LAW ENFORCEMENT TOOLS AND TECHNOLOGIES

FOR

INVESTIGATING CYBER ATTACKS

A NATIONAL RESEARCH AND DEVELOPMENT AGENDA

INSTITUTE FOR SECURITY TECHNOLOGY STUDIES

 This project was supported under Award No. 2000-DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security.

Technical Analysis Group
45 Lyme Road
Hanover, NH 03755
(603) 646-0700
www.ists.dartmouth.edu

LAW ENFORCEMENT TOOLS AND TECHNOLOGIES

FOR

INVESTIGATING CYBER ATTACKS

A NATIONAL RESEARCH AND DEVELOPMENT AGENDA

INSTITUTE FOR SECURITY TECHNOLOGY STUDIES

 This project was supported under Award No. 2000-DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security.

Technical Analysis Group
45 Lyme Road
Hanover, NH 03755
(603) 646-0700
www.ists.dartmouth.edu

Foreword

In 2003, the cascading effects of a computer worm dubbed "Slammer" highlighted weaknesses in several interrelated computer networks that were not predicted beforehand. Exploiting a vulnerability in Microsoft's SQL database software, Slammer degraded performance in airline booking systems, bank Automated Teller Machines (ATMs) and the computer systems that control monitoring at the Davis-Besse nuclear power plant in Ohio. Further, analysis of the Slammer worm revealed that it contained no malicious payload. The damage Slammer caused was from its rapid infection of vulnerable computers, measurably hindering legitimate Internet traffic. Future cyber attacks promise to match Slammer's ability to compromise computer systems with strategically engineered payloads that could significantly impact both American national security and the economy.

In the United States, law enforcement is responsible for investigating and prosecuting the perpetrators of cyber attacks. This is no easy task. Attackers are free to mask their actions using computers in foreign countries. Readily available encryption and anonymizer technologies facilitate secure communications and privacy for law abiding citizens and cyber attackers alike. The data considered necessary to track cyber attackers is often kept for short periods of time if at all. Further, limited resources and inadequate statutes often hamper federal, state, and local agency efforts alike. The men and women who uphold our laws must have access to cutting edge technologies to facilitate their investigations and prosecutions.

This report, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Research and Development Agenda*, is a starting point for addressing United States law enforcement needs. Conducted over two years by the Institute for Security Technology Studies, the top band of technological problems of federal, state, and local cyber attack investigative community are presented. To be certain, there are no easy answers.

The effort that will be required to build solutions for law enforcement is far too large for any one institution to address. Solving these problems will require a collaborative national effort of the leading research and development centers in academia, the private sector, and government. The authors of this report have undertaken the challenge of identifying the priorities by working with law enforcement and with the research community. It is up to decision makers and researchers across government, industry, and academia to initiate and deliver innovative solutions. With no end to cyber attacks in sight, the application of science to investigative problems facing law enforcement is our only option.

Executive Summary

This paper, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Research and Development Agenda*, is the culmination of a multi-year research effort by the Technical Analysis Group at the Institute for Security Technology Studies (ISTS). Building on previous authoritative reports that call for further study of law enforcement needs, ISTS conducted a series of three focused national studies to identify, analyze and prioritize the technology needs of cyber attack investigators and prosecutors. ISTS researchers worked in cooperation with federal, state, and local law enforcement organizations, private sector groups, academic institutions, and government sponsored research and development entities in the United States to produce the *Research and Development Agenda*. The data in this report, third in the series, was collected and analyzed from September 2001 to December 2003.

In this document we present the top band of critical problem areas encountered during cyber attack investigations that may be addressed through research and development.[*] Solving the needs outlined in this work would significantly increase law enforcement's capabilities to investigate and prosecute cyber attack cases. We offer this agenda to serve as a resource for decision makers, developers, and researchers, in government, industry, and academic institutions across the country.

The *National Research and Development Agenda* addresses the following question:

> ***What are the highest priority technological impediments facing law enforcement when investigating and responding to cyber attacks, for which research and development might provide solutions?***

This document presents study data and analysis from all three reports in the series in five topic areas. Summaries of each topic area are presented in the following passages.

The Investigative Process: Preliminary Investigation and Data Collection (page 5)

Several problems faced by law enforcement during the initial stages of a cyber attack investigation may be addressed by research and development.

- Data from multiple computers is often required to proceed with a cyber attack investigation. Although a significant number of tools were identified that purported to address data collection needs, practitioners cited high product costs and lack of law enforcement-specific functionality as current product deficiencies. Solutions that can

[*] Readers wishing to learn more about all of the law enforcement needs discovered during this study may refer to the two reports that preceded the *Research and Development Agenda* titled *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Needs Assessment* and *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: Gap Analysis Report* available from <http://www.ists.dartmouth.edu/TAG>.

automate the collection of data from multiple operating systems may contribute significantly to an investigator's ability to spend investigative time focused on analysis rather than collection.

- Investigators outlined the need to quickly and accurately map a victim's network during the beginning of a cyber attack investigation. The process of manually mapping the physical and logical networks often requires the involvement of systems administrators and other staff who are familiar with the compromised network. Law enforcement investigators continually stressed that insiders may either be a suspect in the case or unskilled in their help with network mapping. Insiders pose a particularly difficult problem for law enforcement. Study participants articulated a desire for automated tools that ***map network topology and graphically represent results*** to speed up the investigative process and alleviate dependence on insiders during data collection tasks.

- Determining the presence and location of log files on multiple computers running any number of software programs is no easy task. There are numerous differences in the way operating systems and software programs log events. Discussions with study participants suggested that locating log files is becoming more complicated as data is written across network area storage and filed remotely in organizations with geographically separated offices. Additionally, investigators have found that other applications, not directly related to operating system, often include some form of event logging that may provide investigative leads. Cyber attack investigators need technology solutions to ***search, recognize, and collect logs regardless of platform or format***.

- In some situations, hard drives or entire computers are seized by law enforcement as evidence. The seizure of computers and their magnetic medium usually entails the computer being turned off. The removal of power does not usually affect hard drives, but volatile resident memory data may only be captured from a computer that is turned on. Although investigators know evidence is present, once the computer is properly seized it can be exceptionally difficult, if not impossible, for them to reliably extract the relevant data for analysis and prosecution. Cyber attack investigators are facing new challenges from software that is designed to run primarily in memory. Solutions to ***capture resident memory data*** may produce an entirely new source of digital evidence for law enforcement.

- Law enforcement officials involved in this study conveyed that working with very large data sets presents problems during cyber attack investigations. The cost of large capacity data storage devices continues to drop with no corresponding advances in technology to facilitate law enforcement's collection, analysis, or storage of large data sets. Feedback from the law enforcement community indicated that current software is not meeting their needs. For example, many of the tools in the current market are designed for forensic work on single machines in traditional crimes, not cyber attacks across networks. Study participants were clear that the amount of data in a typical cyber attack investigation is orders of magnitude larger than found in more traditional types of computer crime. In addition, the rapidly increasing size of digital storage devices is outstripping current software's ability to process the data in

a timely manner. Research is needed into innovative methods to ***analyze very large data sets*** in cyber attack cases.

The Investigative Process: Log Analysis (page 12)

The log analysis process begins after preliminary investigation and data collection tasks are completed. Logs are critical components in cyber attack cases since they often provide technical and temporal information that may further the investigation. Log analysis is a time consuming process often done manually or with the use of simple sorting and editing programs. Although finding log files manually can be difficult, correlating and examining thousands or hundreds of thousands of disparate log entries from multiple networks manually often proves impossible. Overall, solutions to assist law enforcement in processing and compiling logs into relevant case data are few. Law enforcement needs better solutions to: search, collect, and compile logs regardless of platform or format. Automating log analysis tasks would produce an immediate impact on law enforcement's ability to quickly develop investigative leads. Solutions that package logs into a common portable format would allow investigators to broadly share information, a difficult proposition in the current environment.

- Law enforcement also needs solutions that help **present detailed technical information in a graphical format**. For example, a timeline presentation of the events that occurred during a cyber attack is a critical element in the iterative investigative process. Solutions to correct time and date stamps from logs retrieved from machines in different time zones would be useful as this task is often done manually. Study participants noted that some tools were very good at presenting data in a graphical format but their cost was too great for most law enforcement agencies. Other graphical data presentation tools in use by law enforcement were designed for criminal activity analysis and although they do have import features it is unclear if they are useful for analyzing cyber attack data. Cyber attack data may be presented in a trial situation to a judge, jury, and defense attorney. Study participants considered solutions that would facilitate a prosecutor leading a jury, step-by-step, through technical evidence to be essential to successful prosecution.

The Investigative Process: IP Tracing and Real-time Interception (page 14)

- To trace the origins of cyber attacks, law enforcement looks for Internet Protocol (IP) addresses during an investigation. Unfortunately, due to the limitations of the current Internet protocol attackers are able to spoof the IP address from which their attack is launched. Investigators see the development of technology solutions to ***provide the capability to detect, trace, and counter IP spoofing*** as a priority. For the foreseeable future, it will be difficult to use technical methods to reliably detect, counter or trace spoofed traffic over the Internet. Investigators desire solutions to minimize the time spent tracing spoofed traffic so that more effort may be focused on examining legitimate investigative leads. However, the limitations of the current IP make authentication and attribution difficult. New scientific approaches are required to address this difficult, yet essential, research challenge.

- Legally authorized electronic surveillance may also be used by cyber attack investigators to acquire information on cyber attackers. Investigators require technology solutions to ***facilitate real-time interception and analysis of digital data*** including parsing, isolation, and analysis of relevant material from the large volumes of information that may be collected during surveillance. Study participants expressed a need for both speed and clarity to reduce the collected traffic to only that which is essential for the investigation, without losing any relevant data, while also protecting the rights of others whose traffic may simply be sharing the network infrastructure. Ensuring the privacy of law abiding citizens was articulated as a key issue by law enforcement during our research.

Emerging Technologies Requiring Research and Development (page 17)

- ***Encryption*** was the most critical concern of the participants that prioritized the top band of law enforcement needs presented in this study. Encryption technology is easy to use, available for all major computer operating systems, and may be applied to a variety of applications and file types. Currently law enforcement employs "work arounds" or technical means to circumvent encryption. Investigators discussed several past cases where a password was discovered either through witness cooperation or through discovery of the password text within another file. In another case, a keystroke logger was used to capture a password for an encrypted file. However, law enforcement needs additional solutions since the access, opportunity, technical skills, and resources to install a keystroke logger will not be available in many situations. Whether focusing on decryption, password recovery, or discovering other clues on a computer or networked system that ultimately lead to a password or pass phrase, it is clear there is an urgent need for significant research and new solutions in this area. Solutions to circumvent encryption could significantly benefit law enforcement and may require new scientific approaches.

- Digital ***steganography*** is a term used to describe techniques for hiding data within a digital file in such a way that it is difficult to discern the presence or content of the hidden data. Commercial steganographic software programs and home-grown tools use any number of approaches and algorithms to hide messages or data. Study participants were aware of the use of steganography as a method of hiding evidence. They were also aware of the difficulties in detecting its presence. Study participants displayed an awareness of ongoing research concerning the discovery of steganography in digital files. Developing solutions to this challenge will likely involve innovative research and the application of new scientific approaches. The research and software development community faces a challenging task to develop solutions to assist investigators in discovering the use of steganography.

National Information Sharing (page 19)

- Investigators analyzing cyber attack case data look for patterns or profiles in cyber attack data to try and identify attackers. In many cases law enforcement indicated that multiple agencies may be working independently on cyber attack cases that all

originate from a single attacker. Law enforcement would welcome technological solutions such as a ***database for collecting attack profiles in concert with a solution for technical exploit matching to identify attack patterns***. Such information sharing technologies may automate pattern analysis and technical exploit identification across geographic locations. Law enforcement would welcome technologies to act as a ***database for cyber attack signatures that allows law enforcement to assess if their case is a component of larger criminal activity***. Interagency communications are especially important in cyber attack cases due to the relatively narrow window of opportunity available to collect information to further the investigative process. Study participants noted a number of organizational systems are currently available to ***facilitate cross jurisdictional communications;*** but coordination is an ongoing problem. Although the organizations and technical solutions noted in this study are helping, study participants commented that they often rely on personal contacts to meet their needs.

We anticipated development of new solutions during the research period, however, many of the needs promise to be ongoing issues. Several cross cutting themes emerged during our research. First, there exists an immediate and growing need to automate tasks in the investigative processes. As speed is often essential to the success of cyber attack investigations, solutions that allow investigators to spend more time analyzing data rather than collecting and organizing it would be extremely useful. Second, many current tools do not produce evidence-quality data. Many of the tools law enforcement is using are not specifically designed for criminal investigative use. For example, we found that hacking, cracking, and system administration tools were employed by cyber attack investigators. Although evidence-quality data may not be critical in all task or applications, developers of new solutions should be aware of legal requirements. Third, law enforcement noted that many existing tools cost too much for some organizations to acquire. Fourth, solutions that will help alleviate law enforcement reliance on insiders and individuals who may be suspects in cyber attack cases are in short supply. Solutions that are able to collect data without insiders help would be of great benefit to law enforcement.

During this research the resourcefulness displayed by law enforcement when performing complex tasks with limited resources was extraordinary. Creativity and workarounds are often used in an asymmetric fashion to successfully investigate and prosecute cyber attacks. Although progress has been made, study participants noted that improvements are needed in all of the areas noted in this study. For example, public/private research partnerships are continuing to be developed nationwide to combat cyber attacks. These collaborative relationships are reported to partially meet law enforcement, academic, and private sector needs for research, development, and information sharing. However, law enforcement would like to see new scientific approaches and technologies brought to bear on their requirements so they needn't rely on creative, but often temporary, solutions. Solving any of the needs we outline on behalf of the law enforcement community would have a significant impact on our ability to successfully investigate and prosecute cyber attackers and contribute to national security.

Contents

Introduction

This paper, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Research and Development Agenda*, is the culmination of a multi-year research effort by the Technical Analysis Group at the Institute for Security Technology Studies (ISTS).[1] Building on previous authoritative reports that call for further study into law enforcement needs, the ISTS conducted three focused national studies to identify, analyze and prioritize the technology needs of cyber attack investigators and prosecutors.[2] ISTS researchers worked in cooperation with federal, state, and local law enforcement organizations, private sector experts, academic centers, and government sponsored research and development entities in the United States to produce the *Research and Development Agenda*. This report is based on data collected and analyzed from September 2001 to December 2003.

In this document we present the top band of critical problem areas encountered during cyber attack investigations that may be addressed through research and development.[3] Solving the needs outlined in this work would significantly increase law enforcement's capabilities to investigate and prosecute cyber attack cases. We offer this agenda to serve as a resource for decision makers, developers, and researchers, in government, industry, and academic institutions across the country.

Nature of the Problem

Cyber attacks on government, corporate, and academic networks are increasing in number, sophistication, and severity.[4] The rate of information technology advances, the speed at which they are widely adopted, the number of vulnerabilities routinely disclosed for critical programs and the development of new cyber attack tools and methods all show no signs of slowing. For example, the 2000 Code Red worm had a vulnerability-to-exploit cycle of about 10 months. In 2003, the MSBlaster worm vulnerability-to-exploit

[1] The ISTS conducts interdisciplinary research and development projects addressing the challenges of cyber and homeland security. For more information please see <http://www.ists.dartmouth.edu>.

[2] Cyber attacks are defined for this study as computer attacks that can undermine the confidentiality, integrity, or availability of a computer or information resident on it.

[3] Readers wishing to learn more about all of the law enforcement needs discovered during this study may refer to the two reports that preceded the *Research and Development Agenda* titled *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Needs Assessment* and *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: Gap Analysis Report* available from<http://www.ists.dartmouth.edu/TAG>.

[4] Institute for Security Technology Studies, *Cyber Attacks During the War on Terrorism: A Predictive Analysis* (September 2001), <http://www.ists.dartmouth.edu/ISTS/counterterrorism/cyber_attacks.htm>.

cycle was approximately six weeks.[5] Conversely, the development of investigative tools for use by law enforcement has not kept pace with the rapidly evolving instruments employed by attackers.

Research and development of technologies to address the critical needs of law enforcement is sorely needed. Authoritative reports, such as the National Institute of Justice 2001 *Electronic Crime Needs Assessment for State and Local Law Enforcement* have pointed to the need for research on cyber attack tools and technologies:

> There is a significant and immediate need for up-to-date technological tools and equipment for state and local law enforcement agencies to conduct electronic crime investigations. Most electronic crime cases cannot be thoroughly investigated and developed without the benefit of higher end computer technology, which is beyond the budgets of many law enforcement agencies.[6]

Currently, commercial research is largely focused on security products likely to yield near-term profits, and therefore may not adequately address the needs of the relatively small law enforcement market. Government funded research may address some law enforcement priorities, but without a guidebook detailing the critical needs of the community, organizations attempting to develop tools and technologies for cyber attack investigators may expend significant resources in non-critical areas. A significant benefit of the *Research and Development Agenda* may be to focus the efforts of the research community towards the priority needs of law enforcement for use in cyber attack investigations. As of this writing, no other national study identifying or prioritizing research areas in this domain has been identified by ISTS researchers. The completion of the first such document will serve as an initial assessment of current capabilities and research priorities. The completion of subsequent studies will allow the nation to benchmark progress in this critical area and guide future research as new technologies shape priorities.

Object of the Study

The *National Research and Development Agenda* addresses the following question:

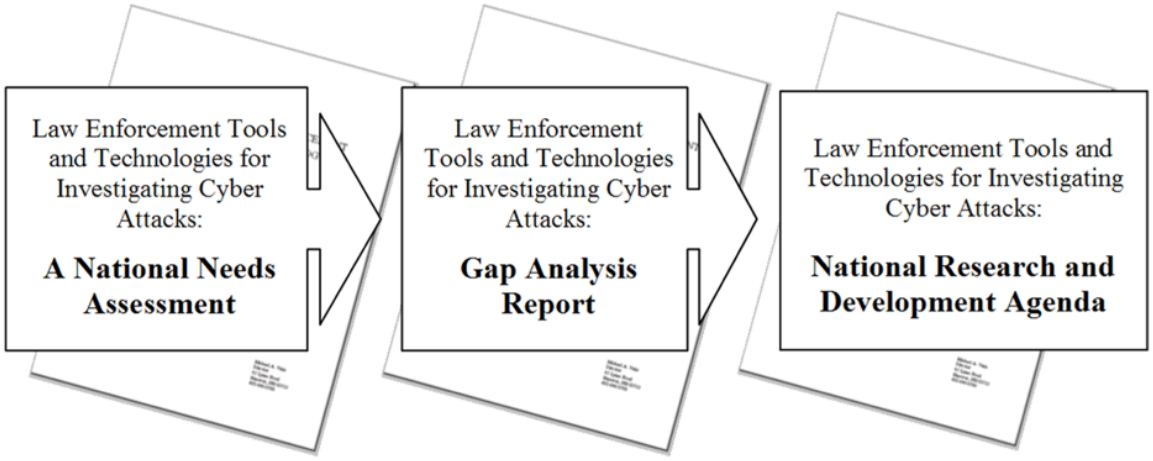
> ***What are the highest priority technological impediments facing law enforcement when investigating and responding to cyber attacks, for which research and development might provide solutions?***

[5] For example see Kevin O'Shea, *Examining the RPC DCOM Vulnerability: Developing a Vulnerability-Exploit Cycle* (October 6, 2003), <http://www.sans.org/rr/papers/index.php?id=1220>.

[6] National Institute of Justice, *Electronic Crime Needs Assessment for State and Local Law Enforcement* (April 2001), <http://www.ncjrs.org/pdffiles1/nij/186276.pdf >.

Methodology

We developed a three-phased approach to identifying, validating, and reporting the technological impediments facing law enforcement when investigating and responding to cyber attacks, for which research and development might provide solutions. The following diagram illustrates the three-step process we employed:



As a first step towards producing the *Research and Development Agenda*, ISTS published *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Needs Assessment* in June 2002.[7] The ISTS *National Needs Assessment* provides a comprehensive look at the problems and technological impediments facing federal, state, and local law enforcement when investigating and responding to cyber attacks. Data for this project was collected through a national web based survey of law enforcement personnel, site visits, and interviews with law enforcement investigators. The findings of the study were validated through a working group of practitioners.

The second step was to identify available technology solutions that purported to address the requirements revealed in the *National Needs Assessment*. A national outreach effort was conducted including contacting practitioners, software vendors, and researchers. We collected over 200 existing and in-development solutions. We mapped the collected tools against the needs from the *National Needs Assessment*, based solely on manufacturers' claims, to determine where 'gaps' in product availability existed. ISTS researchers convened a working group of practitioners to examine the collective data from the *National Needs Assessment* and the tool collection efforts. The group was asked to determine which needs from the *National Needs Assessment* were not satisfied by existing solutions and still required research and development. The tool collection efforts, analysis of the needs versus available tools, and the details of the working group are

[7] Institute for Security Technology Studies, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Needs Assessment* (June 2001), <http://www.ists.dartmouth.edu/TAG/lena.htm>.

presented in the February 2004 ISTS report *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: Gap Analysis Report*.[8]

The third and final stage of this three-part study is detailed in this report: *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Research and Development Agenda*. This paper is the culmination the research and analysis conducted during all three stages, and is based on the prioritized, unsatisfied needs revealed during the *Gap Analysis Report* working group. A major outcome of the Gap Analysis Report working group was the realization that all the needs that had been identified in the *National Needs Assessment* remain law enforcement priorities. The primary focus of the *National Research and Development Agenda* is to present and discuss the top band of critical needs.

Structure of the Report

The *Research and Development Agenda* presents study data and analysis in the following topic areas:

1. Investigative Process: Preliminary Investigation and Data Collection
2. Investigative Process: Data and Log Analysis
3. Investigative Process: IP Tracing and Real-time Interception
4. Emerging Technologies Requiring Research and Development
5. National Data and Information Sharing

Each of the topic areas is introduced with a Background narrative. This section is intended as a high-level overview of the key issues facing law enforcement. The Background narrative in each section is followed by a discussion of the corresponding critical needs and the collective analytical conclusion derived from all three stages of this research.

[8] Institute for Security Technology Studies, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: Gap Analysis Report* (Feburary 2004), <http://www.ists.dartmouth.edu/TAG/gap_analysis.htm>.

Priority Research Needs

The Investigative Process: Preliminary Investigation and Data Collection.

Background

The Preliminary Investigation and Data Collection topic area presents problems law enforcement faces during the initial stages of a cyber attack investigation. From law enforcement's standpoint a typical investigation begins when a victim files a complaint with a law enforcement agency. After the incident is reported, the investigator(s) assigned to the case will contact the victim and usually speak directly to the parties responsible for computer security. The individuals responsible for assisting investigators are often the network system administrator(s), in-house investigator(s), or external security consultant(s).[9]

Ideally, an attack is detected and reported to law enforcement while in progress so that logs[10] and trace data may be collected by the victim in collaboration with investigators. More realistically, investigators are notified after an attack is over and when the recovery of logs and other data from all involved parties, including the victim and Internet Service Providers (ISPs)[11], will be more difficult or impossible to obtain. Since the data that is used to track cyber attacks is often unavailable in relatively short order, timely data recovery will often have a dramatic effect on the success of the investigation.

Cyber attack investigations often involve multiple victims. The process for tracing an attacker's activities in a network is often repeated at multiple locations. It is not uncommon for preliminary investigation and data collection tasks to be revisited several times during an investigation as new materials become available, new attacks occur, and new leads are developed through the analysis of data. An investigation is truly an iterative process, with each cycle leading to the discovery of more relevant information.

[9] See the ISTS *National Needs Assessment* for further background discussion of the five topic areas presented in this study.

[10] Log files: computer files where devices store data about activity on a computer or network.

[11] Internet Service Provider (ISP): A company that offers access to the Internet as a service to individuals or organizations via dial-up telephone lines or direct network connections. ISPs may be local, regional, or national.

Law enforcement requires better solutions to address the following priority needs: Cyber attack investigators have encountered specific problems such as the expedient collection of data from multiple operating systems.[12] Determining the logical and physical topology[13] of a victim's network was cited as another significant impediment to an investigation. Investigators need solutions to support data collection tasks when the quality or integrity of involvement by in-house staff is in question. Additionally, capturing volatile and transient data from computer memory was noted by investigators as an ongoing need. The automation of processes that are currently undertaken manually is a cross cutting theme articulated by law enforcement during this study.

Automate the Collection of Data from Multiple Operating Systems

Data from multiple computers is often required to proceed with a cyber attack investigation. For example, log files can provide information on when and how a cyber attacker compromised a computer. Currently, investigators often collect cyber attack data manually. Investigators want solutions that can automate the collection of data from multiple operating systems across computer networks. Practitioners also articulated the need for solutions to identify and report system configurations and file locations. Solutions that can automate cyber attack data collection may contribute significantly to an investigator's ability to spend investigative time focused on analysis rather than collection.

Modern computer networks are often heterogeneous in nature. The devices present on the network and the operating systems running on the devices are often from different manufacturers. Investigators must be able to collect and correlate relevant data from any operating environment. The *National Needs Assessment* revealed that, although Windows operating system(s) dominate investigator's caseloads, UNIX-based operating systems are found in some cases. As new and updated operating systems are introduced to the marketplace, the challenges facing investigators collecting relevant data from heterogeneous networks will become more complicated. For example, new software such as peer-to-peer and instant messaging clients may produce data that would be of use to investigators. Overall investigators articulated a need for law enforcement specific tools to address their data collection needs.

Although a significant number of tools were identified that purported to address data collection needs during the *Gap Analysis Report*, practitioners cited high product costs and lack of law enforcement specific functionality as current product deficiencies. For example, some practitioners noted that existing tools did not provide robust automation of tasks, gathered only historical information, and collected more data than required

[12] Operating system: The collection of software responsible for booting a computer and providing basic libraries and tools for using the machine. Operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files and directories on the disk, and controlling peripheral devices such as disk drives and printers.

[13] Network topology: The layout and organization of a computer network.

leaving a lengthy analysis procedure for investigators. Practitioners feel that the need to automate and filter the collection of data across disparate operating systems and devices will continue to be an important characteristic cyber attack investigative solutions. The success of a cyber attack investigation often rests on quickly collecting data and following the resulting leads. Automating the identification and collection of relevant case data before it is no longer available would significantly increase law enforcement's investigative capabilities.

Map Network Topology and Graphically Represent Results

In the *National Needs Assessment*, investigators outlined the need to quickly and accurately map a victim's network during the beginning stages of a cyber attack. A critical component of generating a network map is the ability to recognize and detect the myriad devices that are present on the network. As the discussion for the Collection of Data from Multiple Operating Systems illustrates, modern computer networks are complex heterogeneous systems. Networks often have a number of devices such as firewalls[14], routers[15], and even printers that have built-in computing systems sufficiently powerful and autonomous to contribute to a cyber attack.

Any network design has two key elements. First there is the arrangement of hardware devices including connections and cabling called the physical network. Second there are the names and numbers used in software and the software links termed the logical network. The distinction is important because two internet addresses close to each other in the logical network (e.g. 192.168.100.2 and 192.168.100.4) could be physically located thousands of miles apart. Similarly, knowing the physical location of devices will not provide investigators with information on a network's logical structure.

Investigators articulated that determining the structure of a network, both in its physical and logical structures, is becoming increasingly critical to a successful investigation. The process of mapping both logical and physical network relationships is often accomplished manually. Our research indicated that there are network management tools available that can map and monitor a given network. However, it was noted by study participants that these tools do not meet the requirements of the investigative community. For example, the programs or suites of programs that were discovered were often designed to be installed by system administrators as part of their network management functions and are not tailored to investigative purposes. In addition, practitioners noted that network management tools that require pre-installation, of limited forensic capacity, or priced beyond the reach of many investigating organizations are of limited utility. Thus, if a victim does not already have network mapping software in place it is likely that investigator will have to build physical and logical network maps manually.

[14] Firewall: A system designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or a combination of both.

[15] Router: A device that determines the next network point to which a packet should be forwarded toward its destination.

The *Gap Analysis Report* identified few tools to meet investigators' need to automate the process of network mapping. The process of manually mapping the physical and logical networks often involves systems administrators and other staff who are familiar with the compromised network. Law enforcement investigators continually stressed that these insiders may be either a suspect in the case or unskilled in their help with network mapping. Insiders pose a particularly difficult problem for law enforcement.[16] Study participants articulated a desire for automated tools that would alleviate dependence on insiders during data collection tasks.

The ability to quickly visualize a network structure is an essential part of the investigative process, allowing a better perspective of physical access, vulnerabilities, and data flow. Solutions to address law enforcement needs should include graphically represented results for both the physical and logical arrangements of a network, as each lends themselves to different and supportive investigative techniques. New solutions that display relationships in physical and logical maps may provide additional investigative leads. Law enforcement desires network mapping solutions that have the ability to run autonomously or with minimal specialized training for investigators. Solutions that address network mapping needs may also be useful for presenting cyber attack data in the courtroom.

Search, Recognize, and Collect Logs Regardless of Platform or Format

The *National Needs Assessment* reported that cyber attack investigators need technology solutions to search a network for logs, recognize the locations where log files are kept, and collect them regardless of the platform or format of logs on a computer. The first step in log collection is to determine what devices are present on the network. Once devices containing logs have been determined, log files and their locations must be identified.

Determining the presence and location of log files on multiple computers running any number of software programs is no easy task. There are numerous differences in the way operating systems and software programs log events. Discussions with study participants suggested that locating log files is becoming increasingly more complicated as log files are written across network area storage and filed remotely in organizations with geographically separated offices. Additionally, investigators have found that other applications, not directly related to operating system, often include some form of event logging that may provide investigative leads. For example, one study participant noted "logs are tough to find...we've once found an obsolete marketing tool run by a little guy hiding in a small department that actually had valuable information for us." The true breadth of the problem facing investigators was summarized by a participant who wrote: "Another problem is not just the various logs generated by the [operating system], but also application logs." Critical data may be available for only a short period following a cyber attack; therefore, investigators require the ability to collect the relevant data in a

[16] The United States Secret Service in collaboration with CERT/CC of Carnegie Mellon University analyze the physical and online behavior of insiders prior to and during network compromises.

timely manner. Investigators need new approaches to automate this process to reduce the amount of time required to perform this task.

The *Gap Analysis Report* revealed that several solutions purport to address investigators' needs. Investigators indicated that current software may be too complex or expensive—limiting utility to the investigative community as a whole. Study participants were also concerned that the tools must keep pace while the formats of files and programs change. This includes the ability to recognize that a file contains log-type data during a discovery phase, as well as the ability to actually read logs that may be in a proprietary format. One participant noted that they are "not aware of any tool that will look across multi-server networks (domains) for the various types of log files."

Investigators related that they are often forced to rely on victim system administrators or other insiders to locate and retrieve appropriate log files and related data. A study participant remarked that law enforcement has "doubts that any of this software can do this without the cooperation of the system that maintains the logs." This poses two problems. First, when a system administrator is not available to retrieve this data, study participants related that finding relevant log files on unfamiliar computers and networks is a very time consuming process. Second, system administrators or other insiders may be suspect in the investigation or provide incomplete data. An autonomous solution to recognize and collect logs from across a network, that performs independent of the network administrator's input and does so in a forensically sound method is an expressed requirement of study participants. As a participant simply noted "tools that could check the system configuration and determine where log files exist and automatically collect them…would be good." Any solution that reduces the time spent recovering case data and the investigator's reliance on in-house staff will have dramatic, positive effects on the cyber attack investigative community.

Capture Resident Memory Data

Cyber attacks investigations often follow an iterative process while recovered data is examined and analyzed, and leads are developed. Digital evidence recovery tends to focus on recovering data written to a magnetic medium, such as a computer hard drive or floppy disk. In some situations hard drives or entire computers are seized by law enforcement as evidence. The seizure of computers and their magnetic medium usually entails the computer being turned off. The removal of power does not usually affect hard drives, but volatile resident memory data may only be captured from a computer that is turned on. For example, a computer uses its Random Access Memory (RAM)[17] to store temporary data. Data stored in RAM is usually not written to hard drive or floppy. If the electricity powering a computer is turned off, the data stored in RAM is usually lost. Data stored in RAM relevant to cyber attack investigations may include passwords, encryption and decryption keys, and transcripts from online chat sessions.

[17] Random Access Memory (RAM): A type of computer memory that can be accessed randomly. A single byte (a unit of storage capable of holding a single character) in RAM can be accessed without touching the preceding bytes.

Currently, personal computers come equipped with an increasingly large amount of RAM. Cyber attack investigators are facing new challenges from software that is designed to run primarily in RAM. Many programs, including malicious software[18] and encryption[19] programs, are RAM-resident and can run from this volatile memory alone and write little, if any, data to the hard disk. Although investigators know evidence is present in RAM, and the computer is properly seized, it can be exceptionally difficult, if not impossible, for them to extract the relevant data in a manner reliable for analysis and prosecution.

Research for the *Gap Analysis Report* discovered no tools that purported to capture memory-resident data. Investigators noted their frustration at not having solutions to capture data held, or programs residing in, volatile memory. Calculating criminals may not have to employ time-intensive erasing techniques, currently necessary to scrub data from other magnetic media, to remove the evidence of the existence of attack tools, passwords, and other data that resides in volatile memory.

Solutions to capture memory-resident data must address a number of technical and evidentiary issues. For example, it was noted by study participants that the ability to capture memory-resident data, without modification, may not be practical since solutions may change the memory-resident data upon execution. Other investigators noted that the ability to capture any memory-resident data, even if some were lost, could be a valuable investigative aid. Study participants feel that solutions that address capturing memory-resident data should provide consistent, reliable, and auditable actions so that any variations in data caused by the tools would be clearly identifiable. A reliable and defensible solution for capturing memory-resident data may produce an entirely new source of digital evidence for law enforcement. The data resident in volatile memory could provide key information obtained when computers are seized. The benefit to the cyber attack investigative community would be significant.

Analyze Very Large Data Sets

Law enforcement officials involved in this study conveyed that working with very large data sets often presents problems during cyber attack investigations. The cost of large capacity data storage devices continues to drop with no corresponding advances in technology to facilitate law enforcements collection, analysis, or storage of large data sets. Law enforcement investigators need better solutions to help them sift through large volumes of data more efficiently.

[18] Malicious software: Software capable of performing an unauthorized function on a computing device.

[19] Encryption: The process of changing data into an unintelligible code. Decryption is the process of changing encrypted data back to its original state. There are two main types of encryption: asymmetric encryption (also called public-key encryption) and symmetric encryption. Asymmetric encryption uses separate but related keys (sequences of digital data) to encrypt or decrypt data. Symmetric encryption uses the same key to encrypt and decrypt data.

The amount of data collected during cyber attack investigations has significantly increased since the *National Needs Assessment* was conducted in 2001. In our research for that report, the term "excessively large" implied case files or seized RAID arrays[20] with gigabytes[21] or terabytes[22] of data that usually resided within large organizations. In the short time between the publication of the *National Needs Assessment* and this report, the ability to store gigabytes and even terabytes of data has become available to consumers. Consequently, law enforcement agencies are handling larger sets of data even if their case loads do not increase significantly. For example, the number of cases the FBI Computer Analysis and Response Team (CART) examined in FY 2002 to 2003 grew from 5,843 to 6,311—an aggregate increase of 468 cases.[23] The data examined from FY 2002 to 2003 swelled from 337 terabytes to 769 terabytes—an increase of 432 terabytes. In contemporary cyber attack cases network area storage devices that may contain critical evidence may hold hundreds of terabytes or petabytes[24] of data.

The *Gap Analysis Report* identified many tools that claim they are designed to examine large data sets. However, feedback from the law enforcement community indicated that current software is not meeting their needs. For example, many of the tools in the current market are designed for forensic work on single machines in traditional crimes not cyber attacks across networks. Study participants were clear that the amount of data in a typical cyber attack investigation is orders of magnitude larger than found in more traditional types of computer crime. In addition, the rapidly increasing size of digital storage devices is outstripping current software's ability to process the data in a timely manner. An investigator may have to perform multiple analyses on digital media requiring multiple computers. Law enforcement indicated that currently software takes hours or days to complete analytical processes. Investigators called for forensic software that is designed to search large volumes of data more efficiently.

New solutions that speed up the investigatory process in cyber attack cases would make a significant impact. Investigators and prosecutors are required to move quickly in cyber attack cases. In some circumstances, it is critical to move cases forward to prevent unwarranted delays in prosecution. In most cases, law enforcement must quickly develop investigative leads so that data that will be more difficult or impossible to obtain over time is not lost. Law enforcement related that computer forensics facilities, whether

[20] RAID: Acronym for Redundant Array of Independent (or Inexpensive) Disks. RAID usually denotes disk drives that employ two or more drives in combination for fault tolerance and performance. RAID disk drives are used frequently on servers to facilitate large data storage tasks.

[21] Gigabyte: A measure of computer data. A byte usually denotes eight bits which the computer treats as a single unit. Although mega is Greek for a million, a megabyte actually contains 1,048,576 bytes. One gigabyte is equal to 1,024 megabytes. Gigabyte is often abbreviated as G, gig, or GB

[22] Terabyte: A measure of computer data equal to 1024 gigabytes or about one trillion bytes. Terabyte is often abbreviated as TB or T-Byte.

[23] Joint Council on Information Age Crime, Computer-Related Crime Metrics Roundtable PowerPoint presentation (January 30, 2004).

[24] Petabyte: A measure of computer data equal to 1,024 terabytes or about one quadrillion bytes. Petabyte is often abbreviated as PB or P-Byte.

federal, state or local, are often working near or over capacity. When working over capacity the ability to quickly determine investigative leads from large data sets hinders law enforcement's effectiveness.

Overall, the gap between the size of storage media and law enforcement's ability to process large data sets quickly is growing. Law enforcement articulated that the ability to process very large data sets will be seen as a feature that should be integrated into many solutions developed for use in cyber attack investigations. Factual data analysis techniques[25], improved hardware, and increases in operating system speeds will continue to increase the speed of analyzing large data sets. However, it appears to investigators that the rate of increase in the size of storage media and associated volume of seized digital evidence is growing faster than the rate at which the speed of processing software is improving. Research is needed for innovative methods to accomplish the analysis in a timely manner. There may be areas where fundamental data management and innovative processing theories and techniques can be brought to bear on this challenge for the benefit of the investigative community. It is crucial for investigators and prosecutors to receive the recovered information in a timely manner. It is up to software developers and the research community to provide solutions.

The Investigative Process: Log Analysis

Background

After preliminary investigation and data collection tasks are completed the log analysis process begins. Logs are critical components in cyber attack cases since they often provide technical and temporal information that may further the investigation. For example, log files may provide information on when an attack happened and how the attacker gained entry to the compromised system. It is not uncommon for different sets of logs to be generated on a number of devices within a network. For example the web server, mail server, file server and the local machine may all maintain event logs. System administrators usually maintain logs for normal operational tasks and to assist in audits, but operational and audit logs may not contain enough detail for the successful investigation of a cyber attack. Increasing the amount and type of data captured in log files significantly increases the chance of discovering and tracking unauthorized activity. However, the increased storage space and time spent analyzing events by company employees means that in the current environment many network owners do not maintain high audit levels. As a result, logs that cyber attack investigators work with often contain limited data. Further, cyber attackers may erase or modify log files to mask malicious activity. As cyber attack investigations often involve multiple victims, log analysis is often repeated at multiple locations.

[25] Factual Data Analysis: Also known as data mining. Analysis of data using methods that search for trends or anomalies without knowledge of the meaning of the data.

Log file analysis is a critical and complex task requiring investigation by individuals with significant training and experience and a time consuming process often done manually or with the use of simple sorting and editing programs. In the *National Needs Assessment* we reported that survey respondents spent 23 percent of their time in a typical investigation interpreting and analyzing log files. Although finding log files manually can be difficult, correlating and examining thousands or hundreds of thousands of disparate log entries from multiple networks often proves impossible. Overall, solutions to assist law enforcement process and compile logs into relevant case data are few. Law enforcement needs better solutions to search, collect, and compile logs regardless of platform or format. Automating log analysis tasks would produce an immediate impact on law enforcement's ability to quickly develop investigative leads. Solutions that package logs into a common portable format would allow investigators to broadly share information, a difficult proposition in the current environment. Law enforcement also needs solutions that help present detailed technical information in a graphical format. The graphical representation of character and numerical data would help both investigators and prosecutors.

Present Detailed Technical Information in a Graphical Format

There is an ongoing need within the law enforcement community for better analytical tools. The *National Needs Assessment* presents law enforcement's need for solutions to assist and automate the analysis of log data that is currently done manually. For example, law enforcement expressed the need for easy-to-use search functions that would detect patterns and anomalies from collected log files. A study participant commented "flexibility in searching data sets and log files is critical to detecting illegal activities and complex correlations. Large data sets can often conceal much useful information." An important outcome of the log file analysis process is recognizing and understanding complex relationships that may be present within the collected data. Relationships and attack characteristics are often difficult to discern in log files. As noted in the *National Needs Assessment*, many of the data files gathered and produced during the investigative process are in the form of detailed lists of events, network connectivity descriptions and other numeric and character data. In cyber attack investigations, terabytes of data may be collected by investigators. The only way to effectively search such large data sets is through the use of automated tools. The visualization of log analysis data may reduce the time needed to understand how an attack occurred. As repeated in our discussions of law enforcement needs, reducing the time required for any aspect of an investigation will significantly enhance law enforcement's investigative capabilities. Once suspicious events are highlighted by search tools and verified by investigators, solutions are required to present the results in a graphical format.

For example, a timeline presentation of the events that occurred during a cyber attack is a critical element in the iterative investigative process. First, following the path of an intruder through a compromised system is essential for assessing the extent of the attack. Second, analysis of data from one part of a network may provide indications of other areas that were attacked, other sources of an attack, or even other separate attacks. As logs file are generated on individual computers, network events are time stamped. Time

stamps differ when networked computers' clocks are not synchronized or set to different time zones. Solutions to correcting time and date stamps from logs retrieved from machines in different time zones would be useful as this task is often done manually. Investigators need automated tools that compile and present time-corrected logs on a graphical timeline to expedite cyber attack investigations.

The *Gap Analysis Report* discovered a number of tools that purported to meet law enforcements' needs. However, over half of the *National Needs Assessment* respondents (56 percent) were dissatisfied with the tools they had available for interpreting and analyzing log files. Study participants noted that some tools were good at presenting data in a graphical format, but their cost was beyond most law enforcement agencies. Other graphical data presentation tools in use by law enforcement were designed for criminal activity analysis; although they do have import features, it is unclear if they are useful for analyzing cyber attack data. The ability to analyze data may not be critical for solutions that help prosecutors to present complex cyber attack data in a courtroom.

The same data that is used by investigators leading up to an arrest is used by prosecutors. Cyber attack data may be provided in a trial situation to a judge, jury, and defense attorney. Study participants considered solutions that would facilitate prosecutors leading a jury, step-by-step, through technical evidence as essential to successful prosecution. Additionally, the ability to graphically represent the relationships between different technical aspects of an attack is important. The ability for a prosecutor to show the development of an attack, or the progression of an attack through a network, while simultaneously explaining technical issues, would be a significant asset to law enforcement's support of the prosecutorial process. Law enforcement desires flexibility in technological solutions to facilitate graphical timeline analysis. For example, events from IRC[26] and other peer-to-peer networks may have relevance to investigators and prosecutors alike. Graphical presentation solutions should have the capability to tie events to the graphical presentation. Law enforcement recommended the extension of existing solutions, tailored to support prosecutors, as a rapid solution to their pressing needs. Study participants stressed that it was critical that developers included input from the law enforcement cyber attack investigative community to ensure the solutions were relevant to their needs.

The Investigative Process: IP Tracing and Real-time Interception

Background

To trace the origins of cyber attacks, law enforcement looks for Internet Protocol (IP) addresses during an investigation. IP is an abbreviation for Internet Protocol, an agreed upon format for transmitting data between devices. IP specifies the format of packets including the addressing scheme that identifies data's origin and destination.

[26] IRC: An acronym for Internet relay chat, a chat system developed by Jarkko Oikarinen in Finland in the late 1980s.

Unfortunately, due to the limitations of the current Internet Protocol, attackers are able to spoof the IP address from which their attack is launched. Even if an IP address appears to be valid, the perpetrator is rarely in the same geographic location as the victim. The *National Needs Assessment* reported that investigators have difficulty detecting, tracing, and countering IP spoofing, and view the development of technology solutions in this area as a priority.

Legally authorized electronic surveillance may also be used by cyber attack investigators to acquire information on cyber attackers. The *National Needs Assessment* revealed that investigators require technology solutions to assist in the parsing, isolation, and analysis of relevant material from the large volumes of information that can be collected during surveillance. Automated tools are required to alleviate the need for investigators to manually parse this information.

Provide the Capability to Detect, Trace, and Counter IP Spoofing

Cyber attack investigators use IP addresses, the 32-bit number that identifies each origin of destination of data, to trace cyber attacks. Cyber attackers routinely use software specifically created to change their IP address to conceal their identity or the origin of an attack. Malicious actors may use unassigned or unregistered IP addresses or intentionally choose a number from a known range, as they are published and available on the Internet, to mislead investigators. The act of purposely changing your IP address to reduce the ability to reliably trace communications back to their source is termed "spoofing". Attackers find that spoofing is a particularly useful technique, since at the packet level, spoofed packets are indistinguishable from normal traffic. The ease with which packet forging is being facilitated by simple-to-operate programs is bringing sophisticated techniques to cyber attackers with less technical skill. The National Needs Assessment reported that there are technological solutions network owners can employ to limit the effectiveness of spoofing attacks, but current strategies often are underutilized.

During a cyber attack investigation, investigators may uncover IP addresses that are associated with an attacker. Law enforcement may contact the administrator of the network that has registered the IP address to find out who used the IP address in question during a particular period of time. Often that administrator's records may reliably show no outgoing traffic that matches the investigators request since the traffic in question was spoofed, leaving the examiner with few investigative options.

The *Gap Analysis Report* showed that there are tools that may be useful to law enforcement, but they are often the same tools used by attackers and not designed for investigations. Available tools are principally freeware or embedded capabilities in operating systems; for example, nmap, ping, and traceroute can be used for this discovery, but they are not intended as forensically sound utilities. In some cases and uses, the tools may generate data that is not part of the original case. Although this added data may be distinct, recognizable, and minimally invasive, study participants maintained that solutions are required in this realm that produce evidence-quality data and reports.

For the foreseeable future, it will be difficult to use technical methods to reliably detect, counter or trace spoofed traffic over the Internet. Investigators desire solutions to minimize the time spent tracing spoofed traffic so that more effort may be focused on examining legitimate investigative leads, though limitations of the current IP make authentication and attribution difficult. New scientific approaches are required to address this difficult, yet essential, research challenge.

Facilitate Real-Time Interception and Analysis of Digital Data

The National Needs Assessment reported that in instances where digital interception is warranted and legally authorized, local and state law enforcement agencies have found themselves either without the necessary technology or in possession of technology that exceeds their level of training. Study participants noted that real-time data interception often results in large volumes of data. It is common for the warrants authorizing real time interception to place restrictions on what investigators can collect to data pertaining to the ongoing investigation. The need to differentiate traffic sent by those under investigation from other traffic on the network is seen as an increasing problem. Study participants expressed a need for both speed and clarity to reduce the traffic to that which is essential for the investigation, without losing any relevant data, while protecting the rights of others whose traffic may simply be sharing the network infrastructure. Ensuring the privacy of law abiding citizens was articulated as a key issue by law enforcement during our research.

Captured data must be reviewed rapidly to identify suspect IP addresses, retrieve relevant data to follow up investigative leads, and initiate additional legal processes. Investigators noted that the expedient acquisition of data during cyber attack cases often resulted in investigative leads that would over time become unavailable. The *National Needs Assessment* noted that several study participants indicated that the use of real-time data interception technology would increase their effectiveness in combating cyber attacks. A study participant emphasized:

> [… the need for] state and local law enforcement to have software that enables them to conduct real-time court authorized intercepts. … There have been many investigations that we have identified the suspects but have been unable to develop evidence to prosecute. This technology would give us another investigative avenue that we don't have now.

The *Gap Analysis Report* lists several tools to assist in capturing data in a legal manner. Study participants noted that telecommunications vendors may provide law enforcement with solutions that can identify, isolate, copy, and record transmissions in real-time, but the costs involve may be prohibitive. Many existing tools that gather network-wide information are those that are installed by systems administrators for network management. They are not forensic tools by design and tend to be expensive, training intensive, and require pre-installation before an event occurs. Others tools are often components of suites that require significant additional training to be useful. These issues combine to reduce the practical availability of current tools.

Law enforcement requires legally authorized solutions to capture, sort, and analyze data, and to automate capture with robust privacy protections built in. There have been advances in this domain. For example, researchers at the University of Michigan Center for Information Technology Integration have developed an experimental packet vault that may serve as a starting point for further research and development.[27] Tools similar to current protocol analyzers may provide additional starting points for law enforcement specific solutions. For example, protocol analyzers may be matched with an interface to simplify the creation of traffic filters that may be applied to the law enforcement data interception environment. In addition, the market for high-end networking hardware often includes interception capabilities that law enforcement may employ. Ultimately, law enforcement requires solutions specifically designed for investigative use, including forensic and privacy considerations.

Emerging Technologies Requiring Research and Development

Background

During the course of the research for the *National Needs Assessment* we asked study participants to define areas of emerging technologies that they believed would begin to impact investigations in the immediate future. Several emerging technologies were topics within the survey mechanism and other topics were derived from information gathered during the *National Needs Assessment* site visits and workshop sessions. It became clear during the remainder of our research that many technologies were already impacting cyber attack investigations. Encryption and steganography technologies present issues that require immediate attention.

Encryption

Encryption was the most critical concern of the participants that prioritized the top band of law enforcement needs presented in this study. Encryption technology is easy to use, is available for all major computer operating systems, and may be applied to a variety of applications and file types. Law enforcement encounters criminals employing freely available strong encryption technologies. The use of encryption to make data stored in computer files inaccessible may work so well that it is often not possible to discover, regardless of available computing power or investigative resources. As we reported in the *National Needs Assessment*, based on the unacceptable time constraints involved with attempting brute force attacks on encrypted data, the outright defeat of encryption may

[27] The University of Michigan Center for Information Technology Integration prototype is a cryptographically secured archiver of network packet data. This prototype Packet Vault writes captured network packets to long-term CD-ROM storage using strong encryption for later analysis and for evidentiary purposes. The cryptographic organization of the Vault permits selected traffic to be made available without revealing other traffic by encrypting each packet with a key dependent on its source and destination IP addresses. Using commodity hardware, the prototype operates with a 10 Mbps network, but requires excessive manual supervision. For more information see <http://www.citi.umich.edu/>.

not be a reasonable short-term research goal, although it may be a primary focus of mid- to long-term research efforts.

Investigators encounter digital encryption in many forms during cyber attack investigations. For example, the password needed to access a computer, the pass phrase to open a document, the key for opening encrypted email, the key for opening data on a hard drive, data streams for Internet Relay Chat, or file transfers in peer-to-peer networks all may use encryption. The *Gap Analysis Report* listed only one tool currently available that purported to meet investigators needs. A study participant summarized the problem facing law enforcement when they wrote, "The bottom line is that strong encryption works well and is a [real problem] for law enforcement."

Currently, law enforcement employs "work arounds" or technical means to circumvent encryption. Investigators discussed several cases from experience where a password was discovered either through witness cooperation or through discovery of the password text within another file. In another case, a keystroke logger[28] was used to capture a password for an encrypted file.[29] However, law enforcement needs additional solutions since the access, opportunity, technical skills, and resources to install a keystroke logger will not be available in many situations. Study participants spoke almost exclusively about system passwords and not those required for individual file or program access. Technical solutions should, however, include the capability to discover passwords from a variety of operating systems such as Windows, Linux, and Apple. Solutions should collect passwords for software applications. Additional capability to discover passwords or pass phrases used with storage encryption will likely involve greater efforts because of the variety of methods available for encryption. Whether focusing on decryption, password recovery, or discovering other clues on a computer or networked system that ultimately lead to a password or pass phrase, it is clear there is an urgent need for significant research and new solutions in this area. Solutions to circumvent encryption may require new scientific approaches that could significantly benefit law enforcement.

Steganography

Digital steganography is a term used to describe techniques for hiding data within a digital file in such a way that it is difficult to discern the presence or contents of the hidden data. There are at least ten different approaches or algorithms for steganography in use in more than 200 different software programs at the time of this writing. The methods employed may make no perceptible change to the source file. If the presence of an embedded message can be detected, extracting the message without knowing the original algorithm can prove as difficult as decrypting an encoded message. Many steganography programs employ encryption as an added layer of security, increasing the likelihood that messages cannot be found or understood.

[28] Keystroke logger: A program or device that records every keystroke typed at a computer, typically saving the keystrokes to disk or flash memory for later retrieval.

[29] For example, the United States v. Scarfo, 180 F. Supp. 2d 572 (2001) case.

Study participants were aware of the use of steganography as a method of hiding evidence. They were also aware of the difficulties in detecting its presence. Currently, the law enforcement community believes it has not encountered steganography in many cyber attack cases, but as one interviewee stated: "What you don't know about you can't address." Currently, the detection of steganography in the investigative community is generally limited to searching for the programs used to create steganography files. However, many steganography tools can run from a floppy, a USB drive[30], or RAM and leave no trace on the system after use, other than an apparently benign file. Other detection approaches involve a thorough statistical analysis.[31] Study participants displayed an awareness of ongoing research at the Department of Defense Computer Forensic Laboratory and at MITRE in addition to work being conducted by WetStone Technologies. Developing solutions to this challenge will likely involve innovative research and the application of new scientific approaches. As one participant stated, "Many great tools exist, but not one does the trick. You need multiple tools, and still that is not enough." The research and software development community has a challenging task to develop solutions to assist investigators in discovering the use of steganography.

National Information Sharing

Background

During this research a recurring theme emerged: the need for information-sharing services pertaining to cyber attack investigations. The national data and information sharing needs of law enforcement organizations vary. Some law enforcement organizations simply require the ability to identify and reach out to other law enforcement agencies that handle cyber attacks. Other agencies would like to share details of particular attacks to determine if they are part of larger criminal activity. Overall, the sharing of cyber attack information between law enforcement organizations could be improved by systematic information sharing and coordination.

Database for Collecting Attack Profiles in Concert with a Solution for Technical Exploit Matching to Identify Attack Patterns

Investigators analyzing cyber attack case data look for patterns or profiles in cyber attack data to try and identify attackers. In many cases law enforcement indicated that multiple agencies may be working independently on cyber attack cases that all originate from a single attacker. In these cases, investigators at different agencies are often only able to see single events in the overall attack. A component of pattern or profile development is analysis of the technical exploits attackers use. For example, during the analysis of log files certain patterns may become apparent that indicate a known attack profile. Recognizing these profiles during a manual search requires expert log data interpretation

[30] USB drive: A portable storage device, typically a flash memory card, which plugs into the computer's Universal Serial Bus (USB) port.

[31] For example, see the Image Science Group work at <http://www.cs.dartmouth.edu/~farid/group.html>.

skills and knowledge of continually changing exploits. Law enforcement would welcome technological solutions that automate pattern analysis and technical exploit identification across geographic locations to indicate if their case is linked to other cyber attack investigations.

Database for Cyber Attack Signatures that Allows Law Enforcement to Assess if their Case is a Component of Larger Criminal Activity

In many cases, the law enforcement investigators who participated in this study felt that the information they needed was "out there" but they did not have the capability to "get it." Law enforcement indicated that a nationwide information sharing system created for computer attack patterns or profiles would cut the time spent calling other law enforcement agencies trying to identify similar attacks. This capability is especially urgent for smaller agencies with limited resources. A national information sharing solution based on current attacks, or data generated by the analysis of previous attacks, may allow investigators to quickly assess whether their case is a component of larger criminal activity. Law enforcement supervisors commented that information sharing solutions may allow investigators to find or share new data to further stalled cases through a nationwide information sharing system. Solutions to help law enforcement quickly identify patterns and coordinate with outside agencies are sorely needed.

Facilitate Cross Jurisdictional Communications

Law enforcement investigators often need to communicate with investigators across jurisdictions during cyber attack cases. We reported in the *National Needs Assessment* that identifying and communicating with other cyber attack investigators was a recurring frustration for law enforcement. Interagency communications are especially important in cyber attack cases due to the relatively narrow window of opportunity available to collect information to further the investigative process.

The Gap Analysis Report reported that a number of organizational systems are currently available to investigators to assist in cross jurisdictional communications. Participants included the US Department of Justice 24x7 Network, FBI Legal Attaché Network, National Institute of Justice National Law Enforcement Corrections Technology Center system, the Department of Defense Joint Task Force / Computer Network Operations, American Prosecutor's Research Institute, Department of Justice Computer Crime and Intellectual Property Section, National White Collar Crime Center, Internet Crime Complaint Center, and Regional Information Sharing Systems. Listserves such as CFID, HTCIA, IACIS, and Digital-DA are "very useful for communicating with folks in other jurisdictions who can help you." Secure law enforcement-only web portals, such as CyberCop and the newly revamped Cyberscience Lab website were also suggested as great solutions for reaching across jurisdictional lines. The Department of Justice has a number of information sharing and outreach programs that assist law enforcement including the Computer Crime and Intellectual Property Section (CCIPS), the network of Assistant United States Attorneys who have been designated Computer and Telecommunications Coordinators (CTCs), and the thirteen Computer Hacking and

Intellectual Property (CHIP) Units. The National Association of Attorneys General (NAAG) maintains a computer crime contact list comprised of state and local law enforcement personnel. The InterAgency Coordination Cell (IACC), located at FBI HQ, is a federal interagency group formed to coordinate investigative and operational matters among agencies responding to cyber attacks. Lastly, the ISTS Technical Analysis Group is testing a prototype national contact index for cyber attack investigators.[32] Although the organizations and technical solutions noted here are helping, study participants noted that they often rely on personal contacts to meet their needs. Furthermore, upon reflecting on the sheer number and breadth of organizational missions, law enforcement investigators noted that efforts to facilitate cross jurisdictional communications could benefit from greater coordination.

Conclusion

In delivering a *Research and Development Agenda* specifically focused on law enforcement tools and technologies for investigating cyber attacks we have met one of the challenges put forth in the 2001 National Institute of Justice report titled, *Electronic Crime Needs Assessment for State and Local Law Enforcement*. We have conducted a national survey of law enforcement specific research needs in a critical domain.

We clearly recognize that the data presented in the *Research and Development Agenda* is a snapshot in time. As new technologies become available, new needs within the investigative community will emerge. We predict that similar studies will be needed to address future needs. That said, as we compiled the full list of needs for the *Gap Analysis Report* and top band of needs for the *National Research and Development Agenda* it became clear that the needs expressed two years ago during the *National Needs Assessment* have remained unsatisfied within the investigative community. Although we thought new solutions would emerge over the time we conducted our research, many of the needs promise to be ongoing issues. For example encryption, IP spoofing, and information sharing are areas where research and development is sorely needed to develop new solutions to increase law enforcement's capabilities.

Several cross cutting themes emerged during our research. First, there exists an immediate and growing need to automate tasks in the investigative processes. As speed is often essential to the success of cyber attack investigations, solutions that allow investigators to spend more time analyzing data rather than collecting and organizing it would be extremely useful. Second, many current tools do not produce evidence-quality data. Many of the tools law enforcement is using are not specifically designed for criminal investigative use. For example, we found that hacking, cracking, and system administration tools were employed by cyber attack investigators. Although evidence-quality data may not be critical in all tasks or applications, developers of new solutions should be aware of legal requirements. Third, law enforcement noted that many existing

[32] For more information see the Prototype National Contact Index for Cyber Attack Investigators web site at <http://www.ists.dartmouth.edu/TAG/pnci.htm>.

tools cost too much for some organizations to acquire. Fourth, solutions that will help alleviate law enforcement reliance on insiders and individuals who may be suspects in cyber attack cases are in short supply. Solutions that are able to collect data without insider involvement or knowledge would be of great benefit to law enforcement. Lastly, public/private research partnerships are continuing to be developed nationwide. These collaborative relationships aim to meet law enforcement, academic, and private sector needs for research, development, and information sharing.

As we wrote in the *National Needs Assessment,* the entities that develop technological solutions to the obstacles outlined in this study have a singular opportunity. Since existing technologies do not meet cyber attack investigators' requirements, the solutions that are developed may become widely adopted by the law enforcement community. Similar to the rise in popularity of the Windows and Linux operating systems, well designed software for law enforcement that integrates users' needs could revolutionize the speed and effectiveness of cyber attack investigations.

During our research, the resourcefulness law enforcement showed performing complex tasks with limited resources was extraordinary. Creativity and workarounds are often used in an asymmetric fashion to successfully investigate and prosecute cyber attacks. Overall, law enforcement would like to see new scientific approaches and technologies brought to bear on their needs to reduce their reliance on creative, but often temporary, solutions. Although progress has been made, participants noted that improvements are needed in all of the areas noted in this study. Solving any of the needs we outline on behalf of the law enforcement community would have a significant impact on our ability to successfully investigate and prosecute cyber attackers.

Acknowledgments

ISTS Director:

Martin Wybourne

ISTS Research Staff for the Report:
Robert Hillery
Andrew Macpherson
Kevin O'Shea

The *National Research and Development* Agenda is the culmination of a three-part research study. The following individuals directly contributed to the creation of this study:

Leo Arsenault
George Bakos
Vincent Berk
Daniel Bilar
Jay Bregman
Bill Brosius
Dan Burroughs
Kathleen Cassedy
Henry "Chip" Cobb
Julie Cullen
Garry Davis
Edward A. Feustel, Ph.D.
Matt Funk
Paul Gagnon
Trey Gannon
Eric Goetz
Mike Gray
Nicole Hall-Hewett
Colleen Hurd
David Koconis, Ph.D.
Stacy Kollias
David Kotz
Dennis McGrath
Mark Noel
Richard Scribner, Ph.D.
Steve Snyder
William Stearns
Brett Tofel

The following outside organizations and individuals directly contributed to the creation of this study:

Theresa D'Orsi
John Elder
RAND Corporation
Emily Reber, Ph.D.
John King and Tim Owen, Ventana East Corporation

The Institute for Security Technology Studies extends it sincere appreciation to the many individuals and organizations from government, industry, and academia that participated in the development of the *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Research and Development Agenda.*

@Stake
Agora
Air Force Research Laboratory
Bank of America
BOS-NET
Bell Labs
Central Intelligence Agency
California Office of Criminal Justice Planning
Carnegie Mellon / Software Engineering Institute
CERT CC
Computer and Technology Crime High-Tech Response Team (CATCH)
Connecticut Department of Public Safety
Connecticut United States Attorney's Office
Counterpane
County of Los Angeles Sheriff's Department
CyberCop Portal
Cyber Science Laboratory
Decision Strategies
Delaware State Police
Department of Defense, Computer Forensics Laboratory
Department of Defense, Joint Task Force for Computer Network Operations
Department of Justice, Criminal Division, Computer Crime & Intellectual Property Section
Department of Justice, NIJ
Earthlink
Federal Bureau of Investigation
Florida Department of Law Enforcement
Future Focus
Georgetown University - Georgetown Institute for Information Assurance Hewlett-Packard
Hartford, Vermont, Police Department
HTCIA
International Association of Chiefs of Police
Knowledge Solutions
Los Alamos National Lab
Massachusetts Office of the Attorney General Criminal Bureau
Mitre
NASA Office of Inspector General
National Association of Attorneys General
National Institute of Justice
National Law Enforcement and Corrections Training Center – North East
National Law Enforcement and Corrections Training Center – West
National White Collar Crime Center
New Hampshire Attorney General's Office
New Jersey State Police
NYECTF
New York Police Department CITU
North Bay High Technology Evidence Analysis Team
Ohio Bureau of Criminal Identification and Investigation
Pacific Institute for Computer Security, San Diego Supercomputer Center
Pennsylvania Governor's Office
Pennsylvania State Police
Philadelphia Police Department
Purdue University
Sacramento Valley High Tech Task Force
SANS
San Diego Supercomputer Center
San Diego Computer and Technology Crime High Tech Response Team
SEARCH
South Carolina Law Enforcement Division
Southern California High Tech Task Force
State of Connecticut Department of Public Safety
State Street
Stroz Associates
Tenable Security
United States Department of Justice
United States EPA
University of New Haven
United States Secret Service
University of Tulsa - Center for Information Security
Utica College of Syracuse University - Computer Forensic R&D Center
Vermont State Police
Wetstone Tech

Contact Information

Please address comments and questions to:

Law Enforcement Tools and Technologies for Investigating Cyber Attacks:
A National Research and Development Agenda

Technical Analysis Group
The Institute for Security Technology Studies
45 Lyme Rd.
Hanover, NH 03755
Telephone: (603) 646-0700
Fax: (603) 646-0660

Project e-mail: <tag@ists.dartmouth.edu>

The ISTS website is available at <http://www.ists.dartmouth.edu>

The ISTS Technical Analysis Group web site is available at
<http://www.ists.dartmouth.edu/TAG/>

Publication Notice

LAW ENFORCEMENT TOOLS AND TECHNOLOGIES

FOR

INVESTIGATING CYBER ATTACKS

A National Research and Development Agenda

First Printing:

 This project was supported under Award No. 2000-DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security. Points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security.

The authors of this report have made every effort to provide original definitions, use definitions provided in past ISTS publications, and acknowledge the sources of publicly available common knowledge definitions integrated into this document. Footnote definitions were compiled from multiple sources (including <www.cnet.com>, <www.foldoc.org>, <www.techweb.com>, <www.lycos.com>, <www.sans.org>, and <www.techtarget.com>) in addition to ISTS scientists. Due to the complexity and public availability of many of the technical definitions used in this study, ISTS acknowledges the possibility that one or more definitions may resemble definitions offered in other non-ISTS sources. We invite the authors or readers of these non-ISTS sources to notify ISTS in writing if similar language is found in this document. Upon notification we will takes steps to verify the claim. If appropriate, ISTS will insert language crediting the appropriate non-ISTS source or to change our own definitional language.