# Chapter 4. Information Gathering

In this chapter, we will discuss the information gathering phase of penetration testing. We will describe the definition and purpose of information gathering. We will also describe several tools in Kali Linux that can be used for information gathering. After reading this chapter, we hope that the reader will have a better understanding of the information gathering phase and will be able to do information gathering during penetration testing.

Information gathering is the second phase in our penetration testing process (Kali Linux testing process) as explained in the *Kali Linux testing methodology* section in Chapter 2, *Penetration Testing Methodology*. In this phase, we try to collect as much information as we can about the target, for example, information about the **Domain Name System** (**DNS**) hostnames, IP addresses, technologies and configuration used, username's organization, documents, application code, password reset information, contact information, and so on. During information gathering, every piece of information gathered is considered important.

Information gathering can be categorized in two ways based on the method used: **active** information gathering and **passive** information gathering. In the active information gathering method, we collect information by introducing network traffic to the target network. While, in the passive information gathering method, we gather information about a target network by utilizing a third-party's services, such as the Google search engine. We will cover this later on.

> **Note**
>
> Remember that no method is better in comparison to the other; each has its own advantage. In passive scanning, you gather less information but your action will be stealthy; while, in active scanning, you get more information but some devices may catch your action. During a penetration testing project, this phase may be done several times for the completeness of information collected. You may also discuss with your pen-testing customer, which method they want.

For this chapter, we will utilize the passive and active methods of information gathering to get a better picture of the target.

We will discuss the following topics in this chapter:

- Public websites that can be used to collect information about the target domain
- Domain registration information
- DNS analysis
- Route information
- Search engine utilization

## Using public resources

On the Internet, there are several public resources that can be used to collect information regarding a target domain. The benefit of using these resources is that your network traffic is not sent to the target domain directly, so our activities are not recorded in the target domain logfiles.

The following are the resources that can be used:

| No. | Resource URL | Description |
|---|---|---|
| 1 | http://www.archive.org | This contains an archive of websites. |
| 2 | http://www.domaintools.com/ | This contains domain name intelligence. |
| 3 | http://www.alexa.com/ | This contains the database of information about websites. |
| 4 | http://serversniff.net/ | This is the free "Swiss Army Knife" for networking, server checks, and routing. |
| 5 | http://centralops.net/ | This contains free online network utilities such as domain, e-mail, browser, ping, traceroute, and Whois. |
| 6 | http://www.robtex.com | This allows you to search for domain and network information. |
| 7 | http://www.pipl.com/ | This allows you to search for people on the Internet by their first and last names, city, state, and country. |

| No. | Resource URL | Description |
|---|---|---|
| 8 | http://yoname.com | This allows you to search for people across social networking sites and blogs. |
| 9 | http://wink.com/ | This is a free search engine that allows you to find people by their name, phone number, e-mail, website, photo, and so on. |
| 10 | http://www.isearch.com/ | This is a free search engine that allows you to find people by their name, phone number, and e-mail address. |
| 11 | http://www.tineye.com | TinEye is a reverse image search engine. We can use TinEye to find out where the image came from, how it is being used, whether modified versions of the image exist, or to find higher resolution versions. |
| 12 | http://www.sec.gov/edgar.shtml | This can be used to search for information regarding public listed companies in the Securities and Exchange Commission. |

Due to the ease of use, you only need an Internet connection and a web browser, we suggest that you utilize these public resources first before using the tools provided with Kali Linux.

### Note

To protect a domain from being abused, we have changed the domain name that we used in our examples. We are going to use several domain names, such as `example.com` from IANA and a dummy domain name `example.com` as well for illustrative purposes.