Sponsors

LONGWOOD UNIVERSITY

Metropolitan State University

UNIVERSITY of ST. THOMAS
Contents

Committee ........................................................................................................................................... 4
Schedule ............................................................................................................................................... 5
HiGate (High Grade Anti-Tamper Equipment) Prototype and Application to e-Discovery .............................................................. 7
Yui Sakurai, Yuki Ashino, Tetsutaro Uehara, Hiroshi Yoshiura and Ryoichi Sasaki
Developing VoIP Honeypots: a Preliminary Investigation into Malfeasant Activity .... 17
Craig Valli
A Framework to Integrate the Data of Interview Investigation and Digital Evidence ...... 25
Fahad Alshathry
Organizational Handling of Digital Evidence ............................................................................ 33
Sheona Anne Hoolachan and Brad Glisson
Social Networking: A Boon to Criminals ................................................................................. 45
Tejashree D. Datar and Richard Mislan
Hard Disk Storage: Firmware Manipulation and Forensic Impact and Current Best Practice ........................................................................... 55
Gareth Davies and Iain Sutherland
Measuring Whitespace Patterns as an Indication of Plagiarism ............................................. 63
Ilana Shay, Nikolaus Baer and Robert Zeidman
Electronic Discovery: A Fool's Errand Where Angels Fear to Tread? .................................. 75
Milton Luoma and Vicki Luoma
A Layered Framework Approach to Mitigate Crimeware ....................................................... 87
Mathew Nyamagwa
Canvass: A Software Tool for Forensic Steganalysis of JPEG Images ................................ 99
Jennifer L. Davidson and Jaikishan Jalan
Digital Records Forensics: A New Science and Academic Program for Forensic Readiness ............................................................................. 109
Luciana Duranti and Barbara Endicott-Popovsky
Developing a Baccalaureate Digital Forensics Major ............................................................... 123
John H. Riley, Jr.
The Defiance College Undergraduate Major in Digital Forensic Science: Setting the Bar Higher ............................................................................. 131
Gregg H. Gunsch
Computer Forensics for Graduate Accountants: A Motivational Curriculum Design Approach ................................................................. 141
Grover Kearns
Conference Committee

The 2009 ADFSL Conference on Digital Forensics, Security and Law is pleased to have the following members of the conference committee.

Jigang Liu
Jigang.Liu@metrostate.edu
General chair
Metropolitan State University
Minnesota, USA

Glenn Dardick
gdardick@dardick.net
Finance chair, Publication chair & Webmaster
Longwood University
Virginia, USA

Brad Rubin
bsrubin@stthomas.edu
Program chair
University of St. Thomas
Minnesota, USA

Gary Kessler
gck@garykessler.net
Honorary Chair
Gary Kessler Associates
Vermont, USA

Larry Gottschalk
Larry.Gottschalk@metrostate.edu
Local arrangements chair
Metropolitan State University
Minnesota, USA

Bob Boyczuk
Robert.Boyczuk@senecac.on.ca
International Advisor Canada
Seneca College
Toronto, Canada

Denis Edgar-Nevill
denis.edgar-nevill@canterbury.ac.uk
International Advisor Europe
Canterbury Christ Church University
UK

Mohamed Chawki
chawki@cybercrime-fr.org
International Advisor Africa and Europe
University of Aix-Marseille III
France

Sizwe Lindelo Snail
ssnail@snailattorneys.co.za
International Advisor Africa
Snail Attorneys, Attorneys at Law
Legal Consultant / Admitted Attorney of the South African High Court
South Africa

Andy Jones
andrew.jones@kustar.ac.ae
International Advisor Europe and Middle East
Khalifa University, UAE

Tetsutaro Uehara
uehara@media.kyoto-u.ac.jp
International Advisor Asia
Kyoto University
Kyoto

Craig Valli
c.valli@ecu.edu.au
International Advisor Asia-Pacific
Edith Cowan University
Western Australia, Australia

Doug White
doug.white@acm.org
CCE Liaison
Roger Williams University
Rhode Island, USA

Israel Aladejei
i.aladejebi@century.edu
Program Committee Member
Century College
Minnesota, USA

Ezendu Ariwa
e.ariwa@londonmet.ac.uk
Program Committee Member
London Metropolitan University, UK

Gareth Davies
gddavies@glam.ac.uk
Program Committee Member
University of Glamorgan, UK

Kevin Harris
Kevin.L.Harris@nscc.edu
Program Committee Member
Nashville State Community College
Tennessee, USA

Milt Luoma
Milt.Luoma@metrostate.edu
Program Committee Member
Metropolitan State University
Minnesota, USA

Vicki Luoma
vicki.luoma@mnsu.edu
Program Committee Member
Minnesota State University Mankato
Minnesota, USA

Ezendu Ariwa
e.ariwa@londonmet.ac.uk
Program Committee Member
London Metropolitan University, UK

John Bagby
jbagby@ist.psu.edu
Program Committee Member
The Pennsylvania State University
Pennsylvania, USA

Irena Kageorgis
ikageorgis@apus.edu
Program Committee Member
American Public University
West Virginia, USA

Judie Mulholland
judie@acret.org
Program Committee Member
Florida State University
Florida, USA

Diane Barrett
Dbarrett@uaf.edu
Program Committee Member
University of Advanced Technology
Arizona, USA

Igor Kholodov
Igor.Kholodov@bristolcc.edu
Program Committee Member
Bristol Community College
Massachusetts, USA

Marcus Rogers
rogersmk@purdue.edu
Program Committee Member
Purdue University
Indiana, USA

Donny Cheung
cheungdo@metrostate.edu
Program Committee Member
Bureau of Crime Apprehension
Minnesota, USA

Stephen Larson
larsnsp@vcu.edu
Program Committee Member
Virginia Commonwealth University
Virginia, USA

Il-Yeol Song
songiy@drexel.edu
Program Committee Member
Drexel University
Pennsylvania, USA
Schedule

Wednesday, May 19

- 08:00 AM  COFFEE
- 08:30 AM  On-site Registration
- 09:00 AM  Welcome and Introductions
  - Glenn S. Dardick
  - Jigang Liu
  - Larry Gottschalk
  - Brad Rubin
- 09:15 AM  Papers/Presentation session
  - Yui Sakurai: HiGate (High Grade Anti-Tamper Equipment) Prototype and Application to e-Discovery
  - Craig Valli: Developing VoIP Honeypots: a Preliminary Investigation into Malfeasant Activity
- 10:15 AM  BREAK
- 10:30 AM  Papers/Presentation session
  - Fahad Alshathry: A Framework to Integrate the Data of Interview Investigation and Digital Evidence
  - Barry Caplin: Information Security and Security Technologies
  - Brad Glisson: Organizational Handling of Digital Evidence
- 12:00 Noon  LUNCH (provided)
  - Speaker: Chris Andrews, Computer Forensics Specialist, Kroll Ontrack, Minnesota
  - Topic: "Computer Forensics in the 21st Century"
- 01:00 PM  Papers/Presentation session
  - Tejashree D. Datar: Social Networking: A Boon to Criminals
  - Gareth Davies: Hard Disk Storage: Firmware Manipulation and Forensic Impact and Current Best Practice
- 02:30 PM  BREAK
- 02:45 PM  Workshop session
  - Diane Barrett: How to Examine Virtual Environments
- 04:00 PM  Panel session: Internal Threats: Reality and Challenges
  - Moderator: Larry Gottschalk, Metropolitan State University
  - Eric Lucero, Senior Technical Consultant-Security, Midwave Corporation
  - Chris Perkins, Digital Forensic Investigator, Medtronic, Inc
  - Darlene M. Tester, Data Privacy Manager/Attorney, Bluestem Brands Inc.
  - Jeremy D. Wunsch, CEO and founder, LuciData LLC.

Thursday, May 20

- 08:00 AM  CONTINENTAL BREAKFAST
- 08:30 AM  On-site Registration
- 09:00 AM  Welcoming Remarks
  - Sue K. Hammersmith, President, Metropolitan State University
- 09:10 AM  Keynote Speech
  - Marc Rogers, Professor, Purdue University
- 09:45 AM  Papers/Presentation session
  - Wayne Jansen: Assessing Mobile Forensic Tools with SIMfill
  - Nathan J Keith: Challenges and Strategies with Windows 7 Limewire: A Practitioner’s Perspective
- 10:45 AM  BREAK
Schedule

Thursday, May 20 (continued)

• 11:00 AM  Papers/Presentation session
  ▪ Ilana Shay: Measuring Whitespace Patterns as an Indication of Plagiarism
  ▪ Milt Luoma and Vicki Luoma: Electronic Discovery: A Fool’s Errand Where Angels Fear to Tread?

• 12:00 Noon  LUNCH (provided)
  ▪ Speaker: Donny Cheung, Special Agent, Bureau of Criminal Apprehension, Minnesota
  ▪ Topic: "Day in the Life of a MN Bureau of Criminal Apprehension Computer Forensics Officer"

• 01:00 PM  Papers/Presentation session
  ▪ Mathew Orangi Nyamagwa: A Layered Framework Approach to Mitigate Crimeware
  ▪ Jaikishan Jalan: Canvass: A Software Tool for Forensic Steganalysis of JPEG Images

• 02:30 PM  BREAK

• 02:45 PM  Workshop session
  ▪ Gareth Davies: Best Practices for Data Recovery of Computer Hard-Drive Based Forensic Artifacts

• 04:00 PM  Panel session: Internal Threats: Computer Forensics: Report from Trenches
  ▪ Moderator: Milton Luoma, Metropolitan State University
  ▪ Donny Cheung, Special Agent, Bureau of Criminal Apprehension, Minnesota
  ▪ Jane Laurence, Sergeant, St. Paul Police Department, Minnesota
  ▪ Scott Lawson, Managing Director, Stroz Friedberg
  ▪ Chad Nordstrom, Agent, Department of Corrections, Minnesota

• 05:15 PM  SOCIAL EVENT (bus trip to the Mall of America)

Friday, May 21

• 08:00 AM  CONTINENTAL BREAKFAST

• 08:30 AM  Papers/Presentations session
  ▪ Barbara Endicott-Popovsky: Digital Records Forensics: A New Science and Academic Program for Forensic Readiness
  ▪ John Riley: Developing a Baccalaureate Digital Forensics Major

• 09:30 AM  Papers/Presentations session
  ▪ Gregg Gunsch: The Defiance College Undergraduate Major in Digital Forensic Science: Setting the Bar Higher
  ▪ Grover Kearns: Computer Forensics for Graduate Accountants: A Motivational Curriculum Design Approach

• 10:30 AM  BREAK

• 10:45 PM  Panel session: Curriculum Design and Implementation in Computer Forensics Education
  ▪ Moderator: Michael Stein, Metropolitan State University
  ▪ Israel Aladejebi, Century College, Minnesota (AAS program)
  ▪ Diane Barrett, University of Advance Technology, Arizona (BS program)
  ▪ Gregg Gunsch, Defiance College, Ohio (BS program)
  ▪ Gary Kessler, Champlain College, Vermont (MS program)
  ▪ Jigang Liu, Metropolitan State University, Minnesota (BAS program)

• 11:45 AM  Conference Close

• 01:30 PM  Tour of BCA Facilities
HiGate (High Grade Anti-Tamper Equipment) Prototype and Application to e-Discovery

Yui Sakurai  
The Dept. of Information Systems and Multimedia Design  
School of Engineering  
Tokyo Denki University  
2-2 Kanda-Nisikicho, Chiyoda-Ku, Tokyo, 101-8457  
Japan  
sakurai@isl.im.dendai.ac.jp

Yuki Ashino  
NEC Corporation, Service platform Institute

Tetsutaro Uehara  
Associate Professor  
Academic Information Media Center  
Kyoto University

Hiroshi Yoshiura  
Professor  
The University of Electro-Communications

Ryoichi Sasaki  
Professor  
Tokyo Denki University

ABSTRACT

These days, most data is digitized and processed in various ways by computers. In the past, computer owners were free to process data as desired and to observe the inputted data as well as the interim results. However, the unrestricted processing of data and accessing of interim results even by computer users is associated with an increasing number of adverse events. These adverse events often occur when sensitive data such as personal or confidential business information must be handled by two or more parties, such as in the case of e-Discovery, used in legal proceedings, or epidemiologic studies. To solve this problem, providers encrypt data, and the owner of the computer performs decoding in the memory for encrypted data. The computer owner can be limited to performing only certain processing of data and to observing only the final results. As an implementation that uses existing technology to realize this solution, the processing of data contained in a smart card was considered, but such an implementation would not be practical due to issues related to computer capacity and processing speed. Accordingly, the authors present the concept of PC-based High Grade Anti-Tamper Equipment (HiGATE), which allows data to be handled without revealing the data content to administrators or users. To verify this concept, an e-Discovery application on a prototype was executed and the results are reported here.

Keyword: Anti-Tamper, e-Discovery, Bitlocker, APIHook
1. INTRODUCTION

These days, most data is digitized and processed in various ways by computers. In the past, computer owners were permitted to process data as desired and to observe the inputted data and the interim results. However, the unrestricted processing of data and accessing of interim results even by computer users is associated with an increasing number of adverse events. These adverse events often occur when sensitive data such as personal or confidential business information must be handled by two or more parties, such as in the following cases.

(1) Epidemiologic studies: Epidemiologic studies often seek to compare data concerning (a) physical loads, such as the exposure dosage of individuals at a workplace and (b) the incidence of illnesses, such as cancer, obtained from hospitals and field studies for investigating correlations with various factors. However, when this data is transferred to a partner or third party and the usual processing is performed, personal information contained in the inputted data or interim results may be revealed to the person using the computer to perform the data processing. This is a problematic situation from the standpoint of protecting personal information. Consequently, at present, the exchange of this data is not possible, thereby creating a problem in that epidemiologic studies cannot be used to improve the health of citizens. As a current solution, (i) data encrypted by the original owner may be transferred to a partner or third party, and (ii) the partner or third party may input the encrypted data into a computer, which after the data is decrypted, implements certain processing and outputs only correlation values, so that the interim results and other sensitive information are not accessible to even the computer user.

(2) e-Discovery: Japanese companies are often involved in e-Discovery performed for US civil court. Prior to preceding to trial in a US civil court in both Japanese and US civil courts the defendant and the plaintiff sides mutually disclose electronic evidence. At this time, if electronic documentary evidence containing a keyword or other security mechanism by the plaintiff side exists but is not disclosed, the trial will be severely disadvantaged. Conversely, if all accumulated electronic data is disclosed unconditionally, personal information may be revealed and critical business information may be leaked unnecessarily to a rival plaintiff. For this reason, electronic documents may be partially sanitized [4]. The plaintiff side receiving the sanitized electronic documents may want to verify with a computer that the sanitized portions do not contain keywords, but if viewing the sanitized portions is freely permitted, the defendant-side secrets would be revealed unfairly. To resolve these problems, computer processing may also determine whether the sanitized portions contain keywords, but other processing such as deciphering the sanitized portions should not be performed, and the interim results should not be revealed.

Common to the solutions of these problems is that the computer owner performs only specific processing of the data and is able to observe only the final results. One conceivable means to realize such a solution is to use tamper-resistant equipment that prevents the owner from changing the processing or accessing interim results. Smart cards (also known as IC cards) are typically used as the tamper-resistant equipment. A smart card contains a secret key that is not revealed, even to the user, and public key encryption, which is generally used to implement digital signatures.

Smart cards, however, have the problems of slow processing speed, small memory, and are difficult for an average person to program using an ordinary computer language. Thus, we developed a system that overcomes these problems by improving the PC hardware and software. This new system, configured from PC-based hardware and from software such as the boot control function (BCF) [11] previously developed by the authors, is known as High Grade Anti-Tamper Equipment (HiGATE).

This paper presents the HiGATE concept, and reports the results of the application of a prototype to e-Discovery.

With the heightened interest in the handling of personal and confidential information, the range of applications for HiGATE is expected to increase in the future.
To achieve the same objective, the use of an encryption protocol has also been envisioned, but such an approach would only be suitable for an extremely narrow range of applications and could not be applied to the types of problems encountered here. Approaches similar to that used in our research, that is, preventing even a computer user from freely processing data and accessing interim results, have not been reported in other studies.

2 OVERVIEW OF PROPOSED SYSTEM

2.1 HiGATE Requirements

HiGATE has the following five requirements.

1. A tamper-resistant area exists that even the owner cannot access.
2. A memory with sufficient capacity for calculations is provided.
3. Arithmetic computations are processed at a fast speed.
4. Programming by a trusted third-party, rather than the owner, is permitted.
5. The equipment is easy to realization.

A smart card is an example of a device capable of satisfying the first requirement above. The use of a smart card, however, has the following two problems. First of all, the processing capacity presents a problem, especially in cases such as e-Discovery, where large quantities of data are handled and processing must be done within a limited timeframe. Secondly, programming a smart card requires the use of special techniques such as micro-programming, and an ordinary person would not be able to program a smart card as desired. Thus, the development of HiGATE, which satisfies the above five requirements, was needed for applications not suitable for smart cards.

2.2 Measures for Satisfying the Requirements

(Requirement 1)
The HiGATE functions necessary to satisfy requirement 1 are listed below:

1. Hardware function
   - Capability to prove that the equipment case has not been opened
2. Software functions
   - Boot control function for application programs
   - Encryption function for entire hard disk (HDD)
   - Arithmetic processing function
   - File delete function that leaves intact the inputted data for calculations and the interim results.

These functions are described in detail in Section 3.

(Requirements 2, 3, 4 and 5)

Requirements 2, 3, 4 and 5 can be implemented automatically through the process of PC-based development. In other words, PC-based development provides the advantages of fast I/O throughput and computational speeds, the capability to accumulate large quantities of data, and the ability to use a common programming language such as C or JAVA instead of a special programming language. Moreover, PC-based development facilitates installation since there is no significant need for special equipment. In this case, Windows, with which the authors have much experience, was chosen as the OS.
3. HIGATE CONFIGURATION

3.1 Prerequisites

(Prerequisite 1) The BIOS and OS are running properly.
(Prerequisite 2) Unauthorized actions are not performed when implementing the HiGATE settings.
(Prerequisite 3) HiGATE does not contain an unauthorized program.

HiGATE operates under the abovementioned conditions.

3.2 Possible Unauthorized Actions

There are three conceivable methods for attacking HiGATE, and these are listed below:

(Unauthorized action 1) Forcible opening of the case and extraction of information from the memory
(Unauthorized action 2) Removal of the HiGATE HDD and connection to another PC to extract the HiGATE internal data and programs, etc.
(Unauthorized action 3) Activation of an unauthorized program to modify other programs or steal information.

3.3 Functions to Protect Against Unauthorized Actions

(Protection 1) Provide proof that the case has not been opened.

As in the description of unauthorized action 1, it is conceivable that the case can be opened and the memory contents stolen. Accordingly, it is important to prevent the case from being opened. This protection may be realized by the following two methods.

(1) Cause the power supply or other essential device to shut off and the memory contents to disappear if the case is forcibly opened. Methods for switching off the power supply include: (a) microswitch-based detection of the case opening, (b) detection via a lead switch and a magnet on the case cover, and (c) optical detection of the lid opening.

(2) Provide a seal that clearly indicates whether the case has been opened. Tamper-resistant labels are an existing technology. If a tamper-resistant label is peeled off, an indication that the case has been opened remains. By making the data contained in the HiGATE lose its effectiveness when a seal is cut and peeled off, unauthorized actions by the user can be prevented.

Here, we decided to use method (2), which is easily achieved. In the future, a combination of methods (1) and (2) will enable enhanced safety.

(Protection 2) HDD encryption

HDD encryption is implemented to prevent the HDD used in the HiGATE from being removed or connected to another PC and to prevent the HDD contents from being modified or extracted as described in unauthorized action 2. An existing technology is the BitLocker function which is provided only in Windows Vista Enterprise and Windows Vista Ultimate editions. Because Windows Vista Enterprise is only sold through volume licensing, we installed Windows Vista Ultimate as the HiGATE OS.

(Protection 3) Boot control (BCF/Vista)

As described in unauthorized action 3, it is conceivable that an unauthorized program can be activated to modify other programs and steal information. BCF/Vista [11] is an existing technology developed by the authors to prevent such unauthorized actions. The BCF/Vista function is described in detail in Section 3.4.

Other functions include the file delete and arithmetic processing functions. The file delete function
deletes data from the HDD used with HiGATE. This function operates so that the HiGATE user does not leave behind any data after the usage of that data has been completed. The HiGATE system handles data which even the user is not permitted to view. Accordingly, the continuous retention of data in the user’s HiGATE system will lead to an unauthorized action. The arithmetic processing function is used in different ways depending on the application, and can be developed freely by the program developer. Additionally, unauthorized actions through the skillful use of the keyboard are conceivable, but can be avoided through the use of the device driver installation control function in Windows Vista Ultimate.

3.4 BCF/Vista

BCF/Vista was a part of the development of the Dig-Force [2][3] digital forensic system.

BCF/Vista computes hash values for programs that are activated during setup, registers these hash values in a white list, and digitally signs the entire white list. The OS and then BCF/Vista are set to start next. After they are running, the BCF/Vista function computes hash values for application programs as they try to start up, and after confirming the validity of the previously registered white list by signature verification, the computed hash value is compared to the hash value in the white list. If the value is the same as the registered hash value, the program starts, but if not registered, APIHook is used to prevent the program from starting.

As a result, even if an unauthorized program attempts to start up, if Prerequisite 1 that “the BIOS and OS are running properly,” has been established, the unauthorized program can be prevented from running. For further details regarding the BCF/Vista function, refer to reference document [11].

3.5 Functional Configuration

Providing the hardware function and the four software functions listed in Section 2.2 enables a tamper-resistant area to be realized to protect the OS and the arithmetic processing function residing on the HDD, as well as the keys, data, etc., residing in memory. Accordingly, an arithmetic processing program created by a program developer and placed in the tamper-resistant area can be executed correctly by the HiGATE system without any unauthorized actions. The HiGATE functional configuration is shown in Figure 1 and smart card and HiGATE functions are compared in Tables 1 and 2.

![HiGATE Block Diagram](image-url)
Table 1. Differences between Smart Cards and HiGATE

<table>
<thead>
<tr>
<th>Difference</th>
<th>IC Card</th>
<th>HiGATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>EMV specification</td>
<td>Windows Vista Ultimate</td>
</tr>
<tr>
<td>Limitation due to program language development</td>
<td>Limited</td>
<td>Not limited</td>
</tr>
<tr>
<td>Memory</td>
<td>RAM (1 MByte)</td>
<td>RAM (3 Gbytes)</td>
</tr>
</tbody>
</table>

Table 2. Similarities between Smart Cards and HiGATE

<table>
<thead>
<tr>
<th>Similarity</th>
<th>IC Card</th>
<th>HiGATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Startup of predetermined program</td>
<td>Read-only semiconductor or memory</td>
<td>BCF/Vista</td>
</tr>
<tr>
<td>Memory cannot be accessed externally</td>
<td>Encryption circuit or single-chip implementation of memory</td>
<td>Case cannot be opened with tamper-resistant label</td>
</tr>
<tr>
<td>False environment cannot be created</td>
<td>Access controlled by CPU</td>
<td>HDD encryption</td>
</tr>
<tr>
<td>Tamper-resistant area</td>
<td>Access controlled by CPU</td>
<td>• BCF/Vista</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• HDD encryption</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Tamper-resistant label</td>
</tr>
</tbody>
</table>

4. HIGATE OPERATION

This section describes the HiGATE operation:

1. HiGATE manufacturing phase
2. Program installation phase
3. Setting phase
4. Usage phase.

Persons involved in the above four steps of HiGATE operation include the manufacturer, program creator, and user.

1. Manufacturing phase
   The manufacturer preinstalls the OS (Windows Vista) and BCF/Vista required for HiGATE in the HiGATE PC, and then transfers the HiGATE system to the program developer.
Program installation phase
The program developer receives the HiGATE system from the manufacturer, and loads the processing programs necessary for application into HiGATE.

Setting phase
After program installation, the HiGATE system and BCF/Vista are configured with the necessary settings. At this time, the program creator implements the settings with administrator privileges.

1) BIOS settings
To prevent booting of an OS installed on the PC other than Windows Vista, “Set the BIOS password” and “Restrict bootable storage to the HDD” BIOS settings are implemented, and only the program creator knows the BIOS password.

2) User account settings
So that the operator cannot operate Windows services and the task scheduler from the user account that is utilized when operating the PC, the user account is not given administrator privileges.

3) Program installation
Controller program files that comprise BCF/Vista, agent program files, and the white list files are installed. These files are stored on a drive encrypted by BitLocker (described below), and are set as read-only files that cannot be overwritten or deleted by the user account.

4) BitLocker settings
BitLocker is a drive encryption function of Windows Vista Ultimate. BitLocker is used to encrypt all the drives on an HDD so that an attacker is unable to remove the HDD from the PC and use another PC to modify the programs that comprise BCF/Vista.

5) Windows service settings
Nanshiki Corp.’s sexe freeware was used to register the controller as a Windows service. This service was registered under the name MonitoringController (hereafter referred to as MC Service). As a result, the controller starts up automatically after Windows Vista has started. With this setting, however, MC Service will not start if the user account is logged in while in safe mode. Therefore, MC Service information is added to the Windows registry so that MC Service will start even in safe mode.

6) Task scheduler settings
The task scheduler is used so that startup occurs when an agent logs into the user account.

After implementing these six settings, tamper-resistant labels are affixed to all HiGATE parts that could be opened.

Usage phase
The user uses a HiGATE system which has been set up as above. At this time, a user account that does not have administrator privileges is used by the user.

These are the phases of HiGATE operation. By allocating the above roles to the participants involved, the participants can be prevented from performing any unauthorized actions. HiGATE users do not possess administrative privileges, and therefore it would be difficult for them to perform an unauthorized action on the data. Moreover, since the program creator does not possess the HiGATE system, it would be nearly impossible for him or her to implement an unauthorized action on the data handled by the user. Accordingly, none of the participating individuals are able to perform an unauthorized action on the data used and handled by the HiGATE system. The prerequisite that the BCF/Vista administrator does not perform any unauthorized actions is made possible by this system.

5. APPLICATION TO E-DISCOVERY
In this section, the HiGATE system is applied to e-Discovery. The e-Discovery system described is the “e-Discovery System for Sanitizing Disclosure Information and for Securing Evidence” [1]
proposed by Takatsuka et al.

5.1 e-Discovery

The US Federal Rules of Civil Procedure (FRCP) were amended in December 2006 so that in US civil litigation, corporations are obligated to disclose electronic evidence during the discovery phase held before the start of a civil trial. This disclosure of electronic evidence is known as e-Discovery [5][6][8]. Figure 2 shows the basic flow of the e-Discovery procedure.

![Figure 2. e-Discovery Procedure](image)

5.2 e-Discovery System Proposed by Takatsuka et al.

First, the plaintiff side provides the defendant side with keywords which relate to the case. Then, using those keywords, the defendant side discloses files relating to the case. However, these files may include information such as company secrets that the company does not want to disclose. As a solution, electronic sanitizing technology [4] is used. The defendant side completely sanitizes documents that do not contain keywords, and partially sanitizes documents that contain both keywords and passages which the company does not want to disclose. This method permits the disclosure of only the minimum required data. With the present sanitization technology, however, a keyword may be contained in an area to be sanitized, and verification thereof is nearly impossible. Takatsuka et al. solved this problem by encrypting the data contained in a sanitized area. Data that has been subjected to sanitization with encryption is disclosed to the plaintiff side. The plaintiff side accepts and verifies the data. At that time, the plaintiff side verifies that keywords are not contained in sanitized areas.

With this process, however, the decrypted results of sanitized locations may be unfairly observable. Therefore, a tamper-resistant security device is used so that interim results cannot be observed, and the HiGATE system is used for this purpose. Here, a HiGATE system that includes the program part developed for e-Discovery is referred to as HiGATE/e-Discovery.

We developed the minimum needed programs for HiGATE/e-Discovery. The languages used are C# and C++, and constitute a total of approximately 3,000 steps.
5.3 HiGATE/e-Discovery Operation

This section describes the operation when HiGATE is applied to e-Discovery. HiGATE is used as a security device at the application site. Application of the HiGATE system solves the problems of data interference by the plaintiff side and the slow processing speed associated with smart cards.

1. HiGATE/e-Discovery manufacturing and program loading
   The manufacturer transfers the HiGATE system to the program developer. The program developer loads e-Discovery software into the HiGATE system. The e-Discovery software program has five functions for digitally verifying keyword files, decrypting sanitized passages, verifying whether keywords exist in a decrypted passage, verifying evidence, and extracting files relating to the case.

2. HiGATE/e-Discovery settings
   The HiGATE settings were described in Section 4. e-Discovery keywords are also delivered and the plaintiff and defendant sides exchange public keys.

3. HiGATE/e-Discovery usage
   The plaintiff side inputs the data received from the defendant side into HiGATE/Discovery, and runs the program. This enables the plaintiff side to verify that keywords are not contained in the sanitized areas. Next, the plaintiff side reviews only the data related to the case, and moves ahead with the litigation.

Based on these considerations, the applicability of HiGATE to e-Discovery appears promising.

6. CONCLUSION

This paper has proposed hardware that is equipped with a tamper-resistant function that cannot be modified by even the hardware owner, has fast I/O throughput and arithmetic processing speeds; is capable of accumulating large quantities of data that can be programmed similarly to a PC and can be developed inexpensively and easily. A prototype was built and the results of an application of e-Discovery were reported.

The basic HiGATE program developed this time has only minimal functionality, but we intend to expanded the functionality in the future, and improve the ease-of-use and safety.

The HiGATE system is suitable for application to situations requiring technology for proving that an administrator has not accessed files or participated in the processing. In addition to e-Discovery, the application is thought to be possible in many other fields, such as data matching in an epidemiologic study.

REFERENCES


Developing VoIP Honeypots: a Preliminary Investigation into Malfeasant Activity

Craig Valli
Secau – Security Research Centre
Edith Cowan University
Perth, Western Australia
c.valli@ecu.edu.au

ABSTRACT
30 years ago PABX systems were compromised by hackers wanting to make long distance calls at some other entities expense. This activity faded as telephony became cheaper and PABX systems had countermeasures installed to overcome attacks. Now the world has moved onto the provision of telephony via broadband enabled Voice over Internet Protocol (VoIP) with this service now being provided as a replacement for conventional fixed wire telephony by major telecommunication providers worldwide. Due to increasing bandwidth it is possible for systems to support multiple voice connections simultaneously. The networked nature of the Internet allows for attackers of these VoIP systems to enumerate and potentially attack and compromise a wide range of vulnerable systems. This paper is an outline of preliminary research into malfeasant VoIP activity on the Internet.

Keywords: VoIP, honeypot

INTRODUCTION
Voice over IP systems are replacing conventional switched wire telephone devices, these systems rely on Internet connectivity for the transmission of voice conversations. Many large corporations have been successfully using these types of systems over a number of years to lower communications costs relating to telephony. We are now seeing this technology being placed into homes through such initiatives as the UK-based BT Home, Australian based iinet BOB and various American offerings. These systems work by replacing conventional fixed line telephony with broadband connected wireless enabled routers that provide not only Internet access but VoIP services within the home or business.

Some 30 years ago computer hackers compromised PABX systems to facilitate cheap long-distance calling through the rerouting of their modem calls through these compromised systems. This illegal rerouting enabled the hackers for the cost of a local call to stay connected to long-distance phone calls for any period of time that they wished to. The actual costs for such long-distance calls was borne by the owner of the PABX system not the hacker and diagnostics on these systems often only indicated that the calls originated from the PABX. This mode of attack started to decrease as PABX developers hardened their devices against attack by these interloping attackers by placing countermeasures in the PABX. This coupled with long distance telephony becoming cheaper and eventually broadband digital services such as ADSL and ISDN replacing acoustic modems and the PSTN as the main carrier for Internet traffic.

Due to technological restraint it was relatively difficult to have more than a one-to-one connection between the attacker and responding devices. This paradigm has now changed with the introduction of VoIP systems where an attacker can control many telephony sessions from the one device.

VoIP is most typically sent as a clear text transmission and is relatively trivial task to intercept with a simple packet sniffer. The transport for the supporting protocols of VoIP are typically UDP based. In addition there are already a variety of tools that can intercept and replay voice conversations that are conducted over VoIP channels. Not all of the tools capable of interception or enumeration are in the
realm of hacker based tools some are legitimate packages such as Wireshark or NMAP. VoIP systems suffer from the same Achilles' heel that all core services do in that for it to be useful it must be available for connection from unknown parties. In the same way that a Web server serves web pages to requesting clients likewise a VoIP server must allow unknown connections to start for a voice conversation to occur. As a result much of the authentication and authorisation has to be open, disallowing the provider of the service basic protections against malicious activity from parties who are interested in enumerating or compromising a system.

**THE CURRENT LANDSCAPE – FACTORS INTERSECTING**

There are various tools available for the enumeration and subsequent compromise of VoIP systems. The following table represents some of the more common tools and indicates their abilities many of these tools have been freely available since 2005.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description/Modus Operandii</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIPVicious Tool Suite - svmap, svwar, svcrack</td>
<td>svmap lists the SIP devices found in an network. svwar maps active extensions on a PBX. Svcrack is a password cracker for SIP</td>
</tr>
<tr>
<td>sipflanker</td>
<td>Many (if not most) VoIP devices have available a Web GUI for their configuration, management, and report generation. And unfortunately it is also common for the username and password to have the default values.</td>
</tr>
<tr>
<td>sipcrack</td>
<td>Sipdump finds SIP logins, then sipcrack is used bruteforce passwords on the identified logins</td>
</tr>
<tr>
<td>steganrtp</td>
<td>SteganRTP is a steganography tool which establishes a full-duplex steganographic data transfer protocol utilizing Real-time Transfer Protocol (RTP) packet payloads as the cover medium. The tool provides interactive chat, file transfer, and remote shell.</td>
</tr>
<tr>
<td>sipp</td>
<td>Test tool and traffic generator</td>
</tr>
<tr>
<td>sipsak</td>
<td>sipsak is a small command line tool for developers and administrators of Session Initiation Protocol (SIP) applications. It can be used for some simple tests on SIP applications and devices.</td>
</tr>
<tr>
<td>VoIPER</td>
<td>VoIPER is a security toolkit that aims to allow developers and security researchers to easily, extensively and automatically test VoIP devices for security vulnerabilities</td>
</tr>
<tr>
<td>UCSniff</td>
<td>UCSniff is an assessment tool that allows users to rapidly test for the threat of unauthorized VoIP eavesdropping. UCSniff supports SIP and Skinny signaling, G.711-ulaw and G.722 codecs, and a MITM ARP Poisoning mode.</td>
</tr>
</tbody>
</table>

In addition to this, there have been numerous articles in the literature that relate to vulnerability in the VoIP protocol (Bradbury 2007; Herculea, Blaga et al. 2008; Jouravlev 2008) some even describe how
to use the above tools to achieve enumeration or even exploit of a VoIP system. There also several books on the exploitation of VoIP (Endler and Collier 2006). This level of information availability, coupled with the increasing numbers of encoded ready to use hostile tools is a malevolent landscape waiting to happen.

In the last three to five years many large infrastructure providers as previously mentioned have started to offer integrated IP/Telephony/VoIP solutions for the home user. This current status quo now provides the attackers with not only knowledge that has moved from tacit (theoretical attack) to explicit (code tools for exploit) but also an increased opportunity for exploit and compromise of hosts through the deployment of VoIP systems in homes sees for the first time viability of attacking these systems for gain.

**MOTIVATIONS FOR ATTACKING VOIP**

Most of the key drivers for motivation of attacking and compromising VoIP systems is built around financial gain in a similar fashion to the PABX compromises of past. The attacker's motivation is to use the victim system as a conduit for malfeasant traffic and have the victim bear the charged cost of any such activity. The other motivations of course are still classic disruption or denial of service to the victim.

**FINANCIAL GAIN**

There are numerous telephone calling cards one can use to call cheaply overseas that utilize systems legitimately and use technology such as VoIP to lower costs for providers of the service. The low cost is enabled through the use of VoIP technologies. The use of a calling card normally involves calling a legitimate number and then keying in the long distance number you wish to call. As a result of keying in the number you are then normally redirected through a cheaper service typically VoIP and calls are initiated to the requested number. The call quality of these systems is sometimes degraded but then the costs are significantly lower for users of these systems.

The illegitimate use of these systems would see compromised VoIP systems being utilised to route the incoming calls to the destination at the expense of the owner of the compromised system. The other method is to use rerouting to dial premium subscriber-based phone lines that charge a set rate per minute for the connection to the end user typically on a range of between $1 - $8 a minute again typically depending on the type of service offered. The difference in price after servicing fees from the provider of the premium number then becomes the profit that is sent to the entity that set up the premium number service for providing the service normally the compromiser of the system.

Malicious users could hide much of their activity by distributing the call load across a multitude of compromised devices in essence a distributed compromise of service. By spreading the call load across a wide range of compromised devices it affords the malicious users good protections from detection by the owners of the systems. Modifying one of John Paul Getty sayings illustrates this concept “it is far easier to steal a $1 from 100 people than to try and steal $100 from one”. By using this *modus operandi* not only do attackers lower their forensic fingerprint but also make it difficult for intrusion detection systems and other methods of monitoring to detect a malfeasant call.

**DENIAL OF SERVICE**

Denial of Service (DoS) is a simple and well proven attack method which can be achieved by flooding network connections, socket/port exhaustion, or resource exhaustion of a service or server by overloading with the end result being memory or CPU exhaustion. This exhaustion of service can have catastrophic outcomes resulting in servers halting and rebooting or simply service to be useless due to performance degradation. Motivations for this could be blackmail as has been evinced already in businesses where availability is crucial for business e.g online casinos where if you want to be able to collect revenues you have to be available and online. This availability nexus is as previously mentioned an Achilles heel for service of this kind.
The other motivation is to gain a competitive advantage over a competitor remembering not all businesses behave ethically. If people cannot contact you on the phone service you are using it becomes increasingly difficult to conduct business with you and they will seek out an alternative. A DoS in these types of cases does not have to be 24/7 to be effective, attacks at key times during a business calendar can have devastating consequences for any business sustainability or viability.

**DETECTION OF MALICIOUS ACTIVITY**

The use of intrusion detection systems whether they be host based or network based is one preferred way of detecting malicious activity on a network directed towards VoIP systems. Many of the attack or enumeration tools such as SIPvicious tool suite (Guac 2010) when used in default mode are overt, verbose and readily identified. The following extract from a system log file is an example of a transaction between a system and the SIPvicious tool suite:

```
UDP message received [416] bytes:
OPTIONS sip:100@xx3.xxx.xxx.xxx SIP/2.0
Via: SIP/2.0/UDP 10.160.67.18:5061;branch=z9hG4bK-2559388112;rport
Content-Length: 0
From: "sipvicious"<sip:100@1.1.1.1>;
tag=63626131373537653133633401313036333036333336
Accept: application/sdp
User-Agent: friendly-scanner
To: "sipvicious" sip:100@1.1.1.1
Contact: sip:100@10.160.67.18:5061
CSeq: 1 OPTIONS
Call-ID: 1020236891970287777884434
Max-Forwards: 70
```

As can be clearly seen as highlighted there are hallmarks or indicators of a SIPvicious based attack on a system. It is a simple task to write some intrusion detection system rules from this exchange. An example Snort rule could be:

```
alert UDP any any -> any (msg:"sipvicious default scan"; content: "|736970766963696f7573|";)
```

This rule simply traps for any UDP connection that has the string sipvicious in it. Alternatively the string friendly-scanner could also be trapped. With the use of dynamic rules one could also for instance trap the next 500 packets from the attackers IP to be able to examine any connections or activity being undertaken by the attacking IP.

The caveat on these types of attacks is that the use of these tools is as provided by the author and that are typically initiated by inexperienced attackers often referred to as script kiddies or n00bs. The inexperienced attacker does often not know or often understand the tool they are using and that they are leaving behind a large forensic fingerprint. One could further postulate that these types of attacks are not organised criminals or experienced cyber criminals searching for vulnerable VoIP systems. These attackers instead will utilise methods that are relatively anonymous and will avoid detection by intrusion detection rule sets in use. This exact scenario now presents another problem for providers or users of VoIP systems who wish to protect the service. The problem is that to protect a VoIP system from compromise one has to assume that all connections are malicious and subject them to intense inspection and scrutiny and or use a whitelist which defeats the purpose largely of having a open connection. One of the known degraders of any network based system and in particular VoIP based systems performance is latency, by performing any form of packet inspection packet latency will be increased. Hence remedy through the enforcement of large rulesets and packet inspections may in fact be worse than the overall complaint.

**USING SIMPLE HONEYPOTS TO DETECTING SCANNING OR ENUMERATION**

A talk by Sjur Usken (Usken 2009) outlines a method for using SIPP (Gayraud and Jacques 2010) and
a packet logging tool *Daemonlogger* (Roesch 2006) as a simple honeypot for the detection of scanning and enumeration by an attacker of VoIP systems. The *SIPp* suite is a test tool or traffic generator for the SIP protocol that is legitimately used to test systems. It enables call establishment, call flow analysis, message statistics and the testing of a range of features found in VoIP systems. This suite provides the basis for low level interactions with the attacking entity.

*Daemonlogger* is a program that can sniff network traffic and either spool it to disk or redirect it to another network interface. In the executing of the current research it is used to write the potentially malicious traffic to disk and log it for later forensic analysis. *Daemonlogger* uses rulesets in tcpdump syntax an example rule follows

```
dst port 5060 or dst port 16384 or dst port 5061 or dst port 1720
```

will trap for connections on ports 5060, 16384, 5061, 1720. Once there are packets trapped then action is initiated to record the packets.

The system being used in our research utilizes rudiments of the system proposed and used by (Usken 2009). The system is replicated across a number of sensors we have installed in a honeypot system that utilizes Surfnet IDS as its supporting infrastructure. SurfnetIDS uses a collection of sensors that are connected back to logging infrastructure via VPN. In our setup there is a multitude of recordings occurring.

<table>
<thead>
<tr>
<th>Basic SQL</th>
<th>The sensors report their VoIP activity back to a customized SQL database for logging of attack data. No packets are captured.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daemonlogger</td>
<td>File based logging that daemonlogger provides on each sensor, any files produced by daemonlogger are sent back via SCP to the central logging server for storage and analysis. This is in addition to a full dump of the Ethernet connection using tcpdump –w</td>
</tr>
<tr>
<td>SurfIDS_Snort</td>
<td>The sensors report via VPN to a Snort based database using custom IDS based rules that respond as a result of a known VoIP attack. These records are incorporated into the SurfnetIDS reporting mechanisms. In this record the attacking IP numbers are also matched against the GeoIP suite for country of origin information.</td>
</tr>
</tbody>
</table>

As the honeypot research is formative and utilizing empirical learning to modify and adapt the honeypot previous experience in deploying honeypot research indicates that there is a scientific need to capture attack data with a multitude of types. The use of multiple streams or samples also allows the use of multiple tools and techniques to validate findings or investigate observed phenomena.

**TECHNIQUES UNDER DEVELOPMENT**

Apart from deep packet inspection in Wireshark and other packet analysis tools, the research is attempting to find suitable methods for automated alert and response to VoIP oriented attacks.

**AUTOMATED ALERTING DEVELOPMENTS**

This is already enabled by the research modus operandii to some degree by the use of Snort and custom IDS rulesets we are developing for VoIP systems. IDS systems in combination with analysis consoles such as ACID or Base or SnortAlert allow for much of this systematic alerting or monitoring to occur. Or in the case of this research enable timely inspection and review of attacks on the systems.

The research is also looking at methods for using daemonlogger to redirect output to another interface that becomes a trigger response i.e any traffic on that interface indicates potential malicious activity. This could be a idle VPN connection that when triggered will initiate connection that then results in an
escalation of response or monitoring.

The on-demand production of graphical representation of activity using Graphviz (Ellson and Gansner 2008) produced images using Afterglow is also another approach being developed in the research. The need for using graphical methods for analysis in combating the complexity of alert data received by network systems is well established. The ability to generate graphs on demand allows a new lens or view of the dataset.

AUTOMATED RESPONSE DEVELOPMENT

Automated response is largely enabled by IDS in this case Snort and development of dynamic rules in combination with firewall or host.deny responses. This dynamic use of IDS as always is a balancing act between false positives and legitimate usability of the system. Another factor that increases problems is that VoIP systems use UDP for transport and the resulting transmissions lack the formal controls and discipline with respect to packet sequence, latency and other attributes that a TCP based transmission can be afforded.

As mentioned previously the use of dynamic rules in Snort allows for a variety of actions not all of them need be a complete denial of route to be effective. The triggered recording of attack sessions allows for smaller analysis parcels to be produced rather than having to trawl large packet capture for small packet sequences this produces significant efficiencies in analysis.

Response via device fingerprint spoofing is a well used paradigm in honeypots and honeyd (Provos 2007) is just one exemplar of the technique. The honeypot system will use known robust fingerprints of a device to fool the attacker into believing they are in fact scanning or interacting with a real device when in fact it is the honeypot. The current research is attempting to generate reliable system fingerprints using NMAP an active network fingerprinter and p0f a passive network fingerprinter for commonly used ADSL VoIP enabled routers. In addition via the production of simple PERL scripts we are also trying to emulate basic service responses to probes by some of the commonly used tools. The aim of the scripts is not to provide perfect emulation of service but merely to mimic standard responses in an attempt to get the attacker to escalate the attack to a higher level of interaction and attack.

CONCLUSION

The research is currently ongoing and should produce systems capable of a more realistic emulation of vulnerable VoIP systems. The research purposefully focuses on VoIP system vulnerability and not standard honeypot systems as there is a significant and successful body of knowledge based around these systems already. It should however, be noted that much of the research is underpinned by successful techniques and methods from within the existing honeypot knowledge domain.

The research has already uncovered increasing levels of probing and enumeration of VoIP systems from a variety of locations on the Internet. The next stage of the research is to start classifying the level of enumeration and attack being undertaken by these attackers to determine if the activity is mere scanning or escalatory in nature and activity that is genuinely seeking to compromise the systems.

REFERENCES


A Framework to Integrate the Data of Interview Investigation and Digital Evidence

Fahad Alshathry
De Montfort University
Software Technology Research laboratory (STRL)
Leicester, UK
famalsha@gmail.com

ABSTRACT

The physical interview process in crime investigation produces an extremely large amount of data, particularly in big cases. In comparison, examiners of digital evidence have enormous amounts of data to search through whilst looking for data relating to the investigation. However, the links between their results are limited. Whilst investigators need to refute or support their hypothesis throughout, digital evidence examiners often use search based keywords. These keywords are usually created from evidence taken from the physical investigation reports and this basic method has been found to have many shortcomings and limitations. This paper proposes a highly automatic framework to integrate anything suspicious that victims or witnesses have said in their interview with their digital data. The proposed model applies to both physical crime investigations and digital evidence examinations.

Keywords: Computer Forensics, Digital Evidence Analysis, Crime Investigation.

1. INTRODUCTION

One of the most frequently discussed and emerging topics in law and criminology is computer forensics. Computer forensics is one of several expressions about whole procedures that deal with digital evidence when involved with digital crime (e.g. hacking) or physical crime (e.g. terrorism). It supports the investigation hypothesis to get answers for these five questions: what, where, when, why and who did the crime.

The most critical phase in both digital crime and physical crime is the data analysis. A number of analysis tools are available in the market; however, they are designed for digital crime and use simple matching techniques. Consequently “the investigator becomes inundated with data and wastes valuable investigative time scanning through noisy search results and reviewing irrelevant search hits”[1]. Thus, “existing general purpose computer forensic analysis tools are rapidly becoming inadequate for modern analysis workloads”[2].

Although a number of models in digital crime are introduced, the physical crime investigation requirements are diverse and should have a specific framework to compare the conclusion of field investigators and the content analyst regarding the psychological characteristics and motives of suspects of concern.

There are several reasons why this framework is needed. Firstly, “the majority of digital evidence that is processed by law enforcement today is on computers used as instruments of traditional crimes”[3]. Secondly, these days most people have at least one form of digital storage e.g. cell phone. Historically, the amount of physical crime is enormous compared with digital crimes which have come to the fore in the last two decades. In addition, it is possible to reduce digital crime by improving security applications and awareness, but preventing criminals using computers as instruments in physical crime is impossible. Finally, when a number of pieces of digital evidence have been involved in a traditional crime, particularly in big cases, questions emerge: who is supposed to investigate it or lead the investigation team? A person whose background is criminology (a normal investigator) or whose background is computing (a technician)? While the analysis in traditional crime depends on
disciplines that have continued for hundreds of years, using computers in physical crime has only arisen recently. Therefore, confusion among technicians and normal investigators is common and has led to a lowering of the standards of proficiency.

Consequently, the loss of correlation between the statements of the accused or victims and their digital evidence data possibly generates truncated evidence which does not support one another. This can be particularly obvious when the case is manipulated by numerous investigators, different agencies or when the media is examined by different technicians. This issue is repeated in questions and recommended research in number of the Digital Forensic Research Workshop (DFRWS) 2001[4], 2004 [5] and 2005 [6].

The objective of this work is to introduce a framework to integrate these two types of evidence and to improve the coordination between the two investigatory teams.

This paper is organized as follows: the second section describes the methodology; the third section is the research hypothesis. The fourth section is the proposed investigation framework. The fifth section shows a scenario. The sixth section depicts related work. The seventh section presents the discussion. The final section gives the conclusion and future work.

2. METHODOLOGY

The research area is very critical particularly in the collection of the data. Therefore, an ad hoc methodology is applied. To adapt to the issue; however, we divided our research plan into three phases. The author worked at a computer forensic research centre for years. Throughout this experience, he observed the process of an organization technique when dealing with digital evidence. He found that the gap between investigators and technician examiners is a serious problem. It is considered to be the main cause which prevents the investigators from using digital evidence in the investigation process. Therefore, this experience is taken as a valid base for the research. Secondly, a comprehensive revision of previous studies in the research domain is conducted looking for a solution in the literature. Thirdly, the framework has been realized and is currently being tested.

3. THE HYPOTHESIS

The examiner in computer forensics formulates a hypothesis to start the analysis. This hypothesis relies on what the investigator needs. A common search technique is based on keywords, or metadata. However, physical investigation is more than a search by keywords, and requires an accurate observation of the psychological behaviour throughout numerous questions, e.g. does X work on his computer and when? Does X navigate the internet and why? What the user wrote and why? Does X have a relationship with Y and at what level? And for how long?

This work assumes that digital evidence is a number of activities which have been added and accumulated in the past in digital storage. These activities may show the users psychology, behaviours, plans or social relationships. This work, also assumes that the physical crime investigation process is like a spiral cycle. Throughout this cycle, the investigator attempts to answer hypothetic questions. Of course, from this perspective, the entities (e.g. involved persons) in a case profile are increasable. Hence, the investigation continues until the answers are achieved. We assume, also, that when the technician could correlate these entities and activities, answers may be provided to the investigation questions. However, computers may contain thousand of files and include hundreds of errors in various ways, or even have several parts which may be missed. Therefore, a number of questions may be left without answers; the users of digital evidence who are the suspects, victims, or witnesses, of course, have these answers. In criminology, the fastest way to extract these answers is through interrogation. Although investigation data and digital evidence may answer these questions, the barrier is how can we integrate these two sets of data? However, when we achieve this, it is possible to apply this technique to digital crime investigation as well.
4. PROPOSED INVESTIGATION FRAMEWORK

This model divides the whole process into four main blocks: Crime Scene Phase, Inspection Phase, Interview investigation Phase and Integration Phase. Any main phase has sub-phase[s] as shown below: (see Figure 1)

4.1 Crime Scene Phase

Although this phase is not fully technical with the exception of some circumstances, it provides a solid base for other phases. This phase exists in most current models and has sub-phases explained below:

- Identifying an Incident and Crime Scene

The investigators receive the initial information about the incident and issue a written search warrant. The level of formal authorization varies considerably depending on the crime type and its level. At this stage, all equipment, and investigators assistants must be prepared. Although identification is not a technical task, it forms the basis of the case and if there is a mistake the case will have weaknesses which may destroy the legality of the evidence.

- Preservation of and Surveying a Crime Scene

Preservation is the activity in a crime scene to protect the evidence that appears to be interesting for investigation. However, because digital evidence is changeable and can be modified, this requires extra care, e.g. investigators take the hash value then do a full copy image of the irremovable media immediately to avoid possible spoilage. In addition, they document every action by taking many
pictures of evidence; determining their location and using labels which are related to each other, describing the computer, its manufacturer, the data storage, its serial numbers, its interface and so on. Once the evidence has been secured, the [normal] investigator makes a survey and formulates hypotheses to explain what happened. Although the investigator may suppose a hypothesis relating to the visible objects, it is impossible to correlate that with digital evidence until the data has been extracted and displayed from the hardware.

- **Collection and Transportation**

  This phase is not only dedicated to digital evidence, but also includes people who may be involved in the crime. The people, according to their relationship with the crime, may be sent to hospital, to their homes or to the investigation department. On the other hand, digital evidence would be sent to the digital evidence laboratory. The collecting and transporting of digital evidence requires extra attention to prevent any loss or damage. Furthermore, it is important to insure that transport does not affect the integrity of the evidence.

4.2 Inspection Phase

This phase is fully technical; therefore, people who work in this phase must have good experience in computing e.g. file systems, data structures, and encryption. However, they should not engage in data comparison or entities relationships. This phase consists of three sub-phases explained as below:

- **Bit-stream Imaging**

  The best evidence is the original storage media. However, keeping the original evidence in the laboratory is not permitted. The insight behind the operation on a full copy of evidence is to keep the original without any modification. Therefore, getting the hash value is significant to ensure that evidence is not modified. Imaging means a bit-by-bit copy of the original evidence. In other words, the exact data stored in the duplicated media must be the exact data in the original. Once the image is performed, the original evidence should be kept in a secure place out of the laboratory building.

- **Low Level Data analysis**

  In this sub-phase, data analysis is involved with manipulating all text data to be active for the system and readable for the examiner; that includes recovery of deleted files, hidden data, decryption of encrypted data and translation of human language if different. Data analysis that belongs to a network layer should be illustrated in this phase as well, and then the results should be sent back to case profile. In addition, whole picture files, video and audio should be given comments explaining their contents e.g. if a persons name or mobile number has appeared in a video file, these entities must be extracted into a text file and stored as analyst comments.

- **Data Transformation and Store**

  Once the file is readable as text, it should be stored in a digital evidence database. It should have a few details about the case such as the digital evidence’s owner, the crime description, and the information that is required. Therefore, the evidence data should be linked with these details and stored as object files with their hashes inside a directory. Authorization to modify this data is restricted. Users are allowed to read-only the data and copy files when needed. To enable search and comparison later, the file format of all files should be illustrated in this phase. The text file format should be stored as a plain text.

4.3 Interview investigation Phase

Interview investigation (or interrogation) requires correlating all entities with the activities that have been gathered about the persons involved as well as at the crime scene to determine suspects. In addition, it requires documentation for any action, i.e. questions and answers. This data is required to be filtered and analyzed to explore any strange and relevant phrases or words. Therefore, the proposed system requires that the data should be stored in a database, to automate the investigation process and
to allow information to be compared with other resources.

4.4 Integration Phase

This phase is central to the whole approach of the solution when we can determine the two targets, interview investigation as digital texts and digital evidence data, and integration will be possible. The matching of these sides should be presented in a view screen with referencing of each hit and word context to allow an analyst to give their technical opinion as comment. The advantage of our system is the dynamic in exchange information between investigators and technicians. This phase consists of three sub-phases as explained below:

- **Text Mining**
  In this sub-phase, it is expected that the system shows not only similar words, patterns, but also it should present more inferences, e.g. are these patterns significant for investigation, is this word a persons name or does it mean something different. This technique is called semantic text mining.

  In addition, the system uses a number of techniques to extract the knowledge such as machine learning, text chunking based on entities and their context, association rules, clustering entities and ontology. These techniques allow users running a number of representations (e.g. generating a series of questions automatically, presenting Information visually (e.g. entities with relationships).

  The proposed system has two approaches to present its results. The first form is dedicated for the normal user, who wants to run the search manually. The second form is to show the dynamic search results. Using the dynamic search shows any matches in real time during the interview investigation.

- **Decision Support**
  In respect of crime detection, the decision in crime investigation is very sensitive. As a consequence, the system results are not considered as the final decision until they are validated by an analyst. However, time is significant in an investigation and the expected number of hits is enormous. Thus, the system should provide an abstract level of decision to allow the analyst to start commenting on hits that have high priority rather than hits that may be further from the core profile of the case. There are two methods to evaluate matching words to assess if they are significant or not: 1) comparison with a lexical of suspicious words and phrases (e.g. hate words) judged by their context; 2) the case history data is used as the foundation of investigation judgments and it is used widely in a normal process; hence, the system decides the entities that are recorded in the case history database as significant words for investigation.

- **Comment and Reporting**
  Throughout the text analysis, the expert checks all hits and adds the value of hits into the interview investigation database as new questions. These questions should be answered by the suspects, victims or witnesses. Their answers will be searched again until the hits are completed. The comments may also be sent back to the analyst to provide an explanation for any ambiguous technical points. The system stores comments in the digital evidence database as new events associated with the case profile; the events are shown to the technician to allow him to replay their answers. Therefore, the reporting stage is considered like a switch for the results.

5. SCENARIO

This case shows how to apply the framework to the analysis of digital evidence and the integration of its data with the interview investigation. Because it is very difficult to present a real case, the case study is fictional, but is based on frequent cases, which have occurred when a crime has been associated with digital evidence.

A suspicious car was being tracked by an investigator before it had an accident and the driver escaped. The investigator found a computer laptop, three CD’s and two small locks of hair inside the car.

In addition, a witness called (Wit1) described the driver at the accident scene.
Three persons were involved in a drug case number 6544x three weeks ago. This case is considered to be a class C drug activity, and there are two persons who have already been arrested. These people are mentioned as: Sus1 and Sus2. They only know the third person by his photograph because he uses a nickname. According to the witness’ statement, the investigator thought that the suspicious car was being driven by the third person. Therefore, it is necessary for the driver to be arrested and sent to the counter-drug section, to give his statement about the accusation.

Consequently, the digital evidence was sent to the laboratory to be examined along with a letter containing these details: the car plate number, owner’s name, case number in addition to MD5 values\(^1\) of the evidence and their size. The investigators wait to provide them with any information that can supply further knowledge to the case or to the suspect’s relationship to the other citation.

Because the two accused as well as the car owner have already given their DNA and the results have been saved in the database at the forensic laboratory it was possible to show that the DNA analysis of the two little locks of hair were similar but did not match with these persons. The time was very critical after the investigation received a threatening letter.

5.1 Computer Forensic Process (Using traditional Framework)

After the examiner had taken a bit-by-bit image, he began to analyze the file format of the evidence and to recover deleted files for laptop data. The size of its hard drive was 120GB and real size was 81GB. Then, the examiners created all entities based on the provided information in the case and reported these keywords i.e. Sus1, Sus2, Wit1, the driver’s nickname and the plate number.

The examiner used a known tool to extract these entities. He created these as keywords manually. Creating keywords manually, of course, requires time, and the creation of wrong words is common. After he had run the search, thousands of hits were matched and presented in the examiners computer and had to be examined carefully. Checking these entities took around three days.

5.2 Computer Forensic Process (by Proposed System)

Once the examiner had completed the imaging, during which he analyzed file format and recovered deleted files, the data was sorted and delivered to his assistant. His assistant transformed the data into text file format by an ad hoc tool and then saved these files in the digital evidence database as unstructured plain text data. Interview investigation with suspects was started immediately. All questions and answers were stored in the investigation database in short order.

The system (proposed system) is running dynamically. Therefore, it started extracting whole entities and making the integration. The system presented the matches with some supporting information concerning its relevance. The text analyst reviewed the results, made comments and then sent them to the investigator as new questions to be put to the suspects. Nearly, all human names are recognized and all strange words are determined.

5.3 Observation

Although the technical methodologies in both frameworks were similar (i.e. thousands of hits were produced in both methods), the time and way of processing are different. In the normal framework, the examiner did not know what was significant for investigation. Examiners were under stress after the investigators asked the laboratory supervisor to speed up the results because they had received a letter threatening that a man involved with drugs may be planning an imminent attack. Therefore, the examiners decided to send a report every hour. This report included hundreds of pages written in MS-Word. Both the examiners and investigators could not deal with this situation because system integration was missing. Therefore, a number of examiners and investigators had to combine to work as one team. This operation took more than 72 hours of hard work but within 48 hours of receiving the threatening letter a drug fighter was killed. After 62 hours of hard work, this team could integrate a

\(^{1}\) MD5 is a hash function used in a wide variety of forensic applications to check the integrity of files.
deleted Doc. file containing the full name of the driver.

By using the proposed framework, the process runs smoothly, i.e. there is no need to search for people’s names, the system extracts them automatically; there is no need to create keywords to start the search, and the search is working dynamically. Therefore, the proposed system is highly recommended to deal with these types of crimes.

6. RELATED WORK

Despite there being several digital forensic frameworks that have been proposed and have shown the importance of analysis in computer forensics, to date, there are no appropriate results covering the integration between the technician and normal investigators in physical crime. For instance, although Casey and Palmer model [7] gives full details to follow, they did not give a demonstration for their proposed steps. DFRWS[4] report does not explain the model steps in full detail particularly the analysis step. Reith, et al model [8] is considered as an extension to the DFRWS model and the approach is a linear process. However each phase cannot send a feedback to the previous phase; and that is inappropriate for physical crime investigation when involved with digital evidence. Finally, Carrier and Spafford model [9] uses the physical crime process as a successful methodology investigation, however, the results should be passed through 17 phases, which might lead to a time delay. Also, they did not mention how the analysis is going to be accomplished. Particularly, these models do not mention the integration of interview investigation or interrogation with their steps regarding digital evidence. In addition, by reviewing the literature, little data was found on the association between interview investigation and digital evidence. Therefore, this study sets out with the aim of assessing the importance of integration between the data from interrogation and digital evidence.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>... Identification</td>
<td>Identification</td>
<td>Identification</td>
</tr>
<tr>
<td>Preservation, recovery</td>
<td>Preservation</td>
<td>Preparation</td>
</tr>
<tr>
<td>Harvesting, reduction</td>
<td>Collection</td>
<td>Approach Strategy</td>
</tr>
<tr>
<td>Organization and search</td>
<td>Examination</td>
<td>Preservation</td>
</tr>
<tr>
<td>Analysis</td>
<td>Analysis</td>
<td>Collection</td>
</tr>
<tr>
<td>Reporting</td>
<td>Presentation</td>
<td>Examination</td>
</tr>
<tr>
<td>Persuasion &amp; testimony</td>
<td>Decision</td>
<td>Analysis</td>
</tr>
</tbody>
</table>

Table 1 phases of principle models

7. DISCUSSION

This (on-going) research proposes an automated framework for the integration of the information obtained in the non-digital phase of investigation with the analysis of digital evidence with the aim of improving the quality of digital evidence analysis and interview investigation process.

It uses the advantages of these models[4, 7, 8] as a valid base to build the framework, particularly in the crime scene phase and the examination phase. While these two initial phases are similar in digital crime and physical crime, the investigation hypothesis and the requirement evidence are different. For instance, the analysis phase in this framework has been specified for mining the data to extract entities (e.g. person names, location, organizations) text clustering and linking entities. Thus, the integration
The phase is central.
The contribution of this framework can be used for extracting knowledge in different forms e.g. generating questions, presenting entities visually by applying ontology, give reasoning about events by applying association rules, analyze psychological characteristics and motive using cognitive behavior or the creation of automatic keywords based on crime domain. Thus, it will improve the results of Casey model [7] DFRWS Model [4] and Reith, et al. [8] if we add the integration phase (of our framework) instead of the analysis phase in these models.

8. CONCLUSION AND FUTURE WORK
Computers have been involved with crimes on two sides, as a communication tool in digital crime or as an instrument tool (data storage and organizer) in physical crime. Investigation in both of these types is complex. Although there are no major differences in the examination of computers in both of them, the investigation length, investigation methodology, and analysis type are variant. This paper proposes a high automatic framework to integrate the conclusions of the field physical investigators and content analyst of digital evidence and improving the coordination between these two investigatory teams.

The introduced integration model makes the investigation process straightforward. Simultaneously, it means that several cases cannot be handled easily. One such case arises when the file formats are not supported to transformation or data contains text files for two human languages. Another challenge is when we want the system to decide the relationship between words as human names; it requires pre-processing to compare them with other databases such as dictionary names using AI. All these tasks require deep analysis as well as testing on real data. The research of these requirements and other possible improvements of the model presented in this paper are left for future work.

9. REFERENCES
Organizational Handling of Digital Evidence

Sheona Anne Hoolachan
HATII
George Service House
11 University Gardens
Glasgow, Scotland, UK
G12 8QH
pse57812@gmail.com

William Bradley Glisson
Room 402B HATII
George Service House
11 University Gardens
Glasgow, Scotland, UK
G12 8QH
B.Glisson@hatii.arts.gla.ac.uk

ABSTRACT

There are a number of factors that impact a digital forensics investigation. These factors include: the digital media in question, implemented processes and methodologies, the legal aspects, and the individuals involved in the investigation. This paper presents the initial idea that Digital Forensic Practice (DFP) recommendations can potentially improve how organizations handle digital evidence. The recommendations are derived from an in-depth survey conducted with practitioners in both commercial organizations and law enforcement along with supporting literature. The recommendations presented in this paper can be used to assess an organization’s existing digital forensics practices and a guide to Digital Forensics Improvement Initiatives.

Keywords: first responder; forensic readiness; organizations; procedures; digital evidence; Digital Forensic Practice.

1. INTRODUCTION

The digital revolution has profoundly affected how both private and law enforcement organizations handle digital evidence. With the increase of globalization and the advancing nature of technology, criminals are targeting digital media as well as using them as a tool to conduct crimes [2]. Organizations underestimate how often they will be subject to one of these attacks and how often they will have to produce reliable evidence to present in court [14]. Therefore, it is becoming increasingly important to identify potential weaknesses and rectify them [14]. Hence, digital forensic practices and the handling of digital evidence is an issue which is pertinent to many organizations throughout the world. Within the course of a year, the average organization will encounter events such as unauthorized access to the computer system, computer based attacks (e.g. phishing or denial of service), or online defamation [5]. In a study conducted by the Home Office in 2004 [8], organizations identified 101 different types of criminal threats administered by 137 different technological methods. The participants in the Home Office research study included Information Technology Security, Law Enforcement, Academic and Governmental organizations with a range of experience over a variety of security oriented topics [8].

With the exponential growth in technological advances and their utilization for criminal end, one might expect these figures to be higher in 2009. The criminal threat categories delineated in the Home Office study demonstrates the array of knowledge necessary to investigate potential violations. An act
such as fraud would be considered an offence under the Fraud Act 2006 (England) and could, if successfully prosecuted, lead to 10 years imprisonment [4]. Should an offence like this occur during the course of business and the discoverer (herein after referred to as the ‘first responder’) is untrained in the handling of digital materials, vital evidence could be lost or compromised, and a prosecution may not be secured. Incidences of money laundering and child pornography can be the most problematic for the organizations, as untrained first responders can inadvertently taint the evidence. These two particular crimes are the only ones which organizations are obligated, in the United Kingdom (UK), to report to the police [2].

Ueli Maurer discusses, in relation to Digital Rights Management (DRM), the importance of technical, legal, social and business aspects when managing information and information technology systems [7]. Highlighting these four criteria is useful for the examination of digital evidence. During the aforementioned discovery scenario, the relationship between the technology and the first responder is inseparable and, therefore, interdependent. The problematic technological discovery and subsequent evidence recovery is solely reliant on the first responder’s action and the complexity of the technological discovery will affect who has the skills to correctly handle the situation. Implementation of protocols must take into account the culture of the organization, policies and laws applicable to the type of information handled, as well as the size and nature of the organization.

This paper summarizes the results of an initial UK Organizational survey that examines how institutions handle digital evidence when it is discovered by untrained first responders. The results of the survey established the initial justification for Digital Forensic Practice (DFP) recommendations. These recommendations are based on the small sample of individuals interviewed for this research and cannot encompass every type of organisation. The purpose of this study is to initiate discussion surrounding issues that arise within organisations where there is currently no standardised set of protocols in place to protect the integrity of digital evidence.

2. EXISTING RELEVANT WORK

There are three major areas that should be addressed when considering the significance of digital evidence becoming tainted by an untrained first responder: legal aspects including admissibility; the forensic readiness of the organization; and the first responder themselves. This paper will not concentrate on the legal aspects as these are jurisdictional and will vary both nationally and internationally. The latter problems, however, will be briefly discussed in order to contextualise the research and its value.

2.1 Forensic Readiness

Forensic readiness is a concept outlined by Robert Rowlingson from QinetiQ Ltd, an internationally recognised defence technology and security consultancy [15]. From the perspective of law enforcement agencies the forensic process begins when the crime has been committed or it has been discovered and reported. The concept of forensic readiness, according to Rowlingson, is that an organization can pre-empt the occurrence of a crime by gathering evidence in advance and in doing this, organizations will benefit not only in instances where prosecution becomes an issue but also in limiting their own business risks [13]. Although this definition of the forensic readiness strategy acts as the primary thesis of Rowlingson’s paper, he addresses other matters of benefit and contention. These are identified below and in Section 2.2.

One of the main strengths of the forensic readiness model is the recognition of the range of personnel within an organization who can become involved in a legal inquiry and this is something that will be critiqued in Section 2.1.3. Rowlingson identifies no less than eleven different departments and their associated personnel that must be considered in an investigation.

Although the variety of staff involved will vary depending upon the magnitude of the investigation, this observation identifies the crux of this paper’s argument – there are a multitude of people who need to understand the correct protocol within a digital investigation.
2.2 First Responder: Policies

Forensic readiness may be implemented as an organization, but when evidence is discovered the strength or weakness of that evidence is reliant on the first responder and their actions [6]. The United States Department of Justice (USDoJ), in their first responder’s guide, defines a first responder, in a digital evidence context, as:

- Anyone encountering a crime scene that might contain electronic evidence;
- Anyone processing a crime scene that involves electronic evidence;
- Anyone supervising someone who processes such a crime scene;
- Anyone managing an organization that processes such a crime scene [9].

This means that there is a desire to be forensically ready within the organization and in relation to evidence gathering, but also there is a need for forensic readiness amongst all members of staff.

Many organizations implement incident readiness through company policies. However, given the complexity of digital forensics, the expectation of staff members to understand both technical and legal principles is unreasonable based on other areas of expertise that the employee may possess. Reliance on policies is the downfall of Rowlingson’s forensic readiness implementation as a high level of expectation of technical knowledge is forced upon the average user. Relying on this knowledge for the purposes of digital security, particularly in the context of evidence collection, is completely unacceptable. A notable example of unreasonable expectation within current digital security literature is David Harley’s article focusing on the problem of phishing attacks. Harley makes a number of strong remarks regarding what comprises phishing, the pitfalls of the average user and methods by which phishing scams can be educated against and defeated. From the perspective of this paper, however, the major downfall of his proposed solutions comes from the relationship between technological indicators and the untrained end user, or first responder in the context of this research. He expects the average user to identify security cues such as accessing the HTML source code to verify the legitimacy of the senders email address; Domain Name Service (DNS) misdirection; and identification of deceptive embedded Universal Resource Locators (URL’s) – activities many users would be unable to undertake regardless of their desire to act more securely [3]. Harley’s paper, by no means, disregards less-technically capable individuals, as he proposes quizzes to simplify the issues. His paper does, however, act as a useful baseline by which “simple” technological cues can be measured and, when contextualised to an untrained first responder, it becomes apparent why this dichotomy can become problematic for the field of digital forensics.

Regardless of the technical knowledge requirements outlined, policies are problematic at a very basic level. The complexity of jargon and the expectation to read and understand complicated ideas, concepts and rules leads to the insecure behavior demonstrated when policies are presented to many employees – they sign the documents without reading them [16]. Julie Nosworthy identifies some of the key downfalls of a policy culture including the tension between technological solutions and the human factor [10]. She identifies the failings previously indicated with regards to a lack of understanding of external departmental issues, specifically in relation to Human Resources (HR) and Information Security (IS) [10].

2.3 First Responders: Automated Tools

In order to fully explore the various tools currently utilised within the realm of first response, it is important to address automated solutions as well as more human-driven techniques previously discussed, such as policies. These will not play a particular role within the Digital Forensic Practices recommended in Section 4.0, but are beneficial for background understanding of the problems and solutions. FRED is a tool used by the American Air Force Office of Special Investigations and its objective is to handle digital evidence without the intervention of any users [6]. The goal of the actual
FRED tool is to determine whether or not an intrusion had occurred by means of analysis of network connections, active processes, dynamic link libraries (DLL) and open ports with MD5 hash values recorded for critical files on the system [6]. Furthermore, the entire tool fits onto a CDROM and the disk would play host to a batch file created by the tool to create an audit trail [6].

Ideally, an automated tool would remove responsibility from the first responder and allow unified evidence collection regardless of their specific knowledge. It could be argued that on these grounds the aforementioned criticism of policies and their unreasonable expectations placed upon the user are circumvented. However, an automated tool does not remove the human factors from the acquisition process and therefore there is still scope for data to become compromised. An individual will be required to insert the disk and execute the tool. How can it be determined what they did to the computer before the auditing tool began its work? Depending on the size of the system the tool may take some time to complete its functions. What happens if the first responder leaves the room while the tools is executing? There is scope for a suspect or bystander to compromise the tool while it is in operation and issues surrounding the field of live forensics which are outwith the scope of this paper. Some of these issues are addressed in the conclusion on Kornblum’s article where it is indicated that:

“There have been some procedural questions regarding when FRED should be run and by whom, but these are issues of policy, not technology” [6].

This harks back to Nosworthy’s analysis of the problematic areas of policy making where a ‘them and us’ culture exists [10]. The technological implementations in this scenario cannot exist without the human factor and it is, therefore, unacceptable to blame any shortcomings on the individuals who are required to make the whole system operate correctly.

3. SURVEY COMPILATION

The study attempts to investigate whether or not the individuals in charge of implementing policies and dealing with the implications of digital evidence handling have a sufficient understanding of the issues and the practices within the field. The following sub-sections present the experimental approach and demographic information.

3.1 Methodology

The style for this case study was a combination of both structured and semi-structured interviews. Structured interviews have a specific set of questions which are presented to each interviewee and, upon answering the questions, the interviewer moves to the next question or set of questions. Semi-structured interviews begin with a broad set of themes to be discussed and the responses from the interviewee will lead the direction of the interview and generate new lines of questioning [12]. The case study questions were pre-defined and were split into two main thematic groups: policies and general issues, together with technical issues for the organizational questions; organizational evidence and technical issues for the law enforcement questions. The same sets of questions were asked of each individual interviewee. Where their responses afforded the opportunity, further questioning along these lines was followed to gain as much information as possible.

The organizational survey questions were initially validated by an individual who not only possesses corporate environment work experience but has also undertaken research projects at the University level. An ex-senior police officer was approached to validate the law enforcement questions in a similar manner. Both of these validations led to minor adjustments being made to the questions to clarify semantic ambiguities. For example, prior to the validation process, one question read: “How is this communicated to all staff?” This was changed to “How is this communicated to all staff within the organization?” in an attempt to mitigate ambiguity for the interviewee.

Prior to distributing the questions and in addition to the initial validation procedure, individuals who were not due to participate, but yet had relevant background knowledge, were asked to validate the questions to ensure they were easily understood by the target audience. This was done at the
recruitment stage for the organization questions. After ascertaining that an individual was not appropriate for the research project due to their lack of daily contact with the subject matter, they were asked to explain their understanding of each question to ensure that their understanding matched the interview intention.

### 3.1.1 Interview Demographics

A total of ten interviewees were recruited based initially on two broad areas concerned with digital evidence, i.e., organizations and law enforcement. Within these categories a balance was sought between public/government organizations, corporate organizations and educational institutions. A balance was also sought between the organizational sides and the police, lawyer and ‘other’ (such as private investigator) for the law enforcement side. Table 1 provides a detailed insight into the survey participant’s organizations, years of experience and job title.

<table>
<thead>
<tr>
<th>Interviewee Key</th>
<th>Type of Organization</th>
<th>Years Experience</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Academic</td>
<td>Undisclosed</td>
<td>Information Security Coordinator</td>
</tr>
<tr>
<td>HB1</td>
<td>Health Board</td>
<td>24</td>
<td>Head of Information Governance</td>
</tr>
<tr>
<td>HB2</td>
<td>Health Board</td>
<td>11</td>
<td>Information Governance Technical Advisor</td>
</tr>
<tr>
<td>F1</td>
<td>Financial</td>
<td>5</td>
<td>Security and Technical Consultant</td>
</tr>
<tr>
<td>F2</td>
<td>Financial</td>
<td>8</td>
<td>Internal Audit Manager</td>
</tr>
<tr>
<td>F3</td>
<td>Financial</td>
<td>18</td>
<td>Divisional Records Officer</td>
</tr>
<tr>
<td>F4</td>
<td>Financial</td>
<td>4</td>
<td>Information Security Manager</td>
</tr>
<tr>
<td>CS1</td>
<td>Civil Servant</td>
<td>10</td>
<td>Procurator Fiscal</td>
</tr>
<tr>
<td>DE1</td>
<td>Digital Examiner</td>
<td>4</td>
<td>Digital Analyst for a Private Company</td>
</tr>
</tbody>
</table>

The intention behind seeking the balance among the various categories was to gain an objective understanding of issues addressed by organizations and those which are distinctive to law enforcement agencies in order to identify areas of correlation and anomalies. Due to the separation between organizations and law enforcement interviewees, two different sets of interview questions were posed to participants depending on their respective associations. The individual questions are available upon request.

### 3.2 Results

The survey findings identified the following four areas as areas that need to be discussed:

- Problems identified by the organizations surrounding digital security
- Current first responder practices
- Issues surrounding communication and training
- Variables identified by the organizations that act as unique difficulties or points of note.

#### 3.2.1 Problems identified by the organizations surrounding digital security

When interviewees were asked to identify the type of attacks, crimes or general issues which are problematic within their organization, the following were acknowledged:

- Theft of physical objects including hardware
- Loss or the need for encryption of USB keys
- Implementing security measures that only addressed external attacks rather than
considering the internal threats

- First response errors
- Warning colleagues of digital device seizure or search
- Unauthorized access to data
- Emails including spam and viruses
- Phishing attacks on customers and staff
- Making copies of child pornography for the purpose of examinations
- Malicious malware
- Personal use of the organization’s system (e.g. via photographs and music)
- Identity theft
- Human error
- Password complacency.

Although there were overlaps in the problems identified by the interviewees, this list demonstrates that, within a relatively small interviewee group, a vast range of different problems encountered has been identified. This supports the aforementioned critique of automated first response tools - how can an automated tool be expected to handle such a range of subjective issues and cope with human error? Some of these problems can be automatically handled, such as malicious malware or email problems, but many of the issues facing digital security exist in a non-digital medium. For example, HB2 identified unauthorized access of digital patient records to be a problem within their health board. He commented that people access data that they are not supposed to access because “people are nosey”. Although access control software could help prevent this problem, determining what is an act of curiosity and what is a genuine access of confidential records is a subjective issue which cannot be determined by automation. HB1 reinforced this point, saying that:

“You can put as many procedures and policies in place and you can put as many technical measures in place but there is this big grey area in the middle with people in it and they’ll always let you down.”

In terms of policies, interviewees often agreed with the critique outlined in Section 2.2 First Responder: Policies regarding the unreasonable expectation of individuals to understand technical problems. It was, however, noted by F3 and F1 that often leaving the level of education at a basic awareness level is safer for the company’s interest. F3 commented that:

“Personally, I feel that staff don’t need to know the details of digital security, they need to know it at a high level and the importance of it but we would tend to leave that to subject matter experts who are dealing with it on a daily basis…There is a risk that if too many people know too much detail there is potential for something to happen.”

F1 concurred saying “the less they know the better for us”. There is a conflict existing here because much of the literature suggests that the only way for individuals to take digital security seriously is to fully understand why they are taking certain actions rather than a different set of actions [10]. F1 and F3 acknowledge that it is often safer to restrict the amount of knowledge staff have regarding digital security, but there is a limit to the amount of information that can be discretionary and a point in which it will restrict the awareness of issues.

This is a notable conflict of interests between those implementing the security and the end-users and
there were other such incompatibilities identified through the interviews which will be discussed in Section 3.2.3 Issues surrounding communication and training.

3.2.2 Current first responder practices
One of the most conspicuous blanket practices amongst all of the organizations was how to deal with a first responder discovery of evidence on the organization’s system. Every individual in the organization category of interviewees details a policy involving a phone call to another department should an end-user discover something that could potentially be used as evidence. The interesting aspect arises from the fact that each organization suggested contacting a different person. Even individuals within the same organization, in different roles, identified different departments that should be contacted. These included:

- Local I.T. support
- Legal
- Help-lines
- Financial Crime
- Line Manager.

F3 and F2 are employed by the same financial institution within different areas of the team. F3 identified the legal department as the point of contact whereas F2 suggested the financial crime department. This anomaly could have arisen through their differing roles and within their specific role or the type of information they frequently handle, then these departments may be the most suitable. The most striking aspect about the response to this question, across the board, is that there is no definitive standard regarding what end-users should do upon discovery of potential evidence. However, all the organizations provided their staff access to policies regarding general digital security via their intranet or equivalent. The potential problem with this system is addressed in Section 4.1 Do not use the intranet for policies regarding the handling of digital evidence.

The impact of not using the intranet for policy dissemination, from the law enforcement perspective, varied between DE1 and CS1. DE1 identified instances where he had been called to an organization and upon arrival the requesting member of staff was still on the computer and eager to show him information they had found. He explained to the individual that:

“her activity on the computer was going to result inevitably in the loss of recoverable information that could have been useful…from there onwards you’re not looking at the computer that the suspect used…all I was looking at was how her boss was using the computer”

He did, however, explain that this was a small company and perhaps this could account for the method in which it was reported and handled. CS1 explained that the difference in resources and capabilities between large and small organizations often results in different methods of handling and reporting. This, and similar incidents are often reported to a private investigator, perhaps not exclusively but in part, because of the size of the company as smaller organizations are more likely to circumvent any of their own departments and hand the information directly to the police or investigators. Larger companies, CS1 continues, would just sack the individual to avoid reputational damage. On the whole, however, CS1 did not feel the reliability of the evidence would be problematic for admissibility because if the first responder did alter any information forensic analysis would be able to identify and account for any of these changes.

3.2.3 Issues surrounding communication and training
Training was widely identified as a problem area within this study. Two of the most candid interviewees were F4 and HB1 and their perspectives on the subject of training were very interesting.
F4 identified many weaknesses in the field of digital security within his organization and, as he is relatively new to the organization, he has been tasked with rectifying these issues. The difficulty, from his perspective, is that the organization has fallen far behind in its process of reviewing company policies and this is the first review conducted since the policies were written seven years ago; creating an immensely difficult task. The only communication, at present, of policies and digital security training is a Computer Based Training (CBT) scheme which forms part of employees’ initial induction – there is no follow up training.

Furthermore, although policies and procedures can be accessed electronically after this initial induction, most of the staff, according to F4, do not know this capability exists and subsequently do not seek access to them. He explains that there are hard-copies of the policies distributed to each manager, but “I suspect most of them end up in a cupboard”.

HB1 identified an ongoing battle with staff regarding information security training. She highlighted that although the organization as a whole has agreed to annual information governance training, individual staff have not taken this up and, as mentioned in 4.1, the staff must understand the issues in order to heighten their awareness. This breakdown in communication will be discussed further later in the paper. HB1 explained that their health board also has a CBT scheme in place for training which has become outdated to the point of redundancy. She explained that the boost in their training numbers can be credited to the Personal Development Programs (PDPs) in their health board whereby staff must meet certain personal targets for training amongst other aspects of their role in order to progress up their pay-band. This is the theory of behavioral economics where an individual’s internal desire to achieve something is driven by monetary rewards [1]. Although this theory is criticized for driving individuals by non-altruistic goals, HB1 concedes that she is “quite happy to use this system if it means training improves”.

Beyond the scope of training, another recurrent theme about communication was salesmanship. HB1, A1 and F4 all discussed the notion that, as a department, you have to sell your ideals to other departments in order for them to recognize your cause. HB1 and A1 both described instances where they would give lunchtime seminars or take part in other departmental meetings in order to highlight the importance of digital security. F4 discussed it in a much more business sense:

“At the moment, I don’t think we are selling the benefits well enough and, therefore, it gets ignored. But it’s down to us to sell the benefits because we can’t just expect to say ‘you should do it. It’s a good idea’ and not say well hang on, this is what you’re going to get for your money. Everything comes down to money and we have to justify our existence.”

This pragmatic attitude was somewhat refreshing regarding the literature around this subject, such as the example of Harley’s phishing article, outlined in section 2.2, which seemed to expect something from users simply because he tells them to expect something [3]. As identified, Harley’s approach does not take into account different skill levels of users whereas F4 wants to make people understand why digital security is a pertinent issue rather than simply imposing his views upon them. It could be argued that although the literature is often too idealized in its suggestions of best practices, the implementation can sometimes be more pragmatic.

3.2.4 Variables identified by the organizations that act as unique difficulties or points of note
At the end of the formal interview questions, interviewees were asked to raise any issues that they felt were unique to their organization or that were prevalent in the current climate that had not yet been raised. Although the issues raised during this section were not directly related to first response discovery or the underlying training issues, very diverse problems were identified.

Because F3, F2 and F1 were all from financial institutions, their responses were similar acknowledging that problems are fairly uniform throughout the financial world. These included the lack of mandatory training and education as well as risk assessing the organization from a variety of levels and vantage points. Interestingly, F1 identified the organized crime aspect to be a large problem
facing all types of organizations.

HB1 and HB2 both identified patient care as the unique aspect of their health board over other types of organizations. HB1 highlighted that:

“Everything we do will potentially affect the patients and if someone isn’t doing what they’re meant to electronically, they could kill somebody.”

HB2 elaborated upon this point by highlighting the reputational aspect of their health board versus organizations. He indicates that one of the worst outcomes for a corporate organization is reputational damage, whereas within a health care situation there are lives at stake. Furthermore, their health board’s reputation is arguably more valuable than an organization because for many people there is no alternative. He notes that if a company’s reputation is damaged, the customers have the option of moving elsewhere. Other than private medical care, there is no alternative to their health board. HB1 concurs, explaining that within her specific health board “we have to be whiter than white”.

F4 states that the problematic area for his particular organization is their relative infancy within the financial sector. He explains that the organization is only beginning to push branding of the company and his concerns lie in the raised profile of the company outweighing the security measures to prevent attacks:

“If we raise our profile as a brand someone will start looking at us and saying well what’s going on here and what can we get from these people?”

This was not raised specifically as an issue, but perhaps his concerns are more relevant given the current financial climate and the desperation of people to have enough money to survive. He goes on to say that it would have been nice if both brand profile and brand protection had developed in parallel rather than one being required to catch up.

A1’s unique aspect of the academic organization stemmed from the diverse nature of the staff, students and their requirements. Web filtering, for example, is a limited security measure within the University as there is no way to predict what topical areas students and staff will be researching at any particular time. He comments on the financial ties relating to this situation. Often, research grants will be awarded only if the researcher can provide proof that the institution is equipped to allow the research to be conducted. This may involve reducing security measures for particular parts of the University or taking specific computers off the network. He also comments that mandating anything within the University, such as training and education is extremely difficult as academia takes priority over all else.

This lack of standardization and mandating is prevalent within the law enforcement side also, according to DE1. He comments that varied approaches to the handling of digital evidence by different police forces and organizations can make it difficult to work as an independent outsourcing company as different clientele will require information to be presented in a variety of formats. This is definitely a recurring theme within the field of digital security regardless of whether one is implementing it or handling the aftermath.

Although not a unique aspect of his position, CS1 explains that he has recently been told about a police force in Holland that is training all their Criminal Investigation Department (CID) police to handle digital evidence in a forensically sound manner. According to CS1, the problems surrounding digital devices has become so prevalent and widespread that they are ensuring that all their detectives have a basic understanding of how to handle the devices and will pass on particularly difficult cases to experts. He described it as ‘triaging’ the computers and this was also a phrase used by HB1 in a discussion of outsourcing versus handling information internally.

### 4. DIGITAL FORENSIC PRACTICE (DFP) RECOMMENDATIONS

An analysis of the results generated from the interview process, as well as studying current literature,
provided three recommendations that organizations should consider when dealing with digital evidence. Beyond these three recommendations there are too many variations within organizational structure and the unique problems faced by different types of organizations as demonstrated in section 3.2.4 for formal restrictions to be productive. These recommendations provide a solid basis for evidence to be handled at the discovery stage in order to assist the preservation of evidence integrity. The three recommendations are as follows:

1. Do not use the intranet for policies regarding the handling of digital evidence.
2. Have a centralized co-ordination point so staff members are clear on who should be contacted.
3. Use an external company to perform forensic analysis but have internal ‘triaging’ capabilities.

4.1 Do not use the intranet for policies regarding the handling of digital evidence

Many of the organizations used the intranet as a communication point for staff should they need to refer back to policies at a post-training stage, as mentioned in Section 3.2.2 Current first responder practices. Although this approach, for the most part, is a sensible measure to ensure the ability staff to be able to access policies at all times. For the purpose of handling digital evidence, it is unacceptable. After the point of evidence discovery, the staff member should be able to access the information required without asking anyone or touching the suspect computer. In terms of asking anyone, this is not recommended as it may not be clear who has perpetrated the crime and giving warning to individuals within the office may compromise the evidence. F3 highlighted an incident that had occurred involving a laptop whereby the individual owning the laptop had been warned ahead of time that it would be investigated. This allowed the individual to delete information from it and although this was subsequently recovered the potential for affecting the admissibility existed. Although interaction with other individuals could be counterproductive within an investigation, the primary concern would be those in the immediate vicinity of discovery and, therefore, providing staff with a centralized “help” facility would allow explanation of protocol, within minimised risk, to the digital evidence. This is discussed in Section 4.2. A hard copy of policies should also be accessible to all staff members as supplementary, supporting and verification materials, as initial training can be easily forgotten after the fact. These facilities should be available to every employee regardless of whether an incident is likely to occur within their job as anyone can be a first responder to an incident.

4.2 Have a centralized co-ordination point so staff members are clear on who should be contacted

Although this may incur cost to the company, Section 3.2.2 Current first responder practices demonstrated the range of responses by interviewees when asked who a first responder should contact in instances of evidence discovery. Having a centralized point acting as a call centre for coordination will assist in the implementation of the first recommendation.

All that would be required is a card or sticker attached to every terminal or laptop in the company and first responders would have access to the appropriate information rather than having to consult the intranet.

4.3 Use an external company to perform forensic analysis but have internal ‘triaging’ capabilities

F3 and HB1 noted that it is sensible to handle forensic analysis externally to avoid any potential corruption or to prevent people from not reporting issues to protect friends or colleagues. This is a very sensible idea but, in line with CSI1’s information about the Dutch police, internal ‘triaging’ capabilities are essential to an organization. Although those interviewees who noted outsourcing as a company policy said they used local companies to ensure they were on the scene quickly, there will still be a time lapse between discovery and the arrival of an outsourcing team. Having staff members on site who can secure and handle the device at this interim point would prove invaluable to an organization when issues of admissibility are raised. Larger organizations may have dedicated
departments that deal with digital investigations; however, a contingency of outsourcing would be beneficial should an incident arise where the investigative department itself requires examination. Arguably, this recommendation could be split into two issues. The first addresses the use of external forensic companies and the second has the responsibility of internal ‘triaging’. However, given that there will be a natural time-lapse between discovery, reporting the incident and the external company arriving on the scene. Having the plan for handling the evidence between these time-lapses is fundamental in ensuring the continuity of evidence. Based on the outcome of this research, ‘triaging’ capabilities would be recommended for every organization. Although interviewees indicated that their external contractors would endeavor to be on-scene at the earlier point, there are limitations as to how quickly this would occur and the data has the potential to change within this time.

5. CONCLUSION

The continued integration of the digital data into business environments highlights the need for organizations and law enforcement to be able to handle digital evidence from a first responder perspective. A lack of digital forensics practices and procedures cultivates an environment that is conducive to compromising evidence in organizations. This research attempts to identify areas within organizations where the provisions for first responder evidence handling are not at an acceptable level to guarantee the admissibility of items of evidentiary value.

The conclusions derived from the survey resulted in the creation of DFP recommendations. The implementation of these recommendations should assist in the mitigation of inadmissible evidence. The DFP recommendations are as follows:

1. Do not use the intranet for policies regarding the handling of digital evidence.
2. Have a centralized co-ordination point so staff members are clear on who should be contacted.
3. Use (or plan for) external companies to perform forensic analysis but have internal ‘triaging’ capabilities.

Future research should investigate higher level staff members who are responsible for dictating the culture of the organization and setting the budgets which ultimately shape the implemented digital forensic strategies.

Furthermore, it would be interesting to broaden the research scope to include private companies. This is due to the limited financial scope of the current research which perhaps masks a fuller understanding of the bigger picture. It would also make a noteworthy addition to future research to consider businesses that are purely internet based to determine if they face the same problems. In addition, it would be beneficial to investigate companies that operate within a technological field rather than those simply using technology within the course of their business

Additional research would also benefit from an examination of the front-end or lower-level staff’s perspective to verify whether the assessment made, by higher level interviewees, of their ability to cope with discovering evidence was accurate. The scope of the research could also be expanded to incorporate the business issues involved in implementing security matters such as cost risk analysis and management issues.

6. REFERENCES

Social Networking: A Boon to Criminals

Tejashree D. Datar
Computer and Information Technology Department
401 North Grant Street, Knoy 255
Purdue University
West Lafayette IN 47907-2021
tdatar@purdue.edu

Richard Mislan
Computer and Information Technology Department
401 North Grant Street, Knoy 255
Purdue University
West Lafayette IN 47907-2021
rick@purdue.edu

ABSTRACT
With the world getting more and more digitized, social networking has also found a place in the cyber world. These social networking sites (SNSs) which enable people to socialize, and build and maintain relationships are attracting attention of all kinds of people such as teens, adults, sports persons, and even businesses. But these SNSs are also getting unwanted attention from people like sexual predators, spammers, and people involved in criminal and illegal activities. This paper talks about SNSs and how these sites are exploited for criminal or illegal activity. The SNSs are discussed in detail with respect to user profiles, user networks, and privacy and security with respect to these user accounts. The paper also talks about the way available data on these SNSs can be exploited. The paper concludes with a few real life recent criminal cases associated with these SNSs.

Keywords: social networking, social networking sites, Facebook, MySpace, online predators, phishing, social networking crime, social networking models

1. INTRODUCTION
With the advent of the Internet, it is now very easy to be connected to a number of people, groups, and communities which was not this easily possible before the Internet was widely available. The Internet gave rise to online social networking which is mostly done via the use of social networking sites (SNSs) such as Facebook, Twitter, MySpace, Friendster to name a few among the many available SNSs. Today, online social networking has become such a huge phenomenon, that Twitter was declared the most popular English word of 2009 (Parr, 2009). Facebook, one of the social networking sites, ranks third in the overall web traffic in the United States with over 104.2 million users per month (Quantcast, 2009a). MySpace, another social networking site ranks tenth in the overall United States web based traffic with over 55.8 million users per month (Quantcast, 2009b). Online Social Networking has become so much part of our daily lives that it is not uncommon to keep all our contacts posted of what we are doing, if not every minute, but everyday of our life. Social Networking has become a powerful tool for businesses and other things like even the 2008 Presidential Election.

The basic purpose of these SNSs is online interaction and communication and maintaining relationships. SNSs have various models, but the most common model is to present a person’s profile and to visualize the person’s network of contacts to other people (Gross & Acquisti, 2005). These SNSs allow people to put all kinds of personal information on their website. When people join these social networking sites, they have to create their personal profile. This profile contains information such as name, which could be real or pseudo, date of birth, address, hometown, gender, ethnicity,
religion, spouse’s information, workplace details, school details, and the person’s personal interest. This profile could also include photographs, videos, and personal messages. Other members can connect by sending friend requests or messages. When a person is added to the contact list or the friend list, this person gets the privilege to access the friend’s profile and all the personal information put on this profile. These SNSs also have the privacy option wherein the profile user can hide all of the available personal information from other users except their friends in the friend list, that is, users who are directly connected to the profile owner. Even with all this available privacy, users are least bothered about their profile privacy and are happy to share all the personal information with the online world.

With such personal information as name, address, date of birth, gender, information on children, personal messages like updates, and photos made easily available by the users themselves, it is easy for criminals to gain access to this information. But as per these SNSs models, these criminals also get access to the directly connected contacts of the user. These criminals are commonly referred to as “Online Predators”. This paper focuses on three popular social networking sites in the United States, namely, Facebook, MySpace, and Twitter. The paper will describe social networking and its history, and then will describe the above-mentioned three sites in detail concerning the user profile, content, and privacy. It will then describe the possible use of these sites by online predators to conduct their criminal activities.

2. SOCIAL NETWORKING: EVOLUTION

Until the 1990s, Internet was not so widely and commercially available to the common public. As Internet started becoming more available and more popular, people started viewing it as a useful and commercial application. This was the evolution of online social networking. But does it mean that people did not socially connect to each other before the Internet boom? Human being, in itself is called a social animal. Social networks have been studied and analyzed for a long time now. This analysis of social networks is useful in studies of kinship structure, social mobility, science citations, contacts between members of deviant groups, corporate power, international trade exploitation, class structure, and many other areas (Scott, 1998). Internet was a revolution; similar to how telephone was a revolution. As social networking is nothing but maintaining relationships and building new relations, before the advent of the Internet, people used simple methods like snail mail, telegrams, telephones and even actually physically meeting each other to maintain and build new contacts.

In the 1980s and 1990s, a form of social networking called the Bulletin Board System or simply BBS was popular. Here people could send text messages and the BBS ran over the telephone lines (Gigaom, 2008). The first site that could be called as a social networking site came into being in 1997. This was the start of online social networking with SixDegrees.com coming into existence. Users were able to create profiles and list friends using SixDegrees. By 1998, users could also search for friends on SixDegrees. SixDegrees promoted itself as a tool where people could connect with each other and send messages to each other. But SixDegrees failed as a business and in 2000 was finally closed (Boyd & Ellison, 2007). In 2002, social networking sites finally started blossoming with the introduction of Friendster. MySpace was introduced in 2003, while Facebook was open to the general public in 2006. Twitter was also launched in 2006 (Nickson, 2009). Thus started a new age of social networking that we have now come to know.

3. USER PROFILE CONTENT OF THE SOCIAL NETWORKING SITE

As stated earlier, this paper concentrates on the three most popular social networking sites in the US, namely, Facebook, MySpace, and Twitter. An account was specifically created for the purpose of this paper on each of these sites. While Facebook and MySpace have more fields as compared to Twitter, all these sites ask for information like name, birth date, photograph, hometown to just name a few. Compared to Facebook and MySpace, Twitter has more concentration on “chat” for social networking. Facebook and MySpace are more oriented towards maintaining and building new relationships.
amount of personal information that could be put up while creating a user profile on Facebook and MySpace is astounding. Apart from the fields mentioned earlier, a user could put in information like gender, sexual orientation, relationship status, movie or music taste, biological data to name a few. Appendix 11.1 lists and compares all the available fields related to a user profile on all the three sites.

With all this data relating to personal information available on the Internet, privacy is now a huge concern. Most of the user profile fields on these sites have an option of visibility. This means that the user can decide if the specific content should be available to everyone on the network (here network means the entire SNS network) or just the user’s personal network of friends. Even with all this form of privacy available, users tend to keep their profiles open to everyone. This has created a huge security concern as crimes related to these SNSs started rising. These crimes could be anything from cyberstalking, social phishing to sexual assault. These sites are even referred to as “Predators Playground” (Schrobsdorff, 2007).

4. OPERATION OF THE SITES AS A SOCIAL NETWORK

Social networking works in the same way as computer networks. One user is directly connected to a number of users namely contacts or friends and these friends are in turn connected to other contacts. This forms a kind of web or mesh where users act as nodes and every node has multiple branches, which are the user’s contacts.

Since every user on these SNSs is unique, the amount of information put out by each user is different. The way these users behave online is hard to define, but this behavior generates out of trust. Fukuyama, and Lewis and Weigert in their respected papers (as cited in Dwyer, Hiltz, & Passerini, 2007) discuss that trust is a critical determinant in personal or face to face relationships. Similarly, Coppola, Hiltz, and Rotter, Jarvenpaa and Leidner, Meyerson, and Piccoli and Ives in their respective papers (as cited in Dwyer, Hiltz, & Passerini, 2007) discuss that trust is also important for successful online interactions. Metzger in her paper write that(as cited in Dwyer, Hiltz, & Passerini, 2007) trust is a precondition for disclosure in interpersonal exchange situations, because of the reduced perceived risks which are involved while revealing private information (Dwyer et al. 2007).

From the above arguments it can be said that relationships on these sites will not be built without trust. To build up a relationship, the user generally adds other users as their friends only if they know each other, even though it was a very brief interaction. The way these SNSs’ networks work, once a person is being added into the friend list, this person can access all the information of the user including the users other contacts. This way, it will not be very difficult for an online predator to gain trust of an individual by employing the briefest of interactions and once added to the friend/contact list, exploit this individual’s personal data and also maybe search for other potential victims through the now open medium of “Friend List”. Crime via social networking is increasing rapidly and criminals are now viewing these SNSs as a tool for committing crimes. If the user account is not open to everyone, the key point of the user information being available for exploitation lies in gaining trust and access to the user profile via Friend/Contact List.

5. FINDING THE USER INFORMATION FOR EXPLOITATION

There are many ways of finding the user information. The user itself can be found by doing a simple search in search engines like Google or Bing. There are certain privacy features available for users of the SNSs that allow the users to not be found via the search engines. This is called profile searchability. Online predators do not search explicitly for users this way. They prefer to contact potential targets as a user of the SNS. Wolak, Finkelhor, and Mitchell in their paper write that online predators prefer to meet and seduce their victims online. They also say that majority of the victims are aware that they are conversing with an adult (Wolak et al. 2008). In a social phishing study conducted by Jagatic, Johnson, Jakobsson, and Menczer, they found out that students readily give university information to a non-university party. They say that a phisher can mine information about relationships via social networking sites. For this study, Jagatic, Johnson, Jakobsson, and Menczer
used freely available user profile data from SNSs for the phishing attack. This data appeared to originate from a friend on the network. They found that the targets were much more likely to disclose personal information to friends than strangers (Jagatic et al. 2005).

So how much information is easily available? Acquisti and Gross in their study found that Facebook users have more trust in the Facebook privacy settings. These users are not much concerned about the information in itself as they think that they can control the information and the privacy controls as to who can view the data. They found out that users are also mildly concerned about who can access their personal data. Another interesting thing Acquisti and Gross found in their study was that users of Facebook, trusted the system and its members more than compared to MySpace (Acquisti & Gross, 2006).

With the users having this attitude towards privacy and trust, with respect to both the SNS and the users in the contact list, it is very easy for online predators to gain access to personal information of the users or for phishers to use phishing methods to collect personal data.

6. PRIVACY AND SECURITY

SNSs have a lot of privacy features. Users have control over who can search their profile called the profile searchability or who can view their profile called profile visibility. A user’s direct network of contacts has exclusive privilege of viewing all of the users content such as the message posts by the user and by user’s other contacts on the user’s profile. These direct contacts have access to all of the user’s photos, videos, list of the communities, friend/contact lists. These users however, cannot see the messages/mail communication between the user and the user’s other contacts.

The privacy issue arises when some user content is seen or accessed by unintended people. This occurs when “friends of friends” or secondary contacts can view the user’s content like photos, and videos. A user can be connected to thousands of secondary users or the friends of friends and this potentially increases the risk of personal information being available to users who are not even in the contact/friend list of the user. Acquisti and Gross in their paper write that an online social network lists hundreds of direct contacts/friends and include hundreds and thousands of additional contacts which are just three degrees of separation from the subject (Gross & Acquisti, 2005).

With these statistics it is very easy to cross reference a particular user via the open friends/contacts channel. If an online predator gained the trust of a teen and gets added to that teen’s contact/friend list, this opens a big window for this online predator to search for potential victims via this teen’s friend/contact list. This predator will also have access to the teens photos and from there access to any open profiles as well as photos, videos, and personal information of the teen’s other contacts, which essentially become the predator’s secondary contacts.

Phishers work in different way. They gain unauthorized access into a users account and start sending spam to the user’s direct contacts. These messages could be anything like the Nigerian Scam or appear to come from the user and ask to fill information on a third party network or could be even a virus which infects the machine if the link to it is clicked. SNSs are opening new doors for phishers and scammers. One can become a member of these SNSs very easily. Also, most of these sites lack basic security measures like SSL logins. This makes it easy for hackers to access the user data without the site’s direct collaboration (Gross & Acquisti, 2005).

7. EXAMINING THE SNS FOR INVESTIGATIVE PURPOSES

With the vast amount of data that is readily available on the SNSs, similar to criminals, investigators can also use this data for investigative purposes. The ways of finding user information for investigative purposes is very much similar to what the criminals use. For investigative purposes, a specific user will be targeted to gather information from. Shoemaker in her paper (as cited in Lampe, Ellison & Steinfield, 2006) write about a function called ‘surveillance’ which allows an individual to track the actions, beliefs and interests of the larger group, to which they belong to (Lampe et al. 2006).
Lampe et al. classified this type of surveillance by the goals of users as ‘social searching’ or ‘social browsing’. Social searching is where the site is specifically used to investigate specific people. Social browsing helps to find people or groups with whom the individual wants to connect offline (Lampe et al. 2006). Social searching is the type of surveillance that investigators use as they target specific individuals to gather information.

Just like normal people, criminals tend to keep their profiles open to public. Some criminals go as far as to put status updates about the crimes they have committed. Investigators can use these SNSs to verify an alibi, or to even just check up the profile of the particular individual. SNSs are used by the investigators as form of resource. They use these sites more reactively rather than proactively (Klein, 2008). Apart from investigators, people like insurance adjustors, insurance attorneys, prosecutors, defense attorneys are also taking help of the SNSs to check out their clients or their witnesses (McKinney, 2010).

8. SOME CYBER CRIME CASES INVOLVING THE SNSS

With the increase in popularity of the SNSs among general public, there is also an increase of popularity of these SNSs among criminals. These SNSs have actually opened a lot of doors for the crime world and the ways the crime is committed. Now-a-days, reports of lot of criminal activities involving the SNSs can be heard. Facebook and MySpace are especially popular in this area. A few cases relating to SNS are listed below.

Recently in the news was John Forehand, who was arrested for allegedly asking his teen daughter for sex over Facebook. John Forehand started communicating with his teen daughter over Facebook. He then told her that he was having inappropriate dreams about her and then proposed sex with her via posting graphic details of the activity on her Facebook account (The Huffington Post, 2009). In this particular case, the teen daughter added John Forehand to her contacts/friend list, as she must have trusted him since he was her father. If we can call John Forehand a predator, then the daughter’s other teen contacts could be considered potential victims. John Forehand had easy entry in any open accounts via his daughter’s contact/friend list.

In another case, a man named Robert A. Wise was arrested on charges of online solicitation of a minor. Wise was sending explicitly sexual messages to a teenage girl via MySpace. After being contacted the police posed as the 14 year old girl on MySpace and via the MySpace chat arranged a meeting with him. When Wise came to meet the girl at a prearranged spot, he was arrested. The cops also found online evidence against Wise to charge him with the sexual assault of another 14-year-old girl he had allegedly met on MySpace (Schrobsdorff, 2007). In yet another case a Houston man was arrested with sexual assault of a child. This 15 year old teen had been communicating with this man on MySpace and had actually snuck out of the house to meet him in person (Schrobsdorff, 2007).

In yet another incident, Emily Mayhan, a 20 year old Facebook user found that her Facebook account had been hijacked and the password to the account changed due to which she could not access the account. After that several of the contacts in her friend list started getting messages stating that she was stranded in London without cash and in urgent need of cash. Facebook closed her account on account of suspicious activity after a few days but no action was taken on the incident. According to the Federal Bureau of Investigation (FBI), this is a case of online hoax or phishing, which takes place for identity theft or for financial information (Davis, 2009).

In yet another phishing scam on Facebook, Bryan Rutberg’s Facebook account had been hacked into and messages appearing from him were being posted saying that he is in urgent need of help. Many of his contacts also received emails saying he had been robbed at gunpoint while travelling in the United Kingdom and he was in need of money. Rutberg was locked out of his own account and the scammer had even removed his wife from his contact list. The account was de-activated after about 24 hours (Sullivan, 2009).

And lastly but not the least, “Spam King” Sanford Wallace was sued by Facebook for accessing users’
accounts without their permission and then sending phony messages and posts. Facebook claimed that Wallace used phishing sites or other similar means to fraudulently gain access to Facebook accounts of the users. After that, he used these accounts to distribute phishing spam throughout the network (cnet, 2009). Wallace was also charged and fined for the MySpace case in 2008 where he sent junk messages to the MySpace users (USA Today, 2009).

9. CONCLUSION

While social networking sites are a good way of maintaining relationships and building new relationships, a user should be always aware of the existing dangers of using these sites. With the high amount of personal data put on these sites, there is always a risk of this data being exploited. Even with the use of privacy settings, the data is easily accessed through an open account or even through a closed account. Howsoever innocent the personal data is, online predators are always watching and users are targeted via phishing.

10. REFERENCES


(2009), Social Networks, from the 80s to 00s. Retrieved from Gigaom: http://gigaom.com/2008/01/20/social-networks-from-the-80s-to-the-00s/, November.


11. APPENDIX

11.1 User Profile Fields for Facebook, MySpace, and Twitter

<table>
<thead>
<tr>
<th>Facebook</th>
<th>MySpace</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name/alias</td>
<td>Name/alias</td>
<td>Name/alias</td>
</tr>
<tr>
<td>Photograph</td>
<td>Photograph</td>
<td>Photograph</td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td>Lists</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Birthday</td>
<td>Birthday</td>
<td></td>
</tr>
<tr>
<td>Sexual orientation</td>
<td>Sexual orientation</td>
<td></td>
</tr>
<tr>
<td>Hometown</td>
<td>Hometown</td>
<td>Location</td>
</tr>
<tr>
<td>Relationship status</td>
<td>Relationship status</td>
<td></td>
</tr>
<tr>
<td>Friends list</td>
<td>Friends List</td>
<td>Followers/Following/Lists</td>
</tr>
<tr>
<td>Political views</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious views</td>
<td>Religious views</td>
<td></td>
</tr>
<tr>
<td>Activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musical taste</td>
<td>Musical taste</td>
<td></td>
</tr>
<tr>
<td>Television taste</td>
<td>Television taste</td>
<td></td>
</tr>
<tr>
<td>Movie taste</td>
<td>Movie taste</td>
<td></td>
</tr>
<tr>
<td>Books taste</td>
<td>Books taste</td>
<td></td>
</tr>
<tr>
<td>Quotations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email address(es)</td>
<td></td>
<td>Email address</td>
</tr>
<tr>
<td>Telephone number(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instant messenger ID(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational history</td>
<td>Educational history</td>
<td></td>
</tr>
<tr>
<td>Employment history</td>
<td>Employment history</td>
<td></td>
</tr>
<tr>
<td>Group affiliations</td>
<td>Networking Categories</td>
<td></td>
</tr>
<tr>
<td>Photo albums</td>
<td>Photo albums</td>
<td></td>
</tr>
<tr>
<td>Blog entries</td>
<td></td>
<td>Link to online Bio</td>
</tr>
<tr>
<td>Personal 'about me' entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal heroes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'Who I'd like to meet'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zodiac sign</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 11.2 Some Recommended Safety Guidelines for Users of SNSs

- Always be aware that any content once put on the Internet always stays there even though it appears to be deleted.
- Always be aware of the content put out on the SNSs.
- Always use the available privacy features on the SNS. Do not leave the user profile open to be accessed by everyone.
- Do not accept friend/contact requests from unknown people.
- Think twice before putting any information on the SNSs.
- Avoid putting photos that might attract unwanted attention from any online predators.
- Avoid putting detailed information of oneself as well as family members like spouses, children, and parents.
- Be aware of any phishing content that might appear to be posted by any of the friends.
- Always be aware of the risks of social networking and different uses of SNSs.
Hard Disk Storage: Firmware Manipulation and Forensic Impact and Current Best Practice

Gareth Davies, Iain Sutherland
Faculty of Advanced Technology
University of Glamorgan
CF37 1DL
+44(0)1443 480480
{gddavies, isutherl @glam.ac.uk}

ABSTRACT
The most common form of storage media utilized in both commercial and domestic systems is the hard disk drive, consequently these devices feature heavily in digital investigations. Hard disk drives are a collection of complex components. These components include hardware and firmware elements that are essential for the effective operation of the drive. There are now a number of devices available, intended for data recovery, which can be used to manipulate the firmware components contained within the drive. It has been previously shown that it is possible to alter firmware for malicious purposes, either to conceal information or to prevent the drive’s correct operation. We review the general construction of a hard disk drive. In particular we examine the error handling process present within hard disk drives for dealing with failed or failing sectors and detail how this can be manipulated. The potential forensic impact on an investigation of manipulating firmware is then explored. We propose best practice considerations when analyzing a hard drive where firmware manipulation is suspected and detail a possible method to detect this form of modification.

Keywords: Hard Disk, Steganography, Data Recovery, Firmware.

INTRODUCTION
The hard disk drive remains one of the most common storage devices and therefore commonly features in digital investigations. There are numerous papers discussing best evidential practice and a substantial number of procedures, including the Association of Chief Police Officers Guidelines (ACPO 2008) in the UK and the Department of Justice, Prosecuting Computer Crimes guidelines in the USA (DoJ 2007). These guidelines clearly consider hard drive media and define best practice processes and procedures for the collection and general analysis of digital evidence. However, in specific cases where a technically competent suspect has access to particular data recovery hardware and software, there is the potential for the various hard disk drive firmware implementations to be manipulated for malicious purposes. This can allow the user to have the capacity to conceal information on the drive and place this data beyond forensic recovery using standard tools and techniques. There is also the potential for the drive to be sabotaged by these tools and by possible future forms of malware, prohibiting any form of forensic analysis. Therefore, there is a need for an investigator to understand some of the processes that can be undertaken to recover data from a damaged disk drive and also the potential for these techniques to be misused allowing the concealment of potential evidence. This enables the investigator to comprehend the forensic significance / impact of data recovery techniques.

HARD DISK DRIVE FUNCTIONALITY
A hard disk drive is a complex device composed of platters, voice coils, read / write heads, casing,
mountings, a motor and a printed circuit controller board. These are manufactured in a number of form factors, the most common being the 3.5 inch and 2.5 inch disks found in desktop and laptop systems respectively. The data storage area is composed of a stack of metal, ceramic or glass platters coated with a magnetic film. Each disk surface has a separate armature and head assembly. One rotation of the disk at a particular radius is known as a track. For sets of surfaces, a set of tracks at the same radius is known as a cylinder. The sector is the smallest addressable unit, typically containing 512 bytes of data. A specific sector address can be found using the cylinder address (C) the Head (H) and the Sector (S). At a higher level of abstraction the Logical Block Address (LBA) method assigns a sequential number to each sector.

The main hard disk drive manufacturers now offer drives with a maximum capacity of around 2TB. Once the drive has been formatted and contains a file system, the capacity is somewhat reduced. Not all areas of the disk are addressable by the host computers operating system as shown in Figure One. In addition to the user addressable space, there are areas of the drive that are used for the manufacturer to record data. These include the Host Protected Area (HPA) used for holding diagnostics and other utilities required by the PC manufacturer (Gupta et al 2006) and the Device Configuration Overlay (DCO), either or both of which can exist on a hard disk. The Device Configuration Overlay (DCO) is similar to the HPA, but is used by manufacturers to configure drive sizes and may exist at the same time. An excellent overview of the HPA and DCO are provided in Carrier (2005).

![Figure One: Overview of Disk Data Storage Areas](image)

The firmware area / system area of the drive is not accessible during the normal operation of the drive and subsequently is not addressable by the average user or the operating system. An initial portion of the drive firmware is present on the PCB controller board. This is then responsible for loading the platter resident firmware / system area which facilitates full operation. Disk firmware controls all aspects of the internal hard drive operation. The firmware controls the disk startup / self-check sequence when the system is powered on, placing the drive in a ready state that allows the host computer to load an operating system. During operation the firmware ensures the correct operation of the hard drive, allowing it to correctly interact with other components on the system (e.g. the operating system).

**FIRMWARE OPERATIONS**

The firmware in the majority of drives consists of a series of modules; P-list, G-list, SMART Attributes and U-List (Firmware Zone Translator). Each of these performs a key function. An example function is defect control, no disk is manufactured without flaws and there will be some sectors on the drive that cannot be used. At the time of production these flaws are recorded in the disk firmware as the ‘P’ (permanent / primary / production) list. As the disk ages and through wear & tear other sectors may fail; this is recorded in the ‘G’ (growth) list. Reads and writes are automatically redirected (remapped) to spare sectors, see Figure Two below.
P-list and G-list sectors are automatically bypassed by the drive electronics and so do not slow down drive sector access times. By adding or removing a sector from the P-list and/or G-list, we have the ability to hide/make-visible data on the hard-drive.

Figure Two: G-list Remapping

This process is transparently handled by the disk and occurs ‘beneath’ the operating system via the two lists, P-list and G-list (Blyth et al 2008). The firmware in these disks may fail. The G-list may become full on some disk models and as a result the disk may stop working. An error in the firmware can prevent the disk being accessed while still physically healthy and all user data remaining intact.

FORENSIC IMPACT

In a previous paper (Sutherland et al 2009) we have examined the possibility of steganography and data hiding via the manipulation of disk firmware. This paper addresses the issues of the possible forensics impact of these techniques on forensic practice and procedure. The possibility of attacking systems via firmware distribution and the use of microcode exploits has also been discussed by Zhou et al (2009).

To date there are only a limited number tools available tools to perform repair or modifications on firmware. There are a number of free / shareware tools that claim to read some portions of the firmware, usually disk model / serial number (Browsedata 2004). But these tools do not facilitate sufficient control over the firmware to make repairs or exploits possible. In terms of commercial products the authors are aware of two systems available for this type of analysis and repair. Both systems comprise a combination of hardware and software tools. One particularly sophisticated tool originates in Russia and costs in the region of $4000. The full Russian tool suite includes the ability to extract data and work with some solid-state devices and SCSI disks is in the region of $15,000. A more readily available device is offered from China and can be obtained via resellers in Europe for around $350 per disk manufacturer. Either of these tools would enable a competent user to manipulate firmware to conceal data or code from the hard disk dive itself.

There are a number of possible scenarios where this technology could be misused. An individual may use disk firmware steganography to conceal information within the drive, either by using the firmware defect control system or by the manipulation of bad sectors (Blyth et al 2008, Sutherland et al 2009). Another possibility is the use of malware on the drive to prevent the disk ever operating correctly again by attacking unique critical elements of the firmware, denying a user or forensic investigator access to any data. This kind of exploit would be developed and targeted at particular disks and systems and would act as a sophisticated method of sabotage that could render the drive contents irrecoverable.

Firmware manipulation can have a significant impact on the forensic process. Data that has been
hidden using firmware steganography techniques will not appear for analysis in a traditional forensic image. In the event of malware targeting and corrupting the firmware, this will prohibit any forensic images to be acquired from the hard disk drive. As previously outlined, this type of manipulation is not detected by the current suite of forensics tools and so requires both specialised tools and training to detect. However, an investigator having been trained in this area and in possession of the correct equipment is still faced with a number of problems.

In the case of determining if the disk firmware has been tampered with, either to conceal information or as a result of malware targeting and corrupting the firmware, the investigator would need to be able to assess the validity of the firmware. This process has a number of challenges: each family of drives is unique so the investigator would need to evaluate the drive against a comparable disk and perhaps use key firmware modules, even hardware components, from the donor drive to verify the firmware is valid or to enable the repair of the original drive to a fully functional state to allow forensic analysis. The ability to do this work would require a substantial library of firmware / donor drives for comparison and replacement of failed mechanical parts. This library would be difficult to build, as often donor drives are difficult to obtain and match due to strict compatibility criteria, which changes dramatically by manufacturer.

In a previous paper (Sutherland et al 2009), the authors examined and highlighted the possibility to manipulate the firmware, via the defect control system i.e. the error lists; the P-list and G-list. If a malicious attempt has been made to manipulate the error lists then an LBA has been remapped to a reserve location on the drive and subsequently altered. This presents a situation where the original ‘bad’ sector contains one piece of information and the remapped sector may contain different information, this maybe due to time factors i.e. the original file copied has been overwritten and replaced with another file, or, direct editing of the information contained at those remapped LBA addresses has been done purposely. This can be achieved via a hex editor by zeroing out the hex values that correspond to the LBA number or saving an alternate file to that LBA location, overwriting the copy of the original data (see Figure Three below).

![G-list Forensic Impact](image)

**Figure Three: G-list Forensic Impact**

In this situation the user and most forensic tools would see and access the data at the remapped sectors, not the original contents. It is possible using advanced data recovery tools and techniques for the investigator to still access the original location and data, regardless of whether it was a genuine system shift or the cause of system manipulation. In this case, in attempting to access these areas the investigator will have to work on, and alter, the original media, which would require changes to best practice procedure.

Detecting this form of misuse is potentially difficult. The major problem lies in the ability to obtain and verify the error lists. The error lists and certain portions of the data contained in the firmware /
system area are unique and disk model specific and therefore cannot be compared to another version of the disk. A donor drive with matching firmware is required and this can still differ in parts somewhat. However, it is possible to return the hard disk drive back to it’s original ‘factory settings’ with regards to the original list of all available physical sectors that can be used by the disk, whilst still retaining user data. This would enable the investigator to access and forensically analyse all the original sectors. Advanced data recovery tools would be necessary to facilitate the manipulation of the defect control system and also adjust the hard disk drive read / write parameters, enabling the retrieval of data from all sectors regardless of whether they are legitimate bad sectors, as these sectors have not yet ultimately failed but have been marked as bad by the system and discontinued from service.

To access all original CHS locations that were available at the time of manufacture, the G-list module for the defect management system would have to be fully cleared. Carrying out this action would allow the retrieval of any data that may have been maliciously hidden via the firmware steganographic technique previously devised. However, all data that has been saved to the reserved sector area will have been lost at this point, which would result in the loss of potential evidence data.

**FORENSIC BEST PRACTICE**

These types of malicious techniques have the potential to impact upon forensic best practice. There are two possible alternatives when proposing best practice for this type of analysis. The first is to assume that this type of analysis is required in each case, clearly this would be prohibitively expensive, time consuming and unnecessary in the vast majority of cases. The second option is to apply this type of analysis when the evidence indicates the possibility of firmware tampering. The latter would have to be indicated by a combination of the suspect’s technical expertise, the presence of certain hardware and software tools at the scene and suspected incomplete or missing evidential material.

Where there are grounds for suspecting that a suspect may have modified drive firmware, then there are a number of actions that could be constituted as best practice in this type of case. At the crime scene this impacts the seizing of any computer equipment. Any products that have data recovery branding or related documentation need to be seized. The functionality of the device needs to be examined to determine if it has the capability to perform such exploits, informing the subsequent direction of the investigation. If data recovery products are recovered then the following recommendations are made:

Firstly, provided the hard disk drive is functioning correctly and allows access to the user data, a standard image of the drive should be acquired as a baseline for further detailed investigation. All hard disk drives will develop legitimate bad sectors due to natural wear and tear, these bad sectors will invoke reserved space into service and data will be remapped to this area as previously discussed. Over time it is likely the data at those LBA locations will change, so an alternative copy cannot be guaranteed. For this reason it is necessary to obtain a baseline image first with which to work from, otherwise, further firmware modifications will prevent access to this reserve area and its data resulting in the loss of potential evidence.

Secondly, once a baseline image has been created, it is suggested that the investigator should use data recovery tools and techniques to reset the dynamic defect list, the G-list, clearing its contents, which would re-align all of the original LBA numbers to their CHS counterparts. After the realignment has been successfully completed it is crucial that the hard disk drive read parameters are altered, allowing the drive more time to read from all sectors. If this is not altered, the disk may encounter problems when reading the sector and re-mark the sector as failed, subsequently not acquiring any data. This is important as malicious users may take the extra step to not only hide data in physical sectors not accessible by the disk itself, but also to hide data in legitimate failing locations, so that if the defect G list was re-set, the data would still not be fully obtainable due to default drive read time configuration.

The proposed method would allow a comparison of data between the two images. The first forensic image of the drive would contain all data residing in the user data area and in the reserved area; the
second would contain all data from the user area. This would contain all the data in the original physical sectors of the drive that were mapped out due to natural wear and tear, and or through suspected manipulation of the firmware to conceal data. Best practise would then be to separate the duplicated data as much as possible from the two images, highlighting the differences. This could be achieved by creating a known file filter from the first image. This could be facilitated via MD5 or SHA-1 hash recognition systems, and could then be used to disregard all known files from the second image, leaving the variations to be analysed in a separate environment from that of the first original forensic image. Both images together would then make up a complete copy of all data contained on the drive, allowing a comprehensive forensic investigation to be performed.

Where the suspect may be in the possession of data recovery hardware and software tools and in the event of the drive not responding to any forensic imaging attempts, there is the possibility that the drive has been purposely sabotaged via firmware corruption methods. The drive may possibly be repaired; it would depend upon the type of corruption and the model of hard disk drive itself, as all have differing firmware implementations. Best practice would be to use advanced data recovery tools and techniques to firstly diagnose the exact form of firmware or hardware corruption. This will provide an indication as to how the drive was damaged and also how to proceed with the recovery. It is possible that some repairs can be made with data recovery tools without the use of donor disk drives. The drive can then be imaged via standard practise, taking into account the above mentioned current best practise method to gain access to all areas of data.

In other cases where donor drives are required to enable repair, the investigator would have to source a suitable donor hard disk drive and begin the recovery using the original diagnosis as a starting point. If full recovery of the disk and the file system cannot be achieved, it is possible to recover some data through standard data carving techniques e.g. Scalpel. (Richard III & Roussev 2005) Advanced data carving techniques would enable the investigator to carve out tangible user and or system files from the raw data available. The proposed two part imaging method outlined above would be have to be implemented to facilitate carving from all data areas.

Malware engineered to target hard disk firmware can in theory render a full acquisition of the data contained on the drive very difficult to achieve even with sophisticated data recovery tools and donor parts. This could be due to the malware targeting disk specific critical subsystems contained in the firmware, damaging the drive. This is an area for further research. The drive in this scenario would not be able to reach a ‘ready’ state and the firmware / system area would be inaccessible for diagnosis and repair. In this case the investigator should attempt to use advanced data recovery techniques, which may be able to achieve a ‘ready’ state. One example of such a technique would be to emulate the service area of a matched donor disk on to the hard disk needing to be recovered, in some cases enabling access to the firmware so repairs could be made, again, if possible.

**SUMMARY AND CONCLUSIONS**

There are a number of potential problems relating to the forensic analysis of malicious hard disk firmware modification. Without the correct knowledge of the systems it can be very difficult to find or reverse this type of modification. Hardware and software costs supporting this type of analysis are significant. The correct training is not widely available and is expensive. While this remains unlikely to impact the vast majority of forensic cases, the increasing availability of the data recovery tools used to carry out this work makes it a possible area for future concern. In this paper the forensic impacts of such hard disk firmware exploits have been discussed and suggested current best practice has been put forward for the correct handling of such cases.

**ACKNOWLEDGEMENTS**

The authors would like to acknowledge members of the Information Security Reach Group (ISRG) at the University of Glamorgan in particular Prof. Andrew Blyth and Mr. Nick Pringle. And EBS for the proof read.
BIOGRAPHY

Mr. Gareth D. O. Davies is a Ph.D. Student in the Faculty of Advanced Technology at the University of Glamorgan. The main focus of his research is the security and forensic analysis of hard disk technology. He is a part-time lecturer on the Computer Forensics undergraduate degree at Glamorgan University and has been involved in a variety of other research projects in the area of Computer Forensics and Information Security. Mr. Davies has also acted as a consultant and investigator on forensic and disk recovery technology cases in the University’s Computing Forensics Laboratory.

Dr. Sutherland is Reader of Computer Forensics in the Faculty of Advanced Technology at the University of Glamorgan. His main field of interest is computer forensics, he maintains the University’s Computing Forensics Laboratory. Dr. Sutherland has acted as an investigator and consultant on both criminal and civil cases. In addition to being actively involved in research in this area and supervising a number of Ph.D. students, Dr. Sutherland teaches Computer Forensics at both undergraduate and postgraduate level on the university’s computer forensics degree schemes.

REFERENCES

Accessed 9th April 2010

Accessed 11th March 2009

Blyth A.J.C., Sutherland I., Pringle N., (2008) Tools and Techniques for Steganography and Data Insertion onto Computer Hard-Drives, 8th Annual Program Manager’s Anti-Tamper Workshop, Sponsored by US DoD Anti-Tamper Executive Agent SAF/AQL and Department of the Army, Redstone Arsenal, Huntsville, AL, USA.


Accessed 9th April 2010


Browsedata (2004) HDD Firmware Serial Number Source Code 1.01 Free Download  
Accessed 11th March 2009


Accessed 10th April 2010

Measuring Whitespace Patterns as an Indication of Plagiarism

Ilana Shay
Zeidman Consulting
15565 Swiss Creek Lane
Cupertino, CA 95014 USA
phone: 408-203-9715
fax: 408-741-5231
Ilana@ZeidmanConsulting.com

Nikolaus Baer
Zeidman Consulting
15565 Swiss Creek Lane
Cupertino, CA 95014 USA
phone: 805-699-6452
fax: 408-741-5231
Nik@ZeidmanConsulting.com

Robert Zeidman
Zeidman Consulting
15565 Swiss Creek Lane
Cupertino, CA 95014 USA
phone: 408-741-5809
fax: 408-741-5231
Bob@ZeidmanConsulting.com

ABSTRACT
There are several different methods of comparing source code from different programs to find copying\(^1\). Perhaps the most common method is comparing source code statements, comments, strings, identifiers, and instruction sequences. However, there are anecdotes about the use of whitespace patterns in code. These virtually invisible patterns of spaces and tabs have been used in litigation to imply copying, but no formal study has been performed that shows that these patterns can actually identify copied code. This paper presents a detailed study of whitespace patterns and the uniqueness of these patterns in different programs.

Keywords: Copyright Infringement, Intellectual Property, Litigation, Open Source, Plagiarism, Source Code, Source Code Similarity, Whitespace.

\(^1\) Although many in the field refer to plagiarism, this is not accurate. Plagiarism is unauthorized copying. The algorithms defined in this field of computer science can detect copying but not whether the copying was authorized. We will refer to “copied code” instead of “plagiarized code” except where we are quoting from other papers.
INTRODUCTION

When writing code, the programmer is focused on the visual elements: statements, comments, variable names, strings. During the writing process the programmer also uses non-printing characters to separate the programs visual elements. The non-printing characters can be spaces, tabs, or newlines. The sequence of these non-printing characters is the whitespace pattern.

The significant problem of copied code in academia and industry, and ways to address the problem, has been discussed in a number of papers (Abraham, S. and Milligan, G. 2008). Papers that survey the various copy detection methods for detecting copying discuss the examination of language tokens like comments, data types, identifiers and strings but do not mention whitespace (Whale, G. 1990) (Sallis, P., Aakjaer, A., and MacDonell, S. 1996) (Clough, P. 2000) (Goel, S. Rao, D., et al. 2008). Some papers that discuss specific techniques for copy detection mention whitespace in passing, but do not address it in their algorithms (Parker, A. and Hamblin, J. 1989). Several significant papers in the field describe copy detection algorithms that ignore or remove all whitespace before performing a comparison (Baker, B. 1995) (Hamilton, J. 2008). Early papers by one of the authors of this paper replaces all sequence of whitespace characters with a single space character (Zeidman, R. 2006, 2008).

Only one paper investigated copy detection methods by specifically looking at whitespace (Aliefendic, S. 2009), but the methods presented in the paper were not successful at separating copied code from independently developed code. We decided to investigate whitespace file patterns and determine whether comparing whitespace patterns in different files is a reliable method to measure code similarity and thus detect copying.

HYPOTHESIS

Our hypothesis was that if a whitespace pattern is a good method of evaluating file similarity, then a file copied from another file will have similar whitespace patterns, while two independently developed files will have very different whitespace patterns. We would not expect different files to have similar whitespace patterns.

We will score file pairs based upon a percentage of similarity of their whitespace patterns. An effective method should meet the following criteria:

- The whitespace pattern of a file compared to itself should produce a 100% similarity score.
- The whitespace pattern of a file compared to a completely different file should produce a low similarity score.

To test this hypothesis we compared and determined the percentage of similarity of the whitespace patterns of source code files from one program against themselves. Each file should be very similar to exactly one file in the set, itself. If we graph the similarity scores of the whitespace patterns for all file pairs, and if whitespace pattern matching is a good method, the similarity scores should produce a low average similarity score with one large, narrow peak close to 0% and a small peak, representing files compared to themselves, at 100%.

We also compared source code files from one program against another unrelated program. If we graph the similarity scores of the whitespace patterns and obtain one peak, it can be understood according to these conditions:

- Low average means that the whitespace method works really well because low similarity scores indicate that most of the files are different from each other.
- High average and high standard deviation means that the method may still work, but we need to investigate why we are getting high correlation for different files. We may need to filter out some of the files.
High average and low standard deviation means that the whitespace method finds different files as similar to each other, which means that the method does not work well.

If we graph the similarity scores of the whitespace patterns and obtain more than one peak, then this implies that there may be files that are similar for reasons we have not taken into account such as a common author or the use of third party code. For this method to be a good way of finding copied code, we need to find a way to filter out files to get one peak.

**METHODOLOGY**

There are no off-the-shelf whitespace comparison tools, so we had to develop our own. The steps to measure the whitespace similarity are:

- Convert each file in the source code to a whitespace file format.
- Analyze the results.

The CodeDiff tool in CodeSuite compares and scores the similarity of files, but we had to develop tools to convert the source code to a whitespace format, and to analyze the CodeDiff database results. These tools are described below.

**Converting Source Code File to Whitespace File Format**

We created the FileReformatter program that inverts the text in the file. By this we mean that the invisible characters become visible, and the visible characters become invisible, so that CodeDiff can evaluate the whitespace characters.

FileReformatter converts source code files to whitespace file format according the following rules:

- Every continuous sequence of printable characters is converted to one space.
- Every space is converted to the character ‘S’.
- Every tab is converted to the character ‘T’.
- Newline characters are not converted.
- The output file contains only the characters ‘T’, ‘S’, space and the original newline characters.
- Newline characters are analyzed as line separators.

For example, the following line of C code:

```c
int var = prevValue + 5;
```

can be thought of as:

```c
int(S)var(T)=(S)prevValue(S)+(S)5;
```

where (S) and (T) represent space and tab characters; so FileReformatter will translate the line into:

```
S T S S S
```

**Comparing Files Line By Line**

We used CodeDiff to compare directories that have been translated by FileReformatter. It compares files in pairs; one file from the first directory and one file from the second directory. Each line from the file in the first directory is compared to each line in the file from the second directory. CodeDiff
calculates the percentage of matching non-blank lines to the total number of non-blank lines in the first file. This percentage is the file pair similarity score. The output is a database containing all of the file pairs and their similarity scores.

**Tools for Analyzing CodeDiff Results**

Four programs were required to manipulate CodeDiff database and calculate the average score and standard deviation (“STD”):

1. **DB Skimmer**: parses the CodeDiff database and removes all entries except the n\textsuperscript{th} highest scoring file pairs for each file in the first directory. We created this program.
2. **CalcAvrScore**: calculates the average of file pair similarity scores in the CodeDiff database. The user can select the number of highest scoring file pairs to count in the average calculation, and can also disregard file pair scores based on the file size, or the number of lines in the file. We created this program.
3. **Filter DB**: removes files from a CodeDiff database, based on their extension, size, or the number of lines that they contain. This program calculates the average and standard deviation for the remaining scores in the database. We created this program.

**TESTS**

Tests were done on code that was written only in the C programming language. We ran FileFormatter on the source code and converted every single file to whitespace format. When running CodeDiff, we selected the following parameters:

- File reporting threshold: 512.
- Count all scores including 0%.
- Comparison option: percent of first file. We selected this option since we want to have the first directory, as a base for our comparisons. First we compare the same directory to itself then we compare the same directory to another directory.

**Compare One Program to Itself**

The open-source Linux Kernel version 1.0 was selected to be tested. This directory has 487 different source code files in many sub-directories. CodeDiff was set up to compare the same directory against itself, which means each file is compared against itself as well as all the other files in the directory tree. We expected to see 100% similarity when comparing each file to itself, and low similarity when comparing a file to all the other files.

**Compare One Program to a Completely Different Program**

Comparing two different programs should result in low similarity scores. We expect to see low scores, but we may also get some high similarity scores if there is code from a third party or a code generation tool. Using CodeDiff, we compared the Apache HTTP version 2.0.35, which contained 653 files to the Linux Kernel version 1.0, which contained 487 files.

**RESULTS**

In the results we display the percentage of score distributions and also the average and standard deviation (STD).

**Compare One Program to Itself, All Files**

The graph in Figure 1 counts all possible file pairs; we can see that it includes pairs with 100% similarity. All 487 files were compared against each other, for a total of 237,169 file pairs.
There were more than the expected 487 files pairs with scores of 100. There were 4,163 file pairs with 100% score and 2,421 file pairs with scores between 90% and 100%. A few reasons for this could be that header files may be disproportionally similar to each other, small files may be disproportionally similar to each other, and Linux may have duplicate code in different files or duplicate files in different directories that were identified as similar.

We assumed that header files may be similar and disproportionally contributing to the high scores, so we filtered out the scores of the files with extension .h. Because the CodeDiff score is calculated as a percentage of the file size, we thought that filtering small files with less than ten lines may also help to reduce the number of high similarity scores. Filtering both small files and header files left 276 files in the directory and 76,176 total file pairs, as shown in Figure 2. Although the majority of file comparisons fall at the low end, the average score is fairly high (35%) and the large standard deviation (25.2) means there is a big spread among the results as can be seen. Also there are still some 100% matching files.

Looking for copying the investigator looks for the highest matching score for each file in order to find files that are more likely to have been copied. A high matching score indicates high similarity. When comparing one program to itself, for each file we will have at least one 100% matching score. We might have more than one 100% matching scores; this can happen if the same file was copied to another sub directory for example.

We wanted to see what happen when we take out the identical matching scores of each file to itself so we took out the top score from the database and examined this case, which is called "second highest similarity." This case represents a real life scenario when there maybe some similar code, or similar functions, but files are not identical.
Figure 3 is a graph of the highest scoring file pairs for each file in the first directory. All but the highest scoring file pairs were filtered out of the database. The average score is 100 and the STD is 0, which was expected.

![1st high score avg=100 std=0](image)

Figure 3: Compare one program to itself, 1st highest score.

**Compare One Program to Itself - Second Highest Similarity**

A new database was created by filtering out all but the second highest scoring file pairs for each file in the first directory. The newly created database has only one file pair for each file; the highest matching file other than itself. Using CodeSuite we created a summary spreadsheet that had 487 file pairs. Even after removing the top matching files, we can see that we have some 100% matching scores.

![2nd high score avg=79 std=18.4](image)

Figure 4: Compare one program to itself, 2nd highest score.

We then filtered out both header files and files with less than 10 lines to produce the data that contain 276 file pairs score for Figure 5. The average is still high, but we can relate this to the fact that since we compared the same program to itself, maybe the same writer wrote the other files so there still may be some similarity in the code. At this point, however, this whitespace pattern matching has not produced a low, narrow peak near zero that would allow us to confirm that these files are not copied from each other.
Figure 5: Compare one program to itself, 2nd highest score; filter both small files and header files.

**Compare One Program to a Different Program, All Files**

The comparison of the 653 Apache server files to the 487 Linux Kernel files produced 318,011 file pairs. The results are shown in Figure 6. The average is low but interesting to see that we still have 100% matching scores.

Figure 6: Compare one program to a different program, all scores.

Filtering out small files and header files resulted in a database with 118,956 total file pairs. This shows a low average and low standard deviation, as we would like, but again this is looking at all file combinations, which does not help us detect copied files.

Figure 7: Compare one program to a completely different program, all scores; filter both small and header files.
**Compare One Program to a Completely Different Program – First Highest Similarity**

We have assumed that the Linux source code files are completely different from the Apache source code files, so the whitespace similarity scores should be very low. We have also assumed that Linux and Apache have few if any identical files, so the first high similarity graph of two different programs should be comparable to the second high similarity graph of a program against itself that was shown in Figure 4. Figure 8 is a graph of the highest scoring file pairs for each of the 487 files in the Linux directory, with 487 file pairs.

![Figure 8: Compare one program to a different program, 1st high score.](image)

The average is high but it is lower than the average when comparing one program to itself. This graph still contains 100% scores.

Filtering out small and header files resulted in a smaller database that has 274 file pairs. The graph in Figure 9 has a high average considering the fact that the compared files are from different programs and we believe them to be developed independently.

![Figure 9: Compare one program to completely different program, 1st high score; filter both small and header files.](image)

**Reasons for High Similarity of Different Programs**

Looking at the original source code of the similar files, we found that some of the high similarity scores were generated because of:

- Header files. Most of the lines in header file contain variable types and variable names; when translating these lines to whitespace format, these lines are similar.

- Small files. Similarity is calculated as a percentage of the file size, small files will have higher similarity percentages.

- Macros and definition statements. Files that extensively use macros and definitions
may look similar.

- Common program statements. The whitespace translation of two very different lines of code may be the same, especially when common programming statements like ‘if’ or ‘for’ are being used.

- Common author. It is possible that the programs actually have some common authors that worked on both programs.

- Third party code. Using third party code in different programs increases the similarity scores.

- Automatic code generation. Using automatic code generator tools like Visual C++ wizard will probably increase the similarity scores.

- Copied routines. Some common routines may have been copied between the two programs.

- Copied files. Some files may have been copied between the two programs.

That may all be reasons for high similarity but probably cannot account for this significant amount of similarity.

**CONCLUSION**

This whitespace pattern matching method can be used to focus a search for evidence of similarity or copying, but this method cannot stand by itself. When we compared a set of files to themselves, there were many files at the low end of the graph, but there was a wide distribution rather than a thin peak. Even after attempting various filtering methods, we often had a large spread and a number of different files with identical or nearly identical whitespace patterns.

Even when we compared completely different files from completely different software projects, and filtered out header files and small files with less than 10 lines, there were still many files with high similarity scores. Therefore, this method is not precise. High whitespace comparison scores do not necessarily mean that there is similarity between programs, since it has been shown that completely different programs may also have high scores. Low scoring file pairs indicate a low level of similarity, and high scoring file pairs do not imply that copying occurred. High scoring file pairs indicate where further investigation for copying might focus.

**FUTURE WORK**

Future work can be done in the following areas:

- Examine sequences of whitespace. Perhaps similar sequences of whitespace in lines of code are better indicators of copying than line-by-line patterns. For this we will need to use CodeMatch® function of CodeSuite® instead of CodeDiff®.

- When filtering out small files, vary the number of lines in a file that we define as a small file until whitespace pattern matching is effective.

- When comparing the same program to itself, filter out files with the same name. In this paper we eliminate the top score, which was 100%. But if the same file was copied to a different directory we will get other file pairs with similarity scores of 100%. This filtering will eliminate comparing identical files that were copied to different directories.

- Examine all unseen characters in addition to space and tab.
Test this whitespace pattern matching method on different languages.

Compare code generation tools. When using a code generation tool, like MFC Wizard, the whitespace similarity score is expected to be higher than for manually created code. We cannot currently eliminate code from code generation tools because we don’t know how to identify it from its whitespace pattern.

Compare two different versions of the same program as a more rigorous test of whitespace pattern comparisons.

**AUTHOR BIOS**

**Ilana Shay** Ilana Shay is a research engineer at Zeidman Consulting. She has developed software for advanced printers, e-commerce, image processing, and semiconductor equipment. She was the head of the computer science department in Ort Holon, technical high school. Ilana is certified in the use of CodeSuite®. Ilana has won recognition for excellence from Tencor Instruments. She holds a bachelors degree in mathematics and computer science from Tel Aviv University.

**Nikolaus Baer** is a research engineer and projects manager at Zeidman Consulting. He has developed software for marine research, optical testing equipment, military terrain databases, mobile applications, and medical devices. He has written articles and given presentations about software trade secret theft and how to analyze source code for detecting it. Nik is certified in the use of CodeSuite®. He placed first in the Start Cup 2004 business plan competition. He holds a bachelors degree in computer engineering from UC Santa Barbara, where he attended on a Regents Scholarship.

**Robert Zeidman** is the President of Zeidman Consulting. He is the author of the books *Designing with FPGAs and CPLDs*, *Verilog Designer's Library*, and *Introduction to Verilog*. Bob is a Senior Member of the IEEE and was the recipient of the 1994 Wyle/EE Times American by Design Award and the 2003 Jolt Reader's Choice Award in addition to other engineering and scholastic awards. He holds seven patents and has an MSEE degree from Stanford University and BS degrees in physics and EE from Cornell University.

**REFERENCES**


Electronic Discovery: A Fool’s Errand Where Angels Fear to Tread?

Milton Luoma  
Assistant Professor  
Metropolitan State University  
700 East 7th Street  
St. Paul, MN 55106  
651 793-1481  
651 793-1246 fax  
Milt.Luoma@metrostate.edu

Vicki Luoma  
Associate Professor  
Minnesota State University  
145 Morris Hall  
Mankato, MN 56001  
507 389-1916  
507 389-6497  
Vicki.Luoma@mnsu.edu

ABSTRACT
Electronic discovery has transformed the discovery phase of civil litigation in recent years. The expectations of lawyers and parties were initially established in the Rowe and Zubulake cases that led to a complete revision of the electronic discovery rules contained in the Federal Rules of Civil Procedure. Subsequent cases have underscored the importance of document search methodologies and implications for attorneys, IT professionals, and digital forensics professionals. The authors review how electronic discovery has evolved thus far and offer recommendations regarding the electronic discovery process.

Keywords: Electronic discovery, e-discovery, keyword search, concept search,

INTRODUCTION
Technology has changed the legal profession forever and civil litigation will never be the same for lawyers. Information has never been more available but at the same time so overwhelming in volume and less comprehensible to the typical attorney. By 2011 there will 1,800 exabytes of electronic data in existence, or 1.8 zettabytes. Data are growing by a factor of 10 every five years. (Mearian, 2008) In addition, electronic containers for data such as files, images and tags are presently growing 50% faster than the data are growing. (Mearian, 2008) All of the books, tapes, and other documentation contained in the Library of Congress equal only about 10 terabytes (10,000 gigabytes) of data. (Mearian L., 2007) Today, over 94% of all information is created and stored digitally. Regardless of its source, 85% of all produced electronic data eventually passes through a corporate computer, Web site, network or asset, and then that corporation has responsibility for that information that may have to be searched in the discovery process in a lawsuit. (Mearian L., 2007)

Further, information is rarely deleted because storage is easy and inexpensive. The devices needed to store information keep getting physically smaller but larger in capacity. For example, a thumb drive is expected to be available in 2011 with a capacity of up to terabyte of information. (Wired, 2007) This volume of stored data creates an overwhelming dilemma when litigation is commenced. A litigator...
must determine what information should be sought and how to ask for the information all the while trying to determine how to find all of the information the other litigator requests.

Before the digital age attorneys searched through available hard copy documents to retrieve those documents the opposing party requested. If the opposing attorney failed to request the right documents or if they were no longer available the matter was closed. In 2007 new Federal Rules of Civil Procedure were passed to try to clarify court rules concerning digital information. One of the biggest changes is that litigants and their attorneys are required to be cooperative and to become transparent in the discovery process. This required transparency is contrary to the adversarial instinct of every trial attorney and the previous century of litigation practice. In addition, most attorneys have little training, understanding or experience in ESI. Even more alarming is that most attorneys do not know how much they do not know. Thus, discovery has become a very complex process with numerous requirements and many concern parties. Many litigants, software companies, and consultants are attempting to simplify the process at a reasonable cost but full compliance can make the process extremely expensive and time consuming.

THE NEW FEDERAL RULES OF CIVIL PROCEDURE

Procedural law is different from substantive law. Substantive law governs rights and obligations while procedural law specifies the methods, procedures, and practices required to be followed in civil cases, including the disclosure of electronically stored information held by the parties. Beginning with the Rowe (in Rowe Entertainment Inc. v. William Morris Agency, Inc., 2002) and Zubulake (Zubulake v. UBS Warburg LLC, 2003) cases, the Judges Francis and Scheindlin first recognized that the court rules on discovery needed to be reinterpreted in the age of electronically stored information. Both of these cases led to the first real guidelines to litigants and their attorneys as to what would be expected of them and served as the catalyst to change.

On April 12, 2006, the U.S. Supreme Court approved amendments to the Federal Rules of Civil Procedure (FRCP) concerning discovery of electronically stored information. The amendments took effect on December 1, 2006. These revised Federal Rules put litigants and their attorneys on notice that they need to give early attention to electronic data, how and where it is stored, saved and retrieved. Rule 16 (Federal Rules of Civil Procedures, 2006) Rules 26a and f, (Federal Rules of Civil Procedure 26, 2006) require the court and the parties to give early attention to any issues concerning electronically stored information including preservation of evidence.

One of the new rules, Rule 26 (f) (Federal rules of civil procedure, 2006) requires the parties to come prepared to meet and confer to exchange information concerning the type of digital information they have, how it is stored, and how to locate the data. The major problem in this process is that attorneys are not qualified or prepared to conduct such a digital investigation.

The new rules also require that the court begin monitoring the process at the beginning of the discovery phase of litigation. The court must monitor the litigants and require them to confer as soon as possible before the first scheduling conference to discuss preservation of discoverable information and any issues relating to their discovery. (Federal rules of civil procedure, 2006)

The rules further provide that when the electronic information is not reasonably accessible, a party may not have to provide it, however, there are very few cases in which litigants have been excused from providing information, and it is much more likely that the court will sanction the non-providing party. In addition, the rules require that at the beginning of a case, each party must disclose by description, category and location of all documents and things in the disclosing party's possession, custody or control that it may use to support its claims or defenses. (Federal Rules of Civil Procedure 26, 2006).

In addition, the party requesting discovery “may specify the form or forms in which electronically stored information is to be produced.” (Federal rules of civil procedure, 2006) The producing party may then object to the requested form(s) of the information and, in such case, must “state the form or
forms it intends to use.” (Federal rules of civil procedure, 2006).

The rules further try to clarify discovery by providing that the party must produce the electronically stored information “in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable.” (Federal Rules of Civil Procedures, 2006) In the time since the rules were enacted the courts have consistently found there has been enough notice to potential litigants that their data must be stored and retrievable or the litigant faces sanctions.

In one such case, the court expressed displeasure with data downgrading stating, "taking an electronic document...printing it, cutting it up, and telling one's opponent to paste it back together again, when the electronic document can be produced with a keystroke, is madness. (Covad Commc'ns. Co. v. Revonet, Inc., 2009)

THE PRACTICAL DIFFICULTY FINDING RELEVANT INFORMATION

The problem for litigants and their attorneys is how to find the data requested. The most common methodology has been key word searches. A keyword search is a method to locate a document by looking for certain words or terms that would be found in a record. It is a method most web browser users are comfortable with and feel competent in conducting. Some of the problems with this methodology are determining what words to search and who conducts the search.

In a recent case (Asarco, Inc. v. United States Envtl. Prot. Agency, 2009) the plaintiff argued that the defendant's keyword search was inadequate and the court agreed, ordering an additional keyword search utilizing four additional key word terms. The court also noted that "keyword searches are no longer the favored methodology." (Asarco, Inc. v. United States Envtl. Prot. Agency, 2009).

There have been several studies conducted on this methodology to determine whether keyword searches are an adequate methodology. One of the earliest studies conducted in 1985 by David Blair and M.E. Maron found that attorneys were only 20% effective at imagining all of the different ways that document authors could refer to words, ideas, or issues in their case. (The Sedona Conference Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) In this study, attorneys and paralegals who had special skills in a unique area of tort law (subway accidents) were asked to identify appropriate keywords to search through 40,000 documents and 350,000 pages in a case involving a subway accident. The participants were told to stop when they thought they had found at least 75% of all the relevant documents. When the participants stopped they were certain that they had found at least 75% of the relevant documents but in actuality they had located only about 20% of the relevant documents. (Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007)

This study found that even experts searching other peoples’ works are very poor at identifying the right key words to use in a search. (The Sedona Conference Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) This type of study has been repeated several times and continues to be repeated. The Sedona Conference had participants review over one hundred hypothetical complaints to find relevant documents with much the same results. Further, the TREC Legal Track conducted a study in 2008 and in 2009. TREC Legal Track is a conference that researches information retrieval issues. The 2008 study used over 7 million documents available on the tobacco litigation and the 2009 used the Enron data publically available. The searchers in both of these studies were comprised of information scientists and litigators. (Oard, 2008).

There are several problems with key word searches that include determining all of the possible synonyms for a word. For example, there are approximately 120 words for the word “think.” (Best Practices Commentary on the Use on Search and Information Retrieval Methods in E-Discovery, 2007) Another example is that there are 25 words for “early,” 15 words for “expedite,” 16 for “appointment,” and 12 for “interview.” (Rogets, 1911, amended 1991). The results of the 2008 TREC “consensus” keyword searches found that, on average, Boolean keyword searches found only 24% of the total number of responsive documents in the target data set. (Oard, 2008)
Other problems include ambiguity, acronyms, abbreviations, misspellings and typographical errors. Another keyword search methodology is Boolean searches which allow the searcher to combine words and phrases using logical operators such as the words AND, OR, NOT and to limit or increase the search. These studies along with civil cases have highlighted the problem in determining which keywords should be used. The new rules require litigants to have early meet-and-confer sessions and if the parties are going to rely on keyword searches for their ESI discovery, then they should attend with keywords in mind and with an idea of what keywords would find the documents of interest. As consultant Craig Ball states, the meet-and-confer sessions as required by Rule 26(f) consist of “two lawyers who don’t trust each other negotiating matters neither understands.” (Ball, 2006)

Judge Andrew Peck stated, “This case is just the latest example of lawyers designing keyword searches in the dark, by the seat of their pants, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails.” He continued, “Moreover, where counsel are using keyword searches for retrieval of ESI, at a minimum they must carefully craft the appropriate keywords, with input from the ESI custodians as to the words and abbreviations they use. The proposed methodology must be quality-control tested to assure accuracy in retrieval and elimination of false positives. It is time that the bar – even those lawyers who did not come of age in the computer era --understand this.” (William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 2009)

Judge Andrew Peck further stated, “The case should be a wakeup call for them to understand the need for careful thought, quality, control, testing, and cooperation with opposing counsel in designing keywords to produce emails or other electronically stored information (ESI)” (William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 2009)

This case points out two problems indicating that keyword searches are not the best methodology and attorneys are not prepared for ESI searches. Again, in Equity Analytics v. Lundin the court set the protocol for searching his former employee’s computer. The plaintiff Equity Analytics claimed that defendant, its former employee, gained illegal access to electronically stored information after he was fired. Defendant claimed that another employee gave him permission to use his username and password. The defendant by his own admission accessed his employer’s computer system at least 18 times in a 90 day period. When the defendant provided discovery, the information was obtained through keyword searching. The employer argued against the defendant’s use of these terms as inadequate because defendant had uploaded a new operating system (“Leopard”) on his home computer in 2007. Equity further argued that the files could have been compromised that were on his previous computer and further converted easily from one format to another to hide information. Equity also argued if they were not given the right to make mirror image of the hard drives that there were likely fragments of information that keyword search would not find. (Equity Analytics, LLC v Lundin, 2008)

Judge Facciola stated that “determining whether a particular search methodology, such as keywords, will or will not be effective certainly requires knowledge beyond the ken of a lay person (and lay lawyer) and requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence.” (Equity Analytics, LLC v Lundin, 2008). One of the most outspoken Judges on Electronic searches, Judge Facciola found in another case that the issue of appropriate searching electronic data was too complicated for lawyers and judges to address without the aid of expert testimony. The judge stated that keyword search analysis is an area of e-discovery “where angels fear to tread.” (United States v O'Keefe, 2008)

In yet another case Judge Facciola, recommended "concept searching," -- the use of complex search engines that make use of linguistic or statistical patterning to locate responsive e-mails and electronic documents, (Disability rights Council v Washington Metropolitan Transit Authority, 2007) The court found that there are risks involved with conducting keyword searches. The court said it is important to determine which keywords were chosen and how they were they used to search the document population; the rationale for selecting those keywords; the qualifications of the individuals selecting such keywords to design effective searches; and whether and to what degree the results of the search were measured for reliability.
In one civil case, (Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008), the parties could not agree on the keywords to search through 165 documents. The court ordered the parties’ computer forensic experts to meet and confer in an effort to identify a joint protocol to search and retrieve ESI responses. The computer forensic experts met and prepared a five page keyword/phrase search list. A problem arose as to how much of the material retrieved in this method inadvertently produced privileged material or work product. The defendants then came up with seventy keywords to cull out privileged material; however, the defendants were unclear as to how these 70 words were chosen, how the search was conducted, and what safeguards were employed. (Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008). Judge Grimm ruled that the defendants had waived the privilege because they did not employ a search expert, did not test the key word list, did not check the results and did not have a claw-back agreement. (Victor Stanley, Inc. v Creative Pipe Inc, et. al, 2008)

In another case the parties did have a stipulation with a claw-back agreement to return privileged documents inadvertently produced approved by the court. The stipulation provided that in case privileged material was disclosed, the document was to be returned upon written demand. If the recipient of the document wanted to challenge the privilege claim, they were required to do so in writing within five days of the demand for the document’s return. As discovery progressed, plaintiffs were permitted to search certain of defendants’ storage devices where they discovered several relevant documents not previously produced. Those documents were produced to plaintiffs and included one document that defendants later claimed was privileged and demanded it be returned. The plaintiff returned the document. The plaintiff later moved the court for the right to use the document to which the defendant objected on the grounds that the defendant had objected in a timely manner in accordance with the claw-back agreement. The court ruled that the defendant did not have to produce the document because the plaintiffs failed to follow the agreement made with the defendants. (Bro-Tech Corp. v. Thermax, Inc., 2008 , 2008)

In yet another case the plaintiff moved to compel discovery arguing that recently found e-mails and alleged irregularities in the defendant's production of documents justified additional discovery. The defendant argued that any additional e-mails were in backup storage and not reasonably accessible. In this case the court questioned the relevance of the potential evidence and noted that the timing of the request – just two weeks before the trial – made it impossible to complete the searches before the start of trial. As a result, the court granted the plaintiff limited discovery holding that the plaintiff may hire an outside expert at its own expense to search at most five of the defendant's backup tapes; the search would contain a limited number of search terms; the search would include the e-mail records of seven key employees; and the results would be subject to a confidentiality agreement. (Kilpatrick v Berg, Inc, 2009)

**RULE 702 OF THE FEDERAL RULES OF EVIDENCE**

If matters were not complicated enough for litigants, when determining appropriate sanctions, courts are now requiring litigants to justify their methods of search, the reasons for particular search words, and the expertise of the searcher. Judge Facciola held that any challenges to the search methodology in producing e-discovery must be scrutinized under Rule 702, (Equity Analytics , LLC v Lundin, 2008) (United States v O'Keefe, 2008) and decided in hearings under Daubert v. Merrill Dow Pharmaceutical and Rule 702. The Daubert case set the standard for ruling on the admissibility of expert testimony in federal courts. In Daubert, the court found that when scientific evidence is involved the court must determine whether the testimony underlying the reasoning or methodology is scientifically valid and properly can be applied to the facts. (Daubert v. Merrill Dow Pharmaceuticals, 509 U.S. 579 , 1993)

Judge Terrence McVerry in the Smith v. Life Investors Ins. Co. found that the defendant had a duty to demonstrate that its search methodology was reasonable, and the court suggested that proof could be established through identification of the keywords used, an explanation of the qualifications of those selecting the keywords, and proof of assurance testing. (Smith v Life Investors Ins. Co of Ame, 2009).
The court further found that it is the duty of the party to be aware of the literature describing the strengths and weaknesses of various methodologies. (Smith v Life Investors Ins. Co of Ame, 2009). The court also determined that if the parties’ methodology is challenged they "should expect to support their position with affidavits or other equivalent information from persons with the requisite qualifications and experience, based on sufficient facts or data and using reliable principles or methodology." (Smith v Life Investors Ins. Co of Ame, 2009)

Judge McVerry also ruled that any challenges to e-discovery because of search methodology must be viewed under Federal Rules of Evidence Rule 702 and the Daubert case. Rule 702 reads:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. (Federal Rules of Evidence, 2009)

In a later case the Supreme Court ruled that the standard outline in Daubert applies not only to testimony based on ‘scientific’ knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge. (Kumho Tire Company, Ltd. v. Carmichael, 1999) Kumho requires that a court “exercise its gate keeping obligation so that the expert, whether relying on professional studies or personal experience, will, when testifying, employ the same level of intellectual rigor that the expert would use outside the courtroom when working in the relevant discipline.” (Kumho Tire Company, Ltd. v. Carmichael, 1999)

The courts are applying the Daubert case to the e-forensic experts and their methodology. In a 2008 case involving misappropriation of trade secrets and computer fraud and abuse violations, the parties submitted competing expert testimony in the plaintiff’s motion for spoliation sanctions and cross motions requesting that the opposing forensic experts be excluded under Federal Rule 702 and Daubert. (Nucor Corp. v. Bell, 2008) In this case the court found that both proposed experts met the standards under Daubert and accepted them as experts; however, the court took the unusual position of rejecting some of the plaintiff’s expert testimony as not meeting the Daubert standard. Plaintiff Nucor requested sanctions claiming defendant Bell had destroyed evidence found on John Bell's laptop computer and the SanDisk device that Bell discarded. Nucor asked that the court grant a default judgment, or in the alternative, grant an adverse inference charge to the Jury. In support of the request John Jorgensen, an expert in the field of computer forensics, testified that the defendant had intentionally deleted information on its computer by wiping data. Jorgensen, the expert also testified that the missing SanDisk contained Nucor-related documents. Jorgensen also testified that defendant’s continued use of the laptop resulted in the loss of relevant data, and that defendant’s expert spoliated evidence because he engaged in “destructive” testing on the laptop's hard drive. Defendant’s countered Nucor's arguments with the testimony of Dr. Sean McLinden, another expert in the field of computer forensics, who testified that no wiping occurred and that no relevant data was lost as a result of defendants’ continued use of the compute. Further, McLinden claimed that he provided plaintiff with an exact copy of the laptop hard drive. However, Bell has admitted that he destroyed the SanDisk

After a Daubert hearing on the qualifications of the parties’ computer forensics experts, the court ruled that both forensic witnesses were experts under the Daubert standard but excluded the testimony of Jorgensen on the issue of wiping. Jorgensen had testified that the laptop had large blocks of “zeroes” and therefore he theorized that the zeroes could have been created by use of a secure delete function of the computer program. Dr. McLinden testified Bell’s laptop did not have a secure delete function. The court concluded that Jorgensen’s methods to test wiping failed to meet the Daubert standard and
therefore barred him from testifying on that matter. (Nucor Corp. v. Bell, 2008)

Validation under Daubert standards when litigants use software could be extremely difficult. If the person using the software is not a forensic expert it is doubtful that they would be qualified to testify as to the validity of the methodology. Furthermore, search engine software applications are proprietary. It is not reasonable or appropriate to expect the designers and proprietors of such software to testify in court to explain the proprietary methods and source code of their programs, and the actual methodology or other valuable information to satisfy the Daubert requirements. An expert could testify that he or she has validated a program by testing it in the lab obtaining the same results as the results obtained by a program known to be valid.

If litigants are required to validate their methods, they should be prepared to discuss, how many false positives were discovered, how many false positives have been found, what methods were developed to measure what is missing, and the statistical sampling methods used. Attorneys need to be aware that they may be required to defend the methodology they used in their search. Required proof may include providing quality assurance testing and providing measurement protocols to support their search methodology. Rule 34(a) of the Federal Rules of Civil Procedure has mentioned sampling as an accepted methodology for producing information so it could be argued as a method of validation.

In the O'Keefe case Judge Facciola found that when testing search engines against the Daubert test it might make a difference if the expert is attempting to prove something is present versus attempting to prove that something is not present. The problem is that typically the requirement is to find everything. In a Texas appellate case the court found the expert’s use of Guidance Software’s EnCase forensic tools to create a bit-stream forensic image of defendant’s hard drive to search for child pornography was appropriate. Since the purpose was to find whether something existed and it was found, the court accepted the evidence as meeting the Daubert standard. (Sanders v. Texas, 2006) However, in O’Keefe the challenge was that the information was not complete. It is difficult under the Daubert standard to prove that nothing was missed. Successful Daubert challenges can totally destroy the opponent’s case so litigators and forensic experts need to prepare for the eventual challenge.

Since there are better search methods than keyword methods, it would be difficult for an attorney to defend this methodology as the sole search method used after applying the foregoing standards. Litigants must use other search methodologies, such as concept searches or pattern recognition searches. However, even with other search methodologies attorneys and their experts must be able to defend the methodology used.

SEARCH METHODOLOGIES OTHER THAN KEYWORDS

The primary problem in selecting a search methodology is that the method chosen must include the ability to search, to cluster data, to classify, to filter, to analyze social network information, and finally, to extract information. In the Lehman Brothers bankruptcy case there were over 3.2 billion e-mail and instant message records that had to be reviewed. It became clear to the investigators when they realized that keyword searching would not be adequate so they had to rely on other methods including concept searching and visual mapping. (Lehman Brothers Holdings Inc., 2008) In cases like the Lehman Brothers case and others attorneys had to learn to use other search methodologies that should include fuzzy searches, semantic searches as well as concept searches.

1. Fuzzy Searches

Fuzzy searches find misspellings in words so they are particularly effective in finding results where words are typically misspelled. Spelling errors are fairly common in cases involving technical terms. The degree of “fuzziness” is normally adjusted via a numeric value. Many search engines will use the degree of fuzziness to determine the number of letters that can be wrong in the misspelling. For instance, a fuzzy value of 1 would mean that only one letter can be wrong in the word and a value of 3 means three letters can be wrong. For example, using a search term of “rake” and a fuzzy degree of 1
will find the term “fake.” Fuzzy searching is a very valuable ability when dealing with sophisticated terms, where the proper spelling is not widely known.

2. Concept Searches

Concept searching takes the input term and returns results that are related in meaning. This is best explained in this example. If the term ‘car’ was used, then results of ‘automobile’ and ‘vehicle’ would be in the returned hits. One of the most well known of the concept search engines is Attenex. The Attenex engine is very powerful, but not for the faint of heart. The cost is very high and is typically used in large, high profile cases where there is a large volume of electronic information. The key issue to be aware of with conceptual search engines is that they are only as good as the programming logic.

3. Bayesian Network Searches

Essentially, a Bayes network is a simple probabilistic classification method based on applying Thomas Bayes’ theorem regarding conditional probabilities. A Bayes network can be used to answer questions such as, “I am Caucasian male, 42 years old, and am overweight. What is the likelihood that I am diabetic?” This method has been effective in classifying documents for retrieval based on the content of the documents and the concepts that are contained in the document population. For example, based on the content of all of the documents in a given population of documents, what is the likelihood a particular document is relevant to a particular concept or search term? A Bayesian network will retrieve documents that have a high probability of being relevant to a particular class or category. As documents are added to the population of documents, the probabilities are updated. This methodology has proven to be very useful in information retrieval and data mining. The advantage of Bayesian networks in data mining is that the probabilities are updated continually during the search process, and thus, it can connect information with causal relationships. (Heckerman, 1997)

CONCLUSION

In conclusion, whatever the chosen search methodology used, attorneys must obtain the assistance needed whether it is from a linguist, a computer forensics expert, or an IT expert to provide the information requested and to demand the information through discovery needed to properly litigate. Courts have come to recognize the problem with keyword searches before litigants have. Judge Facciola stated, “Whether search terms or ‘key words’ will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics.” (United States v O'Keefe, 2008).

Litigants will need to defend their methodology and it may be necessary to employ more than one methodology to convince the court that the search can be defended. In a typical case there can be more than one hundred thousand documents that must be searched to find all of the necessary and relevant documents without providing irrelevant or privileged documents. If the search methodology seeks to be too precise, that is, a search that results in a high percentage of retrieved documents being considered relevant, there is a risk that some relevant documents may be overlooked. If the search methodology seeks to have a high level of recall, that is, to find every possible relevant document, there will likely be a high percentage of false positives, or documents that are not relevant. There is necessarily a degree of tradeoff between precision and recall that is involved in developing any search methodology. Despite the safe harbor rules in place in the Federal Civil Rules of Procedures, the courts have had little tolerance for litigants’ failure to provide the necessary discovery requests. Noncompliant litigants can be sanctioned by courts imposing fines, attorney fees, discovery costs, adverse inferences, and dismissal of the case it is entirely.

All attorneys must work with the opposing attorneys to determine the method and amount of searching that will be conducted. The new rules and subsequent cases demand that attorneys cooperate with one another and that the discovery process be transparent. However, even if attorneys agree to limit their ESI search there is always the chance of a malpractice lawsuit looming if the demand was inadequate. Attorneys must employ computer forensics experts throughout the discovery phase of the litigation
process. Further, attorneys must understand that searching methodologies require more knowledge than how to Google. Attorneys must take classes in computer forensics as well as information retrieval and electronic discovery. Universities need to develop more programs in computer forensics to prepare specialists who can advise counsel on such matters and who are prepared to take testify in court.

In this digital world attorneys must understand how to discover, produce, preserve and justify electronic data. Attorneys and their experts must be prepared for the eventual challenges to their method and experts. The cases in which the judges have required justification of the methodology has put litigators on notice that choosing a search method and expert may not be an easy matter.

“While it is now possible to store enormous amounts of reported decisions in electronic databases, the retrieval of relevant cases remains an extremely difficult task.” (Smith, 1998) It will continue to be a difficult task as data storage continues to multiply and data searching and retrieving continues to be scrutinized.

Regardless of the methodology chosen, attorneys need to have a plan to determine the best way to find, retrieve and seek electronic data. Attorneys need to know what experts are needed and engage them early in the process. Attorneys need to be prepared to argue what is reasonable to find and produce and what is reasonable to request.

REFERENCES


ClearOne Commc'n's, Inc. v. Chiang, WL 920336 (D.Utah April 1, 2008).


Federal rules of civil procedure, 16(b) (2006).


Lehman Brothers Holdings Inc., No. 08-13555 (US Bankruptcy Court, Southern District of New York September 15, 2008).
Smith v Life Investors Ins. Co of Ame, WL2045197 (W.D. PA July 9, 2009).
Victor Stanley, Inc. v Creative Pipe Inc, et. al, FRD 251 MJG-062662 (District Court of Maryland 2008).
A Layered Framework Approach to Mitigate Crimeware

Mathew Nyamagwa
Metropolitan State University
Mathew.Nyamagwa@metrostate.edu

ABSTRACT
Crimeware attacks are growing at such an alarming rate and are becoming so prevalent that the FBI now rank cybercrime among its top priorities after terrorism and espionage. New studies estimate cyber crimes cost firms an astounding $1 trillion annually. But the good news? Over 80% of them are preventable. Crimeware is not a purely technical threat but more of a socio-technical affair. This clearly brings out the fact that computers do not commit a crime, but we (humans) do! In this paper I propose a layered approach that involves all stakeholders from end-users to service-providers and law enforcement to greatly mitigate the recent proliferation of crimeware.

Keywords: Crimeware, Jurisdiction, International space

1. INTRODUCTION
Malware is now not just about a lone hacker or bored college student, it now built and pushed by technologically sophisticated organizations, aided-by phishing-like deceit tactics and spread via advertisements, social networks and other IT electronic devices for financial gain. By exploiting known and undisclosed vulnerabilities, these malware help cybercriminals capture keystrokes of individuals, spies on corporations and politicians and threatens national security by means of serverjacking, information leakage and a potential deterioration of trust in the infrastructure[1]. Unlike earlier forms of code-based attacks that operated without any human intervention, these programs frequently exploit human credulity, cupidity, or naiveté to persuade or trick the computer user into downloading, installing, or executing them.

Crimeware is not a purely technical threat but more of a socio-technical affair. From the –technical perspective, traditional security mechanisms such as firewalls, antivirus solutions, and intrusion detection and prevention systems have been used to reduce its impact. Different yet sophisticated security frameworks such the Clark-Wilson Commercial Integrity model[2] by David Clark, a professor of computer science at MIT and David Wilson, accounting executive at Ernst and Whitney, and Risk Management Framework[3] by Dr. Gary McGraw of Cigital among others have been proposed and implemented but still agree that computer security is still incapable of thwarting sophisticated threats that morph and mutate rapidly.

From the socio- perspective, Cyber society lacks all of the requisite attributes of a state as outlined in the 1933 Montevideo Convention on the Rights and Duties of States, which specifies that a bona fide country (society) must have a permanent population, a defined territory, a government, and the capacity to enter into relations with other states[4].

This definition implies that cyber society is not a distinct and sovereign state. Rather, it is a composite of loose associations that transcend traditional geopolitical nation-state boundaries. Jurisdiction in cyberspace is therefore contingent on the laws of individual countries that govern human activities in physical places. In many respects, the Internet has obscured the question of national sovereignty and jurisdiction.

To this point, three fundamental questions arise: What if the transgressor resided in a country that does not have computer crime laws? Could they be charged under the laws of nations or states that do? If cyberspace extends beyond the geopolitical boundaries that traditionally define legal jurisdictions, what responsibility do governments have to protect individuals interacting in this transient realm?
Because the Internet has obscured the question of national sovereignty and jurisdiction, crimeware attacks are more precarious. New studies estimate that cybercrime cost firms an astounding $1 billion[5], and now the FBI ranks cybercrime among its top priorities after terrorism and espionage[6]. But the good news! Over 80% of cybercrime is preventable[7]. I propose a layered framework approach that involves all stakeholders including end-users, computer scientists, law enforcement, legal agencies and politicians should combine to develop a computer security framework, computer forensics and investigation techniques, cyber laws to curb and/or deter inter-geopolitical crimeware attacks.

Section 2 looks into related work, section 3 discusses this layered framework taking a defense mechanism that combines end-user literacy, aspects of law and order in cyberspace, the use of trusted computing platform, and secure infrastructure. Finally, section 4 is the conclusion.

2. Related Work

Crimeware can be defined as any malicious piece of software that satisfies at least one of the following criteria: 1). Stealing online credentials, personal data, or any other piece of information necessary for identity takeover, with the intent of using the stolen identity to steal funds; and /or 2). Performing unauthorized online transactions in order to steal funds; this includes Trojans that “hijack” online banking or other secure sessions of infected users and carry out fraudulent transactions after the user has logged out[8].

Technical frameworks[9,10] have been presented to mitigate specific unit areas of cybercrime. In [9], the design (Rubot framework) provides an experiment framework to set up customized botnet architectures on testbeds. The problem with botnet research is the scope (millions of victims) and the evasion techniques deployed by its propagators. Likewise in[10], the phishing-prevention framework is presented, derived from SANS 17799 that focus specifically on the problem of identity theft by phishing attacks. These frameworks among other have provided a secure technical platform but given the nature and current state of crimeware, a social framework needs to be attached.

Like David Johnson and David Post[11] have argued in a number of articles, cyberspace should be its own Jurisdiction, with its own courts (or “Virtual magistrates”) to resolve disputes that occur entirely online.

3. Layered Framework Approach

Fig 1: A Layered Framework representing the degree of defense.
3.1 End User Security Management (EUSM)

Reference [12] showed data that end users were concerned with security; however, they were both unsure of the appropriate action to take and frustrated by security practices extraneous to actual work. As networked computing has become pervasive, the challenges of maintaining secure environments for end-user computing have increased enormously. End users are often cast as the “weak links” in computer security, as they tend not to be aware of the latest security attack techniques nor the rapidly-evolving preventive measures.

EUSM represents a tradeoff. Reducing users’ responsibilities may result too much disengagement, but demanding too much attention may result in fatigue and confusion. Compliance is a similar issue: the needs of the organization and the abilities of the individual need to be balanced.

Most end users conflate security with general functionality; they see a security failure as not significantly different from any other failure. This can be a problem, since some security problems do not cause failures; this can also be a good sign that it is possible to motivate users to act.

While user behaviors for more widely spaced tasks, such as changing passwords and updating virus scanner profiles, were problematic, a number of security ensuring behaviors observed that are more regular and common. Examples include locking the computer when away, keeping passwords secret, and protecting the physical security of the computer.

A standard approach to ensuring more literate end users is to expand the computer literacy curricula that people are exposed to in their work settings. Informal learning (learning that happens during the enactment of everyday behaviors and problem solving) should be recruited to enhance users’ EUSM practices; this type of learning is more amenable to constantly changing information (like EUSM), and easier for organizations to leverage than formal training, due to lower cost and higher value.

Users have no way to understand what is happening within a computer or a computer network; even if they do, they have no interest in a computer’s internals. This leads at times to an inability to distinguish between different problem etiologies. This may be due to willful ignorance, frustration, or limited mental models of computer functionality and failure, but is likely to be caused at times by all three.

The general problem of software invisibility is both a benefit and a drawback to computing. Users are not interested in technical details, and when they are forced to understand these details to operate a computer properly.

As much as users are to get down to these technical details, policies must be designed, implemented and governed by a legal system can greatly deter crimeware.

3.2 Law and Order in Cyberspace

What is the identity nationality into cyberspace? Given the technologically evolving nature of cybercrime, this is a perfectly understandable attitude and a respectable position to take if we are willing to ignore, accept, and perhaps even help perpetuate confusion surrounding the dynamic technological aspects of computer-related crime and its effects throughout society.

The state of empirical research and general body of literature pertaining to these issues are extremely limited. Virtually no studies of a general population have yet been published that take into account the theoretically complex phenomenon of cybercrime, operationalized as activities in which computers or other electronic IT devices are used to facilitate illegal or socially abusive behaviors via the Internet.

Estimates of financial losses due to technology abuse are considered, and how much underreporting by corporations (and other types of victims) is assumed[13]. It has also long been known that most companies are reluctant to report cybercrimes to law enforcement for fear of losing public confidence.
or exposing weaknesses in their information security[14]. Other major reasons for the lack of data and research on crimes committed with computers or other types of IT devices include:

1. The priority given to violent crimes by public officials and victim advocacy organizations
2. Relatively indifferent attitudes of many police and prosecutors to nonviolent crime
3. The technical and intimidating nature of crimes committed with computers and other IT devices
4. Easy treatment of cybercriminals by courts even when prosecuted
5. The inadequacy of training for police officers and security professionals to recognize and respond to, and adequately investigate cybercrimes, and
6. The ever-changing capabilities of computers and those who abuse them.

It is generally believed, for example, that a substantial portion of transnational financial crime constituting economic and security threats to many regions throughout the world originates in former Soviet Bloc countries such as Ukraine, Russia, and the Baltic States. Such crimes can be attributed to a number of factors:

1. Organized crime groups looking for ways to broaden their of influence
2. The lack of a mature model for policing such crimes and the resources to support it,
3. The resulting lack of fear of apprehension and retribution
4. The vast pool of potential victims, and
5. The ease with which financial resources can be transported around the globe.

3.2.1 Laws and Regulations in Cyberspace

Laws and regulations are extremely important because to some extent they affect virtually every aspect of administering criminal justice and increasingly to managing the security of data, information systems, and facilities that operate critical information infrastructures.

Laws and regulations can be thought of as tools governments’ uses to maintain societal order, productivity, and well-being. This situation is exacerbated by cybercrime and information security being international concerns shared by many nations having different legal systems, traditions, and laws.

It is important to understand from the onset that construction and enforcement of cyber-related laws and regulations is rapidly evolving in ways that cannot be fully anticipated, and that analysis of these issues depends on understanding the challenges computing and telecommunications technology pose relative to legal standards of behavior, and with respect to due care and diligence in the design and management of information security products and systems.

Most people agree that an individual, who sends a worm, Trojan, or virus into cyberspace, wreaking worldwide havoc on information systems and infrastructures, should be punished severely.

Matters of law arising from using, misusing, and even abusing computers, other electronic IT devices, and information systems are not as clear cut. The concept of law has various definitions and applications but is essentially a set of rules that defines standards of behavior and use of technologies by and among individuals in society.

Laws and accompanying regulations reflect the values, needs, and beliefs of the members of a society and they are designed to provide for the continuity of society, as well as safe, predictable, and reliable relationships among its members under changing conditions.

Laws are also intended to deter people from committing wrongs; to protect individuals, groups and organizations from harm; and to inform members of society of desirable behaviors and preferred
courses of action as well as punishments for not voluntarily complying with society’s rules.

Laws and regulations pertaining to illegal use of computers or other electronic IT devices are designed for the protection of everyone in computerized societies. Cyber laws and regulations also provide for ranges of penalties and compensation that may be owed to victims of cybercrimes. They also typically specify legal remedies for victims of computer abuse or cybercrimes to seek damages for losses associated with damaged, manipulated, stolen, or destroyed data or information systems.

The trans-border nature of cybercrime presents additional conditions and challenges to concepts of jurisdiction. Administering justice in international cases is predicated on identifying, and charging or arresting defendants suspected of committing crimes and/or civil wrongdoing and finding a court system with legal authority to bind litigants to its authority that is also cost-effective and geographically accessible. This can be enormously difficult.

Cyber society lacks all of the requisite attributes of a state as outlined in the 1933 Montevideo Convention on the Rights and Duties of States, which specifies that a bona fide country (society) must have a permanent population, a defined territory, a government, and the capacity to enter into relations with other states.

This definition implies that cyber society is not a distinct and sovereign state. Rather, it is a composite of loose associations that transcend traditional geopolitical nation-state boundaries. Jurisdiction in cyberspace is therefore contingent on the laws of individual countries that govern human activities in physical places.

International agreements about managing crime are usually very difficult to establish because nations often have very different views as to what constitutes justice.

In 1983, the Committee of Experts on Computer Related Crime of the Paris-based Organization for Economic Cooperation and Development (OEDC) became the first international group to study what could be done to prevent, control, and reduce the harmful impact of computer-enabled abuse and crime internationally[15].

The committee’s report, released in 1986, specified several types of computer-enabled activities that countries should consider making illegal, such as computer-enabled fraud and forgery, unauthorized alteration of computer programs or data, interception of communications, theft of trade secrets, and computer hacking – activities that constitute Crimeware. The report also noted several barriers to international management of computer-enabled crime problems, including:

1. Lack of global consensus regarding a legal definition of criminal conduct and by extension behavior constituting computer-related crime,
2. Lack of expertise of police, prosecutors, and the courts to understand computer-enabled activities in relation to existing national laws and legal principles,
3. Inadequacy of international legal powers to investigate and access computer systems and seize intangible data needed as evidence in national criminal hearings,
4. Inconsistent bodies of national laws applicable to computer-related crime matters and for guiding, supporting, or mandating investigation of computer-related crimes, and
5. The transnational character of many computer crimes coupled with lacking extradition and mutual assistance treaties, plus the inability of existing treaties to take into account the dynamic and special requirements of computer-enabled crime investigations and prosecution.

In this report the OECD committee called attention to the world’s enormous computer-enabled abuse and crime problem, made worse by limited and incongruent legal mechanisms among nations for preventing and controlling such high-tech crimes.
All these attempts to create international agreements to manage cybercrime have been more or less effective. Clearly they indicate that progress is being made toward overcoming legal barriers to investigating and prosecuting cybercrime internationally.

This clearly indicates that laws deter crime but a trusted level of computing is needed to take care of the technical details that users have less interested in.

### 3.3 Trusted Computing

Trusted Computing technology partially addresses this question by providing a means for end-users (and third-parties) to derive increased confidence in the platforms with which they interface, as well as providing standardized mechanisms to protect user data and information from software attack[16].

Trusted Computing technologies can be used to impede the distribution, infection and execution of crimeware applications.

Cybercrime can broadly be defined as any crime that is facilitated or committed using a computer, network, or hardware device, where the computer, network or device may be the agent of the crime, the facilitator of the crime, or the target of the crime.

Irrespective of the actual motivation for such activity, a crimeware attack, or more generally a malware attack, must typically pass through three stages to fulfill its goal. These are distribution, infection and execution.

1. **Distribution**: Distribution refers to the means by which malware arrives at a platform.
2. **Infection**: Infection is the process by which malware penetrates a platform.
3. **Execution**: It is during this stage that the malicious objectives of the malware are revealed. The malware may attempt to gain unauthorized access to information, capture user-entered details or steal proprietary data. This data is collated by the crimeware and transmitted back to the attacker for processing.

A Trusted System is one that will behave in a particular manner for a specific purpose. The Trusted Platform Module (TPM) specifications form the core of all Trusted Computing implementations. These specifications describe a microcontroller with cryptographic coprocessor capabilities that provides a platform with the following functionality: A number of special purpose registers for recording platform state; a means of reporting this state to remote entities; secure volatile and non-volatile memory; random number generation; a SHA-1 hashing engine; and asymmetric key generation, encryption and digital signature capabilities[16].

Trusted Computing has become synonymous with three fundamental concepts: Integrity measurement and storage, attestation, and protected storage. However, recently the definition of what constitutes Trusted Computing functionality has been revised and extended to incorporate the concepts of secure boot and software isolation.

An integrity measurement is the cryptographic digest or hash of a platform component. Platform attestation enables a TPM to reliably report information about the current state of the host platform. Attestation provides a powerful technique to combat crimeware distribution and infection. A platform, upon requesting access to a company's intranet, may be required to demonstrate through attestation that it has up-to-date anti-virus software with the latest signature definitions, that its spam filters are operating correctly and that it has installed the latest OS security patches. Similarly, a client could request that a server attests to its operating environment prior to the disclosure of sensitive data.

Protected storage functionality uses asymmetric encryption to protect the confidentiality of data on a TPM host platform. The notions of binding and sealing are of fundamental importance to Trusted Computing. Binding refers to the encryption of data with a public key for which the corresponding private key is nonmigratable from the recipient's TPM. Sealing takes binding one step further. Sealing is the process by which sensitive data can be associated with a set of integrity metrics representing a
particular platform configuration, and encrypted.

A secure boot process extends the integrity measurement and storage functionality. During a secure boot, a platform's state is reliably captured, compared against measurements indicative of a trustworthy platform state and stored. If a discrepancy is discovered between the computed measurements and the expected measurements then the platform halts the boot process.

Secure boot functionality can detect the malicious or accidental modification or removal of security-critical software at boot time. Secure boot functionality could be used to prevent a maliciously modified server from helping to distribute crimeware; this would reduce the effectiveness of a server modification attack to a denial of service. Secure boot functionality is not currently described as part of the TPM specifications. However, there is a TCG specification describing how it can be enabled on a trusted mobile platform.

An isolated execution environment, independent of how it is implemented, should provide the following services to hosted software:

- No interference: Ensures that the program is free from interference from entities outside its execution space.
- Trusted path: Ensures the presence of a trusted path between a program and an input device.
- Secure inter-process communication: Enables one program to communicate with another, without compromising the confidentiality and integrity of its own memory locations.
- Non-observation: Ensures that an executing process and the memory locations it is working upon are free from observation by other processes.

Hardware-enforced software isolation enables the segregation of security-critical software and data so that they cannot be observed and/or modified in an unauthorized manner by software executing in parallel execution environments. Additionally, the presence of isolated execution environments can ensure that any infection is contained within the execution environment which the crimeware has infected. In this way, user I/O data can be secured in transit to protect it from crimeware, such as keyloggers, which may have infiltrated the platform.

Trusted Computing mechanisms may also be used as a means of enhancing crimeware functionality. For example, Trusted Computing provides trivial means for crimeware to launch denial of service attacks, of various types, against a platform. If a crimeware application can be installed on a platform, thereby altering the system state, then first, access to networked services may be denied, since the platform will (correctly) not be considered trustworthy. Second access to data may be denied, since the current state of the platform will not match the integrity metrics to which the data has been sealed. Third, system startup may be suspended if the presence of crimeware is detected during the boot sequence.

Trusted Computing, as currently deployed, can do little to protect an end-user platform from crimeware attack. Trusted Computing, as it currently stands, provides a limited, but useful, set of cryptographic functionality. Problems associated with software vulnerabilities will not be ameliorated by the presence of Trusted Computing.

An abundance of information exists on the potential positive applications of Trusted Computing. However, as the technology becomes more widely deployed, it seems likely that Trusted Computing functionality will be increasingly targeted by crimeware.

Today's users are accustomed to installing all kinds of plug-ins on their client systems, so distributing a malicious plug-in wouldn't be particularly difficult. All an adversary would need is a good (and
reasonable-looking) cover story to have the user install the plug-in or otherwise exploit software vulnerability. Many social-engineering techniques can be used for this purpose.

A combined use of these mechanisms can greatly mitigate these attacks. As networked computing has become pervasive, the challenges of maintaining secure environments for end-user computing have increased enormously.

3.4 Secured Communication

By shrinking time and distance, the Internet has accelerated globalization, connecting people and businesses worldwide. The Web is emerging as the dominant interface for information exchange and service delivery, and e-mail is becoming the communication tool of choice. At the same time, however, there is a growing perception of the Internet as an insecure environment, and these concerns may prevent the Internet from realizing its seemingly limitless potential. The recent proliferation of malware—including viruses, worms, Trojan horses, spam, phishing schemes, distributed denial-of-service (DDoS) attacks, spyware, and adware—has made the Internet a harrowing experience for many individuals and a severe headache for organizations[17].

Ironically, what has made the Internet so successful—its open and decentralized structure—is also what sustains malicious online activity. Information on newly discovered vulnerabilities propagates quickly, and tools to launch ever more sophisticated attacks are readily accessible.

Controls implemented by Internet service providers (ISPs), which are interested in protecting their own network—and customer base—from external attacks, predominantly target inbound traffic[9]. However, there is no similar economic incentive to control outbound traffic, as the potential damage is to other networks. This lack of clear lines of accountability derives from both the decentralized nature of datagram routing in the Internet as well as its decentralized organizational structure.

What the Internet needs is an institutional structure that strongly motivates ISPs, network service providers, equipment vendors, and users themselves to control attacks at their origin as well as to maintain security on a dynamic basis. One way to accomplish this is to introduce a certification mechanism that induces service providers to voluntarily accept some degree of accountability, without interfering with the underlying decentralized protocols.

Such a mechanism could propagate incentives through the network, ensuring that distributed participants coordinate their efforts to increase security as well as reduce congestion.

To improve Internet security, it is essential that service providers control outbound as well as inbound traffic. Outbound traffic control stamps out attacks at the source and thus stops them from spreading, without subjecting the network to congestion. Outbound control is especially effective when done by ISPs, which can leverage the direct relationship with their customers to hold them accountable and take punitive action against violators.

The most effectively and efficiently way to secure the Internet is to block malware as it leaves a network. Some ISPs and e-mail providers do curtail outgoing malicious content, but the practice must be universal or nearly universal to work[17].

3.4.1 Service Provider Certification

This is a security mechanism for service providers based on the notion of a certifying authority (CA). Membership in the scheme is voluntary: Providers that choose to join pay a subscription fee to the CA and are called certified providers, while those who opt out are known as noncertified providers; traffic originating from each type is labeled similarly. The CA requires certified providers to compensate: remote providers that receive malicious traffic from the certified providers’ users; and their own customers who receive malicious traffic, regardless of the source.

The CA holds any certified source provider accountable for an attack originating from its domain, regardless of whether a human customer or a zombie node initiated the attack. To minimize
compensation payments, certified providers are motivated to filter all outgoing traffic.

This scheme encourages service providers to take up an offensive, rather than a defensive, posture against intrusions and spurious traffic that exploit the Internet’s distributed structure. Tying penalties to the source of malicious activity shifts the responsibility for security to the originating provider. This is particularly important for attacks that rapidly escalate by replicating and targeting multiple destinations. The certification mechanism currently applies only to service providers supplying Internet connectivity, whether to residential or enterprise customers.

The effectiveness of inbound traffic control depends on the ability to correctly identify the source of incoming packets. This proposed certification mechanism will reduce such attacks, which rely on backbone networks for transportation, by motivating access networks to implement their own outbound traffic controls.

The game-theoretic analysis is used to evaluate the viability of this incentive mechanism and to determine its implications in terms of provider actions and collective security. The game-theoretic analysis uses the concept of the Nash equilibrium to characterize the likely choice of strategies by agents. In a Nash equilibrium, each agent chooses the best response to strategies that other agents employ, implying that agents’ expectations are mutually correct and that they act rationally based on these expectations.

Following standard economic theory, service providers are divided into two types, those with a low-risk security profile (A) and those with a high-risk security profile (B), with the distribution of types but not the classification of individual providers being common knowledge. A high-risk profile indicates that the provider’s customer base is more prone to sending out malicious traffic, either intentionally or by having less securely configured machines.

This analysis indicates that with a nonprofit certifying authority, all service providers will choose certification, leading to a net increase in system surplus. With a profit-maximizing CA, different equilibria may exist depending on the proportion of A-type providers in the network. When this proportion exceeds a certain threshold, only A-type providers get certified, leading to a separating equilibrium. When the proportion of A-type providers is below the threshold, all providers subscribe to the certification scheme, leading to a pooling equilibrium. These are Nash equilibria, wherein each participant chooses the best response to others’ actions.

In the case of a separating equilibrium, the CA has enough A-type subscribers that it can afford to exclude the B-type providers by setting high subscription fees. For B-type providers, these fees, combined with the expectation of costly penalties due to the number of A-type providers that potentially receive traffic from them, makes certification prohibitively expensive.

In the case of a pooling equilibrium, the CA sets the subscription fee low enough to induce everyone to join, the high number of subscriptions compensating for the low fee. B-type providers choose to be certified along with A-type providers because they can draw on a potentially larger pool of providers for compensation, their need for insurance is lower, and their customers stand to benefit from being able to communicate with customers of all types.

This voluntary participation-based incentive mechanism achieves the benefits of segregation through economic means, robustly supporting communications without interfering with the Internet’s basic structure.

Failure to combat the growing scourge of malware could lead to real fragmentation of the Internet. Service provider certification improves overall security without undermining the fundamental design philosophy of the Internet as an open, decentralized network.

Some certification service providers, CSP assessors, and browser manufacturers have jointly defined the notion of an extended validation SSL certificate[18]. An EV-SSL certificate is a normal server certificate that meets more stringent security requirements. As such, it aims to tell more reliably
whether the certificate is valid and meaningful in a given context. It’s too early to tell whether EV-SSL certificates will significantly improve the level of protection against online channel-breaking and MITM attacks. Mainly because EV-SSL certificates may replace normal certificates and inherit their shortcomings and problems in everyday life.

4. CONCLUSION

This clearly brings out the fact that computers do not commit a crime, but we (humans) do![6] Security researchers have made tremendous progress in keeping pace with Internet threats, but there are limits to what technology alone can accomplish.

To effectively mitigate crimeware both technical and social aspects must be combined to form this layered defense framework. In the first layer, end users should be more literate by expanding the computer literacy curricula in their work setting. In the second layer, we find that the Internet has obscured the question of national sovereignty and jurisdiction, but a legal system is needed to control and hence deter crime. Trusted computing that constitutes the third layer increases the confidence in the platforms with which end users interface as well as provide a standardized mechanism to protect user data and information from crimeware. Finally, the growing proliferation of crimeware is raising doubts about the Internet’s future. To improve Internet security, it is essential that service providers control outbound as well as inbound traffic. This forth layer, offer a certification scheme to motivate providers to control outgoing traffic, thus efficiently increasing overall security while preserving the Internet’s open, decentralized structure.

5. REFERENCES


CANVASS - A Steganalysis Forensic Tool for JPEG Images

Jennifer L. Davidson
Department of Mathematics
Iowa State University, Ames, IA 50011
Phone: (515) 294-0302
Fax: (515) 294-5454
davidson@iastate.edu

Jaikishan Jalan
Department of Computer Science
Iowa State University, Ames, IA 50011
jai.ism@gmail.com

ABSTRACT

Steganography is a way to communicate a message such that no one except the sender and recipient
suspects the existence of the message. This type of covert communication lends itself to a variety of
different purposes such as spy-to-spy communication, exchange of pornographic material hidden in
innocuous image files, and other illicit acts. Computer forensic personnel have an interest in testing for
possible steganographic files, but often do not have access to the technical and financial resources
required to perform steganalysis in an effective manner. This paper describes the results of a funded
effort by a grant from the National Institutes of Justice to develop a user friendly and practical
software program that has been designed to meet the steganalysis needs of the Iowa Division of
Criminal Investigation in Ankeny, Iowa. The software performs steganalysis on JPEG image files in
an efficient and effective way. JPEG images are popular and used by a great many people, and thus are
naturally exploited for steganography. The commercial software that is available for detection of
hidden messages is often expensive and does not fit the need of smaller police forensic labs. Our
software checks for the presence of hidden payloads for five different JPEG-embedding steganography
algorithms with the potential of identifying stego images generated by other (possibly unknown)
embedding algorithm.

Keywords: steganography, steganalysis, JPEG images, GUI software

INTRODUCTION

Steganography is the practice of communicating a hidden message in such a way that no one, apart
from the sender and intended recipient, suspects the existence of the message. The goal of
steganography is to embed a payload into a cover object to obtain a stego object in such a way that the
presence of hidden information cannot be detected by either perceptual or statistical analysis of the
stego object. The counterpart of steganography is steganalysis. The main goal of steganalysis is to
identify whether a given object has a payload embedded in it. Other information about the payload is
often sought, including identification of the steganography algorithm, estimation of payload length,
recovery of the payload, or obliteration of the payload.

With the advent of digital media and the Internet, multimedia objects such as still images and videos
have become popular and are shared easily. Image and video data make a good choice for hiding
payload. These objects are readily available and their broad presence on the Internet makes it difficult
to check each one for hidden payload and thus difficult to detect the use of steganography. A single
image can hold a reasonable amount of information, and a video file can hold more. In addition, there
is a plethora of freeware available for hiding secret information, as can be seen by visiting the site
stegoarchive.com (Stegoarchive, 1997). MSU StegoVideo is a video steganographic tool that is freely
available online (MSU, 2004). In this paper, we restrict steganalysis of image data to Joint Photographic Experts Group (JPEG) format because of its wide use in consumer cameras and on the Internet. It also has the advantage of low bandwidth for storage and transmission, unlike raw or other uncompressed formats.

There is a growing concern within the community that steganography is being used for illicit purposes. The USA Today (Kelly, 2001), the New York Times (Kolata, 2001), and the United States Institute of Peace (Weimann, 2004) have reported that terrorists may be using steganography and cryptography on the web as a means of covert communication. A recent report from National Institute of Justice encourages investigators to look for steganographic information while dealing with child abuse or exploitation and terrorism cases (NIJ, 2008). A more recent online report from the New York Times (Kerbaj, 2008) reported that a raid on computers from a terrorist group produced child pornographic images that contained secret messages hidden in them. These reports have lead local police departments to be concerned about the use of steganography for crimes committed within their local jurisdictions. For example, the Iowa Crimes Against Children Taskforce (ICAC) in Ankeny, Iowa, has expressed such concerns. This paper presents results of the authors collaborating with ICAC to address their steganography detection needs. We found that while steganalysis algorithms abound in the academic literature, there are few software programs that address the needs of local police departments who perform computer forensic functions for steganalysis. Here we describe Canvass, a software package that has been developed to make our research accessible to the Iowa Department of Criminal Investigation forensic lab.

There are several major stego-detection tools in existence today. StegoSuite is a commercially available software program developed by WetStone Technologies for the U.S. Air Force (Wetstone, 2010). StegoSuite performs a variety of steganography detection actions including previously installed software, image filters, and other features. The cost for a single user license is approximately $1495. Another group of commercial software is available through Steganography Analysis and Research Center in West Virginia. For example, StegAlyzerRTS is an advanced data leak protection software product that is capable of detecting the use of digital steganography in real time (SARC, 2010). It scans files entering and leaving a network for signs of steganography applications. It is also very expensive: as of March, 2009, the price was listed as $14,995.00 (PublishersNewswire, 2009). Neither of these software products presents the product’s false alarm rates of detection. A false alarm occurs when an innocent image (without hidden content) is flagged as a stego image erroneously. High false alarm rates can lead to manual inspection by a computer forensic analyst, which can require large amounts of time by the human. Thus, a low false alarm rate is necessary for practical use in forensic labs. A third stego-detection software is the freeware StegDetect, developed in 2001 by Neil Provos (Provos, 2001) to perform steganalysis on suspected stego images. His experiment was to steganalyze millions of JPEG images from sites like eBay and USENET (Provos & Honeyman, 2002; Provos & Honeyman, 2003) to determine if his program could detect hidden content. Not a single image with hidden data in it was detected. False alarms rates for the experiment performed with Stegdetect ranged from less than 10 percent to more than 20 percent. In short, commercial software is often too expensive for local police departments while at the same time, reliable false alarm rates of the software are not available for evaluating practical use.

It is of course desirable to have high detection rates of stego images (high true positive rates) while keeping the false alarm rate low, and this is the area of performance that CANVASS is focused to addressed. In this paper, we present true positive rates for the Canvass steganalyzer using a public steganography database. A true positive rate equals one minus the false positive alarm rate, discussed in more detail below.

The main goal of this research was to develop a software package that addressed the needs of local police departments who perform computer forensic functions for steganalysis. The authors collaborated with personnel at the state of Iowa’s Division of Criminal Investigation lab in Ankeny,
Iowa, to develop a user-friendly software package. The lab required software that was easy for computer forensic personnel who are not experts in steganalysis to use. After numerous meetings with the Internet Crime Against Children (ICAC) workforce members, we developed Canvass as a cross-platform software that specifically addressed the image forensic application of steganalysis for the lab. It is designed with an intuitive graphical user interface that implements a state-of-the-art steganalyzer, discussed below briefly and in more detail in (Davidson & Jalan, 2010). The “brain” of the steganalyzer is a classifier designed to have low false alarm rate on known testing data; indeed, most of the research time was spent on developing a state-of-the-art pattern classifier that produced simultaneously higher true positive rates and low false alarm rates. The testing data was drawn from a collection of standard image databases (see (Davidson et al., 2010)) used by the steganalyzer community for steganalysis development. A table of false alarm rates calculated from more than 115,000 different images is available in Canvass, where many classifications have a low false positive rate.

The remainder of this paper is organized as follows. In “Steganography and Detection,” we give a short description of how steganography operates and how the detection in JPEG images is performed. In “Canvass Steganalyzer,” a description of the classifier that distinguishes between stego and cover images is given. In the section, “The Graphical User Interface (GUI) of Canvass,” the software package is described from a user’s perspective. We then discuss our findings and point out future directions of our research in the Conclusions section. The last three sections give Acknowledgements, Author Biographies, and the References.

STEGANOGRAPHY AND DETECTION

Written records of humans communicating covertly in plain sight go back over 2400 years ago, to the Greek Herodotus. His written records titled “The Histories” (de Selincourt, 1996) document the use of slaves to send messages secretly tattooed under scalp hair grown out to cover evidence of the tattoo. The tattoo was shaved off to reveal the hidden message. More recently, microdots and DNA have been used for covert communications (White, 1992; Clelland, Risca, & Bancroft, 1999). The explosive use of steganography in digital media in the past 15 years has occurred because of the ease with which it is possible to hide files in images. Hiding in audio files such as .mp3 and .wav is also relatively easy. In particular, JPEG formatted image data is easy to get and easy to embed in, using readily available freeware. For this reason, our efforts are focused on JPEG image data, and our software processes only JPEG image data.

A steganography algorithm consists of two parts: the embedding algorithm and the extraction algorithm. The embedding algorithm takes the message and alters a “cover” image in such a way that the message is part of the image file but invisible to the eye. This produces a stego image. The extraction algorithm takes the stego image, and inverting the process of embedding, recovers the hidden message. Knowledge of the original image is not typically necessary for proper extraction of the message. If the alterations to the original cover image are relatively small, it is impossible to notice visually that changes have been made to an image. In practice, the sender destroys the cover image after generating the stego image and therefore it is not possible to get access to the cover image. Thus, if a third party such as a computer forensic investigator would like to determine if a possible image contains a hidden message, other more sophisticated techniques than visual inspection must be used.

There are several different approaches for detection of hidden data. One way would be to inspect a computer disk for steganography freeware-related files, the presence of which may lead to more thorough searches for stego images. This in fact can be a very good indicator of the use of steganography freeware (Zax, 2009). Another way is to produce a hash file for a known stego image, and store the hash values in a database along with other hash values of known stego images. This will work if specific images are being passed from user to user, but will not work if a user creates a new stego image whose hash value is not in the database. A third way is to perform blind detection on an image. A blind detector takes an image as an input and produces a YES (stego) or NO (innocent)
Most successful blind detectors use sophisticated pattern classifiers that extract statistical measures from known examples of stego and cover image data to generate “signatures” able to distinguish between the two classes, stego and cover images. The steganalyzer in Canvass has such a classifier, based on a state-of-the-art set of statistical measures (Davidson et al. 2010). The use of the classifier is transparent to the user, and thus the user does not have to know the inner workings of the steganalyzer.

How does steganography embedding work? Most steganographic algorithms embed bit values into an image. Thus, the digital file to embed must be represented as a bit stream of zeros and ones. Since all digital data on a computer is stored in this way, any file in theory can be embedded into an image. The basic process is as follows. An image consists of gray values at pixel locations. The gray values are integers represented by a certain number of bits. Many images use 8 bit integers to represent their colors. The first bit of the message is compared with the least significant bit (LSB) of the first gray value at a specified pixel location in the image. If the bit values match, then nothing is done and the second bit of the message is compared with the LSB of the gray value at the next specified pixel location. If any message bit does not match the LSB of the current gray value, then the gray value’s LSB is changed to match the message bit. Each message bit is thus “embedded” in this manner until there are no more message bits or there are no unused pixel locations left. Extraction involves the reverse process: the scanning order of the embedding process is followed and the LSB of each gray value is inspected and written down. The bit insertion process is pictorially represented in Figure 1.

Knowing the scanning order for embedding plus any keys used to encrypt the bit stream allows that user to extract and decrypt the message. However, this information is not usually available to anyone except the sender and receiver. Changing enough bits can, however, render the stego image statistically different from the original cover image, and it is this property that permits pattern classifiers to be developed that can differentiate between stego and cover images. The interested reader is guided to (Provos et al. 2003) for more detailed information on this topic.

In the case of JPEG images, the situation is slightly more complicated. The values that are changed to match the message bits are the quantized discrete cosine transform (DCT) coefficients, the important information representing the visual content of the image. They are integers that reside in a transform domain (of the discrete cosine transform) and are subsequently Huffman encoded in a lossless format to represent the image data in a compressed format. The representation of the image data using the quantized DCT coefficients allows much of the redundant information to be discarded, and thus allows a compressed representation of the original data to be stored typically with many fewer bits. The
interested reader is directed to (Bhaskaran & Konstantinides, 1995) for further details of the use of the DCT in image compression. The pertinent information for steganalysis is that statistical analysis of the distributions of these quantized DCT coefficients can offer very good steganalysis detection and the steganalyzer in Canvass uses this approach.

CANVASS STEGANALYZER

As mentioned previously, Canvass uses a pattern classifier as the “brain” of the software. This is a sophisticated software program that is trained on known examples of stego-images and innocent or cover images. The classifier used in Canvass is the Support Vector Machine (SVM) (Vapnik, 1995) and provides very good classification results in steganalysis. There are two stages to using an SVM: 1. Training and 2. Testing. In the training stage, input values in the form of features are processed by the SVM algorithm iteratively until an acceptable solution is found. In the testing phase, a possibly unknown image is input to the classifier and very quickly it is classified as one of two classes, stego or innocent. The training stage takes several hours to several days of computing time to find a solution and is typically compute-intensive.

The features that are used to train the SVM are crucial to the performance of the classifier. If the features extracted from the image are not representative of the class then the performance of the classifier will be poor in both phases. Thus, selection of good features is essential to good performance of the classifier. The features in Canvass are based on conditional probability density functions derived from modeling the image as a partially ordered Markov model (POMM) (Davidson, Cressie, and Hua, 1999). Using the difference of quantized Discrete Cosine Transform (DCT) coefficients in four pixel directions, the differences are modeled as POMMs and the empirical values of their probability density functions are used as input to the pattern classifier. More details can be found in (Davidson et al., 2010). The number of features used is 98. Another state-of-the-art steganalyzer in (Pevny and Fridrich, 2007) used the quantized DCT coefficients directly plus values from a Markov transition matrix representing differences of the coefficients. Their classifier required 274 features. The paper (Davidson et al., 2010) describes detection rates and false positive rates of both steganalyzers, with false alarm rates were typically between 0 and 10 percent for both Pevny’s and Davidson’s steganalyzers. Thus, the sophisticated and accurate POMM steganalyzer developed over many months of research effort was used in the core of Canvass for classification. As mentioned earlier, it is important not only to have high detection rates of stego images, but low false alarm rates of innocent images.

Once trained, Canvass detects the presence of five different embedding algorithms: Jsteg-jpeg (Upham, 1995), Outguess (Provos, 2001), F5 (Westfeld, 2001), Steghide (Hetzl, 2003), and JPHide (Latham, 1999). The general detection process is described pictorially in Figure 2. The suspect image is input to a steganalyzer. The steganalyzer extracts the statistical features called a feature profile, and inputs the features to Canvass. The feature values are used by the classifier to produce a YES (stego) or NO answer. Figure 2 shows a sample statistical profile which consists of feature number and feature value extracted by a steganalyzer. In the case of Canvass, an embedding algorithm that was likely used is also identified in the case that the classification of the image is stego.
A classifier has an associated false alarm rate, which is calculated on known classes of data. A false positive occurs when the steganalyzer classifies a cover image as a stego image. A true positive occurs when a stego image gets classified as a stego image. Obviously, the lower the false positives the better off the forensic analyst is. After using 106,571 images for training the SVM classifier for Canvass, a set of 115,603 different test images was used to produce the Confusion matrix shown in Table 1. Each pair of classes has an SVM that chooses between the two classes. The six classes are: cover, Jsteg, Steghide, Outguess, F5, and JPHide. There are a total of 15 such binary classifiers, and the best answer is chosen by majority vote of the 15 answers. This approach gives reasonable results and is applied to the Canvass steganalyzer.

Table 1. Confusion matrix for Canvass's true positive rates expressed in percentages.

<table>
<thead>
<tr>
<th></th>
<th>Cover</th>
<th>Jsteg</th>
<th>Outguess</th>
<th>F5</th>
<th>Steghide</th>
<th>JPHide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover</td>
<td>86.78</td>
<td>0.29</td>
<td>0.15</td>
<td>5.69</td>
<td>3.96</td>
<td>4.58</td>
</tr>
<tr>
<td>Jsteg</td>
<td>0.05 bpc</td>
<td>2.09</td>
<td>97.24</td>
<td>0.84</td>
<td>0.13</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>0.10 bpc</td>
<td>0.19</td>
<td>99.78</td>
<td>0.03</td>
<td>0</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.20 bpc</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.40 bpc</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OutGuess</td>
<td>0.05 bpc</td>
<td>1.86</td>
<td>0.54</td>
<td>90.38</td>
<td>1.11</td>
<td>8.17</td>
</tr>
<tr>
<td></td>
<td>0.10 bpc</td>
<td>0.02</td>
<td>0.08</td>
<td>99.63</td>
<td>0.05</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>0.20 bpc</td>
<td>0</td>
<td>0.02</td>
<td>99.98</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F5</td>
<td>0.05 bpc</td>
<td>44.64</td>
<td>0.17</td>
<td>0.19</td>
<td>42.12</td>
<td>5.53</td>
</tr>
<tr>
<td></td>
<td>0.10 bpc</td>
<td>4.94</td>
<td>0.10</td>
<td>0.13</td>
<td>86.22</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>0.20 bpc</td>
<td>0.19</td>
<td>0.03</td>
<td>0.05</td>
<td>99.26</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.40 bpc</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>99.56</td>
<td>0.08</td>
</tr>
<tr>
<td>Steghide</td>
<td>0.05 bpc</td>
<td>21.88</td>
<td>0.29</td>
<td>0.91</td>
<td>3.61</td>
<td>73.58</td>
</tr>
<tr>
<td></td>
<td>0.10 bpc</td>
<td>7.73</td>
<td>0.32</td>
<td>1.20</td>
<td>1.94</td>
<td>89.17</td>
</tr>
<tr>
<td></td>
<td>0.20 bpc</td>
<td>1.01</td>
<td>0.19</td>
<td>1.18</td>
<td>0.40</td>
<td>97.30</td>
</tr>
<tr>
<td></td>
<td>0.40 bpc</td>
<td>0.05</td>
<td>0.07</td>
<td>0.94</td>
<td>0.03</td>
<td>98.94</td>
</tr>
<tr>
<td>JPHide</td>
<td>0.05 bpc</td>
<td>30.08</td>
<td>0.14</td>
<td>0.07</td>
<td>3.43</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>0.10 bpc</td>
<td>28.99</td>
<td>0.15</td>
<td>0.08</td>
<td>4.12</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>0.20 bpc</td>
<td>3.61</td>
<td>0</td>
<td>0.03</td>
<td>3.85</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>0.40 bpc</td>
<td>0.03</td>
<td>0</td>
<td>0.03</td>
<td>0.59</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Different amounts of payload (hidden content) were embedded into the images, expressed in bits per nonzero coefficient, or bpc. This is an average of amount of bits embedded over all the possible bits allowing embedding. The confusion matrix gives the percentages of images that were correctly classified into the proper category: Cover, Jsteg-jpeg, Outguess, Steghide, and JPHide. If an image was misclassified into a different category then it appears as part of the percentage of all images that were misclassified in that category improperly. For example, 0.19 percent of the 5,936 images embedded with 0.05 bpc using the F5 algorithm were improperly classified as being embedded using Outguess. These numbers are as good as or better than the state-of-the-art steganalyzer in (Pevny et al., 2007). The false positive rate for Outguess embedded at 0.10 bpc is 100 – 99.63 = 0.37%. Thus, many of the image classes have low false positive rates. Note that although many classes have low false positive rates, a classifier cannot be trained on the universe of jpeg data and so there will always be some error. The goal is to use as large a database of training images as possible to find the “best” solution to the classification problem.

THE GRAPHICAL USER INTERFACE (GUI) OF CANVASS
Canvass was written in Java to provide complete portability to different platforms. It provides the following features to the user:

1. Ability to process multiple images with one command either on a computer or at a specified website.
2. Display of processing information in real time. It shows a variety of information such as which steganography algorithm was likely used for embedding.
3. An option to save the processing information at any time.
4. It displays the current image for visual inspection.
5. It has the ability to run on multiple platforms.

A Model-View-Controller (MVC) architecture has been used to design this software. Because of this, it can be easily extended using a different steganalyzer from the backend, for example, if additional binary classifiers are added to extend the ability to classify other embedding algorithms or even using other feature sets as inputs. This software will be made available from the Midwest Forensic Resource Center, Ames Laboratory for limited distribution to recognized police departments, after June 1, 2010.

CONCLUSIONS AND FUTURE DIRECTIONS
The software package Canvass has been developed to address the limited steganalysis needs of the Iowa Division of Criminal Investigation. The GUI is easy to use and effective in processing many image data in batch mode. The steganalyzer itself is designed to detect innocent images and stego-embedded images from five different algorithms - Jsteg-jpeg, Outguess, F5, Steghide, and JPHide. Since it uses a blind steganalyzer, Canvass has the possibility of detecting other similar embedding algorithms that use the JPEG domain. There are several directions for improvements. One is to add the detection of more embedding algorithms by training more binary SVMs and including them as part of Canvass’s steganalyzer. Another extension of Canvass’s capabilities would be to include a system scan of the computer to look for file artifacts residing on the computer, which indicate stego software was used for embedding. Including capabilities for password attack on identified stego images from within Canvass could also extend the practical use of Canvass. Identifying double-compressed images accurately and then passing to appropriate classifiers would also be very useful.

Another area that could use improvement is the accuracy of low embedding rate. The detection accuracies are quite high when an appreciable fraction of the largest possible message is embedded, such as at 0.2 bpc and 0.4 bpc. However, for the lower rates of 0.05 bpc and 0.1 bpc, detection rates particularly for F5 and Steghide are quite low. Increasing the accuracy of the steganalyzer will require
research into creating a better pattern classifier, and will include looking for better features, and developing better classifiers models. Complementing the current POMM features with other different features could increase the classifier accuracies at the lower embedding rates. Addition of other features sets, such as those in (Pevny et al., 2007), might give better complementary detection. Another approach might be to estimate message length using the POMM. Since the POMM provides a closed form for calculating the joint probability density function, new techniques could be investigated by assuming a parameterized model of POMM to estimate the message length and maximizing the joint probability density function conditioned on parameters of the model. This could provide an indirect measure of message length, which in turn could be used to provide detection of embedding above a low threshold representing the message length. Another way to estimate the message length might be to use the current features and length of message embedded along with SVM regression, a variant of the support vector machine, to predict the length of messages in unknown images.

ACKNOWLEDGEMENTS

This work was funded by the National Institute of Justice, through the Midwest Forensics Resource Center at Ames Laboratory under Interagency Agreement number 2008-DN-R-038. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University, under contract No. DE-AC02-07CH11358. We would also like to thank Dr. Gwaneal Doerr for providing easy access to a rich database of JPEG images that facilitated the success of this research.

AUTHOR BIOGRAPHIES

Dr. Jennifer Davidson is the Associate Chair of the Department of Mathematics at Iowa State University. She has been performing research in image processing for the past 25 years and steganalysis for the past six years. She is a member of SPIE and is a coordinating editor for the Journal of Mathematical Imaging.

Jaikishan Jalan holds a Bachelor's degree in Computer Science and an M.S. degree in Computer Science from Iowa State University. His research interest lies in image processing, multimedia and information security.

REFERENCES


Digital Records Forensics: A New Science and Academic Program for Forensic Readiness

Luciana Duranti
School of Library, Archival and Information Studies
The University of British Columbia
IKBL 470, 1961 East Mall
Vancouver, British Columbia V6T 1Z1 CANADA
e-mail: luciana.duranti@ubc.ca

Barbara Endicott-Popovsky
Center for Information Assurance and Cybersecurity
University of Washington
4311 11th Ave NE Suite 400 Box 354985
Seattle, Washington 98105
e-mail: endicott@u.washington.edu

ABSTRACT
This paper introduces the Digital Records Forensics project, a research endeavour located at the University of British Columbia in Canada and aimed at the development of a new science resulting from the integration of digital forensics with diplomatics, archival science, information science and the law of evidence, and of an interdisciplinary graduate degree program, called Digital Records Forensics Studies, directed to professionals working for law enforcement agencies, legal firms, courts, and all kind of institutions and business that require their services. The program anticipates the need for organizations to become “forensically ready,” defined by John Tan as “maximizing the ability of an environment to collect credible digital evidence while minimizing the cost of an incident response (Tan, 2001).” The paper argues the need for such a program, describes its nature and content, and proposes ways of delivering it.

Keywords: digital records, records authenticity, graduate education, record theory, records forensics science, records forensic discipline, forensic readiness, Digital Records Forensics, digital preservation

1. INTRODUCTION
Two of the most challenging issues presented by digital technology to the law enforcement, records management, archival and legal professions, researchers, business, government and the public are the identification of "records" among all the digital objects produced by complex dynamic and interactive systems, and the determination of their "authenticity.” The first issue—the identification of digital records—is addressed by Digital Diplomatics, a contemporary development of a centuries-old discipline that studies the nature, genesis, formal characteristics, structure, transmission and legal consequences of records (Duranti, 1996, 1998). The second issue—the assessment of the authenticity of digital records—is only indirectly addressed by Digital Forensics, which is defined by Ken Zatyko as “the application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence after proper search authority, chain of custody, validation with mathematics, use of validated tools, repeatability, reporting, and possible expert presentation” (Zatyko, 2007). More specifically, the Digital Forensics Research Workshop, in 2001, defined “digital forensics” as “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 facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned
operations” (Digital Forensics Research Workshop, 2001).

The determination of the authenticity of individual medieval records of questionable provenance was the original reason for the development, in the 17th century, of the science of Diplomastics, and, in the context of the development of a Digital Diplomastics, its theory and methods have been successfully applied to contemporary digital records (Duranti, 2009a, Duranti, 2005; Duranti and MacNeil, 1997; Duranti, Eastwood and MacNeil, 2002; Duranti and Thibodeau, 2006). Thus, in several ways, the objects of study of Digital Forensics and Digital Diplomastics overlap and their methods of inquiry complement each other (Duranti, 2009b). At the same time, their perspectives are very different and the sum of their bodies of knowledge is not at this time able to address all the issues of ‘recordness’ and authenticity with which our legal system is constantly confronted, due to the extremely rapid obsolescence of information technologies and to the manipulability, mutability and fragility of the digital entities that these technologies produce and store, especially after those entities have been removed from the original system.

Thus, a team composed of diplomatics, archival science, information science, evidence law and digital forensics specialists has undertaken a research program, the purpose of which is to develop a new science called "Digital Records Forensics" (Digital Records Forensics Project, 2008-2011) by integrating the concepts and methods of all these bodies of knowledge. This integration will 1) enable those who need to assess the trustworthiness of digital records that no longer reside in the original system in which they were made or received and maintained to ascertain whether they are accurate and authentic, having preserved their original identity and integrity; 2) foster development of methods for maintaining the authenticity of these records over the long term, regardless of their format; 3) ensure that the Law of Evidence maintains an awareness of the changing nature of documentary evidence determined by digital technologies and adjusts its requirements and procedures to the changing characteristics of such evidence; 4) contribute to organizational forensic readiness as firms and agencies anticipate the need to support legal action with admissible digital evidence (Nevins, et.al., 2008; Endicott-Popovsky, et.al. 2007, 2005; Endicott-Popovsky and Frincke 2007a, 2006; Taylor, et.al., 2007); and 5) allow for the development of education programs forming professionals capable of acquiring, as well as creating, assessing, controlling and maintaining reliable, accurate and authentic records for as long as they are needed.

2. THE DIGITAL RECORDS FORENSICS PROJECT

The legal systems, both common and civil law, consider records to be a very special kind of documentary evidence. Records are defined in archival science as any document made or received in the course of a practical activity by a natural or an artificial person (or, physical or corporate, moral, or juridical person, depending on the country) and kept for action or reference. In civil law environments, a record is admissible as evidence in court simply on the basis of the recognition of its record nature. In common law environments, in addition to relevance, disputed records may require further steps to gain admissibility, such as proof of authenticity, and compliance with the best evidence and the hearsay rules. Thus, it is vital to establish clear and stable parameters for the identification of records among all the digital entities that may exist in a digital system, be it a document management system, a geographic information system, an assembly of separate applications, like e-mail, or any other form of information technology. This issue keeps coming up at trials and in political discussions. In an example, the British Columbia Rail case, where the judge pointed out that legislation speaks of preserving “records,” the Liberal MLA Ralph Sultan asked “What is the definition of a record?” referring “to the controversy over to what extent e-mails qualify” (Palmer, 2010). In another example, the Supreme Court of Canada is deciding whether hyperlinks in a text are akin to footnotes or make of the material to which they connect the reader a component of the document being read (Tibbetts, 2010).

At common law, to be admissible, records must constitute the “best evidence.” The best evidence rule requires that the original of any document, regardless of medium or form (e.g., a letter, a recording),
be used as evidence at trial. A copy will be allowed into evidence only if the original is unavailable for a legitimate reason. However, in the digital environment, we no longer have originals. In fact, we cannot keep digital records. We can only maintain our ability to reproduce or even to re-create them as needed. As a consequence, the authenticity of digital records is difficult to establish on the records themselves and becomes an inference that one draws regarding the integrity of the system (Electronic Transactions Act (B.C.), s. 8, 2001; Canada Evidence Act, s. 31.2(1), 1985).

Nevertheless, the law requires that an authentication of the record submitted as evidence is made by a competent third party who either recognizes the record, if it is an original, having seen it before, or provides an expert opinion on its authenticity. Traditionally, archivists have been able to provide the needed expertise, but their body of knowledge is inadequate to assess the authenticity of digital records and would profit from an understanding of the methods Digital Forensics uses to analyze and evaluate the environment in which the records existed. In turn, Digital Forensics methods will not only benefit from, but be extraordinarily enriched by Digital Diplomatics, which has established very sophisticated methods for assessing record trustworthiness and developed a strong conceptual model of an authentic record rooted in jurisprudence, administrative history and theory, and on recordkeeping practices in bureaucratic organizations (MacNeil, 2004). This model is especially important for answering questions about the authenticity of digital records extracted from their original environment and about procedures for extracting records so that their identity and integrity can be maintained intact, thus allowing them to be later authenticated.

The identification of "records" among all the digital objects produced by complex dynamic and interactive systems, and the determination of their authenticity, are issues that have been and continue to be directly dealt with by a research project called InterPARES (InterPARES Project, 1999-2012), the goal of which is to develop the knowledge necessary to support the reliable and accurate creation and the long-term preservation of authentic digital records (MacNeil, 2000, 2001, 2002; Duranti, 2005; Duranti and Thibodeau, 2006). The objects of InterPARES research are digital records that exist as large aggregations in live systems and are still in the hands of the creating organizations. These organizations must anticipate the possibility that digital records they produce will be relied upon as evidence in civil and criminal trials, thus necessitating advanced preparation to ensure admissibility (Nevins, et.al., 2008; Endicott-Popovsky, et.al. 2007, 2005, Endicott-Popovsky and Frincke 2007a, 2006, 2007). InterPARES recommendations and guidelines ensure that it will be possible to preserve authentic copies of these records permanently.

The additional problem that needs to be addressed, using the knowledge of digital records trustworthiness developed by the InterPARES project and the new concepts and methods of Digital Diplomatics derived from it, is that presented by records that have been extracted from the system in which they were generated and/or maintained either by the creating body itself or by third parties, such as police departments or archival organizations or units. These records may have been removed from the original system and placed on portable media by the creator for storage elsewhere, or by other parties, such as law enforcement officers, for use as evidence in criminal investigations. Thus, they may end up on CDs or DVDs accumulated in an office drawer, or on backup tapes in an off-site warehouse. They may also end up being acquired at auctions, either inadvertently, for example by individuals who, after buying what they assumed were blank, used tapes, later discover that they actually contain records, or intentionally, for example by collectors of digital art, unaware of the difficulty of assessing the authenticity of such art when separated from its original technological context. These records are often of uncertain origin and/or exist in proprietary formats that are hard to maintain over time, yet often must be maintained intact with their identity and integrity for long periods of time (e.g., while waiting to serve as documentary evidence in a trial, or for their ongoing research value).

The objectives of the Digital Records Forensics research program are:
1. to develop concepts and methods that will allow the records management, archival, legal, judicial, law enforcement and digital forensics professions to recognize records among all digital data objects produced by complex digital technologies once they have been removed from the original system;

2. to develop concepts and methods to determine the reliability, accuracy and authenticity of records no longer in the original digital environment;

3. to identify, develop and organize the content of a new science and discipline called “Digital Records Forensics;” and

4. to develop the intellectual components of a new program of education for Digital Records Forensics experts.

2.1 Relevant Scholarly Literature

The legal profession has been fully aware of the problems presented by documentary evidence in digital form for a long time, to the point that a dedicated group of legal experts, the Sedona Conference, has already issued two editions of principles to be followed in the production of digital records, realizing, along with the InterPARES project and the archival profession at large (International Council on Archives, 2008; European Commission, 2008), that the key to having trustworthy record sources is to generate them according to specific authenticity requirements and maintain them in the correct way throughout their existence (Sedona Conference Working Group Series, 2007). However, this does not solve the problem of documentary evidence that has already been created in systems that do not satisfy authenticity requirements, especially if it no longer resides in the original system or in any system at all, having been stored in external media. This is a major issue for the law enforcement and legal professions, and the judiciary, as demonstrated by the large number of scholarly writings on the subject (Gahtan, 1999; Arkfeld, 2002-2006; Rice, 2005). The judiciary has tried to address the problem by specifying minimum requirements for admissible digital evidence and by providing guidelines for meeting these requirements, but has not provided guidelines for assessing material that does not obviously correspond to the requirements (British Columbia Electronic Evidence Project, 2006; Guidelines for the Discovery of Electronic Documents, 2005, Supreme Court of B.C., 2006). Digital Forensics does not focus on the documentary evidence per se, but on the environment of its creation and maintenance, regardless of the efforts made by scholars in the field to find appropriate methods to assess the digital entities themselves (Casey, 2004; Carrier, 2005; Pollitt and Sheno, 2005).

The importance of using Diplomatics to acquire an understanding of digital entities as records and assess their authenticity on the basis of their characteristics is widely recognized by archival and diplomatic scholars (Bearman, 1992 and 2006; Barbiche, Blouin, Delmas, Delmas and Blouin, Guyotjeannin, 1996; Ansani 1999; Guyotjeannin, 2002, 2003). Standards developing bodies, like the Canadian General Standards Board, have attempted to address these needs by issuing requirements based on archival concepts (CAN/CGSB-72.34—2005), and scholarly archival literature on the subject has pointed out the pitfalls of leaving such responsibility to legislators rather than to researchers (Iacovino, 2005; Cox, 2006). Very recently, archival educators have recognized that education on digital records requires the contribution of a variety of disciplines and professional fields (Duff, Marshall, Limkilde and van Ballegooie, 2006). As well, writers on digital forensics have identified the need for interdisciplinarity in the formal programs aiming to educate professionals in their field (Irons, 2006; Boucher and Endicott-Popovsky, 2008; Irons, Stephens and Ferguson, 2009; Casey 2007; Nance, Armstrong and Armstrong, 2010). This requirement has also been amply demonstrated by a plethora of research projects on digital preservation, all of them inter- or multidisciplinary (e.g. ERPANET, www.erpanet.org; Digital Curation Centre (DCC), www.jisc.ac.uk; Digital Preservation Europe, www.digitalpreservationeurope.eu; CASPAR, www.casparpreserves.eu; and, mostly, InterPARES, www.interpares.org, which is the only digital preservation project entirely
focused on records).

All these digital preservation projects have identified the evidentiary issues discussed earlier, but have not dealt with them—their priority being long-term preservation of aggregations of materials—and have not included digital forensics knowledge in their research. However, some of the concepts developed by the InterPARES project constitute the necessary foundation for understanding digital documentary evidence, assessing its record nature and determining its authenticity, including the concepts of ‘formal record elements’ versus ‘attributes’ and ‘digital components,’ ‘accuracy’ versus ‘authenticity,’ ‘digital authentic copy,’ ‘fixed form’ versus ‘bounded variability,’ ‘manifested record’ versus ‘stored record,’ and ‘instructive record’ versus ‘enabling record;’ all of which will support the determination of what constitutes a record among the various digital entities extracted from computer systems, and of which instance or manifestation of those entities has the force of an original and can be assessed as authentic. Traditionally, police evidence rooms and public archives have implicitly guaranteed that the records kept by them are as authentic as they were when first acquired, but this presumption is no longer tenable for digital documentary evidence; thus, it is essential to develop procedures that can reassure the public and the court system that no undetectable manipulation of such evidence can occur throughout the time of its maintenance, especially when the keeping of the evidence is entrusted to one of the parties having a stake in it (i.e. the police).

2.2 Methodology

To develop a Digital Records Forensics science, the research team has conducted an in depth literature review and developed a data base of annotated writings from all the areas of knowledge covered by the research (see http://www.digitalrecordsforensics.org/drf_biblio_db.cfm). This database serves as a fundamental resource for the research program, and will be continuously enriched and maintained as a resource for the education program under development. At the same time, the team has built a data base of case law related to the issues identified above, such as authenticity, integrity, recordness, etc. (not yet accessible to non team members); and a terminology database which aims at defining the key terms for the project in each of the disciplines involved, and indicating the preferred definition in the context of the Digital Records Forensics area of knowledge (in progress, but available at http://www.digitalrecordsforensics.org/drf_term_db.cfm). From this accumulated knowledge, the team has developed an activity model of the digital records forensics processes, and prepared separate questionnaires for semi-structured interviews of digital forensics experts, law enforcement officers, records managers for law enforcement departments, court clerks, lawyers and judges (see http://www.digitalrecordsforensics.org/drf_questionnaires.cfm). The purpose of the interviews is to discover what are the criteria that these professions consider to be the basis for determining the trustworthiness of digital evidence, what methods they believe must be used to maintain digital evidence trustworthy from the moment they start interfering with its original digital environment, and what kind of education program would best serve the needs of Digital Records Forensics experts. To date we have conducted more than fifty interviews. A web administered questionnaire aimed at establishing shared beliefs about the fundamental means of establishing and maintaining digital record trustworthiness and discovering the gaps and the problematic areas in existing digital forensics knowledge will target members of the digital forensics, law and records professions, and selected members of the public (e.g., journalists and scholars).

The research team is also conducting ethnographic investigations involving the examination of the forensic and recordkeeping procedures of the Vancouver Police Department (VPD), which is our test bed partner. In addition, graduate students under the direction of the team’s experts in the Law of Evidence and in Digital Forensics are examining and describing the hierarchy for policy changes and decision-making, the current court procedures governing the admission of digital evidence, and the problems noted by the personnel responsible for digital evidence. The team is using the preliminary findings from the interviews as points of reference for studying the environment and as the basis for their discussions with the professionals working at the VPD. The graduate students, under the
supervision of the researchers from the VPD, are also examining the digital evidence identified as records preserved by the VPD, which, in the view of the Department itself, is problematic for various reasons (e.g., unreadable because of obsolescence; produced in a legacy system no longer available; of unknown lineage). Inspection of the material is conducted using the analytical methods of all disciplines involved in this research program. Solutions to the identified problems will be implemented on copies of the material and tested in light of the Law of Evidence. As needed, the team will elaborate on existing concepts and procedures or develop new ones. As both the ethnographic approach and the case studies constitute action research, the team collaborates with the subjects of the investigation, who are co-participants and stakeholders, to develop together practical methods and new knowledge. The team will synthesize its findings into structured content for a Digital Records Forensics science as part of a proposal for a new interdisciplinary program of graduate education in Digital Records Forensics Studies.

3. THE DIGITAL RECORDS FORENSICS BODY OF KNOWLEDGE

In order to determine the content of the body of knowledge that would identify Digital Records Forensics as a science and a discipline, it is appropriate to reflect on the characteristics of both. A science comprises the ideas about the nature of the object of its study (i.e., theory) and about the principles and procedures for handling, controlling, examining, and maintaining such an object (i.e., methodology). The analysis of these ideas, principles and methods; the history of the way they have been applied over time in different contexts (i.e., of practice); and the literary criticism of both analysis and history (i.e., scholarship) are also an integral part of a science. Thus, a science can be defined as a system inclusive of theory, methodology, practice, and scholarship, which owes its integrity to its logical cohesion and to the existence of a clear purpose that rules it from the outside, determining the boundaries in which the system is designed to operate.

If we regard a science of Digital Records Forensics as an organic and unitary system, we have to accept that we would be dealing with a special type of discipline. A discipline encompasses the rules of procedure that discipline the search of the scholar, and the knowledge so acquired. In the case of a digital records forensics system, however, the rules that will guide the investigation of scholars into issues, problems or concepts would have to be determined by its theory and methods. This is especially noticeable when research aiming to develop methods, strategies and/or standards for the treatment of new types of material looks for a starting point, or fundamental terms of reference.

To explain, it is useful to identify the components of the system in the case of a Digital Records Forensics science. The object of its study would be digital records. Consequently, its theory would be constituted of ideas about the nature of records in the digital environment, their characteristics, components, relationships and behaviour. Its methodology would encompass ideas about location and acquisition of digital records, identification and analysis, evaluation and interpretation, maintenance, transmission and preservation. Its practices would comprise accepted standards and the specific processes followed in various cases in different contexts, as well as the tools and instruments selected to carry out those processes and their performance. The purpose ruling this system from outside and determining its boundaries would be the acquisition/production of digital records capable of serving as reliable, authentic and accurate evidence, and their preservation for as long as required by the relevant juridical system. Scholarship would therefore aim at gaining an understanding of types of records and systems, of methods and practices, of legal, administrative and technological issues, and, on the basis of such understanding, developing more effective methods and practices, solutions, proposals for changes to the law, for design of new tools, etc. However, it is clear that, in order to be useful, such scholarship would have to be guided by the theoretical and methodological ideas that constitute the foundation of the system, such as the concepts of record, authenticity, evidence, forensic process or digital record systems.

Digital Records Forensics as a field of study is highly interdisciplinary. Some of the disciplines/sciences/practices whose knowledge is to be brought to bear on Digital Records Forensics
are centuries old, while others may be very recent but are entrenched in their very established views of things. To make a new science out of a field of study cross-fertilized by several bodies of knowledge requires a very detailed work of comparison and reconciliation of concepts, carefully aimed at maintaining consistency with the ultimate purpose of the new field. Thus, the selection of terms, definitions, principles, etc. should not occur on the basis of what is best in absolute terms, but of what best serves the purposes of Digital Records Forensics and is consistent with the other accepted ideas within it. Again, it is necessary to regard this new science as a system made up of parts, structure and processes. The parts are theory, methodology, practice and scholarship, each of which is, in turn, composed of parts. The structure is a hierarchical one, where each level descends from and depends on the previous one, with theory being the determinant and cohesive element. The process most relevant to us, at this stage of scientific system development, is that of feedback, a process by which our hypotheses, ideas, findings or realities are brought into the system, confronted with the ideas ruling the system from the inside and with the purpose guiding it from the outside, and either absorbed by and integrated within the system, renewing and enriching it, or rejected.

An example of the process described above can be seen in an article aiming at comparing the concepts of Digital Diplomatics with those of Digital Forensics (Duranti, 2009b). As mentioned earlier, Digital Diplomatics is really a branch of Diplomatics, rather than a separate discipline, developed as a result of the application of the knowledge of the latter to the analysis of digital records (Duranti, 2009a). The article discusses the concepts of trusted custodian, digital record, reliability, authenticity, accuracy, integrity, etc. from the perspective of both fields, and the methods of identification and analysis used by each, showing the similarities and the divergences and identifying the areas in which each can benefit from the other. In the course of this comparison, also the perspective of the law of evidence in North America is kept into account. It is very important to continue this type of investigation to develop a Digital Records Forensics science that can form the core of an academic program for digital records forensics professionals.

But it is not necessary to wait for a full-fledged science to be developed before delivering the knowledge that already exists in the form of a graduate university program. While it is true that a graduate program is given legitimacy in the eyes of a university by the existence of a substantial body of knowledge in a well defined area, it is equally true that the development of such a body of knowledge is the consequence of the existence of a graduate program that educates both professionals and scholars in conducting ongoing theoretical and applied research. Thus, it is possible to start now in a small way, but “thinking big” and maintaining our focus on the ultimate goal.

4. A DIGITAL RECORDS FORENSICS GRADUATE PROGRAM

At this stage of development of the body of knowledge of a Digital Records Forensics Science, we have established that its theory, methodology and practice would mostly derive from:

- The Law of Evidence, which rules the whole system from outside and provides its purpose;
- Diplomatics (and specifically Digital Diplomatics), which embodies the theory of the record;
- Digital Forensics, which comprises the core methodology related to the acquisition, analysis and evaluation of digital evidence and the related practices;
- Archival Science, which provides the theoretical and methodological knowledge related to recordkeeping and long term preservation;
- Information Technology, which offers the necessary understanding of systems concepts, computer architecture, computer network communication, discrete mathematics, database design, algorithms and data structures, imperative
programming, mark-up languages, and end-user programming tools; and

- Organizational Information Assurance, a relatively new field that examines concepts, elements, strategies, skills related to the life cycle of information assurance -- involving policies, practices, mechanisms, dissemination and validation -- that ensure the confidentiality, integrity, availability, authentication and non-repudiation of information and information systems (Endicott-Popovsky and Frincke, 2005a, 2004).

How can this body of knowledge be delivered in the context of a graduate program without having to establish at the outset a full fledged degree in Digital Records Forensics? In one example that the Master of Archival Studies at the University of British Columbia and the School of Information at the University of Washington are pursuing at this time, the two schools would make an agreement according to which the students enrolled in the Master of Archival Studies take a semester in the School of Information receiving credit for their courses in their home program, and vice versa. The combination of courses would encompass the entire body of knowledge outlined above and the University of British Columbia students would receive a Master of Archival Studies degree with specialization in Digital Records Forensics, while the students of the University of Washington would have the same specialization attached to a Master of Information.

It will be necessary to develop one new course in one of the two programs to provide the intellectual framework for the specialization, and to adjust some of the course content in both, but this collaboration offers an opportunity for testing the viability of a graduate program in Digital Records Forensics, proving to universities that there is a demand for such a program, and developing an integrated body of knowledge on which to build innovative, original knowledge. This can happen because such a program would be based on the three educational principles already shared by the two programs contributing to it: 1) professionals must be educated in the core theoretical and methodological knowledge that identifies their profession; 2) they must be educated in international standards as well as in the specific, local and unique aspects of the juridical-administrative environment in which they will work; and 3) they must be educated in the scholarly as well as the practical nature of their work.

The third of these principles is the most important for university programs. Research is a critical component of a graduate level program, because it is an expression of the intellectual nature of the study, the scholarly substance of the work that professionals do, and the status of the program with respect to other graduate programs. Several course offerings can enable students to engage in scholarly enquiry of various kinds, from the thesis to directed research projects involving in-depth investigation of a specific issue or problem. Moreover, it is a requirement for every faculty member to conduct scholarly research and granting agencies are more than willing to provide funds for the participation of graduate students in research, thus, they may work as paid research assistants on faculty members’ research projects. It will be this engagement in research that will produce new knowledge and support the creation of dedicated full-fledged programs in Digital Records Forensics.

However, in a Master’s level program, the cultivation of research skills must be balanced with the development of professional knowledge. Accordingly, it is important to inculcate in students engaged in research a sense of the relevance of their investigations to their professional lives. This is why the study of research methods should be a required component of any program of education, as it will equip students with the knowledge necessary not only to produce new knowledge, but also to understand and interpret research conducted by others.

Graduate programs are judged to a significant degree by the quality and quantity of the research produced by faculty and students, thus, expanding the opportunities for research is vital to their success and growth. Students benefit enormously from the opportunities research projects provide for acquiring research skills and contributing to the advancement of disciplinary knowledge. Once the students graduate and begin their working lives, the knowledge and experience they have gained
through their participation in research translates into a benefit to the institutions and organizations that employ them.

Talking of the practical component of such a graduate program, it is important to emphasize that experiential learning in the context of the education of professionals is not an exercise to discover theory and methods empirically. Its main purpose is to provide future professionals with a way of applying the theoretical and methodological knowledge learned in class and testing it in the professional arena. This is the best way of demonstrating to the students that theory and practice feed each other and neither could have value without the other. Recently, some programs have introduced co-operative work experience opportunities for their students. Co-operative education is a learning method that, through pre-employment workshops, coaching by career specialists, and workplace experiences, offers students the opportunity to combine real world experience with their classroom education and develop employment skills specific to the records professions. Simply stated, universities and employers co-operate to provide students with an opportunity to learn in a workplace setting by alternating practical, paid work experience in various fields of interest with their academic studies. Most importantly, at this stage of program development, the practical experience would allow digital records forensics students and their professors to assess the value of their education, to identify gaps, and to work towards a course and curriculum development that better serves the needs of professionals. At the same time, the students and their program of education would be visible to professionals, who will appreciate the value of both and generate the demand required by universities to support such programs.

5. CONCLUSION

The Digital Records Forensics Project began two years ago with the objectives of producing much needed new knowledge and creating dedicated graduate programs of education delivering it. The research conducted to date has demonstrated the need for Digital Records Forensics specialised knowledge among several different professions: digital forensics experts, lawyers, law enforcement officers, judges, court clerks, records managers, archivists, systems designers, etc. In addition, the research has shown that, in light of recent court decisions that have increased the length of retention of digital evidence used in trials, in some cases requiring permanent retention, long term digital preservation has become a major issue, to the point that recordkeeping and archival knowledge must become part of the intellectual armour of every professional responsible for digital evidence. That the type of educational program we envision would produce a professional in high demand in a variety of environments has been abundantly demonstrated to our research team by the responses given in the course of our interviews by judges, lawyers, court services administrators, and last, but definitely not least, digital forensics specialists and members of forensics units within police departments. As Mark Johnstone, Sergeant, Forensics Services Division, Financial Crime Unit, Vancouver Police Department, put it, “people need to understand what exactly a record is. And then understand the manner in which it’s maintained. So you’d have to have the knowledge of what it is you’re trying to maintain and then the knowledge of the systems that are maintained. So, yes, there’s some very specific knowledge needed” (transcript of interview, part 2 of 2, 12-09-2009). It is our hope that, in the next year, we will have moved quite far in reaching our goals and will have earned the support of the digital forensics profession for establishing a Digital Records Forensic science in academia, in whatever form will be most appropriate and useful.

6. REFERENCES


Canada Evidence Act, R.S.C. 1985, c. C-5 as am.


118


European Commission (2008), Model Requirements for the Management of Electronic Records (MoReq2). Available at http://www.project-consult.net/Files/MoReq2_body_v1_0.pdf


International Council on Archives, ICA (2008), Principles and Functional Requirements for Records


Tan, J. (2001), Forensic Readiness, Cambridge, MA: @Stake.


Developing a Baccalaureate Digital Forensics Major

John H. Riley, Jr.
Dept. of Mathematics, Computer Science and Statistics
Bloomsburg University
400 East Second Street
Bloomsburg, PA 17815
570-398-4627
jriley@bloomsburg.edu

ABSTRACT
As colleges and universities consider instituting a bachelor’s degree in digital forensics or computer forensics, there are numerous questions to be addressed. While some of these normally occur in the development of any new major, there are aspects of digital forensics which do not often (if ever) occur in other majors. We discuss the issues that should be resolved in the development of a baccalaureate degree program in digital forensics.

Keywords: Digital forensics major. Computer forensics major.

1. INTRODUCTION
For a number of reasons, a college or university may consider offering a baccalaureate degree program in digital forensics. We do not examine these reasons, but caution both institutions and individuals to check that a digital forensics major fits the institution. Mission statements and strategic plans must be consulted to determine this fit. The questions and alternatives we describe in this paper not only help the development of a digital forensics major, but also give those considering such a major a sense of what it entails. At the very least, this should help institutions and individuals to be realistic about what a digital forensics major will require.

There are always questions to be answered when new programs are being discussed and developed. Since this happens regularly in higher education, institutions have procedures that direct faculty and administrators through curriculum development and ensure questions common to all new programs are addressed. These institutional procedures are valuable and should be the starting point for anyone who is thinking about starting a digital forensics major.

Some of the areas we discuss may be covered in a general way in an institution’s curriculum development process. In these cases, the specific alternatives mentioned will be useful in meeting institutional requirements. However digital forensics, as an academic major, has aspects that are not included as part of an institution’s normal curriculum development process. How much of what follows normally is considered in a particular institution’s process will vary. The parts that don’t should not be ignored.

There are very few baccalaureate degree programs in digital or computer forensics. Utilizing the sites (Christine 2009 and Morris 2010), it appears that there are fewer than ten such programs in the United States. Furthermore, some of these are combined majors (e.g. Computer Investigations and Criminal Justice at St. Ambrose University, Davenport, Iowa) or concentrations within majors (e.g. Information Technology with a concentration in Computer Forensics, online at American Intercontinental University). Examining the curricula of these programs reveals that decision makers have faced many of the questions that we discuss.

While we have grouped the questions in six general areas, the questions and their answers are not independent. Proposers will need to address questions from all areas in conjunction with one another.
2. PROGRAM EMPHASIS OR EMPHASES

Just as a career in digital forensics can follow many different paths, majors in digital forensics can have different emphases. In this section we describe some of the underlying emphases of digital forensics to be considered in building a digital forensics major. While these emphases are not disjoint, which are in the minds of the designers and implementers will influence the major, particularly as a program begins.

While law enforcement (LE) is often the first area of digital forensics that comes to mind, there is a need for digital forensics in corporate information technology (IT). While LE and IT digital forensics overlap, a conscious or unconscious slant toward one or the other will influence choices. A curriculum with an LE slant would probably have more emphasis on legal issues (e.g. the need for search warrants and chain of custody issues) than one with IT in mind. An IT emphasis would necessitate more education about networks. An LE based curriculum would also need to acquaint students with the hierarchical (quasi-military) organization of LE. With IT in mind, courses in business should be considered. Finally, an institution’s culture may be more accepting of LE or IT and proposers should know their audience when using references to either LE or IT.

Similarly, designers of a major in digital forensics will need to decide how much emphasis to put on the technical and legal facets of the discipline. While neither can be omitted, students in a program which is more technically oriented may find employment in IT outside of digital forensics. In contrast, students graduating with a good background in legal matters may work in ediscovery.

Higher education curricula also can be designed to emphasize theory over practice or vice-versa. While a digital forensics major is inherently practically oriented, some theory must be present and decisions about how much must be made. For example, while binary and hexadecimal numbers are a sine qua non it may not be essential to cover two’s complement arithmetic. An operating systems course which focuses on the design principles or criteria of an operating system is not inappropriate but is not the same as giving students the ability to use Linux. Again, an institution’s culture should be considered in this discussion.

These emphases may not be explicitly mentioned or considered in the development of a program. They may arise as consequences of other decisions. However, realizing that they exist can eliminate difficulties. Using an operating system course taught by a traditional computer scientist may be acceptable if the digital forensics program is theoretically oriented but not if the program needs practical skills (e.g. familiarity with a command line environment). Discussing where incidence response (more of an IT emphasis) fits in the program with high level administrators who are expecting an LE orientation might be difficult. Establishing some understandings or being explicit about a program’s emphasis or emphases will make the program development process much easier by providing a framework for decisions.

3. PROGRAM CONTENT

As soon as one begins to consider what might be good to put in any curriculum, one discovers there is always too much. The main challenge is to decide what not to include. In this section, we discuss many of the areas that might be included in a digital forensics major. These must be approached in the framework outlined above.

Traditionally, colleges and universities have thought of a major as being defined by the courses in the major. Towards the end of the twentieth century, a shift toward defining curricula, including majors, by the desired outcomes for students occurred. Consequently, designers of a digital forensics curriculum should not start by thinking about the courses needed for their major, but about what students should know or be able to do when they finish a course or complete the major. Proposers should become familiar with their institution’s outcomes assessment procedures and requirements. Ultimately the outcomes of the digital forensics major will need to be assessed.
Two aspects of hardware are essential to a digital forensics major. First, the need to preserve evidence means students must have some sense of the ability of media to lose or retain data. Second, since storage and other digital devices must be connected to examination machines, experience doing this is necessary. The wide variety of digital devices (consider cell phones) means that even the coverage of even these basic topics will have some omissions. Additional topics from computer and network architecture certainly can be considered for inclusion.

Since operating systems and file systems determine what evidence may or may not be present on a storage medium, forensics examiners need a grounding in these areas. File systems are particularly problematic for curriculum designers. Understanding them at some level is needed, but it is impossible to include everything about a modern file system in a curriculum.

Operating systems and applications produce and leave many artifacts. Students must have some experience finding and recovering artifacts but decisions about how many and which ones must be made. For example, the registry in a Windows system has thousands of entries. Some can be crucial in an investigation. Curriculum designers must determine what will be included in their major.

Almost all of the programs in existence include some coverage of the preceding topics.

We have intentionally used the descriptor digital forensics instead of computer forensics. A forensics curriculum which only covered computers (in the strictest sense) is feasible. However, the proliferation of digital devices (cell phones, GPSs, PDAs, …) and the data they contain makes the inclusion of digital devices in a forensics major desirable. The wide variety of devices and the lack of standards (particularly among cell phones) makes complete coverage of this domain impossible. To a certain extent, the coverage will be determined by logistic constraints, e.g. which types of cell phones are available for study.

The incorporation of forensics tools such as EnCase® or FTK® must be done carefully in an undergraduate curriculum. A college or university should not be training students in how to use a specific tool or tools. Rather the use of a tool should occur in the context of some other end such as case work or the recovery of artifacts. Consequently, one question to answer is how will a curriculum use a forensics tool to meet other objectives. It is also important to consider when in the major (first semester, second semester, …) students will be introduced to these tools. Because they are so powerful, if the tools are introduced too early, students may not be motivated to understand some important concepts. For example, if a tool recovers deleted files, students may not appreciate what file deletion entails. Worse yet, when the tool fails in some task, they may not have any sense of what to do next. On the other hand, slogging through manual tasks unnecessarily will discourage some students. Deciding the place of forensics tools in the curriculum is very important.

While a network in the most literal sense only transfers data, the impact of networks on computers and the data they contain means some aspects of networks should be considered for inclusion in a digital forensics major. Designers of a digital forensics major may want to discuss the use of networks in forensics work, e.g. to transfer data from a target machine to an examination machine. Network forensics and incident response are viable topics for the major. All existing programs include an introduction to networks course. One program (at Rochester Institute of Technology) includes four (quarter) courses in networking with an advanced track that centered on networking.

An introduction to programming should help digital forensics students have some sense of the algorithmic nature of programs and give them some background in data structures. Both of these are useful in understanding the functioning of digital devices. Most collegiate level programming courses meet these ends. Beyond or instead of general purpose programming, programming in a digital forensics major may be specialized for two other purposes. First, being able to write scripts (e.g. in Perl) to automate some tasks is useful. Second, C or assembly language programming can be used to delve deeply in computer workings and is useful in malware investigations. Despite the centrality of programming to computing, most forensics degrees do not include a programming course.
There are several areas of computer science which may be incorporated in a digital forensics curriculum. Data bases, data mining, computer and network security are possibilities. Information systems topics, such as systems administration, can be useful. One way of handling this, particularly when courses already exist, is to allow students to choose from a list of courses. Most programs do this, but there is wide variation among the actual computer science and information systems courses that may be used in the digital forensics major.

Topics from criminal justice and legal areas may be included in a digital forensics major. Evidentiary issues are pertinent for both LE and IT. When law enforcement is being considered, search and seizure are important. White collar crime is another area which is useful for all students. Only two programs (both at very technically oriented institutions) do not include courses in this area.

Business topics can be useful for students majoring in digital forensics. As accounting discrepancies are often important, an introduction to accounting and fraud accounting fit well into a digital forensics major. At least three programs include an accounting course. A basic of understanding of organizations can be obtained from the area of management and a management course is required by one program.

Most undergraduate degree programs have a general education component. There are areas in general education which can be specified for a digital forensics major. A vital area is communication, both written and verbal. Technically oriented students are often weak in communication skills, so a digital forensics major must provide students the means of improving their communication skills. Since most undergraduate general education programs require communications courses, these are present in most programs. It’s reasonable to expect that many digital forensics graduates will work with sensitive issues, so practical ethics may be part of the major. Mathematics topics such as probability, statistics and cryptography should be reviewed to see if they support the major. Most programs have an explicit mathematics requirement. Typically the requirement is discrete or finite mathematics and introductory statistics.

The preceding outlines the main considerations for the areas for inclusion in a digital forensics major. Within the areas chosen, many more decisions about material will need to be made.

Furthermore, while the designers of a digital forensics major at an institution may desire a particular set of topics for the major, some decisions and options will be dictated by local resources. Utilizing existing courses, when appropriate, is generally helpful.

4. STUDENTS

One reason an institution may consider starting a digital forensics major is to enroll or retain more students. The major should be realistically designed for the students of the institution. Students in a digital forensics major will most likely be similar to students already at an institution.

Faculty in science, engineering or technical areas may not be aware of the range of abilities of students at their institutions. Since digital forensics is a new field, students may enter the major unaware of its demands. Administrators may also be unrealistic and expect that a digital forensics major will not have the attrition rates of other STEM (science, technology, engineering and mathematics) fields. Resolving questions about expectations of the major in terms of enrollment and retention will help to minimize problems when the major is in place.

Depending on the institution, students may enter the major as new freshmen, transfer students from other two and four year institutions or as internal transfers from other majors at the institution. Transfer students (both internal and external) often believe they should be able to finish the major fairly quickly, say in four semesters (two years). If an institution has significant numbers of transfer students (especially external transfers) proposers should determine how the program will accommodate transfer students. A digital forensic major, like other STEM majors, will have some sequential structure. Minimizing course dependencies will help transfer students complete the major
Digital forensics graduates will have knowledge and abilities that can be applied malevolently. It is tempting to consider background checks for students entering digital forensics. There are many difficulties with background checks. It is much better to be clear about professional and individual responsibilities and duties in the curriculum. At the same time, some employers do require background checks and students should be aware of this.

Students and their families will often inquire about employment and internships in digital forensics. The institution’s program development process will probably require the proposers to address employment prospects. These answers are useful, but cannot be considered definitive. Like any new program, hard data about digital forensics graduates’ employment and internships will not be available for years. It’s also wise to keep in mind that economic conditions years in the future are always uncertain and will influence employment opportunities. The U.S. Department of Labor’s Occupational Outlook Handbook (U.S. Dept. of Labor, 2010) is a useful source of data in this regard.

Finally, the number of students expected in the major should be stated, almost certainly as part of the program development process. Enrollment in entry level courses will include nonmajors, particularly if a digital forensics minor is established. Anticipating these demands is important for planning.

5. FACULTY

The implementation of a digital forensics major depends essentially on the faculty involved. The digital forensics faculty can (and should) be valuable, contributing members of the institution. However, the expectations of these faculty may need to be different from other faculty.

An institution’s usual program development process will determine such matters as number of faculty, use of adjuncts, course load, advisement load and so on. To a certain extent these numbers will be driven by the number of students in the major.

Recruiting (either internally or externally) the faculty for the digital forensics major can pose problems. What level of credentials (masters, doctorate) will fit the institution? If the institution desires doctorate level faculty (particularly for permanent appointments) it needs to determine which fields (e.g. computer science, criminal justice) are acceptable. If the institution accepts other credentials (e.g. work experience in digital forensics) for digital forensics faculty, those faculty may be at a disadvantage in obtaining tenure or advancing in rank.

Faculty involved in a digital forensics program will most likely be changing or adapting from their original areas of expertise. Training will normally not be in an academic venue, particularly if the program has a practical emphasis. The faculty must be comfortable being with nonacademics and recognize that they may be receiving instruction from individuals who do not have academic credentials. Administrators must recognize the value of training offered by commercial entities. Similar comments apply to conferences that digital forensics faculty will attend.

Teaching digital forensics courses imposes burdens on faculty that are atypical. For example, text books are almost nonexistent. As a consequence, faculty must develop exercises and assignments on their own from scratch. Laboratory assignments can be very time consuming to develop, if only for the sheer volume of data that is needed. The institution should determine how much help to give digital forensics faculty. An extremely valuable, but costly way, of helping faculty is to reduce their teaching load. Agreements about teaching loads need to be in place as a program begins and develops.

Institutions frequently expect faculty to do research or produce scholarly results. Other activities may be more relevant for digital forensics faculty, particularly as they develop curricula for students. For example, working with local police or a district attorney’s office will help a faculty member understand the needs of law enforcement and provide the police or district attorney with technical advice they might not have otherwise. Obtaining a forensics certification is also a worthwhile endeavor for faculty as it provides insight to the practice of digital forensics. Digital forensics faculty
need to know beforehand how such activities will be viewed by academic administrators and other faculty at the institution to enable digital forensics faculty to obtain tenure and advance in rank.

These issues are less thorny if existing faculty, particularly tenured associate and full professors, are associated with the program. Such faculty will not be concerned about meeting tenure and promotion criteria. In addition, having well regarded faculty working in the digital forensics program will legitimize it within the academy. However, even these faculty will feel the pressures outlined above. They will also be responsible for recruiting and then mentoring new digital forensics faculty. As new faculty may have a background not in digital forensics, mentoring them may include educating them in digital forensics, which is usually not part of mentoring.

Institutions and individuals developing a digital forensics major should understand that faculty work in this field will be quite demanding in ways that probably not have been experienced in other disciplines. Recognizing and accommodating these different facets of faculty work is important. Furthermore, they may not be covered in an institution’s program development process. Developers of a digital forensics major should be conscious of the distinct role of faculty in the program.

6. LABORATORIES AND RESOURCES

A digital forensics major should have a substantial laboratory or hands-on component. All campus laboratories involve space, cost and maintenance so discussions about digital forensics laboratories will occur naturally. However, there are special considerations for digital forensics laboratories that need to be addressed.

Computer forensics software often needs access privileges more like an administrator than a user. Understandably, campus technology departments are uncomfortable with this. One solution to this is to isolate the computer forensics laboratory from the rest of the campus. There are many advantages to doing this, but it does make administering the laboratory more difficult. Another solution is to use virtual machines.

Commercial computer forensics software can be very expensive. Decisions about what software to use in a program will need to be made. Since licenses can be restrictive, a program must ensure it can provide access for the number of students it enrolls. If students are to do work beyond the scheduled laboratory hours, they will need access to the laboratories when they are not in use.

Free computer forensics software exists and can be utilized. Beyond its lack of cost, freeware has the advantage that it can be used by students on their own machines outside of the laboratories. Since freeware is not advertised, it takes special effort by faculty to find it and evaluate it before utilizing it. Support (e.g. updates and documentation) for freeware is often minimal which also limits its utility. Finally, a good deal of free computer software is single purpose (e.g. an MD5 hash utility) so assembling all the tools needed for a computer forensics laboratory from freeware can be problematic.

The amount of data that must be handled by practicing digital forensics examiners can be large (terabytes). A computer forensics laboratory should be equipped to handle large amounts of data. In particular, it must be able to store and serve disk images. While there are ways of storing a disk image in pieces, storage requirements will be extensive and must be anticipated.

Computer forensics examiners encounter a range of devices and software. The laboratories for the digital forensics major should not be completely monolithic. This means decisions about, for instance, which operating systems (or versions of operating systems) exist in a laboratory must be made.

Forensics laboratories may include actual work with hardware, e.g., extracting a hard disk from a computer. If a campus has a replacement cycle for its computers, computers that have been taken out of use can be put in a laboratory for students to work on. Of course, the data on the hard disk must be wiped to meet privacy concerns.

The preceding addresses the laboratories for computer forensics. If networks or digital devices are
part of the curriculum, they should exist in laboratory environments. A laboratory network which already exists for a computer science program may be used for the digital forensics program. Computer science and digital forensics faculty should cooperate (and learn from one another) in this endeavor. A laboratory for digital devices may be harder to establish.

Cell phones illustrate most of the problems that occur with digital devices in a laboratory setting. First, a cell phone doesn’t sound like a piece of laboratory equipment. Without some actual use (air time) there will be no data on a cell phone, so a cell phone plan may be needed. Finally, experience with different cell phones is useful. A funding request for twenty different cell phones along with a cell phone plan for them will raise many questions. Donated cell phones are of limited utility. If there’s data on them, it needs to be securely erased to eliminate privacy liabilities. They then need use (as before). In addition, the rapid evolution of cell phones means donated cell phones may not be realistic examples. A plan for maintaining the physical security of the laboratory’s cell phones must also be established. Finally, the specialized equipment and software needed for cell phone forensics must be in the laboratory.

It is possible to have combined classrooms/laboratories or separate classrooms and laboratories. An advantage of the first is that instruction can mix class lectures and discussions with hands-on experience fluidly. The disadvantages of the combined classroom/laboratory are students will be distracted by the laboratory equipment (computers) and when it is being used as a classroom, it is not available as a laboratory.

Laboratories need support. This support can be provided by the faculty, the institution’s IT staff or a lab technician. If the faculty do this, this must be compensated by a reduction in their other responsibilities. Institutional IT staff or lab technicians will need guidance and assistance from the forensics faculty about the particular, and perhaps peculiar, needs of digital forensics laboratories.

As the laboratory component of a digital forensics major is very important, developers should have an initial laboratory plan in place when the major begins. As the major progresses and digital forensics changes, laboratories will need to evolve.

7. ADMINISTRATION

A digital forensics major must be administered in a way that is suitable for its institution. Usually this means that it is administered by a department. It is unlikely that a new department will be created for the major.

If the digital forensics major is proposed by faculty in an existing department, that department will almost certainly be the host department for the major. If it is proposed by faculty from several departments, one of those departments may be chosen as the administrative unit for the department. This situation will require coordination among the departments and a dean’s assistance may be required. A clear understanding of faculty members’ commitment to the digital forensics major (particularly in terms of course assignments) must be in place. If new faculty are being hired for the digital forensics program, there may be several possible departments for the major. Before agreeing to host the digital forensics major and its faculty, a department should review the concerns that have been outlined in the preceding sections. If a department cannot satisfactorily address these concerns, particularly those about faculty, it should not host the digital forensics major.

In addition, any new major is an opportunity for publicity. Typically this involves the faculty in the new major. Faculty in the host department who are not involved with the digital forensics major should be comfortable with the attention the new program will attract.

More generally, while the institution as a whole must accept and value the different kind of work the digital forensics faculty do, the host department has a special obligation to understand and support digital forensics. There is an enormous amount of work involved in developing and maintaining a digital forensics program. Doing this with indifferent or hostile colleagues is an almost impossible
burden. For all of these reasons, a good host department for the digital forensics major must be found. While the digital forensics major does not necessitate the creation of a new department, the program should have one faculty member coordinating it. Titling a person as Assistant Chair for Digital Forensics or Digital Forensics Program Coordinator may be helpful. Compensation, especially in the form of a reduced teaching load, should be considered. Ideally, this person knows the field of digital forensics, has good organizational skills and is able to work with colleagues, administrators and external stakeholders.

Determining the host department for the digital forensics major may occur naturally as the major is developed. If it does not, a host department should be chosen carefully. Currently, digital forensics majors are hosted in a variety of departments, divisions, or schools: Computer Information Science; Information Technologies, Networking, Security and Systems Administration; Business and Justice Studies; Information Technology & Sciences; Behavioral and Applied Social Sciences; Mathematics, Computer Science and Statistics.

8. CONCLUSIONS

Any new major poses opportunities and challenges for an institution and the individuals involved. The first question that should be addressed when a new major is being considered is how well it will fit the institution. If a major in digital forensics major seems desirable for an institution, the institution should recognize that digital forensics poses some characteristics which are not typical for a baccalaureate major. The areas discussed in this paper outlines the questions that arise from these characteristics. An institution should address these questions as it develops its digital forensics major.

ACKNOWLEDGEMENTS

The author is grateful for the opportunity to have worked with Scott Inch in the digital forensics curriculum at Bloomsburg University. Discussions with Dr. Inch have been very valuable in developing this paper.

BIOGRAPHY

John Riley teaches computer forensics at Bloomsburg University. He has worked on curriculum and accreditation at both the department and university level.

REFERENCES

The Defiance College Undergraduate Major in Digital Forensic Science: Setting the Bar Higher

Gregg H. Gunsch
Defiance College
701 N. Clinton St.
Defiance, OH 43512
419-783-2460
419-784-0426 fax
ggunsch@defiance.edu

ABSTRACT
This paper provides background information to accompany the panel discussion on Curriculum Design and Implementation in Computer Forensics Education. It is specifically focused on the content and delivery of Defiance College’s undergraduate (B.S.) program majoring in Digital Forensic Science (DFS). The genesis and evolution of the Defiance College DFS program are described, along with its successes, challenges and known opportunities for improvement. The desired outcomes of the panel discussion include articulating the necessary components of an undergraduate program, refining expectations of knowledge and skills required of students upon graduation, and suggesting strategies for achieving those expectations despite inevitable resource limitations and diverse student demographics.

Keywords: education, undergraduate, digital forensics, computer forensics, pedagogy, skills development

1. GENESIS
The Defiance College major in Computer Forensics was launched in August 2006 after several years of development effort by an Advisory Board comprised of local law enforcement personnel, representatives from industry, the Ohio Peace Officer Training Academy, and members of Defiance College. Creation of the program was largely driven by the urging of Defiance County Sheriff David Westrick, who recognized that the workload of digitally-related cases was ever-increasing and unsustainable, and that entry-level practitioners were needed to help alleviate that load locally and nationally. At that time, there was a single detective in this corner of northwest Ohio with the knowledge and skills to process digital evidence: Deputy Steven Mueller had worked on cases for thirty separate agencies in the region (presently, over forty). Defiance College was in a unique position to build upon a long-established and vibrant criminal justice program, firmly founded on the college’s culture of active engagement with the community and law enforcement.

Concurrent with the development of the computer forensics major, Defiance College (DC) was also partnering with several community agencies to establish the Family Justice Center of Northwest Ohio (FJCNWO), a collection of centers providing comprehensive services and support for victims of domestic violence, stalking, and sexual assault [4]. The FJCNWO officially launched in Oct 2006 with a substantial grant from President George Bush’s $20M Family Justice Center Initiative [3]. A portion of the grant was allocated to DC for the establishment of a computer forensics laboratory, both for delivery of the major as well as training of law enforcement personnel in support of the FJCNWO mission. Involvement with the FJCNWO provides both a context for the education of DC students as well as opportunities to apply that education in community service, particularly for those students in the social work, education, business administration, art, criminal justice, and computer forensics programs.

By the spring of 2006, DC had secured tentative approval of the computer forensics major from the
Ohio Board of Regents, pending the hiring of a qualified faculty member to manage the program and deliver the courses. In the SP06 semester, several students attended the first offering of an introductory computer forensics course held by existing DC faculty and visiting guest speakers. Six of those students were sophomores who “stepped out on faith” that they would be able to graduate in 2008 with a degree in computer forensics, or at least a self-made major resembling one. It was then that Dr. Gregg Gunsch and Defiance College discovered each other. He was hired to start August 2006, and final approval of the program was immediately granted by the Ohio Board of Regents.

Delivery of the program required creating content for six new lecture courses and their associated labs, and the arrangement of two senior capstone activities: an internship and earning a nationally-recognized computer forensics certification. It is important to note that this program is not a concentration or track of an existing computer science or similar program. It is a self-contained integrated major, drawing from other programs to be interdisciplinary, but is more than a handful of specialty courses added to an existing core.

The six pioneering students were entering their junior year of college, but content-wise, they were entering the sophomore year of study in their major. Rather than have them extend for a fifth year, or graduate with self-made majors consisting of only a portion of the computer forensics program, we accelerated the “natural” sequence of course offerings so that they could graduate in May 2008. Through hard work and personal attention we were able to implement the accelerated program and to graduate all six students with the full major in Computer Forensics. Since 2006 the program has grown to 41 full and part-time students, with an expected inbound enrollment of 35-40 new students in Fall 2010.

2. EVOLUTION

The original CF program, as designed by the Defiance College Computer Forensics Advisory Board and approved by the Ohio Board of Regents in 2006, consisted of the following courses. All of the courses beginning with “CF” were created for this new major; the rest previously existed.

- IT 110 – Programming I
- IT 130 – Database Design
- IT 320 – Networking Fundamentals
- IT 330 – Information Technology Ethics
- CF 110 – Intro to Computer and Digital Forensics
- CF 120 – PC System Software
- CF 130 – Operating Systems
- CF 230 – Seizure and Forensic Examination of Computer Systems
- CF 310 – Advanced Topics in Computer Data Analysis and Recovery
- CF 320 – Network Forensics
- CF 340 – Intrusion Detection
- CF 450 – National Certification
- CF 497 – Forensic Internship
- CJ 111 – Intro to Criminal Justice
- CJ 155 – Criminal Law
- CJ 217 – Criminal Investigation
CJ 221 – Criminal Evidence and Procedure
CJ 471 – Criminology
MA106 – Pre-Calculus Mathematics
BA363 – Business Law
AC221 – Financial Accounting
AC222 – Managerial Accounting

It had been the expectation of all involved that improvements would be forthcoming after the major’s first offering. The most apparent shortcoming was that there existed no “program prerequisites” to ensure that students have the background and academic qualifications to succeed in this major and career. This was especially crucial to remedy given the incongruity between the type of student who would flourish in the technologically-challenging computer forensic major and the type of student typically attending DC, a liberal arts-based college. Students coming to DC generally don’t anticipate pursuing a technological degree, and those students looking for a technological degree typically don’t look to a liberal art-based college. This presented both a challenge for recruitment of the right students, as well as a challenge to ensure they were adequately prepared for the core forensics courses of the major.

This shortcoming was also articulated by one of the first graduates: although she felt competent in computer forensic knowledge and had developed useful practical skills, she was uncomfortably unsure of her understanding of computer hardware/software fundamentals. Finally, along with these realizations came an additional impetus for change: due to steadily-falling student demand and anticipated departure of the IT faculty, the Management Information Systems (MIS) major was to be terminated. Rather than retain the existing IT courses, we used this as an opportunity to restructure the computer forensics program to accomplish the following:

- provide a strong computer technology foundation at the front end of the program,
- filter out students whom are not well-suited for the field of computer forensics, and
- align course content and prerequisites for smooth flow through the program.

The biggest change was to replace the two tangentially-relevant 100-level IT courses with a pair of new foundation-building courses to prepare students to earn the CompTIA A+ Certification with the IT Technician endorsement. Obtaining this certification was to become the prerequisite to the more advanced CF courses, effectively creating a “gate” to the core computer forensics program for only properly prepared students. Advantages of requiring students to become entry-level service technicians before starting the meat of the computer forensics program include:

- Conformance with National Best Practices: New computer analysts joining the FBI’s Regional Computer Forensics Laboratories must earn this same A+ certification before being sent to the FBI Computer Analysis and Response Team (CART) forensic training.
- Preparation: Students will have established crucial foundational knowledge and be significantly better prepared to succeed in the core computer forensics courses.
- Screening: This certification requirement would do for the computer forensics program what freshmen chemistry does for forensic science – ensures that only students with the affinity and background for the discipline move forward into the major’s core courses. This prevents a student from heading down a path of likely failure, as well as helps ensure that classes don’t get dragged down by having to backfill remedial knowledge.

The other relevant IT courses were repurposed to better prepare students for computer forensics, and the remaining CF courses were restructured/renumbered to smoothen the program’s flow. The
following courses constitute the current major [1], requiring 67 semester credit hours:

- CF 105 – CompTIA A+ Computer Essentials Exam Preparation
- CF 106 – CompTIA A+ 220-602 Exam Preparation
- CF 110 – Introduction to Computer and Digital Forensics
- CF 205 – Computer Security Fundamentals
- CF 210 – Operating Systems
- CF 215 – Computer Forensic and Security Ethics
- CF 305 – Seizure and Forensic Examination of Computer Systems
- CF 310 – Advanced Topics in Computer Data Analysis and Recovery
- CF 315 – Fundamentals of Computer Networks
- CF 405 – Network Forensics
- CF 410 – Intrusion Detection
- CF 450 – National Certification
- CF 497 – Computer Forensic Field Experience and Seminar
- CJ 111 – Introduction to Criminal Justice
- CJ 155 – Criminal Law
- CJ 217 – Criminal Investigation
- CJ 221 – Criminal Evidence and Procedure
- CJ 471 – Criminology
- MA106 – Pre-Calculus Mathematics
- BA363 – Business Law
- AC221 – Financial Accounting
- AC222 – Managerial Accounting

In early 2009 we made one other significant change: renaming the major to “Digital Forensic Science.” The justification was that “computer forensics,” while a commonly-used term, was inadequate to describe the scope of devices that process, store or transmit information in digital form. Nationally and internationally, this difference is being recognized in the careful selection of titles of organizations and activities such as the Digital Forensic Research Workshop, the Digital Forensic Certification Board, the Journal of Digital Forensic Practice, and the International Journal of Digital Forensics and Incident Response. Of greater significance is that Digital Forensics is now a recognized science according to the American Academy of Forensic Sciences (AAFS). In addition, we emphasize the application of scientific principles: in the classroom, students are taught to apply methods that are disciplined, systematic and repeatable in the collection, preservation, analysis, and reporting of digital evidence. The processes of hypothesis generation, experiment planning and execution, observation, and hypothesis support, refutation and revision are routinely discussed and practiced. Materials from the Scientific Working Group on Digital Evidence (SWG-DE) are included in our curriculum, and reflected in the catalog. Therefore, it is appropriate to claim that the students are taught to be scientist-practitioners, and that the most descriptive title of our program is Digital Forensic Science. This change was promptly approved by the Ohio Board of Regents. The CF course
designations remained the same (as opposed to becoming DF or DFS) at the request of the registrar.

3. PROGRAM DESCRIPTION

The four-year, in-residence, undergraduate DFS program fits the niche of developing entry-level practitioners who are ready for apprenticeship and have the well-rounded background expected from a liberal arts education to become life-long self-learners: oral and written language skills, civic and social sciences, natural sciences, arts and humanities, religion, physical education and life skills, etc. While there is some element of vocational training involved, the focus is on the development of critical reasoning skills, understanding the underlying principles of the tools and systems to which they are applied, and self-directed research to adapt to new situations. This embodies a tension between education and training, between preparing minds for life-long learning and a vocation. The program achieves this through a mix of traditional lecture, hands-on laboratory exercises, and engagement learning where the student amplifies his/her education through service to the community.

Students in the DFS program are actually being prepared to enter one of two career fields: digital forensics and computer/network security. The tools and techniques of digital forensics for information in motion are also applicable while recognizing and responding to intrusions into a company’s computer network. During the first year of the DFS program, students complete the two foundational A+ preparation courses, an introduction to digital/computer forensics, an introduction to criminal justice, and criminal law. During the sophomore year, students learn computer security fundamentals, general principals of operating and file systems, criminal investigations, criminal evidence and procedures, and ethics focused on computer security and forensics. In the junior year, students dig into the real meat of forensics work, starting with post-mortem acquisition and analysis of hard drives, and progressing through to live system analysis, mobile device forensics, password recovery, and criminology. They also learn the fundamentals of network communications and packet analysis. In their senior year, the students tackle network forensics and intrusion detection, and apply themselves towards accomplishing the two senior capstones. One is to complete 120+ hours of field experience (loosely called an “internship”) with one or more agencies performing digital forensics or network security tasks; typically, the forensics experience has been with law enforcement, and network security with industry. The second capstone is to earn a nationally-recognized, vendor-neutral, digital forensic certification such as the GCFA, CCE, or CFCE. Throughout the program the students make extensive use of the dedicated digital forensics laboratory, becoming familiar with a variety of commercial and freeware tools, including FTK, EnCase, WinHex, TSK/Autopsy, Wireshark, Snort/Squil, flavors of dd, etc., and tool collections such as Helix. [1]

4. SUCCESSES

4.1. Service Learning

The dominant defining characteristic of Defiance College is the emphasis on service or engagement learning: putting education into practice through betterment of the local or global community. Engagement learning is a natural teaching technique for creating problem solvers, and it is an effective means of incorporating real-world context into the classroom and developing a mature attitude of service before self. Students in the DFS program seek ways of integrating service experiences into the classroom. Some of the activities they have accomplished include:

- Examined the computer of a client of the Family Justice Center of Northwest Ohio (FJCNWO) who believed she was a victim of electronic stalking, then removed the existing malware and fortified the computer’s security suite
- Performed on-site security assessments of FJCNWO facilities
- Developed Internet safety awareness training materials and educated the incoming freshmen
- Provided Internet safety and identity theft awareness training to clients of the FJCNWO
Recovered personal data (music files, financial records, family photos) from a crashed hard drive of a community member. We wrapped crime tape around the computer and treated it as a suspected music piracy case, so that the students exercised the sequence of events from collection through forensic analysis, while performing a useful service in the process.

Hand-carved a large number of photographs from a corrupted digital camera memory card belonging to the nine-year-old son of an adult DC student. During this exercise we explored the limitations of and differences among the data carving (recovery) routines of FTK and foremost, and in the process, were able to recover the only existing photos of the child’s recently deceased pet cat.

Assisted the prosecutor’s office of a major city with a cold case homicide investigation. This was the first of what we expect to be an oft-repeated service activity.

Assisted high-school age students develop Internet-safety presentations for a FJCNWO Faith-Based Youth Initiative

The students also engage in community service opportunities outside of the classroom, including:

- Through DC’s Citizen Leader program, one student is organizing a computer repair and secure data wiping service for the community
- DC students have formed a school charter of the High Technology Crime Investigation Association, and begun the process of developing the next incarnation of HTCIA’s Internet Safety for Children Campaign public website [2]
- Students in the DFS, criminal justice, and (traditional) forensic sciences programs are helping to organize and conduct our second “Got Clue” crime investigation summer camp for area high school students
- Many DFS students are members of the Criminal Justice Society, where they have supported local law enforcement training sessions and campus disaster preparedness exercises

4.2 Certification Experiences

The pioneer class of 2008 chose to challenge the SANS GCFA certification for their capstone. To prepare, they all enrolled in SEC508 through the SANS Mentor program, held on campus and led by Dr. Gunsch. We spent significantly more contact time together than the Mentor program required, integrating the SANS materials into the other on-going courses and going far beyond that programmed into the SANS training. All six pioneers passed their certification exam on the first try.

As a result, greater emphasis on TSK as a tool for understanding FAT and NTFS file system structures has been incorporated into the junior-level courses, with routine side-by-side comparisons to other tools such as FTK and WinHex, and additional discussions on computer-related laws have been included throughout the program. The current seniors have formed a self-study group, purchased practice exams, and are actively preparing to challenge the GCFA examination before graduation in May 2010.

4.3 Internship Experiences

Each senior is required to complete 120 or more hours of field experience under the supervision of a forensic examiner or network security analyst. The students journal their experiences and produce a final report to benefit their fellow and future students. To date, internships have been held with the Toledo, Fairborn and Lima Police Departments, the Defiance County Sheriff’s Office, the Ohio Peace Officer Training Academy, the Ohio Bureau of Workman’s Compensation, and two insurance companies. The students have actively participated in tool validation, case management, and analyzing cases involving child pornography, drug, homicide, fraud, and employee misconduct.
4.4 Graduate Placement

All of the graduates who actively and intentionally sought employment in the digital forensics field so far have readily found employment. Hiring agencies include the DEA, the Ohio Attorney General’s Office, and private companies contracting their forensic services to law enforcement and industry.

4.5 Program Growth

Since the initial launch in 2006, the DFS program has experienced steady growth in terms of the number and preparedness of incoming students. There are currently 41 students actively enrolled in the program, ranging from freshmen straight out of high school to adult transfer students and military veterans through the Yellow Ribbon Program. There has been some degree of the “starry-eyed CSI effect” on the part of the less mature students, but those students quickly realized they have neither the aptitude nor predilection for DFS and moved elsewhere. The ideal new student would already be comfortable with computer hardware; for example, someone who repairs computers or has built a high-end gaming machine. Prospects are contacted early and encouraged to start or join a high school computer club and earn the A+ certification before or shortly after arriving at Defiance College. One of the sophomores did just that prior to arriving at DC in 2008, and several of the incoming freshmen appear to be able to achieve that goal.

A targeted recruiting campaign has been launched to enroll 35 to 40 qualified freshmen to enter in August 2010. As mentioned previously, simply drawing from the demographic typically attracted to a rural liberal arts-based college is not a practical approach; recruiting has to reach students with an affinity for digital technology. Through deliberate outreach in an ever-widening network, the Office of Enrollment Management has been able to connect with and attract a growing pool of qualified student candidates. To meet this anticipated growth spurt, Defiance College has begun the search for an additional full-time faculty member to supplement the current permanent and adjunct members.

Defiance College enjoys a rather unique status by having majors in Criminal Justice, Forensic Science, and Digital Forensic Science. The two applied sciences programs build heavily on the strong, 30-year old Criminal Justice major, which richly adds to the students’ knowledge and provides context for their education. Development of a fourth field, Forensic Accounting, is underway, starting first as a concentration for business students but later will grow into a complete major as demand dictates.

4.6 Advisory Board

The advisory board that drafted the original plan for the computer forensics program was recently augmented with a great deal of additional local and national talent. The new DFS Program Advisory Board has 25 members, covering a wide range of perspectives, including local law enforcement agencies, the FBI, DoD and DEA, Fortune 500 security practitioners, forensic tool developers, insurance and banking industry, investigative firms, graduate and undergraduate academia, and LE training. The roles of the Advisory Board are to provide guidance on the academic directions of the DFS program, assist in marketing the program to prospective students and employers, assist in recruitment of qualified students and faculty, help to identify funding opportunities for capacity-building and sustainment, help to identify or create internship opportunities for the students, and assist in job placement for the graduates. The board has also provided great advice directly to the students regarding what employers value, and what will make the students stand out during the job search process – an appreciated side benefit of which has been the unsolicited affirmation of what their professor tells them.

5. OPPORTUNITIES FOR IMPROVEMENT

This section discusses several open issues whose resolution should further strengthen the DFS program. It is hoped that the panel discussion on Curriculum Design and Implementation in Computer Forensics Education at the ADFSL conference could help to resolve some of them.
5.1 Focus of the Program

Students in the DFS program are actually being prepared to enter one of two career areas: digital forensics and computer/network security. For example, topics such as seizure and examination of computer systems prepare a student for the digital forensics field, while intrusion detection of ongoing activity is generally a network security issue. A case can be made for recognizing the significant amount of overlap between the two fields, particularly in the tools and techniques used for analysis as well as the methods used by the offenders. However, it has been suggested that we may be unnecessarily broad, requiring students to assimilate knowledge from too many areas at once. The open question is, “Are we biting off more than we should chew?”

Should we drop some of the network security-oriented content and corresponding internship flavor, and focus strictly on digital forensics? Is that even possible? Alternatively, should we develop two distinct tracks through the program as capacity and demand allow? Or is it the case that the original designers of the program were correct and the two fields are so intertwined that someone entering either field at the apprentice-level should have familiarity with both?

5.2 Learning Objectives

As part of the cyclic re-accreditation process, DC operates a self-study program for continual, deliberate improvement. Part of this self-study includes the establishment of and assessment against a set of learning objectives, both for the general education process as well as each of the majors. The current objectives for the DFS major are:

- Digital Evidence Scene Documentation: The student will demonstrate competence in crime scene documentation by photographing, creating diagrams, and tagging evidence items in a mock crime scene.

- Digital Evidence Image Acquisition: The student will demonstrate competence in digital evidence collection by successfully duplicating three digital evidence devices (hard drive, USB drive, floppy disk), using two distinct methods under two operating systems, and proving that exact copies were made while the original evidence remains unchanged.

- Digital Evidence Analysis and Reporting: The student will demonstrate competence in digital evidence analysis and reporting by successfully completing an examination of a set of digital evidence images and documenting the examination in a report suitable for legal proceedings.

- Network Security Monitoring: The student will demonstrate an understanding of network monitoring by installing and configuring an intrusion detection system, then monitoring, detecting, analyzing, and reporting on irregular network events created by the instructor and classmates.

- Testing Against an External Standard: The student will demonstrate competence in the breadth of foundational computer forensics knowledge by successfully passing the qualifying examination for a nationally-recognized digital/computer forensics certification; e.g., SANS GCFA, ISCFE CCE, or IACIS CFCE (law enforcement).

The open question is, “Are these the right objectives?” Do these capture the essence of the major and help determine if the students are being taught – and retain – the correct information to be successful in the field? If these are not the right objectives to assess, then what are? What process can be used to determine the proper set of objectives?
5.3 Certifications

Each student is expected to earn two industry certifications while in the DFS program: the CompTIA A+ certification at the front end, and a nationally-recognized digital forensics certification as a senior capstone. The open question is, “What happens if they don’t?”

The CompTIA A+ certification requirement was envisioned to be the gate to the sophomore-level DFS courses, and is currently listed as the prerequisite to these courses. It was assumed that after the two preparation courses were completed, the students would be able to pass the pair of examinations required to earn the certification. However, in practice it has not been that simple. The certification was designed for someone with 500 hours of hands-on experience in computer troubleshooting. Two semesters of coursework, no matter how intense, does not provide the same assimilative experience. So far, students who have received hard-earned “A”s in the courses (based on homework/lab assignments, quizzes and tests) have succeeded on the first try of the certification exams with only about a 50% pass rate. More assimilation time seems necessary to complete the certification, but this tends to spill well beyond the start of the subsequent courses. Our current practice has been to allow students who have passed the A+ prep courses to proceed to the sophomore courses in the DFS program; it seemed unnecessarily punitive to “hold them back” a year because they weren’t ready to successfully pass the certification exams. An additional concern has been the out-of-pocket cost to take the exams ($168 each) and any necessary retakes.

A similar question arises regarding the capstone digital forensics certification. Should failure of the externally-issued examination prevent or postpone a student’s graduation? This might be true for externally-accredited programs like education and nursing, but no external organization has imposed a certification requirement for the DFS program. There is a two-credit course associated with earning the certification. Can or should a grade be assigned for work done in preparation to take the examination, even if the student fails? A grade can be given based on participation in group-study activities, practice examinations, and even the degree of failure of the certification examination itself. Our current practice has been to award an “A” to students who pass the certification examination; so far, we’ve been fortunate to not have to address the “didn’t pass” situation. It will only be a matter of time, however, before a “C” student, who otherwise meets all graduation requirements, will not pass the certification examination in time for graduation. Would forcing completion of the certification process be unnecessarily punitive?

The feedback from the advisory board has been mixed. Some suggest that if the student cannot pass a certification examination, that student isn’t prepared to progress or enter the field; others recognize the realities of the academic process and need for assimilation time. The board was consistent to encourage the students to earn both certifications, even after graduation, because of the value they add to the résumé and employer. Even the A+ certification sets prospective employees apart as it shows foundational knowledge has been achieved and testimony is more trustworthy.

We feel that the DFS student should graduate with a very strong résumé and be highly marketable. The record should consist of a diploma, two certifications from industry (foundational and capstone), and a rich internship/field experience. In practice we are not certain that we can or should require students to earn the certifications to graduate, but we strongly encourage them to do so to achieve maximum competitiveness.

6. CONCLUSION

All positions on the training/education spectrum deserve attention and provide important contributions to address the growing, critical, collective need of knowledge and skills within industry and law enforcement. From short training sessions for tool proficiency and continuing professional development, all the way through advanced graduate research, high-quality programs are essential to meet that need. The four-year, in-residence, undergraduate Digital Forensic Science program at Defiance College fits the niche of developing entry-level practitioners who are ready for
apprenticeship and have the well-rounded educational background to become life-long self-learners. The ideal freshman student would enter DC already comfortable with computer hardware and be ready to dig into technical topics quickly. The most successful graduate will leave with a very strong résumé consisting of a B.S. degree, the A+ certification, a nationally-recognized digital forensics certification such as the GCFA or CCE, a rich internship/field experience, plenty of hands-on lab experience with a variety of tools, and a balanced education fully grounded in service-oriented learning in the liberal arts tradition.

BIography
Dr. Gregg Gunsch is a Professor of Digital Forensic Science at Defiance College, one of only a small handful of colleges and universities in the US to offer an undergraduate major in digital forensic science. He is a retired USAF LtCol, is a registered Professional Engineer (OH), and holds the CISSP, GCFA, CCE, and DFCP (Founder) certifications. Prior to arriving at Defiance College in 2006, Dr. Gunsch led graduate research and taught courses in information system security/assurance for thirteen years at the Air Force Institute of Technology, Wright-Patterson AFB, OH.

REFERENCES
ABSTRACT

Computer forensics involves the investigation of digital sources to acquire evidence that can be used in a court of law. It can also be used to identify and respond to threats to hosts and systems. Accountants use computer forensics to investigate computer crime or misuse, theft of trade secrets, theft of or destruction of intellectual property, and fraud. Education of accountants to use forensic tools is a goal of the AICPA (American Institute of Certified Public Accountants). Accounting students, however, may not view information technology as vital to their career paths and need motivation to acquire forensic knowledge and skills. This paper presents a curriculum design methodology for teaching graduate accounting students computer forensics. The methodology is tested using perceptions of the students about the success of the methodology and their acquisition of forensics knowledge and skills. An important component of the pedagogical approach is the use of an annotated list of over 50 forensic web-based tools.

Keywords: Accounting, auditing, forensics, learning styles, MBA, curriculum design.

1. INTRODUCTION

The field of digital or computer forensics has become increasingly important to various disciplines and has had a dramatic impact on the accounting discipline. The AICPA and PCAOB (Public Company Accounting Oversight Board) have issued statements regarding auditors’ increased responsibility for IT knowledge, fraud recognition and the importance of evaluating IT controls during a financial audit. The Sarbanes-Oxley Act of 2002 (SOX, 2002), increased evidence of business fraud and advancements in IT networks and systems should lead organizations to a higher expectation of auditors’ IT skills (Sumners and Soileau, 2008).

The difference between what the public expects from external auditors and the auditors’ own views of their responsibilities is referred to as the “expectations gap.” Lynn Turner, former SEC chief accountant, states that past audit failures and well publicized financial scandals have widened this gap (Turner, 2002). Enron, WorldCom, Adelphia, and Qwest (the Big Four) alone cost investors and pensioners more than $460 billion. These scandals which involved “cooking the books” went undetected by auditors and have created greater distrust of accountants in general. This perceived distrust in the quality of financial information may actually affect the proper functioning and integrity of the capital markets (Rezaee and Crumbley, 2007).

The AICPA, in its Statement on Auditing Standards No. 99 (SAS 99), Consideration of Fraud in a Financial Statement Audit (AICPA, Professional Standards, vol. 1, AU sec. 316), states that an “auditor may respond to an identified risk of material misstatement due to fraud by assigning …forensic specialists.”

SAS 99 suggests several procedures that are forensic in nature. These involve the performance of substantive tests or the application of methods and techniques of evidence collection based upon the
possibility of fraud at various levels of management, including override of internal controls, falsification of financial statements, misappropriation of assets, and collusion. Examples include extended interviews of financial and non financial personnel, surprise audits including recounts of inventories, tests of low risk accounts, and special tests not ordinarily performed (SAS 99). Even prior to the Big Four scandals the “Panel of Audit Effectiveness” on August 31, 2000, stated that “auditors should perform some ‘forensic-type’ procedures on every audit to enhance the prospects in the detection of financial statement fraud” (AICPA, 2000).

In response to the crisis of distrust, the AICPA has committed to six leadership roles to help restore confidence in the financial reporting process (Melancon, 2002). These include a role to promote academic research in how to strengthen antifraud education in the accounting curriculum.

Recent studies found that accounting students lack the requisite IT knowledge and skills to perform satisfactorily in their careers positions (Ahmed, 2003; Abu-Musa, 2008). Foundation knowledge should include topics focusing on IT security issues, IT auditing, IT governance, computer based analytical methods and general forensic and fraud investigative auditing knowledge techniques. Textbooks for a foundation class today typically devote separate chapters to each of these topics (Buckoff and Schrader, 2000; Crumbley et al., 2007).

An understanding of IT and related controls is important to accounting students (Van Grembergen, De Haes and Moons, 2005). Internal control, as defined by COSO (1992), is the process designed to help firms achieve objectives in the effective and efficient use of resources, reliable financial reporting, and compliance with applicable laws and regulations. IT controls increase an organization’s requirement for specialized knowledge and skill and are thus more costly to implement than other types of controls (ITGI, 2004; Cerullo and Cerullo, 2005). Understanding these controls is paramount to the effectiveness of both internal and financial auditors. While the adoption of computer-based auditing systems has steadily increased, a lack of IT education and background has prevented many auditors from integrating the necessary IT knowledge and skills with their professional knowledge (Davis, Schiller and Wheeler, 2007). This impairs the ability of the auditor in conducting appropriate tests on the relevant IT controls (Li, Huang and Lin, 2007).

Auditors with forensic IT skills have been in increased demand as a result of new regulatory requirements for compliance and higher emphasis on IT governance (Hoffman, 2004). The knowledge and skills for these professionals extend well beyond those for traditional auditors and, ideally, are a blend of accounting, forensic investigative and IT knowledge and skills. Most programs in accountancy, however, have not addressed the integration of the more traditional accounting with forensics and IT knowledge and skills Auditors who are educated and experienced in all of these areas become a valuable resource as either an IT auditor or internal auditor (Kearns, 2006).

Graduate accounting students often lack interest in information technology and may not understand the relevance of IT to their own career success (Kearns, 2009). Thus, the teacher is challenged to motivate the students and to encourage them to acquire IT knowledge by developing a positive attitude (Cangemi, 2000). Merhout and Buchman (2007) state that the education of IT auditors requires a blending of skills and educators “should strive to cultivate such a positive attitude in their students, and they should also make their students aware of the potential opportunities in the challenging IT audit career path.”

The purpose of this paper is to present a methodology, based upon established learning theories, for motivating accounting students to acquire skills and knowledge in a body of forensic tools commonly used in auditing. These tools are available on the Internet and can be used to develop an understanding of their application in a forensic auditing context. The paper will also present a summary of the application of several tools and an appendix will list over 50 potential tools for use by accountants.

2. LEARNING STYLE THEORIES

Learning style theories state that different individuals perceive and process information differently.
Therefore, to be successful, educators must strive to present information in different ways in order to support different learning styles. Two main learning styles are perceivers and processors. The **concrete perceivers** absorb information through direct experience while **abstract perceivers** rely upon analysis and observation. **Active processors** make immediate use of information in order to make sense of it while **reflective processors** reach cognitive awareness by simply thinking about the information. Traditionally, teaching has favored abstract perceiving and reflective processing. To be more effective, teachers must adopt styles that can motivate all students and provide for higher levels of interaction to increase the benefits for concrete perceivers and active processors. A brief discussion of three models expected to support and motivate accounting students of all learning styles follows.

### 2.1 ARCS Model of Motivational Design

The ARCS Model of Motivational Design is a method for improving the motivation of instructional material by presenting material in a way that is aligned with the learner’s motivational needs. It suggests that student motivation can be influenced through development of strategies in four areas: **Attention**, **Relevance**, **Confidence**, **Satisfaction** (Keller, 1987).

Attention can be gained through incorporating novel approaches, stimulating questions or problems for the student to solve, interactive methods, specific examples, and varying the material. **Relevance** is defined as the extent to which information is perceived as significant and satisfies student expectations (Keller, 1983). It can be established by using language and examples familiar to the learner and relating the knowledge to future usefulness. Role playing and games that force students to think creatively and independently create relevance. Using analogies and visualization of real-world problems, providing examples of the importance of information to the students’ careers, inviting guest lecturers who are experienced in the subject, and acting enthusiastically can also provide relevance. Research has shown students are most motivated by “interest in and perceived relevance of the material” (Gorhan and Christophel, 1992, p. 247).

Confidence can be instilled by creating incremental successes in the learning process, making success meaningful, and giving the learner control over their progress and assessment. **Satisfaction** is supported by methods that provide constructive feedback to the learner and provide examples of how the material can assist in real applications. It is also supported by an organized learning environment that empowers students and is closely related to confidence. Once confidence is established, satisfaction will follow if the learner is challenged. Creating learning goals that are attainable and in which the learner can demonstrate competency, especially in the classroom, instills both confidence and satisfaction.

### 2.2 Problem Based Learning

Problem-Based Learning (PBL) is a method of instruction that stresses an interactive hands-on approach to solve real-world problems. PBL uses open-ended context specific problems with no single correct answer. Students are encouraged to seek out the best tools and methodologies and seek creative ways to unstructured or semi-structured problems and to learn through exploration.

Students work in self-directed collaborative groups and teachers assume the role of facilitators. The teacher’s role is to provide direction and motivate the learner to inquiry through challenging questions and problems. Focus is on creative thinking whereby students apply their knowledge to new situations in order to solve unstructured problems and discover new and creative pathways to solutions (Hmelo-Silver and Barrows, 2006). PBL is thought to increase learner’s motivation, improve problem-solving skills, and develop critical thinking (Margetson, 1994).

### 2.3 Discovery Learning

Discovery Learning (DL) is an inquiry-based approach to instruction that makes the learner primarily responsible for uncovering new knowledge through inquiry-based instruction. Using problem-solving situations, the learner draws upon past knowledge and skills to construct new pathways to problem-
solving in order to discover new facts and relationships. It is an interactive approach in which students grope with real-world problems performing activities that lead them to new knowledge and skills. DL recognizes the importance of mapping new information to the problems it can address (Bruner, 1967). As a result, the learner is more likely to retain the new knowledge and skills. DL encourages active engagement, motivates the student, and promotes creative and critical thinking.

A benefit of the ARCS model is that it provides explicit suggestions and is highly applicable for a course of instruction that is both skills-based and conceptual. PBL is also appropriate for a course that introduces computer forensics to accounting students. Accountants are accustomed to working in groups and are generally self-motivated. DL is useful where the students solve auditing problems using a selection of web-based tools they deem appropriate. PBL and DL both support the inquiry-based, problem-solving approach where the student is primarily responsible for acquiring skills and knowledge. In this environment, the teacher acts as facilitator, providing direction, tools and assignments and intervening when the student is not making sufficient progress or is confused about the learning outcomes.

2.4 Audit Problems

Students will be assigned audit problems that simulate the IT control problems that can be encountered by external, internal and IT auditors working with private and governmental organizations. The problems will reflect real-world assignments and provide the student the wherewithal to investigate using computer forensic tools in a logical and structured manner. Student responsibilities will include uncovering weak internal controls and instances of fraud. Assignments will include: establishing that the network system is secure, establishing that the network has anti-virus software, performing an inventory of system assets, establishing that the local host is secure, locating hidden files, erasing or wiping sensitive file space, creating check-sums, recovering passwords, recovering damaged files, encrypting files, checking file integrity, mapping ports to applications, and verifying IPv4 and MAC addresses. Working in groups, students will be responsible for identifying the correct audit approach and incorporating tools that support their analysis. They will present their findings based on the use of these tools. The instructor will introduce a sample of the tools in the classroom and use quizzes and in-class collaborative problem-solving to gain attention. Role and game playing and guest speakers will be used to support both attention and relevancy. The relevancy of the assignments will be explained to the students citing the usefulness to their own careers.

Individually, each student will be called upon to formally present the use of two of the forensic tools and discuss their investigative potential. Students will also be given short description of audit problems and be asked to determine which tools might be applicable. These exercises are expected to create active discussions and create confidence and satisfaction.

2.5 Forensic Tools

An annotated list of forensic tools that can be used by accountants in their audit investigations is shown in Appendix 1. Students will determine which of these tools are applicable to their assigned problems and show how they can be used in the investigation. Collaborative exercises involving specific tools and DOS commands will be assigned. The accompanying links are available at http://www.stpt.usf.edu/bayview/forensics/.

3.0 HYPOTHESES AND METHODOLOGY

Nine hypotheses are presented. The hypotheses are intended to reflect the students’ interest in the subject and their perception of the class a priori and ex post and identify what curriculum design strategies are successful. First, it is theorized that accounting students will perceive computer forensic education as positively affecting their careers. The student’s desire to pursue knowledge and skills in this area is important to classroom success. Thus,

H1: Graduate accounting students perceived interest in computer forensics will be positive.
Whereas many accounting subjects are basic to a successful career, computer forensics knowledge is not essential unless the student wishes to gain advanced certification or become an IT auditor. Otherwise, the class may lack relevance and students would not be motivated to take it. Students who are interested in IT auditing and gaining certification are expected to perceive the class as more relevant to their needs and display higher interest. Thus,

H2: Graduate accounting students perceived interest in computer forensics will be positively associated with the intention to acquire advanced certifications.

H3: Graduate accounting students perceived interest in computer forensics will be positively associated with the intention to become an IT auditor.

Students are expected to work interactively in groups identifying and acquiring the specific forensic tools they need to solve auditing problems. This approach is expected to support both concrete and abstract learning styles. Analysis of the audit problem and selection of appropriate tools would represent the abstract while application of the tools and analysis of the results would represent the concrete. As such, a majority of the students will express satisfaction with the exercises that involve forensic tools and will be successful in their selection and application.

H4: Graduate accounting students will be successful in the selection and application of web-based forensics tools to audit applications.

Curriculum design and pedagogical techniques based on the ARCS Motivational Model are expected to be associated with higher levels of attention, relevancy, confidence and satisfaction. Using the learning strategies in Table 1, it is expected that students’ perceived levels for course outcomes will be higher where they perceive these techniques to be present and successfully applied. Thus,

H5: Graduate accounting students will indicate a positive association between the attention indicators and the indicators of course outcomes.

H6: Graduate accounting students will indicate a positive association between the relevancy indicators and the indicators of course outcomes.

H7: Graduate accounting students will indicate a positive association between the confidence indicators and the indicators of course outcomes.

H8: Graduate accounting students will indicate a positive association between the satisfaction indicators and the indicators of course outcomes.

Graduate accounting students are generally highly self-motivated and comfortable with abstract reasoning. Therefore, attention strategies are assumed to be of less importance. Because they are in a career that is in demand, have attained a high level of education, and are seeking an emphasis in an advanced field of study, it is assumed that confidence and satisfaction strategies will be of less importance. However, because graduate accounting students are under pressure to attain advanced certifications must have a fifth year in accounting to be a CPA, and are pursuing an MBA degree, time management is important to them and they are less likely to invest time in courses that are not on their career path. Therefore, relevancy strategies are expected to be more important in curriculum design. Thus,

H9: Graduate accounting students will indicate a higher positive association between the relevancy indicators and the indicators for course outcomes than for the other three design indicators of attention, confidence and satisfaction and the indicators for course outcomes.
<table>
<thead>
<tr>
<th>Planned Learning Approach</th>
<th>A</th>
<th>R</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide various and novel approaches to learning reflecting different learning styles.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Use multimedia in instruction and provide material in lecture and readings.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Clearly state how students will be evaluated.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Base assignments on individual needs and provide examples based on student’s experience and goals.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide clearly stated weekly goals and explain why it is important to achieve these goals.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide immediacy with the students.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Allow flexible approaches to problem solving.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Make personal contact with each learner to insure assignments and modalities are reflecting personal needs.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Assign a group project and encourage collaborative work.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Collect student demographic information on first day of class and tailor assignments to learning styles.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide assignments that reflect real-world forensic auditing tasks.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide a comprehensive set of forensic tools and instructions.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide early assignments that reflect entry experience and provide student some control over outcome.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide concrete and constructive feedback on progress in a timely manner.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Establish guidance on how students can successfully attain goals.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide ongoing technical support in a timely fashion.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Where A–Attention, R–Relevance, C–Confidence, S–Satisfaction
Adapted from Chyung, 2001.

Table 1. Planned Learning Approaches using the ARCS Motivational Model
3.1 Data Collection and Analysis

Data will be collected from graduate accounting students taking the class *Computer Forensics for Accountants* at the University of South Florida St. Petersburg. This class is offered as part of a four-course emphasis on fraud accounting in the MBA program. Data will be collected twice: on the first day and last day of class. Questions will measure the precourse (*a priori*) and postcourse (*ex post*) attitudes for Attention, Confidence, Relevance and Satisfaction. The two research instruments are shown in Appendix 2. Analysis will involve correlation of the indicators and ANOVA to test the hypotheses. Analysis will be performed using the indicators shown in Table 2 that are taken from the post course survey (Appendix 2.b). Measures are collected for each of the four motivational indicators and course outcomes. Measuring the association between the indicators and course outcomes will provide insight in the ability of the curriculum to impact learning outcomes. The following analysis will be performed.

*A priori* and *ex post* students’ perceptions of computer forensics and ANOVA to determine if there is a significant difference.

*A priori* and *ex post* association between students’ intention to acquire advanced degrees and perceptions of computer forensics and ANOVA to determine if there is a significant difference.

1. *Ex post* association between students’ perception of *attention* and course outcomes.
2. *Ex post* association between students’ perception of *relevancy* and course outcomes.
3. *Ex post* association between students’ perception of *confidence* and course outcomes.
4. *Ex post* association between students’ perception of *satisfaction* and course outcomes.
5. ANOVA analysis to determine if the association between *relevancy* and course outcomes is significantly different that between the other three indicators and course outcomes.

4.0 DISCUSSION

Computer forensics has become increasingly important to accountants and to the accounting profession. Educating accountants in the use of forensic tools is essential but difficult. Students perceive forensic tools as complex and find there are no firm guidelines telling them what tools best fit what audit problems. Consequently, the education in forensics is much less structured than the typical accounting course. As such, the instructor must motivate the students to learn in an interactive mode that fits various learning styles. An approach to curriculum design is to develop learning strategies that will motivate students with different learning styles. This paper presents that methodology in an attempt to improve the ability of accounting students to acquire knowledge and skills in computer forensics and learn to apply it to problems that parallel real-world situations. A set of over 50 forensic tools, mostly web-based, are also presented for use in solving realistic audit problems.
<table>
<thead>
<tr>
<th>Survey Qtn.</th>
<th>Measures / Indicators</th>
<th>A</th>
<th>R</th>
<th>C</th>
<th>S</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The course goals were explicitly stated</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I understood why the course goals were important to my career</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>The course evaluation methodology was clear</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A variety of classroom teaching techniques were used</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PowerPoint slides were interesting and useful</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The professor used a variety of methods of instruction</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>The professor related well to the class</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>The professor showed interest in my learning progress</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>The professor responded to my emails promptly</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>The professor engaged the class in active discussion</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>The professor provided assistance when I had questions</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I felt comfortable asking questions and contributing in-class</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>The group assignment helped me to master the material</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>The professor encouraged collaborative work</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Grading of assignments and exams was fair and clear</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>I felt in control of the material and my success in this class</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>I found the guest speakers to be valuable and interesting</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>I am comfortable with computer tools in general</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>The course met or exceeded my expectations</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>I learned new forensic skills in this class</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>I believe IT knowledge and skills are important to my career</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>I learned relevant forensic accounting knowledge in this class</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>The web-based forensic tools were important learning aids</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Applying the forensic tools to audit problems increased my understanding of the forensic techniques</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>I would recommend this class to a friend</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Where A–Attention, R–Relevance, C–Confidence, S–Satisfaction, CO–Course Outcomes

Table 2. Measures of ARCS Indicators Using Post Course Survey Questions
5.0 REFERENCES


APPENDIX 1

ANNOTATED LIST OF COMMANDS

Most of the following tools are available free online. Most require download but some can be run online. All links are provided in a section following the descriptions. Information for downloading and using each tool is provided online. The accompanying links are available at http://www.stpt.usf.edu/bayview/forensics/

DOS Commands

**Disk Operating System** commands are useful for determining the IP and MAC address and examining activity on the computer ports. Some commands offer a variety of switches for providing added information. A full list of commands is available on the web. These commands are typed in the command prompt window that is accessible in the Windows operating system in the Accessories folder.

- `ipconfig /all` - provides system information including the IP and MAC addresses
- `ping [url]` - sends a packet to the url to determine if the address is alive
- `tracert [url]` - lists all hops between the local and the destination computer
- `netstat -b` - lists all open ports and the program that opens each one
- `nslookup [url]` - lists all IP addresses associated with the url
- `psinfo` - list system information, including installed service packs and drive information
- `bootcfg /query` - lists contents of boot.ini file
- `fport` - view of processes mapped to ports

Free Web Tools

**Autoruns** - Windows utility shows you what programs are configured to run during system bootup or login, and shows you the entries in the order Windows processes them. These programs include ones in your startup folder, Run, RunOnce, and other Registry keys.

**BgInfo** - Assists administrators who manage multiple computers. It automatically displays relevant information about a Windows computer on the desktop's background, such as the computer name, IP address, service pack version, and more. You can edit any field as well as the font and background colors, and can place it in your startup folder so that it runs every boot, or even configure it to display as the background for the logon screen.

**CCleaner** - Used to securely delete any browsing history including temp files, cookies and browsing history. Can also wipe your Recycle Bin, Clipboard and memory. For the truly security conscious and paranoid.

**Cookies** - To view and remove cookies in Internet Explorer, click Tools/Internet Options and in Browsing History click Settings and View Files (these are temporary files). Sort and view cookies (type is Text Document). You may delete cookies but may lose some functionality if you have stored passwords.
**Disk Investigator®** - Disk Investigator helps you to discover all that is hidden on your computer hard disk. It can also help you to recover lost data. Display the true drive contents by bypassing the operating system and directly reading the raw drive sectors. View and search raw directories, files, clusters, and system sectors. Undelete previously deleted files. Verify the effectiveness of file and disk wiping programs.

**EasyCrypto®** - You can build self-extracting encrypted and compressed archives and send them to others. The only thing a recipient needs to extract the encrypted files is the correct password. The decrypting and decompressing engine is encapsulated inside the archive. After you encrypt files, EasyCrypto securely wipes originals automatically.

**Eraser®** - Effectively wipes disk areas containing previously deleted files so they cannot be recovered. Can be used for a single file or multiple files.

**Forensic Toolkit®** - Contains several Win32 Command line tools that can help you examine the files on a NTFS disk partition for unauthorized activity. Can identify unauthorized changes to files and hidden files. Provided by Foundstone (a division of McAfee).

**Galleta V L.0®** - Supports investigation of cookie files by importing them into Excel for examination. Provided by Foundstone (a division of McAfee).

**HashCalc** - Checksum calculator with a simple dialog-size interface that creates 13 common checksum algorithms. A checkbox next to each makes it easy to indicate whether the app will calculate the file checksum for that algorithm.

**Hidden Files** - For Windows, go to Control Panel/Appearance and Personalization/Folder Options and check Show Hidden Files, Folders and Drive.

**John the Ripper® (JtR) and FS Crack** - Password auditing tool. JtR is a command driven program and FS Crack is a graphical user interface (GUI) front-end for JtR and includes a user guide.

**K-9®** - Content filtering software for home computers and based on the same technology used in industry. Allows parents to control what content is accessible. Internet content is divided into 60 distinct categories. Administrators can configure the software to block or allow specific categories.

**LCP®** - Password auditing tool with graphical front-end. Individuals and administrators can use it to test the strength of their own passwords.

**Microsoft® Windows® Malicious Software Removal Tool** - computers for and helps remove infections by specific, prevalent malicious software—including Blaster, Sasser, and Mydoom. When the detection and removal process is complete, the tool displays a report describing the outcome, including which, if any, malicious software was detected and removed. The tool creates a log file named mrt.log in the %WINDIR%\debug folder.

**Microsoft® Windows® Computer Investigation Guide for Windows** - Provides a computer investigation model as well as process and best practice information. The guide also provides a fictitious example of an investigation that involves unauthorized access to confidential information. An appendix provides information about how to prepare for computer investigations, sample worksheets, contact information for reporting different types of computer-related crimes to appropriate law enforcement agencies, and lists of useful tools.

**Ophcrack®** - A stronger password recovery tool for alphanumeric passwords. Takes considerably longer to download but valuable for tougher recoveries.

**Fgdump** - Utility for dumping passwords on a Windows system. The user must have administrator access rights. Once dumped, the passwords can be cracked with JtR.

**Free Word and Excel Recovery®** - A tool for recovering passwords to Word, Excel and Access files that have been password protected. Great for security minded individuals who have misplaced their
passwords.

Nessus® - Vulnerability scanner commonly used today by many companies and government organizations. Requires installation of plug-ins. Scans are very thorough. Configuration auditing, asset profiling, sensitive data discovery and vulnerability analysis of your security posture. Performs configuration auditing, asset profiling, sensitive data discovery and vulnerability analysis of your security posture. Always run locally unless you have permission to scan the network.

Network Locator - A utility that approximates and displays the geophysical location of your network address on a Google Map. Not a download.

Nmap - An open source tool for network exploration and security auditing designed to rapidly scan large networks but can be used on single hosts. Displays what hosts are available on the network, what services (application name and version) those hosts are offering, what operating systems (and OS versions) they are running, what type of packet filters/firewalls are in use, and other characteristics. Commonly used for security audits, and routine tasks such as network inventory, managing service upgrade schedules, and monitoring host or service uptime.

Paraben® 2.0 - Allows you to mount your forensic image and explore it as though it were a drive on your machine while preserving the forensic nature of your evidence. In fact, P2 eXplorer is one of the only programs that mounts images as logical and physical disks. This means all the deleted, slack, and unallocated space is accessible. An image isn't just mounted to view logical files; it is mounted as an actual bit-stream image, preserving unallocated, slack, and deleted data.

Pasco® - Supports investigative research of browsing activity by examination of Internet Explorer cache files (index.dat files). Parses data for input into spreadsheet programs for easier analysis. Provided by Foundstone (a division of McAfee).

Phone Number GeoLocator - Provides the geographical location of U.S. phone numbers. Not a download.

Proxify - A web-based anonymous proxy service which allows you to surf the Web privately and securely. There is no software to install. Just enter a URL in the online form. Proxify hides your IP address and the encrypted connection prevents monitoring of your network traffic.

Pwchecker - Evaluates the strength of your current passwords. Provides detailed information about problems that make the password less secure.

Pwdump - Dumps Windows password hashes into a format usable by nearly all password crackers. Different versions are available. Some prefer fgdump.

Recuva® - Recovers files from hard drives, external drives (USB drives, etc.) and memory cards. Recuva can even undelete files from your iPod!

Refog® - Described as being impossible to be seen or removed by your teenage kids or the spouse. Refog is a surveillance tool that collects all key activity on the computer. A freeware is available.

Reverse IP Domain Check - Enter a domain name or IP address and search for other sites known to be hosted on that same web server. A fast and easy way to find the IPv4 address for a domain name. Not a download.

Sam Spade - A network suite of tools including one that will help track unsolicited email.

Sentinel 2.2.1® - Advanced file integrity checker audits the system folder (as well as up to 20 other folders) for the slightest file changes and/or additions. If any files fail the integrity check, the antivirus/trojan program will be notified. Has been deployed at both the enterprise and government sectors to ensure system compliance and security. Easy to use GUI format.

Shields Up® - Free Internet vulnerability profiling and provides your IPv4 address. Easy to use and provides a fast graphical display of the status for all common ports.
Show Hidden Files - Security utility designed to find password protected files of various types - .zip, .rar, Word .doc, Excel .xls, PGP disks. 30 day free trial version.

ShowWin - Identify hidden windows, force windows to stay on top and, in special cases, identify passwords behind those asterisks (******). Provided by Foundstone (a division of McAfee).

Snare for Windows® - Can be used to filter and organize events from Windows Event Viewer and custom rank events by level of severity. Works remotely and makes administration more efficient and effective.

Superscan 4® - Identifies which ports on the computer are open and what programs opened them. Too many open ports makes it easier for hackers to infiltrate your computer. After an application has closed, the associated port should also close. Certain ports are automatically associated with specific applications (Port 80 – Web, Port 20 – FTP, Port 25 – Email, etc.). Caution: You can used Superscan on your own computer but do not scan networks without alerting the network administrator who will be monitoring your own activity. This could appear to be a hacker seeking unprotected ports. Provided by Foundstone (a division of McAfee).

Sysinternals Suite - A large bundle of troubleshooting utilities have been rolled up into a single suite of tools. This file contains the individual troubleshooting tools and help files. A must.

Tcpview - A Windows program that will show you detailed listings of all TCP and UDP endpoints on your system, including the local and remote addresses and state of TCP connections.

Trout® - A tracert and Whois program that can identify hosts on the route to a destination computer. Provided by Foundstone (a division of McAfee).

Vision V 1.0 - Lists all open ports, maps them to their applications and provides supplementary detailed information for analysis of activity. Interrogates ports and identifies possible probes. Provided by Foundstone (a division of McAfee).

Web Sleuth® - Sleuth is a Web Application Assessment Tool. It can be used to great effect by both auditors and developers alike to debug/analyze Web Applications for function & security. 20 day free trial version.

WhoIs Lookup Tool - Performs a WHOIS lookup to provide information such as the owner of a domain name or an IP address on the Internet. This tool is limited to .com, .net, and .edu domains. Not a download.

Windows Event Viewer® - Organizes logged events on Windows OS. Go to Control Panel/Administrative Tools/Local Security Policy/Audit Policy and click Audit Account Logon Events and select both Success and Failure. For more detail go to the Microsoft page for this utility.

Wireshark Packet Sniffer - Intercepts packets on the network including those that are not yours. Primarily used by network administrators for monitoring network activity, they are also used by hackers for intrusion into private files. Download Wireshark I for an introduction. You must also download WinPCap to actually capture the network packets. Wireshark II allows you to filter out unwanted packets, and Wireshark III can be used to intercept email. Caution: Do not use these without alerting the network administrator who will be monitoring your own activity. These tools should never be used to capture packets that do not belong to the user.

Comercial Software for Forensic Analysis

Anonymizer Anonymous Surfing - Hides your IP address, protecting your privacy and securing your identity. Online snoopers are unable to track the sites you visit and build profiles on your Internet activities. 30 day free trial version.

EnCase® - Utility to acquire data in a forensically sound manner, using software that has been
accepted in courts worldwide. Limit incident impact and eliminate system downtime with immediate response capabilities. Efficiently collect only potentially relevant data upon eDiscovery requests.

**Hide the IP 2010®** - Allows user to hide the IP address to anonymize web activity.

**OnTrack Data Recovery®** - Commercial data recovery including disk, tape, email, all file formats and operating systems, lost and damaged files.

**Phone Search Central** - Not a web tool but a web-based investigative service that guarantees results. Specializes in the reversal of unlisted/non published cellular and landline telephone numbers throughout USA, Canada, Great Britain and Ireland. Locates telephone number information from active or disconnected numbers, mobile phones, home phones, unlisted numbers, US based phone numbers and Canadian phone numbers. Any kind of phone number can be reversed to its owner.

**SafeBack** - An industry standard self-authenticating computer forensics tool that is used to create evidence grade backups of hard drives. Creates bit-stream backup files of hard disks or a mirror-image copy of an entire hard disk drive or partition. The image files cannot be altered or modified to alter the reproduction.

**Spector 360®** - A surveillance tool that records web sites visited, emails sent and received, chats and instant messages, keystrokes typed, files transferred, documents printed and applications run. Shows exact visual detail employee online activities. Monitors web and computer activities from each of your employees, feeds that information into a database and provides you with more than 50 built-in reports. A lower-priced version is available for home use. All employees should be aware of the detail that employers can amass for possible use in a theft-of-services suit. Site includes informative demos.
APPENDIX 2.A Pre Course Survey

GRADUATE ACCOUNTING STUDENT PRECOURSE IT INTEREST SURVEY

Instructions: The purpose of this survey is to gather information that will allow your professor and the Program of Accountancy to make better informed choices about the materials included in the curriculum. Your individual answers are confidential and will not be made available to anyone other than the professor. Under no circumstances will an individual student’s responses be identified or used as part of his or her grade. Please answer all questions as accurately as possible.

Do you plan to pursue advanced certifications such as a CPA, CMA, or CIA?  YES  NO  MAYBE

Would you consider pursuing a certification as an IT Auditor?  YES  NO  MAYBE

Have you taken the course in Fraud Accounting?  YES  NO

Have you taken the course in Forensic Accounting?  YES  NO

| Circle the number that best describes your level where 1 = Low, 4 = Neutral, 7 = High |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1  I prefer lectures using PowerPoint presentations                  | 1  2  3  4  5  6  7 |
| 2  I prefer in-class discussions                                     | 1  2  3  4  5  6  7 |
| 3  I prefer working independently                                    | 1  2  3  4  5  6  7 |
| 4  I prefer working in a group                                       | 1  2  3  4  5  6  7 |
| 5  I prefer in-class problem solving                                 | 1  2  3  4  5  6  7 |
| 6  I prefer active participation in class                           | 1  2  3  4  5  6  7 |
| 7  I prefer learning material that has immediate application        | 1  2  3  4  5  6  7 |
| 8  I am comfortable with abstract material                          | 1  2  3  4  5  6  7 |
| 9  I need to know how the material is relevant to my career          | 1  2  3  4  5  6  7 |
| 10 I prefer to approach problem-solving in a structured manner       | 1  2  3  4  5  6  7 |
| 11 I need constant feedback on my performance                       | 1  2  3  4  5  6  7 |
| 12 I generally have difficulty in setting up problem solutions      | 1  2  3  4  5  6  7 |
| 13 I enjoy learning new skills                                       | 1  2  3  4  5  6  7 |
| 14 I can usually learn a subject by reading about it                 | 1  2  3  4  5  6  7 |
| 15 I learn best in a highly structured environment                   | 1  2  3  4  5  6  7 |
| 16 I value the opinions of my classmates on the subject material     | 1  2  3  4  5  6  7 |
| 17 I am comfortable with computer tools in general                   | 1  2  3  4  5  6  7 |
| 18 I enjoy learning about computer tools                             | 1  2  3  4  5  6  7 |
| 19 I believe IT knowledge and skills are important to my career       | 1  2  3  4  5  6  7 |
| 20 I believe computer forensics knowledge is important to my career   | 1  2  3  4  5  6  7 |

| Circle the number that best describes your level of knowledge or skill where 1 = None, 7 = Very High |
|-------------------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 19  MS Windows Operating System                   | 1  2  3  4  5  6  7 |
| 20  MS Windows Word, PowerPoint, Excel            | 1  2  3  4  5  6  7 |
| 21  Web-based computer utilities                 | 1  2  3  4  5  6  7 |
| 22  Accounting Software Packages (any)           | 1  2  3  4  5  6  7 |
| 23  Financial auditing                            | 1  2  3  4  5  6  7 |
| 24  Forensic auditing                             | 1  2  3  4  5  6  7 |
APPENDIX 2.B Post Course Survey

GRADUATE ACCOUNTING STUDENT PRE-COURSE IT INTEREST SURVEY

Do NOT write your name on this form. This survey is anonymous.

Instructions: The purpose of this survey is to gather information that will allow your professor and to make better informed choices about the materials included in the curriculum. All responses are completely anonymous. Please answer all questions as accurately as possible. If a question does not pertain to your classroom experience, leave it blank.

Do you plan to pursue advanced certifications such as a CPA, CMA, or CIA?
Would you consider pursuing a certification as an IT Auditor?
Have you taken the course in Fraud Accounting?
Have you taken the course in Forensic Accounting?

Circle the number that best describes your level of agreement where 1 = Low, 4 = Neutral, 7 = High

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The course goals were explicitly stated</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>I understood why the course goals were important to my career</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>The course evaluation methodology was clear</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>A variety of classroom teaching techniques were used</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>PowerPoint slides were interesting and useful</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>The professor used a variety of methods of instruction</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>The professor related well to the class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>The professor showed interest in my learning progress</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>The professor responded to my emails promptly</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>The professor engaged the class in active discussion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>The professor provided assistance when I had questions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>I felt comfortable asking questions and contributing in-class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>The group assignment helped me to master the material</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>14</td>
<td>The professor encouraged collaborative work</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>Grading of assignments and exams was fair and clear</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>I felt in control of the material and my success in this class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>I found the guest speakers to be valuable and interesting</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>I am comfortable with computer tools in general</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>The course met or exceeded my expectations</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>20</td>
<td>I learned new forensic skills in this class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>21</td>
<td>I believe IT knowledge and skills are important to my career</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td>I learned relevant forensic accounting knowledge in this class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>The web-based forensic tools were important learning aids</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>24</td>
<td>Applying the forensic tools to audit problems increased my understanding of the forensic techniques</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>I would recommend this class to a friend</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
Subscription Information

The Proceedings of the Conference on Digital Forensics, Security and Law is a publication of the Association of Digital Forensics, Security and Law (ADFSL). The proceedings are published on a non-profit basis.

The proceedings are published in both print and electronic form under the following ISSN's:

ISSN: 1931-7379 (print)
ISSN: 1931-7387 (online)

Subscription rates for the proceedings are as follows:

Institutional - Print & Online: $120 (1 issue)
Institutional - Online: $95 (1 issue)
Individual - Print: $25 (1 issue)
Individual - Online: $25 (1 issue)

Subscription requests may be made to the ADFSL.

The offices of the Association of Digital Forensics, Security and Law (ADFSL) are at the following address:

Association of Digital Forensics, Security and Law
1642 Horsepen Hills Road
Maidens, Virginia 23102
Tel: 804-402-9239
Fax: 804-680-3038
E-mail: office@adfsl.org
Website: http://www.adfsl.org