

# REMOTE COMPUTER FINGERPRINTING FOR CYBER CRIME INVESTIGATIONS

Jon Novotny, Dominic Schulte, Gavin Manes and Sujeet Shenoj

*Center for Information Security, University of Tulsa, Tulsa, Oklahoma 74104, USA*

**Abstract** This paper describes a novel tool for remotely “fingerprinting” computers used in criminal activity. The tool employs network scanning and machine identification techniques to acquire a fingerprint of a computer over the Internet. The fingerprint includes identifying information about the operating system, banners, enumerations and services. Once the computer is seized, it is connected to a closed network and scanned again to produce a second fingerprint to check against the original. Two scenarios – one related to investigating pedophiles and the other involving an illegal website – are examined. Legal issues pertaining to the use of computer fingerprinting in criminal investigations are also discussed.

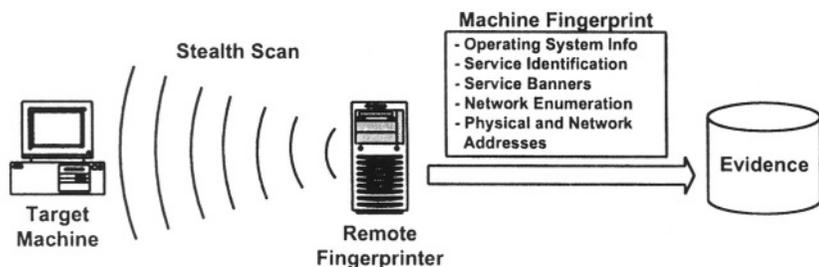
**Keywords:** Computer Forensics, Digital Evidence, Computer Fingerprinting

## 1. Introduction

The goals of a criminal investigation are to acquire evidence of illegal activity and to conclusively establish the role of the suspect. In a cyber crime investigation, it is necessary to identify the devices (computers, cell phones, PDAs, etc.) used in the illegal activity and to establish that the suspect used those devices in the crime [20].

Networked computers are increasingly used in illegal activities [6]. For example, Internet chat rooms are often used by pedophiles to reach potential victims: according to one survey, 19% of young people between the ages of ten and seventeen who used the Internet at least once a month received unwanted sexual solicitations [13].

Law enforcement agents sometimes pose as minors in chat rooms to catch pedophiles [2,11]. In atypical scenario, an undercover detective and suspect meet in a chat room, communicate using chat room applications and/or direct messaging (e.g., e-mail), and arrange a sexual liaison at which time the pedophile is arrested [17]. The evidence includes recorded chat room conversations and direct messages, which demonstrate an intent to engage in sex with the “minor”



**Figure 1. Remote machine fingerprinting.**

and the details of the rendezvous [12]. The suspect is tied to this evidence when he/she is arrested at the rendezvous point.

Remote fingerprinting can be used to obtain evidence that the suspect's computer was the one used to communicate with the undercover detective. This "machine fingerprint" includes identifying information about the operating system, banners, enumerations and services of the suspect's computer. After the suspect is arrested and his/her computer is seized, it is scanned again to produce a new fingerprint. Matching the new fingerprint with the remotely acquired one can help establish that the seized computer was used in the crime.

The remote fingerprinting strategy is presented in Figure 1. The fingerprinting tool employs network scanning techniques [1,7,8] to acquire a fingerprint of a target machine over the Internet. The fingerprint is cryptographically sealed along with other electronic evidence (e.g., chat room and e-mail transcripts). While it is preferable that the target machine be configured to permit network scans, it is possible to obtain useful identifying information even when it employs network defense mechanisms.

## 2. Network Scanning Tools

The hacker community was the first to use network-based machine exploration. Their goal was to identify hosts with specific vulnerabilities that could be exploited. Interestingly, generalized versions of early hacker tools are at the core of current strategies for mapping networks, identifying machines and discovering vulnerabilities – all intended to secure computing assets. Popular tools like Nmap [5] and ENUM are in fact the direct result of ideas, scripts and programs originally developed by hackers.

Nmap is an open source utility for network exploration and security auditing [5,8]. It transmits IP packets in carefully designed sequences to identify network hosts, open ports, services, operating systems and versions, packet filters, and firewalls, among other information. Nmap, however, has two shortcomings. First, Nmap was not developed to uniquely identify single machines; rather, it offers a broad view of multiple machines to locate common attributes. Second,

Nmap is not appropriate for cyber investigations because it does not ensure the accuracy and integrity of the data it acquires. Indeed, this limitation is true for all the tools discussed in this section.

ENUM, along with the more recent ShareScan and Netcat tools, can enumerate NetBIOS and Samba information, including network shares, user accounts and password policies. The Unix `finger` command similarly provides information about network/host users, including full name, default shell and login times.

Two other popular tools are Nessus [16] and LANGuard [10]. Nessus uses a set of plug-ins to test for the presence of various services and their vulnerabilities. It can distinguish between an SMTP server running on port 25 and port 3005; however, it uses a brute force approach that requires frequent updates. The LANGuard scanner can identify operating systems and obtain operating system specific information. It also enumerates NetBIOS groups on hosts running Microsoft operating systems and SNMP properties on network printers.

Although current network scanning tools gather useful information about target machines, no single tool provides all the desired capabilities and ensures information integrity, which is required for remote fingerprinting. What is needed is a comprehensive, reliable tool for remote machine fingerprinting that will meet the rigorous standards of evidence.

### **3. Remote Fingerprinting System**

The remote fingerprinting system refines existing network scanning techniques to acquire a fingerprint of a target machine over the Internet. All machine-specific information constituting a fingerprint is available without authentication or is available anonymously, and can be acquired by any machine connected to the Internet. No intrusion into personal data, including files and their contents, is attempted. When a machine connects to the Internet, it uses various protocols (TCP, UDP, etc.) and ports (1-65535) and applications running on these ports (`ftp` 21, `telnet` 23, etc.). This information and data gathered from the network stack and active services are used to produce a fingerprint. All the information gleaned is time-stamped, cryptographically sealed and stored in an evidence container (e.g., hard disk).

Network enumeration is the most intrusive technique used by the fingerprinter, as it potentially provides a list of files and computer settings. Note, however, that an enumeration only yields the list of resources that the user has chosen to share with other computers. Storing a hash value of the resource list instead of the list itself can mitigate the intrusiveness of this technique; fingerprint comparison would then check the stored hash value.

Remote machine fingerprinting has three components: open port identification, operating system detection and service analysis (Figure 2). Service

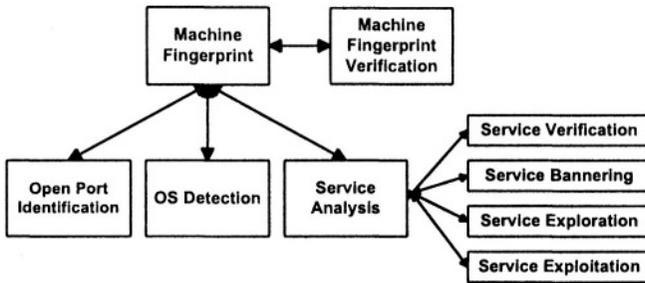


Figure 2. Machine fingerprint components.

analysis has four subcomponents: service verification, service bannering, service exploration and service exploitation. One or more of these components can constitute a machine fingerprint. Obviously, incorporating more components produces a better (more detailed) fingerprint.

### 3.1. Open Port Identification

Open ports are identified using a specially designed Java-based host port mapping tool. The list of open ports provides a foundation for operating system detection and service analysis.

The port mapping tool uses the full connect and SYN stealth techniques to identify open ports [8]. The full connect technique attempts to open a Java socket with the remote port, which requires a full TCP handshake. The SYN stealth technique transmits the initial SYN packet in the TCP handshake and waits for a SYN-ACK packet. SYN scanning requires fewer packets (it is faster) and can go undetected (half-open connections are less likely to be logged than full TCP connections). The port mapper efficiently scans all 65,535 ports, alert to the fact that users “hide” certain services in uncommon port ranges to avoid detection.

### 3.2. Operating System Detection

The TCP stack of a target machine is probed by transmitting a set of seven TCP packets. The responses are analyzed to infer operating system (OS) information.

TCP packets transmitted to the target machine vary slightly in their control bits and whether they are sent to open or closed ports [9,15]. The responses to these packets are dependent upon the TCP stack of the target machine, and help identify the OS, often down to version and/or kernel build.

TCP/IP stack analysis relies on two closely related facts. First, network protocol specifications are evolutionary in nature. For example, explicit congestion notification (ECN) proposed in RFC 3168, gives routers the ability to notify

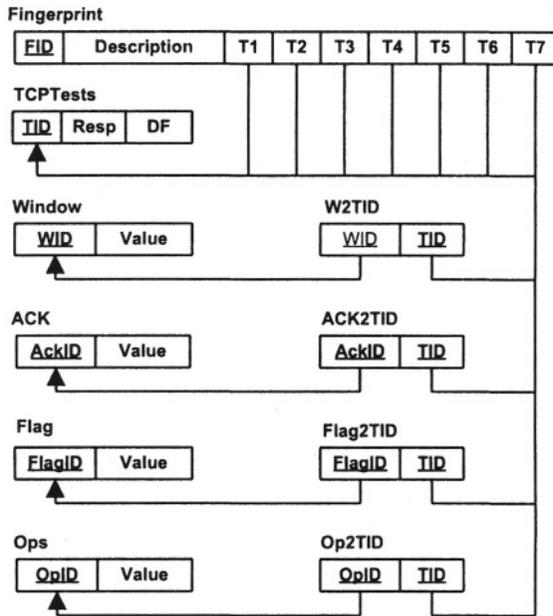


Figure 3. Operating system identification.

sending and receiving parties of persistent network congestion using certain IP and TCP header fields [18]. ECN is now filtering its way into network OSs, and is being used by some (not all) implementations. Checking the ECN field distinguishes OSs that use it from those that do not. Second, vendors often view protocol standards as recommendations, resulting in implementation variations. One example is the use of reserved space within packets, where some vendors evidently interpret reserved to mean “saved for vendor use.” Such variations help differentiate between OS specific network protocol implementations.

Figure 3 displays the database schema used in OS detection. The main table (Fingerprint) links each of the individual tests into a single fingerprint, which identifies the OS based on the test results. The other nine tables are used to store the defining attributes of a response packet, like Window Size and TCP control bits. These tables are linked using the test identifier (TID) attribute, where each TID represents a unique response received from the probe packets. This organization improves efficiency because many responses are identical according to our criteria.

### 3.3. Service Analysis

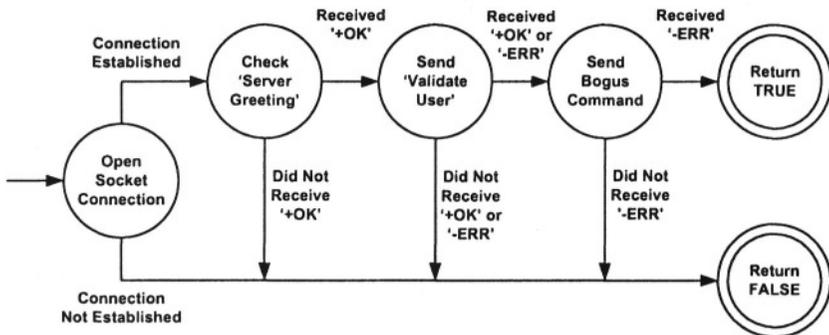
Service-related information is acquired by service analysis, which includes service verification, service bannering, service exploration and service exploitation, and uses the list of open ports provided by the port mapping tool as its

starting point. Analyzing the network services offered by the target machine provides information which differentiates the machine and is an important component of its fingerprint.

Service verification confirms the identity of services, regardless of their port numbers. This guards against cases where users mask possibly malicious intentions by hiding common services on uncommon ports. Verification is implemented using finite state machines. By focusing on RFC and vendor (e.g., ssh.com) specifications, service verification is only a matter of identifying and encoding core commands and responses. Figure 4 presents the finite state machine used for POP3 service verification.

Service banners provide a wealth of information about a machine and its users [9]. Default banners for telnet and ftp may identify the operating system down to its kernel build. Personalized banners, e.g., call signs and computer names, provide valuable identifying information. The service bannering utility connects to the available services and captures all the information offered before authentication, e.g., identifiers, greetings and prompts.

The service enumeration utility lists anonymously available data provided by a service or services. The data may include files represented in tree form from an http server or an ftp service, usernames, drives and directories on a machine, SMTP server settings, operating system password policies, routing tables, printer trays and paper types.



**Figure 4. POP3 service verification.**

Service exploitation has limited (if any) investigative applications as it potentially involves illegal activity. This technique works by identifying and exploiting vulnerabilities to obtain private information.

### 3.4. Fingerprinting Results

Identifying specific machines in networks driven by standardization and interoperability is difficult. However, malicious individuals often mask their identities and activities, e.g., by hiding chat room or file sharing services on uncommon ports, by modifying service banners, or by employing network security

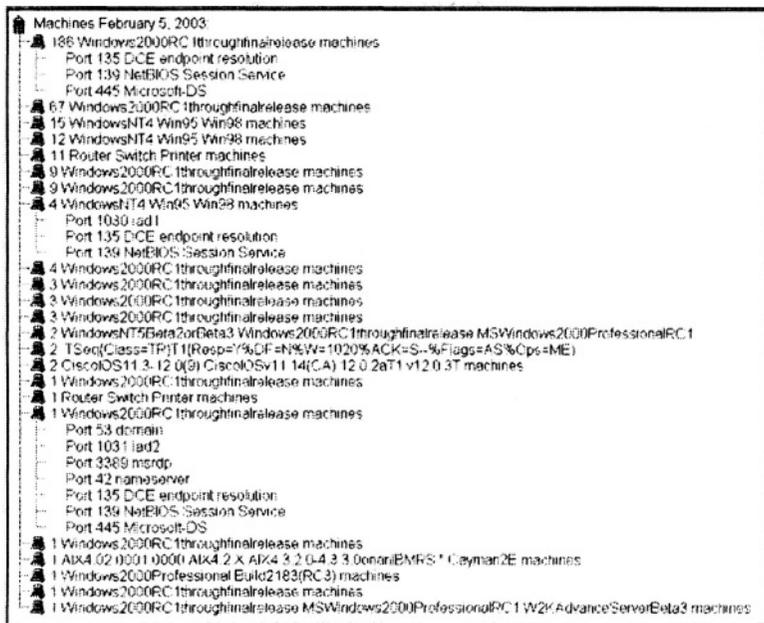


Figure 5. Fingerprint results for a production network.

mechanisms. The remote fingerprinter thrives on the uniqueness introduced by such actions.

Figure 5 summarizes the results of remote fingerprinting performed on a production network of 340 computers. Only two components – commonly used (2,048) open ports and operating system – are considered. More than 50% of the machines have identical profiles based on the two fingerprint components. Even so, eight machines are uniquely identified, and many more are grouped with three or fewer other machines with the same fingerprint. This is significant because computers used in illegal activities would likely fall in one of the smaller fingerprint groups or be uniquely identified. Indeed, when additional identifying components are used, the uniqueness of fingerprints is even greater.

#### 4. Investigative Scenarios

The fingerprinting technique can be used to good effect in a variety of investigations. Two common scenarios, one related to investigating pedophiles and the other involving an illegal website, are discussed.

Children are increasingly exposed to pedophiles on the Internet [6,11]. In the past, pedophiles would prey on children at school playgrounds, parks and neighborhoods; now they target children from the privacy of their own homes. According to one survey [13], 19% of young people between the ages of ten

and seventeen who used the Internet at least once a month received unwanted sexual solicitations.

Pedophiles frequently use Internet chat rooms to reach potential victims [6,17]. The chat room conversations and e-mail exchanges may escalate to meetings intended to consummate the online relationships. Sometimes these meetings progress beyond sexual abuse and rape, including kidnapping and murder [13].

One strategy used by law enforcement to apprehend online pedophiles is illustrated in Figure 6. An undercover detective posing as a minor communicates with a suspect in an Internet chat room. The detective builds a relationship with the suspect through repeated chat room communications and via direct messages (e.g., e-mail or instant messages). During the communications, the detective arranges a liaison with the suspect. Often, the detective asks the suspect to bring something (e.g., stuffed animals, alcoholic beverages or sexual paraphernalia). The item requested in Figure 6 is Duff Beer. When the suspect is apprehended at the rendezvous point with the requested items, there is evidence that the suspect was the individual who communicated with the undercover detective in the chat room.

The evidentiary strategy can be viewed as completing the triangle in Figure 6. First, all chat room conversations and direct messages, including the request to bring Duff Beer, must be captured and sealed as evidence. This evidence is acquired using chat room logging tools, monitoring software or by simply videotaping over the detective's shoulder [12,17]. Then, it is necessary to tie the suspect's computer (after it has been seized) to the sealed chat room transcripts (side 1 in Figure 6). Next, the suspect must be linked to the seized computer (side 2).

The remote fingerprinter and a simple chat room monitoring tool help complete sides 1 and 2 of the evidence triangle. The chat room monitor logs conversations and direct messages, and the remote fingerprinter links the suspect's machine to the recorded transcripts. Matching the remotely acquired fingerprint with that of the seized computer establishes that the suspect's computer was used for the online communications. Side 3 of the triangle is established when the suspect is arrested at the meeting place with the items (e.g., Duff Beer) that were specifically requested by the detective in the logged communications.

Another potential application of remote fingerprinting involves acquiring information about a website host, e.g., one offering child pornography or one belonging to a suspected terrorist group. In such a case, the fingerprinter can be used to identify the web server for when it is eventually seized.

Remote fingerprinting has many other applications. In all cases, it is important that the fingerprint be sealed (e.g., using MD5 hashing) like other electronic evidence [20]. Moreover, chain-of-custody and evidence storage standards must be guaranteed [20,22,23].

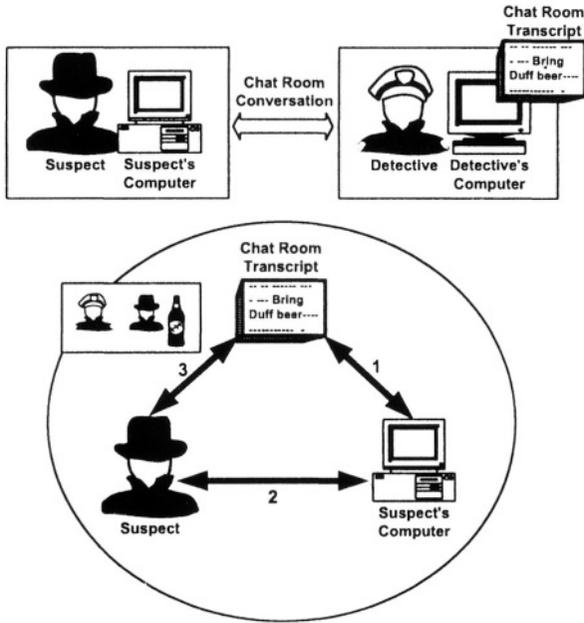


Figure 6. Pedophile investigation.

## 5. Legal Issues

No evidence acquisition technique or tool should be employed without examining the legal implications. This section considers the admissibility of remote fingerprinting evidence and the legality of its use by law enforcement.

### 5.1. Admissibility of Evidence

Admissibility refers to the principles determining whether or not particular items of evidence may be admitted into a court of law [22]. In the United States, Supreme Court decisions and the Federal Rules of Evidence (F.R.E.) define the criteria for the admissibility of novel scientific evidence [20,22].

In the landmark 1993 judgment, *Daubert v. Merrell Dow Pharmaceuticals*, the Supreme Court cast upon trial judges the duty to act as gatekeepers charged with preventing junk science from entering the courtroom [14]. To assist judges in this role, the decision specified four factors: testing, peer review, error rates and acceptability in the relevant scientific community. The *Daubert* case emphasized that a judge's inquiry is a flexible one, and its focus must be solely on principles and methodology, not on the conclusions that they generate. Although other cases have recognized that not all four *Daubert* factors apply to every type of expert testimony (see, e.g., *Kumho Tire Co. v. Carmichael*; Tyus

v. Urban Search Management), the factors can be helpful in estimating the admissibility of evidence produced by remote fingerprinting [14].

It is also necessary to consider issues related to F.R.E. requirements. In Rule 702, the requirements state: “If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case” [14].

Thus, the reliability of scientific evidence is supreme in determining its admissibility. This rule guides the use of DNA evidence in federal courts [14]. For example, since 1987 - when DNA evidence was first used in a criminal case - the evidence has generally been admitted in federal courts when it can be verified that DNA testing has been reliably applied. In the majority of cases in which DNA evidence has not been admitted, the reasons were that scientific techniques were not applied properly and/or chain-of-custody standards were not met.

Remote fingerprinting is obviously not as mature as human fingerprinting or DNA testing. As with these techniques, the technology should advance with time to become sufficiently reliable. But unlike human fingerprinting and DNA testing, remote fingerprinting may not provide a “unique” identification. Rather, we believe that a remote fingerprint – properly acquired and preserved – would constitute one more indication that the seized machine was used in the illegal activity.

## 5.2. Federal Law

If law enforcement fails to respect the privacy of individuals in its investigations, evidence will be suppressed and criminals will escape prosecution [3,20]. More importantly, the citizenry’s confidence in government will be eroded.

U.S. federal law governing law enforcement’s acquisition of electronic evidence in criminal investigations has two primary sources: the Fourth Amendment to the U.S. Constitution, and the statutory privacy laws in the Wiretap Statute (18 U.S.C. Sections 2510-22), the Pen/Trap Statute (18 U.S.C. Sections 3121-27) and the Electronic Communications Privacy Act (ECPA) of 1986 (18 U.S.C. Sections 2701-2712) [19,22]. The ECPA regulates how the government can obtain stored account information from network service providers; remote fingerprinting does not involve such information, therefore the ECPA is not relevant to this discussion.

**5.2.1 The Fourth Amendment.** The Fourth Amendment limits the ability of government agents to search and seize evidence without a warrant. We

consider the constitutional limits on warrantless searches in cases involving computers. According to the Supreme Court, a warrantless search does not violate the Fourth Amendment if: (i) the government's conduct does not violate a person's "reasonable expectation of privacy," or (ii) the search falls within an established exception to the warrant requirement [4,22,23].

The most relevant exception to the warrant requirement pertaining to remote fingerprinting is plain view. To rely on this exception, "the agent must be in a lawful position to observe and access the evidence, and its incriminating character must be immediately apparent" [22]. The exception merely permits an agent to seize evidence that he/she is authorized to view under the Fourth Amendment.

The Fourth Amendment also restricts the government's use of innovative technology to obtain information about a target [4,22]. In *Katz v. United States*, the Supreme Court ruled that a search is constitutional if it does not violate a person's "reasonable" or "legitimate" expectation of privacy [22]. Therefore, in the case of remote fingerprinting, it is important to assess whether or not an individual with a computer connected to the Internet has a reasonable expectation of privacy regarding his/her services and banner information.

*Kyllo v. United States* is also relevant [22]. In this case, the Supreme Court held that the warrantless use of a thermal imager to reveal the relative amounts of heat released from the various rooms of a suspect's home was a search that violated the Fourth Amendment. In particular, the Court held that where law enforcement "uses a device that is not in general public use, to explore details of the home that would previously have been unknowable without a physical intrusion, the surveillance is a 'search' and is presumptively unreasonable without a warrant." Use by the government of innovative technology, such as remote fingerprinting, which is not in general use to obtain information from a suspect's computer may implicate the *Kyllo* ruling. This suggests that law enforcement may not employ remote fingerprinting without a warrant.

The Supreme Court, however, restricted its holding in the *Kyllo* case to the use of technology to reveal information about "the interior of the home" [22]. Therefore, it can be argued that the *Kyllo* case would not apply to the use of the remote fingerprinter because the information gathered would metaphorically lie outside the home, within the public domain of the Internet.

**5.2.2 The Wiretap and Pen/Trap Statutes.** The Wiretap Statute (18 U.S.C. Sections 2510-2522), generally known as "Title III," is the most important federal statute governing real-time electronic surveillance in federal criminal investigations [21,22]. Title III broadly prohibits the "interception" of oral communications, wire communications and electronic communications. Failure to comply with this statute may result in civil and criminal liability, and in the suppression of evidence. However, Title III contains an exception

permitting law enforcement officers to intercept communications in which they are a party to the communication. This “under color of law” exception makes Title III a relative non-issue for remote fingerprinting [22].

The Pen/Trap Statute (18 U.S.C. Sections 3121-3127) governs the use of pen registers and trap and trace devices to collect address information and other non-content information for wire and electronic communications [22]. Because law enforcement officers are legally permitted to access the entire communication by the “under color of law” exception to Title III when they are party to the communication, it is reasonable to assume that they would be permitted to intercept the address information as well. Therefore, the Pen/Trap Statute would not pose a barrier to law enforcement officers using the remote fingerprinter.

### **5.3. State Law**

In addition to the applicable federal statutes, state law must also be considered. Although the “under color of law” exception to Title III may permit law enforcement officers to intercept communications in which they are one of the parties to the communication, not all state wiretap laws have this same exception. Indeed, some state wiretap laws require the consent of all parties to a communication. Because state law varies considerably in this regard, we recommend that law enforcement officers be familiar with the particular state laws in effect where an investigation takes place, aware that remote fingerprinting may require a warrant.

## **6. Conclusions**

As the scope and magnitude of cyber crime increase and the specter of cyber terrorism rears its head, it is imperative to design sophisticated forensic tools for acquiring evidence that will help identify, apprehend and prosecute suspects. Remote fingerprinting is a novel technique for identifying computers on the Internet. The machine fingerprint, which includes detailed information about the operating system, banners, enumerations and services, may not provide a “definitive” identifier like a human fingerprint or DNA sample. Still, a remote machine fingerprint, legally acquired and properly preserved, is one more objective indication that a seized computer was used in a crime.

## **References**

- [1] M. Andress. Network scanners pinpoint problems. *Network World*, February 2, 2002.
- [2] B. Bell. Secrets and lies: News media and law enforcement use of deception as an investigative tool. *University of Pittsburgh Law Review*, 60:745-837, 1999.
- [3] K. Connolly. *Law of Internet Security and Privacy*. Aspen, New York, 2003.

- [4] M. Elmore. Big brother where art thou? Electronic surveillance and the Internet: Carving away Fourth Amendment privacy protections. *Texas Tech Law Review*, 32:1053-1083, 2001.
- [5] R. Farrow. System fingerprinting with Nmap. *Network Magazine*, November 5, 2000.
- [6] Federal Bureau of Investigation. *A Parent's Guide to Internet Safety*. www.fbi.gov, 2002.
- [7] J. Forristal and G. Shipley. Vulnerability assessment scanners. *Network Computing*, January 8, 2001.
- [8] Fyodor. The art of port scanning. www.insecure.org, 1997.
- [9] Fyodor. Remote operating system detection via TCP/IP stack fingerprinting. www.insecure.org, 2002.
- [10] GFI Software. LANGuard Network Security Scanner. www.gfi.com.
- [11] J. Leonard and M. Morin. Stalking the web predator. *Los Angeles Times*, January 17, 2002.
- [12] A. Meehan, *et al*. Packet sniffing for automated chat room monitoring and evidence preservation. *Proceedings of the 2001 Workshop on Information Assurance and Security*, 285-288, 2001.
- [13] K. Mitchell, D. Finkelhor and J. Wolak. Risk factors for and impact of online sexual solicitation of youth. *Journal of the American Medical Association*, 285(23):3011-3014, June 20, 2001.
- [14] A. Moenssens, editor. Amendments to the Federal Rules of Evidence. www.forensic-evidence.com, 2003.
- [15] J. Nazario. Passive system fingerprinting using network client applications. Crimelabs Security Group. www.crimelabs.net, January 19, 2001.
- [16] Nessus. Documentation. www.nessus.org.
- [17] J. Novotny, *et al*. Evidence acquisition tools for cyber sex crimes investigations, *Proceedings of the SPIE Conference on Sensors and C3I Technologies for Homeland Defense and Law Enforcement*, 4708:53-60, 2002.
- [18] K. Ramakrishnan, S. Floyd and D. Black. The addition of explicit congestion notification (ECN) to IP. RFC 3168, September 2001.
- [19] E. Sinrod, *et al*. Cyber-crimes: A practical approach to the application of federal computer crime laws. *Santa Clara Computer and High Technology Law Journal*, 16:177-232, 2000.
- [20] P. Stephenson. *Investigating Computer-Related Crime*. CRC Press, Boca Raton, Florida, 1999.
- [21] R. Strang. Recognizing and meeting Title III concerns in computer investigations. *United States Attorneys' Bulletin*, 49(2):8-13, 2001.
- [22] U.S. Department of Justice. *Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations*. www.cybercrime.gov, 2002.
- [23] R. Winick. Searches and seizures of computers and computer data. *Harvard Journal of Law & Technology*, 8:75-128, 1994.