can be entered into other tools so that they can search Pulsedive.

Image from https://sec487.info/og, January 4, 2019.

© SANS Institute 2019

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Domains and IPs"

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. DNS
4. **Findings Online Devices**
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. Government Data
8. Researching Companies

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Internet-connected Systems

There are systems that scan the internet for computers and allow us to search their findings

We can examine the ports open on these devices, the services they provide, whether they have authentication, and look for "interesting" things

We will also examine two specialist sites that search for systems and the data they display

Censys

SHODAN

Internet-connected Systems

In previous sections of the class we learned about using whois and DNS to gather information about a domain or IP address. Now we can also use the specialized search engines at shodan.io and censys.io to reveal the systems we may need to visit. Both of these sites are constantly scanning the internet looking for computer systems of all shapes, sizes, and features and cataloging what they find. Whether someone has opened up the Remote Desktop Protocol (RDP) ports on their Windows 10 computer so that they can remotely log in to their computer or someone has placed a new web camera online, these sites will find them, probe what services are being offered, and store the results.

Images from https://www.shodan.io/ and https://censys.io/, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Censys.io

Scans the internet and catalogs the results for us to search

Searchable content includes page content, HTTPS certificate properties, and server responses

Start with what you know: a term, a domain, a subnet, or an IP address

Create a free account instead of using the anonymous mode and get an API key for programmatic access to Censys

They allow for downloading ALL of Censys' data for offline analysis

Cheat sheet with examples at https://sec487.info/t9

Censys.io

The Censys.io web site actively scans the internet and catalogs the results. The data that is retrievable is everything from contents of SSL and TLS HTTPS certificates (where we may find email addresses, names of employees, and internal domain names) to what ports are open on the systems to the HTML source code of any web pages that are returned.

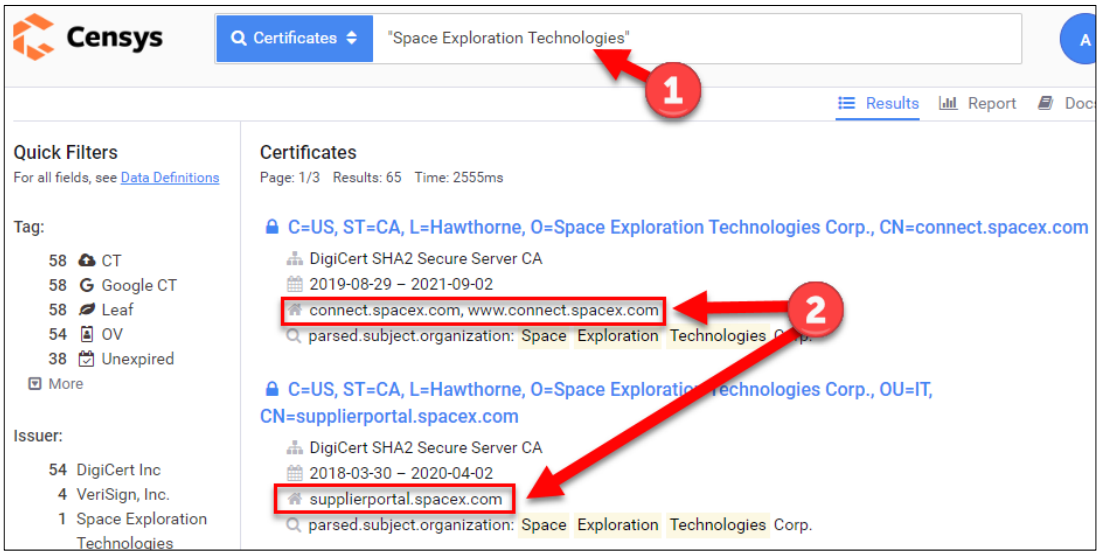
For help with searches, visit the https://sec487.info/nu site and examine the examples and suggestions. With no training, though, you can still use the search engine by typing in a domain name of interest (for example, spacex.com) and IP address or range, or a term of interest (for example, "Space Exploration Technologies").

The site allows anonymous searches but encourages users to use a free account for their work. A benefit of registering is that you can then use the API key given to registered users to access the site using tools such as SpiderFoot and Recon-ng.

Examine the contents of the https://sec487.info/nv page for information about what scans Censys.io is and has been performing and to download its data for offline analysis. You can also download the cheat sheet from https://sec487.info/t9 to learn about the powerful searches that can be performed.

© SANS Institute 2019

Censys.io Certificate Search



SANS Open Source Intelligence (OSINT) Gathering and Analysis 64

Censys.io Certificate Search

Encryption certificates contain data that may be sensitive in nature (email addresses, names of employees, internal domain names), and we can search them using the "Certificates" tab of the site. In the image above, we are searching for the string "Space Exploration Technologies" (more commonly known as "SpaceX") in any encryption certificate that Censys.io came across while scanning. Typically, we think of certificates with HTTPS traffic, but they also are present in VPNs (Virtual Private Networks) and encrypted mail. The results show domain names (in the red boxes) for systems we may not have known about before.

In the above slide, we entered the string "Space Exploration Technologies" into the certificate search. This made Censys examine the data it has collected from HTTPS certificates for this string. We can see two of the 65 results, and they both have this string in the organization "O" section of the HTTPS certificate. Using this technique, we may be able to identify systems that have HTTPS certificates from our target organization but are scattered across many cloud providers like Amazon, DigitalOcean, and Linode.

For help on the certificate searches, visit the data definitions at https://sec487.info/t8.

Image from https://sec487.info/az, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Censys.io IPv4 Results

Autonomous System:

- 13 SPACEX - Space Exploration Technologies Corporation
- 2 DATANOC - Lanset America Corporation
- 2 OVH
- 2 TOTAL-SERVER-SOLUTIONS - Total Server Solutions L.L.C.
- 2 UNIFIEDLAYER-AS-1 - Unified Layer

More

Protocol:

- 28 80/http
- 26 443/https
- 10 22/ssh
- 6 21/ftp
- 3 110/pop3

More

IPv4 Hosts "Space Exploration Technologies"

158.103.0.131 (cetactlab.soe.morgan.edu)
MORGAN-STATE-UNIVERSITY - Morgan State University (33170) Baltimore, Maryland, Unit
Ubuntu 443/https, 80/http
CESET | Center of Excellence in Systems Engineering for Space Exploration Technologies
atmocenter.org, www.atmocenter.org
80.http.get.body: Exploration Technologies</title> <!-- InstanceEndEditabl

67.201.58.238 (vm-www01.spacex.cl.zerolag.com)
TOTAL-SERVER-SOLUTIONS - Total Serve lutions L.L.C. (46562) El Segundo, California, States
443/https, 80/http
Coming Soon shop.spacex.com
443.https.get.body: Exploration Technologies Corp

107.21.249.222 (ec2-107-21-249-222.compute-1.amazonaws.com)
AMAZON-AES - Amazon.com, Inc. (14618) Ashburn, Virginia, United States
Amazon 22/ssh, 80/http
Lars Blackmore's Research
80.http.get.body: Exploration Technologies</a> (SpaceX

SANS Open Source Intelligence (OSINT) Gathering and Analysis 65

Censys.io IPv4 Results

If we keep the terms "Space Exploration Technologies" in the search field and switch to the IPv4 Hosts search, we get a variety of results where those terms appear in some content returned to Censys.io. We can see the networks where the systems were found on the left image: SPACEX, DATANOC, etc. In the middle image we can see a breakdown of the TCP ports (and their protocols) where Censys.io found the terms. And finally, the right-side image shows the details of where in the results the terms were found. We can, of course, pivot to any of the retrieved systems and gather more information about them by clicking on their IP addresses.

Examining the far-right image a bit more, we see places around the world where SpaceX may have business relationships (arrows at 2), other groups are working on SpaceX projects (arrow 1), and cloud servers (arrow 3).

For more details on IPv4 host searches, visit https://sec487.info/t6 for data definitions and https://sec487.info/t7 for examples.

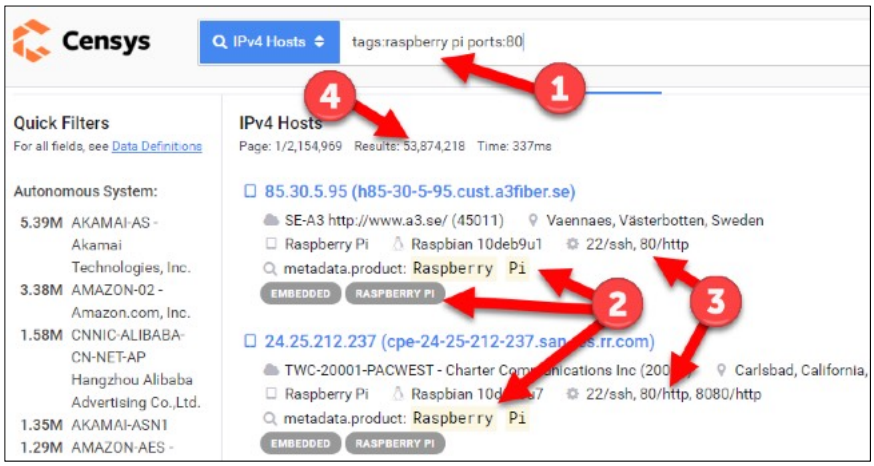
Images from https://sec487.info/b0, September 23, 2019.

© SANS Institute 2019

Censys Details (1)

Once you select a result, you can drill down and gather more details from Censys

Here, we performed a search for "tags:raspberry pi ports:80" [1]



Censys Details (1)

Let's take an example search from Censys and dive deeper into what data the site can reveal. We are specifically interested in small computers called Raspberry Pi's.[2] These embedded computers can be used for a variety of purposes and are sometimes internet-facing. We performed a query for them using the Censys "tags" parameter. We will add the web port 80 to this query to find Raspberry Pi computers on the internet that have web sites. Our search is "tags: raspberry pi ports:80" in the IPv4 section of Censys (https://sec487.info/td), shown in the slide above at arrow 1.

Our results revealed over 53 million hosts discovered (arrow 4). We can see that Censys tagged these devices (arrows at 2) and that they are running services on the common web port 80 (arrows at 3). Let's continue to perform our passive OSINT by clicking on one of these devices and see what else Censys knows about it.

Image from https://sec487.info/td, September 23, 2019.

References:

[1] https://sec487.info/td

[2] https://www.raspberrypi.org/

©2019 Micah Hoffman

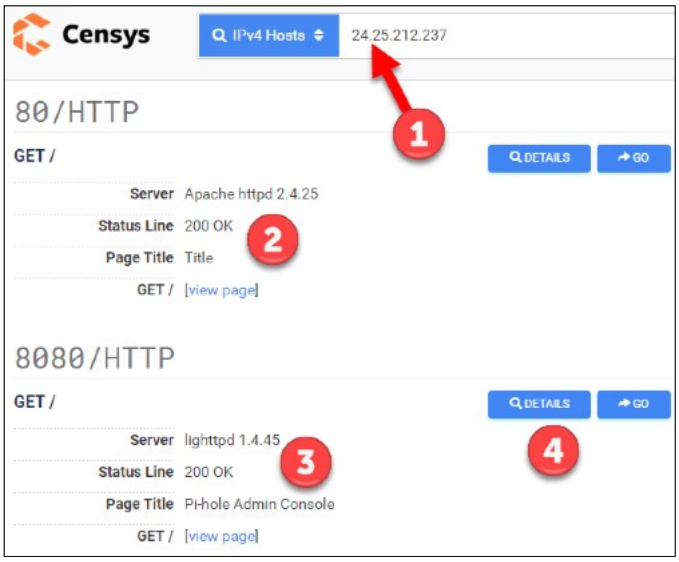
© SANS Institute 2019

Censys Details (2)

We selected a Raspberry Pi computer running a server on port 80

Detailed view shows the Apache web server running on port 80 (arrow 2) and lighttpd web server running on port 8080 (arrow 3)

Let's look at that 8080 server



Censys Details (2)

We selected the 24.25.212.237 host[1] to get details on. In the image above, we see the services running on ports 80 and 8080 are web servers. We see some of the parsed data that was returned from the servers. Now we drill down even further by clicking on the "Details" button (number 4).

Image from https://sec487.info/te, September 23, 2019.

References: [1] https://sec487.info/te

© SANS Institute 2019

Censys Details (3)

This page shows the HTML that the lighttpd server sent (1)

It appears to be a Pi-hole server (2)

8080.http.get.body	<!DOCTYPE html> <!-- Pi-hole: A black hole for Internet advertisements * (c) 2017 Pi-hole, LLC (https://pi-hole.net) * Network-wide ad blocking via your own hardware. * * This file is copyright under the latest version of the EUPL. * Please see LICENSE file for your rights under this license. --> <html> <head> <meta charset="UTF-8">
8080.http.get.body_sha256	ff30b9caec827a420651b8822c6a2ab3525b0284a820d9fb3acf
8080.http.get.headers.cache_control	no-store, no-cache, must-revalidate
8080.http.get.headers.content_length	14595
8080.http.get.headers.content_type	text/html; charset=UTF-8
8080.http.get.headers.expires	Thu, 19 Nov 1981 08:52:00 GMT
8080.http.get.headers.pragma	no-cache
8080.http.get.headers.server	lighttpd/1.4.45

SANS Open Source Intelligence (OSINT) Gathering and Analysis 68

Censys Details (3)

In this view, we see the raw data that was returned to the Censys scanners. Arrow 1 points to the location where the HTML contents of the response are stored. In this web data, we can read the HTML comment "<!-- Pi-hole: A black hole for Internet advertisements" and see further down the link to the Pihole project's web page (https://pi-hole.net). This page also shows content in the JSON format.

This completes the example of performing remote reconnaissance on a host by examining the contents cached within the Censys.io web site. Our next step here might be to visit the web site, to examine it for vulnerabilities from the version of the Pihole software disclosed, or to move on to another system. It depends on our OSINT goals.

Image from https://sec487.info/tf, September 23, 2019.

©2019 Micah Hoffman

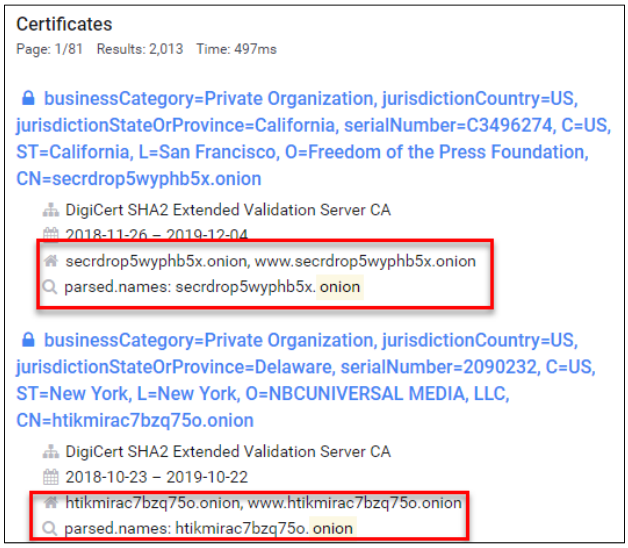
© SANS Institute 2019

Censys Searching for Onions

The Hunchly Dark Web Tor Dark Web setup guide[1] notes surface web sites sometimes have HTTPS certificates with Tor dark web names in them

We can refine this suggested Censys query to search certificates for:

"parsed.names: onion"



Censys Searching for Onions

Justin Seitz's Hunchly web site has a Dark Web setup guide[1] that instructs users how to use Hunchly within the Tor Dark Web network. He notes that sometimes you do not need to visit the dark web to discover hosts. Censys catalogs HTTPS certificates that may contain Tor Onion service names in the "parsed.names" parameter.

He provides a Censys search (https://sec487.info/tc) that shows this content within the IPv4 host search. Moving the search to the Certificates search, we change the search string to be "parsed.names: onion" (https://sec487.info/ta) and retrieve thousands of results.

Images from https://sec487.info/ta, September 23, 2019.

Resource: [1] https://sec487.info/tb

© SANS Institute 2019

Shodan.io

Shodan is a search engine for **internet-connected devices**

Like Censys.io, it **scans the internet** and catalogs the systems and services it finds

It takes **screenshots** of the web pages and login screens too!

Search header information sent from the server

Designed for engineers and requires use of **special search query terms** to get desired content

Highly recommended to get a **free or paid account** to run filters

Shodan.io

Shodan, unlike Google, scans and indexes internet devices and not just the web pages that they serve. It will interact with SSH (Secure Shell), FTP (File Transfer Protocol), web, mail, Internet of Things, and any other system and protocol it comes across.

For many of the services it discovers, it will take a picture of what the login screen or web site looks like and present that to you so that you can view the contents of a system without visiting it. In addition to recording what content is sent back to Shodan's crawler via the data portion of the responses, it also gathers the data in the headers of the traffic. So, if you are searching for a term and it appears in an HTTP cookie, Shodan can highlight that data.

Further differing from regular search engines, Shodan uses specific query properties, syntax, and filters for its queries. You can still submit "spacex" as a search term, but you may not receive all the results you need, or that Shodan contains, unless you employ the search filters. These filters can be found at:

- https://sec487.info/b1
- https://sec487.info/b2
- https://sec487.info/b6
- https://sec487.info/e9
- https://sec487.info/t4

Some of Shodan's features are only available to those people with registered user accounts, either paid or at the free level.

©2019 Micah Hoffman

© SANS Institute 2019

Shodan Filters of Interest

title: Content scraped from the HTML tag

html: Full HTML content of the returned page

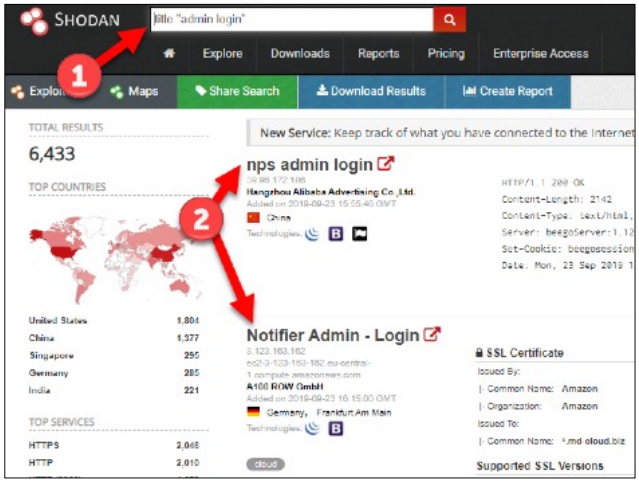
net: Network block

country: Search for results in a given country (2-letter code)

city: Restrict results to a city

postal: Search a US ZIP code

hostname: Host or domain name



Shodan Filters of Interest

The SANS blog post[1] highlights some of the Shodan filters that may be of interest to penetration testers. For our OSINT purposes, we can examine a subset of these for how they may be used in our investigations.

- `title`: If you know what the title of a web page is, you can use this filter to search for it. In the HTML source code of the page, the content will be between the <title> and </title> tags. The image above uses the `title:"admin login"` filter (arrow 1) to specify web pages that have words "admin login" in their title (arrows at 2 in the image above).[2]
- `html`: Examines the contents of the HTML web page for your terms.
- `net:` If you discovered the network addresses (subnet) of your targets, you can instruct Shodan to return results from those systems.
- `country:` This filter limits results to a specific country and is usually used in conjunction with our filters.
- city: Limits results to a specific city. If you know that your target has a system in certain city, use this.
- `postal:` Uses the United States postal codes to limit the area (similar to "city:") that results are returned from.
- `hostname:` Restrict the results to only those containing the given host or domain name.

References:

[1] https://sec487.info/b6.

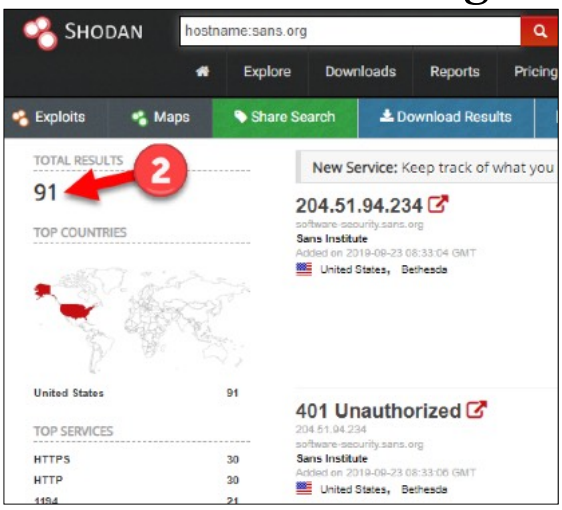
[2] https://sec487.info/t2, September 23, 2019.

© SANS Institute 2019

Shodan.io Filters Example

sans.org

hostname:sans.org



Shodan.io Filters Example

To illustrate using filters, consider that our customer may want us to find systems that have sans.org in the hostname or domain. Entering the "sans.org" string into Shodan (left image[1]) gets us 11 results (arrow 1), whereas searching with the "hostname:" filter on the right,[2] we get almost 9 times as many results (arrow 2).

Images:

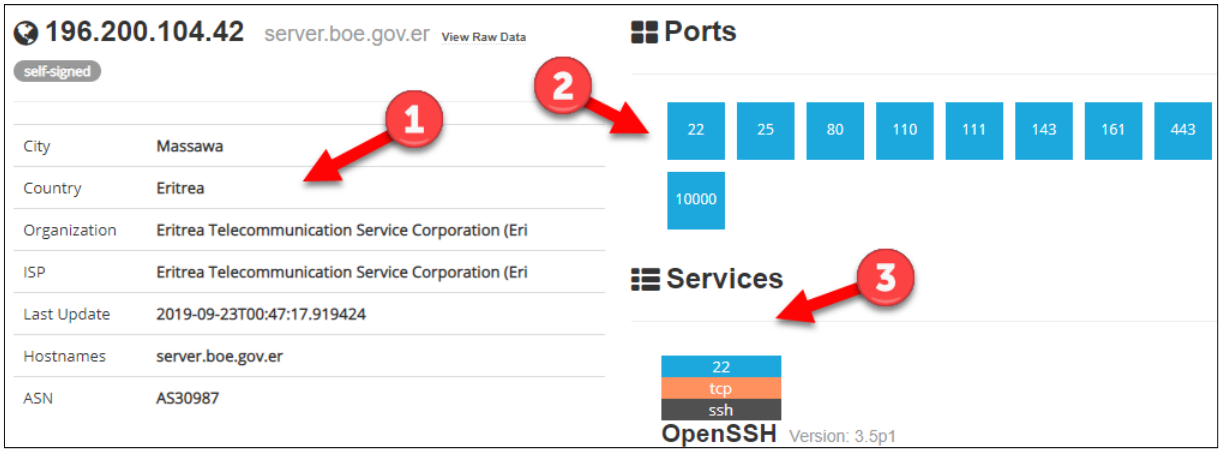
[1] https://sec487.info/b3, September 23, 2019.

[2] https://sec487.info/b4, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Shodan Details (1)



Shodan Details (1)

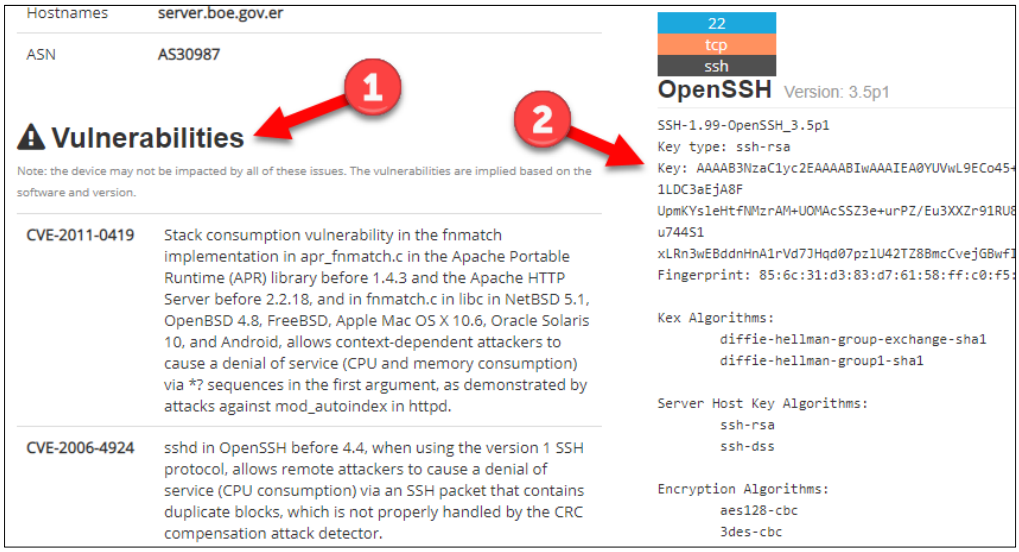
The image above shows the detailed results from a Shodan query. The information about a given host varies according to what services were found on the system, whether Shodan was prevented from accessing the content due to authentication, and other factors.

In the above image, we see general information about the IP address space where this host was discovered (arrow 1), what computer ports were open on the specific device when it was scanned (arrow 2), and the first service that was interrogated (arrow 3).

Image from https://sec487.info/b3, September 23, 2019.

© SANS Institute 2019

Shodan Details (2)



Shodan Details (2)

As we continue to scroll down the page, some paid users will see the vulnerabilities section (arrow 1). This section outlines what CVEs (Common Vulnerabilities and Exposures, https://cve.mitre.org/) have been associated with the data that the server sent to Shodan. For instance, if a web server was discovered through a Shodan scan, that device might be disclosing in the HTTP header response that it is using PHP version 5.3. That version of PHP may be known to be vulnerable to exploitation, and a specific CVE might be tagged to that issue. Shodan alerts us to these issues.

In the right column (arrow 2), we view a glimpse of the services section. Here Shodan displays the information that came back from the server when it was scanned. Many times this information contains helpful OSINT data that we should collect and analyze.

Image from https://sec487.info/b3, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Shodan Image Filters

When Shodan scans a target, it may take a picture of the response from the system

These responses can be web pages, VNC servers, and remote desktop connections

As a paid user, visiting https://images.shodan.io and using filters can reveal much

Image Filters:

- `screenshot.label:`**`desktop`**
- `screenshot.label:`**`ics`**
- `screenshot.label:`**`loggedin`**
- `screenshot.label:`**`login`**
- `screenshot.label:`**`pos`**
- `screenshot.label:`**`terminal`**
- `screenshot.label:`**`webcam`**
- `screenshot.label:`**`windows`**

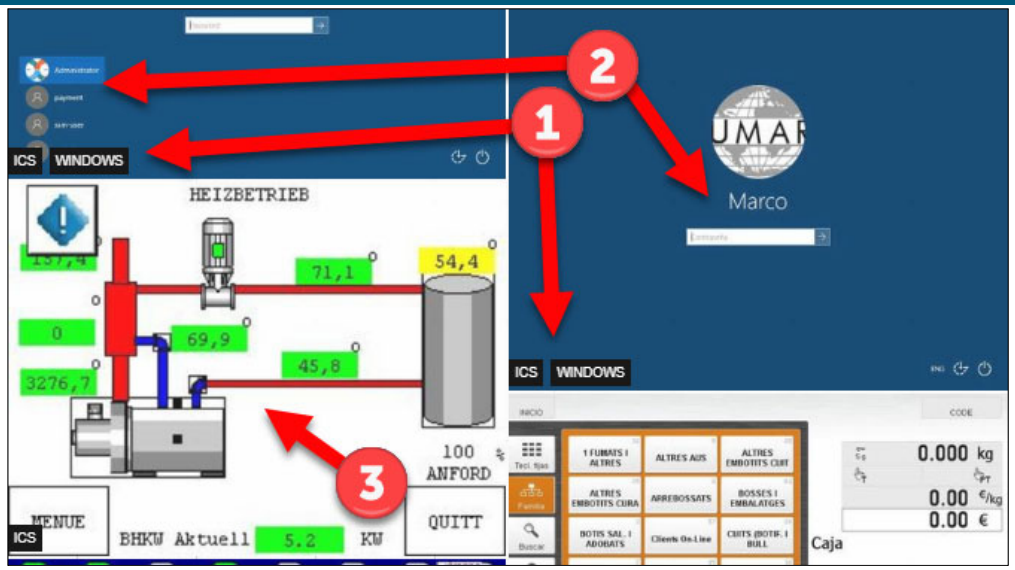
Shodan Image Filters

As Shodan's scanners interrogate and collect data from systems attached to the internet, they sometimes will encounter a device that presents a web page or remote login screen or might be logged in already. In these cases, Shodan takes a picture of what its scanner bots "saw" when visiting the page and then categorizes those images and allows authenticated, paid users to browse and search through them.

Visit the https://images.shodan.io web site as an authenticated user, and you can parse the images using the above filters to extract pictures that are interesting for your OSINT work.

© SANS Institute 2019

Shodan Image Example



Shodan Image Example

Here is an example of what the Shodan Images pages looks like. The arrows labeled “1” are pointing to the specific filters that are tagged to the images. In this case, the "ICS" and "WINDOWS" filters will display these images. Clicking on these images will take you to the system details page (already discussed). You can then examine which open port Shodan retrieved the image(s) from.

Some of the images display logged-in systems or login prompts (arrows at 2, above). We can harvest valid users and Windows domains from these systems.

As shown, some industrial control systems (ICSs) are attached to the internet and not protected with authentication or other access controls. In these cases, Shodan takes pictures of whatever it views on their systems. Arrow 3 points to a manufacturing monitoring system that appears to have interactive buttons at the bottom. Depending upon what the system does and if those buttons are enabled, anyone visiting this page could control this system.

Image from https://sec487.info/t5, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

More Shodan Resources

Jake Jarvis maintains a GitHub repository similar to the Google Hacking Database but for Shodan queries

Categorized queries like:

Traffic Light Controllers / Red Light Cameras →

```
mikrotik streetlight
```

Voting Machines in the United States →

```
"voter system serial" country:US
```

Home Devices

Yamaha Stereos →

```
"Server: AV_Receiver" "HTTP/1.1 406"
```

Apple AirPlay Receivers →

Apple TVs, HomePods, etc.

```
"\x08_airplay" port:5353
```


More Shodan Resources

If you are looking to find more within Shodan, consider visiting Jake Jarvis' Awesome Shodan Search Queries GitHub repository (https://sec487.info/v0). Here, he has categorized many interesting Shodan searches (as shown above). From ink jet and 3D printers on the internet to traffic lights and web cameras, Jake's site categories Shodan queries and provides links so you can easily search for content on Shodan.

Some of the queries use filters that only registered (free or paid) Shodan users can use.

Images from https://sec487.info/v0, September 30, 2019.

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. DNS
4. Findings Online Devices
5. **Wireless Networks**
6. Recon Tool Suites and Frameworks
7. Government Data
8. Researching Companies

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Wireless Wi-Fi Primer

Term	Explanation
SSID	Service Set Identifier. This is the name of the wireless network. Ex: ThePrincessBride
BSSID	Basic Service Set Identifier. Unique ID of a wireless transmitter in MAC address format. Ex: `00:20:91:a1:33:11`
beacon	Announcements from access points broadcasting SSIDs. Ties SSID to BSSID (access point).
probe requests	Announcement from client requesting a certain SSID. Ties request for SSID to a BSSID (client).

Wireless Wi-Fi Primer

Before we get into using wireless devices in our cases, we need to understand the terms we might see. The above slide explains a few of the terms. Let's get a little more detail on them.

The first 3 octets (3 groupings inside the colons (:)) of a BSSID represent the organization that made the device. This OUI, or Organizationally Unique Identifier, can be researched to determine the type of device or where it might have come from. Check out the text document at https://sec487.info/v3 for a reference of OUIs. Head over there and search for "00:20:91" from the above example and see who owns that OUI. The last 3 octets of the BSSID are mostly unique to the transmitter.

Note that both clients (mobile devices, laptops, etc.) and access points have unique BSSIDs.

SSIDs are broadcast from devices offering Wi-Fi services, such as an access point (AP), mobile hotspot or MiFi, or a device like a printer that has an ad-hoc wireless network configured. These devices offering wireless networks generally will announce their SSIDs by sending out wireless packets called beacons. Client wireless devices listen for these beacons and understand what SSIDs are near.

Instead of listening for the beacons from Wi-Fi network devices, the clients we use can send out probe requests asking if certain networks are nearby. If an AP receives a probe and can service the network the client is asking for, it will reply to the probe with a probe response.

For more information on wireless protocols, visit https://sec487.info/v4.

© SANS Institute 2019

Wireless 802.11 Wi-Fi Networks

Wireless or Wi-Fi networks surround most of us: in cars, on planes, in homes, in offices…

Wireless for OSINT:

- Wireless network names are generally unique
- Wireless networks can be geolocated
- Network names can be revealing

With unique networks that are known to be at a specific place, we can now better understand where a device may have been

Is it 100% accurate? No.

Some networks are not logged, and some are located incorrectly

Wireless 802.11 Wi-Fi Networks

Our modern world has become swathed in computer 802.11a, b, g, n, and ac wireless networks. Our computers, phones, tablets, watches, TVs, and other devices use these Wi-Fi networks to send and receive data. The network names or SSIDs (Service Set IDentifiers) for each network are generally unique. We say "generally" because there is no requirement for a network SSID to be unique to be valid. Within a location, especially at work sites, you will see multiple access points that have the same network name.

But across most of the world, Wi-Fi network names can be unique to a store, a hotel, a coffee shop, or a home. When we find these network names in our investigations, we need to see if we can locate that network somewhere in the world. If we can, then there is a good chance that the device was at that location. Of course, there are limits and exceptions to locating a network to a certain geographic location. We will see that the techniques that people use to collect this information can be skewed and erroneous.

We also should pay attention to the name of the wireless networks we find. They may include family names, sentiments, locations, and more.

©2019 Micah Hoffman

© SANS Institute 2019

Google Collected Wi-Fi Data

In 2010, Google was found to be collecting Wi-Fi data from their Google Maps vehicles

As they drove around, they would "sniff" wireless networks and log their GPS positions

This would help Google geolocate devices via Wi-Fi

Google admits collecting Wi-Fi data through Street View cars

German request for data audit reveals the web giant 'accidentally' stored payload information from open networks

A Google mapping car with which the web giant admits accidentally collecting wifi data. Photograph: Ben Birchall/PA

Google Collected Wi-Fi Data

In 2010, Google was found to be "accidentally" collecting wireless network names and data with their Google Maps vehicles that drove around.[1] Collecting the data became a big issue for the company due to privacy laws in many of the regions where Google collected the signals.[2] As Google's mapping vehicles drove around roads, they would also sniff the wireless traffic from the air, collect it, and record the GPS locations of those signals. They wanted to use this data to increase the accuracy of their location services. Understanding all of the wireless networks in a region allowed them to more effectively locate devices.

Unfortunately, people did not like the idea of someone driving a vehicle around their towns and collecting the wireless data from the air. Little did they know that people have been doing this privately for over a decade before Google got caught doing it.

References:

[1] (Image) https://sec487.info/9y

[2] https://sec487.info/9z

© SANS Institute 2019

WarXing

Since **Peter Shipley**'s work in 2000, people have been war-walking, war-driving, war-boating, war-droning...

War [INSERT VERB] is moving around an area, **collecting wireless network information**, and **geolocating** it to a specific place

Collecting Wi-Fi data is simple using freely available tools like Kismet

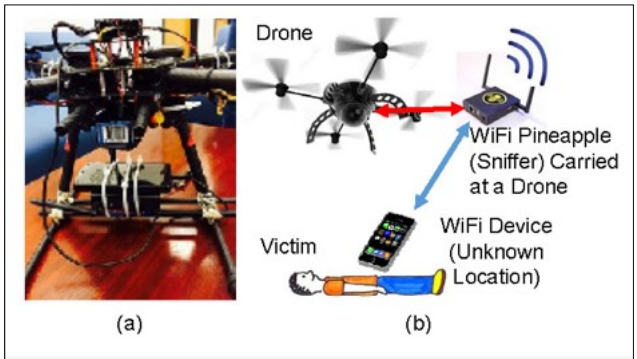


Fig. 1. (a) Experimental setup with WiFi Pineapple mounted on a drone, and (b) Model used to access packets from WiFi Pineapple remotely to localize a WiFi device.

WarXing

In 1999–2000, Peter Shipley created a technique to automatically log wireless network information along with the inputs from a GPS as he drove his vehicle around an area.[1] He coined the term "war-driving"[2] for this activity, and both the term and the technique took off! People bought wireless devices (at that time the Orinoco PCMCIA cards were in high demand), serial (and later USB) GPSs, and antennae and set off in their vehicles to find and map networks. Using the Wayback Machine (https://sec487.info/a0), we can look at some of the old images that Peter Shipley created from his war-driving.

The premier tool for collecting this data has been the Kismet (https://kismetwireless.net/) free software. Over the years, other, more specialized clients developed to do this work. The easiest to use of these clients is WiGLE.net's Android client (https://sec487.info/v5).

Since that time, this practice has evolved with better applications and more-flexible platforms for moving about. People regularly use boats, bikes, and even flying drones [3] [4] to collect wireless network data in this fashion. Some of the better free tools are Kismet, Kismac, and the Wigle.net Android client.

Image from https://sec487.info/v7, September 30, 2019.

References:

[1] https://sec487.info/a2

[2] https://sec487.info/a1

[3] https://sec487.info/v6

[4] https://sec487.info/v7

 ©2019 Micah Hoffman

© SANS Institute 2019

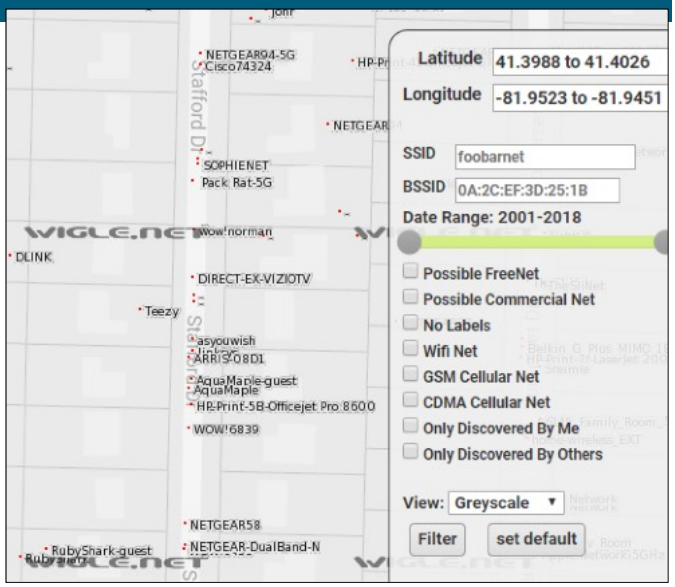
WiGLE.net

The WiGLE.net web site allows us to search for submitted geolocated Wi-Fi networks

Has data on wireless networks:

- Across the world
- From 2001 to present day

Anonymous and free, registered user access levels



WiGLE.net

The WiGLE.net web site allows those war-walkers, drivers, droners, and boaters to submit their wireless network and geolocation data so that it can be archived and searched. The data contained in this site goes back to 2001 and covers many areas of the world. Wireless networks can be searched for name network name (SSID), unique MAC address of the wireless transmitter (BSSID) location, and other attributes.

While you can use the WiGLE.net site without creating a user account and logging in, it is recommended that analysts sign up for the free accounts so that they can perform more-advanced searches and use the WiGLE API for automated queries using scripts.

One thing that WiGLE.net allows us to do is see what the network names of specific wireless networks were over time. Take the example of two neighbors having an argument. To communicate with each other, they change the SSIDs of their wireless networks to things like "MyNeighborIsAJerk" one day and then "Stop Playing Loud Music" another. When someone war-drives that neighborhood, they could collect those network names and submit them to WiGLE. Since the BSSID remains constant when the SSID changes, we can map a network's name changes over time and may be able to get parts of the wireless argument.

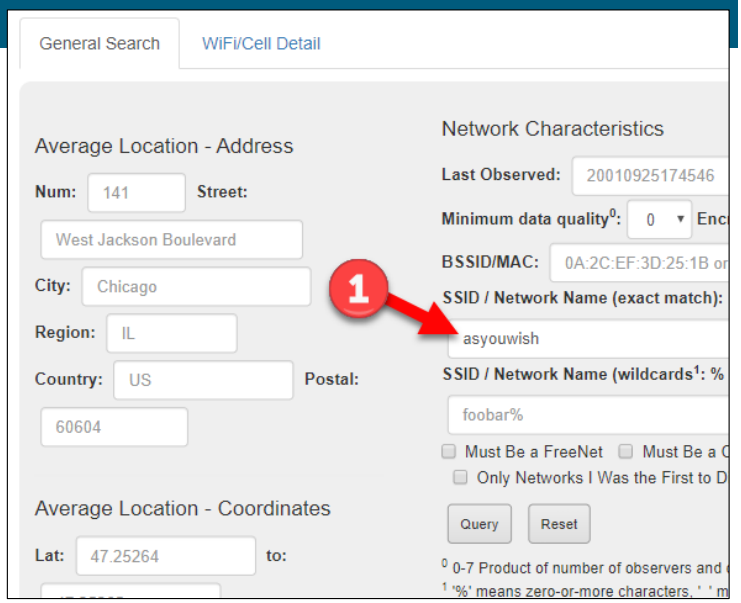
© SANS Institute 2019

WiGLE.net Searches

Highly recommended that you use the WiGLE Advanced search for your queries

It is more flexible and includes wildcard searches

WiGLE searches require a valid (free) login



WiGLE.net Searches

There are both basic and advanced search forms on the WiGLE.net site. For our purposes, the advanced page is usually the best choice, as it allows detailed searching of WiGLE's data via a number of characteristics. As shown in the above image, whether you wish to search a given latitude and longitude, a network name, or an address, the advanced search will help you locate the data you need. We searched for the network name (SSID) "asyouwish" (arrow 1).

To perform these searches on the site, you need to create a free user account.

Image from https://sec487.info/ng, September 30, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

WiGLE Search Results

<< showing records 100 to 161 of 161 >>

Map	Net ID	SSID	Name	Type	First Seen	Most Recently	Crypto	Est. Lat	Est. Long	Channel
map	84:00:2D:DB:CF:C0	asYOUwish		infra	2018-02-22T16:00:00.000Z	2019-03-20T14:00:00.000Z	WPA	40.85277557	-96.87345123	153
map	88:AD:43:55:11:D0	asyouwish		infra	2019-08-26T12:00:00.000Z	2019-08-26T10:00:00.000Z	WPA	42.30733109	-83.39735413	36
map	88:D7:F6:B7:C7:40	AsYouWish		infra	2019-04-13T15:00:00.000Z	2019-05-12T14:00:00.000Z	WPA	40.41051483	-111.86353302	1
map	90:50:CA:B6:D7:D8	AsYouWish		infra	2017-12-19T15:00:00.000Z	2017-12-20T13:00:00.000Z	WPA	45.45804977	-98.48860931	1
map	90:E6:BA:A9:9E:06	asyouwish		infra	2011-07-10T17:00:00.000Z	2012-06-01T19:00:00.000Z		55.64312744	37.36116409	1
map	94:10:3E:04:49:B5	AsYouWish		infra	2017-01-20T20:00:00.000Z	2018-03-12T12:00:00.000Z	WPA	30.29276085	-97.73362732	2

SANS Open Source Intelligence (OSINT) Gathering and Analysis 85

WiGLE Search Results

Executing a search for networks named "asyouwish," we received 161 records that were matches in WiGLE.net. Let's look at the results view to understand what we see.

1. The "map" field will launch a popup window with the estimated location of the device.
2. The Net ID column, commonly called a MAC address or BSSID, is a unique address for the wireless transmitter.
3. Pay attention to the first and most recently seen columns, as some of the data in WiGLE is very old. The "Most Recently" column shows the last date when the device was detected.
4. The estimated latitude and longitude columns give the rough coordinates of where the device might be located.
5. The channel column allows us insight into what wireless spectrum the device uses. Channels 1–14 use 2.4GHz and others use 5GHz.

Finally, let us get additional information about one of these networks. We put a rectangle (6) around the network device "90:50:CA:B6:D7:D8". The next slide will show what the details page looks like when you click the hyperlink on the BSSID.

Image from https://wigle.net, September 30, 2019.

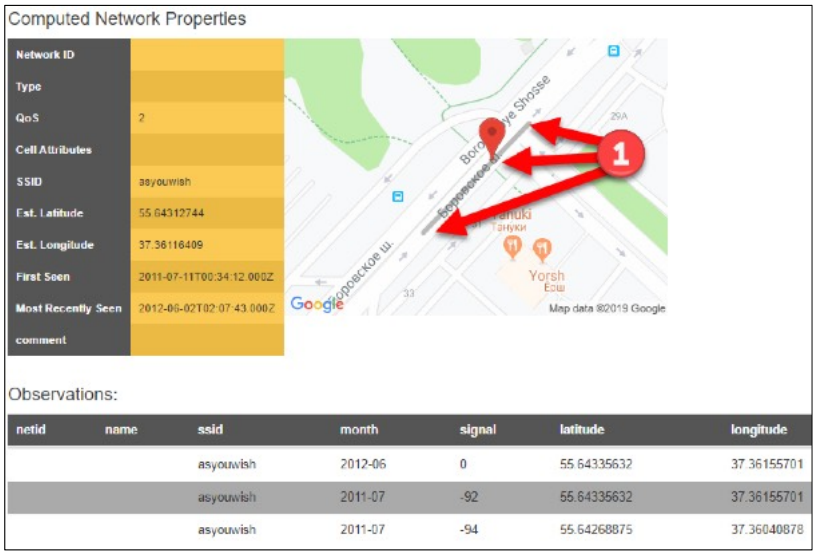
© SANS Institute 2019

WiGLE Detailed Network Data

The SSID is plotted on Google Maps based on signal strengths (1)

The dates the network was discovered in each submission

Clicking on the map launches an interactive map



netid	name	ssid	month	signal	latitude	longitude
		asyouwish	2012-06	0	55.64335632	37.36155701
		asyouwish	2011-07	-92	55.64335632	37.36155701
		asyouwish	2011-07	-94	55.64268875	37.36040878

WiGLE Detailed Network Data

The detailed network screen displays the approximate latitude and longitude of the wireless network transmitter, what dates the network was seen, and possibly other information from the collection. Clicking on the Google Maps image will launch an interactive map at the location of the network so that you can see what other buildings and networks are nearby and turn on satellite view to see more of what that area looks like. Keep in mind that the locations of these networks are just estimations based on the collected signals.

Since cars drive on streets near houses, the collected data may show that a network is in the middle of the road when it is actually in the buildings on one side or the other. The plotted location of the wireless access point, in most cases, will be on the path that the person collecting it drove, biked, or walked. Exceptions to this would be if the site was doing any triangulation of the data points for a given wireless device.

The above image shows the "asyouwish" SSID located in a suburb of Moscow (https://sec487.info/v9).

Image from https://sec487.info/v8, September 30, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Wireless and OSINT Scenarios

In the coming slides, we will investigate several scenarios where understanding Wi-Fi traffic will help your OSINT investigations.

1. Stationary location Wi-Fi AP
2. Stationary location with different networks over time
3. Hotspot phone or MiFi device
4. Same BSSID in multiple places

Wireless and OSINT Scenarios

We are going to examine several reasons why understanding the wireless networks and devices in a location is important for OSINT. The slide above shows the different situations we will be investigating in the coming slides.

© SANS Institute 2019

Stationary Location Wi-Fi AP

Some SSIDs do not change

Find SSID at a location and search to find other places

Can also find details about properties from SSID names

Gray's Point, New South Wales, Australia

Stationary Location Wi-Fi AP

It is common for people and businesses to not change their Wi-Fi SSIDs. WiGLE's network location allows us insight into where there systems are and other details.

For instance, if we look at a certain company's headquarters office building and discover that its wireless network is called "Acme Wireless Network," we can search for other locations that have that SSID and possibly discover additional places where that company has offices even if those offices are not published publicly.

Another issue that arises from SSIDs at certain locations is that sometimes the SSIDs are distinctive and give away information. Take, for example, the image of 3 houses from Gray's Point, New South Wales, Australia, shown in the slide above. Each of these houses has an SSID mapped to it starting with "OPTUSVD3". A quick search of "internet Sydney Australia" on a search engine reveals that the Optus (https://sec487.info/vc) internet provider services homes in this area with broadband internet. These devices also seem to have 6 characters at the end of the SSID that look like they could be unique identifiers or the last portion of BSSIDs. Performing a search in WiGLE for "OPTUSVD3%" revealed over 115,190 results!

Image from https://sec487.info/vb, September 30, 2019.

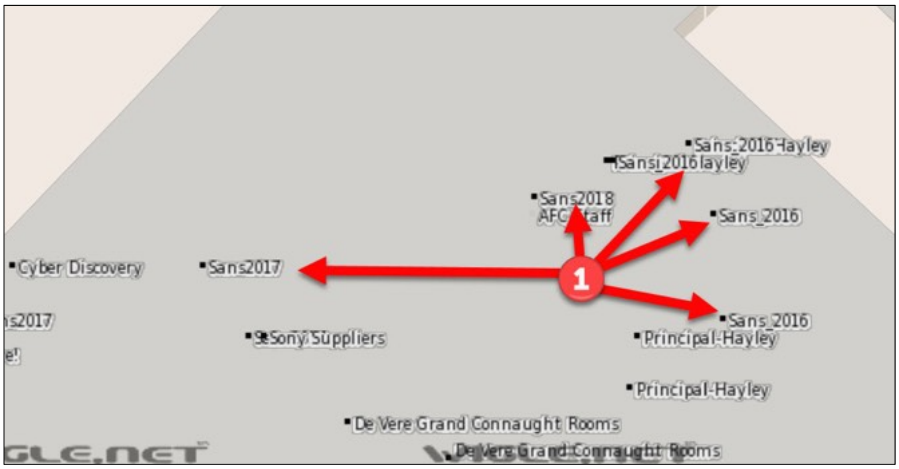
©2019 Micah Hoffman

© SANS Institute 2019

Stationary Location Over Time

We can also watch a location over time and view what SSIDs are shown in and near it

Here we see the SSIDs from SANS events in 2016–2018



De Vere Grand Connaught Rooms, London, UK

Stationary Location Over Time

What if we looked at the SSIDs that have been seen at conference centers and offices over the years? Could we find interesting OSINT implications? Of course we can. In the example above, we extracted the WiGLE.net data from 2016–2018 at the De Vere Grand Connaught Rooms in London, UK (https://sec487.info/ve). This is the conference space that the SANS Institute has used for classes for several years. And we can see this based upon the SSIDs collected in WiGLE (Sans_2016, Sans2017, Sans2018). So from the SSIDs of their networks, we can locate where SANS conducts its trainings.

What if we examined a company's office building over time? Could we see when vendors, bringing their own wireless hotspots and MiFi devices, visit certain locations? Yes. And, by extension, if we know that a MiFi device with the SSID of "Acme Forensic Investigators" was seen at several different locations during certain time periods, could that show that the companies that were being visited by that vendor may have had data breaches? It is possible.

Image from https://sec487.info/ve, September 30, 2019.

© SANS Institute 2019

Wi-Fi Hotspots and MiFi Devices (1)

A target using a MiFi device or using a "hotspot" on a phone could be detected and have data uploaded to WiGLE.net

map	00:04:0E:E1:AF:61	KUBICKI iPhone		2011-01-25T17:00:00.000Z	2011-01-25T11:00:00.000Z
map	00:04:DF:DE:07:27	alex do iphone	infra	2018-08-16T10:00:00.000Z	2018-08-16T09:00:00.000Z
map	00:04:ED:77:4F:6A	Bluberry iPhone	infra	2015-05-16T20:00:00.000Z	2016-08-11T00:00:00.000Z
map	00:04:ED:96:AE:07	Roberta iphone	infra	2015-10-21T19:00:00.000Z	2015-10-21T10:00:00.000Z
map	00:05:4F:87:42:D8	xiaoyu\342\200\231s iPhone	infra	2017-07-25T19:00:00.000Z	2017-07-25T19:00:00.000Z
map	00:05:4F:A4:57:F4	Carlos Iphone	infra	2017-07-01T20:00:00.000Z	2017-07-01T19:00:00.000Z

We can track their devices

Wi-Fi Hotspots and MiFi Devices (1)

We have seen that WiGLE.net collects and archives wireless network names and the BSSIDs of wireless devices. Let's shift our focus from those wireless networks that are in a fixed location and move to those that people carry with them. MiFi devices and personal hotspot tethering is a common practice for users to get their devices connected to the internet. If we were war-driving and detected someone's wireless network for their device, then we may be able to use the data at WiGLE.net to track their device, and them, over time.

The above slide shows a few of the over 1 million results from a WiGLE.net search for networks that had the words "iphone" in them (authenticate to WiGLE.net and then use the link https://sec487.info/of).

©2019 Micah Hoffman

© SANS Institute 2019

Wi-Fi Hotspots and MiFi Devices (2)

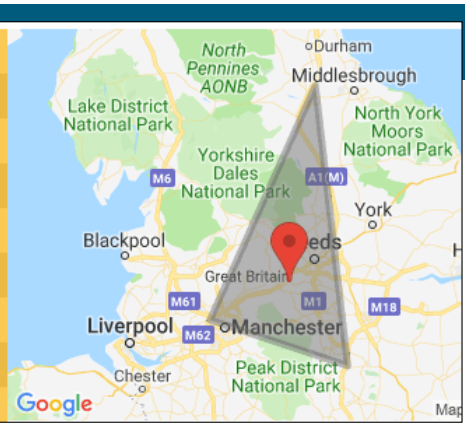
BSSID - `00:06:C6:8F:A9:22`

SSID - `PAUL'S IPHONE`

There are at least 30 locations for the device

And the SSID changes to `Audi_MMI_347`

5	
PAUL'S IPHONE	
53.69110489	
-1.73905265	
2017-09-12T15:06:44.000Z	
2019-03-08T14:42:11.000Z	



ssid	month	signal	latitude	longitude
	2017-09	-75	53.76216507	-1.58720827
	2017-09	-75	53.76258087	-1.586115
PAUL'S IPHONE	2018-04	-79	54.59629822	-1.54191673
Audi_MMI_3477	2018-02	-82	53.67152405	-1.83332837

SANS Open Source Intelligence (OSINT) Gathering and Analysis 91

Wi-Fi Hotspots and MiFi Devices (2)

If we pivot from looking at locations to looking at individual devices, like the one noted above, we can sometimes find devices that have been discovered in WiGLE data many times over the years. The device emitting the "PAUL'S IPHONE" SSID shown above has over 30 locations where it was noted in WiGLE data. The map above plots the major WiGLE data coordinates for this device and shows a table (below the map) with individual coordinates so you can get more precise locations.

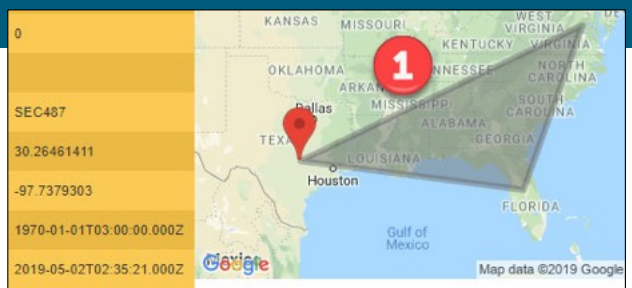
Not only did we potentially find the region where "Paul" travels with his iPhone, but we also found that the SSID for this device switches between "PAUL'S IPHONE" and "Audi_MMI_347". The second SSID seems to be for an Audi vehicle Multi Media Interface (MMI). Perhaps now we are looking for a person named Paul, driving an Audi, in the Leeds area of the United Kingdom?

© SANS Institute 2019

Same BSSID in Multiple Places

Sometimes the same BSSID will be reconfigured to have different SSIDs in different locations

SANS event device is for one class Wi-Fi then is moved to new location and used for another class Wi-Fi



ssid	month	signal	latitude
SEC487	2019-05	-74	30.26455688
SEC487	2019-05	-79	30.26467705
XIPHOS	2018-10	-84	38.27582932
HST777	2018-04	-87	28.46193314

2

Same BSSID in Multiple Places

Another example of using WiGLE data is to track a single device across time and note when it has been repurposed. Perhaps a person moved and changed their Wi-Fi from "DreadPirateRobertsRules" to "InigoMontoyaRules." By tracking the BSSID of the device (which should not change), we can watch monitor different configurations of the same device.

In the example above, we searched for the "SEC487" SSID that is used in the live versions of the SANS SEC487 class. We found a device and clicked on its BSSID to discover that the device has been noted in different parts of the United States (1) and has been using different SSIDs (2).

Image from https://sec487.info/vf, September 30, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

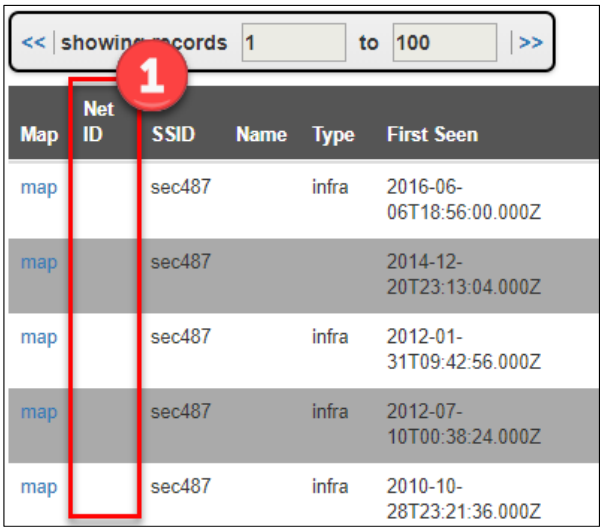
WiGLE Account Problems

When you create a new account on WiGLE, it may be corrupt

It will not give accurate results

Two ways of knowing if your account is effected:

1. The **Net ID** column is empty in results
2. You receive too many results



WiGLE Account Problems

There is no doubt that WiGLE is an amazing resource for accessing wireless from two decades of data collection. One problem that sometimes happens to people creating new accounts is that the accounts somehow get corrupted when they are created. Two methods can show that a user account was not created correctly in WiGLE: First, the Net ID column (box 1 in the slide above) is empty when doing a search. If you see this, know that your account is most likely corrupted and you need to create a new one. Second, the Net ID column not showing data is sometimes coupled with receiving a huge number of results back from your searches. However, the Net ID column being blank is the best indicator we know of to determine if you need to create a new account.

Image from https://sec487.info/va, September 30, 2019.

© SANS Institute 2019

Other Types of Wireless Devices

Wi-Fi is not the only wireless technology

We have RFID, Bluetooth, Zigbee, cellular, and NFC transmitters in our devices

WiGLE.net's API can get you access to Bluetooth data that people submitted to their site

Our mobile devices can be transformed into Bluetooth-sniffing devices with applications

iOS

LightBlue ® Explorer

Android

nRF Connect for Mobile

Other Types of Wireless Devices

While Wi-Fi networks can have major impacts on our OSINT investigations, nowadays we have many other wireless transmitters that we carry with us on a daily basis. These RFID, NFC, cellular, Zigbee, and Bluetooth devices can be used to track us and our targets. Each device has a unique identifier for the wireless radio. While your target moves through a busy marketplace, you can be following behind them, staring at your phone, watching the signal strength of the transmissions from their smart watch. Sounds like something out of a spy novel, yes? Well it is not. Tracking devices and the people holding them is big business for marketing and advertisers and is incredibly helpful for law enforcement to put objects at certain locations.

LightBlue® Explorer for iOS devices (https://sec487.info/vg) and the Android nRF Connect for Mobile (https://sec487.info/vh) are excellent examples of Bluetooth and Bluetooth Low Energy (BLE) sniffing applications for your mobile devices. Download and install them on your devices and then start them scanning. They will discover a huge variety of devices from televisions and Fitbit step counters to smart watches and Bluetooth beacons. They will note the GPS coordinates where your device detected the other systems and will plot the relative signal strength indicator (RSSI), which can be used to roughly determine how close to the Bluetooth transmitter you are.

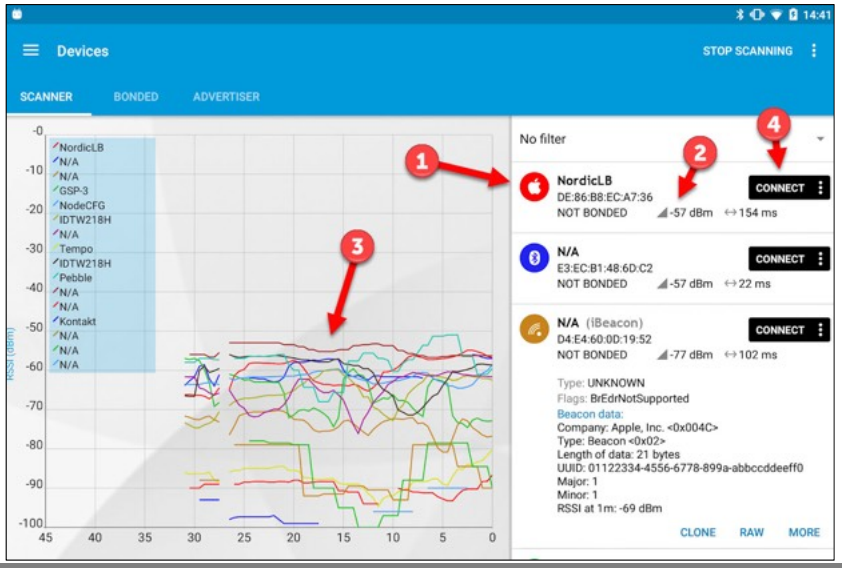
Images from https://sec487.info/vg and https://sec487.info/vh, September 30, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Screenshot nRF Connect for Mobile

1. Type and name of device
2. RSSI tells how far from transmitter
3. Plots "nearness" over time
4. Connect to device (not recommended!)



Screenshot of nRF Connect for Mobile

So you can get a better idea of how simple it is to track devices and the depth of data we receive from these Bluetooth-sniffing applications, let's take a look at the sample image from the nRF Connect for Mobile Android application.

In the image above, we see the device's name and type (arrow 1). There is a MAC address (unique device transmitter ID) next to that. Then we have the RSSI (arrow 2), which can be used to determine distance from the transmitter (as the number gets closer to 0, you move closer to the transmitter). The application plots the RSSI over time in a plot (arrow 3). Finally, if you are allowed to, you can click the "Connect" button (arrow 4) to interact with the device. We do not recommend this because that moves well beyond OSINT.

Images from https://sec487.info/vj, September 30, 2019.

© SANS Institute 2019

DFIR Android Mobile Device Geolocation

Mobile data from cell towers and Wi-Fi networks can help geolocate devices

In DFIR, mobile device data is extracted using software like Cellebrite

Matt Edmonson's Anaximander script processes and maps this data in Google Earth

DFIR Android Mobile Device Geolocation

The OSINT Curious Project blog has a guest post (https://sec487.info/sw) written by SANS Instructor Jeff Lomas (@bleubloodhound) and a link to a recorded webcast on using data dumped from a mobile device to geolocate where that device has been. Using historical wireless and cellular connection data collected on the devices and extracted using DFIR software like Cellebrite, Jeff shows methods of finding where the device was.

Matt Edmonson's Anaximander script (https://sec487.info/sx) takes data dumped from Cellebrite and looks it up against cell tower data and wireless networks. It then creates a Google Earth KML file that plots the locations of those towers. Opening the file in Google Earth can show dates and times when a device was near a certain location (bottom slide image).

Images from https://sec487.info/sw, September 25, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Wireless OSINT"

This page intentionally left blank.

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. DNS
4. Findings Online Devices
5. Wireless Networks
6. **Recon Tool Suites and Frameworks**
7. Government Data
8. Researching Companies

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Reconnaissance Tools and Frameworks

There are a variety of excellent open source collections of modules for performing IP, domain, and people OSINT

These save us time by reusing the data they retrieve in one module in other modules

Allow for us to extend them through the creation of custom modules

Reconnaissance Tools and Frameworks

There are several tool suites or frameworks that OSINT analysts use to rapidly gather information on their targets. Most of these tools have modules that do the work of making web, DNS, and other network requests to a variety of sources locally and on the internet. Consider them a terrific starting point to your investigation. Many of them will also use the info they gathered in other modules. A good example of this data reuse is if one module finds usernames or email addresses, other modules may be able to look up additional data about those accounts.

If you can write some Python or Ruby, then you can create your own modules that do what you want them to do. You can use them on your system or submit them to the framework owners so that others can use them in their work.

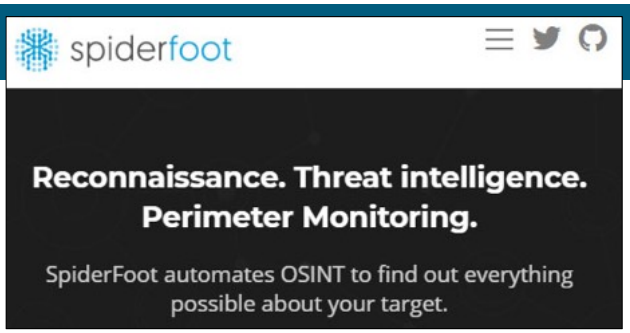
© SANS Institute 2019

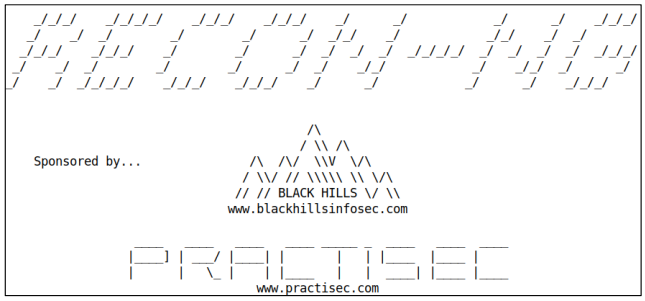
Recon Frameworks

We will briefly look at a few frameworks:

- SpiderFoot
- Recon-ng

Many of the features and modules overlap, but there are differences to explore





Recon Frameworks

While this is by no means an exhaustive list of OSINT frameworks, the SpiderFoot and Recon-ng software suites provide analysts a variety of creative, reliable, and free methods to retrieve information about their targets. The two frameworks mentioned here are stable and produce excellent results. While there are many similarities between these frameworks, there are some differences in how they work, the modules that they contain, and how you get information out of them.

References:

- SpiderFoot – https://spiderfoot.net/
- Recon-ng – http://recon-ng.com

©2019 Micah Hoffman

© SANS Institute 2019

Reconnaissance Tools and APIs

There are two main methods of retrieving data from web applications using scripts: use an **API** and **scrape data** from human-readable pages

Most sites prefer scripts use the API method

Some APIs are free, and others **cost thousands of dollars**

The recon tools in the section ahead all can use APIs to gather more data

You are responsible for registering and retrieving the API keys from the destination web sites

Then you enter those into the tools

Reconnaissance Tools and APIs

When humans surf web pages using their browsers, they view human-readable content that might have images and JavaScript files to make the web-browsing experience a pleasant one. When we use scripts to retrieve data from web sites, the sites usually prefer that we use an API (Application Programming Interface) that provides script-formatted (XML, JSON, etc.) data.

To get an API key from a web site, you may need to register for an account and/or pay a monthly or yearly fee. It all depends on the web site you are using. Most of the recon frameworks have pages in their documentation showing how to get the API keys from various sites. Once you register for the API keys, you enter them into the recon tool and can use that module.

© SANS Institute 2019

SpiderFoot Overview

Name	SpiderFoot
URL	https://spiderfoot.net
Technology	Python
Owner	Steve Micallef (@binarypool)
Number of Modules	160+
Use	Command line, web GUI
Versions	Open source, HX version (free and paid)

SpiderFoot Overview

Our first framework, SpiderFoot, is owned and maintained by Steve Micallef (@binarypool). It is found on the https://spiderfoot.net web site, where Steve has versions of SpiderFoot and installers for various operating systems. This Python-based tool has over 160 modules and can be run from a terminal/command prompt or a web browser.

SpiderFoot has two versions: open source and HX. The SpiderFoot HX (https://sec487.info/nh) is hosted in the cloud and has an API, a more-robust database, and enhanced visualizations.

Image from https://sec487.info/uo, September 27, 2019

©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot: Open Source vs HX

Feature	Open Source	SpiderFoot HX
Targets	Single	*Multiple
Traffic Location	Installed system	*SpiderFoot cloud
Scan Duration	*Unlimited	Depends on subscription level
Scan Retention	*Unlimited	Depends on subscription level
Data Export	Yes	Yes
Graphing	Basic	*Four different graphing styles
Risk ID	No	*Yes
Scheduled Scans	Manual using CLI	Depends on subscription level
Maintenance	You patch/upgrade	SpiderFoot team maintains

SpiderFoot: Open Source vs HX

The table above highlights some of the bigger differences between the free, open source version of SpiderFoot and the cloud, HX version. We colored and put an asterisk (*) in front of settings that we believed to be better on one platform or the other. Below, we will clarify some of the above information.

- **Targets** -The open source version of SpiderFoot allows for a single target to be submitted to scan. This can be a domain, IP address, hostname, subnet, ASN, email, phone, or human name. The HX version allows for multiple targets and can import them from a file.
- **Traffic Location** - When you run a SpiderFoot scan from the version of SpiderFoot you download and install on your own system, that is where the scanning traffic and web requests come from. If you couple this with a VPN to hide your IP address, there is an increased level of anonymization. Alternatively, SpiderFoot HX runs from cloud servers and is less attributable to you since the SpiderFoot project owns or rents those servers.
- **Risk ID** - SpiderFoot HX tags higher risk findings with special icons in a specific column of the output. The open source version of SpiderFoot does not. Tagging high-risk data in this manner allows it to be quickly detected and investigated.
- **Maintenance** - A final reason that you might choose one platform over the other is who maintains the system. In the open source version of SpiderFoot, you have to install it on your computer or server and, therefore, must maintain, patch, and upgrade that system to keep it "cyber healthy." When you use the HX version, you transfer that responsibility to the SpiderFoot team, lessening your administrative burden.

© SANS Institute 2019

SpiderFoot HX Unique Features

Cloud Hosted
Accessible any time, anywhere requiring no maintenance effort or setup costs. Always have the latest, most stable modules.

Detect Changes
Automatically scan targets according to a daily, weekly or monthly schedule and get notified when changes in their OSINT footprint are detected.

Scale & Speed
Get results much faster than the open source version, and handle over one million data elements easily thanks to a completely overhauled and distributed architecture.

Scan Comparisons
Compare any two scans to see the commonalities and differences between them.

API
Further automate your OSINT by integrating with the fully documented RESTful API, including examples to get you up-and-running quickly.

Team Collaboration
Create user accounts enabling your team to collaborate, also supporting role-based access control.

Reporting & Drill-Down
Automatically identify risky data elements and see the data in ways that are more insightful.

Multi-Target Scanning
A target rarely consists of just one domain name or subnet; with SpiderFoot HX you can specify multiple targets in a single scan. Save time and easily see all the inter-relationships.

Support
E-mail support to help you get the most value out of SpiderFoot HX.

Correlations
Over 30 built-in correlation rules that help you navigate to the most important and potentially interesting data from your scan.

Visualisations
Many new visualisations to help you gain more insight into your scan results and interactive node graphs to explore the data.

Revamped Interface
SpiderFoot HX builds upon the same principles of the open source interface by providing a snappy and easy-to-use interface for all these advanced features.

SpiderFoot HX Unique Features

Aside from the differences between the open source and the HX versions of SpiderFoot raised on an earlier slide, the HX version has many more advantages, as shown above. Some of the features like scan comparisons, multi-target scanning, and correlations enable individuals and organizations to examine the interplay between data points more effectively. Additionally, these and other features like SpiderFoot monitoring assist people looking to track changes over time.

Image from https://spiderfoot.net/hx/, September 29, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot HX Subscription Levels

SANS Promo code for 15% off "spotlight"

	Hobby	Basic	Standard	Professional	Enterprise
	Free	€19 / mo	€59 / mo	€179 / mo	€599 / mo
Unique targets per scan	1	4	256	1,024	4,096
Scans per month	3	5	10	50	200
User accounts	1	1	2	5	Unlimited
Scan duration limit	15 minutes	4 hours	24 hours	72 hours	120 hours
Scan data retention	7 days	30 days	3 months	12 months	24 months
Customer email support	✗	✓	✓	✓	✓
Monitoring & Change Notifications	✗	✗	✓	✓	✓
SpiderFoot HX API	✗	✗	✓	✓	✓

SpiderFoot HX Subscription Levels

The HX version of SpiderFoot has a variety of subscription levels moving, from the free "Hobby" level through to the Enterprise level that costs thousands of euros per year. The Hobby level should be appropriate for those wanting to try out the web tool and experiment with its features and functions. As you pay more, you get access to more features, such as an API, and longer durations for the scans to be held and to run. This translates into you being able to run scans that are both deeper and broader (more targets).

Want to try out HX? Visit the https://sec487.info/nh page and enter a valid email address, and the SpiderFoot staff will create a SpiderFoot HX instance for you and send you a link with how to access it.

Image from https://spiderfoot.net/hx/, September 29, 2019.

© SANS Institute 2019

SpiderFoot Open Source in Action

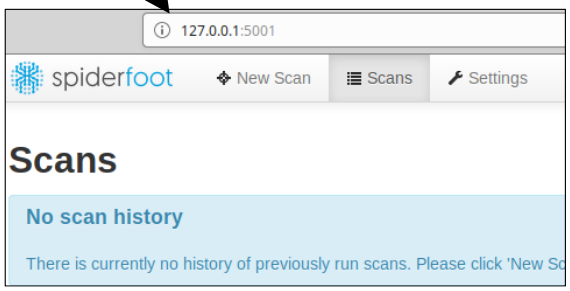
```
Starting web server at http://127.0.0.1:5001 ...


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

```
$ python sfcli.py

                Open Source Intelligence Automation.
                by Steve Micallef | @binarypool

[*] Version 3.0.
[*] Server http://127.0.0.1:5001 responding.
[*] Loaded previous command history.
[*] Type 'help' or '?'.
sf>
```



SpiderFoot Open Source in Action

Let's examine how the open source version of SpiderFoot that is locally installed works.

Depending on what operating system you are running the SpiderFoot server from, you may launch it by simply double-clicking a file (on Windows you use the sf.exe file to start the server) or by manually running the sf.py script. The documentation at https://sec487.info/4a is quite detailed in case you need more verbose notes.

The server will then "listen" for requests and commands on a local port, which, by default, is on TCP 5001. You can then use the command-line method of accessing the server by running the `sfcli.py` Python client in another terminal (shown in the left image above). Alternatively, you can access the SpiderFoot server via a web browser by visiting the http://127.0.0.1:5001 page (or whatever IP address and port you have configured SpiderFoot to run on), which is shown on the right in the slide.

©2019 Micah Hoffman

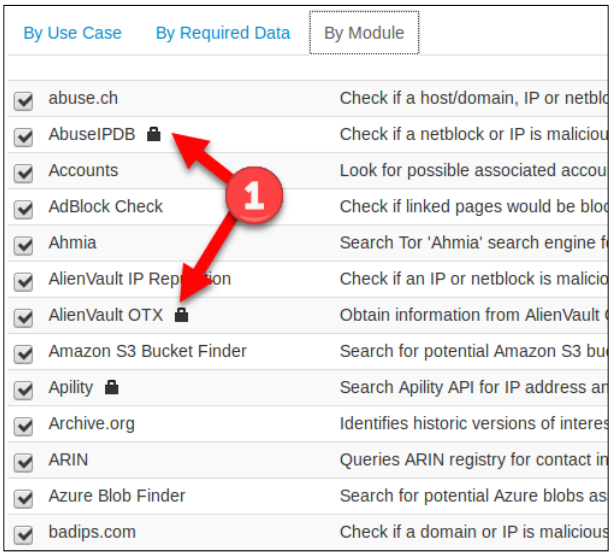
© SANS Institute 2019

SpiderFoot Open Source Modules

The workhorses of SpiderFoot are its modules

They send requests to a variety of web resources and then store the results in the SpiderFoot DB

Some modules require an API key indicated with a lock icon (arrows at 1)



SpiderFoot Open Source Modules

Once the server is running and you have connectivity to it via a CLI (command-line interface) or web browser, you can start configuring the overall system and eventually start scans. As mentioned previously, SpiderFoot has modules that users can leverage to retrieve data for their work. While we will examine some of the useful modules in the coming slides, it is important to note that many of the modules may need additional configuration before they can be used. The lock icon (arrows labeled “1” in the image) indicates that the module requires an API key from the web site that it gathers information from before it can be used. It is the user's responsibility to apply for and install the API keys in SpiderFoot prior to using those modules that require them. In the web GUI, this data and other configuration settings can be accessed in the "Settings" page of the application (located by default at http://127.0.0.1:5001/opts). If you choose to use a module that requires an API key that is not configured in SpiderFoot, the module will be skipped.

© SANS Institute 2019

SpiderFoot Interesting OSINT Modules

People:

- **Accounts** - Checks account names across over 150 different web sites
- **Social Networks** - "Identify presence on social media networks such as LinkedIn, Twitter and others."
- **Instagram** and **MySpace** - Gather information from those social media platforms

Domains/IPs:

- **Censys** and **Shodan** - Search the Censys.io and shodan.io web sites (require API keys)
- **Archive.org** - "Identifies historic versions of interesting files/pages from the Wayback Machine."
- **Whois** - "Perform a WHOIS look-up on domain names and owned netblocks"

SpiderFoot Interesting OSINT Modules

The SpiderFoot framework is a powerhouse for OSINT recon. While we don't have space to describe all its modules, rest assured that if you are trying to find something about a person, host, or domain, SpiderFoot can help. Above are a few of the modules to give you an idea of how it can assist your efforts.

Reference:

All quotes are from https://sec487.info/z3, November 10, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

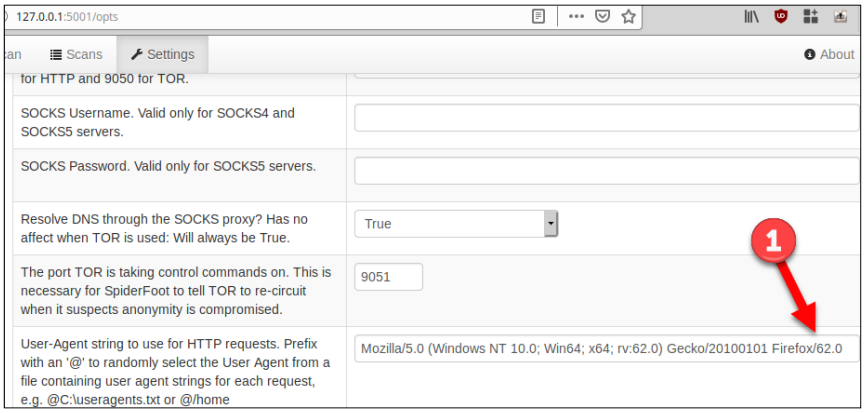
SpiderFoot Settings

Examine settings in SpiderFoot

Sometimes they are inappropriate

Examples:

- HTTP timeout of 5 seconds
- User agent string



SpiderFoot Settings

Within both versions of SpiderFoot (open source and HX), you should examine the settings. They determine how the application scans and processes your results. Some of the settings are per module and allow you to customize each module's functioning and enter an API key for it if needed. These settings are found and configured in https://sec487.info/up in the web interface for the open source version of SpiderFoot and via the "Configure" menu at the top of the main window in the HX version.

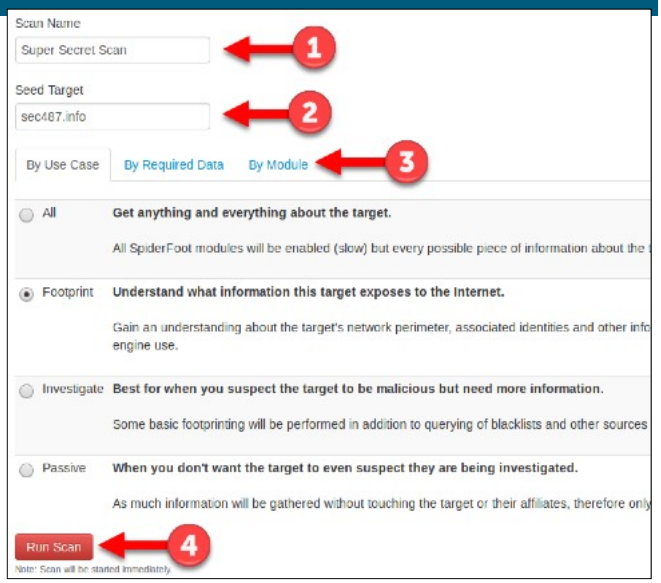
Before running any scans, look for old or inappropriate settings in these settings areas so that you do not waste time running scans that do not do what you want. An excellent example of why you need to look in these settings is that the default timeout for web requests is set to 5 seconds in the open source SpiderFoot. If you run a scan from your local system with this value, SpiderFoot will give up on responses that are not returned within 5 seconds. While this value may be appropriate for high-speed internet connections, we highly suggest you increase this value to 30 seconds or more for high-latency internet connections. We need to pay particular attention to this parameter when scanning over a VPN, as the VPN can introduce delays.

Another example of why you need to change default values can be found in the default SpiderFoot user agent string (shown at arrow 1 above). Notice that the string ends with "Firefox/62.0"? According to https://sec487.info/uq, Firefox 62.0 was released September 5, 2018 and, while it is possible that someone has not updated their web browser since then, it is unlikely since most web browsers auto-update now. Scanning using this user agent string could allow observant system administrators to infer that you might be using SpiderFoot to visit their web sites.

© SANS Institute 2019

Starting a SpiderFoot Open Source Scan

1. Enter scan name
2. Enter seed target data
3. Choose the modules or "use case" you wish to use
4. Click "Run Scan"



Starting a SpiderFoot Open Source Scan

To start a SpiderFoot scan of a target within the web GUI, click the "New Scan" button and fill out the "Scan Name" field (arrow 1). Next, you will need to enter the starting information that you want researched (arrow 2). Then, choose the profile (collections of preselected modules) or individual modules that you wish to run (arrow 3). Once the scan is configured as you wish, click the red "Run Scan" button at the bottom of the page (arrow 4), and the tool will begin its work.

In this example, we ran the SpiderFoot application via the web GUI and entered "sec487.info" as the seed target data. We then selected the "Footprint" profile of modules.

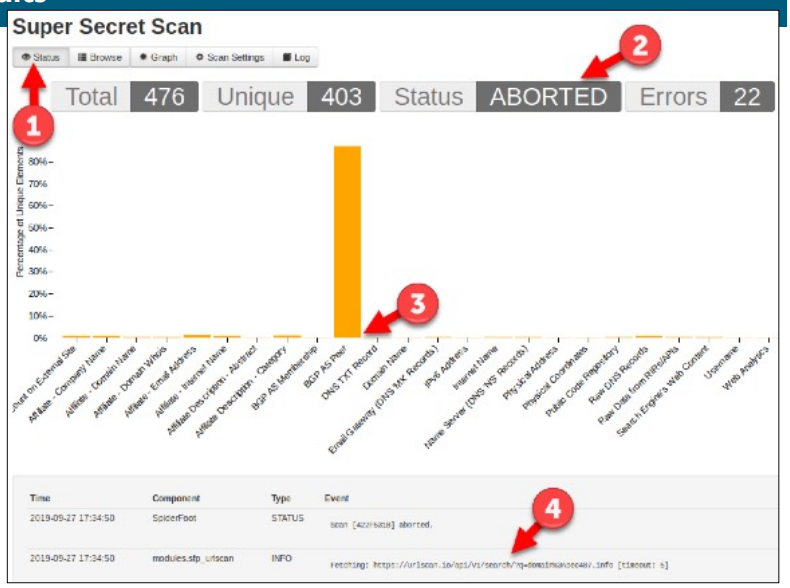
©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot Open Source Results

As the scan runs, analysts can view the progress in the "Status" area (1)

Results are posted to the GUI incrementally, so you can begin analyzing data before the scan completes



SpiderFoot Open Source Results

SpiderFoot provides the analyst ample information about what it is doing and what it found while the scan continues to run. This allows analysts to begin working with the retrieved data before the completion of the scan. As SpiderFoot retrieves more data, it will continue to populate the status page with a graph (shown above) as well as a log of what occurred during the scan (below the graph).

Arrow 1 points to what data we are viewing (Status). Arrow 2 highlights the current status of the scan. In this case, we aborted or stopped the scan prior to it completing. The bar graph in the area of arrow 3 shows what data sets returned the most information. We can also view what SpiderFoot is currently doing or has done by viewing the log information where arrow 4 is located. This is helpful when the scan appears to have hung or when you are getting many errors. The URLs shown in this area allow you to visit the pages manually to better understand what SpiderFoot is seeing.

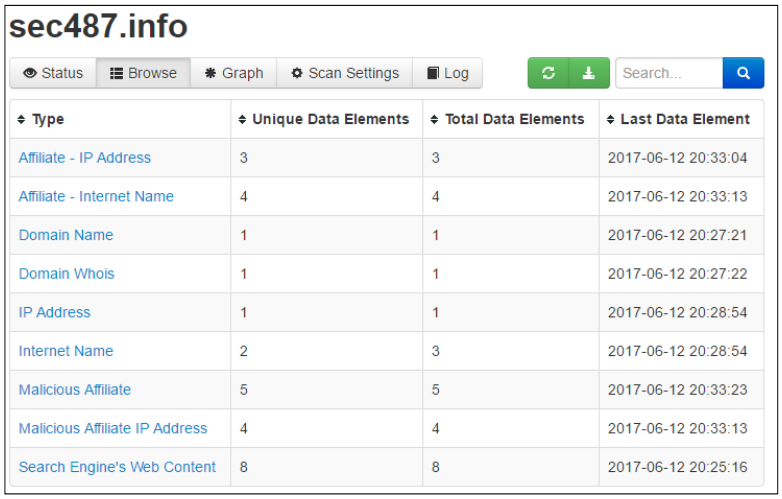
© SANS Institute 2019

SpiderFoot Open Source Browsing Data

Results can be viewed while the scan is running

The "Browse" button shows results discovered in a table format

Clicking the hyperlinks allows the analyst to see details



Type	Unique Data Elements	Total Data Elements	Last Data Element
Affiliate - IP Address	3	3	2017-06-12 20:33:04
Affiliate - Internet Name	4	4	2017-06-12 20:33:13
Domain Name	1	1	2017-06-12 20:27:21
Domain Whois	1	1	2017-06-12 20:27:22
IP Address	1	1	2017-06-12 20:28:54
Internet Name	2	3	2017-06-12 20:28:54
Malicious Affiliate	5	5	2017-06-12 20:33:23
Malicious Affiliate IP Address	4	4	2017-06-12 20:33:13
Search Engine's Web Content	8	8	2017-06-12 20:25:16

SpiderFoot Open Source Browsing Data

During and after the scan runs, analysts can view the data SpiderFoot accumulated in the "Browse" tab. The user is presented with different categories of information and can click the hyperlinks to see the detailed content within each. This data can also be exported in a CSV (Comma Separated Value) file so that it can be used in other systems.

©2019 Micah Hoffman

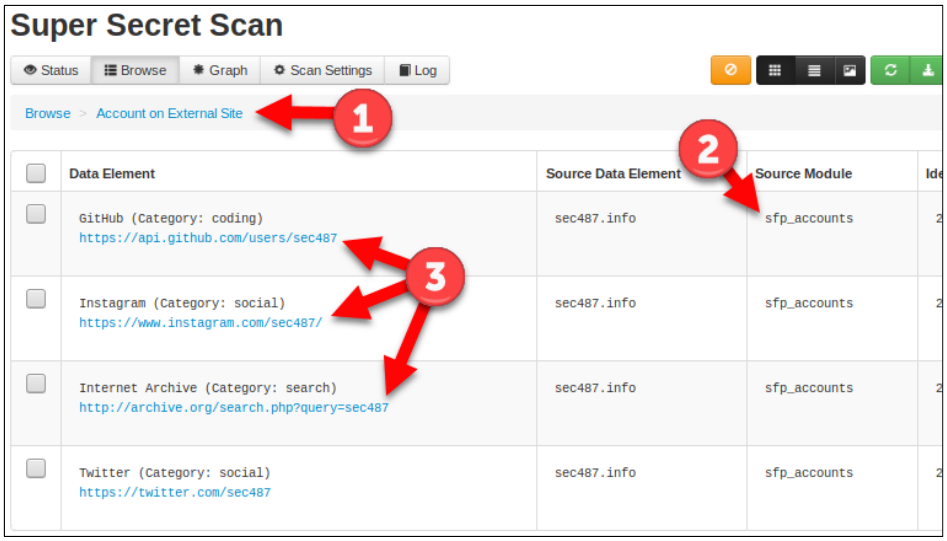
© SANS Institute 2019

SpiderFoot Open Source Detailed Data

Getting details on data type (1)

Can see what module data came from (2)

View data (3)



SpiderFoot Open Source Detailed Data

So what did SpiderFoot open source actually retrieve about the sec487.info domain? In the "Browse" tab, we clicked on the "Account on External Site" link to bring us to this page of data (arrow 1). We see that this content came from running the sfp_accounts module (arrow 2) and retrieved four locations where the sec487 string was potentially used as a username (arrows at 3).

Understanding what module the data came from helps us to work backward and false-positive-check the data. It may require you to review how a specific module collects its data before you can move forward, but it points you in the direction of the source of the content.

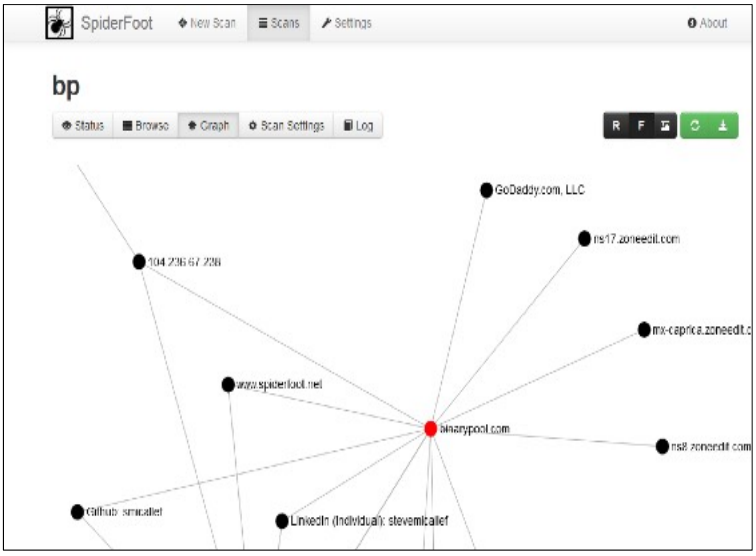
Each data element shown (arrows at “3”) may be in a variety of formats, depending on the module output. In this case, the sfp_accounts module outputs the site name, category, and the URL to the account. This may look similar to the WhatsMyName project discussed previously in the class because this module uses the JSON data file from that project. Other modules may show text DNS results, JSON, or HTML output.

© SANS Institute 2019

SpiderFoot Open Source Graphed Data

Each significant data point (IP, domain, email) can be shown connected to other points using the "Graph" view

Data can be exported to be used in external visualization tools such as Gephi and Casefile



SpiderFoot Open Source Graphed Data

Another distinguishing feature in SpiderFoot is that it will graphically link the different piece of data that it retrieved and display them in the "Graph" view. If your target is a domain and SpiderFoot discovers email addresses, IP addresses, and more, it will connect them and allow you to interact with each data point. Depending upon the target and the wealth of data on the internet about it, this data could be quite numerous. SpiderFoot has data-exporting features that allow the data points and their connections to be post-analyzed in other applications such as Gephi (https://gephi.org/) and Casefile (https://sec487.info/4b).

Image from https://sec487.info/ni, January 3, 2019.

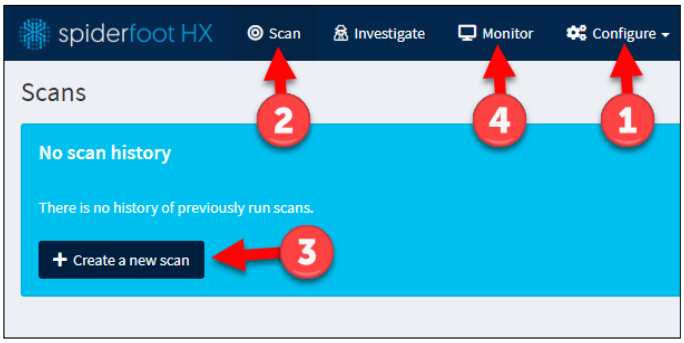
©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot HX

Sign up for free account using valid email

Start at the Hobby level with 15-minute max scans and 1 scan at a time



Once logged in, you can configure settings (1), look at old scans (2), or create a new one (3)

Depending on your subscription level, you can use other features like "Monitor" to run scheduled scans

SpiderFoot HX

The polished, cloud-based HX version of SpiderFoot can be a good alternative to using the open source version. To sign up, you must submit a valid email to the web form so that the SpiderFoot team can email you the server name and your account details once they have set up your account. Sometimes the process to receive this login information is quick (minutes to hours), and sometimes it can take a day or two to receive.

The free tier of HX is called the Hobby subscription level and has several limitations (already discussed). Once you log in to the SpiderFoot HX application, you will be presented with a screen looking similar to the one shown above. We always suggest you examine and change the settings (arrow 1) before running the application. Arrow 2 points where you can view previously run scans, and arrow 3 shows the place to click to create a new one.

Paid subscription levels to SpiderFoot HX provide access to additional features such as the scheduled scans, called Monitor (arrow 4) on the site.

© SANS Institute 2019

Starting a SpiderFoot HX Scan

1. Choose profile or modules
2. Enter scan name
3. Enter seed target data
4. Or import targets from a file
5. Click "Run Scan Now"

Starting a SpiderFoot HX Scan

Shown above is the process to start a SpiderFoot HX scan. Since our example uses the Hobby subscription level, we can only submit a single target for scanning. Higher levels allow for a comma or newline separated list of targets to be pasted into the field or uploaded using the Import button (arrow 4).

Note that the options for which modules we run are different here in HX than in the open source version of SpiderFoot. In the HX version we have three choices: all modules, passive modules, and choose your own modules. The open source version of SpiderFoot had two additional levels that you can re-create by selecting the modules you wish.

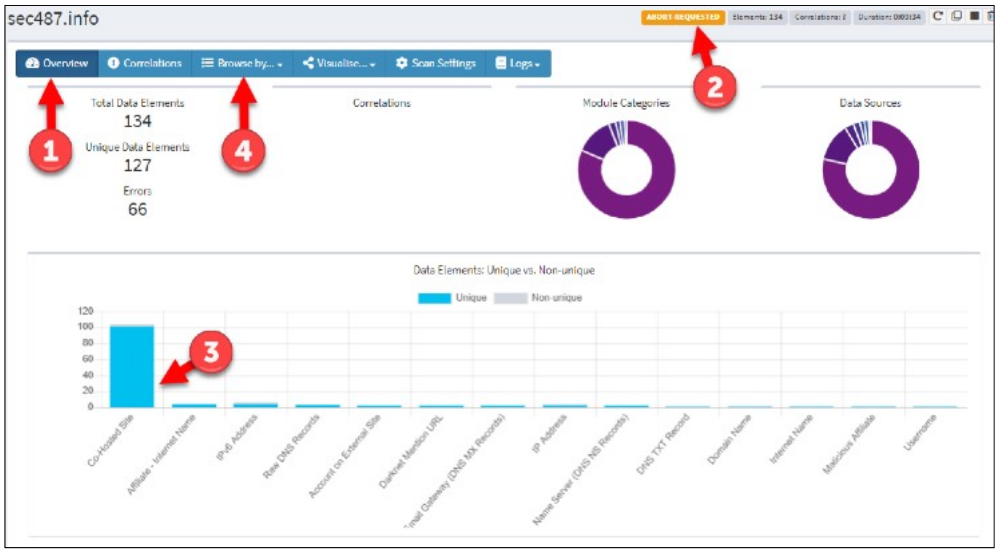
In this example, we ran the SpiderFoot HX application (at the Hobby level), chose to run the Passive modules, and entered "sec487.info" as both the name of the scan and the seed target data.

©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot HX Results

The Overview tab shows overall status (2) and which types of data have returned the most results (3)



SpiderFoot HX Results

Similar to the open source version of SpiderFoot, this "Overview" display shows what the scan is doing and is refreshed at 5-second intervals. We can view the status of our scan (arrow 2) and see which types of data have been returned. The names of the tabs are different from the open source version of SpiderFoot (Overview instead of Status, etc.) but the functions are similar.

For our work, this screen is not as exciting as the "Browse by…" (arrow 4). Let's examine this tab, as that is where we can see the actual data collected.

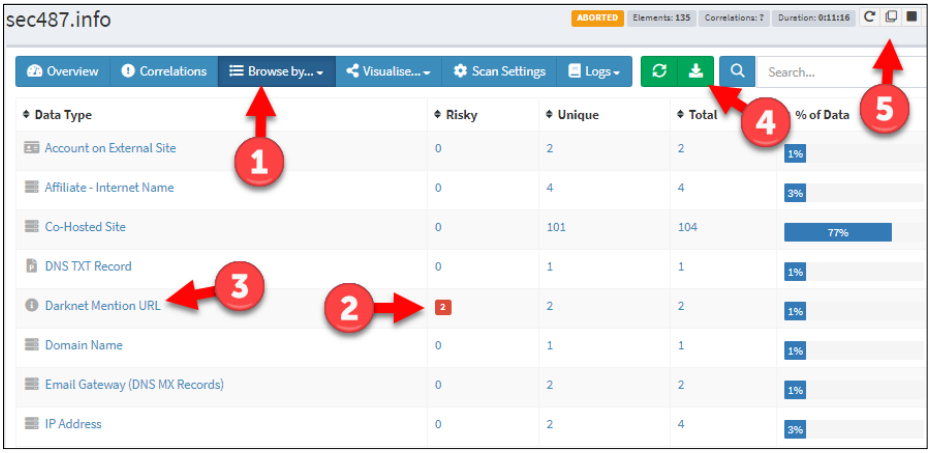
© SANS Institute 2019

SpiderFoot HX Browsing Data

Shows "Risky" results (2)

Examine the data by clicking the Data Type (3)

Allows for exporting data (4)



SpiderFoot HX Browsing Data

In the "Browse by…" tab (arrow 1), we see data a bit differently from the open source SpiderFoot. Most noticeably is the addition of the "Risky" column for results presenting a high risk (arrow 2). In the scan we ran we can see that there were two mentions of sec487.info on the Darknet. To investigate this, we select the category header "Darknet Mention URL" (arrow 3) to see details.

Other features found on this and other SpiderFoot HX screens include the export button (arrow 4), the status of the scan, and controls to stop, clone, and delete the scan (arrow 5).

©2019 Micah Hoffman

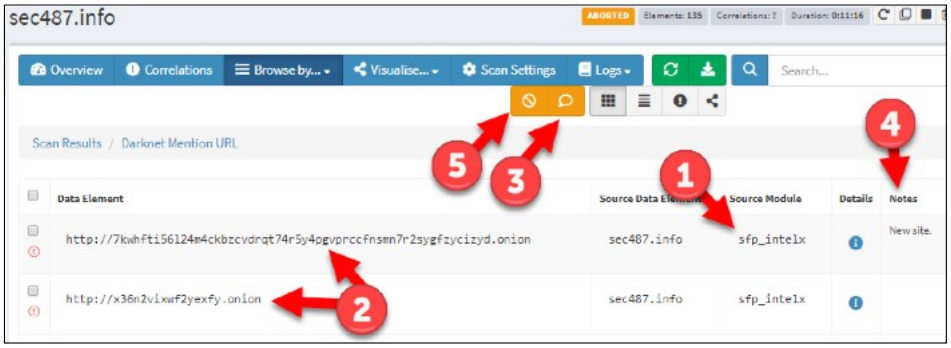
© SANS Institute 2019

SpiderFoot HX Detailed Data

See the module (1) and data (2)

Can add notes (3 & 4)

In both versions, can mark false positives (5)



SpiderFoot HX Detailed Data

The results shown in HX are in the same format as shown in the open source version of this tool. We still see the module that was used to acquire the data (arrow 1) and the detailed content retrieved (arrows at 2). In HX, we can write notes about entries using the icon at arrow 3. You can see we entered a note of "New site." for the first record (arrow 4).

Both HX and open source versions of SpiderFoot allow the user to select entries and mark them as false positives (arrow 5) so that they are effectively removed from results and data analysis.

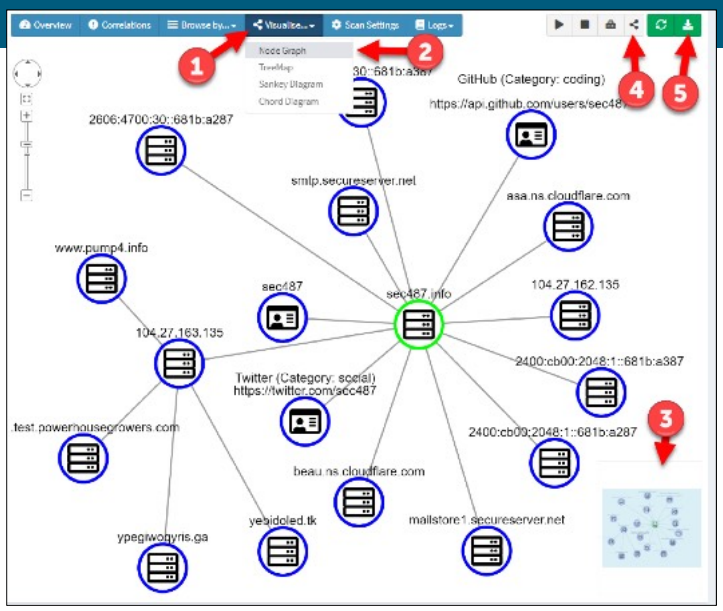
© SANS Institute 2019

SpiderFoot HX Graphed Data

SpiderFoot HX has several different visualization types (2)

In this Node Graph, we see connectedness of the data

You can change the format (4) and export it if desired (5)



SpiderFoot HX Graphed Data

Another distinction between the open source and HX versions of SpiderFoot is the visualizations of the scan results. In the image above, we selected a "Node Graph" type of visualization, which shows links between data objects. The interface has a global view (arrow 3), which can be helpful for very large datasets. There are also buttons to change the layout of this graph (arrow 4) and export the data to Gephi (5).

©2019 Micah Hoffman

© SANS Institute 2019

SpiderFoot Pros and Cons

Pros:

- HX version has excellent features
- Easy start up (especially for HX)
- Multiple methods of accessing data
- Many modules (create your own!)
- Output data to use in other tools
- Visualizations!

Cons:

- Many modules require APIs to retrieve data
- Some modules have hard-coded user agents with the "SpiderFoot" string

SpiderFoot Pros and Cons
Some of the pros and cons (benefits and challenges) with the SpiderFoot tool are presented here. Keep in mind that you need to evaluate the tool for your needs, your team, and your method of working, so these pros and cons may not be as important to your work.

Pros:

- Easy start up - Downloading and running SpiderFoot is simple to do, or you can create a free account and use the HX version.
- Multiple methods of accessing data - With the open source CLI, web GUIs, and the cloud HX version, SpiderFoot has a variety of useful methods of accessing the tool, running scans, and using the data collected.
- Many modules - SpiderFoot most likely has scripts that will help you in your OSINT research. If it doesn't, you can create your own modules and, if you want, share them with the community by submitting them to the SpiderFoot project page at https://sec487.info/m.
- Output data to other tools - Many of the SpiderFoot web pages have options to export the data into formats that can be used in other tools.
- Visualizations - For a rapid understanding of what results data is connected to which other pieces of data, the simple visualizations in SpiderFoot can be helpful.

Cons:

- As with other frameworks, many of SpiderFoot's modules require API keys to work. Some of these keys may require accounts to be created and possibly may cost money to obtain.
- Analysts should know that the user agent string that is set in the global settings of SpiderFoot is not used on about 17 modules. These modules have their user agent strings hard-coded to the string "SpiderFoot". When using these modules, SpiderFoot will send the "SpiderFoot" string and therefore the site will know that you are running a tool and gathering data about a certain person or system. These modules are Python, and analysts have the ability to change those modules' values by hand.

© SANS Institute 2019

SpiderFoot OPSEC Gotcha

While SpiderFoot has a field in the Global Settings to set the user agent string that it uses when making requests, it does not always use this

Some of the modules have hard-coded user agent strings that are set to "SpiderFoot"

If you are concerned about letting a site know that you are using the SpiderFoot tool, change those entries

It can be done in the local code or by sending the requests through a proxy and having that perform the rewriting

What about HX?

SpiderFoot OPSEC Gotcha

Looking through the code for SpiderFoot open source, an observant person would note that several modules have a hard-coded user agent string of "SpiderFoot". When the scripts make web calls to remote systems to gather data, that string would be presented to the site instead of the user agent string that is customizable in the Global Settings. We asked about why some modules had hard-coded user agents and some did not, Steve explained that some of the destination sites that SpiderFoot pulls data from requested that the tool use a specific user agent. You cannot make these changes using SpiderFoot HX, but it may not matter since you are using a server you don't own for the scans.

If you are interested in modifying this in the code, download it, change into the spiderfoot directory, then run `grep -r 'useragent="Spi' *| sed s/" "//g|sort -u`. You should see output similar to what's shown in the image below. The files that would need to be altered are on the left. Another alternative to change it would be to send all the SpiderFoot traffic through a web proxy and have that application alter the strings.

```
student@sec487 (10:50:23) :/opt/tools/spiderfoot$ grep -r 'useragent="Spi' *| sed s/" "//g|sort -u
modules/sfp_accounts.py:data=self.sf.fetchUrl(url,useragent="SpiderFoot")
modules/sfp_alienvault.py:useragent="SpiderFoot",headers=headers)
modules/sfp_azureblobstorage.py:res=self.sf.fetchUrl(url,timeout=10,useragent="SpiderFoot",noLog=True)
modules/sfp_binaryedge.py:useragent="SpiderFoot",headers=headers)
modules/sfp_builtwith.py:useragent="SpiderFoot")
modules/sfp_censys.py:useragent="SpiderFoot",
modules/sfp_circllu.py:useragent="SpiderFoot",headers=headers)
modules/sfp_clearbit.py:useragent="SpiderFoot",headers=headers)
modules/sfp_commoncrawl.py:useragent="SpiderFoot")
modules/sfp_digitaloceanspace.py:res=self.sf.fetchUrl(url,timeout=10,useragent="SpiderFoot",noLog=True)
modules/sfp_duckduckgo.py:useragent="SpiderFoot")
modules/sfp_fraudguard.py:useragent="SpiderFoot",headers=headers)
modules/sfp_fullcontact.py:useragent="SpiderFoot",postData=json.dumps(data),
modules/sfp_greynoise.py:useragent="SpiderFoot",headers=header)
modules/sfp_haveibeenpwned.py:useragent="SpiderFoot",headers=hdrs)
modules/sfp_hunter.py:useragent="SpiderFoot")
modules/sfp_iknowwhatyoudownload.py:useragent="SpiderFoot")
modules/sfp_psbdmp.py:useragent="SpiderFoot")
modules/sfp_pulsedive.py:res=self.sf.fetchUrl(url,timeout=30,useragent="SpiderFoot")
modules/sfp_riskiq.py:useragent="SpiderFoot",headers=headers,
modules/sfp_s3bucket.py:res=self.sf.fetchUrl(url,timeout=10,useragent="SpiderFoot",noLog=True)
```


©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Overview

Name	Recon-ng
URL	http://recon-ng.com [1]
Technology	Python 3
Owner	Tim Tomes (@lanmaster53)
Number of Modules	70+
Use	Command line, API and RPC
Versions	Free (Version 5.x is latest)

Recon-ng Overview

Tim Tomes' (@lanmaster53) reconnaissance framework, Recon-ng, is a modular Python application that is easy to run and extend with your own modules. It has an interactive command-line interface, similar to Metasploit's msfconsole, and an API so that it can be called from other scripts. Over the years, Tim and many others have created many YouTube videos (https://sec487.info/a4), and others in the community have performed talks at conferences or written blog entries (https://sec487.info/a5) to explain how to use this framework.

Its modules run the OSINT spectrum, from collecting data about hosts, IP addresses, and domains to finding user accounts and credentials on sites across the internet.

One thing to note is that in 2019, Recon-ng underwent an upgrade from Python 2.x to 3.x. Tim and a developer also went through and overhauled how the tool worked, changing everything from how the modules are installed to the command names used in the tool. We mention this because you may encounter older videos, blog posts, and conference talks showing how to use the older 4.x version of Recon-ng. Understand that there is a big difference between 4.x and 5.x versions in Recon-ng and, while some of the tool stayed consistent across versions, a huge number of positive changes also occurred.

For more information about how to use Recon-ng v5, visit https://sec487.info/ux.

[1] The recon-ng.com domain redirects to https://github.com/lanmaster53/recon-ng, where Tim Tomes hosts the source code for Recon-ng. Some users get concerned (but should not) when their browsers get redirected to this code repository.

© SANS Institute 2019

Recon-ng in Action

Recon-ng is currently using the version 5.x branch of the tool (1)

By default, Recon-ng v5 does not install any modules (2)

```
Sponsored by...
www.blackhillsinfosec.com
www.practisec.com
[recon-ng v5.0.1, Tim Tomes (@lanmaster53)]
[*] No modules enabled/installed.
[recon-ng][default] >
```


Recon-ng in Action

The most common method of using Recon-ng is via the CLI (command-line interface). Once Recon-ng is installed on your system, in a terminal window or at a command prompt, you type `./recon-ng --no-analytics` and you should see a screen similar to the one above. For help on installing Recon-ng, visit https://sec487.info/uy. This is the interactive method of using Recon-ng, and you can see at arrow 1 that we ran version 5.0.1 of the tool.

Note that, Recon-ng version 5.x (v5) does not install any modules when you install the base tool. This is shown in the slide at arrow 2 and is one of the more important distinctions between version 4.x and 5.x of the tool. We will explore this later.

At any point you can type "help" (no quotes) and hit Enter, and the tool will show its help screen (shown below).

```
[recon-ng][default] > help

Commands (type [help|?] <topic>):
---------------------------------
back            Exits the current context
dashboard       Displays a summary of activity
db              Interfaces with the workspace's database
exit            Exits the framework
help            Displays this menu
index           Creates a module index (dev only)
keys            Manages third party resource credentials
marketplace     Interfaces with the module marketplace
modules         Interfaces with installed modules
options         Manages the current context options
pdb             Starts a Python Debugger session (dev only)
script          Records and executes command scripts
shell           Executes shell commands
show            Shows various framework items
snapshots       Manages workspace snapshots
spool           Spools output to a file
workspaces      Manages workspaces
```


©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Options

Viewing the settings of how Recon-ng works is performed at the global and module levels using the "`options list`" command (1)

```
[recon-ng][default] > options list

  Name        Current Value  Required  Description
  ----------  -------------  --------  -----------
  NAMESERVER  8.8.8.8        yes       default nameserver for the resolver mixin
  PROXY                      no        proxy server (address:port)
  THREADS     10             yes       number of threads (where applicable)
  TIMEOUT     10             yes       socket timeout (seconds)
  USER-AGENT  Recon-ng/v5    yes       user-agent string
  VERBOSITY   1              yes       verbosity level (0 = minimal, 1 = verbose, 2 = debug)
```

```
[recon-ng][default] > options set TIMEOUT 30
TIMEOUT => 30
[recon-ng][default] > options set USER-AGENT
"Mozilla/5.0 (Windows NT 10.0; Win64; x64;
rv:69.0) Gecko/20100101 Fefirox/69.0"
USER-AGENT => "Mozilla/5.0 (Windows NT 10.0;
Win64; x64; rv:69.0) Gecko/20100101
Firefox/69.0"
```


Recon-ng Global Options

At the global level, when you first run the app, you have access to tool-wide preferences, as shown in the above slide. These govern the DNS server the tool should use and the user agent string sent to web sites, among other settings. Tune these to meet your OSINT and OPSEC needs before your first use.

We will change the TIMEOUT (the amount of time in seconds that Recon-ng waits for an HTTP response from a system) to 30 seconds. We find that 10 seconds, the default value, is appropriate to use on fast, reliable networks. Using VPNs or other networks may introduce latency and cause Recon-ng to not wait long enough for server responses. In the above slide, we show you the command to give Recon-ng to change this setting: `options set TIMEOUT 30`.

Next, we highly recommend that you change the default Recon-ng user agent string from "Recon-ng/v5" to something looking a bit more common like Firefox's user agent. As shown above, we send the following command to Recon-ng to change this parameter: `options set USER-AGENT "Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:69.0) Gecko/20100101 Firefox/69.0"`.

You may decide that other global settings should be changed too. The format of the command you send will be in the same pattern as shown above: `options set VARIABLE YOURDATA` (where VARIABLE is the name of a parameter in Recon-ng and YOURDATA is the new value you wish to set).

© SANS Institute 2019

Internal Database and Workspaces

Recon-ng has internal SQLite database(s) it uses to store data it retrieves

The database, by default, is in your home directory in the `.recon-ng` directory

Workspaces separate case data by creating new directories and database files for each space

```
[recon-ng][default] > workspaces create case1
[recon-ng][case1] > workspaces list

  +------------+
  | Workspaces |
  +------------+
  | case1      |
  | default    |
  +------------+
```


Internal Database and Workspaces

When Recon-ng is gathering data from the internet, it needs to store it somewhere. It does this in a SQLite database in the user's home directory. The tool creates the `.recon-ng` directory in Linux and macOS operating systems to store settings, API keys, and results. One important warning is that this directory is hidden (see the leading "." on the directory name?). You need to know about it so that you can remove old cases and workspaces.

Since users may use Recon-ng across a variety of clients and cases, and we don't want one customer or target's data getting mingled with another's, Recon-ng employs workspaces to segment retrieved data. In the tool, we can pass it commands to create new workspaces, select existing ones, and delete old workspaces as needed. Each workspace has its own directory in the user's home directory and SQLite database to store results. When you delete an old workspace through the tool, it removes the workspace and the data it contains from Recon-ng's database and file system.

We manage workspaces using the "workspaces" command. In the slide above, we created a new workspace using the `workspaces create case1` command. Then we checked our work to see if it created successfully by listing all the workspaces with `workspaces list`.

©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Database Tables

To view all tables and their schemas, type: `db schema`

To view a table, type, `show TABLENAME` (for example, to view hosts table: `show hosts`)

Can make custom database queries:

```
db query select host,ip_address
from hosts where host like
'%sans%' limit 5
```

```
[recon-ng][default] > db query select host,ip_address from hosts where host like
 '%sans%' limit 5

  +---------------------------------------------+
  |             host             |   ip_address  |
  +---------------------------------------------+
  | sans.org                     | 45.60.103.34  |
  | digital-forensics21.sans.org | 66.35.59.133  |
  | www21.sans.org               | 66.35.59.103  |
  | pre-ondemand31.sans.org      | 204.51.94.121 |
  | digital-forensics31.sans.org | 204.51.94.133 |
  +---------------------------------------------+
```


Recon-ng Database Tables

Everyone brings their own experiences and knowledge into their OSINT work. If yours happens to have some database experience, then you will want to know about the other `db` commands you can send to Recon-ng.

1. To view the schema of all the database tables, type `db schema`. This will output the table names, what fields can be found within each, and what types of data can be stored in those fields.
2. To view all the existing data in a database table (all the "rows"), type `show TABLENAME` (where you replace TABLENAME with the name of the table you wish to view). So typing `show vulnerabilities` will dump the vulnerabilities table data to the screen. While you might want to see all the data, this is sometimes suboptimal because the data may wrap around the screen or show information that you are not interested in.
3. Recon-ng allows us to use common SQL statements (more info at https://sec487.info/v1 and https://sec487.info/v2). For instance, perhaps our hosts table had 4,000 hosts in it and we only wanted to see five hosts that had the string "sans" in them? Oh, and we don't want to see all the other columns in the table (region, city, etc.), just the host and IP address. In that case, we can use an advanced query like the one shown above. The more you understand about making SQL queries, the more powerful you will be within the `db` commands of Recon-ng.

© SANS Institute 2019

Recon-ng Module Marketplace

All modules have been moved to a marketplace

No modules are installed by default

```
[recon-ng][default] > marketplace search

+-------------------------------------------------------------------------------------+
|                          Path                          | Version |     Status    |  Updated   | D | K |
+-------------------------------------------------------------------------------------+
| discovery/info_disclosure/cache_snoop                  | 1.0     | not installed | 2019-06-24 |   |   |
| discovery/info_disclosure/interesting_files            | 1.0     | not installed | 2019-06-24 |   |   |
| exploitation/injection/command_injector                | 1.0     | not installed | 2019-06-24 |   |   |
| exploitation/injection/xpath_bruter                    | 1.1     | not installed | 2019-08-19 |   |   |
| import/csv_file                                        | 1.1     | not installed | 2019-08-09 |   |   |
| import/list                                            | 1.0     | not installed | 2019-06-24 |   |   |
| import/nmap                                            | 1.0     | not installed | 2019-06-24 |   |   |
| recon/companies-contacts/bing_linkedin_cache           | 1.0     | not installed | 2019-06-24 |   | * |
| recon/companies-contacts/pen                           | 1.0     | not installed | 2019-06-24 |   |   |
| recon/companies-domains/pen                            | 1.0     | not installed | 2019-06-24 |   |   |
| recon/companies-domains/viewdns_reverse_whois          | 1.0     | not installed | 2019-08-08 |   |   |
| recon/companies-multi/github_miner                     | 1.0     | not installed | 2019-06-24 |   | * |
| recon/companies-multi/shodan_org                       | 1.0     | not installed | 2019-06-26 |   | * |
| recon/companies-multi/whois_miner                      | 1.0     | not installed | 2019-06-24 |   |   |
| recon/contacts-contacts/mailtester                     | 1.0     | not installed | 2019-06-24 |   |   |
| recon/contacts-contacts/mangle                         | 1.0     | not installed | 2019-06-24 |   |   |
| recon/contacts-contacts/unmangle                       | 1.0     | not installed | 2019-06-24 |   |   |
| recon/contacts-credentials/hibp_breach                 | 1.2     | not installed | 2019-09-10 |   | * |
| recon/contacts-credentials/hibp_paste                  | 1.1     | not installed | 2019-09-10 |   | * |
| recon/contacts-domains/migrate_contacts                | 1.0     | not installed | 2019-06-24 |   |   |
```

```
[recon-ng][default] > marketplace search
```


Recon-ng Module Marketplace

Recon-ng uses modules to perform its OSINT work. Each module has a purpose, which could be to retrieve data from an internet source or manipulate existing data in some fashion. In Recon-ng v4, all modules were installed by default, but that changed with v5 when all modules were pushed to the "Marketplace." Centralizing the modules in one place makes it easier to maintain, deprecate, and update them, which were all problems that plagued Recon-ng v4.

To see all the modules you can install, type `marketplace search` (arrow 1). With no additional parameters, this command will list all of the modules available, their names (arrow 2), versions, statuses (arrow 3), when they were last updated, if there are dependencies (the "D" column) that need to be installed on your system before you can use the modules, and if the module requires an API key (the "K" column) to work.

©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Module Installation

Install one or more modules using the `marketplace install` command

```
[recon-ng][default] > marketplace install XXXXX
```

```
[recon-ng][default] > marketplace install profiler
[*] Module installed: recon/profiles-profiles/profiler
[*] Reloading modules...
```

Install all of them by replacing XXXXX with all

```
[recon-ng][default] > marketplace search

  +-------------------------------------------------------------------------------------+
  |                          Path                          | Version |   Status   |  Updated   |
  +-------------------------------------------------------------------------------------+
  | discovery/info_disclosure/cache_snoop                  | 1.0     | disabled   | 2019-06-24 |
  | discovery/info_disclosure/interesting_files            | 1.0     | installed  | 2019-06-24 |
  | exploitation/injection/command_injector                | 1.0     | installed  | 2019-06-24 |
  | exploitation/injection/xpath_bruter                    | 1.1     | installed  | 2019-08-19 |
  | import/csv_file                                        | 1.1     | installed  | 2019-08-09 |
  | import/list                                            | 1.0     | installed  | 2019-06-24 |
  | import/nmap                                            | 1.0     | installed  | 2019-06-24 |
  | recon/companies-contacts/bing_linkedin_cache           | 1.0     | disabled   | 2019-06-24 |
```

1

Recon-ng Module Installation

We can install as many of the modules as we wish. Keep in mind that importing and exporting data to and from Recon-ng are controlled by modules, too. The command to install modules is `marketplace install XXXXX` (where you replace XXXXX with the name of the module you wish to install or use the word "all" to install all the modules).

Why doesn't Recon-ng come with the modules installed by default? One reason is that when you install a module that requires an API key that you have not added to Recon-ng yet, it will print out an error and disable the module (see arrows at 1 above). If you only have the modules installed that you have your system configured to run, then you do not get those warning banners.

I like having all of the modules installed locally on my system, but sometimes you may wish to be more tactical in your module installation. Above, we installed a single module using the `marketplace install profiler` command. Then we went ahead and installed them all by replacing "profiler" with "all".

To show that the modules are installed, use the `marketplace search` command with no strings and it will now show that the modules are installed (image above). Some, requiring API keys or dependencies that your system may not have yet, might show as disabled.

© SANS Institute 2019

Recon-ng Modules

Recon-ng's modules are what perform the retrieval of data for us

The modules in the "recon" branch are OSINT-related

Modules organized by what you have and what you want: "domains-contacts"

```
recon/contacts-credentials/hibp_paste
recon/contacts-domains/migrate_contacts
recon/contacts-profiles/fullcontact
recon/credentials-credentials/adobe
recon/credentials-credentials/bozocrack
recon/credentials-credentials/hashes_org
recon/domains-companies/pen
recon/domains-contacts/metacrawler
recon/domains-contacts/pen
recon/domains-contacts/pgp_search
recon/domains-contacts/whois_pocs
recon/domains-credentials/pwnedlist/account_creds
recon/domains-credentials/pwnedlist/api_usage
recon/domains-credentials/pwnedlist/domain_creds
recon/domains-credentials/pwnedlist/domain_ispwned
recon/domains-credentials/pwnedlist/leak_lookup
recon/domains-credentials/pwnedlist/leaks_dump
recon/domains-domains/brute_suffix
recon/domains-hosts/binaryedge
recon/domains-hosts/bing_domain_api
```


Recon-ng Modules

Recon-ng uses modules to perform its OSINT work. Each module has a purpose that could be to retrieve data from an internet source or manipulate existing data in some fashion. With over 70 modules in the project, you are going to need to go through and figure out which provide you the data you need.

The modules are organized into branches: recon, discovery, exploitation, reporting, and import. Within the recon branch, which is most important to us in our work, modules are further grouped by "what you have" and then a "-" and then "what you want". So, if you had a domain (for example, apple.com) and wanted contacts from that domain (for example, first and last names and email addresses), you would look in the "recon/domains-contacts/" modules for one that would retrieve the data you needed.

©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Module Information

To see what a module does and other details of an installed module, load it and then get info (1)

Shows description (2) and where the module pulls seed data (3)

```
[recon-ng][default] > modules load netcraft
[recon-ng][default][netcraft] > info

      Name: Netcraft Hostname Enumerator
    Author: thrapt (thrapt@gmail.com)
   Version: 1.0

Description:
  Harvests hosts from Netcraft.com. Updates the 'hosts' table with the results.

Options:
  Name    Current Value  Required  Description
  ------  -------------  --------  -----------
  SOURCE  default        yes       source of input (see 'show info' for details)

Source Options:
  default        SELECT DISTINCT domain FROM domains WHERE domain IS NOT NULL
  <string>       string representing a single input
  <path>         path to a file containing a list of inputs
  query <sql>    database query returning one column of inputs
```


Recon-ng Module Information

So, how do we find out what a module actually does and where it pushes or pulls data to or from? Those questions are found in the module information. The easiest method of accessing this is, for installed modules, to load the module and then ask for info. In the above image, we loaded the netcraft module by typing `modules load netcraft`. Then we asked for information by typing `info` (arrow 1).

The response contains a description of what the module does (arrow 2). This might also tell you where the data that is retrieved will be stored in Recon-ng (in the above example, the data will be stored in the "hosts" table).

One other important piece to understand is where the seed or starting data is going to come from. In the above example, by default, this netcraft module will pull all of the unique domains from the domains table (arrow 3). You can override this and tell Recon-ng to only use a certain domain by changing the SOURCE variable in the module. Type `options set SOURCE test.com` (replace test.com with your domain of interest).

© SANS Institute 2019

Starting a Recon-ng Scan

1. Create a new workspace
2. Choose the module
3. Enter seed target data
   - IP address
   - Domain name
   - Subnet
   - Email address
4. Run the scan

```
1. workspaces add case1
2. modules load whois_pocs
3. options set SOURCE faa.gov
4. run
```


Starting a Recon-ng Scan

The process for starting a Recon-ng scan is fairly straightforward. Once you have entered all the API keys you want and decided on which module to use, you can then create a new workspace (or use an existing one), select the module to use, add data or allow Recon-ng to pull content from its internal database if you have already run other scans and retrieved data, and then execute the scan.

The example on the right side of slide shows us:

1. Creating a new workspace called "case1."
2. Loading the already installed module named "whois_pocs," which searches for data in the ARIN Whois registry.
3. Setting the seed data or source to "faa.gov," which is what the module will look up.
4. Then we "run" the module.

©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Results

Can be retrieved with the `show TABLENAME` command

Export using reporting modules (see notes)

```
[*] 32 total (31 new) contacts found.
[recon-ng][case1][whois_pocs] > show contacts

 +-----------------------
```

© SANS Institute 2019

Recon-ng OPSEC Gotchas

In order to see which modules are used most, Recon-ng uses Google Analytics at launch

When run, the app will send information with your IP and modules you run from your system to Google

You may not want this information disclosed

To turn off this feature instead of launching the script using:

```
$ recon-ng
```

Use the "no analytics" flag:

```
$ recon-ng --no-analytics
```

Additionally, the default Recon-ng user agent is `Recon-ng/v5`

Recon-ng OPSEC Gotchas

If you do choose to use Recon-ng, there is something you need to know about it. Anyone can submit a module to the Recon-ng framework. If it is something that Tim Tomes believes is useful and reliable, he will include it in the framework. Sometimes people make modules and get them included in the framework and never update them…even when they no longer work. Tim then needs to figure out which of the modules he will maintain. To do this, he leverages Google Analytics to understand which modules are being used most. Those are then the modules that get fixed/maintained first. This presents the OSINT analyst with a problem since, by default, your Recon-ng system will send this traffic to Google and essentially tell Google and Tim that a person from your IP address ran Recon-ng and ran certain modules. If you are running the software from a client site, the client's IP is reported to Google. You may not want this data being sent.

To remedy this, Recon-ng allows the user to turn off the Google Analytics traffic using the "`--no-analytics`" flag when Recon-ng is started. It is highly suggested that you always run Recon-ng using this flag.

Lastly, the default user agent string that is sent out to every web site that is visited by Recon-ng is `Recon-ng/v5`. If you are trying to be stealthier and not let web sites know that you are using the Recon-ng tool, consider changing it using "`options set USER-AGENT XXXXX`" (where XXXXX is replaced with the user agent you wish to send).

©2019 Micah Hoffman

© SANS Institute 2019

Recon-ng Pros and Cons

Pros:

- Easy to start up
- Large variety of modules for host and people OSINT
- Many modules (create your own!)

Cons:

- Many modules require APIs to retrieve data
- Google Analytics shares your use with the tool owner
- Many blogs, posts, and video tutorials are using version 4 of Recon-ng

Recon-ng Pros and Cons

Some of the pros and cons (benefits and challenges) with the Recon-ng tool are presented. Keep in mind that you need to evaluate the tool for your needs, your team, and your method of working, so these pros and cons may not be as important to your work.

Pros:

- Easy to start up - Downloading and running Recon-ng is simple to do.
- Excellent module variety - There are modules for performing person and host OSINT.
- Many modules - With over 70 modules, Recon-ng most likely has modules that will help you in your OSINT research. If it doesn't, you can create your own modules in Python. Tim Tomes always welcomes new modules or alterations on existing ones via the Recon-ng git repository.

Cons:

- As with some of the other tools, Recon-ng leverages APIs to interact with some services. The API keys need to be obtained and entered in the tool for those modules that rely on them to work.
- Usage data is sent to Google Analytics by default.
- Documentation - Recon-ng has been a solid tool for a number of years and, as such, there is a plethora of videos, blog entries, and walkthroughs showing how to use the tool. Unfortunately, the commands used within Recon-ng changed significantly between version 4 and version 5. Most of the documentation out there, as of this writing, is for the older version 4.

© SANS Institute 2019

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"SpiderFoot"

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. DNS
4. Findings Online Devices
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. **Government Data**
8. Researching Companies

This page intentionally left blank.

© SANS Institute 2019

Government Data

Many governments give their citizens (and others) access to data that they collect

The data may be well-publicized and organized or challenging to find

Quantity and quality of data vary by geography

Example types of data:

- Contact directories of employees
- Salary information of employees
- Businesses registered in that country
- Military organization data
- Health information
- Life events (birth, marriage)
- Court data

Government Data

With the internet connecting people and governments, more information is being shared from governments via web applications. The quantity and quality of the data varies from country to country and province to province within those regions. Some governments have excellent categorized web resources and some you may have to wade through multiple pages to find the data you want.

©2019 Micah Hoffman

© SANS Institute 2019

GIS (Geographic Information System)

"A framework for gathering, managing, and analyzing data. Rooted in the science of geography, GIS integrates many types of data." [1]

GIS overlays layers of data on maps

Many governments share this data openly

Consider performing a search engine search for them with a search string of:

```
gis site:.gov.* -filetype:pdf
```

SANS | Open Source Intelligence (OSINT) Gathering and Analysis 139

GIS (Geographic Information Systems)

Governments possess huge amounts of data, and much of it can be geolocated. Whether you are looking to find where overhead power lines are strung or what counties have the most crime, these GIS platforms can help visualize and localize government data.

The question then becomes, "How do I find these GIS sites?" We suggest performing searches using your favorite search engines for strings like "gis site:XXXXX -filetype:pdf". Replace the "XXXXX" in that string with the highest level of the domain you wish to search. For instance, in the United States, we might make that "site:.gov" but looking for GIS sites in India might change our search to contain "site:.gov.in". In Google's search site, this will retrieve over 3.4 million results! Some of these results are not the views with data overlaid upon maps.

Image from https://sec487.info/u0, September 25, 2019.

References:

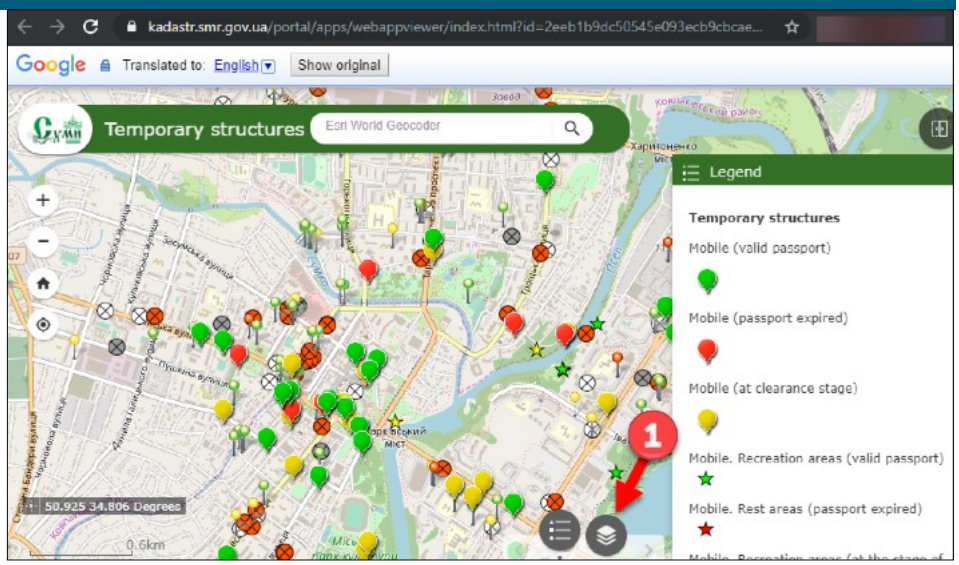
[1] https://sec487.info/t-, September 25, 2019.

© SANS Institute 2019

GIS Example

Look for the "layers" icon (arrow 1) to select different data on the map

This is a map of temporary structures in Ukraine



GIS Example

Here, we look at an example of what a public GIS might look like. This is the Ukrainian web site at https://sec487.info/u1. We always look for the layers icon (arrow 1) when viewing these maps. This is the place where we can turn on and off additional information on the map. In the above example from Ukraine, we see temporary structures plotted on the map. Not shown in the image but displayed on the web site is a table with additional data about each structure plotted on the map.

Image from https://sec487.info/u1, September 25, 2019.

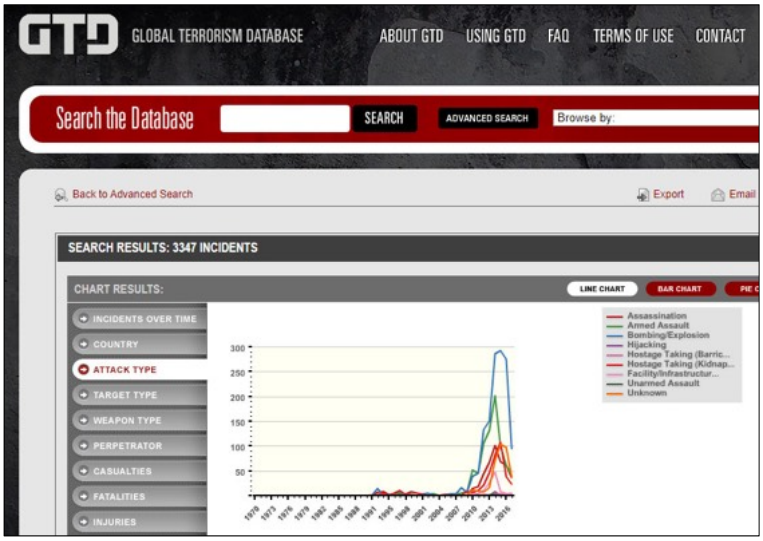
©2019 Micah Hoffman

© SANS Institute 2019

Open Data

Open Data (or "OpenData") is simply that—data that a government or organization shares without restrictions[1]

Variety of data in these systems, from terrorism databases to IP addresses and domains to people data



Open Data

The Open Data movement is about sharing the huge amounts of data some organizations and governments collect.[1] The content of the shared data can vary from Rapid7's TCP scans of the internet[2] to the RAND Database of Worldwide Terrorism Incidents[3] to information about governments and companies in Africa[4].

Many of these sites contain search engines that can be used to parse their huge repositories of information. Other sites allow users to download some or all of this open data to their local computers for offline analysis. Sector035 wrote a "Week in OSINT"[5] post about these international Open Data sites if you would like additional information.

Image from https://www.start.umd.edu/gtd/, September 25, 2019.

References:

[1] https://sec487.info/u2

[2] https://sec487.info/u4

[3] https://sec487.info/u5

[4] https://sec487.info/u6

[5] https://sec487.info/u3

© SANS Institute 2019

United States Governments

There are 3 different levels of government in the United States:

- Federal
- State
- Local

At each level, certain data is published to the internet for government transparency and freedom of information

United States Governments

Within the United States, there are three levels of government, each having certain jurisdictions and responsibilities. Depending upon what you need for your investigation, you will want to target certain levels of these governments. As you move from federal to state and state to local, the services offered and information available becomes narrower and requires you to understand more of the details about what you need from where.

An example of this is looking for a person who is a federal employee and lives in Norfolk, Virginia. At the federal level, we may have records of government contracts the person works on for the United States Navy. At the state level, there may be records about the Limited Liability Company that this person created as a side business. Down at the local level, we might see volunteerism or community participation in a local event.

Images from https://www.usa.gov/, https://www.virginia.gov/, and https://www.norfolk.gov/, December 2, 2017.

©2019 Micah Hoffman

© SANS Institute 2019

Federal Level

The federal level of government contains overarching data about the United States government, policies, and people

Understanding the branches of government and possible data they may contain is helpful

Huge number of places to search; let's examine a few

Examples of Agencies and Data:

- Federal Election Commission
  - National elections and campaigns
- Department of Justice / FBI
  - Suspects and criminals
- Federal Communications Commission
  - Air waves, media, internet providers

Federal Level

At the federal level of the United States government, we have large agencies, departments, and commissions that are focused on certain categories of work. For instance, the State Department focuses on relationships with other countries and provides notices about the political stability of their governments. If your customer is looking to perform some business in a non-US country, these briefings can be beneficial.

Other agencies focus on the citizens and non-citizens within the country. The https://www.usa.gov web site shows the large breadth of government organizations that you could harvest data from on the page https://sec487.info/cn. Each page of this alphabetically organized site contains tens of links to federal organizations (the "D" page has 40 alone)! This is why, when looking to the federal government sites for data, you need to understand what you are looking for before you begin the search.

In the coming slides, we will show you a sampling of some types of data you can access. There are so many web applications offering so much data that we could create an entire course called "Retrieving Data from the US Government." The content in the following pages is to get you started thinking about, for your work, what federal government resources you could leverage to augment your assessment.

© SANS Institute 2019

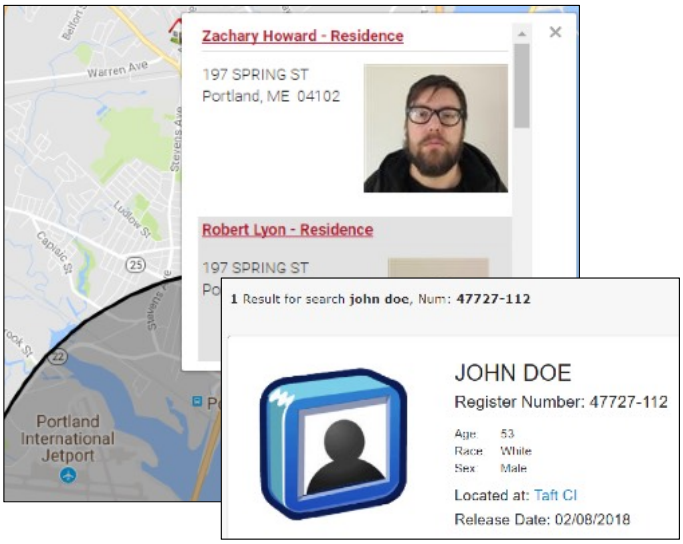
Federal Crime and Punishment

Bureau of Prisons

- Inmate locator (search)

Find registered offenders

- National Sex Offender Public Web site
- FBI Sex Offender Registry



Federal Crime and Punishment

The judicial branch of the federal government, along with the Bureau of Prisons (https://sec487.info/co) and the Department of Justice, has searches where people can look up those people who have been in or released from the United States prison system.

Some people must register with the federal (and state) governments as sex offenders, providing their addresses of where they live. This information is also searchable using the NSOPW (https://sec487.info/cp) and FBI web sites (https://sec487.info/cq). Enter an address or a person's name, and you can view the offenders who are registered in that area (example shown above) or the details about the person who registered, respectively.

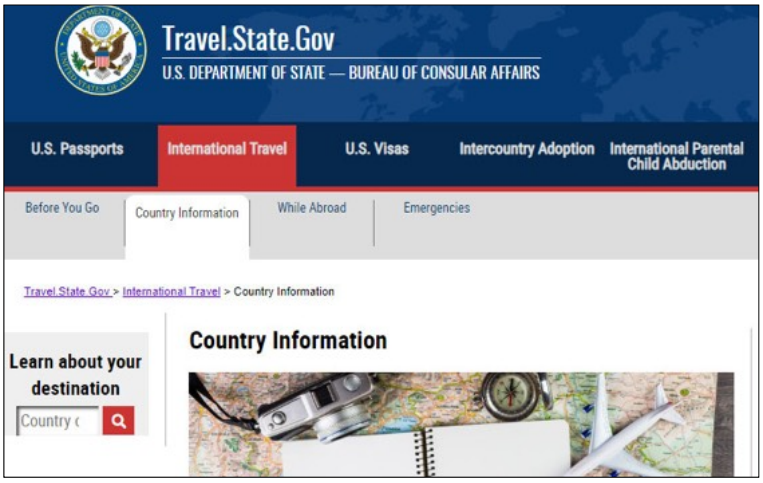
©2019 Micah Hoffman

© SANS Institute 2019

Federal State Department Guidance

The State Department's Travel Country Information pages aid in understanding political, social, and economic conditions in countries around the world

They also publish travel warnings for countries



Federal State Department Guidance

Doing OSINT on a country? Looking for information about the political scene or what tourism is like in a non-US country? The State Department's Travel Country Information Pages[1] can help you understand everything from the culture and religious beliefs of a country to security concerns a visitor may face. They provide details and analysis of many countries and can be valuable if you are researching a location for your assessment.

Image and Reference:

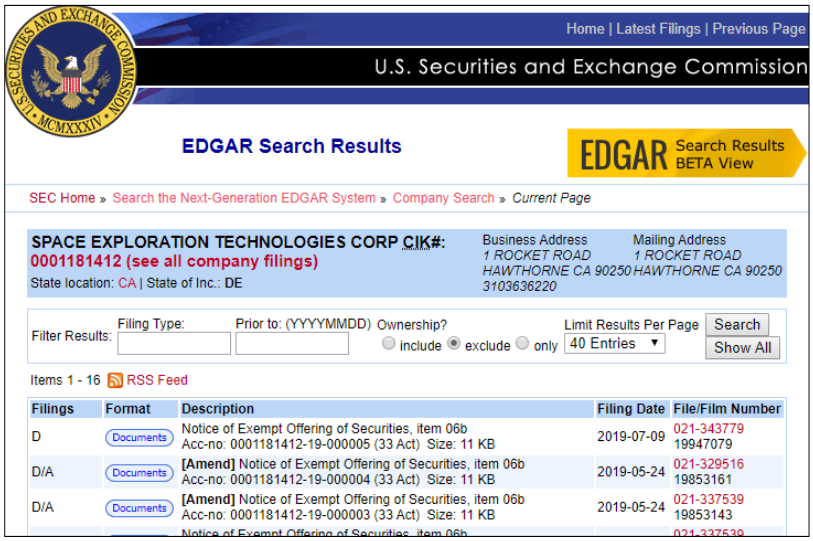
[1] https://sec487.info/nj, January 3, 2019.

© SANS Institute 2019

Securities and Exchange Commission

Business OSINT? Use EDGAR.

"EDGAR...performs automated collection, validation, [and] indexing...of submissions by companies and others who are required by law to file forms with the U.S. Securities and Exchange Commission (SEC)." [1]



Securities and Exchange Commission

The SEC requires public companies, those traded on stock markets, to publicly disclose information about their companies and the leaders within. These disclosures are collected and analyzed in EDGAR (Electronic Data Gathering, Analysis, and Retrieval system)[2] and made available for public searches. The data in EDGAR mainly focuses on businesses but can highlight interesting activities by its owners and leaders.

For person-based OSINT, the forms that can be searched here detail the buying, earning, and selling of company shares and leadership changes.

A detailed guide to using EDGAR for investors (or OSINT analysts!) is available at https://sec487.info/ct.

Image from https://sec487.info/ts, September 25, 2019.

References:

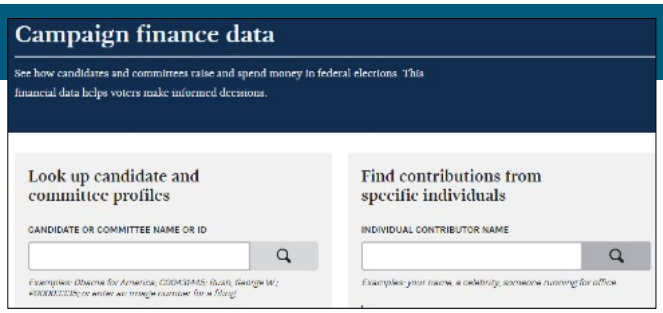
[1] https://sec487.info/cu

[2] https://sec487.info/ag

 ©2019 Micah Hoffman

© SANS Institute 2019

Federal Election Commission



The FEC tracks who contributed to what campaign

It allows you to search[1] by candidate and by contributor

Looking into political contributors? FEC will help

Contributor name	Recipient	State	Employer	Receipt date	Amount
WASHINGTON, EBONY	MACOMB COUNTY DEMOCRATIC COMMITTEE	MI	NONE	11/24/2018	$37.00
WASHINGTON, EBONY	PHILIP A. HART DEMOCRATIC CLUB	MI	NONE	11/23/2018	$42.00
WASHINGTON, EBONY	MACOMB COUNTY DEMOCRATIC COMMITTEE	MI	NONE	11/23/2018	$37.00
WASHINGTON, EBONY	OAKLAND COUNTY DEMOCRATIC PARTY	MI	NONE	11/17/2018	$39.00
WASHINGTON, EBONY	PHILIP A. HART DEMOCRATIC CLUB	MI	NONE	11/16/2018	$38.00

Federal Election Commission

This commission tracks aspects of campaign finance—both who is donating and to whom are they sending the money. All this data is available for downloading and/or searching online. Since the information is made available for download, other sites such as Political Money Line (https://sec487.info/cv) pull the data and use other sources to deliver similar "Who is donating to whom?" resources.

Images from:

https://sec487.info/cw, January 3, 2019.

https://sec487.info/oi, January 5, 2019.

Reference:

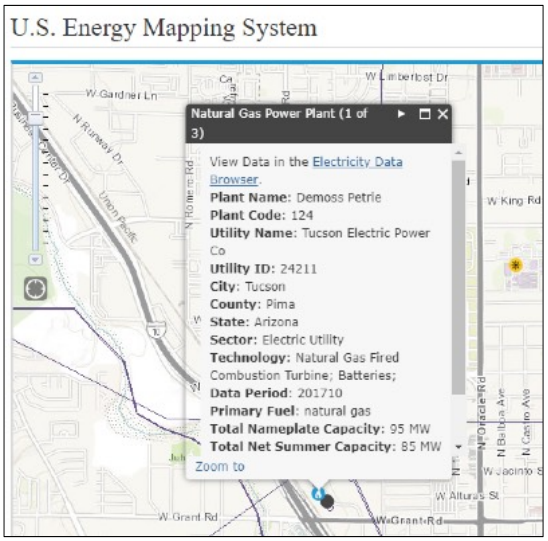
[1] https://sec487.info/cw, January 3, 2019.

© SANS Institute 2019

U.S. Energy Information Administration

The U.S. Energy Administration site displays interactive maps of power generation plants, power lines, pipelines, and fuel storage tanks.

Details include owners, size/capacity, and locations



U.S. Energy Information Administration

Geographic Information Systems (GIS) show information mapped to geographic locations. The United States Energy Information Administration maps (https://sec487.info/gz) tie electrical, petroleum, and natural gas power generation and transmission plants to locations. It provides an interactive map that displays power lines, right of ways, pipelines, petroleum storage tanks, and other energy-related buildings.

Image from https://sec487.info/gz, March 27, 2018.

©2019 Micah Hoffman

© SANS Institute 2019

Federal Communications Commission

The FCC manages the US's air waves

People and companies need to register for licenses to transmit on certain frequencies or have specific equipment

	Call Sign/Lease ID	Name	FRN	Radio Service	Status	Expiration Date
61	DO00015955	HERNANDEZ, MARTIN	0025492786	CM	Active	
62	DO00016696	HERNANDEZ, DAVID E	0026139535	CM	Active	
63	DO00017389	HERNANDEZ, LUIS A	0026884957	CM	Active	
64	DO00017824	HERNANDEZ, MATTHEW T	0027457662	CM	Active	
65	DOGB017413	RNANDEZ, HUGO G		CM	Expired	02/01/2001
66	DOGB025294	DEZ RODRIGUEZ, AGUSTIN V		CM	Expired	09/20/2001
67	DOGB034427	ORTIZ, AGUSTIN		CM	Expired	06/09/2002
68	DOGB051665					
69	DOGB056737					
70	DOGB074370					

MAIN ADMIN

Call Sign	DO00017389	Radio Service	CM - Comn
Status	Active	Auth Type	Regular
Dates			
Grant	10/05/2017	Expiration	
Effective	10/05/2017	Cancellation	
Licensee Information			
FRN	0026884957	Type	Individual
Licensee Name			
HERNANDEZ, LUIS A 189 WEKIVA COVE DESTIN, FL 32541			

Federal Communications Commission

Thinking about all of the different areas of government that collect user or people data, we come to the FCC (Federal Communications Commission). They manage the wireless spectrum and regulate media such as radio and TV stations. If your OSINT target is one who loves amateur radio (commonly called ham radio) as a hobby, then the FCC may have some data for you. Amateur radio operators must take and pass an exam to be able to transmit on the amateur radio frequencies. Where there is a test, there is registration, and where there is government registration, there is usually a searchable web page.

The amateur, aircraft, ship, and other radio operators search page can be found at https://sec487.info/cx.

For our OSINT purposes, we can find names and addresses as well as call signs in these databases.

Images from https://sec487.info/tt and https://sec487.info/tu, September 25, 2019.

© SANS Institute 2019

Federal Agency Directories

Many government agencies make the work phone numbers of their employees searchable

Know someone working in the federal government? Go to the agency page and look for a phonebook/directory

Federal Agency Directories

Many federal agencies and commissions maintain publicly accessible phone books of their employees. These searchable, online repositories of OSINT goodness can contain names, work locations, phone numbers, email addresses, and titles of the workers within an organization.

Not all agencies have them on the internet, and others only publish some content. If you have a target within a federal agency, searching on these sites can yield valuable data to collect and pivot upon.

Images:

https://sec487.info/nk, December 2, 2017.

https://sec487.info/nl, December 2, 2017.

©2019 Micah Hoffman

© SANS Institute 2019

Federal Government Worker Pay

There are several third-party sites with searchable databases of some federal government employee pay histories

Results retrieve work locations, pay grades, and how much money people make

Rachel

Title: Medical Officer

Agency: Food and Drug Administration

In 2017, Rachel was a Medical Officer at the Food and Drug Administration in Silver Spring, Maryland. As our dataset only goes as far back as 2004, it is likely that has worked in the federal government prior to 2004.

Rachel is a RF-00 under the code is for use by hhs only. payscale and is among the **highest-paid ten percent of employees** in the Food and Drug Administration.

Year	Occupation	Paygrade	Base Salary	Bonus	Location
2017	Medical Officer	RF-00	$390,000	$0	Silver Spring, Maryland
2016	Medical Officer	RF-00	$285,000	$0	Silver Spring, Maryland
2013	Medical Officer	AD-00	$275,600	$3,445	Rockville, Maryland
2012	Medical Officer	AD-00	$275,600	$0	Rockville, Maryland

SANS Open Source Intelligence (OSINT) Gathering and Analysis 151

Federal Government Worker Pay

Some federal government workers' pay is public record. Visit sites like https://www.federalpay.org/ (https://sec487.info/i2) and https://www.fedsdatacenter.com/ (https://sec487.info/i1), enter a first and last name of a worker, and retrieve their organization and pay (salary or hourly wage). Not all employees of the federal government will appear in this data, and sometimes the data is not current, but it can give you historical data that may be useful.

Much of the data on these sites may be redacted, but the sites often aggregate data and display what a person of a certain position might make.

The above image came from https://sec487.info/tv, September 25, 2019.

© SANS Institute 2019

PACER Court Case Locator

Information from US district, bankruptcy, and appellate courts (civil and criminal cases)

Requires authentication and fee for heavy use

Retrieve details about cases, plaintiffs, defendants, court documents, audio clips

CLOSED,IFP

United States District Court
District of Nevada (Las Vegas)
CIVIL DOCKET FOR CASE #: 2:18-cv-01766-APG-NJK

Gunn v. Hillin et al
Assigned to: Judge Andrew P. Gordon
Referred to: Magistrate Judge Nancy J. Koppe
Demand: $5,000,000
Cause: 15:1692 Fair Debt Collection Act

Date Filed: 09/11/2018
Date Terminated: 04/11/2019
Jury Demand: Plaintiff
Nature of Suit: 480 Consumer Credit
Jurisdiction: Federal Question

Plaintiff
Jason Gunn represented by **Jason Gunn**
10620 Southern Highlands Parkway #153
Las Vegas, NV 89141
702-356-3810
PRO SE

V.

Defendant
Sean P. Hillin

Defendant
Locate Source America LLC

Defendant
TY Capital Management, LLC

PACER Court Case Locator

PACER (Public Access to Court Electronic Records) is run by the United States Administrative Office of the US Courts (https://sec487.info/tz). It is a national index of United States appellate, district, and bankruptcy court cases. The data inside is outstanding if you need to research court proceedings about a target person or company at the US federal level.

PACER requires a valid login with a valid address for it to send an account token that you use on your first login. However, providing a credit card to the service will immediately activate your searching privileges on the site. The services on PACER cost $0.10 per page, but there is no fee if usage is less than $30 per quarter.[1]

What types of data can be retrieved from PACER? Essentially everything that happened in the court record that was public. You can retrieve documents used in cases, see who the debtors of a person filing for bankruptcy were, examine audio files of court conversations, and discover much more.

Image from https://sec487.info/ty (requires valid account), September 25, 2019.

Reference: [1] https://sec487.info/tx

©2019 Micah Hoffman

© SANS Institute 2019

Freedom Of Information Act (FOIA)

Under the Freedom of Information Act of 1967,[1] anyone can request a federal US government agency to share information

Some agencies have exclusions that restrict access to their data

Sites like MuckRock catalog public FOIA requests

Facial Recognition - Buffalo (NY) No Responsive Documents	Freddy Martinez	Buffalo Police Department	Buffalo, NY	09/25/2019	07/08/2019
Expenditures relating to visits from federal officials (Phoenix Police Department) Awaiting Response	Emma Best	Phoenix Police Department	Phoenix, AZ	09/25/2019	09/17/2019
Summit Showdown Awaiting Response	Carlos Gamez	U.S. Department of State	United States of America	09/25/2019	05/24/2018
Hyperloop One Global Challenge Finalist Documents Awaiting Response	Paxtyn Merten	Miami-Dade County Transportation and Public Works	Florida	09/25/2019	10/29/2018
Dark Matter/NESA Awaiting Acknowledgement	Jurre van Bergen	U.s. Cyber Command	United States of America	09/25/2019	09/25/2019
RI Labor and Training Completed	Caroline Haskins	Department of Labor and Training	Rhode Island	09/25/2019	09/11/2019

Freedom Of Information Act (FOIA)

The 1967 Freedom of Information Act[1] allows anyone to request records from the United States government. It was an attempt to bring more transparency and accountability to government activities. Over 50 years later, we still have the ability to request government data using the FOI requests. The government may not return certain data that Congress has identified as "excluded" from FOIA requests, and agencies may redact some of the data as well.

The above image is an excerpt from https://sec487.info/tz on September 25, 2019 that shows several requests that the MuckRock.com web site contained. This site allows people to submit FOI requests through its system, which it then tracks the status of. When responses are filled, MuckRock hosts the responses and any documents sent with them on its site. We can use this web site to view other people's FOI requests and use the data retrieved for our OSINT projects.

References:

[1] https://www.foia.gov

© SANS Institute 2019

State Level

All of the 50 states (along with the US territories) have web sites that host data

State sites are excellent for finding business registration, court records, and real estate data

Examples of Departments and Data:

- Department of Corrections
  - Search for prisoners
- Board of Bar Examiners
  - Who is a lawyer in good standing?
- Department of Motor Vehicles
  - Driver's licenses and vehicle registrations
- Business permits and licenses

State Level

State governments and US territories are responsible for providing a range of services to their populations. They are also tasked with a variety of registration and regulation responsibilities that require them to receive, record, and sometimes publish data about people in their regions.

To find relevant data at the state level, you will need to understand the geographic location of the person or business you are performing the OSINT research upon. Where does a person reside? What state was a business registered in? These are common questions that should be examined prior to performing state-level government data searches. Without understanding where your target is, you run the risk of wasting time searching sites that will not have the information you need.

The slide above notes a tiny sample of the types of data we might be able to find at the state level. We cannot, obviously, visit and review the content on all of the state web pages. Instead, over the next few slides, we will examine some of the content on New Hampshire's, Arizona's, and Montana's web sites.

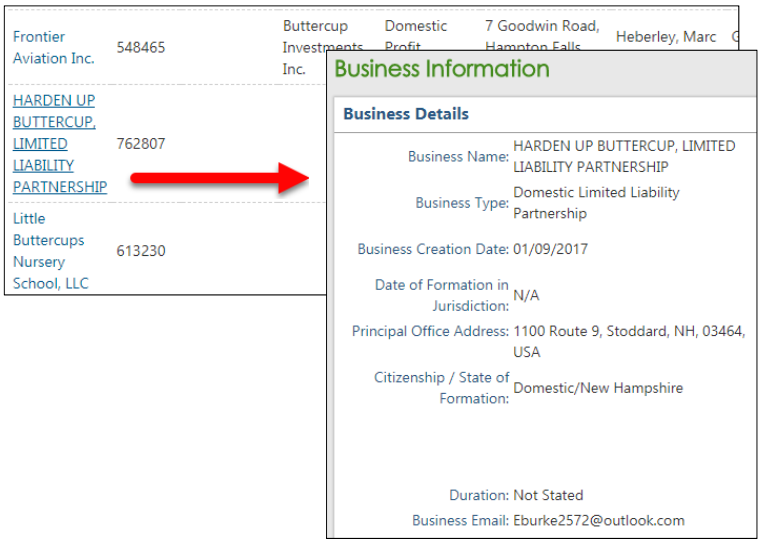
©2019 Micah Hoffman

© SANS Institute 2019

New Hampshire's Corporate Division

Corporate and business registration searches can yield:

- Names
- Addresses
- Phones
- Emails
- Partner names
- Trademarks



New Hampshire's Corporate Division

Businesses and corporate entities send a variety of forms and registrations to state governments. We can search these documents (or web pages where that content is displayed) for information about a business. While in the above example we see mostly contact information, we can record this data and then pivot on much of it to expand the OSINT research area. Additionally, using the above data, we could search state government sites that showed licenses and permits to find out what the business does, who are its clients, and more.

Images from:

https://sec487.info/nm, December 3, 2017.

https://sec487.info/fm, December 3, 2017.

Reference:

[1] https://sec487.info/cy

© SANS Institute 2019

Arizona's Court Cases

Many states have made their court cases (judicial system) searchable

You can find historical data, open warrants, details about plea deals, and more

Case Number	Name	Party Type	Birth Date	Address
M-0945-CR-92000130	A MICHELLE SMITH	DEFENDANT		SHOW LOW, AZ 85901
S-1100-DO-201601004	A THOMAS SMITH	ATTORNEY		
M-0763-CR-20040197	A VIRGINIA WILLIAMS SMITH	DEFENDANT	10/1942	SUN CITY, AZ 85351

Case Information

Case Number:	M-0763-CR-20040197		
Title:	ST OF AZ VS SMITH A VIRGINIA	Category:	Criminal
Court:	Youngtown Municipal	Filing Date:	7/14/2004
Judge:		Disposition Date:	8/11/2004

A VIRGINIA WILLIAMS SMITH DEFENDANT - D 1 Date of Birth: 10/1

Citation	Count	Description	Disp. Date	Disposition
00700014774	1	GAME AND FISH TITLE VIO	8/11/2004	PLEA GUILTY/RESP SENT IMPOSED

Arizona's Court Cases

Court records from criminal violations, lawsuits, and civil proceedings can be found on many state (and local) government sites. These records can provide valuable OSINT data for your work, as they contain dates and times, charges against people, lawsuit details and references, the history of the proceedings and historical information about people/subjects, current open warrants for people, ages, contact information, and locations.

Images from https://sec487.info/fn, December 3, 2017

©2019 Micah Hoffman

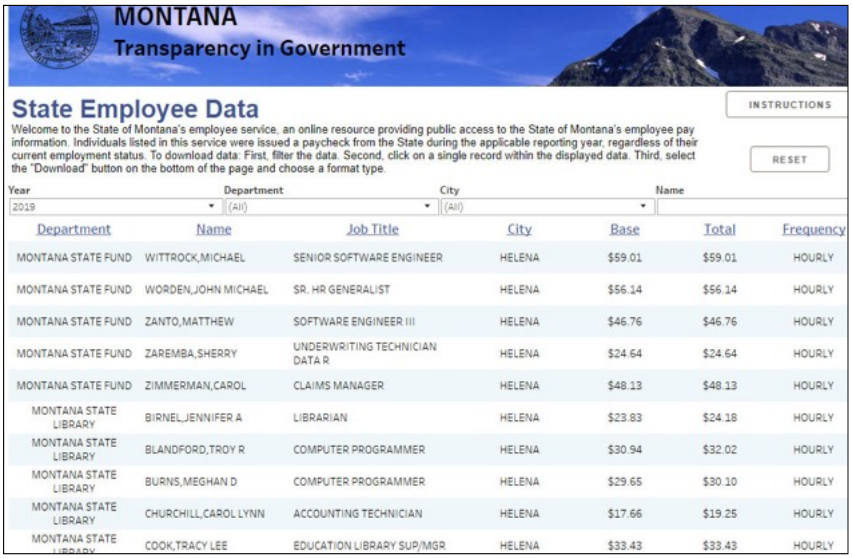
© SANS Institute 2019

Montana State Employee Data

Your subject work in the state government?

Their current and historical pay data may be accessible

Many states show this public data



MONTANA
Transparency in Government

State Employee Data

INSTRUCTIONS

Welcome to the State of Montana's employee service, an online resource providing public access to the State of Montana's employee pay information. Individuals listed in this service were issued a paycheck from the State during the applicable reporting year, regardless of their current employment status. To download data: First, filter the data. Second, click on a single record within the displayed data. Third, select the "Download" button on the bottom of the page and choose a format type.

RESET

Year: 2019 | Department: (All) | City: (All) | Name:

Department	Name	Job Title	City	Base	Total	Frequency
MONTANA STATE FUND	WITTROCK,MICHAEL	SENIOR SOFTWARE ENGINEER	HELENA	$59.01	$59.01	HOURLY
MONTANA STATE FUND	WORDEN,JOHN MICHAEL	SR. HR GENERALIST	HELENA	$56.14	$56.14	HOURLY
MONTANA STATE FUND	ZANTO,MATTHEW	SOFTWARE ENGINEER III	HELENA	$46.76	$46.76	HOURLY
MONTANA STATE FUND	ZAREMBA,SHERRY	UNDERWRITING TECHNICIAN DATA R	HELENA	$24.64	$24.64	HOURLY
MONTANA STATE FUND	ZIMMERMAN,CAROL	CLAIMS MANAGER	HELENA	$48.13	$48.13	HOURLY
MONTANA STATE LIBRARY	BIRNEL,JENNIFER A	LIBRARIAN	HELENA	$23.83	$24.18	HOURLY
MONTANA STATE LIBRARY	BLANDFORD,TROY R	COMPUTER PROGRAMMER	HELENA	$30.94	$32.02	HOURLY
MONTANA STATE LIBRARY	BURNS,MEGHAN D	COMPUTER PROGRAMMER	HELENA	$29.65	$30.10	HOURLY
MONTANA STATE LIBRARY	CHURCHILL,CAROL LYNN	ACCOUNTING TECHNICIAN	HELENA	$17.66	$19.25	HOURLY
MONTANA STATE LIBRARY	COOK,TRACY LEE	EDUCATION LIBRARY SUP/MGR	HELENA	$33.43	$33.43	HOURLY

Montana State Employee Data

If the subject of your investigation is a state government employee, then you may be able to find their historical (and sometimes current) pay data online. Above, we show the Montana State Employee Data page,[1] but other states, such as North Carolina,[2] make this data available, too, sometimes coupled with their history of promotions and bonuses.

References:

[1] https://sec487.info/cz (URL not shown due to size)

[2] https://sec487.info/d0

© SANS Institute 2019

New York Vehicle Lookup

Open Data systems can be accessed through helper sites

These allow us to search and parse the data efficiently

NY:GWF8627 (Passenger)

Lookups: 3 | Recent: 09/13/2019

Violations: 19 *(hide violations)*

Date	Type	Location	Fines
07/09/2015	No Parking - Street Cleaning	Brooklyn (739 Dekalb Ave)	N/A
05/17/2016	Improperly Displayed Plate	Brooklyn (194 Tompkins Ave)	$0
08/02/2016	No Parking - Street Cleaning	Brooklyn (4719 Ave H)	$0
08/26/2016	Failure to Display Meter Receipt	Brooklyn (467 Myrtle Ave)	$0
12/27/2016	Fire Hydrant	Brooklyn (4723 Avenue H)	$115
01/11/2017	Expired Meter	Brooklyn (888 Fulton St)	$35

New York Vehicle Lookup

The City of New York Open Data project (https://sec487.info/u7) provides a vast amount of data about things that happen within New York City. Whether it is lists of people who have successfully completed the requirements to drive taxi cabs, permits from the Department of Buildings to start construction, or vehicles that have received parking or other tickets, this site allows you to download all of it.

Since the data is open, some helper web sites like https://www.howsmydrivingny.nyc/ have begun to leverage it for the public good. This site (and others around the country like the HowsMyDrivingDC Twitter bot at https://sec487.info/u8) make searching for data in these open data sets simple. For the howsmydrivingny.nyc site (shown above), simply fill in the form with what you wish to search for and the site will retrieve records of interest. For the DC Twitter account that does the same thing, simply tweet a license plate to the account, and it will pull the data from Washington DC's Open Data store.

This data is so open, you can also google it! Take the New York license plate "GWF8627" and enter it into Google. The first result is the New York City Open Data set (https://sec487.info/u9).

Images from https://www.howsmydrivingny.nyc/, September 25, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Local Level

The local level of government contains specific information about the people and businesses within its borders

"Local" is anything under the state level:

- County
- City
- Village

Examples of Departments and Data:

- School systems
  - Find content about teachers and administrators of the schools
- Department of Corrections
  - Find people in jails and that have had traffic tickets
- Office of Permits and Licenses
  - Who is remodeling their home?
- Transportation - Buses, trains

Local Level

The local level of US governments is the most specific to its small region. This category is really a "catch-all" for all the smaller governments under the state level. This includes, but is not limited to, county, city, village, and other government bodies that may have authority over an area.

Examples of the types of data we can find here are noted in the slide above.

© SANS Institute 2019

City of Charlottesville, VA

One example of local building permit searches at the local level is the City of Charlottesville, Virginia's Neighborhood Development Services

Search for a name and pull up the permits

Owner Search Results

Click the Owner's Property Address or the permit/project# to review detailed information.

111 RIVERBLUFF CIR	BN-14-0305	SMITH, JEFFREY D & ALISON I	55A088619	07-15-2014
111 RIVERBLUFF CIR	PN-14-0321	SMITH, JEFFREY D & ALISON I	55A088619	09-25-2014

111 RIVERBLUFF CIR

New Search

Click the permit/proj. # to view the application information

4 Records | Pages --> 1

Permit/Proj.#	Type	Sub Type	Description	Entered	Status
BN-14-0305	BUILDING	C101 - SINGLE FAMILY DETACHED	NEW SINGLE FAMILY RESIDENCE - DETACHED W/ PORCH IN REAR OF STRUCTURE	07-15-2014	CO ISSUED
EN-14-0448	ELECTRICAL	NEW SERVICE	NEW ELECTRICAL SERVICE FOR NEW SINGLE FAMILY RESIDENCE - DETACHED W/ PORCH IN REAR OF STRUCTURE	10-27-2014	COMPLETED

City of Charlottesville, VA

Taking a look at a single example of some of the local information that can be retrieved from local governments, we will use the City of Charlottesville, Virginia, USA as an example. They have a "Neighborhood Development Services" section of their web site (https://sec487.info/oj) where anyone can search for a person's name and pull up permits that that person filed. Alternatively, you can search on an address and find the owner and their permits.

Images from https://sec487.info/oj, January 5, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Link Aggregators

There are several web sites that have compiled extensive links to state and local government web sites and data

If you cannot find what you are looking for using search engines, try them

Link Aggregators

While not official government sites and not search engines, several link aggregator web sites can help you find links to state and local government web sites in case you cannot find them in other manners. Below is a list of sites and their URLs.

- NETR Online - https://sec487.info/d2
- Free Public Records Directory - https://sec487.info/d3
- BRBPub.com - https://sec487.info/no
- SearchSystems.net - https://sec487.info/d4
- PublicRecordCenter.com - https://sec487.info/nn

© SANS Institute 2019

International Governments

While the United States governments share a huge amount of data about their projects and citizens, other governments do too

Some governments are more reserved or protect their data, requiring authentication

Let's look at some examples

International Governments

Across the world, governments share a variety of data with people on the internet. Some, like the United States, share a large amount of content (as we have seen). Other governments may share an equal amount to the USA or less, depending on their systems, goals for sharing, and laws. If you are researching businesses or people in other countries, it may be useful to look for official government information about them to aid your investigation.

Image from https://sec487.info/ok, January 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Mexico Social Security

The Instituto Mexicano Del Seguro Social (IMSS) has an excellent search page for finding data on employees

Search for a name!

Export to Excel New Search Send by e-mail
77 records, showing 1 to 50 .

Position	First name	Surname	Second surname
SECRETARY...	JUAN CARLOS	VELASCO	PEREZ
HEAD OF A ...	JUAN MANUEL	ALMAZAN	PEREZ
HEAD OF A ...	JUAN LEONARDO	MENES	SOLIS
COORDINATED...	JUAN ANTONIO	MU & OZ	CONSUEGRA
DIRECTOR			
DIRECTOR			
COORDINAT			
COORDINAT			
SUBDELEGA			
SUBDELEGA			

Name (s)	JUAN JOSE
Surname	STAIRS
Second surname	GONZALEZ
Public Server Charge	HEAD OF AFFILIATION AND COLLECTION SERVICES IN EDO DE MEX ORIENTE
Type of Personnel	Trust
Job title / Salary Compensation	102
Market Stall	M33 HEAD OF DELEGATION SERVICES A
Direct phone	(01) 55765956
Switch	53591138-53593912-53594171-55766404
Extension number	28958-28959

SANS Open Source Intelligence (OSINT) Gathering and Analysis 163

Mexico Social Security

The Mexican government's Instituto Mexicano Del Seguro Social (IMSS) site has an advanced search (https://sec487.info/ol) feature that allows a user to type in a person's name and retrieve their government contact information, including name, title, office, salary compensation, and phone numbers.

Images from https://sec487.info/ol and https://sec487.info/om, January 5, 2019.

© SANS Institute 2019

Australian Business Registry

Switching to Australia's Business Registry

We can search for business names and retrieve owner, type, location, and other information

ABN	Name	Type	Location
45 004 205 449 Active	BUTTERCUP	Business Name	2113 NSW
91 155 659 677 Active	MISS BUTTERCUP	Business Name	2038 NSW
68 882 195 323 Active	BUTTERCUP BEARS	Trading Name	3754 VIC

Entity name	From	To
GOULD, MICHELLE ANNE	02 Aug 2002	(current)
VINCENT, MICHELLE ANNE	03 Aug 2001	02 Aug 2002
VINCENT, MICHELLE ANNE	17 Jul 2000	03 Aug 2001
VINCENT, MICHELLE ANNE	01 Jun 2000	17 Jul 2000
VINCENT, MICHELLE ANNE	01 Nov 1999	01 Jun 2000

Australian Business Registry

If you are looking into businesses registered in Australia, the Australian Business Registry database will be a valuable resource. Search for the name of a business and retrieve owner, location, and type information. You can then pivot on those pieces of data and dive deeper to gather more information.

Images from https://sec487.info/on and https://sec487.info/oo, January 5, 2019.

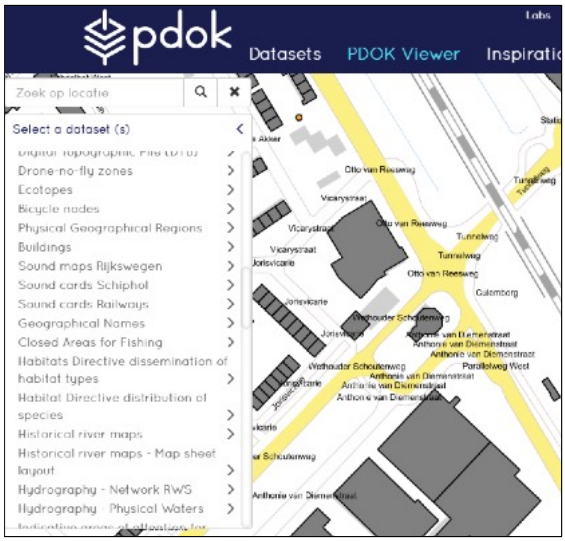
©2019 Micah Hoffman

© SANS Institute 2019

Publieke Dienstverlening Op de Kaart (PDOK) - Netherlands

Government GIS data in the Netherlands

Excellent mapping resource with names, regions, overlays of economic and other data



Publieke Dienstverlening Op de Kaart (PDOK) - Netherlands

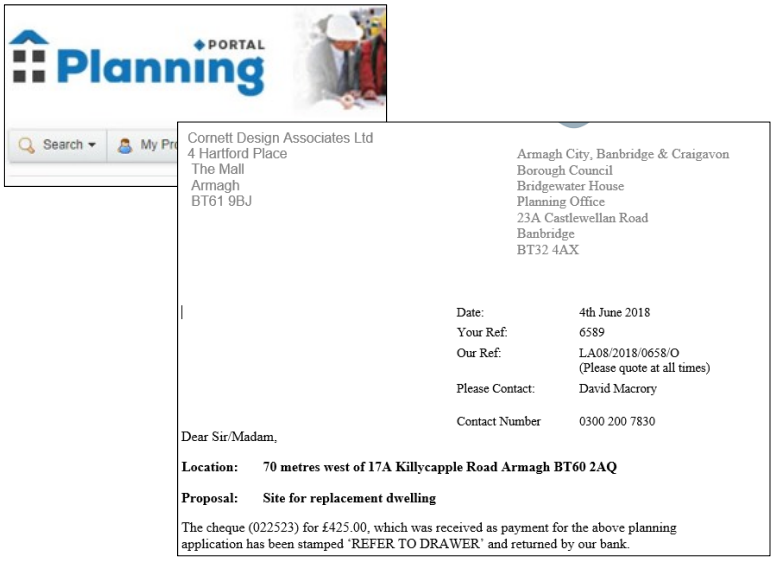
An excellent government GIS site in the Netherlands is PDOK. It contains a variety of "layers" that can be enabled to view a variety of data on the maps.

Image from https://sec487.info/op, January 5, 2019.

© SANS Institute 2019

United Kingdom Planning Portal

We can use a government search engine in the UK to find permits to modify structures, who modified what, and pull up digitized documents from conversations about the permits!



United Kingdom Planning Portal

This site is similar to other building and property remodeling sites that other governments have. This one provides not only the information about permits granted to people and companies to perform work on properties but also has a section of each entry where you can retrieve digitized documents that were sent or received while the permits were active. These documents, as shown above, can contain information about who performed what work for how much on which dates. Valuable data!

Images from https://sec487.info/oq and https://sec487.info/or, January 5, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Government Trivia"

This page intentionally left blank.

© SANS Institute 2019

Course Roadmap

- Day 1: Foundations of OSINT
- Day 2: Gathering, Searching and Analyzing OSINT
- Day 3: Social Media, Geolocation, and Imagery
- **Day 4: Networks, Government, and Business**
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

NETWORKS, GOVERNMENT, AND BUSINESS

1. Whois
2. IP Addresses
3. DNS
4. Findings Online Devices
5. Wireless Networks
6. Recon Tool Suites and Frameworks
7. Government Data
8. **Researching Companies**

This page intentionally left blank.

©2019 Micah Hoffman

© SANS Institute 2019

Business OSINT

Many OSINT investigations will require investigating businesses as our customers may be looking to:

- Invest in a company
- Partner with a company
- Merge or acquire a company
- Compete with a company
- Investigate possible illegal or unethical behaviors

When performing OSINT on a company, you will want to gather:

- Basic data (who, what, where)
- Profiles and news of the company
- Formal/government business filings about the company
- Public reviews/perceptions
- What business systems are accessible the internet

Business OSINT

In some of our OSINT assessments, we will need to gather data about businesses, organizations, and companies for our customers. It may be that the business of interest is the focus of the work, such as in cases where your customer is looking to grow their company by investing in, partnering with, merging with, or acquiring a company. Alternatively, your customer may wish you to examine a competitor's company and find what they are working on, who they are hiring, and what contracts they may have won. We also see cases where our clients want us to help them piece together content about possible wrongful, illegal, and unethical activities that a business may have performed.

Regardless of the "why" you are looking into a company, you will need to scour the data on the internet to gather certain data:

- Basic data about the company, such as who works there, what they do, and where their offices are
- Profiles of the company that are published by aggregator sites and news about the company, both positive and negative
- Formal, government submissions and business filings made about the company, if applicable.
- Any public reviews and perceptions about the company
- Finally, what systems the company has that are internet-accessible

© SANS Institute 2019

Business Basic Data - Who

Within the basic data we will need to gather, we need to answer the "who" questions

Some of those concern the company itself; others concern their customers, partners, and suppliers

General "who" questions to look for data on:

- Who runs the company?
- Who works at the company?
- Who are the customers?
- Who are the business partners?
- Who are the suppliers?
- Who are the people in the news?

Business Basic Data - Who

When a company is our target and, depending upon our client's OSINT goals, we may need to understand the answers to several "who" questions about that business. Questions that we will look for answers to include:

- Who are the leaders that run the company at the executive level and, if the company is large enough, lower levels as well?
- Who are the people that work at the company? What is their citizenship?
- Who are the customers of the company? This may or may not be obvious depending on the company we are examining.
- Who are the companies that the target company is partnering with?
- Who is supplying parts and/or labor to the target company?
- Who from the target company is making headlines in the news and what are those headlines about?

©2019 Micah Hoffman

© SANS Institute 2019

Business Basic Data - Where

Within the basic data we will need to gather, we need to answer the "where" questions

The "where" questions might be simple to answer if we are targeting a single small business or extremely complex with large, multinational corporations

General "where" questions:

- Where are the company's offices/locations?
  - Gathering addresses, phone numbers, and emails for each office can be helpful
- Where are they incorporated or registered?
- Where do they do business?
- Where are their internet-facing computer systems?

Business Basic Data - Where

The "where" questions not only concern the physical locations of the target business, but also cover legal and computer areas. Understanding the "where" helps us better understand breadth and scope. Questions that we will look for answers to include:

- Where are the company's offices or locations?
  - Here we may gather physical addresses, phone numbers, images from Google Street View, and email addresses.
- If the company is a type that requires registration or incorporation, where did they perform those official actions?
- Where does the company do business? This could be virtual if their business is focused on online commerce, or it could be across numerous countries around the world.
- Finally, we are most likely going to need to understand which computer assets and systems the company uses to conduct business.

© SANS Institute 2019

Business Basic Data - What

Within the basic data you will need to gather, we need to answer the "what" questions

Answers to these questions may overlap with the "who" and "where" questions

General "what" questions to look for data on:

- What does the company "do"?
  - The primary and secondary functions of the company
- What computer systems do they use and how?
  - Email, cloud, web, IM, billing
- What are the important business relationships and ties to and from the company?

Business Basic Data - What

We now examine the "what" questions, which add information about the goals of the business, how they conduct their business, and relationships to other entities. Questions that we will look for answers to include:

- What business functions does the company "do"?
  - What is it known for?
  - What does it do but is not known for?
  - Record the primary and other business functions of the business.
- Regarding computer systems, what do they use to conduct their business?
  - What providers or technologies do they use for email, cloud, invoicing systems? What about instant messaging and web hosting? Other online activities? Map them out to look for weakness and/or exposure.
- Look for the business relationships that the company and its employees have with other organizations. Those ties can prove powerful influencers in how they conduct business.

©2019 Micah Hoffman

© SANS Institute 2019

Business Profile Sites

Buzzfile is a business intelligence site showing data about companies and their leaders

Some sites focus on the financial aspects of the company and others on more general information

http://www.buzzfile.com

Displaying 1 - 37 of 37 companies

Action Format Filter

Company	Type	Street Address	City	State	Zipcode	Empl	Revenue
Facebook	Single	200 Jefferson Dr	Menlo Park	CA	94025	1	33,851
Facebook	Single	1101 Dexter Ave N	Seattle	WA	98109	2	110,294
Facebook Inc.	Single	551 5th Ave	New York	NY	10176	5	197,651
Opticians On Facebook	Single	3121 Bedlington Place	Southampton	PA	18966	3	82,957
Facebook, Inc.	Branch	225 Park Ave S	New York	NY	10003	2	N/A
Facebook Inc	Single	200 Share Way NW	Altoona	IA	50009	8	918,062
Facebook, Inc.	HQ	1601 Willow Rd	Menlo Park	CA	94025	800	40,653,000,000

SANS Open Source Intelligence (OSINT) Gathering and Analysis 173

Business Profile Sites

Our work of profiling businesses can be accelerated by using sites that aggregate data about businesses. Some of these web applications pull data from a variety of online locations and can compile a good overview of a company. Consumers sometimes rely on these sites when looking to invest in a company, or not, so some are more focused on the financial aspects of a business than on other parts.

This module of the course will focus on those profiling sites that provide meaningful data for free. If your team has money to purchase business-profiling information from vendors, you may be able to go further in your investigation more rapidly than relying on the free data.

The slide image is from one business profile company called Buzzfile. We will look more into this company in upcoming slides.

Image from https://sec487.info/fo, September 25, 2019.

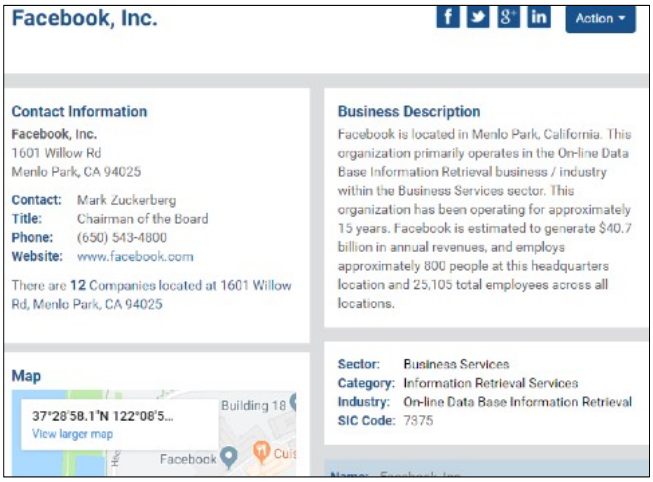
© SANS Institute 2019

Business Profile Sites - Buzzfile

The free Buzzfile.com web site compiles data about businesses

Public data about the location(s) type and categorization of the business are evident

Other helpful features include other businesses at location and other companies nearby



Business Profile Sites - Buzzfile

The Buzzfile site, shown in the slide, displays a summary businesses. Everything from who runs it to what it "does" to where it is located is displayed. Buzzfile can also give analysts a better understanding of other businesses that are registered at the target company's location or in the surrounding area.

In the detailed view of the company Facebook shown in the slide above, we see its corporate office location, contact information, and a general business description. Not shown on the slide but farther down on this web page is data about total revenue earned by this company, how many employees it reported, and demographics about the location of its corporate headquarters.

Image from https://sec487.info/fp, September 25, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Business Profile Sites - manta

The manta.com site also aggregates data about US small businesses

In addition to name, address, phone, bio, and web site, it also reports number of employees and company category (SIC and NAICS)

Whitefish Energy Holdings, LLC
530 19th Street # 200
Whitefish, MT 59937
Phone: (406) 471-4668
Web: www.whitefishenergy.com
Whitefish Energy Holdings, LLC is a
Single Location business.
Categorized under Electric Power L
in 2015 and incorporated in MT. Cu
revenue of 240745 and employs a s

Own This Business?

Detailed Information

Location Type	Single Location
Year Established	2015
Annual Revenue Estimate	240745
Employees	2
SIC Code	1623, Water, Sewer, Pipeline, and Communications and Power Line Construction
NAICS Code	237130, Power and Communication Line and Related Structures Construction
Business Categories	Electric Power Line Construction in Whitefish, MT Electric Power Line Construction Water, Sewer, And Utility Lines

SANS Open Source Intelligence (OSINT) Gathering and Analysis 175

Business Profile Sites - manta

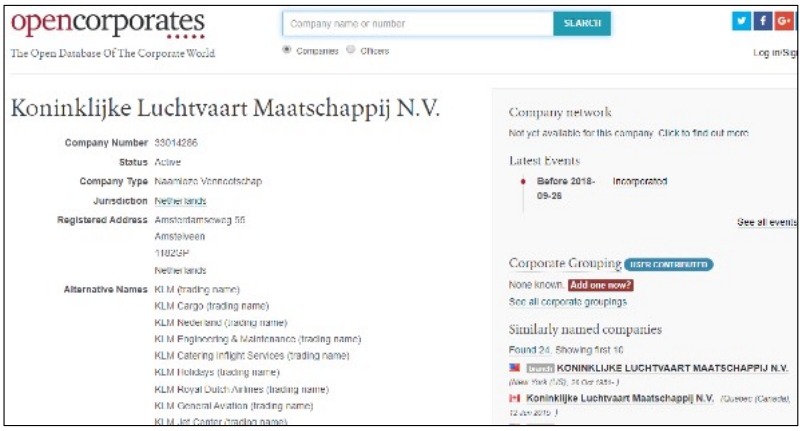
The manta.com web site aggregates a variety of data about US small businesses, as shown in the images above. We searched for the Whitefish Energy Holdings, LLC, company on the site. This company was of interest after it was awarded a $300 million contract to fix the electrical power grid in Puerto Rico following destruction from Hurricane Maria in 2017. The SIC (Standard Industrial Classification) and NAICS (North American Industry Classification System) categories help us understand the primary functions of the business.

Images from https://sec487.info/af, November 18, 2019.

© SANS Institute 2019

opencorporates.com

"As the largest open database of companies in the world, our business is making high-quality, official company data openly available." [1]



International repository of public corporate data

opencorporates.com

The main purpose of the opencorporates site is corporate data transparency. It collects, aggregates, and normalizes publicly available data from companies and organizations around the world. With either an anonymous or free registered account, you can search for information about a target person or entity. The site's true strength is in showing the user relationships to other organizations, branch offices, logos registered to the organization, and office locations.

Clicking on the "OPEN DATA JSON" link on the web page shows the user even more information about the target record, including creation and update dates of the record, the unique company number of the entity, and all of the data found in the human-readable web page. The corresponding API page for the above entry for the KLM Airline company shown in the above slide can be found at https://sec487.info/uh.

Reference:

[1] https://sec487.info/ug

©2019 Micah Hoffman

© SANS Institute 2019

Nonprofit Research

Some entities that you investigate may be nonprofit organizations

Details include overviews of the organization, scans of its tax documents, and sometimes brief analyses of the data

International and USA-only organizations

Nonprofit resources:

- charitynavigator.org (Int'l)
- guidestar.org (USA)
- projects.propublica.org (USA)
- gov.uk/find-charity-information (UK)

Nonprofit Research

Drawing data from nonprofit organizations' US tax returns, ProPublica (https://sec487.info/b8) maintains an excellent site that allows the examination of this data. The data goes back over 16 years, with detailed analysis of some of the nonprofits

Another, similar resource is GuideStar (https://sec487.info/ba), which not only dives into the tax returns of nonprofits, but also extracts contact information and tax ID numbers and posts nonprofit-submitted mission statements. Some of the data requires a login and may require purchasing reports, but much is free.

The Charity Navigator site (https://sec487.info/np) is similar to the above in the content it shares but also displays a rating on financial and accountability for the charity's resources and has an international scope.

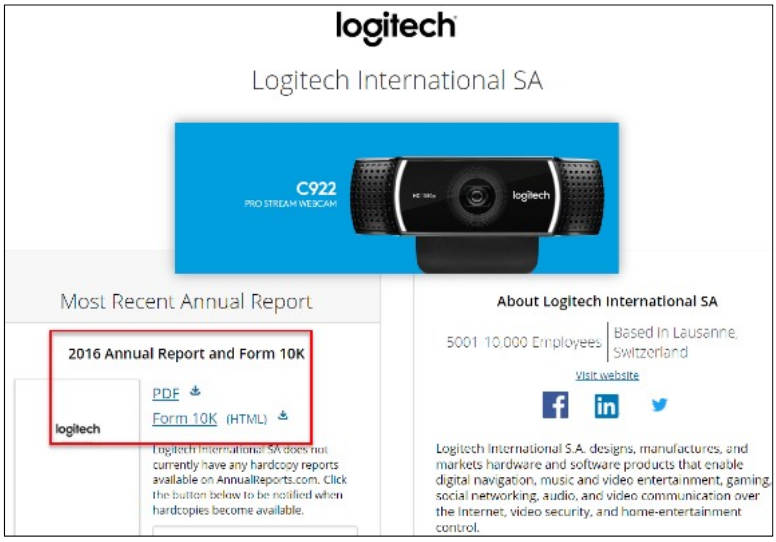
The United Kingdom has a site (https://sec487.info/sv) where people interested in examining nonprofit organizations registered in their country can perform their research.

© SANS Institute 2019

Business Filings - AnnualReports.com

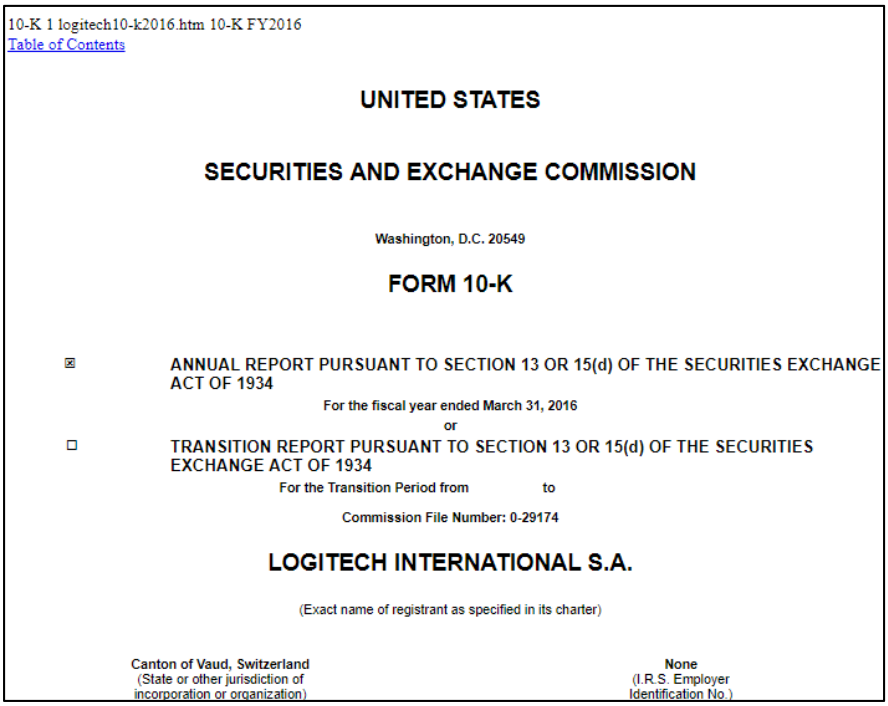
Companies, especially publicly traded ones, may need to submit corporate filings (SEC 10-K)

Sites like AnnualReports.com make it easy to find this data



Business Filings - AnnualReports.com

Companies of certain sizes, with certain corporate structures may need to file yearly reports about their organizations. In the slide above, we used the annualreports.com web site (https://sec487.info/nq) to retrieve the data about the company Logitech. While this page shows us basic data about the company, if we click on the links in the lower left of the page, our browser is redirected to the official United States Securities and Exchange Commission (SEC) 10-K form (https://sec487.info/fq and below).



10-K 1 logitech10-k2016.htm 10-K FY2016
Table of Contents

UNITED STATES

SECURITIES AND EXCHANGE COMMISSION

Washington, D.C. 20549

FORM 10-K

☒ ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934

For the fiscal year ended March 31, 2016

or

☐ TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934

For the Transition Period from to

Commission File Number: 0-29174

LOGITECH INTERNATIONAL S.A.

(Exact name of registrant as specified in its charter)

Canton of Vaud, Switzerland (State or other jurisdiction of incorporation or organization)	None (I.R.S. Employer Identification No.)

©2019 Micah Hoffman

© SANS Institute 2019

Business Filings - EDGAR

The United States SEC has been tracking company filings since 1934 and has 21 million posted

It is data on public companies and is intended on advising potential investors

You can search these filings for business and organizational content

Business Filings - EDGAR

US-based and international companies that are publicly traded (on stock markets and in securities such as mutual funds) must submit certain documents to the US SEC (Securities and Exchange Commission). These documents are searchable on the EDGAR web site (https://sec487.info/ag), shown above.

OSINT analysts can use the contents of these corporate filings to discover information about the business, its assets, management, and work. Below are some of the documents viewable about the McDonalds Corporation (https://sec487.info/ah).

SEC Home » Search the Next-Generation EDGAR System » Company Search » *Current Page*

MCDONALDS CORP CIK#: 0000063908 (see all company filings)
SIC: 5812 - RETAIL-EATING PLACES
State location: IL | State of Inc.: DE | Fiscal Year End: 1231
(Assistant Director Office: 5)
Get insider transactions for this issuer.
Get insider transactions for this reporting owner.

Business Address
ONE MCDONALD'S PLAZA
OAK BROOK IL 60523
630 623 3000

Mailing Address
ONE MCDONALD'S PLAZA
OAK BROOK IL 60523

Filter Results: Filing Type: Prior to: (YYYYMMDD) Ownership? include exclude only Limit Results Per Page 40 Entries Search Show All

Items 1 - 40 RSS Feed Next 40

Filings	Format	Description	Filing Date	File/Film Number
10-Q	Documents Interactive Data	Quarterly report [Sections 13 or 15(d)] Acc-no: 0000063908-17-000053 (34 Act) Size: 7 MB	2017-11-02	001-05231 171171441
8-K	Documents	Current report, items 2.02 and 9.01 Acc-no: 0000063908-17-000046 (34 Act) Size: 2 MB	2017-10-24	001-05231 171149969
8-K	Documents	Current report, items 7.01 and 9.01 Acc-no: 0000063908-17-000041 (34 Act) Size: 662 KB	2017-09-21	001-05231 171096078

© SANS Institute 2019

International Registries and Searches

There are web sites that have links to registries of businesses across the world

Find the place the company has offices and look for documents

These sites sometimes catalog names, addresses, and phone numbers and other times documents exposed on the net

Some web sites have databases with references to companies from across the world

They may not be official government sources, but these sites may give you access to data you need

International Registries and Searches

For international and multi-national businesses, we can use catalog sites to find data. Some of these web applications are similar to a phone book, where you enter a business name and a country, and it will show you the address, phone, and web site of the company. Others contain databases that can be searched for relevant content.

As with many other web sites on the internet, when you search them, you get what is in the database. We need to examine the content skeptically by asking ourselves analytic questions such as:

- Is the content factual?
- Is it still reliable?
- When was it updated?
- Where did it come from?

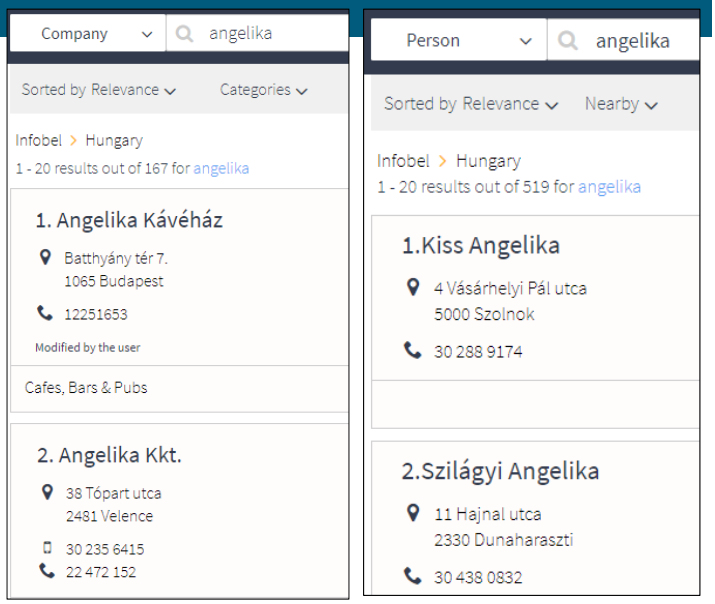
©2019 Micah Hoffman

© SANS Institute 2019

Infobel

Companies or people can be searched by name or category

Shows phone numbers, addresses, and map when details are requested



Infobel

Infobel (https://www.infobel.com/) is a Belgian company that hosts a worldwide database of companies and people. The company states on its web site that it has "the largest databases on the market with more than 186 million businesses and 350 million telephone subscribers worldwide."[1] This is one of those applications that is similar to a phone book (or https://www.whitepages.com/). You enter the country and the type of records that you want to search (people or companies), and it shows the results.

In the images above, we searched in Hungary for a business with the name Angelika and then a person with that name. The results are shown and, for those with additional information, we can see those details by clicking the "More Information" button.

Images from https://sec487.info/am and https://sec487.info/ak, September 25, 2019.

Reference:

[1] https://sec487.info/ns, November 10, 2017.

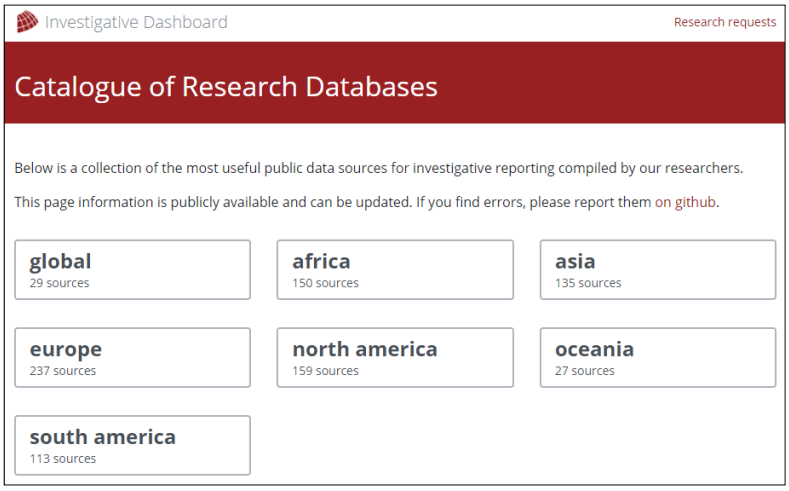
© SANS Institute 2019

Investigative Dashboard

Databases for journalists (and OSINTers!) to search

Focused on known databases and leaked/public documents

Search for company, person, or keyword



Investigative Dashboard

Operated by the Organized Crime and Corruption Reporting Project, the Investigative Dashboard site and its related https://sec487.info/nt site are focused on sharing information that could be used to investigate misconduct, crime, and corruption around the world. Its online databases contain tagged information by country and topic and allow keyword searching.

Access to some of the content may be restricted to paid or registered users, but much of the site links to free documents and content.

Images from https://sec487.info/an, September 25, 2019.

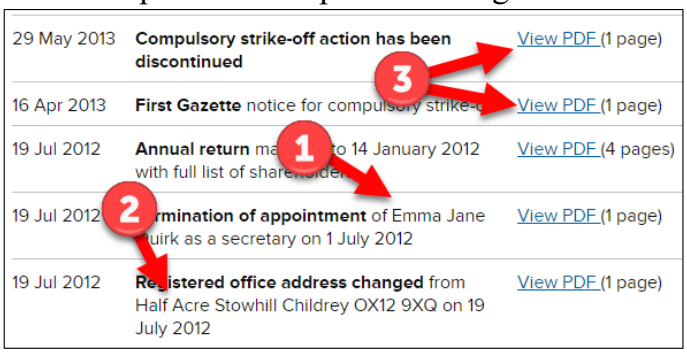
©2019 Micah Hoffman

© SANS Institute 2019

Gov.uk Companies House

Companies House allows the searching of companies that have registered with the UK government

https://beta.companieshouse.gov.uk



You can retrieve:

- People names (arrow 1)
- Locations (arrow 2)
- Historical scanned documents (arrow 3)

Gov.uk Companies House

Companies that register to do business in the United Kingdom will appear in the Companies House UK government web site. This site is rich with information about businesses, where they do business, and who runs them, and it contains historical filings that go back decades.

In the above slide, we pulled up the historical records of the Regulus Limited company, which was dissolved 31 March 2015. We see the names and locations in the past entries, but there is other data not shown in the slide that is on the site. There is information about the officers and directors of the company, such as Ian Hugh Quirk who was the last director of the company. The site shows his country of residence, nationality, and date of birth (month and year).

Image from https://sec487.info/ud, September 25, 2019.

© SANS Institute 2019

Gov.uk Overseas Registries

Not just for UK data, this web site has a solid set of links to business registries in many countries

Links take you off the site to live web sites with the content

Gov.uk Overseas Registries

This United Kingdom government web site (https://sec487.info/ao) is focused on cataloging overseas business registries. The page is updated regularly and has most of its links pointing to official government web sites in other countries (and those generally do not change URLs often). If your OSINT investigation takes you into the international world of businesses, these links can be helpful.

Images from https://sec487.info/ao, September 25, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

C-Level Analysis

Let us shift the discussion from the company to those people in power who run the organization

These sometimes powerful people have connections to others

For publicly traded companies, there is a wealth of data about the upper tier of management

Creating profiles of the C-level people may include:

- Relationship mapping
  - Who do they know and how?
- Compensation analysis
  - How much money do they report making?
- Social media analysis
  - Where are their social media accounts and what are they posting?

C-Level Analysis

Now that we understand about the business entity (what it does, with whom, and how), we should move to examining the upper-level management (sometimes referred to as the C-Suite or C-Level[1]) that runs the business. Analyzing who these people are, who they know, what networking groups and clubs they belong to, their compensation, and their social media personas can lead to important discoveries.

In many countries, businesses listed on a stock market disclose information about their company to government regulators. This data usually contains content about who the upper management is and their compensation packages. We might have seen these documents while researching the overall company in the previous section.

Reference:

[1] https://sec487.info/aq

© SANS Institute 2019

Executive Business Relationships

"It's not what you know, but who you know"

This axiom was first posited in 1914[1] and it is still true today

Business relationships and connections span across countries, businesses, and groups

Understanding how people are connected can lead to other businesses and people who we might be relevant to our work

Let's examine some web sites that can automate some of these relationship maps using OSINT

Executive Business Relationships

Many people have heard the old axiom "It's not what you know [that will get you ahead in this world], but who you know." Business is about relationships, and in discovering and mapping those relationships, we can better serve our customers. While we will focus on executives and their business networks, relationship mapping of social media networks, family members, and groups is an essential skill of OSINT analysts.

Reference:

[1] https://sec487.info/ar

©2019 Micah Hoffman

© SANS Institute 2019

Corporation Wiki

The Sagewire Research LLC created the Corporate Wiki for transparency in US businesses

It displays connections and relationships between people and companies that researchers may not have known about

Suggest registering for a free account to view and export data

Corporation Wiki

The corporationwiki.com web site hosts data about companies, people, and their relationships. Through an easy interface, you can explore these connections to find meaningful links. This site was created by the Sagewire Research LLC to foster transparency in business and government. While much of the data is accessible without registering for an account, it is highly beneficial to sign up for one since you can then gain access to exporting the data in Excel and PDF formats and can use the interactive mapper. Concerned about your privacy and don't want to register? This site does not validate email addresses when you sign up, so using an "@example.com" domain email account or another throw-away email account will work fine to get you access.

© SANS Institute 2019

Corporation Wiki - Elon Musk

Let's take a look at Elon Musk's profile on the site as an example

In the overview we get name, company (not Tesla or SpaceX?), phone, and a summary of the person

Corporation Wiki - Elon Musk

As an example, we will look up Elon Musk and see what data is revealed. First thing we notice is that the record does not show Tesla, Inc., or SpaceX as Elon Musk's main businesses. Instead, it reports "Chairman for Solarcity Corporation." When we scroll down the page, we will see his relationship to these other, more prominent companies noted.

Here we get the scope of what we are going to see: 15 companies, 63 connections to other people, and 10 locations. This should be interesting!

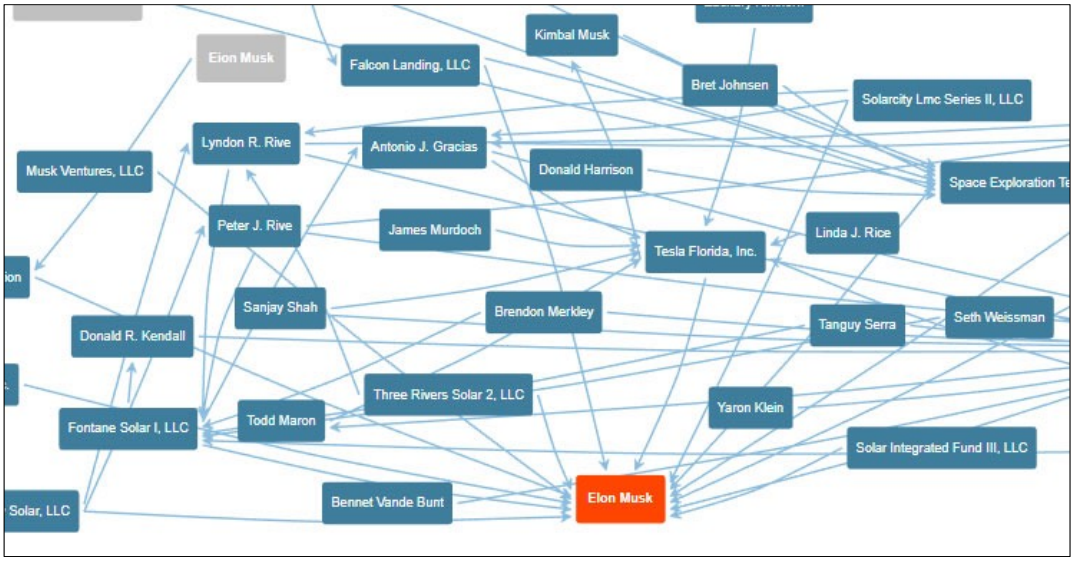
Create a free user account on the platform, and you can export this data to an Excel file. This will export what is displayed on the screen in the appropriate formats.

Image from https://sec487.info/at, September 23, 2019.

©2019 Micah Hoffman

© SANS Institute 2019

Relationship Map



Relationship Map

Scrolling down the page, there is a section that displays an auto-generated relationship mapping of our target (Elon Musk) to all his companies and connections. While this data is displayed on the site, it may be more useful to our investigation to download the data in an Excel document and upload to Gephi or another link analysis tool.

Clicking on any one of the boxes will redirect our browser to Corporation Wiki's page about that entity or person—a kind of pivoting around that data point. The map does not show what the relationships are (owner, employee, contributor, founder, etc.) between objects, merely that they exist.

Image from https://sec487.info/at, September 23, 2019.

© SANS Institute 2019

Companies and Connections Tables

Companies for Elon Musk — Excel

Name	Status	Incorporated	Key People	Role
Fontane Solar I, LLC	Active	2010	27	Chairman
Solarcity Corporation	Active	2006	25	Chairman
Solarcity Lmc Series II, LLC	Active	2014	19	Chairman
Solarcity Finance Company, LLC	Active	2013	17	Chairman
Three Rivers Solar 2, LLC	Active	2014	17	Chairman
Andrews County Solar, LLC	Active	2015	15	Chairman
Space Exploration Technologies C...	Active	2002	7	President
Tesla Florida, Inc.	Active	2003	26	President

« 1 2 » Showing 8 records out of 15

Connections for Elon Musk — Excel

A connection is made when two people are officers, directors, or otherwise associated with the same company. Elon has sixty-three known connections and has the most companies in common with Brad Buss.

Brad Buss 7

Antonio J. Gracias 7

Antonio J. Gracias has the following companies in common with Elon Musk

Name	Status	Incorporated	Key People	Role
Solarcity Corporation	Active	2006	25	Director
Tesla Florida, Inc.	Active	2003	26	Director
Fontane Solar I, LLC	Active	2010	27	Director
Solarcity Finance Company, LLC	Active	2013	17	Director
Solarcity Lmc Series II, LLC	Active	2014	19	Director
Three Rivers Solar 2, LLC	Active	2014	17	Director
Andrews County Solar, LLC	Active	2015	15	Director

John H N Fisher 6

SANS Open Source Intelligence (OSINT) Gathering and Analysis 190

Companies and Connections Tables

Continuing down the Elon Musk page, we can see several tables containing the data that was used to generate the relationship map. On the left is a table of companies Mr. Musk is associated with, along with the role he has at those businesses, and on the right is an image of the connections section of the page. This section illustrates how Elon Musk is related to another person. Mousing over an entry here (as we did for Antonio J. Gracias) causes the details to drop down under it.

Clicking on either the company entry or a person entry will pivot your browser to the record Corporation Wiki has for that entity or person. Additionally, the "Excel" button is visible in the upper right of each picture, reminding us that we can download this data for post-processing in our own applications (shown in image below).

Image from https://sec487.info/at, September 23, 2019.

	A	B
1	Link	Connection
2	https://corp.wiki/e/68841642	Brad Buss
3	https://corp.wiki/e/57050794	Antonio J. Gracias
4	https://corp.wiki/e/61047958	John H N Fisher
5	https://corp.wiki/e/123792879	Donald R. Kendall
6	https://corp.wiki/e/148627561	Brendon Merkley
7	https://corp.wiki/e/61047934	Nancy E. Pfund
8	https://corp.wiki/e/38254496	Lyndon R. Rive
9	https://corp.wiki/e/38254500	Peter J. Rive

©2019 Micah Hoffman

© SANS Institute 2019

Known Addresses and Sources

Known Addresses for Elon Musk

- 1 Circle Star Way San Carlos, CA 94070
- PO Box 10195 Palo Alto, CA 94303
- 3055 Clearview Way San Mateo, CA 94402
- 1 Rocket Rd Hawthorne, CA 90250
- 393 Vintage Park Dr San Mateo, CA 94404
- 1105 Bel Air Pl Los Angeles, CA 90077
- 6800 Dumbarton Cir Fremont, CA 94555
- 1050 Bing St San Carlos, CA 94070
- 1053 W Grand Ave Chicago, IL 60642
- 1050 K St NW Washington, DC 20001

Sources

Florida Department of State
Data last refreshed on Sunday, August 18, 2019

California Secretary of State
Data last refreshed on Sunday, September 22, 2019

California Secretary of State
Data last refreshed on Sunday, September 22, 2019

Texas Secretary of State
Data last refreshed on Friday, September 20, 2019

SANS Open Source Intelligence (OSINT) Gathering and Analysis 191

Known Addresses and Sources

Continuing to scroll down the page, we find the addresses section and the sources. Addresses may be an important piece of your investigation and should be collected and pivoted upon to see if they are home or business addresses (most here are business addresses). Clicking on the links to an address will redirect your browser to a page showing what businesses appear at those addresses. This could be used to discover related businesses.

Keep in mind that the documents you find for registered businesses may be registered in the name of an agent. The addresses, phone numbers, and other data displayed could be for a Certified Public Accountant (CPA) or lawyer who is a representative of the business and not data for where the company does business or the person you are researching lives.

Finally, we have the sources that the site used to gather all of its data from. Displaying this data to the user is important for requesting official documents and in understanding the authenticity of the data displayed.

Image from https://sec487.info/at, September 23, 2019.

© SANS Institute 2019

Corporation Wiki for Businesses

Key People Who own Space Exploration Technologies Corp. Excel

Name	
Gwynne Shotwell 1 ~ Background Report ~	President
Elon Musk 14 ~ Background Report ~	President, CEO, Chief Executive Officer, Chairman
Bret Johnsen ~ Background Report ~	CFO, Chief Financial Officer
Tim Hughes ~ Background Report ~	Treasurer, Secretary, Assistant Secretary, Vice President
Donald Harrison 27 ~ Background Report ~	Director
Kimbal Musk 2 ~ Background Report ~	Secretary
Gaynne Shotwell ~ Background Report ~	President

Other Companies for Space Exploration Technologies Corp.

Space Exploration Technologies Corp. is listed as an officer of another company.

Name	Status	Incorporated	Key People
Falcon Landing, LLC	Active	2011	3

California Secretary of State

Filing Type:	Statement & Designation By Foreign Corporation
Status:	Active
State:	California
Foreign State:	Delaware
State ID:	02414622
Date Filed:	Monday, May 6, 2002
Registered Agent	Elon Musk

SANS Open Source Intelligence (OSINT) Gathering and Analysis 192

Corporation Wiki for Businesses

We just looked at Elon Musk on the corporationwiki.com site, but the web site also can help with your business-related searches. Let's pivot from Elon Musk to Space Exploration Technologies Corporation, more commonly known as SpaceX. The page for it shows everything from who the key people are, to addresses where it does business, to other companies that this business may have a role in, as well as those official corporate filing documents where much of this data was retrieved from. If the company has trademarks, there is a section on the page for that as well.

Overall, the Corporation Wiki provides meaningful, reliable data about people and businesses and gives us pivot points to find even more.

Image from https://sec487.info/au, September 23, 2019.

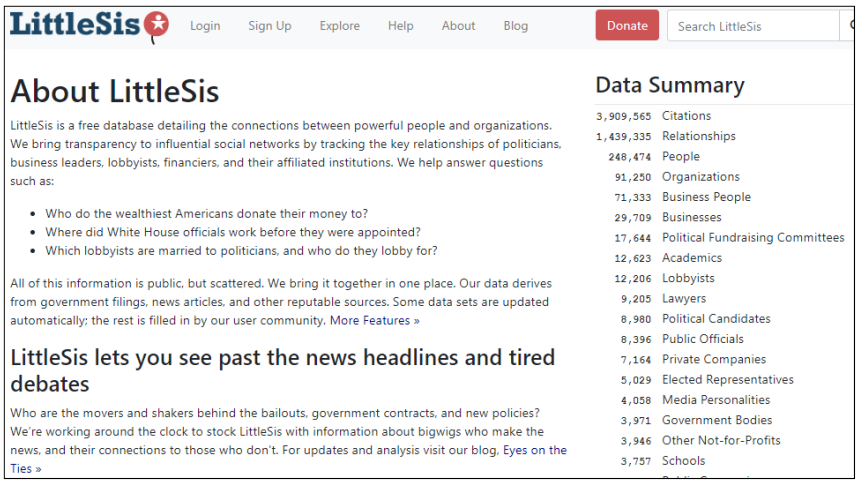
©2019 Micah Hoffman

© SANS Institute 2019

LittleSis.org

Another site dedicated to publishing data in support of factual journalism is LittleSis.org

Some content such as the "Maps" is user-created and not useful



LittleSis.org

The non-profit group Public Accountability Initiative created the LittleSis.org site in 2009. It provides both data and analysis about people and businesses. It allows people to create accounts, contribute to groups, and generate maps based upon the data in the site. This is something to keep in mind as you use this resource; since anyone can become an "analyst" on the site, we do not know who is creating the maps and content, and some of it may contain bias or other analytical fallacies.

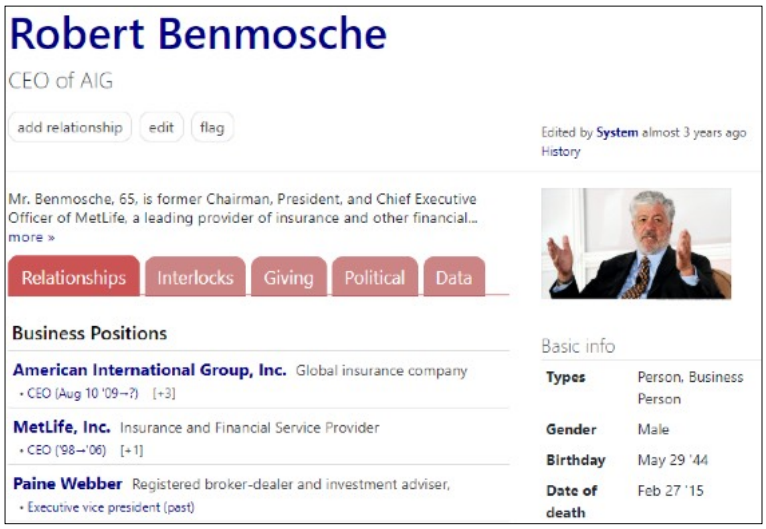
Image from https://littlesis.org/about, September 23, 2019.

© SANS Institute 2019

LittleSis and Robert Benmosche

LittleSis shows data for who the person knows, where they donate, and the person's possible political affiliations

We also get biographical information about the person



LittleSis and Robert Benmosche

Per the example shown in the image above, we can get a general biography of the target[1] from his page. There are five main tabs of data:

- The Relationships tab highlights businesses the target has an interest in, family, memberships, education, holdings, places he made donations, and any government positions he may hold.
- The Interlocks tab shows other people who have a work or other relationship with the target.
- The Giving tab shows what political organizations the target donated money to.
- The Political tab shows a graph of every person/group the target donated money to and what their political affiliations are. Using this data, we may be able to ascertain what the target's political affiliations are.
- The Data tab is an excellent place to view or download all the data that was used to create the bio. We can, once again, download to a CSV file and import into Gephi or another tool for post-processing.

Not shown in the image above but also appearing on the page is a list of the sources (including links to them) that were used to collect the data shown on the target's page. This is helpful for pivoting to other sources and also in understanding the validity of this content.

Image from https://sec487.info/ue, September 25, 2019.

Reference: [1] https://sec487.info/uf

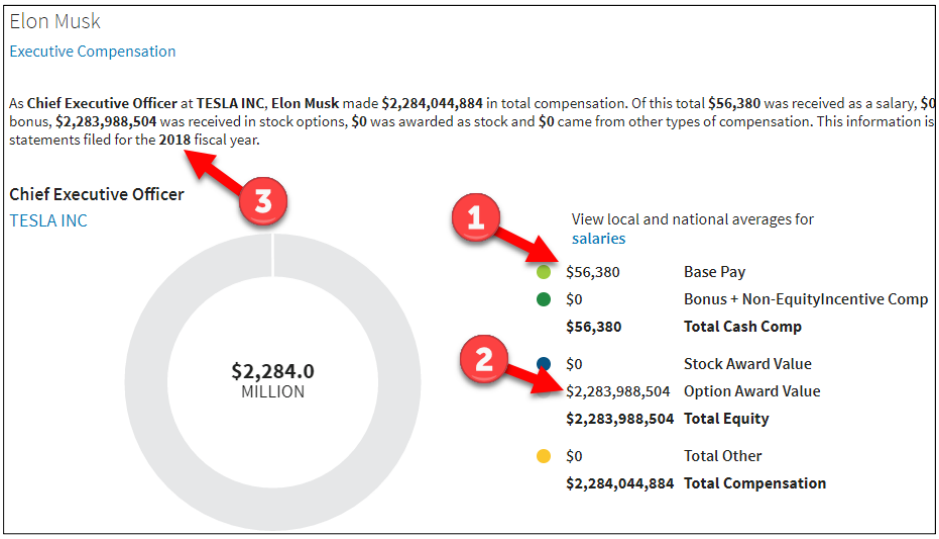
 ©2019 Micah Hoffman

© SANS Institute 2019

Executive Salaries - Salary.com

Many executive salaries and compensation packages need to be reported in regulatory documents

Salary.com collects them



Executive Salaries - Salary.com

Moving from relationships to compensation, our customers may want data on what salaries and bonuses their competitors' executives are earning. If the company we are researching is a publicly traded one, we can go back to the corporate filings and look for ourselves or we can use a site such as salary.com.

On the salary.com web site, we can look up an executive by name or we can browse by companies to find what the top people in an organization are being paid. Taking a look at the image above, we see that Elon Musk, CEO and Chairman of Tesla, Inc., received a salary of $56,380 in 2018. Doing more OSINT, we find an article (https://sec487.info/ay) describing that this salary is what someone who makes minimum wage in California makes...and that Mr. Musk never claims his salary. Why? Looking at the image we can see that he received over $2.2 billion in stock options. He also may have income from his other businesses.

Image from https://sec487.info/ax, September 25, 2019.

© SANS Institute 2019

What Do Employees Say?

Examine what current and past employees of companies divulge at company rating sites

Keep in mind that this population of submitters may be biased in their subjective submissions (disgruntled, overly liking, etc.)

Sites where employees can review their employer include:

- Careerbliss.com
- Comparably.com
- Glassdoor.com
- Indeed.com

Scour the content and then apply your analysis to temper the subjective nature of the posts

What Do Employees Say?

Everyone on the internet has an opinion, and many will be happy to share theirs with you. The focus of our research here is to glean what employees or former employees are saying about their companies. To do this, we could use Google searches and such, or we could visit some of the sites listed above since they are focused on providing platforms for people to review their employers.

When reading opinions and user-supplied content, we always consider the motivations and biases of the submitter. Perhaps the person was fired for incompetence or insubordination and they left a scathing review of the company that may not reflect the opinions of others at the business. Instead of taking each of the entries as valid, look for trends in the data.

©2019 Micah Hoffman

© SANS Institute 2019

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Business OSINT"

This page intentionally left blank.

© SANS Institute 2019

SEC487 Day 4 Summary

Today we explored IP addresses, whois, DNS, recon suites, government data, and business OSINT.

Tomorrow we will continue by diving into the dark web and international OSINT, and we will have a solo CTF!

This page intentionally left blank.

©2019 Micah Hoffman