Common steps in the information-gathering process include:

1. gathering information from general resources (such as Google or the organization’s website)

2. determining the network’s logical and physical dimensions

3. finding open ports, active services, and access points

4. identifying active computers and devices

5. detecting operating systems

6. researching known vulnerabilities of running software

Of these steps, footprinting covers the first 2 steps in the process. Note that step 1 and at least part of step 2 are passive in nature; they do not require direct interaction with the victim. This is one of the key characteristics of footprinting: to gather information about a victim without directly interacting and potentially providing advance notice of the attack. Footprinting also generally focuses on gathering information externally, from outside the target organization. When footprinting an organization, a hacker can…

• examine the company’s website

• identify key employees

• analyze open positions and job requests

• assess affiliate, parent, or sister companies

• find technologies and software used by the organization

• determine the network address and range

• review the network range to determine whether the organization is the owner or the systems are hosted by someone else

• look for employee postings, blogs, and other leaked information

• review collected data

Footprinting may be the easiest part of the hacking process because most organizations generate massive amounts of information that is made available online. Before a skilled hacker fires up an active tool, such as a port scanner or password cracker, he or she will meticulously carry out the footprinting process to plan and coordinate a more effective attack.

The state of a website at a particular point in time may still exist somewhere out in cyberspace. One of the tools that a security professional can use to gain information about a past version of a website is the Wayback Machine. It is a web application created by the Internet Archive that takes “snapshots” of a website at regular intervals and makes them available to anyone who looks. It is available at www.archive.org. Although it doesn’t keep exhaustive results on every website, the websites it does archive can stretch back to 1996. Currently the Internet Archive has a sizable amount of content cataloged, estimated to be in excess of 327 billion web pages and related content. A potential drawback: a site administrator, using a file called robots.txt, can block the Internet Archive from making snapshots of the site, denying anyone the use of old information.

Of particular interest in the EDGAR database are the items known as the 10-Qs and 10-Ks. These items are quarterly and yearly reports that contain the names, addresses, financial data, and information about acquired or divested industries. Closer examination of these records indicates where the company is based; detailed financial information; and the names of the principals, such as the president and members of the board. EDGAR is not the only source of this information though:

• Hoover’s – www.hoovers.com

• Dun and Bradstreet – www.dnb.com/us/

• Yahoo! Finance – http://finance.yahoo.com

• Bloomberg – www.bloomberg.com

In a process known as Google Hacking, the goal is to locate useful information using techniques already provided by the search engine but in new ways. The GHDB is merely a database of queries that identifies sensitive data and content. Some of the items an attacker can find are available using the following techniques:

• advisories and server vulnerabilities

• error messages that contain too much information

• files containing passwords

• sensitive directories

• pages containing logon portals

• pages containing network or vulnerability database

Currently, many tools are available that can be used for obtaining types of basic information, including:

• Whois to provide information about registered users, or their assigned agents, of domains, IP addresses, or systems

• Nslookup to find information about a resource stored in the Domain Name System (DNS), including domain name, DNS server, and IP address(es)

• Internet Assigned Numbers Authority (IANA) and regional Internet registries (RIRs) to find the range of Internet Protocol (IP) addresses

• Traceroute to determine the location of the network’s

The Internet Corporation for Assigned Names and Numbers (ICANN) is the primary body charged with management of IP address space allocation, protocol parameter assignment, and DNS management. Global domain name management is delegated to the Internet Assigned Numbers Authority (IANA). IANA is responsible for the global coordination of the DNS root, IP addressing, and other IP resources.

Obtaining the network range requires the attacker to visit at least one or more of the Regional Internet registries (RIRs), which are responsible for management, distribution, and registration of public IP addresses within their respective assigned regions. Currently, there are 5 primary RIRs.

Automatic Registrar Query – The manual method of obtaining network range information is effective, but it does have the drawback of taking a significant amount of time. You can speed up the process using automated methods to gather this information faster than can be done manually. Several websites are dedicated to providing this information in a consolidated view. Numerous websites are also dedicated to providing network range information automatically. Some of the more common or popular destinations for searches of this type include:

• www.betterwhois.com

• www.geektools.com

• www.all-nettools.com

• www.smartwhois.com

• www.dnsstuff.com

• http://whois.domaintools.com

Underlying all these tools is a program known as Whois, which is software designed to query the databases that hold registration information. Whois is a utility that has been specifically designed to interrogate the internet DNS and return the domain ownership, address, location, phone number, and other details about a specified domain name.

Nslookup is another useful program to query internet domain name servers. Both UNIX and Windows come with an nslookup client. If you provide nslookup with an IP address or a fully qualified domain name (FQDN), it will look up and show the corresponding IP address and hostname. Nslookup can be used to do the following:

• find additional IP addresses if authoritative DNS is known from Whois

• list the MX (mail) server for a specific range of IP addresses

One of the missions of the IANA is to delegate internet resources to RIRs. The RIRs further delegate resources as needed to customers, which include internet service providers (ISPs) and end-user organizations. The RIRs are organizations responsible for control of internet protocol version 4 (IPv4) and IPv6 addresses within specific regions of the world. The 5 RIRs are as follows:

• American Registry for Internet Numbers (ARIN) – North America and parts of the Caribbean

• Reseaux IP Europeens Network Coordination Centre (RIPE NCC) – Europe, the Middle East, and Central Asia

• Asia-Pacific Network Information Centre (APNIC) – Asia and the Pacific region

• Latin American and Caribbean Internet Addresses Registry (LACNIC) – Latin America and parts of the Caribbean region

• African Network Information Centre (AFRINIC) – Africa

Per standards, each RIR must maintain point-of-contact (POC) information and IP address assignment.

Traceroute is another handy software program that helps to determine the path a data packet traverses to get to a specific IP address. Traceroute, which is one of the easiest ways to identify the path to a targeted website, is available on both Unix/Linux and Windows operating systems.

Information that can be uncovered online can include:

• posted pictures, video, or information

• posted content about personal activities, political or activist affiliations, and beliefs

• posting derogatory information about previous employers, coworkers, or clients

• discriminatory comments or fabricated qualifications

Many applications were not built with security in mind. Insecure applications, such as Telnet, File Transport Protocol (FTP), the “r” commands (rcp, rexec, rlogin, rsh, etc), Post Office Protocol (POP), HyperText Transfer Protocol (HTTP), and Simple Network Management Protocol (SNMP), operate without encryption. What adds to the problem is that some organizations even inadvertently put this information on the web.

Here are some of the defenses that can be used to thwart footprinting:

• website – any organization should take a hard look at the information available on the company website and determine whether it might be useful to an attacker. Any potentially sensitive or restricted information should be removed as soon as possible along with any unnecessary information.

• Google hacking – this attack can be thwarted to a high degree by sanitizing information that is available publicly wherever possible. Sensitive information, either linked or unlinked, should not be posted in any location that can be accessed by a search engine, as the public locations of a web server tend to be.

• Job listings – when possible, use third-party companies for sensitive jobs so the organization’s identity is unknown to all but approved applicants. If third-party job sites are used, the job listing should be as generic as possible, and care should be taken not to list specific details or versions of applications or programs. Consider carefully crafting job postings to reveal less about the IT infrastructure.

• Domain information – always ensure that domain registration data is kept as generic as possible and that specifics, such as names, phone numbers, and the like, are avoided. If possible, employ any one of the commonly available proxy services to block the access of sensitive domain data.

• Personnel social media posts – be especially vigilant about information leaks generated by well-intentioned personnel who may post information in technical forums or discussion groups that may be too detailed. It is not uncommon for information leakage to occur around events such as layoffs, mergers, or contract terminations.

• Insecure applications – make it a point to regularly scan search engines to see whether links to private services are available (Terminal Server, Outlook Web App [OWA], virtual private networks, and so on). Telnet and FTP have similar security problems because they both allow anonymous logon and passwords in cleartext. Consider replacing such applications with a more secure application, such as Secure Shell (SSH), or comparable wherever possible.

• Securing DNS – sanitize DNS registration and contact information to be as generic as possible (for example, “Web Services Manager”, main company phone number 555-1212, techsupport@hackthestack.com). Have 2 DNS servers – one internal and one external in the demilitarized zone (DMZ). The external DNS should contain only resource records of the DMZ hosts, not the internal hosts. For additional safety, do not allow zone transfers to any IP address.