

## OSINT-Investigation – Lab 1.0

*The purpose of this guide is to perform an open-source intelligence investigation using manual tools, automated tools, and keen analytical skills! This lab requires Internet access.*

**IMPORTANT:** We will be performing “passive” analysis on our target/s, as we do not have permission to perform “active” analysis.

### OSINT Investigation Case Study:

You have been tasked with investigating the Internet attack surface for a global organization via “passive” open-source intelligence gathering.

You need to answer the following questions:

1. What are the primary geographical regions for Internet Infrastructure [country]
2. What Cloud Services are in use [m365, azure, aws, etc.]
3. What technologies are in use [firewalls, web servers, operating systems]
4. What remote-connectivity endpoints are identifiable
5. Are any vulnerabilities known/seen publicly
6. Are any leaks/breaches known/seen publicly
7. Is there any evidence of fraudulent/unauthorized public systems/services

**NOTE:** The following sections of the lab will correspond to these questions.

#### A. Select a Target Business

Select a target business. Here are some suggestions, but you are welcome to any one or more than one for investigation, with the IMPORTANT caveat that we are performing “passive” analysis only!

Please note that the organization size/scope dramatically impacts the size/scope of your analysis. Too small and there is minimal open-source data (OSD), too big and the volume of OSD can be overwhelming.

Example Targets:

- Ageas [insurance company]
- Infobel [tele directory]
- Luminus [electricity]
- Fluxys [natural gas]
- Tapptic [software]

#### B. Prepare for Manual Investigation

You are welcome to perform browser-based investigation steps from your HOST computer or from within one of your class VM's.

For maximum value, OSINT should be a bit of “science” and a bit of “art,” meaning a set of familiar/capable tools are essential, a basic methodology is beneficial, while also leveraging creativity, flexibility and “human” intelligence to think like an attacker, discover nuanced relationships, correlate data/sources, etc.

The following lab sections are suggestions for a basic methodology. Your analysis should be “iterative” (step 1, step 2, step 3...learn a few things...return and repeat step 2, step 3, step 4...etc.).

## 1.0 - Primary Geographical Regions

Often, *Wikipedia* is a great place to get a high-level overview of your target. We'll use "BESIX" for our example target. [Wikipedia.org]

Target Business: \_\_\_\_\_

Target Business Primary Domain (website): \_\_\_\_\_

Type of Business: \_\_\_\_\_

*DNSDumpster* is a great way to learn about data that may help with multiple OSINT questions/queries, and it provides a map-of-the-world with "green" areas representing areas with "target-domain" related Internet infrastructure [dnsdumpster.com]:



**GeoIP of Host Locations**

Target Primary Geographical Regions:

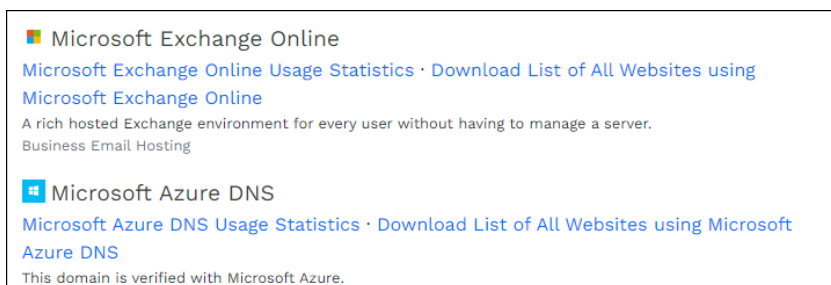
## 2.0 - Cloud Services in Use

*DNSDumpster* often provides clues about "target" cloud infrastructure via name review and/or ASN review. Look for entries like the following [dnsdumpster.com]:

- MX Record: \*.protection.outlook.com (m365 email)
- ASN: microsoft-corp-msn-\* (m365/azure)
- IP Block Owners: Microsoft, Proximus, AWS, Azure, etc.

Review the "Hosting (IP block owners)" diagram at the top of your results. Then review the MX record, Host Records (A), etc.

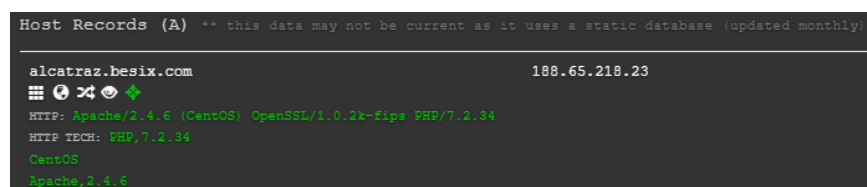
*Builtwith.com* is designed to show you website interactions and technologies, but it can also show things like "Microsoft Azure DNS" and "AWS CloudFront," revealing "cloud-service providers." [builtwith.com]



Target Cloud Service Providers:

### 3.0 - Technologies in Use

Once again, *DNSDumpster* often reveals information about web servers, underlying operating systems, and even other services accessible from the Internet. Under the “Host Records (A)” section, review the “green” entries: “HTTP:”, “HTTP Tech:”, “SSH:”, etc.



```
Host Records (A) ** this data may not be current as it uses a static database (updated monthly)

alcatraz.besix.com 188.65.218.23
[Icons: DNS, Ping, Traceroute, etc.]
HTTP: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.2.34
HTTP TECH: PHP/7.2.34
CentOS
Apache/2.4.6
```

**NOTE:** This reveals web-server software type/version (Apache 2.4.6), underlying OS (CentOS), web-server applications (PHP 7.2.34). This could be very useful for future/further analysis!

As per “step 2.0,” *Builtwith.com* gives more specific insight into web technologies in use, like “asp.net” or “PHP” or “WordPress” (a somewhat infamous web platform, widely used and often using vulnerable plug-ins!).

Technologies in Use (Web Server):

Technologies in Use (Operating Systems):

Technologies in Use (Web Server Applications/Components):

Technologies in Use (Other):

### 4.0 - Remote Connectivity/Endpoints are Identifiable

In this step, we are looking for names, services and/or TCP ports that identify remote-connectivity solutions, like a VPN or a remote-access web portal. These are often named to be easily recognizable by “authorized” users.

Review *DNSDumpster.com* output for names like “remote.company.com,” “vpn.company.com,” and any other DNS record that falls into possible remote-connectivity categories.

*DNSDumpster* also gives us information about IP addresses/ranges in use by the company. If there are multiple “A Record” entries in an adjacent address range, we can pivot to additional “IP Searches” to review TCP ports related to remote connectivity (port 22 for SSH, port 3389 for RDP, etc.).

**EXAMPLE:** “besix.com” has servers at 213.246.249.207 and 213.246.249.243, which is a reasonable indicator that they own/use the range “213.246.249.207 to 213.246.249.243.”

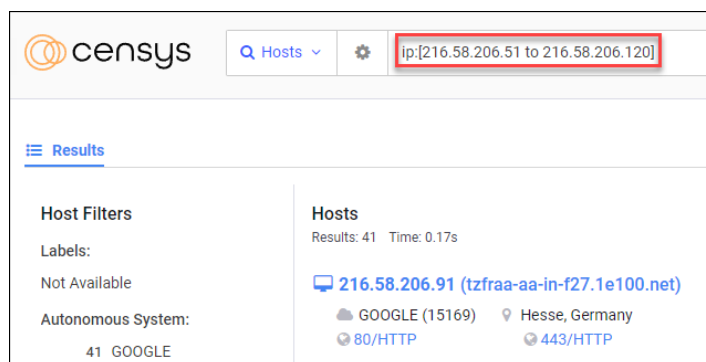
NOTE: If we identify “interesting/concerning” findings in the aforementioned range, we’ll definitely want to CONFIRM the addresses are related to our target (not just assume based on inference)!

Let’s see what we can learn based on IP address/ranges via *search.censys.io*. To search for a range of addresses, select “Hosts,” then enter the query:

NOTE: Change “starting IP” and “ending IP” to your desired search range, eg “1.1.1.1 to 1.1.1.10” (this is just an example range)

*ip:[1.1.1.1 to 1.1.1.10]*

NOTE: Run this query first to make sure you get search results returned, then run the next query to “test” for port 80 (standard web port), then we’ll query for “remote connectivity” ports.



*ip:[1.1.1.1 to 1.1.1.10] and services.port:80*

NOTE: You should get returned results for this, as “port 80” is most likely open on web servers.

Now run the following queries, looking for “remote-connectivity” ports (22 for SSH, 3389 for RDP, 445 for SMB).

*ip:[1.1.1.1 to 1.1.1.10] and services.port:22*

*ip:[1.1.1.1 to 1.1.1.10] and services.port:3389*

*ip:[1.1.1.1 to 1.1.1.10] and services.port:445*

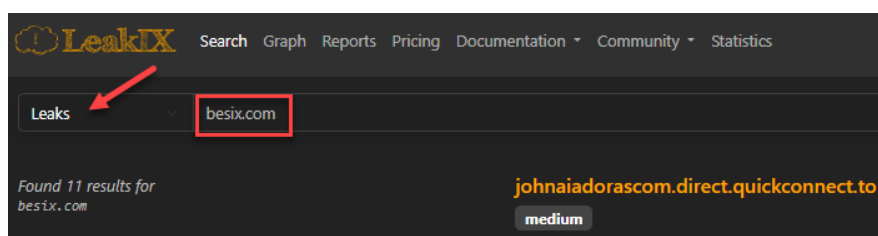
Target Remote Connectivity Endpoints Identified:

## 5.0 - & 6.0 Vulnerabilities/Leaks Identified

The next two steps/questions rely on some flexibility/creativity! To identify public information about vulnerabilities and/or leaks, we can use the information we’ve gathered so far and visit a few different sites, using “company name,” domain name, IP address, etc.

Let’s visit *leakix.net* where, as the name implies, we can lookup “leaks” for a company, domain, IP address.

EXAMPLE: We’ll search for besix, besix.com, etc. Select “Leaks,” and then input your search term:

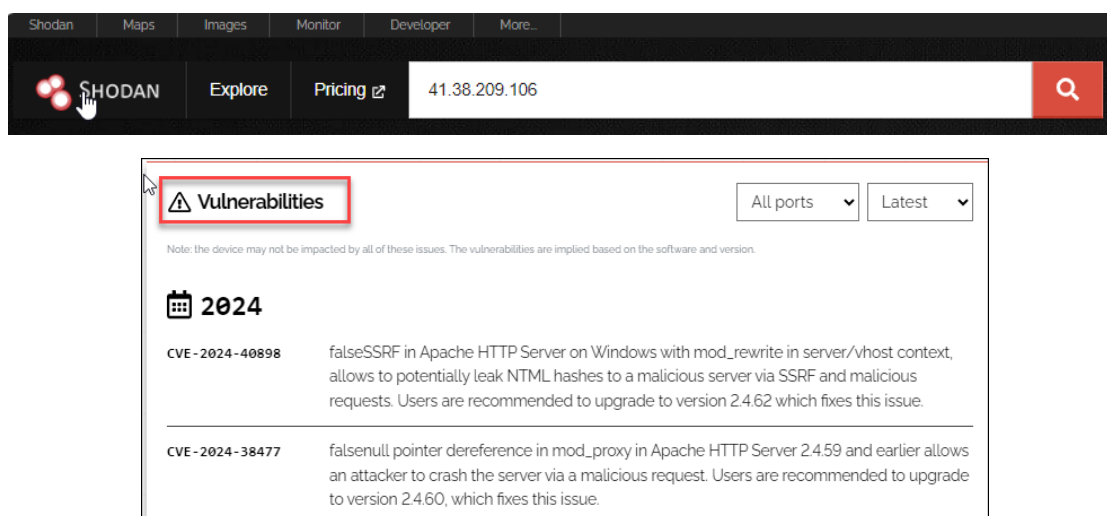


**NOTE:** We have 11 returned results. What is a “leak” exactly? It just means some data exposure related to our search terms. In the first example for besix.com, the severity rating is “medium,” and the on-screen snippet shows a URL ending in besix.com. Each of these would need to be fully researched for validation, target applicability, severity, etc.

You can enter the company name, company domain, one or more IP addresses and continue to search *leakix.net* for your target.

If you find an IP or domain name of interest, you can also visit *shodan.io*, enter the domain name or IP address and see a slightly different view of “evil search engines.” [*shodan.io*]

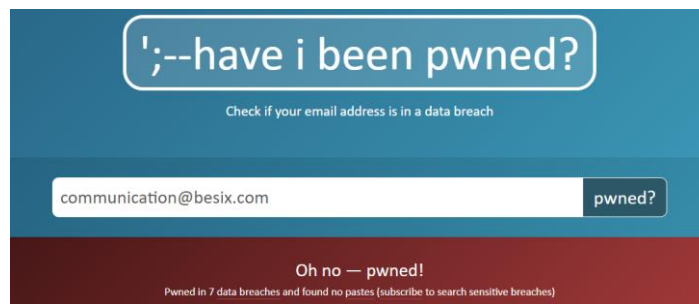
EXAMPLE: When I take the IP address from the second “leak” result on *leakix.net* and search for it via *shodan.io*, there are possible “vulnerabilities” listed, requiring review!



Again, be creative in how you “query” Shodan, Leakix, Censys, etc. to see if you can identify anomalies, concerns, exposures, etc.

For our last “leak/breach” analysis, let’s visit *haveibeenpwned.com*. We’ll need to have discovered an email address in our previous research. This is often as simple as visiting the company website, clicking “about,” and looking for a generic email. We’ll enter that in the search at *haveibeenpwned.com*, which will search a LARGE dataset of publicly-known breaches to tell us whether the email was potentially exposed in a data breach!

EXAMPLE: The besix.com site lists “communication@besix.com” on their “contact us” page, so we’ll search for that:



“Oh no - pwned!” - You can drill down and learn a little more about what breaches may have involved this account.

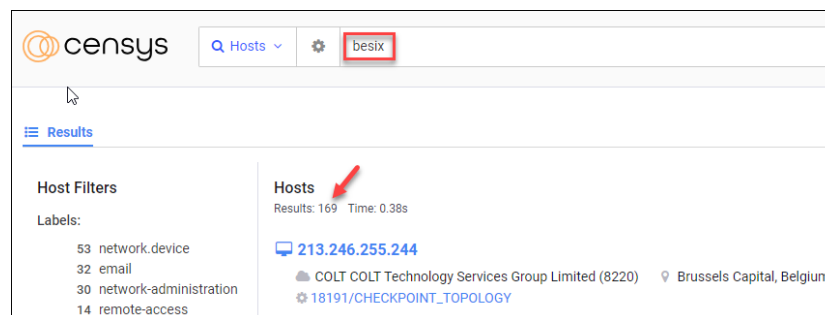
**CHALLENGE QUESTION:** Is this “actionable intelligence?” What actions might follow if you discover your email or a TARGET email address was involved in a breach?

### 7.0 - Evidence of Fraudulent or Unauthorized Endpoints

The final steps are the most “open to interpretation” and creative investigative approaches. What we are looking for is sites/services that are supposed to appear like they are related to our target but are in fact unauthorized. They may or may not be “evil” but might require reporting and additional investigation.

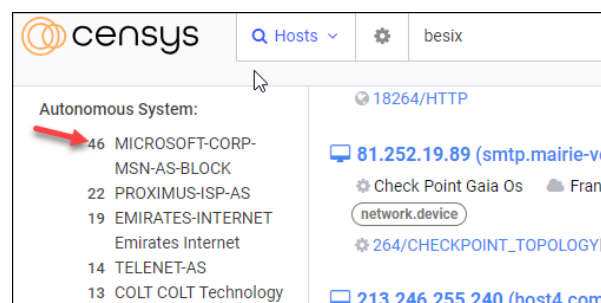
This is where I like to use keyword, ASN and certificate providers to look for “abnormal.” Sometimes keywords are easy, depending upon your target. If you are researching “Zetes” for example, you can search for “zetes” and results are reasonably likely related to the target organization. If your target is “Godiva,” searching for “godiva” or “chocolate” will return many results, most of which are not useful!

**EXAMPLE:** Let’s go back to *search.censys.io* and search for keyword “besix.”



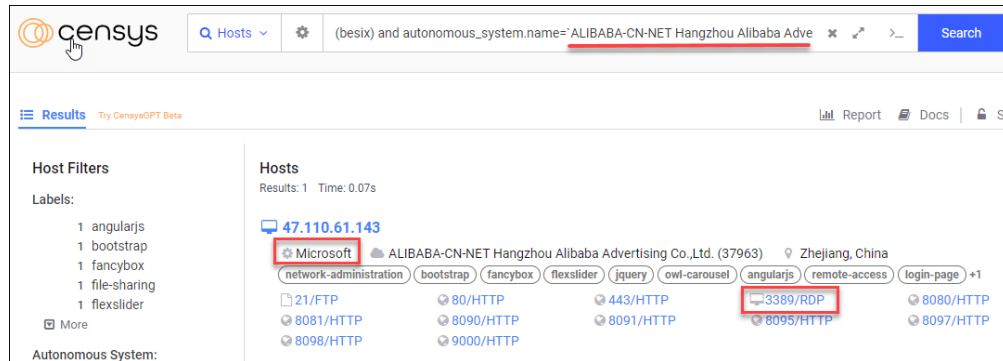
We get 169 results returned. If we look at “Host Filters,” we get a categorical/quantitative view of labels, ASN’s, ports, etc. From here, I like to review “least frequency of occurrence” to see if anything sticks out as abnormal.

For example, there are 46 results running on Microsoft Infrastructure (identified via a MICROSOFT-CORP ASN). That is “frequent,” as are Proximus and COLT COLT.

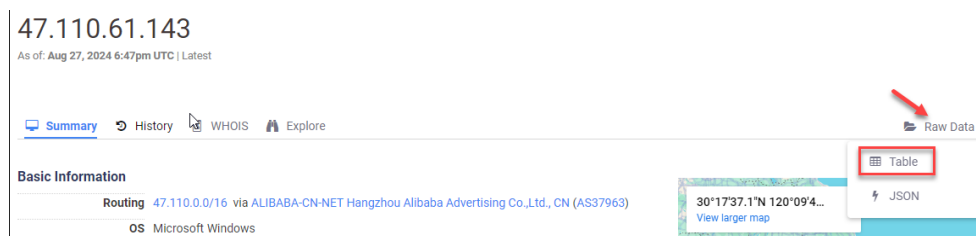


NOTE: Before continuing, you will need to “register” for a free Censys account (feel free to do so with a burner email!) or you can use the “class” account [instructor will provide CREDENTIAL].

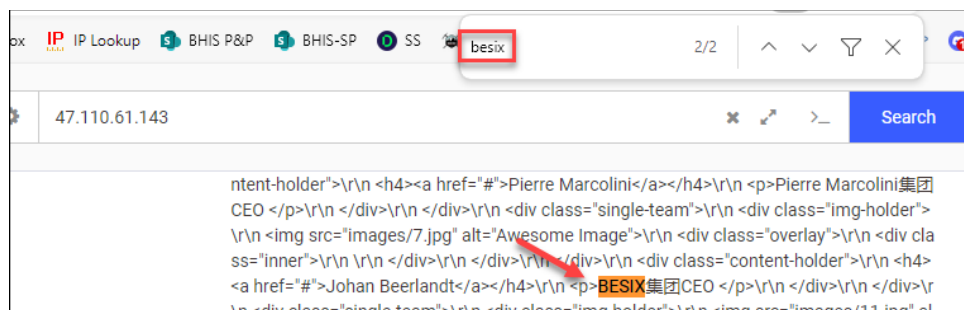
What about the ASN’s with a single entry, like ALIBABA based out of China? If we click the ASN, Censys will filter and show us that endpoint:



IMPORTANT: This looks highly suspect, but we need to validate relationship to the target. If we click on “47.110.61.143” in our example, we can then click on “Raw Data\Table,” then search for “besix” to see where/why/how this is related?



“Ctrl+F” should open up the search option in your browser. I searched for “besix,” and it appears the “relationship” is a news article about the BESIX CEO!



You can repeat the steps in this section looking at “least frequency of occurrence” for ports, services, certificate providers, etc.

**BE PREPARED TO VERBALLY SHARE YOUR RESULTS WITH THE CLASS!**

This concludes this lab!