

A Road Map for Digital Forensic Research

Ву

Collective work of all DFRWS attendees

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2001 USA

Utica, NY (Aug 7th - 8th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment.

As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

DFRWS TECHNICAL REPORT

A Road Map for Digital Forensic Research

Report From the First Digital Forensic Research Workshop (DFRWS)

August 7-8, 2001 Utica, New York

Sponsored By AFRL/IFGB
Air Force Research Laboratory, Rome Research Site
Information Directorate/Defensive Information Warfare Branch

Document authored from the collective work of all DFRWS attendees by: Gary Palmer, The MITRE Corporation

November 6th, 2001 - Final Approved For Public Release

Executive Summary

The search for truth is in one way hard and in another easy - for it is evident that no one of us can master it fully, nor miss it wholly. Each one if us adds a little to our knowledge of nature, and from all the facts assembled arises a certain grandeur.

Aristotle

On August 7⁻⁸, 2001, the first Digital Forensic Research Workshop was held in Utica, New York. Over 50 university researchers, computer forensic examiners, and analysts attended the workshop. The Air Force Research Laboratory, Directorate/Defensive Information Warfare Branch, sponsored the workshop from the Rome Research Site in Rome, New York. This gathering was intended to spark discussion among academics and practitioners with experience and interest in the field of Digital Forensic Science. The objectives of this first workshop were to begin forming a community of interested individuals and to start a meaningful dialog for defining the field and identifying the difficult, high-priority challenges that lie ahead. The major goal was to establish a research community that would apply the scientific method in finding focused near-term solutions driven by practitioner requirements and addressing longer term needs, considering but not constrained by current paradigms. The results of this and future gatherings are intended for wide distribution among sponsors and consumers of digital forensic technology, such as military, civilian, and law enforcement professionals.

The workshop discussions were initiated by presentations from five invited keynote speakers, each representing a different perspective related to forensic analysis. Views on Digital Forensic Science and supporting research from law enforcement, military operations, infrastructure protection, commercial development, academia, and government were presented on the morning of the first workshop day. These presentations provided a broad spectrum of information for attendees that would form the basis of the workshop discussions to follow.

The remainder of the workshop was devoted to group discussions on the following four topics chosen both to define this new field of Digital Forensic Science and to address high-priority technology challenges:

- 1. Define a Framework for Digital Forensic Science
- 2. Discuss the Trustworthiness of Digital Evidence
- 3. Discuss Detection and Recovery of Hidden Data
- 4. Discuss Digital Forensic Science in Networked Environments (Network Forensics)

Attendees divided into five groups, discussed each topic, and then briefed findings to all present. The Workshop 1 discussion resulted in a new definition of Digital Forensic Science

that incorporates objectives of analysis beyond just those used by law enforcement. The Workshop 2 discussion concluded that digital evidence was not inherently untrustworthy but that additional research in underlying technologies was needed to attain the desired integrity and fidelity. Workshop 3 was a fascinating educational experience as many experts in the field of Steganography (a technique whereby messages of many types may be embedded within still images and video with little or no visible effect) explained ongoing research and identified additional areas needing coverage. Finally, the Workshop 4 discussions produced a working definition of Network Forensics as well as identified several major areas of concern that could be candidates for fundamental and applied research.

This document contains the DFRWS proceedings. It is intended to show those in authority with interest in and concerns about the future of law enforcement, intelligence, and information collection and analysis where expertise in Digital Forensic Science exists and how those experts see the problems that lie ahead. To that end, this document serves as an informative foundation for all future work in this area. In the interests of summarizing the results, the Road Map has been included in summary form with a glimpse of a possible timeline for the research the DFRWS would like to perform. This Road Map can be found in Appendix A. Appendix B lists workshop attendees and contact information.

DFRWS attendees consider this initial meeting a great first step toward discovery, understanding, and research in the new discipline of Digital Forensic Science. Also, the tragic events of September 11, 2001, reinforce the critical need for further research in this field. The organization will continue to sponsor additional workshops and focused gatherings as long as there are challenges to address and community interest remains keen. The DFRWS will continue to maintain and enhance the organization's web site (TBD), where members may interact via the pages and the DFRWS list server.

Table of Contents

PURPOSE	1
INTRODUCTION	1
WORKSHOP SCHEMA	3
DIGITAL FORENSIC RESEARCH PERSPECTIVES	6
"BIG COMPUTER FORENSIC CHALLENGES," BY DR. EUGENE SPAFFORD	MAN .8 ER10 11
WORKSHOP DISCUSSIONS	14
Workshop 1 - A Framework for Digital Forensic Science The Definition The Process	16
Building Expertise	
WORKSHOP 2 - THE TRUSTWORTHINESS OF DIGITAL EVIDENCE	
The Description	20
Issues	21
Research Solutions	22
WORKSHOP 3 - DETECTION AND RECOVERY OF HIDDEN DATA	23
The Definition	23
Hidden Data Placement	24
Trends in Concealment	24
Research in Detection and Recovery	
WORKSHOP 4 - DIGITAL FORENSIC SCIENCE IN NETWORKED ENVIRONMENTS (NETWORK FORENSICS)	
The Definition	
Major Issues	
Time	
Performance	
Complexity	
Collection	
Paradigm Distinctions	
Collaboration	
Legal Hurdles	
Emerging Technologies	30
CHAMADY AND CONCLUCIONS	22

PPENDIX A – A DIGITAL FORENSIC ROAD MAP		
BUILDING A FRAMEWORK FOR DIGITAL FORENSIC SCIENCE	33	
Objective	33	
Research Areas		
Payoff	33	
Timeline		
ISSUES OF TRUST IN DIGITAL EVIDENCE		
Objective	35	
Research Areas		
Payoff		
Timeline		
DETECTION AND RECOVERY OF HIDDEN DATA	36	
Objective	<i>36</i>	
Research Areas		
Payoff		
Timeline		
DIGITAL FORENSIC SCIENCE IN NETWORKED ENVIRONMENTS (NETWORK FORENSICS)		
Objective		
Research Areas		
Payoff		
Timeline		
APPENDIX B - ATTENDEES AND CONTACT INFORMATION	4(

List of Figures

rigure	Page
Figure 1 - Nucleus of Digital Forensic Research	4
Figure 2 - DFRWS Organizational Objectives	5
Figure 3 - Keynote Speakers and Forensic Perspective	6
Figure 4 - DFRWS Discussion Topics	14
Figure 5 - A Definition for Digital Forensic Science	16
Figure 6 - A Definition for Network Forensics	
List of Tobles	
List of Tables Table	Page
Table	Page
Table Table 1 - Suitability Guidelines for Digital Forensic Research	3
Table Table 1 - Suitability Guidelines for Digital Forensic Research Table 2 - Investigative Process for Digital Forensic Science	3
Table Table 1 - Suitability Guidelines for Digital Forensic Research	3 17 19

Purpose

The laws of physics, with all their logical apparatus, still speak, however indirectly, about the objects of the world.

Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (6.3431)

The purpose of this paper is to present the proceedings of the first Digital Forensic Research Workshop (DFRWS), held in Utica, New York, and sponsored by the Air Force Research Laboratory, Information Directorate/Defensive Information Warfare Branch. The goal of the workshop was to provide a forum for a newly formed community of academics and practitioners to share their knowledge on Digital Forensic Science. The intended audience is military, civilian, and law enforcement professionals who use forensic techniques to uncover evidence from digital sources.

Introduction

Providing accurate information derived through the use of proven and well-understood methodologies has always been the goal of traditional forensic analysis. Forensic Science applied in courts of law has sought to use commonly applied techniques and tools only after rigorous, repetitive testing and thorough scientific analysis. One only has to look at DNA analysis to see evidence to support this statement. Most citizens today accept DNA evidence without question. To most this type of evidence is irrefutable and uncontestable. However, DNA didn't just appear out of the blue. From all accounts it was first presented in U.S. courts in 1987,¹ a full two years after Dr. Alec Jeffreys surmised that DNA could be used to identify an individual from serological analysis. This discovery came 32 years after Watson and Crick described the DNA molecule, which in turn followed the first indication that this substance existed by 84 years.² The point here is not how long this process took but that it was, in fact, a process. Discoveries built on the solid, repeatable finding of others. Factual discovery takes time and an insatiable desire for accuracy of results as well as precision in the methodologies employed in its production. Without the rigorous process that leads to

¹ The DNA Revolution, by Katherine Ramsland: http://www.crimelibrary.com/forensics/dna/6.htm

Johann Friedrich Miescher, 1869, identified a weakly acidic substance of unknown function in the nuclei of human white blood cells. This substance would later be called deoxyribonucleic acid, or DNA. Milestones in DNA History, About BioTech, http://www.accessexcellence.com/AB/WYW/wkbooks/SFTS/sidebarmilestone.html

proven scientific discovery, decision-makers in the courts and elsewhere are left to rely on supposition or worse yet intuition in the pursuit of justice.

Because the courts are forums where information may persuade us to restrict or remove individual liberties, they have proven to be a serious testing ground for scientific research. Even after all the rigorous research that preceded that first U.S. DNA case, the Florida court still held a pretrial hearing to assess the suitability of this new type of evidence. Before DNA evidence could be admitted, a decision-maker³ had to attempt to understand what this science represented and if he/she trusted it as proof that could support due process as charged by our Constitution. Addressing suitability both in and out of the courts is at the heart of all Forensic Science, and it was, indeed, a key factor in discussions held during the workshop.

Finally, although the tragic events of September 11, 2001, had not occurred at the time of the workshop, their impact on the field of Digital Forensic Science is especially applicable to the future activities of this community. The result will be continued development of improved digital forensic tools and techniques.

The remainder of this document is structured as follows. The next section describes the workshop schema. The following two sections detail the briefings presented by keynote speakers and summarize workshop discussions, respectively. Following a summary and conclusions, the document presents two appendices: Appendix A contains the Digital Forensic Road Map developed by workshop attendees, and Appendix B lists contact information.

2

In courts of law this is either the judge or the jury and is referred to as the "trier-of-fact."

Workshop Schema

The workshop focused on three major areas in which forensic analysis is currently being employed in some form. Table 1 shows these areas, associated with a primary and secondary objective of forensic analysis as well as the temporal environment required for any analysis to be of use in supporting the primary objective.

Table 1 - Suitability Guidelines for Digital Forensic Research

Area	Primary Objective	Secondary Objective	Environment
Law Enforcement	Prosecution		After the fact
Military IW Operations	Continuity of Operations	Prosecution	Real Time
Business & Industry	Availability of Service	Prosecution	Real Time

Investigators employ a different paradigm for each area when performing analyses. That is, law enforcement can't act (or analyze) until there is sufficient reason to believe that a crime has occurred. Alternatively, military and civilian managers strive to anticipate, and take action to thwart, anomalous activity before their mission or service is interrupted. A military commander's decision to pursue the secondary objective of prosecution may involve complex political criteria and coalition factors. Confronted with that same decision, civilian management would, no doubt, have to weigh economic and financial outcomes before proceeding. All three areas listed in Table 1 are necessary to achieve total security for our nation, and all are actively pursuing forensic solutions to meet their disparate investigative goals toward that end.

Similarly, practitioners working in each area have different perspectives about what digital forensic research must offer. The intent here was to incorporate these different views into workshop discussions to allow participants to hear views and opinions they may not have otherwise considered. The hope is that the consideration of all perspectives will allow practitioners to see the benefits of long-term research and allow the academic to perform effective research by identifying realistic applications. As shown in Figure 1, to be effective, fundamental digital forensic research must provide suitable solutions with the widest possible applicability to Homeland Security. To do that the focus must be the foundation science at the root of the technologies we aim to analyze.

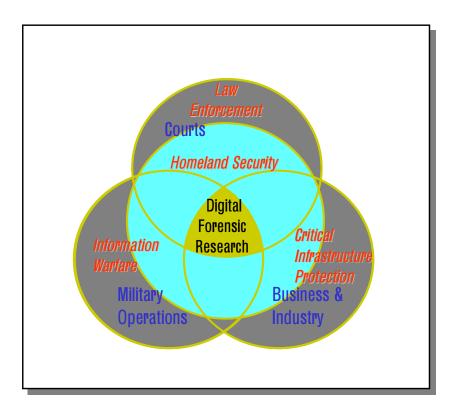


Figure 1 - Nucleus of Digital Forensic Research

The world is increasingly dependent on digital sources of information and the computerized systems and networks involved in data storage, processing, and transmission. This growing dependence drives development to advance required technology. This development results in technologies that will allow for data volumes unprecedented in our history. The pranksters, criminals, terrorists, and other nefarious members of society have not overlooked these facts. So words like *cybercrime*, *cyberwar*, and *cyberterror* have started to become more commonplace, and organizations are being formed to stop the activity these terms define.

The majority of current computer forensic analysis is focused on assisting the law enforcement community. The criteria that define suitability for forensic evidence in this area are the most clearly defined since computer forensic analysis must follow the same long-standing statutory and regulatory guidelines imposed on other, more traditional forensic disciplines. Existing technologies and those that are evolving, in support of law enforcement, will come under increasing scrutiny as technical knowledge expands in scope. For this reason, it is imperative that sound research steeped in the scientific method becomes fundamental to the discovery and enhancement of all tools and technologies employed to assist the courts, including digital forensic evidence.

Forensic analysis in the civilian and military areas is moving quickly to find ways to identify anomalous activity on networks and hosts. In these circles, you will more often hear terms like *network forensics*, *virtual crime lab*, *remote forensics*, or *cyberforensics* to describe types of analysis. Here, prosecution is, at best, a secondary objective for the evidence or proof presented to decision-makers. Although they may decide to seek legal action at a later time, their primary concerns are service availability and mission continuity. Also, managers, or those in authority, oversee clearly defined processes using finite resources within strict budgets. Any decision they make to modify steps, reallocate resources, or deviate from expected product delivery must have the most accurate information available. Therefore, the suitability criteria for forensic analysis here involve the production of high-confidence results in the shortest possible time or it serves little purpose in support of primary objectives.

Fundamental digital forensic research will serve these and other paradigms should they arise. The same core technologies are present for all computer and network users. Existing applications for digital forensic analysis all require rigor to be of any use. They will benefit from repeated testing and published error rates to limit interpretive battles. They will gain when decision-makers can point to generally accepted findings to aid in the choices they must make. To that end, this DFRWS was called to perform the tasks outlined in Figure 2.

- ⇒ Initiate a dynamic community of experts from academia and practice
- ⇒ Promote scholarly discussion related to digital forensic research and its application
- ⇒ Involve experienced analysts and examiners from LE, military and civilian sectors
- ⇒ Define core technologies that form the focus of useful research
- ⇒ Establish a common lexicon so the community speaks the same language
- ⇒ Engage in regular collaborative activity to keep focus sharp and interest high

Figure 2 - DFRWS Organizational Objectives

These tasks form the beginnings of a mission statement for the DFRWS. The remainder of this paper documents the highlights of that event as they relate to the items listed above. The speaker presentations, workshop topics, workshop briefings, and material from question-and-answer sessions are available at the DFRWS web site (TBD). Due to the sensitive nature of some of the material discussed, the site is configured for limited access.

Digital Forensic Research Perspectives

Normal science does and must continually strive to bring theory and fact into closer agreement, and that activity can easily be seen as testing or as a search for confirmation or falsification.

Thomas Kuhn, The Structure of Scientific Revolutions

Workshop activities began with a series of briefings presented by invited speakers chosen to represent those academic, operational, and commercial segments searching for clarity in digital forensic analysis. Speakers were asked to present a twenty-minute briefing outlining their perspective on the greatest challenges our maturing forensic research community would have to face in the coming months and years as well as their thoughts on community focus that would produce the greatest impact. The goals were to set the tone for workshop discussions and to create an atmosphere that addressed law enforcement's needs during group interactions but did not limit discussions to that area alone. The list of speakers, their affiliation, and associated perspectives are captured in Figure 3.

Speaker	Association	Perspective
Dr. Eugene Spafford	Purdue University	Academic Research & Government
Mr. Charles Boeckman	USTRANSCOM (Mitre Corp.)	DOD Operations
Mr. Chet Hosmer	Wetstone Technologies	Commercial Tools Development
Mr. David Baker	Mitre Corp. (NIPC)	Critical Infrastructure Protection
Dr. John Hoyt	National Institute of Justice	Law Enforcement

Figure 3 - Keynote Speakers and Forensic Perspective

The remainder of this section presents major highlights from each of these presentations. The full text and video⁴ for each speaker's presentation have been approved for limited public release and are available to authorized accounts at the DFRWS web site.

The video was created using Windows Media Tools and is therefore only playable using Windows Media Player. Version 7 or higher is recommended.

"Big Computer Forensic Challenges," by Dr. Eugene Spafford

Academic research in support of government, as well as commercial efforts to enhance our analytical capabilities, often emphasizes technological results. Although this is important it is not representative of a full-spectrum approach to solving the problems ahead. Research must address challenges in the procedural, social, and legal realms as well if we hope to craft solutions that begin to fully "heal" rather than constantly "treat" our digital ills. This full-spectrum approach employs the following aspects:

- **Technical**: "Keeping up" is a major dilemma. Digital technology continues to change rapidly. Terabyte disks and decreasing time to market are but two symptoms that cause investigators difficulty in applying currently available analytical tools. Add to this the unknown trust level of tools in development⁵ and the lack of experience and training so prevalent today and the major problems become very clear.
- Procedural: Currently, digital forensic analysts must collect everything which in the
 digital world leads to examination and scrutiny of volumes of data heretofore unheard
 of in support of investigations. Analytical procedures and protocols are not
 standardized nor do practitioners and researchers use standard terminology.
- **Social**: Individual privacy and the collection and analysis needs of investigators continue to collide. Uncertainty about the accuracy and efficacy of today's techniques causes data to be saved for very long time periods, which utilizes resources that may be applied toward real problem solving rather than storage.
- **Legal**: We can create the most advanced technology possible, but if it doesn't comply with the law it is moot.

Research Focus: Work is needed to incorporate forensic hooks into tools rather than use our current band aid approach⁶ that produces point solution tools. We need technology that isn't so easily compromised. Also, to begin to answer the problem of training and experience much more effort is required in producing user interfaces that address deficiencies in skill levels that will always be with us and will no doubt get worse as the problems grow. We need to know how much information and what type, exactly, we must collect to afford the most accurate analysis under particular circumstances. Common terms of reference are needed as well as common analytical standards and practices.

⁵ By well-meaning, hardworking persons answering an urgent call for help by investigators.

In many ways this seems to mirror the age-old problem of designing security into systems, which still presents major problems.

The social aspects of our analytical endeavors are in need of focus, too. We need tools that zero in on truly useful information and quickly deduce whether it is material to the investigation or not. We need to identify a social "end-game". Are we prepared to take serious action to thwart wrongdoing in all its forms? Do we know what the cost will be if we take that action?

All aspects of the problem are essential. Therefore, it is imperative that each collaborates with the other. Researchers, investigators, legislators, and jurists must all work toward a central goal. This requires constant discussion within groups that have representation from all essential parties.

"A Defensive Info Ops Perspective on Forensic Analysis Requirements," by Charles Boeckman

What are the major investigative contingents that drive requirements for forensic techniques and tools? How do they differ and how are they alike? From the perspective of Department of Defense (DOD) operations, on which requirements must the community of researchers focus to satisfy operational needs for current and future military missions employing digital systems and networks?

The case was made that four distinct categories of forensic consumers are currently posting requirements for consideration. The emphasis was on the DOD requirements and what they mean to researchers. The list of four is:

- Law enforcement
- Business or e-commerce
- Research and academic
- U.S. DOD

The first three categories can be described in the following way. Law enforcement requirements focus on gathering evidence for use in prosecution that will be scrutinized against established, strict judicial standards. Business requirements are driven more by economics for use in keeping the business on track using reasonably effective techniques that are cost justified and, more importantly, fast. The academic requirements are still being drafted but should focus on accuracy of result derived from precise, repeatable methods that have wide application to all forensic consumers. All represent a distinctly different approach using varied criteria.

DOD requirements assume information superiority and continuity of mission-critical operations based largely on the fairly new concept of Defensive Information Operations

(DIO)⁷. This category of system protection grew out of the long-standing work in Information Warfare,⁸ where the focus was information attack. The requirements for this new view of operational systems in the military are guided by concerns about how well techniques perform in the following areas:

- Assessing the **impact** of system compromise
- Assessing the scope of that compromise
- Assessing the **intelligence value** of the data collected
- Performing **Battle Damage Assessment** (BDA)

What differentiates these requirements from those in law enforcement is a willingness or, more correctly, the need to sacrifice absolute or even measurable accuracy for quickness in order to serve the mission's timeline. With this new directive, research efforts designed to serve DOD operations should concentrate on the following activities: (1) optimize data collection (don't collect everything because you can; know about mission-essential information), (2) minimize risk of data corruption or destruction, and (3) strive to accommodate operational time constraints.

Just as important here is the fact that systems must be analyzed while active. In most cases, this is the opposite of the current law enforcement view. Research must provide answers as to what data can be collected safely on active systems and networks and what data has the most benefit. Until forensic analysis becomes effective enough to anticipate attacks and prevent compromise, there will always be the threat that compromise may force the shutdown of mission-essential digital components. Research is needed to identify those items that can be safely collected before a shutdown and maybe even more important what operations (if any) must **not** be performed.

The author's search for any tools to help in this regard has produced few, if any, real solutions. Those that were found⁹ are point solutions with little or no claims to accuracy. Therefore, there seems to be an abundant set of forensic requirements from which researchers may draw to help guide their academic activities in support of DOD operations. One caution

DIO represents a multi-disciplinary approach to protecting digital systems. It includes COMSEC, COMPUSEC, INFOSEC, OPSEC, Physical Security, and other tactics used in active systems protection.

⁸ IW has been growing since 1989, and the advent of the Morris Worm and publication of Cliff Stoll's groundbreaking work, *The Cuckoo's Egg*, both served as a wake-up call alerting many to the need for increased network security.

See Briefing on DFRWS web site. General categories are (1) tools to help find rootkits, (2) loadable kernel modules, and (3) spurious *ps* outputs.

is that perhaps some of this research is better performed in communities that afford limited access to only those participating in the work¹⁰.

"Digital Forensic Technologies: Are We Overlooking Key Fundamentals?," by Chet Hosmer

The need for clear requirements is perhaps most apparent from the perspective of those who want to build tools to aid forensic practitioners. Certainly profit is at issue, but anyone who develops tools for this niche community knows that word of mouth about a tool's value has a direct effect on any tool's financial worth.

The perspective detailed here asked many thoughtful questions in two major categories: What are the fundamental truths of this thing we are calling *digital evidence* and what characteristics must be evident across the board for things we deem to be *cyberforensic* technologies?

Fundamental questions surrounding evidence in the digital world begin with identity, providing some digital link between binary data we collect and analyze, and the human being we call a suspect. Currently, mere possession of a digital computer links a suspect to all the data it contains. Will that be sufficient as we continue to become networked in our homes as well as businesses? Can digital data itself (excepting context) ever provide clues to motive of a crime or incident? How do we expand our digital forensic view from disk to network? Profiling, identifying, tracing, and apprehending *cybersuspects* are key issues facing research, but can digital forensic analysis supply answers to the following questions: Who?, What?, Why?, Where?, and When? Are there ever any *cyberwitnesses* to a *cybercrime*? These are some of the questions that should be considered by researchers striving to provide useful solutions for near-term use.

Other factors relate to desirable characteristics of the technologies themselves. Words like *reliable*, *precise*, *accurate*, *non-reputable*, *secure*, *flexible*, and *inexpensive* all make the short list. Researchers must also factor these into concepts, designs, experiments, and prototypes. The use of these technologies begs other questions as well. Where are the approved standards? Did the tools used apply those standards? Does digital forensic analysis employ investigative disciplines that require certification? Were the investigators certified? How is digital evidence integrity assured? Some of these questions go to practice rather than research. But in all cases we must identify and then focus on the fundamentals.

_

This call for reasonable confidentiality may have been scoffed at before the events of September 11, 2001, but its importance is now clear.

"Digital Forensics," by David Baker

The technologies we strive to understand and analyze are everywhere. We are hard pressed to identify a single item in, or segment of, normal world culture that isn't touched by a *digital* hand. Therefore, the environment that researchers and field workers encounter each day poses some very complex dilemmas for Digital Forensic Science.

Study of the digital forensic arena allows a set of fundamental problems to emerge. Most are related to missing or unconsidered steps in the investigative approach. Because of the digital ubiquity mentioned above, all crimes would soon have a *cyberdimension*. A large body of proven investigative techniques and methods exists in more traditional forensic disciplines. Most are applicable in cyberspace but are not yet considered strongly. Jurisdictional boundaries have effects outside cyberspace, and the approach needs to be reviewed and applied in the digital realm. A major related problem is how to determine and prove the real value of digital loss. Last but not least is the rapid pace of technological advancement. This reality, coupled with the variability of applications, drives researchers and investigators alike to continuously try to hit a moving target.

Acts categorized as *cybercrimes* fit a defined pattern of activity. Child pornography, identity theft, fraud, and denial of service are but a few examples. Evidence is mounting that suggests digital surveillance or reconnaissance should be added to this list as well.¹² These events have shown our nation's critical infrastructure to be vulnerable, ranging from electric power to emergency services. Digital forensic research must consider and address these new apparent threats and craft solutions that begin to move out of the after-the-fact mentality.

Across all approaches, three distinct types of digital forensic analysis can be applied:

- Media Analysis: Examining physical media for evidence
- Code Analysis: Review of software for malicious signatures
- Network Analysis: Scrutinize network traffic and logs to identify and locate

Research challenges exist in all categories listed above. Media analysis tools tend to suffer from a lack of operating system and media versatility. They are point solutions. A lack of approved standards compounds this problem.¹³ The constant appearance of new and

This problem has been at the crux of our inability to assess and manage risk in digital systems outside of financially based systems.

The analysis of events leading up to the tragedy of September 11, 2001, may make this a certainty. This will lead to digital forensic research and practice coming out of the *ex post facto* paradigm and into the world of intelligence gathering.

Tool development continues as the practitioner demand increases. Their caseloads continue to rise. They can't be expected to wait for the perfect solution. Therefore,

improved technology (e.g., cellular phones, personal digital assistants [PDAs], the Global Positioning System) has moved the target of media analysis tools way out of range for quick response. Code analysis is very important but suffers from various issues as well. Reverse engineering of software is difficult, specialized, and time consuming. Small pools of expertise are available to use an arcane tool set. In addition, these tools were not intended for *forensic*¹⁴ use. Finally, the legal ramifications of this technique are still being debated. Network analysis requires strong, specific expertise. In civilian circles analysis is dependent upon third-party data collection that is out of your control. Legal issues must be addressed quickly. Researchers and developers can help with time synchronization, which is vitally important.

To be effective, digital forensic research should consider some focus in the following areas: (1) advances in automated malicious detection and legal reverse engineering technology; (2) heightened focus on wireless technology, its vulnerabilities, and the forensic indicators that will assist operations personnel and investigators in identifying questionable activity; and (3) continue to work toward the establishment of approved standards and best practices to strengthen the foundation for Digital Forensic Science.

"Electronic Crime Technology Program: NIJ/OS&T," by Dr. John Hoyt

The National Institute of Justice (NIJ), Office of Science and Technology (OS&T), is chartered to give the state and local criminal justice community the means to address electronic crimes by bridging federal, state, and local government agencies, industry, and academic strengths toward useful solutions. They accomplish this task by applying their resources in six core program areas:

- Tools and technology development
- Technical assistance
- Training
- Standards and certification
- Policy and legal issues
- Outreach and education

Each area must address the realities of Internet expansion and anonymity that have become a home to a new breed of criminal. These wrongdoers are now more technically

applicable (proven) research will need to be phased in to avoid upsetting judicial process or economic stability. Programs like the National Institute of Standards and Technology (NIST)/NIJ Computer Forensic Tool Testing Program (www.cftt.nist.gov) are a must.

Meaning to persuade decision-makers in courts and civilian or military operations.

adept and determined as well as socially isolated, making them harder to catch outside the *cybersphere*. At the same time, our society is so much more dependent on information technologies seemingly full of vulnerabilities. That makes everyone targets. In some sense the crimes have remained constant; it's the exploitation tools and techniques employed that have moved into *cyberspace*.

NIJ/OS&T sees several trends for the consideration of researchers and practitioners alike. First, organized crime and cybercrime have begun to overlap. Traditional "crooks" have begun to see the benefit of more advanced means of exploitation. Second, nuisance tools have begun to be used for overt criminal purposes. Why should crooks design their own tools? Third, tremendous volumes of personal information are available online. This makes *cyberextortion* much more likely. Fourth, encryption is growing in use all over the Internet. In its current state, this growing use benefits the crooks as well as the public. Finally, the major problems our law enforcement system is having trying to overcome jurisdictional inequities have not gone unnoticed by the criminal element. These problems are being used against investigators.

At the crux of the problem NIJ/OS&T sees a need for research describing how to merge the intrusion analyst's process using traditional law enforcement techniques. This will allow for the use of a more general investigative tool set that respects the rigor of after-the-fact analysis but attempts to be more anticipatory or predictive. That is the new paradigm researchers and practitioners will face in the next few years. To help more research along, NIJ/OS&T is sponsoring work in several key programs in which many interested parties can be actively involved:

- Computer Forensic Tool Testing Program— (CFTT): Creating Standards for the Digital Forensic Sciences.
- National Software Reference Library– (NSRL): Accurate, available identity of known files and applications
- Quick Reference Guides for E-Crime Investigations
- E-Crime Initiatives: Working groups (law enforcement, private sector, and academia working together)

These programs, together with continued collaboration among other academic and investigative bodies, will help guide development of useful capabilities for solving crime over the digital landscape.

Workshop Discussions

It would appear that we have reached the limits of what is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in five years.

John von Neumann, 1949

After hearing from the keynote speakers, the remainder of the workshop was devoted to concentrated roundtable discussions. The topics were chosen to stimulate dialog in the areas related to foundations for our new scientific discipline and also to help enhance research already underway in certain key areas of a particularly challenging nature. Figure 4 lists the categorized topics. The emphasis of this first DFRWS involved concentrating on the fundamentals involved in the new organization as well as beginning to establish the common language upon which we could all agree and eventually apply.

In addition to expected dialog, the group discussions uncovered new, unanticipated challenges as well as potential approaches to problem solving. However, the groups knew that it is essential to consider the research already underway in important areas. Therefore, Hidden Data¹⁵ was chosen as the first DFRWS challenge topic. There is no doubt that research and development in this area holds tremendous importance for national security and specifically for Homeland Defense.

Foundations

- Framework for Digital Forensic Science
- Trustworthiness of Digital Evidence
- Definition of Network Forensics

Challenges

- Detection and Recovery of Hidden Data

Figure 4 - DFRWS Discussion Topics

¹⁵ Steganography, which is discussed later, is a subset of this problem set.

Attendees were divided into five groups. Each topic was presented for group discussion that lasted approximately one hour, during which time each group prepared charts to describe their findings. After group interaction was completed, each group presented its briefing to all in attendance and followed up with a question-and-answer period. This approach allowed for a variety of perspectives drawing from the wide spectrum of knowledge and experience levels in each group.

The workshop descriptions below begin with a boxed summary of the opening statement intended to guide each group discussion. The summary is followed by a definition of the topic area and a synopsis, which will present a unified statement of finding that represents all group presentations.¹⁶

Workshop 1 - A Framework for Digital Forensic Science

"Build a taxonomy to guide and direct research. Identify the areas or categories that define the "universe" of Digital Forensic Science"

To be considered a discipline, Digital Forensic Science must be characterized by the following associated entities:

- Theory: a body of statements and principles that attempts to explain how things work
- Abstractions and models: considerations beyond the obvious, factual, or observed
- Elements of practice: related technologies, tools, and methods
- Corpus of literature and professional practice
- Confidence and trust in results: usefulness, purpose

The current state of Digital Forensic Science exhibits only some of these characteristics, and they are not tied to specific disciplinary practices considered by any group as scientifically rigorous. There are elements of practice as defined by a myriad of tools developed with the best intentions, none of which are held to any specific scientific standard. There is, however, a level of trust and precedence established for some of these tools and techniques in common use. The fidelity of the trust placed in these tools and techniques is yet to be tested. None of these characteristics exist for any part of Digital Forensic Science applied outside of the courtroom (i.e., in civilian and military circles).

15

As stated earlier, all group presentations are available to authorized accounts on the DFRWS web site.

Until more formal research is performed, formal hypotheses leading to established theories remain unavailable. Theory is typically antecedent to abstraction and again, theories are largely absent. More formally established research efforts typically lead to scholarly publications for review by peers in academia and practice.¹⁷ Elements of practice and the level of trust in results follow naturally from rigorous research. In an effort to begin to formulate theories, the workshop groups started by trying to place bounds on the domain through definition.

The Definition

To begin establishing credibility for Digital Forensic Science as a scientific discipline, several groups proposed a definition for consideration. An effort was made to consider the topic in light of the multiple perspectives discussed earlier. Figure 5 represents a compilation from group suggestions.

Digital Forensic Science

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.

Figure 5 - A Definition for Digital Forensic Science

The new definition expands the scope outside the courtroom, allowing researchers to consider criteria from a much wider spectrum of practitioners. This distinction brings with it at least one cautionary note meant to place boundaries on the expanded scope:

Digital Forensic Science is not in the business of protection.

Efforts to incorporate digital forensic analysis into civilian and military operations will work in concert with computer and network security research and operations. Digital Forensic Science and analysis will depend (in some cases heavily) upon output from protective security components but must not be in competition with these methods.

16

¹⁷ There was much discussion about establishing a refereed journal for discussing topics on Digital Forensic Science.

The Process

The proposed definition must be applied to all activities in research and practice performed for this new discipline. According to the proposed definition, these activities are investigative in nature, and those practitioners who will employ these tools and methods will follow some form of investigative process in the performance of their duties. If properly categorized, the processes can enable practitioners to visualize where they need to add capability from what is available. Likewise, academic researchers will use the process to look for shortfalls in technology, helping them to focus on areas where research is needed the most. Several groups proposed this fact and offered possible approaches that may be applied. There is still room for debate regarding the breadth of the process and what portions should be considered "forensic." However, most agreed that a process must be established.

The items captured in Table 2 begin to establish the linear process, from Identification to Decision, that appears to be used in digital forensic analysis. The major categories or classes are noted at the top of the table. The contents of the columns below each category are candidate techniques or methods belonging to that class. As mentioned above, it is debatable whether or not all the major categories are to be considered "forensic." Those items in gray are subject to the least confusion, although there is still some discussion about the use of the term *collection and preservation*, and if one is really a subcategory of the other.

Table 2 - Investigative Process for Digital Forensic Science

Mentification	Presentation	Collection	Esamination	sischeup	Pesentation	Denision
Event/Crime Detection	Case Management	Preservation	Preservation	Preservation	Documentation	
Resolve Signture	Imaging Technologies	Approved Methods	Traceability	Traceability	Expert Testimony	
Profile Detection	Chain of Custody	Approved Software	Validation Techniques	Statistical	Clarification	
Anomalous Detection	Time Synch.	Approved Hardware	Filtering Techniques	Protocols	Mission Impact Statement	
Complaints		Legal Authority	Pattern Matching	Data Mining	Recommended Countermeasure	
System Monitoring		Lossless Compression	Hidden Data Discovery	Timeline	Statistical Interpretation	
Audit Analysis		Sampling	Hidden Data Extraction	Link		
Etc.		Data Reduction		Spacial		
		Recovery Techniques				

The debate will continue to move toward a more accurate definition of the process, and follow-on meetings of the DFRWS will act as a forum for the dialog. In this ongoing activity, it is important that these categories and definitions be accompanied by constant process refinement in light of the following topics:

- Developing standards
- Emerging technologies
- Practitioner needs
- Adversarial tactics
- Research discoveries

Of particular interest is the need for continued development of standards for techniques and methodologies employed within the discipline. The standardized approaches produced by groups like the Computer Forensic Tool Testing (CFTT) program will go a long way in accelerating product development and evaluation in the future. That will certainly allow more time for effective research as well as allow practitioners to spend more of their valuable time solving cases rather then testing new software products. Some standards may incorporate metrics that in some cases will help raise the confidence level for evidence presented in courts as well as civil and military forums.

Although it is clear that work remains to be done in fully defining and populating this process matrix, Table 2 represents a positive beginning. There are a few notes worth mentioning in regards to the present structure and how it might develop:

- Preservation must be a guarded principle across "forensic" categories.
- Traceability (cross referencing and linking) is key as evidence unfolds.
- Although presented as linear, the process must entail feedback for research to be totally effective.
- Real-time analysis must be considered as an essential research objective.
- A repository for digital forensic knowledge must be established.
- Effort must focus on developing applied collaborative technologies.
- To identify research areas, a task must be started to align existing tools and technologies with the current process definition. Shortfalls will then begin to appear.

Building Expertise

Long before the year 2000, the entire antiquated structure of college degrees, majors and credits will be in a shambles.

Alvin Toffler

This workshop question included a query as to what expertise would be necessary to address the needs of Digital Forensic Science now and into the future. With an effort just begun to define this discipline, the groups seemed to say the better question would be, "What expertise wouldn't be needed?" Nevertheless, the overall consensus was that a majority of the expertise required would come from the three areas described in Table 3.

Table 3 - Sources of Expertise in Digital Forensic Science

Source of Expertise	Specific Benefit
University undergraduate and graduate	
level curriculum	To fuel ongoing research and field work requiring extraordinary expertise
Focused certification programs	To add to the already credible actions of practitioners in the field
Scholarly, peer reviewed, publication	A widely available forum for discussion an debate on matters of science

Core competencies at the university level would come from fields like Computer Science, Engineering Sciences, Material Sciences, Physics, Mathematics, Criminal Justice, Psychology, Sociology, and certainly many of the existing Forensic Sciences. As a start, the following areas were also proposed as valid candidates for applicable specialization:

- Data Mining
- Languages/Linguistics
- Logic
- Statistics and Probability
- Signal Processing
- Image Analysis
- Encryption
- Evidence Preservation
- Network Engineering

Focused certification activities will build upon reputable programs that already exist. These are primarily geared toward serving law enforcement.¹⁸ However, as research and practice advance, new techniques, tools, and methods will find their way into these training areas. Jurisdictional insistence¹⁹ on using certified professionals will help the effort advance.

Programs, such as those conducted at SANS conferences, catering to more distributed forensic analysis are emerging.

¹⁹ Perhaps at the national level for some specialties.

A referred journal was a topic of some discussion during sidebars at the DFRWS. Many attendees expressed high interest in making this a reality. However, the amount of work required to initiate and maintain an effort such as this caused many some concern. Because this was only the first meeting of its kind, there was some concern about the ongoing "critical mass" (or number of submitted works) needed to sustain a journal focused on such a new, emerging discipline. As a workaround several academics with ties to other existing journals suggested that perhaps those journals could be used to test the interest in the field. This continues to be discussed among participants and will very likely be a major discussion point at the next DFRWS meeting.

Workshop 2 - The Trustworthiness of Digital Evidence

"Is the abstract, transformed nature of digital data troublesome in terms of integrity and fidelity when viewed as evidence? If so, can it be overcome?"

The Description

Many tools and methods exist that allow almost anyone to modify almost any attribute associated with digital data. This is reason to cast doubt on or at least occasionally suspect the integrity of digital evidence. However, coupled with the fact that what we view in electronic or hardcopy form is the result of many layers of transformation and translation, we seem to have an enormous problem with trusting anything we might consider evidence from digital sources. The following statements were implied in the group's discussion:

- Integrity answers the question:
 Has the structure of the data remained unchanged since it was first obtained?
- Fidelity addresses the question:
 How closely does the data accurately or truthfully represent fact or factual events?

A high level of data integrity is usually associated with all forensic analysis. In many ways, the purpose of assuring data integrity is to leave little room for technical argument regarding the fidelity uncovered in the data. The goal of high confidence and trust in evidence produced from analysis is of great concern, especially in courts of law but also in military and civilian operations. To a large degree, it is the confidence in the veracity of the evidence that allows decision-makers to act. Veracity relies upon fidelity and fidelity depends on integrity, as listed below:

Attributes affecting data integrity:

- Digital data is more easily forged than physical data.
- The form of digital data subjected to analysis is usually transformed in some way.
- The form of digital data subjected to analysis is always processed before scrutiny.
- Most analysis is performed using a digital copy or clone of the suspect information.
- Explanations of analytical methods can be misunderstood and cause confusion.
- Attributes affecting data fidelity:
 - Lack of standards
 - Correctness of translation and transformation mechanisms
 - Dependence on subjective reasoning

Issues

Group discussions led to some interesting conclusions. There was some consensus regarding the statement that, "All evidence analysis relies on interpretation." Therefore, fidelity of digital evidence in particular was no more troublesome than fidelity encountered by more traditional forensic sciences. What is missing in the digital realm is any real theoretical data about the details of transformations involved in moving from reality to a digitally processed representation. For example, what happens, exactly, to transform an arrangement of ferrous molecules on a disk to a document displayed by a word processor on a computer monitor? What is the mechanism used to record a scene captured by video camera in compressed video data format? Of course, someone knows the mechanisms in both instances, but can we comment on the "correctness" of the processes involved? Trained and certified forensic serologists can comment on the correctness of DNA evidence via explanations that incorporate findings from molecular biology, population genetics, and probability theory. Most analysis in Digital Forensic Science cannot make similar claims.

Issues related to digital integrity are less abstract and in most cases can be addressed by process creation or improvement. Discounting any unethical behavior by investigators and/or analysts, the application of technology can help greatly in this area. Some key procedural factors to consider that may help are as follows:

- Work from an exact copy of the original data.
- Preserve the data pedigree (hash, source, collection time).
- Employ cryptography as a partial solution (digital signatures and encryption).
- Understand metadata clues in key (or commonly used) applications.
- Rely, to some degree, on corroborating evidence as verification.

Research Solutions

What can academic research in Digital Forensic Science do to help guarantee integrity and reduce the analytical subjectivity in our discipline? Several avenues were discussed in various groups. However, a common theme was that technological applications alone would not fully solve the problems. Human interaction with digital evidence was determined to be a fact of life in Digital Forensic Science into the foreseeable future. Therefore, attention to criteria surrounding that interaction must be a focal point as well. These avenues are listed below:

Technological:

- Methods designed to detect digital tampering would be of vital importance. Watermarking and the judicious application of cryptography are high priorities.
- Securing or assuring protection of repositories from tampering is also tremendously valuable from a point of view of trust.
- Standards for correctness in digital transform methodology are needed to form a baseline for trust in many areas like data and video compression. Many more repeatable methods, models, and statistical analyses are needed to accredit these and other processes.
- Studies of hardware imperfection or electronic signature may produce data that links a piece of data to a source platform with higher confidence.
- Synchronizing and securing time as well as assessing measurable temporal drift
 per platform are very important, especially in the event of reconstruction in
 networked environments.

• Procedural:

- Accepted, standardized procedures are essential to maintain integrity and foster fidelity in digital evidence collection. Research discovery must underlie standards.
- Adequate training based upon those standards must follow, so full certification of laboratory and field personnel should be mandatory.
- Some focus should shift toward interpretive skills. Observing the syntax, semantics, and pragmatics of the data we analyze may yield superior results.

Workshop 3 - Detection and Recovery of Hidden Data

"Identify hiding methods and hiding places likely to be employed in digital realms. Discuss the related detection and extraction challenges we face."

The Definition

This workshop discusion addressed detection and recovery of hidden data. It has been known for some time that malicious attackers, such as terrorist cells, use computers and networks to help organize and plot activities.²⁰ [Also, since the September 11 events, open source information available from the ongoing investigation has begun to indicate that these groups may be using data hiding techniques²¹ to help ensure the secrecy upon which they depend.] Along with anonymous mailers and publicly available Internet access to help deter tracability, they may be using Steganography²² (from the Greek, meaning *covered writing*) to hide communications without drawing attention to themselves.

The DFRWS was held over a month prior to the henious acts of September 11, 2001. Steganography and other data-hiding techniques were already a high-profile item of interest to law enforcement and intelligence agencies alike. The present situation only heightens the resolve of the researchers and practitioners to find ways to effectively detect, extract, and trace the hidden information in all its forms. As it happened, this workshop was attended by representatives of universities and agents in the field who are activly studying this area so crucial to our nation's defense.

A cautionary note precedes these dicussions. Attendees with long-standing experience in computer and information security have noticed that this problem is closely related to the covert channel analysis challenge presented in the late 1980s.²³ Both represent unsolvable problems if the goal is total erradication. Data hiding cannot be prevented in all its forms, and detection is generally effective only for known algoritms. With that said, there is

²⁰ Countering the New Terrorism, Ian Lesser, RAND Press, 1999.

²¹ At varying levels of sophistication and stealth.

A technique whereby messages of many types may be imbedded within still images and video with little or no visible effect.

For many years researchers, operating in an environment of risk avoidance, sought a way to eliminate all possibility for covert channels in computers and communications.

sufficient evidence leading the conclusion that certain techniques are perferred over others and that focus in those areas is certainly justified in research and practice. What follows are descriptions of some of this ongoing work and a detailed look at the spectrum of challenges that must be addressed by research and practice as quickly as possible.

Hidden Data Placement

Components of the complex systems that compose our computers and networks offer many places for data to be hidden for purposes of evasion. That complexity makes the task of identifying and categorizing those places difficult if not impossible. However, the combined expertise at the DFRWS tried to do just that.

Table 4 lists some candidate locations in several general categories. The list is by no means complete, but a review of its contents sheds light on the larger problems confronting the research community. The current focus on Steganography has concentrated activities in the categories of Graphics and Signals. Although evidence seems to indicate that a large percentage of data-hiding techniques used by criminals are in this arena it still leaves a myriad of possibilities largely unchecked. Much more work is necessary in expansion and refinement of the categories and finally populating the resulting matrix with as much data as possible.

Disk Geomerty File Systems Solid State Graphics **Applications** Structures Structures Programming Non-Digital Signals Compression Compound Doc Marked Bad Reserved Packet Heap Space RAM Slack Email Spam CMOS RAM timing channels ile Slack Protocols System Libraries Plain sight lidden partitions Inallocated Spac magery

Table 4 - Categories of Data Hiding

Trends in Concealment

Possible or reasonably accessible locations must be associated with the popular or workable methods used to store and retrieve the data being hidden. The ways data is concealed are as varied as the places used:

• Steganography is the most advertised. Several open source tools are available that make it possible for anybody to hide messages of many types in graphics and images. All one needs is the identical tool on the sending and receiving ends of transmissions and data can be sent undetected. Trends that make this method even more worrisome are as follows:

- Encryption of the hidden message will make recovery and analysis more difficult even if extraction becomes possible.
- More sophisticated hiding techniques are being used other than LSB.²⁴
- Adaptive models are being used to determine a file's construct and modify the concealment approach.
- Exploitation of metadata structures within applications like word processors or spreadsheets is prevalent.
- Data decomposition can be employed. This entails chopping data into parts small enough to bypass filters or other monitors designed to detect known signatures.
- Phase differentials in signals passed on audio and video tracks may be used to transmit context.
- Database indexing can be used to transmit meaning, albeit at very low data rates.

Research in Detection and Recovery

The first phase of work in this interesting and essential area has been on detecting hidden data (primarily Steganalysis²⁵) in images and graphic data streams. The goals expressed most often by researchers are to enhance and develop techniques in the following categories:

- **Blind detection**: This is a method that doesn't require an original (known to be free from any compromise) graphic for comparison. Comparing copies of suspect digital imagery or graphic data representations with known uncompromised versions of the same is a common detection method. To be fully successful this would require a repository that contained all the images in existence. Instead, researchers are looking for ways to analyze the structural elements of the data (the signal characteristics) to look for anomalies.
- Watermarking: Some fine paper products contain barely noticeable indicators of authenticity, called watermarks. The same concept is being studied and researched to decide if it can be reliably applied to the digital data we create and transmit daily. The digital watermark would have a highly detailed and defined structure that could be analyzed to assess if any compromise has occurred for that digital product.
- Image Quality Standards: Detailed metrics regarding resolution, size, aspect ratio, color pallet, and other graphic attributes could be applied to images of differing types. The standard metrics could then be analyzed and compared to those narrow tolerances

Least significant bit - a commonly used technique that is well understood and relatively easy to detect.

²⁵ The detection of data that was hidden using steganographic techniques.

known or expected for that graphics type. Any flaw or potential compromise would be detected. This has obvious limitations for purely textual data.

- **Hashing and encryption**: These techniques are also being studied as stopgap measures. They do have performance impacts associated with them, as do other techniques.
- **Signature Analysis**: Open source Steganography utilities continue to be authored and distributed over the Internet. Indications are that a majority of hidden data being transmitted via these methods use tools that anyone can get and use. Pranksters, professional criminals, and terrorists will tend to use tools they can find rather than author their own. Therefore, we should understand the available tools. If we can identify their existence on a suspect computer, that may reduce the types of signature we need to find.

Further enhancements in this field will depend upon the availability of tools and technologies from the following list:

- Full network packet capture capability (performance and storage are issues)
- File content and structure analysis that looks for chaos in order or order in chaos
- Reliable (and perhaps legal) reverse engineering
- Cryptanalysis
- Signal processing and analysis
- Statistical analysis
- Data pedigree (and origin) tracking and verification
- Awareness training incorporating all the latest methods of detection

Recovery or extraction of hidden data was of interest to all in attendance. Although there was some sideline discussion on this issue, most agreed that it is very difficult to achieve given our present state of understanding in the field. However, the consensus was just as strong, among those involved in research in this discipline, that it is by no means impossible. The DFRWS will continue to focus on this important issue in upcoming meetings and will move to find appropriate funding to expedite needed research in the areas described above.

Workshop 4 - Digital Forensic Science in Networked Environments (Network Forensics)

"Define Network Forensics and identify relevant research areas required to address current shortfalls. Consider the effect of emerging technologies in your discussions."

The term *network forensics* has been used with increasing regularity for some time. Although no official definition exists, the term is commonly used to describe the task of analyzing information collected on active networks from various intrusion detection, auditing, and monitoring capabilities for the purpose of protection. From various discussions it appears to have prosecution as a high-level objective. It is sometimes described as an after-the-fact technique but also seems to desire an as near to real-time an environment as possible. These and other charactreistics seem to place network forensics somewhere in the path between the law enforcement model and the analytical requirements of civilian and military operations.

The monitoring and analysis of data from live systems and networks will become essential to law enforcement as caseloads increase and juristictional boundaries blur. It is already of high interest in civilian and military quarters where the need for quick and reasonably accurate analysis overshadows some legal constraints in other paradigms. This workshop discussion set out to discover just what network forensics meant to attendees: to devise a framework for this discipline in much the same manner as in Workshop 1.

The Definition

Network Forensics

The use of scientifically proven techniques to collect, fuse, identify, examine, correlate, analyze, and document digital evidence from multiple, actively processing and transmitting digital sources for the purpose of uncovering facts related to the planned intent, or measured success of unauthorized activities meant to disrupt, corrupt, and or compromise system components as well as providing information to assist in response to or recovery from these activities.

Figure 6 - A Definition for Network Forensics

Although not specifically mentioned in the definition shown in Figure 6, enterprise management solutions like intrusion detection systems will play a key role as input data sources for network forensic analysis into the foreseeable future. This is especially true because they are very commonly used methods of capturing data from a wide variety of digital sources and storing that data in a centralized repository that could be made accessible to analytical processes with forensic objectives. However, the distinction was made in workshop group discussions between intrusion detection as a monitoring or auditing capability and the separate and perhaps as yet unrealized analysis necessary to fulfill the spirit of the definition above.

Major Issues

Aside from proposing a definition, many group discussions focused on their candidates for major issues in this very new discipline, as discussed below.

Time

Synchronization and integrity of the date and time information associated with events being analyzed are important for all forensic activity. It is especially important when your environment may span jurisdictions and time zones. Work in this area is exceedingly important.

Performance

As mentioned at the beginning of this document, as research begins to accommodate analysis needed in the civilian and military sectors, speed will be a crucial factor in determining effectiveness and overall success. Of course, performance and speed are almost synonymous terms in computing. The major item affecting overall performance is data volume: the amount of data collected for analysis of this type is very often quite large. As stated before, intrusion detection systems will be depended upon to provide input to analyses. Historically, they have been very good collectors of bulk data but less than efficient in filtering and intelligent data reduction that will be necessary in network forensic analysis. Therefore, research activities must focus in part on increasing the efficiency of those techniques designed to scrutinize reams of data and cull out whatever falls outside of investigative requirements.

Complexity

The move from careful analysis of already complex standalone computers to scrutiny of data in networked environments significantly raises the level of detail, especially for the following investigative aids:

• **Tools** designed for use in singular environments cannot easily transition to networks hosting multiple operating systems, protocols, applications, and data formats. Research is needed in providing tools and techniques that accommodate this mixture.

• Correlation of the volumes of data into usable, understandable chunks or sections is necessary to afford useful analysis. Work in areas that help analysts and examiners understand data relationships in the volumes of information stored in intrusion detection databases is crucial. Data mining will be a key area of research.

Collection

Performing network forensic analysis will require scrutiny of large amounts of highly filtered data from varying sources, many times over large geographic spaces. This complexity causes concerns in related areas:

- Who will collect the data? Traditional forensic approaches dictate a certain level of trust in the collector of evidence, especially for use in the courts. The transition to, or at least acknowledgment of the existence of, other forensic paradigms may reduce the requirement for having sworn officers do the capture. However, credible certification is still at issue and the parameters of that define that credibility needs the benefit of research.
- When or how often should data be collected? The issue of "freshness," especially of snapshots of network or system state, will become increasingly more important as this discipline matures. Also, the security and integrity of this remotely collected data will be called into question, so researchers need to address these attributes as well.
- What is to be collected? This is also a major issue implied in statements above. Related to this issue are the non-standard format offerings by the many different intrusion detection systems available for use to feed the network forensic analysis.

Paradigm Distinctions

Even within the new discipline itself, attendees found some distinctive issues that need additional research. Law enforcement has a reluctance to enter into this realm largely because it is out of the typical control that their investigations require:

• Research to aid law enforcement in starting to trust this approach will center on data integrity, chain of custody, and proven methods.

Analysis of this type seems closer to fulfilling the needs of civilian and military operations largely because it doesn't insist on the existence of an *ex post facto* state. Therefore, the following guidance resulted:

- Researchers should focus on how new analytical methods will satisfy the various internal objects of operational networks, such as:
 - Intelligence Needs: identify sources, methods, and intent
 - **Network Operations**: examine traffic patterns to assess their effect on performance and availability

• Law Enforcement: identify and trace or locate suspects

Collaboration

Due to the wider array of expertise required for collaboration analysis, it becomes even more of a concern that data regarding tools and techniques to aid analysis will be shared among all responsible parties in a timely and accurate fashion. Research is sorely needed to spread knowledge in the wide range of skills needed to analyze networks in near real time.

Legal Hurdles

What are the Rules of Engagement necessary for analysts to operate effectively and in some areas legally? How do we operate legally over multiple jurisdictions? How do we identify where the crime or unauthorized activity really originated? What are the legal ramification of personnel or user profiling, and can these techniques be used effectively? What are the collective legal views regarding "expectations of privacy"? How does new legislation impact methods of research used now as well as those yet to emerge?

Legal issues were of special interest to researchers attending the DFRWS. There is special concern over the *Digital Millennium Copyright Act* (effective since 1998) and the new *Security Systems Standards and Certification Act* (SSSCA), known as the *Hollings Bill*. Strong feeling was voiced by many serious researchers that, alone, each act has the ability to stifle research in the digital forensic arena for fear of lawsuits and possible imprisonment they exact. Together, they would almost assure that research would be reduced by a measurable amount. This does not bode well for the future of digital forensic research. This will be a major topic of discussion now and in the very near future.

Emerging Technologies

The groups all proposed candidate technologies that promise to create additional challenges for the field of network forensics. Of course, a challenge also represents an opportunity for research as well. Examples follow:

- Wireless technology will add further layers of analysis by incorporating more
 protocols and services that must be understood and factored into the mix. This
 technology tends to blur the distinction between computers and telephony, which may
 cause additional legal problems. New and additional research tasking will be needed
 to help, especially regarding the accurate location of wireless devices.
- The merging or absorption of wired services into wireless architectures will add complexity. PDAs, smart appliances, 3G and 4G cellular service networks, and peer-to-peer applications must all be studied to assess how they impact forensic analysis in general and network forensics in particular.
- Holographic signal processing and cross-channel bleeding and linkage were also mentioned though never fully discussed.

Summary and Conclusions

Science, ever since the time of the Arabs, has had two functions: (1) to enable us to know things, and (2) to enable us to do things. The Greeks, with the exception of Archimedes, were only interested in the first of these...The Arabs wished to discover the philosopher's stone.

Bertrand Russell, *The Impact of Science on Society*, 1953

In a day and a half of focused discussion, those attending the first Digital Forensics Research Workshop made significant progress toward the establishment of a community and definition of a discipline based on solid, scientific foundations. The hope is that these and other committed professionals from academia, government, industry, and civilian organizations will continue to meet, debate, discuss, and document items that will enhance this field we all believe to be so important to our nation and the world.

In the near future, follow-on workshops and other organizational gatherings will be announced. These meetings will take up where the five groups left off in August of 2001, as follows:

- Continue to define terms and technologies associated with Digital Forensic Science to make communications effective and research more applicable.
- Refine the investigative process so it can be better applied to identifying needed research by closely mirroring steps used in field operations.
- Look to accommodate operational as well as law enforcement perspectives by including more representatives from the e-commerce and e-business communities in our discussions.
- Find the new technical challenges, generate scholarly debate, conceptualize approaches, help identify and align the needed expertise to conduct research, assist in the publication of findings, and help to prototype and, if needed, guide implementation.

The first DFRWS is the initial step in a coordinated effort to bring together the best minds to discuss challenging problems associated with the suitability of digital evidence in all its forms. The digital sources we must analyze to obtain suitable evidence in all its forms have three items in common. First, they are increasingly complex and less understood overall. Second, they are constantly morphing in form and function. Third, at the root of all their increasing functionality and detail are fundamental technologies that can be explained scientifically. It is these foundation technologies we must identify, fully understand, and continuously monitor to support research in digital forensic analysis from all its perspectives.

Appendix A – A Digital Forensic Road Map

Building a Framework for Digital Forensic Science

Detailed Workshop Results

Objective

DFRWS membership agrees that to help this research area become effective and useful a lexicon containing clear definitions and terminology is essential. Academic researcher and practitioner alike should use the DFRWS as a vehicle to achieve that goal through continued group dialog to build consensus.

Research Areas

- **Definition**: Finalize the definition of Digital Forensic Science to give bounds to the research performed within this core discipline. (See Figure 5)
- **Process**: Define a generic investigative process that can be applied to all (or the majority of) investigations involving digital systems and networks. The Digital Investigative Process (DIP) must be defined from highest level categories to the individual steps necessary for complete analysis of all potential digital evidence. (See Table 2.)
- **Expertise**: Build a clear picture of the academic and vocational expertise necessary to perform analyses given the DIP. Start to form academic programs as well as certification efforts to provide the level of proficiency that will be required.

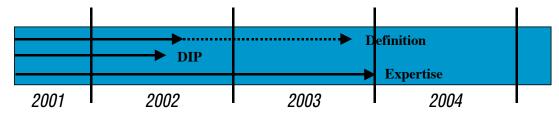
Payoff

Focus on the three research areas above will provide the following payoff:

- Clarity: Bring all interested parties together with little conflict of terminology, allowing more energy to be applied to solutions.
- **Focus**: Use the DIP to show all steps needed for full analysis, what is currently covered manually and what is automated, and what remains unaddressed that should be a focus of basic research.
- **Improvement**: Open the door to process improvement, creativity, and the ability to respond quickly to new threats and challenges. Also, provide measurable proficiency, which raises the level confidence in all results.

Timeline

With the necessary support and focus, the Digital Forensic Science Framework definition should progress at a rapid pace, especially through the use of collaborative measures like DFRWS. As the figure below shows, a solid beginning should be realized with the next workshop. This is a dynamic process and change should be expected well into the next two years. What would suit the cause best is if a complete picture of the DIP could be finalized just before the next DFRWS so it can be fully debated and hopefully finalized just after that event. As for the building of expertise, that will be an ongoing task that will mature over the next two to three years. Some colleges are already offering focused individual course work as well as some certificate programs in the field. Full curriculums will follow if interest and national need remains high.



Issues of Trust in Digital Evidence

Detailed Workshop Results

Objective

We must identify and address those items or characteristics of digital technology that could subtract from the trust needed to use analytical findings as evidence in courts and other venues. Digital technology has provided the ability to mimic reality with the push of a button. However, that same tremendous benefit raises real doubts if we are asking decision-makers to view digital evidence as reality. We must be able to fully understand and explain what happens from raw bits to visual display and we must know how exactly that varies from what is perceived in reality. In addition, we must be able to assure those who ask that the integrity of the information remains unquestioned.

Research Areas

- **Anti-tampering**: Develop methods to thwart or record tampering of files, disks, systems, and networks
- **Correctness**: Develop standards for correctness in digital transform technology. Develop a baseline of understanding in this area.

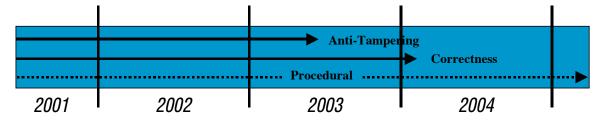
• **Procedural**: Translate standards into proper laboratory protocols. New findings must find their way into digital forensic curriculum.

Payoff

In criminal and civil proceedings, both the defense and the prosecution are advancing in their knowledge of digital technology. The less we understand the technology we are analyzing, the easier it will be to refute any testimony we offer as evidence. For example, is an MPEG or AVI video film clip a true representation of what actually happened? Does it adequately represent the reality it shows? Has it been modified from its original form? Research in this area will begin to answer these questions. All and more will be needed to give decision-makers in the courts, military, and civilian sectors the confidence they will need to do their jobs in the very near future.

Timeline

Work on anti-tampering techniques is already well underway, but there is reason to believe that tasks will continue for another year or more before products begin to emerge. Little work has been undertaken, to date, to define the complex processes behind the transformation from magnetic force to display on a monitor and everything in between. The task is quite daunting and some thought must be given to organizing the problem into units that are solvable and prioritized as to importance and usefulness. As seen in the figure below, the task of crafting procedures from research is essentially ongoing but no less important if we expect the research to be practical.



Detection and Recovery of Hidden Data

Detailed Workshop Results

Objective

All segments of society use computers and that will continue to be true. The less savory segments will no doubt find ways to communicate with their compatriots, ways that aren't monitored in any mainstream sense. The task of research is to try to identify major places as well as the nooks and crannies where date can be hidden in these complex systems and networks. Also, a goal is to craft discovery mechanisms that will detect and extract the data, in all its forms, being stored and transmitted in this fashion.

Research Areas

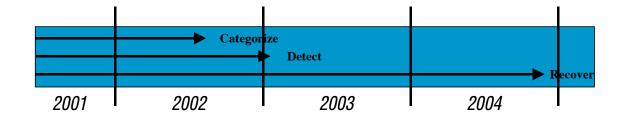
- Categorization: Continue DFRWS efforts to document the large number of places and mechanisms that have potential for use in data hiding. Even though it is a large, almost infinitely variable problem set, this work must continue especially outside the popular area of Steganography and the graphic data realm. (See Table 4)
- **Detection**: Continue to devise mechanisms whereby detection is possible with original material for comparison (i.e., blind detection). High-confidence evidence of the presence of hidden data is the first step. This evidence will help stop the message transmission.
- **Recovery**: Examine technology to extract or recover data that has been hidden. This will allow consumers of these technologies to understand and get answers to the following questions: Who?, What?, Why?, Where?, and When? This is vital research that has many academics interested and ready to start.

Payoff

Detection and recovery of hidden data is another area in which a full understanding of the complexities of digital systems is required. There is a great opportunity to those working on issues of trust to collaborate with researchers in this field. Due to the popularity and availability of Steganographic tools and techniques, this category of data hiding seems to have the largest payoff. [Since our adversaries may be using it to harm our liberties in the current conflict, we were correct in focusing our efforts here.] That coupled with the good and valuable research already underway says that major concentration in this area is a very valuable endeavor. At the same time, as we get closer to solutions that will close off this hiding technique as a viable tool, those same adversaries will adopt other means. We must try to get as far ahead of them as possible.

Timeline

The categorization of hiding places and techniques began with the DFRWS. This activity will continue between now and the next workshop given sufficient collective effort. Major work in detection, for both blind and comparative methods, is underway now. However, researchers implied that much work remains before the techniques can be applied with the necessary confidence required to be effective. That work will continue for the remainder of 2002 and perhaps into years to follow. Funding is a key issue for effective research to flourish. Recovery is quite another matter altogether. Although many key researchers felt it was a distinct possibility, little work has been performed to test current hypothesis in the area. This area is of particular interest and vital importance to national security and law enforcement. Indications are that research will continue on recovery and extraction of hidden data well into 2004 and beyond.



Digital Forensic Science in Networked Environments (Network Forensics)

Detailed Workshop Results

Objective

The objective was to determine if Digital Forensic Science can and should be applied to live networks. If that determination is affirmative (and all indications show that it is), then it must include the beginnings of the identification and definition of yet another branch of Digital Forensic Science. This new branch, network forensics, will borrow tools and techniques from its stand-alone, media-intensive relative, and it will move to produce the same level of accuracy and scientific rigor in a much shorter timeline. In fact, it will strive to exist in the real-time environment of military and civilian operations. It will also present challenges heretofore not addressed, and it will sometimes be at odds with the long-established paradigm applied to forensic science in support of law enforcement and the courts. A major challenge is also presented by current and existing legislation that may jeopardize research in this area and in digital forensic research in general (see Legal Hurdles)

Research Areas

- **Definition**: Answer the following questions: What are the similarities and differences between "normal" Digital Forensic Science and its new network counterpart? How broad is the scope of the problem set and to which environments does it apply? These questions must be answered first for real work to proceed. (See Figure 6.)
- **Performance**: Apply digital forensic analysis to networks in real time. Decision-makers desire information and answers about anomalous or malicious activity before they cause damage or disrupt continuity. Taking entire networks back to a laboratory for analysis is out of the question. So as system and network complexity increases and data volumes rise steadily, this issue becomes an even greater challenge. Research that addresses how to transition media analysis methods, as well as how to craft new techniques to collect and guarantee the integrity of network data, is paramount.

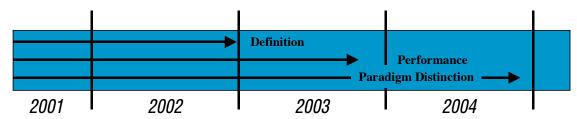
• Paradigm Distinctions: Conduct innovative research to break with the law enforcement convention. Sworn officers cannot be the only trusted agents allowed to collect forensic data for analysis. Trusted collection processes are needed. The field must consider new "suitability" criteria for new analytical objectives. Intelligence, network operations, command and control, and law enforcement must all be considered in the world of Digital Forensic Science applied to networks.

Payoff

A standardized and accepted definition is necessary here as it is in any real scientific activity. It will allow all researchers to understand each other and communicate new ideas in light of shared foundations and principles. Techniques that strongly consider speed and performance as well as accuracy will allow researchers and practitioners to stay up with increasingly complex systems and possibly one step ahead of potential adversaries. The realization that boundless communications and mobile computing are here to stay will allow researchers to help all investigators make the transition from the brick and mortar digital crime lab to a "virtual" model. Research in true Digital Forensic Science applied to core computer technologies will serve all paradigms, including those yet to come.

Timeline

The DFRWS membership has already begun the work of starting to define network forensics. The intent is to continue dialog in future workshops to enhance and finalize this work. As shown in the figure below, there is reason to believe it could be completed some time in late 2002. Research related to analytical performance in network forensics is yet to be started. The work is absolutely necessary and a good amount of the tasking can borrow from work related to intrusion detection and data mining. Because there is little current work and no work seen to be starting soon, definitive research is likely to begin late in 2002 and proceed well into 2003 and beyond. Paradigm distinctions will follow work that shows conclusive scientific results. Research is necessary to define what network data should be collected and at what intervals. The methods must be proven to be secure and also guarantee data integrity. Issues of synchronized time will be paramount and legal hurdles may be significant (especially regarding privacy). The prediction is that accepted paradigm shifts due to research in network forensics may go beyond 2004.



Appendix B - Attendees and Contact Information

Organization	Name	Email Address
Dartmouth University	Dennis McGrath, Ph.D	dmcgrath@ists.dartmouth.edu
Dartmouth University	Vincent Berk, Ph.D.	vberk@ists.dartmouth.edu
Syracuse University	Shu-kai Chin, Ph.D.	chin@cat.syr.edu
Syracuse University	Kamal Jabbour, Ph.D.	jabbour@cat.syr.edu
Syracuse University	Steven Chapin, Ph.D.	chapin@cat.syr.edu
Cornell University	Johannes Gehrke, Ph.D.	johannes@cs.cornell.edu
University of Central Florida	Erol Gelenbe, Ph.D.	erol@cs.ucf.edu
University of Central Florida	John Leeson, Ph.D.	leeson@cs.ucf.edu
Yale University	Eoghan Casey	eoghan.casey@yale.edu
Louisiana State University	Peter Chen, Ph.D.	chen@bit.csc.lsu.edu
Air Force Institute of Technology WPAFB	Gregg Gunsch, Ph.D.	Gregg.gunsch@afit.edu
SUNY Binghamton	Jessica Fridrich, Ph.D.	fridrich@binghamton.edu
University of Texas	Larry LeiBrock, Ph.D.	larry.leibrock@bus.utexas.edu
Villanova University	Bijan Mobasseri, Ph.D.	mobasser@ece.villanova.edu
SUNY Institute of Technology	Heather Dussault, Ph.D.	dussauh@sunyit.edu
SUNY Institute of Technology	Sam Sengupta, Ph.D.	sengupta@sunyit.edu
Utica College of Syracuse University	George Curtis, JD	gcurtis@utica.ucsu.edu
Purdue University	Eugene Spafford, Ph.D.	spaf@cerias.purdue.edu
NY Polytechnic University	Nasir Memon, Ph.D.	memon@duke.poly.edu
NY Polytechnic University	Kulesh Shanmugasundaram, Ph.D.	kulesh@vip.poly.edu
North Carolina State University	Doug Reeves, Ph.D.	reeves@csc.ncsu.edu
George Mason University	Sushil Jajodia, Ph.D.	jajodia@gmu.edu
George Mason University	Neil Johnson, Ph.D.	njohnson@gmu.edu
University of Washington	Dave Dittrich	dittrich@cac.washington.edu

National Institute of Standards and Technology	James Lyle, Ph.D.	jlyle@nist.gov
National Institute of Standards and Technology	Gary Fisher	gary.fisher@nist.gov
DOD Computer Forensic Laboratory	Mark Luque	mark.luque@dcfl.gov
DOD Computer Forensic Laboratory	Robert Griesacker	rob.griesacker@dcfl.gov
National Institute for Justice	John Hoyt, Ph.D.	hoytj@ojp.usdoj.gov
USTRANSCOM - MITRE Corporation	Charles Boeckman	boeckman@mitre.org
NIPC - MITRE Corporation	David Baker	bakerd@mitre.org
G030 - MITRE Corporation	William Dowling	wad@mitre.org
National Center for Forensic Science	Carrie Whitcomb, Director NCFS	whitcomb@mail.ucf.edu
AFRL/IFEC	Richard Simard	richard.simard@rl.af.mil
AFRL/IFEC	Scott Adams	scott.adams@rl.af.mil
AFRL/IFGB	Len Popyack, Ph.D.	leonard.popyack@rl.af.mil
AFRL/IFGB	Joe Giordano	joseph.giordano@rl.af.mil
AFRL/IFGB	John Feldman	john.feldman@rl.af.mil
AFRL/IFGB	John Faust	john.faust@rl.af.mil
AFRL/IFGB	Jim Sidoran	james.sidoran@rl.af.mil
AFRL/IFGB	Andrew Karam	andrew.karam@rl.af.mil
Logicon TASC - Cyberforensics Science & Technology Center	Dan Kalil	daniel.kalil@rl.af.mil
CACI - Cyberforensics Science & Technology Center	Frank Cole	Frank.cole@rl.af.mil
Logicon TASC - Cyberforensics Science & Technology Center	John Del Medico	john.delmedico@rl.af.mil
MITRE Corporation - Cyberforensics Science & Technology Center	Gary Palmer	gary.palmer@rl.af.mil

Emergent Information Technologies - Cyberforensics Science & Technology Center	Derek Bronner	derek.bronner@rl.af.mil
Emergent Information Technologies	Jim Riccardi	jim.riccardi@emergent-it.com
Emergent Information Technologies	Robert McOrmond	robert.mcormond@emergent- it.com
Logicon PRC	Charles Green	charles.green@rl.af.mil
Wetstone Technologies	Chet Hosmer	chet@wetstonetech.com
Wetstone Technologies	Chris Hyde	chris@wetstonetech.com
Wetstone Technologies	Mark Reilly	N/A
Chen & Assocoates, Inc.	Li Jo Ph.D.	Lihobr@aol.com
Odyssey Research	Frank Adelstein, Ph.D.	fadelstein@oracorp.com
Cyber Forensics Research & Development Center @ Utica College	Christine Siedsma	csiedsma@utica.ucsu.edu
Cyber Forensics Research & Development Center @ Utica College	Matt Ward	mward@utica.ucsu.edu