**OSINT**

*Lab Requirements:*

- Kali Rolling VM with the following installed:

- ipwn (https://github.com/altjx/ipwn) (git clone https://github.com/altjx/ipwn)

- also run apt-get install extract

*System Setup:*

* Install Kali in a Virtual Machine
* Run apt-get update && apt-get upgrade
* Run git clone https://github.com/altjx/ipwn
* Run apt-get install extract

**\*\*\* If you are running an older version of Kali (Older than Rolling) grab FOCA from https://www.elevenpaths.com/labstools/foca/index.html and install on Windows for the FOCA part. \*\*\***

Open Source Intelligence (OSINT) is the process of gathering intelligence using open source means. OSINT is one of the most important aspects of cyber security, but goes beyond just the cyber security field. OSINT can be, and should be, an integral part of internal, external, web application, physical and electronic social engineering testing, as well as interviewing. So how do we perform OSINT? Well, we search the web looking for publicly accessible information on our target. For the purposes of this lab, we will be gathering information on Google.

**\*\*\*It is IMPORTANT to remember that during OSINT, there are no scans or other testing against the target. OSINT is completely passive\*\*\***

To start any engagement, I spend a good amount of time on OSINT. As a penetration tester, the longer I spend on OSINT, the better the engagement is. When I spend little to no time on OSINT, I often fail, but when I spend a half a day or more on OSINT, I am very successful. So to begin, we need to find out who we are dealing with. At the end of this document, I have provided some questions for you to answer as you progress. One of the first things to look for is information on the domain the company uses. Since we are gathering information on Google, it is not an unknown company, and we know they own www.google.com. But what else can we find about Google?

Let's start by looking up information on google.com. We can start by issuing a ping command. Open your terminal and type 'ping google.com'. You will receive an IP address. Take that IP address and migrate over to arin.net and put the IP address in the search box. You see the Google owns the 172.217.0.0/16 net block. I am SURE that Google owns more of the Internet than just those addresses. So time for question number 1. How many network blocks are owned by Google? (A trick to find out is by clicking the link next to the company name. and looking at the available information from that site. It looks like a stock ticker.)

Great, now we want to see what Google is using for the Domain Name System (DNS) servers. Any whois site or tool will do, including the command line tool whois. Issue the command 'whois www.google.com'. Looking at the information returned, we can see Google has blocked access to specific people in various contact roles, however, this is not always the case. We are looking for the Google DNS name servers (NS). What are the name servers used by Google?

Another way to gather information is using Google Dorking. Google Dorking is just issuing specific search terms to find information that you are interested in. Some popular "Google Dorks" are inurl:, intitle:, site:, filetype:, intext:, and using the minus symbol (-). Search for information on google.com, specifically see if you can find any .pdf files in google sites, containing the word password. (The Google Hacking Database (GHDB) is a great resource for creating Google Dorks. Check it out by going to https://www.exploit-db.com/google-hacking-database)

Once you find an interesting file, look at the metadata. Metadata is information contained within the document that gives information such as creation date, creator, and information about the computer or software used. The easiest way to look at the metadata is by using exiftool. Run exiftool <filename> and see what metadata comes back. (If you do not have exiftool, type in ‘apt-get install exiftool’) What software was used to create the document? (Hint: If the company is doing it right, there may not be any metadata that is useful. Keep looking at other files to find metadata.)

So as we go, we start putting together more and more information about the company we are gathering information about. At this point we know what network blocks are owned by Google, what DNS servers are in use by Google, and some information about technology used by Google. Next we may want to see what people work for Google. A great way to do this is by looking at LinkedIn. Manually try to find people that work for Google that are on LinkedIn, and as a bonus, see how closely connected you are to one of them.

Gathering information from LinkedIn can be time consuming, so thankfully there are tools to help, and we will get to tools in a little while. LinkedIn showed over 93,000 "Employees" of Google while writing this. Let's see if we can find an accurate number of employees. Use Google Dorking or another method to find how many employees work for Google.

As we have seen, manual OSINT takes time. A LOT of time. There are tools that will help with this data gathering though. Some tools we will look at are theharvester, FOCA (or pyFOCA), recon-ng, maltego, dnsrecon, and Shodan. theharvester is a tool for gathering e-mail accounts, subdomain names, virtual hosts, open ports/ banners, and employee names from different public sources (search engines, pgp key servers). Running 'theharvester' will give you the command line options. See how many email addresses you can find for google.com.

Now let's move to FOCA. FOCA is a tool used to gather information about a company, such as document metadata. I prefer pyFOCA (part of ipwn (https://github.com/altjx/ipwn)) but you can use the Windows tool as well if you would rather. Using pyFOCA (or FOCA), let's see if we can gather information from the metadata of documents. How many documents contained metadata? Any contain the author? Does that person work for Google?

Since we gathered the Google Domain Servers, dnsrecon may not be that important to us, but there are some items we may have missed manually so let’s run dnsrecon to see what servers are stored in DNS. Look for the mail servers used by Google. (Hint: the MX record is for Mail Exchanger)

Shodan is a search engine like Google, but for vulnerabilities. Navigate to www.shodan.io. Create an account or login and let's see what we can find out about a Google server. Take an IP address you have seen from the OSINT we have done so far and search for it. You will see some information about that host, including what ports are open and information about that service. What ports are open on the host you chose? (Remember, you are not scanning the host, Shodan is.)

recon-ng is an awesome tool to perform different types of recon/OSINT. Some modules will touch the network and many require API keys to use. As a quick primer, open recon-ng and run 'use recon/domains-vulnerabilities/xssed', then 'set SOURCE google.com'. This module will look at xssed.com, a site which lists found XSS vulnerabilities, for any XSS found on google.com sites. Type 'run' and see all the sites Google had that were susceptible to XSS. How many sites did recon-ng find?

Using the skills learned in this lab will help you on your way to performing thorough OSINT, and help you on your next interview, or engagement. **Remember, use these skills for good, never for evil.**

1. How many network blocks are owned by Google? - 35 IPv4 & 4 IPv6
2. What are Google's DNS servers? – ns1.google.com, ns2.google.com, ns3.google.com, ns4.google.com
3. What is the search term used to find a .pdf file in google sites containing the word "password"? - inurl:google.com filetype:pdf intext:password
4. What software was used to create the .pdf file? - Varies
5. How closely connected to a Google employee are you?
6. How many employees work for Google? – 61,814
7. How many email addresses did you find using theHarvester? - I found 1
8. How many documents contained metadata? Any contain the author? Does that person work for Google?
9. How many Mail servers does Google use? - 5 (5 IPv4, 5 IPv6, but only 5 hostnames)
10. What ports are open on the host you chose for your Shodan search?
11. How many sites were found vulnerable to XSS according to xssed recon-ng module? - 20