Use offense to inform defense. Find flaws before the bad guys do.

Copyright SANS Institute
Author Retains Full Rights

Interested in learning more?
Check out the list of upcoming events offering "Hacker Tools, Techniques, Exploits, and Incident Handling (SEC504)"
at https://pen-testing.sans.org/events/
The Information We Seek

GIAC (GCIH) Gold Certification

Author: Jose Ramos, joseramostrd@gmail.com
Advisor: Adam Kliarsky
Accepted: October 16, 2016

Abstract

Whether you are performing a penetration test, conducting an investigation, or are skilled attackers closing in on a target, information gathering is the foundation that is needed to carry out the assessment. Having the right information paves the way for proper enumeration and simplifies attack strategies against a given target. Throughout this paper, we will walk through some strategies used to identify information on both people and networks. Some people claim that all data can be found using Google’s search engine; but can third party tools found in Linux security distributions such as Kali Linux outperform the search engine giant? Maltego and The Harvester yield a wealth of information, but will the results be enough to identify a target? The right tool for the right job is essential when working with any project in life. Let’s take a journey through the information gathering process to determine if there is a one size fits all tool, or if a multi-tool approach is needed to gather the essential information on a given target. We will compare and contrast many of the industry tools to determine the proper tool or tools needed to perform an adequate information gathering assessment.
1. Introduction

It is the moment of truth. You are a penetration tester and tasked with finding and exploiting vulnerabilities for a new high profile client. This particular client has specifically requested for you to approach the testing like you were an attacker. Therefore, you were not provided with an IP range or contact details. Your palms begin to clam up as thoughts of losing your job start to circulate in your head due to the complexity of this task. Then you take a moment and think if an attacker performs this very function without detailed contacts and ranges, why can’t a penetration tester do the same? The simple answer is that they can. With knowledge of what to look for and where to locate the information, this task can easily be achieved. Throughout this paper, we will go through many different methods of enumerating and identifying information. We will go through the many ways that information can be obtained (in transit and also at rest). It is common for people to attempt to safeguard their personal information, but the same individuals do not realize that the trusted organizations that they provide their information to might be at risk. We will use a trusted executive for a fortune 50 company as an example. The executive may know that attackers have been trying to spear fish him and the attackers will stop at nothing until they retrieve the information that his organization has entrusted him with. As a precaution, the executive use multi-factor authentication to access his email, he does not use public networks and is very adamant about discussing company secrets over a secure land line. The executive may believe that his and his company’s information is secure. Assuming his information is secure may be his biggest problem. An attacker seeking information will not use conventional methods of infiltration when they are attempting to close in on their target. In fact, the attacker may wait till their target feels secure before they perform the attack. It is very rare that a burglar would walk through the front door of a home in broad daylight. The same concept applies to a hacker brute forcing the executive’s email while making it obvious that the attack is occurring. Instead, an attacker may think outside the box. A stealthier approach would be to gain information on the target’s family via social media and attempt to hack the weakest line of defense, for example the children. If an attacker can convince one of the children to open a crafted email containing a malicious payload on their home machine, this may be the way to gain access to the target information. A Meterpreter shell can be used to as a pivot point to enumerate the executive’s home network further. Furthermore, the shell can also act as a key logger or be used to dump passwords out of memory.

Jose Ramos, joseramostrd@gmail.com
If an attacker wants to get the information, they will think of out of the box methods to achieve it. So let embark on a journey through the world of information and discuss various ways to obtain the data. This paper will discuss tools, techniques, and manual methods strategies used by both penetration testers and attackers. There are two goals of this paper; offensively identifying the right tools for the situation, and defensively how to protect against these types of attacks.

2. The Network

The first step during an information gathering assessment is to understand the information that you are attempting to obtain. For example, if a person was looking to get an IP range for a targeted network, they must first grasp the concept of how networks operate. Going in blindly would be similar to driving a car without the knowledge of traffic signs or street lights. Incorporating IP addresses that are not in a target's range can provide inconclusive results, or even worse, an assessment against a network that you were not authorized to perform. Let’s drill down into this concept as it plays an important factor when it comes to gathering information about a targeted network. Networks are broken down into three different classes (A, B, and C). In addition to this breakdown, each of the classes has a reserved group of IP addresses that are meant for private use. The Microsoft TechNet website (Microsoft, n.d.) provides useful charts that can be utilized as a quick reference when deciphering between the different IP classes. Figure one identifies the various IP classes, and figure two shows the ranges that are reserved for private use.

<table>
<thead>
<tr>
<th>Address Class</th>
<th>First Network ID</th>
<th>Last Network ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>1.0.0.0</td>
<td>126.0.0.0</td>
</tr>
<tr>
<td>Class B</td>
<td>128.0.0.0</td>
<td>191.255.0.0</td>
</tr>
<tr>
<td>Class C</td>
<td>192.0.0.0</td>
<td>223.255.255.0</td>
</tr>
</tbody>
</table>

**Figure 1: IP Class range (Address classes, n.d.)**

Jose Ramos, joseramostrd@gmail.com
The private address space specified in RFC 1918 is defined by the following three address blocks:

- **10.0.0.0/8**
  The 10.0.0.0/8 private network is a class A network ID that allows the following range of valid IP addresses: 10.0.0.1 to 10.255.255.254. The 10.0.0.0/8 private network has 24 host bits that can be used for any subnetwork scheme within the private organization.

- **172.16.0.0/12**
  The 172.16.0.0/12 private network can be interpreted either as a block of 16 class B network IDs or as a 20-bit assignable address space (20 host bits) that can be used for any subnetwork scheme within the private organization. The 172.16.0.0/12 private network allows the following range of valid IP addresses: 172.16.0.1 to 172.31.255.254.

- **192.168.0.0/16**
  The 192.168.0.0/16 private network can be interpreted either as a block of 256 class C network IDs or as a 16-bit assignable address space (16 host bits) that can be used for any subnetwork scheme within the private organization. The 192.168.0.0/16 private network allows the following range of valid IP addresses: 192.168.0.1 to 192.168.255.254.

**Figure 2: Private address space (Public and private addresses, n.d.)**

### 2.1 Search Engine Manipulation

Understanding a network is an important part of information gathering. However, the network does not always need to be scanned to enumerate information. Many of the times information is publically accessible; you simply need to search for the information differently. The Google web browser is a very powerful tool. Some ways have been discovered to manipulate this tool and find targeted information. In fact, the Exploit Database houses some of these hacks that can be used to search for detailed information (Google hacking database, 2009). The image in Figure 3 illustrates a portion of the information found in the Network and Vulnerability section of the website. The "Inurl" text may not appear to be dangerous since it will generate random results. However, if you were to pair the exploitable system “Inurl” with the companies or a person's name, the results can present an immediate threat.

Jose Ramos, joseramostrd@gmail.com
Some other methods can be used to gather information about a particular target. The Exploit Database image shows a number of strings that are tailored for specific devices; however, there are some other options used to find information on the target. An attacker can search for items in the body of a site, by a particular person or even a specified file type. The options are limitless, as an attacker can narrow down the specified information by combining two or more of these search criteria's into a single search. The SANS Institute has done a fantastic job of putting some of the key search items into a single which can be found at www.sans.org/security-resources/GoogleCheatSheet.pdf. The image in Figure 4 shows some of the items that can be used to manipulate the search results when using the Google search engine.

Figure 3: Exploit database
Figure 4: Google hacks

2.2 Discovered Information

Google is a very powerful tool for gathering information. A simple search result on a person’s name often provides results about the individual. Social media information such as Facebook and LinkedIn are often top search results for name searches. Some would assume that an attacker would only use such information for identity theft of an individual; however, they are highly mistaken. Facebook accounts can be utilized as generate a password file on a target with tools that we will later discuss. It is common for people to use their animals or family names in passwords. Birthdays, anniversaries and maiden names are often mentioned in people’s social media posts. The average person often chooses a name or number that they already remember when creating a password for an account. If all the names and numbers are posted within their social media account, an attacker can just pull the context of the posts and replay the

Jose Ramos, joseramostrd@gmail.com
information against a login. Another problem is that people are known to use the same password across many accounts. Therefore, there is a high chance that the same password used to breach a personal account can be used to compromise their corporate account as well.

LinkedIn is another great avenue for information gathering, especially when it comes to enumerating a network. LinkedIn is used primarily to network with professionals and to showcase an individual’s talents and experiences. The problem is that the LinkedIn community often lists too much information about an organization that they are or have worked for. To validate this concept, a random listing was pulled from LinkedIn from the profile of a network administrator. For security reason, the person’s name and the company in which they work for will not be displayed in the image. The information should always be approached from an attacker’s perspective to determine if it holds any value for an attack.

![LinkedIn post](image)

**Figure 5: System engineer LinkedIn post**

The first thing that an attacker would notice is that the network has a combination of Cisco and Palo-Alto products. Not only did this individual mention that they are using Cisco devices, but they also told the models which are deployed. Such information smooth's the path for the attacker. Anyone looking to penetrate this company can search for vulnerabilities against the listed model types. As a proof of concept, I have listed vulnerabilities against the Cisco Catalyst 3750-X device mentioned in the listing (CVE details, n.d.).

Jose Ramos, joseramostrd@gmail.com
The social media sites used today can provide a wealth of information. However, it may be difficult to find information about several accounts owned by an individual. Websites such as namechk.com can resolve this issue. Instead of checking against one social media website, namechk runs a query against many of the traditional media and communication domains used today (namechk, n.d.). The dashboard can be used to connect to the existing page if the account does exist. One additional feature of the website is the ability to use email addresses as well as nicknames used by an individual. Figure seven shows a search against accounts that use the name ‘Hacker’.

Jose Ramos, joseramostrd@gmail.com
Our country has seen their fair share of data breaches. In many data breaches, the information is either sold off or posted on the Internet for everyone to see. Many of the companies send a notification alerting individuals of the breach and then recommend that they change their password. While this does correct the issue with the current website, it doesn't change the fact that people use the same password across many sites. An attacker can easily pull find account information from a breach on a paste site and then replay that information into a different account that they discovered by using namechk.com. To put things into perspective, an image from informationisbeautiful.net is shown below. The website shows an informational timeline of breaches over the past decade. Look at each of these breaches and visualize a complete database of information being dumped on the Internet.

Jose Ramos, joseramostrd@gmail.com
Figure 8: A Decade of Breaches

As one would imagine, anything that is used for good can be used for evil. People are concerned about these breaches and would like to know if their information has been posted on the Internet. A website has been developed to assist with this very need. The website checks to see if your email address was part of a breach and also crawls paste sites to view if your credentials have been posted (Have I been pwned, n.d.). The site is called haveibeenpwned.com and sounds like a lifesaver, right? Wrong, an attacker would use such a website for the very opposite reason. Someone seeking a password can plug the email address of their victim into the website to see if their credentials have been compromised. If confirmed, they can then navigate to the paste site to gather the credentials and attempt to enumerate the account. Figure nine shows a query against a Gmail account named "johndoe." The website responded with results stating that the email has been seen in 31 breaches and found in over 29 paste sites.

Jose Ramos, joseramostrd@gmail.com
Figure 9: Have I been pwned

2.4 The Toolbox

Using Internet web browsers and websites to find information are great ways to seek information. However, many tools have been created to automate the manual discovery of information. The Kali Linux suite is a security distribution that houses some tools that can be used to enumerate both networks and people. In this section, we will go over a few of these tools and discuss their functionality.

One of the most powerful information gather tools in the Kali distribution is Maltego created by Paterva. The tool can enumerate information on both networks and individuals. Maltego comes in three different flavors the Community Edition (free and included in Kali), Maltego Classic, and Maltego XL. All three of the flavors work similarly; however, the paid versions carry additional options. The image below illustrates the value Maltego adds to the information gathering process.

Jose Ramos, joseramostrd@gmail.com
Figure 10: Maltego graph

The Maltego graph is the workspace that is used within the tool. Figure ten starts with the domain and then continues to branch down as it discovers information. One of the greatest features of this tool is that you can start from any item (i.e. person, domain, account, or IP address). Starting from paterva.com, the attacker was able to enumerate the DNS servers related to the domain. Further enumeration identified employees and network addresses connected to the DNS servers. This tool allows an attacker to connect the dots on a visual graph in real time. Furthermore, emails that are found can be viewed on the console itself. One of the caveats about Maltego is that it gathers information from various sources based on your search criteria. In the event you go beyond the target range, you may start to intercept data that does not pertain to the target at hand. For example, if you are checking for a user by email address, you might come across a secondary person that was copied on the message. The secondary email address of the individual may lead you to a domain that does not have any relevance to the initial target being investigated. They say that on the Internet we are all connected; this tool shows just how interconnected we are.

Another useful information tool found in the Kali distribution is The Harvester. This tool allows an attacker to input a company or domain name and cross reference the information across known data sources such as Google, Bing, and LinkedIn. A search for Yahoo.com against a LinkedIn database will provide results of individuals on LinkedIn who have listed Yahoo as their employer. The list can then be exported and

Jose Ramos, joseramostrd@gmail.com
used in additional attack strategies against a targeted network. The Harvester is a simple text-based tool that takes the footwork out of finding employees at a particular company. The best part of this tool is that it can search and provide results in seconds.

Urlcrazy is another tool that is incorporated into the Kali distribution. This tool allows for an attacker to view all registered variations of a domain name. For example, let's say that we own a domain called mywebsite.com. Urlcrazy will display any related domains to the one I typed. Figure eleven shows results on this query.

![Figure 11: Urlcrazy results](image)

The tool provides results on character omissions, repeated characters, additional characters, as well as swapped and replaced characters. An attacker can use this information in the attempt to identify mail servers (mail.mywebsite.com) and other servers that are related to the targeted network. In addition to seeing what is available, an attacker can also register a similar domain to use for spam against the target organization. It wouldn’t be too difficult to register mywebste.com (removing the "l") and sending an email posing as an administrator requesting a user’s password for system maintenance. It is very common for the human eye to overlook such changes, especially when it is a busy day and they have hundreds of emails in their inbox.

### 2.5 Defense and Identification

Jose Ramos, joseramostrd@gmail.com
We have seen how easy it is for an attacker to gather information. The best method to protect your information is by limiting how much information is publicly posted on the Internet. In cases with social media, it is essential for individuals to know where they are posting the information and that privacy settings are configured. A corporation can have a bit more responsibility, especially when dealing with client side information. In such cases, a company must make sure that to protect their client’s information in transit and at rest.

It is easy to assume that someone is in your network viewing and extracting information. However, tools such as Wireshark can assist in viewing your network traffic and evaluating how machines are communicating on the network. Once again, a clear understanding of networks will be needed to view and analyze the Wireshark traffic. This tool can show source and destination address and the protocols that were used across the network. Let's take this into perspective. If you see that your one of the machines on the network has been transferring data to an IP address in China via FTP every night at 2:00 am, this should something to look into.

The traffic on a Wireshark capture can be filtered to narrow down to specific information. For demonstration purposes, I have filtered the traffic capture on my network to view data going from one machine to another. The "ip.src" command represents the source address, and the "ip.dst" shows the destination address. As mentioned, if traffic is being sent to a foreign location that is unintended, it is not a mistake. Such a threat should be looked into.

![Wireshark Capture](image.png)

**Figure 12: Wireshark capture**

Wireshark is a great tool when used for good. In addition to seeing a source and destination, the tool can also capture messages in clear text when secure data transfer

Jose Ramos, joseramostrd@gmail.com
methods are not used. The below image shows a capture of traffic from an HTTP (not secure session). This particular message has little use for an attacker; however, account ID’s and passwords can also be captured via the same method.

![Image of HTTP traffic capture]

**Figure 13: Clear text Information**

Coffee shops and airports are huge targets for these types of attacks. Attackers often spin up free Wireless access points in these locations to perform such an information enumeration attack. Once the client connects to the network, the attacker can stand in the middle of the Internet connection and the client. This attack is commonly referred to as a man in the middle attack. Since the attacker is the middle man, they can virtually view the traffic going to and from the server. Also, the attacker can then use tools such as SSL strip to remove the security of a protocol like HTTPS and still view the traffic going across the wire. The best defense against this type of attack is not to connect to networks that you do not know. In the event the connection is a must, a VPN connection should first be established.

### 2.6 Putting it all Together

The truth is information cannot be gathered using one tool or resource. To efficiently collect information on a targeted network or person, multiple resources need to be used. This paper was meant to expose the reader to the various sources information can be gathered, but the format was intentionally written to bounce back and forth between network and person to simulate an actual investigation. Many times information on an individual is needed to further enumerate a network and vice versa.

It is essential to validate information when it is being gathered. Incorrect subnets and not reviewing the received information can lead the investigation off on a wild goose chase.

Jose Ramos, joseramostrd@gmail.com
chase. Knowing what you are looking for and where to find it will always save time during an investigation. One may choose to manually search LinkedIn for hours finding Yahoo employees, whereas people who read this paper may use The Harvester. There is no simple way to locate the information one might be seeking; but with enough patience and dedication, they are bound to find results.

3. The Closing Act

We have discussed a wealth of information in this paper. The discussions ranged from attackers gaining information when our guard is down, to using tools and attack methods to gather the information ourselves. When it comes to information gathering, it is important to understand both sides of the playing field. A wise man once said the best way to catch a criminal is to think like one. This is a concept that we should keep deer to ourselves when attempting to protect our information. If you take anything away from this paper, it should be that attackers are very crafty and can be very persistent when it comes to achieving their goal. For this reason, alone, we should always keep a defense in depth mindset and also exercise information awareness to our friends and families. One last recommendation would be to use tools and techniques that were discussed to find out what information is available about you on the Internet. You may find some accounts opened in your name, or even personal information easily accessible. Just remember if there is a will, there is a way. Our information is out there just waiting to be found. Some may call it data; but in the end, this is the information that we seek.
References


Jose Ramos, joseramostrd@gmail.com
Upcoming SANS Penetration Testing

<table>
<thead>
<tr>
<th>Event</th>
<th>Location</th>
<th>Dates</th>
<th>Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS London April Live Online 2020</td>
<td>London, United Kingdom</td>
<td>Apr 20, 2020 - Apr 25, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>Instructor-Led Training</td>
<td>Apr 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANS Pen Test Austin 2020</td>
<td>Austin, TX</td>
<td>Apr 27, 2020 - May 02, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>Live Online - SEC504: Hacker Tools, Techniques, Exploits and Incident Handling</td>
<td>United Arab Emirates</td>
<td>May 04, 2020 - Jun 10, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>Instructor-Led Training</td>
<td>Apr 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live Online - SEC560: Network Penetration Testing and Ethical Hacking</td>
<td>United Arab Emirates</td>
<td>May 05, 2020 - Jun 11, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>Hong Kong SE Asia Live Online 2020</td>
<td>Hong Kong, Hong Kong</td>
<td>May 11, 2020 - May 22, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>SANS Security West 2020</td>
<td>San Diego, CA</td>
<td>May 11, 2020 - May 16, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>Autumn Australia Live Online 2020</td>
<td>Sydney, Australia</td>
<td>May 18, 2020 - May 29, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>Live Online - SEC560: Network Penetration Testing and Ethical Hacking</td>
<td>United Arab Emirates</td>
<td>May 18, 2020 - Jun 06, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>Live Online - SEC504: Hacker Tools, Techniques, Exploits, and Incident Handling</td>
<td>United Arab Emirates</td>
<td>May 19, 2020 - Jun 06, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>Live Online - SEC660: Advanced Penetration Testing, Exploit Writing, and Ethical Hacking</td>
<td>United Arab Emirates</td>
<td>May 26, 2020 - Jul 02, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>2-Day Firehose Training</td>
<td>May 26</td>
<td></td>
<td>CyberCon</td>
</tr>
<tr>
<td>CS Cybersecure Catalyst Women Academy SEC504</td>
<td>Brampton, ON</td>
<td>Jun 01, 2020 - Jun 06, 2020</td>
<td>Community SANS</td>
</tr>
<tr>
<td>Instructor-Led Training</td>
<td>Jun 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS-Cybersecure Catalyst New Career Academy SEC504</td>
<td>Brampton, ON</td>
<td>Jun 01, 2020 - Jun 06, 2020</td>
<td>Community SANS</td>
</tr>
<tr>
<td>CS-Cybersecure Catalyst New Canadians Academy SEC504</td>
<td>Brampton, ON</td>
<td>Jun 01, 2020 - Jun 06, 2020</td>
<td>Community SANS</td>
</tr>
<tr>
<td>Rocky Mountain HackFest Summit &amp; Training 2020</td>
<td>Virtual - US Mountain,</td>
<td>Jun 04, 2020 - Jun 13, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>SANSFIRE 2020</td>
<td>DC</td>
<td>Jun 13, 2020 - Jun 20, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>Instructor-Led Training</td>
<td>Jun 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyber Defence Australia Online 2020</td>
<td>Australia</td>
<td>Jun 22, 2020 - Jul 04, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>SANS Japan Live Online July 2020</td>
<td>Japan</td>
<td>Jun 29, 2020 - Jul 11, 2020</td>
<td>CyberCon</td>
</tr>
<tr>
<td>2-Day Firehose Training</td>
<td>Jul 1</td>
<td></td>
<td>CyberCon</td>
</tr>
<tr>
<td>Live Online - SEC660: Advanced Penetration Testing, Exploit Writing, and Ethical Hacking</td>
<td>United Arab Emirates</td>
<td>Jul 06, 2020 - Aug 12, 2020</td>
<td>vLive</td>
</tr>
<tr>
<td>Instructor-Led Training</td>
<td>Jul 6</td>
<td></td>
<td>CyberCon</td>
</tr>
</tbody>
</table>