CHAPTER 7

Reconnaissance

INFORMATION IN THIS CHAPTER

- Website Mirroring
- Google Searches
- Google Hacking
- Social Media
- Job Sites
- DNS and DNS Attacks

CHAPTER OVERVIEW AND KEY LEARNING POINTS

This chapter will explain the basics of the reconnaissance phase of the penetration testing life-cycle. This process will help the ethical hacker discover information about the target organization and computer systems. This information can be used later in engaging the computer systems.

INTRODUCTION

Just as military planners closely analyze all of the available information available to them before developing battle plans, a successful penetration tester must closely analyze all of the information that can be obtained before conducting a successful penetration test. Many times this information can be gained by searching the Internet using Internet sites like Google and others including those that are focused on information sharing and social media. Information can be found on the Internet's name servers that provide direction to user's browsers as well. Email messages can be tracked through an organization and even returned email can help the penetration tester. Creating and examining an off-line copy of the target website can provide a source of valuable information and can be used later as a tool for social engineering tasks, if allowed by the tests ROE.

Hacking with Kali. DOI: http://dx.doi.org/10.1016/B978-0-12-407749-2.00007-0

This phase starts with the test team knowing little about the target. The level of detail provided to the team can range from knowing only the organizations name and possibly a website address to detailed and specific system information including IP address space and technologies used defined in the ROE to limit or scope the test event. The ROE may also limit the test team's ability to conduct activities including bans on social engineering and destructive activities like denial of service (DoS) and distributed denial of service (DDoS) attacks.

The goal of this phase is to find out how much information you can about the organization.

Some things that should be determined about the organization include:

- organizational structure including detailed high-level, departmental, and team organizational charts;
- organizational infrastructure including IP space and network topology;
- technologies used including hardware platforms and software packages;
- employee email addresses;
- organizational partners;
- physical locations of the organizational facilities;
- phone numbers.

Trusted Agents

The trusted agent may be the person that hired the penetration test team or an individual that was designated by the organization that will be able to answer questions about the engagement and will not divulge the fact that a penetration test is occurring to the organization at large.

START WITH THE TARGETS OWN WEBSITE

The targets own website holds vast information for developing the profile for the engagement. For example many sites proudly display organizational charts and key leader's profiles. These should be used as a basis for developing a target profile and information about key leaders in the organization can be used for further harvesting of information on social media sites and for social engineering, if allowed in the stated ROE.

Many organizational websites also include a careers or job opportunity page. This page can be indispensable in determining the technologies used in the organization. For example, listings for systems administrators that are familiar with Active Directory and Windows Server 2012 would be a strong indicator that the organization is at least using Windows Server 2012. The same listing for administrator's familiar or expert in the administration of Windows Server 2003 or 2000 should make any penetration testers ears perk up as these platforms are more vulnerable than newer operating systems.

Each site should be checked for a link to webmail and if found it should be evaluated. If clicking the link results in an Outlook Web Access page being displayed, it would be a good assumption that Microsoft Exchange servers are being used for email. If an Office 365 page is displayed, it is a good indicator that email services are being outsourced and the mail servers would probably be out of bounds based on most ROEs. This would be true of Google webmail as well; however, this should all be detailed in the boundaries defined before the engagement began. If questions on the possibility of crossing a boundary exist, the engagements trusted agent should be used to resolve the question.

WEBSITE MIRRORING

There are times it is more effective to copy the organizations entire website to evaluate offline. This could be to use automated tools to search for terms or just to have a copy in case changes should be made to sensitive information that is on the current site. It is useful just to have a copy of the website to continue reconnaissance when offline. Tools like the command line wget will copy all of the html files from a website and store them on the local hard drive. The tool wget is installed by default in Kali Linux and is a simple tool to use. By using the following command line in the terminal window all of the html files from an entire website will be downloaded. It is important to note that wget will not copy server side programming for pages such as those created with a PHP script.

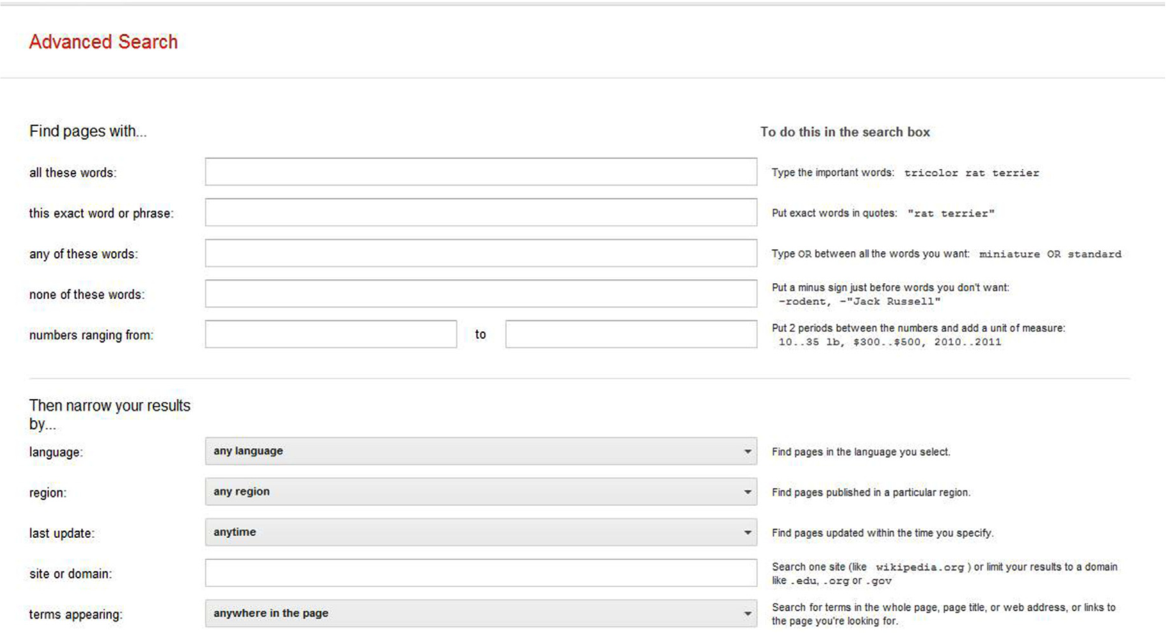
```
wget –m –p –E –k –K –np -v http://foo.com
```

In this example, the wget command is followed by a number of switches or options. As in any case with the tools on Kali Linux, the user manual or man pages can be referenced to determine the bets use of the tool for the engagement being conducted. To view the wget man pages, use the following command.

```
man wget
```

Once in the man pages review the contents by using the up and down arrows and the page up and page down buttons. Press the h key for help and press q to exit the man pages. A review of the wget man pages for this set of switches reveals the following:

– m mirror, turn on options that are suitable for mirroring the website;
– p page or prerequisites, this option ensures required files are downloaded including images and css files;
– E adjust extension, this will cause all pages to be saved locally as a html file;
– k convert links, this enables the files to be converted for local viewing;
– K keep backup converted, will back up the original file with a.orig suffix.



Advanced Search

Find pages with...		To do this in the search box
all these words:		Type the important words: `tricolor rat terrier`
this exact word or phrase:		Put exact words in quotes: `"rat terrier"`
any of these words:		Type OR between all the words you want: `miniature OR standard`
none of these words:		Put a minus sign just before words you don't want: `-rodent, -"Jack Russell"`
numbers ranging from:	to	Put 2 periods between the numbers and add a unit of measure: `10..35 lb, $300..$500, 2010..2011`
Then narrow your results by...		
language:	any language	Find pages in the language you select.
region:	any region	Find pages published in a particular region.
last update:	anytime	Find pages updated within the time you specify.
site or domain:		Search one site (like `wikipedia.org`) or limit your results to a domain like `.edu`, `.org` or `.gov`
terms appearing:	anywhere in the page	Search for terms in the whole page, page title, or web address, or links to the page you're looking for.

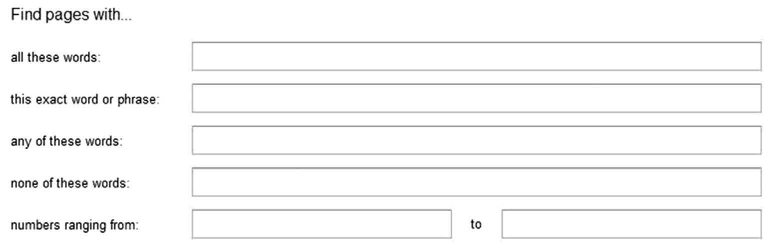
FIGURE 7.1
Google advanced search page.

The files transferred from an organizations web servers will be stored in a folder with the name of the website that was copied. When copying a website, errors may occur when pages created with or containing PHP or are downloaded. This is because much code to create the page is created by a script that runs on the server behind the web page in a location that most website cloning applications cannot access.

Once the files are downloaded it is important that they are not made available for viewing by others, such as reposting the website as this would constitute a violation of copyright law.

GOOGLE SEARCHES

The search Google technique leverages the advanced operators used to conduct detailed searches with Google. Those new to searching with Google can start with the Google Advance Search page located at http://www.google.com/advanced_search as illustrated in Figure 7.1. This page will help walk novice searchers through basic searches. The top half of the page, illustrated in Figure 7.2, will help find web pages by including and excluding words, terms, and numbers. The bottom half of the page will help narrow the results



Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from: to

FIGURE 7.2
Google advanced search (continued).

using Google's operators. The searcher can use any combination of fields on this page to construct the search string that will be used. Using more than one field will make a more complex but more focused search string.

All These Words

This field can be used to find pages containing the words typed in the dialog box regardless of where they are on the web page, in fact the words do not even need to be in the order typed or together, just somewhere on the web page. To conduct this search, type a number of terms in the dialog box and click the Advance Search Button, by doing this the words typed in the advance search page are translated into a search string, and then sent to Google as if they were typed directly in the search field on the main Google page.

This Exact Word or Phrase

Typing a search term in the field to the right of this option will cause the Google search engine to find the words or phrase in the exact order typed and in the order typed. Unlike the "all these words" search only web pages that contain the phrase or words in the exact order and together will be included in the result set. This search works by placing the search terms inside quotes.

Any of These Words

When using this field the Google search will find pages that contain any of the words. Unlike the "all these words" field the pages returned do not have to have all of the words that were typed. This search works by placing the OR connector between terms in the search box.

None of These Words

The words typed in this text box will be used to omit pages from the resulting Google search. Any pages containing the words typed will be removed from the result set. This search works by placing a minus sign in front of the words or terms you do not want in the result set.

Numbers Ranging from

By using the two text fields in this area the search will find pages that have numbers that in the range typed. This type of search can be enhanced by including units of measure, such as pound (lb), miles, or millimeters (mm) or currency like $ or €. This search can be conducted in the main search box by placing two periods between the numbers.

Language

By selecting a language from the drop down selector, the resulting pages will mostly be in the language selected. This search restrictor can be helpful to narrow results to pages that are written in the language most prevalent in the area that the target is located, for example by focusing on German sights a team conducting a penetration test on a German firm can better search for information relevant to this particular engagement.

Region

By selecting a region from the drop down selector the resulting pages will be from web pages published in the region selected. If no selection is made from the languages drop down the results from a search with a region selected will include pages published in that region regardless of the primary language used. By selecting both a language and region, a more focused search can be conducted.

Last Updated

By selecting a time limit in the drop down of these area only pages updated within the selected time frame will be included in the search. This will ensure older pages are not included in the result set and can be used to make sure the resulting pages are after a key event. For example, if the organization that is the focus of the penetration test recently completed a merger with another organization or adopted a new technology the search could be limited to the time since the event to ensure the search results are more relevant.

Site or Domain

This text box can be one of the most helpful when narrowing search results on the target. For example, searches on a government organization may benefit from restricting the results to only.gov domains, while searches on Foo Incorporated may benefit from limiting results to the foo.com domain. This type of restriction can also be conducted in the main Google search text box by using the search restrictor site: followed by the domain or domains that should be returned in the results set, for example use site: foo.com to restrict results to only pages from the foo.com domain.

Terms Appearing

By using this drop down the search query can be targeted at a specific part of the page. Obviously selecting "anywhere on the page" would run the search on entire pages of Internet sites with no real restrictions on where the search query was targeted.

A search on using "in title of the page" will only target the title of web pages. To be specific the title of the page is the part of the web page that is displayed in the tabs of the web browser. This search can also be conducted on the main Google page by using the intitle: operator in the search box.

Using the limiter "in the text of the page" will limit the search to only the text of the page and will exclude things, such as images, documents, and page structure like the title, however, if these items are written in the text of the page the search will return these items in the results. For example, if an image is referenced in the text of the page that image will be returned in the search results, this is true for image markup and links in text as well. Using the intext: operator in the Google search box is equivalent to selecting this option from in the drop down.

Using the "in URL of the page" will restrict searches to the page uniform resource locator (URL). The URL is the address of the web page that appears in the address box of the web browser. Finally, using the "in links to the page" will find pages that link to the search criteria. This search can be conducted in the main Google search box by using the inurl: operator.

Safe Search

Safe search has two options: "show most relevant results" and "filter explicit." The filter explicit setting can reduce sexually explicit videos and images from the search results. Selecting the show most relevant results will not filter the results for sexually explicit content.

Reading Level

The reading level option will filter results by the complexity of the text in the web pages that will be returned from the search. The "no reading level displayed" will execute the search with no reading level filter applied. The option "annotate results with reading level" will display all results; however, the reading level of each page will be displayed in the search results. The Google algorithm is not as scientific or fine grained as other grade level reading tools, including the Lexile level, but is quite efficient in filtering results into these three categories; basic, intermediate, and advanced. This can be helpful when conducting a penetration test by focusing the results on the reading level of the target. For example searches on a scientific organization could be limited to those pages with an advanced reading level. Trying all three levels might be beneficial to see different search results and important information can be gained from searches using the basic reading level.

File Type

File type can be one of the most important searches that a penetration tester can use. This setting contains the search results to a specific file type, for example,.doc and.docx for Microsoft Word Documents of.pdf for Adobe documents. Many times users will use different file types for different types of information. For example many times user names, passwords, and other types of account information will be stored in spreadsheets with.xls or.xlsx extensions. The drop down offers many of the most common file types and any extension can be used in the Basic Google search box by using the filetype: operator, e.g., filetype:xls.

Usage Rights

Usage rights limits the search results by the ability to reuse the content based on copyright and other reuse restrictions. By selecting "Free to use, share, or modify" the results returned will be content that can be reused with restrictions that stipulate how the content can be reused, such as the content cannot be modified, mostly without a fee. Free to use, share, or modify will return in search results that have pages that can be modified within the license restrictions, again the results will allow the content to redistributed normally without a fee. The options with the term commercial in the selection work as those without the term commercial but return results that can be used commercially.

Compiling an Advanced Google Search

Using the fields individually on the Google advanced page returns some impressive search results, but using many of these fields together will improve the way a penetration tester finds relevant information. For example, assume that Foo International (an American Company) merged with another company a month ago and requested a penetration test from your team. In times of transition like this many documents are created to help members of each company in the transition, it may be possible that an employee posted organizational charts to the company's website. One possible search could use the following fields and terms:

- this exact word or phrase: organizational chart
- language: English
- region: United States
- last update: past month
- site or domain: foo.com
- file type: pdf.

The results could then be further refined by adding or removing search fields or changing the options. For example changing the file type to PowerPoint (.ppt) or removing the file type altogether may return the results needed.

GOOGLE HACKING

Google Hacking is a technique that was pioneered and made famous by Johnny Long that uses specific Google operators and terms in Internet searches to return valuable information using the Google search engine. This technique focuses on using specifically targeted expressions to query the Google databases to harvest information about people and organizations. This technique takes the Google searches described earlier and supercharges their results.

Google Hacking makes extensive use of advanced operators and linked options to create targeted queries that can be run in the Google search engine. Many times the searches will be targeted at assembly information about specific technologies such as web management services and other searches will target user credentials. Several great books have been written that fully explain Google Hacking, the most famous is *Google Hacking for Penetration Testers* written by Johnny Long and published by Syngress.

Google Hacking Database

A great number of Google Hacking search query strings have been compiled into the Google Hacking Database (GHDB). The original database is located at http://www.hackersforcharity.org/ghdb/, Offensive Security also has a GHDB at http://www.offensive-security.com/community-projects/google-hacking-database/ that expands on the original database, and coining the term "Googledorks" a moniker for inept or foolish people as revealed by Google [1]. At the time of this writing the GHDB, maintained by Offensive Security, contained over 3350 Google Hacks divided into 14 categories. Over 160 of these search strings can be helpful for finding files that contain passwords. An example of one of these search strings that would attempt to find Cisco passwords is illustrated below.

```
enable password | secret "current configuration" -intext:the
```

Running this search returned almost a million and a half sites, and while some of the files returned may not contain actual passwords a great number of the results actually did contain password lists. This search could be further refined to meet the needs of individual penetration tests by adding additional operators, such as the site or domain operator as follows.

```
enable password | secret "current configuration" -intext:the site:foo.com
```

SOCIAL MEDIA

Social media has become an integrated part of many people's daily lives. This fact makes social media a treasure trove for gathering information in this phase of the penetration testing lifecycle. Information that is fiercely protected by people in the physical world is posted freely by those same people on social media sites using sites, such as Facebook, Instagram, Twitter, LinkedIn, and others a full profile of individuals working at the target location can be developed. This can help in social engineering engagements.

LinkedIn is particularly helpful in developing organizational charts. Built for connecting professionals LinkedIn will often help to fill in blank spots on the target profile, including a better defined organizational chart and even email address lists, although this latter step will often require social engineering as email addresses are not publically displayed on LinkedIn. Finding individuals that once worked for the organization are great sources of information if social engineering is allowed by the ROE. Finally LinkedIn has started to post job opportunities on its site, making it possible to use these listings to understand the technologies used at the target organization.

Create a Doppleganger

A doppelganger in folklore is a ghostly copy of an individual. It is common practice to develop a persona before beginning reconnaissance in the social media world. It is usually not effective to conduct research on a target using the profile of a security expert or penetration tester. If the penetration tester is able to establish social interactions with individuals from the organization through social media it would be far more effective if the penetration tester had a persona that claims to have once worked in the target organization or went to the same college as the CEO that the penetration tester is trying to connect with on LinkedIn. Obviously the penetration tester must be wary of completely taking over a real person's identity an act that could lead some believe that identity theft has occurred, but it is not uncommon for two people to have similar names. For example developing a fictions persona with the name of John Smith that went to Wisconsin University and a background totally made up is not the same as stealing the identity of the actual John Smith that went there. In any case ensure your persona does not bleed over into identity theft or fraud. This means, among other things, not filling out that credit card application that arrives with your personas name on it or using this persona for entering into legal agreements with the persona.

The lines for using a doppelganger should be specified early in the engagement and if social engineering is allowed the doppelganger should be developed that will be effective when social engineering comes into play. When filling out registration for social media sites the penetration tester should pay attention to the usage policy to ensure policies, rules, or in the worst case laws are not being broken by using a doppelganger persona.

JOB SITES

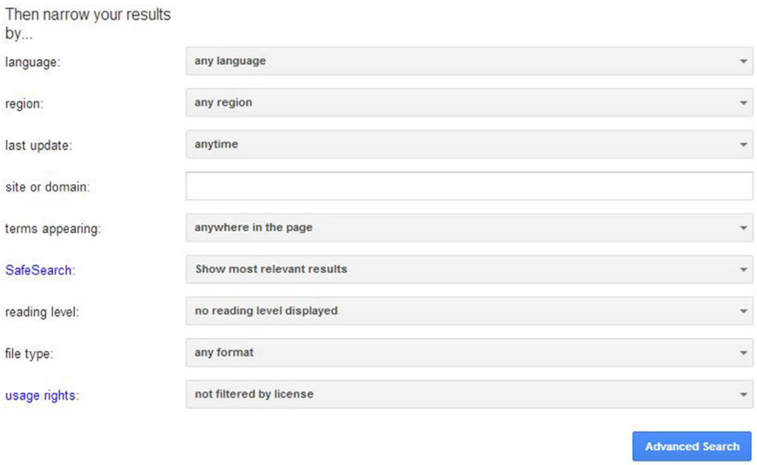
Searching job boards, such as Monster, Career Builder, and Dice, can sometimes result in interesting findings as well. Like the targets own website, these websites can shed light on the technologies used at the target site. Searching these pages with the organization in question will often result in the positions that need to be filled, helping the penetration tester better understand the target. In recent years many firms have begun to understand this weakness and are now listing positions as "company confidential" or other statement in the organization or company area of the job postings.

DNS AND DNS ATTACKS

Domain Name Services, or DNS, provides addressing help for the Internet. Generally people have a better time remembering and using names, like Google.com, while computers have an easier time using numbers like 173.194.46.19 (one of Google's addresses). The hierarchical structure of the Internet also makes the use of numbered octets more efficient. This creates a problem where the best addressing scheme for people does not match the best scheme for computers. Name servers help to solve this problem by serving as translators between computers and people.

These name servers are set up in a hierarchical order with top-level domain (TLD) servers, serving main domains, such as.com,.gov,.edu, and many others. At the other end of the name server hierarchy each network can have its own name server that allows local services and computers to be accessed by name instead of by IP address.

Possibly the easiest way to understand the basic functionality of name servers is to walk through how a computer and web browser interact and work with the entire name server system. From the local name server to the root, or name server that is above the TLDs, each name server can query the next name server above it or provide information to the name server below it, as illustrated in Figure 7.3. If the computer user was to type the address for Google into a web browser a chain of events would be triggered to translate the human readable name to one more useful to a computer. This starts with the user's computer asking the local name server if it knows the IP address relates to www.google.com, if this name server has had this request in the recent past and has cached the answer or Google was registered with that name server the IP address could be returned immediately. If that name server does not have the information cached or otherwise stored it asks the



Then narrow your results by...

language: any language

region: any region

last update: anytime

site or domain:

terms appearing: anywhere in the page

SafeSearch: Show most relevant results

reading level: no reading level displayed

file type: any format

usage rights: not filtered by license

FIGURE 7.3
Filtering Google searches.

next name server, if the next upstream name server does know the information it is returned if not this continues until the request reached the TLD name server, in this case the name server for.com.

Name servers contain a lot of useful information, well beyond web pages. For example, the name server will contain the mail server, or MX record, for the domain, other named computers or "A" records and other helpful information.

QUERY A NAME SERVER

By the nature of their design most name servers are open to the public. The following command entered in the Kali Linux terminal will query the name server assigned to the local computer.

```
nslookup
```

This will result in a carrot (>) being displayed in the terminal indicating the system is awaiting input. Type the following command to query the local name server to determine the IP address of the Google web page.

```
> www.google.com
```

This will return a number of IP addresses both authoritative (the first responses) and nonauthoritative, those following the nonauthoritative note. Nonauthoritative answers are a great source of information as this term only indicates the information is provided from the server's cache.

To exit from nslookup use the following command.

```
> exit
```

The nslookup command will use the name server defined for the local machine. To display the name servers being used for the current nslookup commands use the following command.

```
nslookup
> server
```

The command nslookup can return other information as well. For example, to search for all of the mail servers type the following commands.

```
> set type = MX
> google.com
```

This will return all of the known mail servers for the Google domain.

Identifying the different types of records about the target can be an important part of completing reconnaissance. As stated earlier the nslookup command, by default, uses the locally defined name server. In Kali Linux, the name server is defined in the resolv.conf file located in the /etc directory. Use the following commands to identify the locally defined name server.

```
cat /etc/resolv.conf
```

The name server used by nslookup can be changed to the target domains name server. First identify the targets name server with the following command.

```
r
nslookup
> set type = ns
> google.com
```

Table 7.1 DNS basic record types

Record Type	Default Port	Server Type
mx	25	Mail (email)
txt	n/a	Text message used for human readable notes
ns	53	Name Server
cname	n/a	Alias for another server (conical name)
aaaa	n/a	IP version 6 (IPv6)
a	n/a	Domain or Sub-Domain record

Once the target name servers have been identified, the name server used by nslookup can be changed to one of the targets name servers using the following command. This example sets the name server to one of Google's name servers.

```
nslookup
 > server 216.239.32.10
```

There are a number of records that can be discovered using nslookup. Many of the main record types are defined in Table 7.1.

ZONE TRANSFER

While it is possible to gain a lot of information by using programs like nslookup to manually transfer information it is possible to get much more information in a shorter time using a zone transfer. A zone transfer literally dumps all of the information from a name server. This process is useful for updating authorized name servers. Misconfigured name servers allow zone transfers not only to authorized clients for updates but anyone that requests the transfer.

The Domain Internet Gopher (DIG) is a program that can be used to attempt zone transfers. To attempt a zone transfer use the following command.

```
dig @[name server] [domain] axfr
```

Most transfers will fail, however, if the target name server is misconfigured. The entire name servers record set will be transferred to the local Kali Linux computer. When using this command the domain will be the domain minus any host, for example, foo.com not www.foo.com. The axfr command indicates dig should request a zone transfer. If the transfer is successful the information displayed can be used to add to the targets profile. This will provide valuable information for the future phases of the penetration test.

REFERENCE

[1] http://www.exploit-db.com/google-dorks/.