Chapter XLII
Identification and Localization of Digital Addresses on the Internet

André Årnes
Norwegian University of Science and Technology, Norway

ABSTRACT

A central issue in assessing and responding to an attack on the Internet is the identification and localization of the attackers. In information warfare and cyber terrorism, an attack can be launched using a large number of hosts, in which case fast and accurate identification and tracing is crucial for handling and responding to the attack. In the digital world of the Internet, however, there are many cases where a successful trace is difficult or impossible. The design of the Internet, as well as services that hide the origin of communication and provide anonymity, complicate tracing and create a need for a wide range of tools for tracing. In this chapter, we provide a survey of different tools and services available for tracing the geographic location of hosts and users on the Internet. We consider both active and passive methods of identification and tracing. A passive trace uses information that is available through public sources, in log data, or through commercially available databases. Active methods involve the use of tools for probing the attacking party directly, for example, through scanning and pinging. Some of the methods for locating addresses on the Internet have been developed for use in electronic commerce and marketing applications, but the basic principles are equally applicable to digital investigations and information warfare. We consider only tracing of addresses on the Internet. Consequently, this chapter only considers the Internet Protocol (IPv4 and IPv6), as well as higher level protocols using IP (such as TCP, UDP, and HTTP). We refer to the host that we try to identify as the target host and its address as the target address. The system used to execute the tracing is referred to as the trace host.

AN INTERNET PRIMER

The Internet is the descendant of the U.S. Defense Advanced Research Projects Agency (DARPA) project ARPANET, whose first node was connected in 1969. The core protocol suite, TCP/IP, was introduced when the National Science Foundation (NSF) established a university network backbone in 1983. In 1991 Tim Berners-Lee at CERN in Switzerland publicized the basic protocols for the World Wide Web (WWW). The Internet was publicly known by the mid-nineties, and it is now an integral part of our society. As we have grown more dependent on Internet technologies, our society has also become more vulnerable to attacks; both on the digital infrastructure itself and on critical infrastructure connected to the Internet.

The Internet is a network of networks communicating according to a suite of standardized protocols. The physical network consists of a wide range of physical media, including optical fiber, copper cable, and wireless networks. The communication on the networks is governed by layered protocols, according to the applications in use. Most applications on the Internet rely on the Internet Protocol (IP) and the transport protocols TCP and UDP. IP is a packet-based, connectionless protocol, designed to transmit packets of data between a source address and a target address. It provides no reliability in itself, but the ability to use different routes between hosts makes the protocol very resilient to changes and disruptions on the network. An IP packet is routed between two hosts by intermediate routers. Each router makes a decision of how to route its packets based on its routing policy.

We refer to a digital address as any address that identifies a user, host, or service on the Internet. Examples of digital addresses are Ethernet MAC addresses, IP addresses, AS numbers, DNS domain names, URLs, and e-mail addresses. IANA (Internet Assigned Numbers Authority) is the highest authority for the allocation of IP addresses and AS numbers. A host on the Internet is associated with multiple registration databases. In particular, its IP address is registered in an IP WHOIS database, its domain name is registered in a DNS WHOIS database, and information about its location on the Internet is provided by the routing tables. All of this information can be used to obtain information about the location and identity of addresses and users on the Internet.

In order to perform a successful trace on the Internet, it is necessary to understand the interaction between different protocols. Each protocol may have its own addressing scheme, but it may be necessary to uncover the lowest level addresses, that is, the hardware address on the physical network, in order to associate an address with a physical user or location. There are several published accounts of computer attacks that have been traced successfully. Cheswick (1990) shows us how a hacker is studied in order to learn his intent and identity, and in a book by Stoll (1989) an attack is successfully traced to an espionage agent operating in West Germany.

PASSIVE TRACING

There are multiple sources of information that can be used for passive tracing on the Internet. The most important sources are the structured databases for DNS and IP registration, as well as the routing policies of the network operators. In addition, valuable information exists in unstructured sources, such as on the WWW and on Usenet. Network operators often provide information about their network and routing policies through Looking Glass services. A passive trace implies that there is no communication with the target system.

DNS binds domain names to their respective IP addresses (Mockapetris, 1987). A DNS lookup provides the IP address of a given domain name, whereas a reverse DNS lookup provides the domain name(s) associated with a given IP address. An important tool for identifying the contact persons of a domain name is the DNS WHOIS service. DNS WHOIS is a set of publicly available databases with contact information for DNS addresses. However, there is no authoritative DNS database; many top level domains have their own DNS WHOIS service. Some top level domains also provide anonymity for their customers and will not disclose a user's identity.