Conference on Digital Forensics, Security and Law

Proceedings of the Conference on Digital Forensics, Security, and Law 2009

Burlington, Vermont
May 20-22

Conference on Digital Forensics, Security and Law

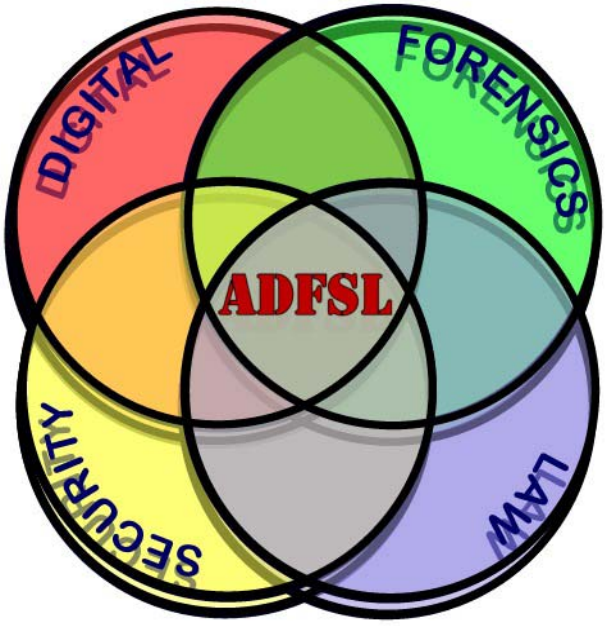
Burlington, Vermont
May 20-22, 2009

Conference Organizer

Glenn S. Dardick
Longwood University
Virginia, USA

Chair

Gary C. Kessler
Champlain College
Vermont, USA



ADFSL

Association of Digital Forensics, Security and Law

ISSN 1931-7379

Sponsors


1878
CHAMPLAIN
COLLEGE


LONGWOOD
UNIVERSITY


COURSE TECHNOLOGY
CENGAGE Learning

Contents

Conference Committee

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

Gary Kessler
gary.kessler@champlain.edu
General chair
Champlain College
Vermont
USA

Linda Volonino
volonino@canisius.edu
Program chair
Canisius College
New York
USA

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

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

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

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

Jane Faust
Faust@champlain.edu
Local arrangements chair
Champlain College
Vermont
USA

Andy Jones
andrew.28.jones@bt.com
International Advisor Europe
British Telecom
UK

Jigang Liu
Jigang.liu@metrostate.edu
Program Committee Member
Metroplitan State University
Minnesota
USA

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

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

Pedro Luís Próspero Sanchez
pedro.sanchez@poli.usp.br
plpsanchez@gmail.com
International Advisor South America
University of Sao Paulo
Brazil

Sizwe Lindelo Snail
SizweS@couzyn.co.za
International Advisor Africa
Couzyn Hertzog & Horak
South Africa

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

Iain Sutherland
isutherl@glam.ac.uk
International Advisor Europe
University of Glamorgan
Wales
UK

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

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

Schedule

Wednesday, May 20

- 08:30 AM CONTINENTAL BREAKFAST
- 08:30 AM On-site Registration
- 09:00 AM Conference workshop
  - *Diane Barrett: Workshop on How the Acceptance of Anonymous Surfing and Tor in Communications has Changed the Evidence Landscape*
- 11:45 AM LUNCH (provided)
- 01:00 PM Conference Welcome: Dr. Dave Finney, President, Champlain College, Gary Kessler, Linda Volnino
- 01:15 PM Keynote presentation: Fred Lane, attorney, author, practitioner
- 02:15 PM BREAK
- 02:30 PM Papers/Presentation session
  - *Owen O'Connor: Cloud Computing & Digital Investigations*
  - Craig Valli: Visualization of honeypot data using Graphviz and Afterglow
- 03:40 PM BREAK
- 03:55 PM Papers/Presentation session
  - *Grover Kearns: Graduate Accounting Students' Perception of IT Forensics: A Multi-Dimensional Analysis*
  - *Anil Aggarwal: Pedagogical Issues in Digital Forensics: A Case Study* -

Thursday, May 21

- 08:30 AM CONTINENTAL BREAKFAST
- 08:30 AM On-site Registration
- 09:00 AM Keynote presentation (Paul van de Graaf, Acting U.S. Attorney, District of Vermont)
- 09:45 AM BREAK
- 10:00 AM Papers/Presentation session
  - *Gareth Davies: The Impact of Hard Disk Firmware Steganography on Computer Forensics*
  - *Damir Kahvedzic: Analysis of the 'Db' Windows Registry Data Structure*
  - *Damir Kahvedzic: Correlating Orphaned Windows Registry Data Structures*
- 11:45 AM LUNCH (provided)
- 01:00 PM Papers/Presentation session
  - *Linda Volonino: Don't Touch That! and Other E-Discovery Lessons*
  - *Andy Jones: Why are we not getting better at Data Disposal?*
  - *Vicki Luoma and Milt Luoma: The Computer Fraud and Abuse Act and the Law of Unintended Consequences*
- 02:10 PM BREAK
- 02:30 PM Papers/Presentations session
  - *Stephen Larson: Concerning File Slack*
  - *Abbas Cheddad: Data Hiding Tools for Digital Forensics Experts*
  - *Dennis Browning: Bluetooth Hacking: A Case Study*
- 05:15 PM SOCIAL EVENT (sunset hor d'oeuvres cruise on Lake Champlain)

Friday, May 22

- 08:30 AM CONTINENTAL BREAKFAST
- 09:00 AM Papers/Presentations session
  - *Nigel Wilson: Digital Forensics Law in Australia – What's Happening Down Under*
  - *Denis George Edgar-Nevill: Cybercrime and the 2012 London Olympics*
  - *Jonathan T Rajewski: Methodology for Investigating Individuals Online Social Networking Persona*
- 10:45 AM BREAK
- 11:00 AM Conference Close: Gary Kessler, Glenn S. Dardick

Workshop: How the Acceptance of Anonymous Surfing and Tor in Communications has changed the Evidence Landscape

Diane Barrett
Associate Professor
University of Advancing Technology
Tempe, AZ
Dbarrett@uat.edu

ABSTRACT

This workshop will explore how the acceptance of anonymous surfing and Tor in communications has changed the way investigators look for browsing and IP address evidence. It will begin by explaining how these technologies are currently used and then move on to a demonstration of several of programs with the intention of fostering discussion on the use of these programs as they become more main stream. It will then demonstrate some of the programs, evidence of remnants and encourage discussion on the best way to find evidence when these types of programs are used. Finally, it will conclude with discussion of some best practices for finding and tracking evidence when these programs and method are used.

DESCRIPTION

Anonymous browsing and the use of Tor (The Onion Router) at one time were considered anti-forensic methods. Onion routing was touted as a way to defeat being tracked by protecting the user from traffic analysis. Running a Tor node, including a Tor exit node that allows people to anonymously send and receive traffic, is lawful under U.S. law. More recently, the use of these tools appear to be acceptable and are used in devices such as the IronKey. The IronKey website explains Tor as a useful tool: "Tor (The Onion Router) helps organizations and people to improve their safety and security on the Internet." The IronKey works by tunneling web browsing communications through the Tor-based Secure Sessions proxy on the IronKey. This Tor-based network consists of high-speed dedicated router nodes for IronKey customers only. All IronKey products are FIPS 140-2 Level 2 certified. Clearly the use of such tools had changed. Devices such as these are touted to law enforcement and governments to keep date secure. Additionally, IronKey partnered with Moka5, a desktop virtual computing company targeting consumers and small to medium-sized businesses, allowing Moka5 to bundle and resell IronKey Secure Flash Drives with the Moka5 Engine loaded on the devices. We now have secure, virtual environments running on a Tor network.

Anonymous browsing has also become widely accepted. Applications and browsers that run from USB devices keep all browsing history on the drive. Companies that provide anonymizers market them as a way to protect your privacy and to keep records from being stored by Internet Service Providers (ISP's) for the purpose of being bought and sold to interested parties.

Currently there are about 100 different anonymizer programs available as well as virtual environments that can either mask or conceal Internet traffic. Anonymizers can be broken down into two categories: networked and single-point. Networked anonymizers transfer communications through a network of Internet computers between the user and the destination. Single-point anonymizers pass surfing through a single web site to protect the user's identity. This is often done through an encrypted communications channel.

Traditionally, a user's browsing activity is stored locally in cached web page with the corresponding URL the user visited. This relationship is mapped in the Index.dat file. The other files that reveal activity about the users Internet activity are cookies, history, and temporary Internet files. When using an anonymizer, this evidence no longer exists. The investigator now has to find different ways to find

evidence to tie the suspect to the activity.

Three anonymizer programs will be demonstrated: one portable application, one networked and one single-point anonymizer. All history, temporary Internet files, and cookies will be cleared and a portable secure browser such as QTweb will be installed, set to private mode, some Internet surfing will be done and then a program such as Mandiant Web Historian will be used to extract information. The demonstration will show what files QTWeb creates and how what it does with the user browser data. A similar demonstration will be done using Tor as the networked anonymizer and one single-point anonymizer such as Anonymouse. After the demonstrations, there will be discussion about experiences from the conference participants.

Next there will be a presentation of the evidence and remnants from the programs demonstrated. For example, tools can be used to dump the contents of all registry keys opened or modified by Tor. For each of these keys, the state before and after Tor is started can be saved. Then, by comparing the two files it is possible to find registry keys modified by Tor. Discussion will follow in regard to how to track users. According to Tor's website, there is nothing that can be done to trace Tor users. The same mechanisms that keep criminals from breaching Tor's anonymity prevent investigators from having access to this information. If by default, a Tor server keeps no log of the packets it transmits, it may be possible to go to the previous hop in the chain for the examination of a node to determine whether the suspected computer is the request originator or only acted as a relay. The Tor authors have decided against putting in any type of backdoor and instead suggest using traditional police techniques such as interviewing suspects, surveillance and keystroke loggers. Discussion should ensue about the viability of information requests since properly configured Tor relays most likely will not have useful data for the inquiring parties. Log files maintained by Tor relay operators may not be required to be turned over due to violation of the Electronic Communications Privacy Act or other data protection laws

One way of tracing anonymous traffic is to use the Metasploit decloaking engine. The engine is a system for identifying the real IP address of a web user, regardless of proxy settings. A quick demonstration of this will follow along with discussion on whether this is a viable solution based on how the process work. Although no vulnerabilities are exploited by this tool, if the Tor setup is properly configured, no identifying information will be exposed.

Lastly, some alternate methods of tracing traffic and an ensuring discussion on best practices to find evidence when these types of programs are used will complete the workshop.

Presentation: Cloud Computing & Digital Investigations

Owen O'Connor
Champlain College Dublin
Ireland

Cloud computing is being adopted for a wide range of business functions, from replacing traditional applications such as Customer Relationship Management to performing large processing tasks using low-cost resources accessed over the Internet. The efficiency benefits and cost savings from cloud computing have led to the development of a large ecosystem of cloud computing providers based on the concepts of Software-as-a-Service (e.g., Salesforce.com, Google Apps, Microsoft Hosted Exchange), "Platform-as-a-Service" (e.g., Google App Engine, Microsoft Azure, Force.com) and Infrastructure-as-a- Service or utility computing (e.g., Amazon Web Services, Slicehost / Mosso, Flexiscale).

Just as business processes and typical IT tasks can benefit from these services, digital investigations can also be made more efficient by adopting utility computing and other aspects of cloud computing. In particular many of the mathematically-intensive functions in digital forensics can be carried out far more quickly using utility computing services such as Amazon EC2, for example calculating cryptographic hashes, performing optical character recognition and carrying out text searches (either directly or via indexing). Digital investigators should also be aware of the potential uses of cloud computing in online investigations, for example to enable complex searches of online content, to assist with covert investigations and to provide temporary virtual servers for online monitoring or content hosting.

This presentation will review the current state of cloud computing, outline the services of leading services and present examples of cloud computing usage in business. The relevance of cloud computing to digital investigations will then be explained, covering the potential for bulk data processing "in the cloud", the use of utility computing services in online investigations and the investigative benefits of Amazon's Alexa Web Services and other open web services. Finally the potential risks of cloud computing will be discussed, focusing on the need for risk analysis process to determine where cloud computing may be appropriate and how risks to confidentiality and forensic integrity can be addressed.

VISUALISATION OF HONEYPOT DATA USING GRAPHVIZ AND AFTERGLOW

Craig Valli
secau – Security Research Centre
Edith Cowan University
c.valli@ecu.edu.au

ABSTRACT

This research in progress paper explores the use of Graphviz and Afterglow for the analysis of data emanating from a honeypot system. Honeypot systems gather a wide range of data that is often difficult to readily search for patterns and trends using conventional log file analysis techniques. The data from the honeypots has been statically extracted and processed through Afterglow scripts to produce inputs suitable for use by the DOT graph based tools contained within Graphviz. This paper explores some of the benefits and drawbacks of currently using this type of approach.

Keywords: honeypot, network forensics, visualization, Graphviz, Afterglow

1. INTRODUCTION

Honeypots generate a large amount of raw data for analysis and investigation by network security professionals. This raw data is typically rich in content and volume due to the *modus operandi* of network honeypots. This captured data can include but not be limited to system log files, network intrusion detection system log files, binary capture files and also malware candidate files.

On a relatively low interaction honeypot this data can run to several megabytes of textual and binary data per day. The analysis and trapping of malfeasance for which a honeypot system is designed is not an ideal match to traditional logfile analysis tools and techniques. Standard log file processors can only achieve so much in their ability to interpret the textual data that is captured by a honeypot into its log files. While log file processors have traditionally rich analysis algorithms they tend to display the top 10 or 100 in any selected investigated category or activity enabled by a suitable configuration of the underlying engine to ascertain this. While this *modus operandi* maybe a viable method for log file analysis based from a per use perspective such as world wide web access monitoring by business users it is typically not suited to analysis of malicious activity. This lack of viability is because often malicious activity by individuals and even the malicious code (malcode) they produce to enable an attack is performed or developed so as not to be detected. This posture of low detection is deleterious to how conventional log file analysis works.

Even dedicated intrusion detection query engines such as ACID, BASE or Surfnet IDS can have trouble locating relevant data within the mass of data that a honeypot will produce on a daily basis. The *modus operandi* of some attacks for instance may only utilize a single bot or compromised host for 1 or 2 interactions on a target. Some of the malcode will use the same attack vector typically the same port or initializing data stream but control 1000s of bots to cause a denial of service or overwhelm a host. Textually this is often hard for the human to see the associations and due to sometimes the high level of traffic this generates significant text is also generated.

The use of customized text scripts while a potentially optimal outcome for a particular scenario typically does not scale well in a live analysis scenario or is often not suitable for reuse. The exploration of alternative methods for interpreting and responding to honeypot data is necessary to

progress the state of network security. Firewalls and other perimeter countermeasures are starting to show their age in the same way as the bastilles and castles of the middle ages similarly became in effective against new attacks. This paper is one such exploration of using Graphviz and associated log file processing scripts based around the AfterGlow suite to map certain types of honeypot data for interpretation by an investigator or researcher.

2. WHAT IS GRAPHVIZ AND AFTERGLOW?

Graphviz is an open source visualization utility developed to generate a variety of graph layouts (Ellson and Gansner 2008). There are several base utilities *neato* which makes "spring model" layouts, *twopi* which generates radial layouts and *circo* which generates circular layouts. They all interpret files that have been described using the DOT language. Afterglow is a series of PERLscripts designed to be used with Graphviz to generate link graphs from Comma Separated Values (CSV) formatted files (Marty 2007). The Afterglow conversion scripts provided take a raw input file (tcpdump, Snort, iptables and Argus logfiles), analyze it and output a comma separated list of records for each of the particular formats. The resulting records are based on the data present in the file and what the investigator would like represented from the complete record via configuration. Essentially the process is one of normalisation of the data into meaningful sets for analysis and subsequent graphical representation.

These customized scripts are not the only method of extraction of the records. Extraction can be accomplished manually using a spreadsheet program a slow and somewhat cumbersome event and also limited to processing 65535 or the row limit of the spreadsheet package in use. The use of spreadsheets is not suited to automatic generation of the graphs direct from the network interface or log file stream which is a long term aim of this research. The use of scripts could also process these interactions live into a database structure in similar fashion that occurs already with Surfnet IDS.

After being processed by the conversion scripts or processes the extracted CSV files are then fed through scripts that produce one of two formats. Either it generates a DOT attributed graph language file - the input required by the graphviz library - or it can generate input for the large graphing library (LGL) that is used extensively in bioinformatics. This experimentation was focused around the production of DOT attributed graphics for processing via Graphviz.

Afterglow accepts either 2 or 3 columns of raw input data to map. For network forensics the 3 column format is the most useful as it can be used to map interaction/events between two entities or attributes within a networked exchange. This operation is particularly useful in tracking or visualizing behaviors that are hard to interpret from text files or represent readily by other graphical or statistical methods.

3. THE HONEYPOT SYSTEM

A nepenthes honeypot system was used as the source data for this research (Baecher, Koetter et al. 2009). A nepenthes based honeypots purpose is to pose deceptively as a vulnerable target platform for malcode to interact with. It achieves this by the emulation of known vulnerability for example MS03-26 (Microsoft 2003) and other known security exploits that allow for upload of malcode or the execution of arbitrary code on the victim computer. Nepenthes via subsequent selective emulation of the known exploit to the malfeasant agent attempts to enable the successful download or transmission of malcode payload for subsequent capture and post incident analysis. Post incident analysis is normally the static and dynamic analysis of the malcode using specialized tools such as disassemblers.

Nepenthes logs and traps a wide range of data types and can store it in a wide range of formats for use by analysts. Nepenthes systems are also typically hardened and use firewalling to redirect or deny connections of interest to the honeypot system. The utilization of a system such as the Surfnet IDS system also extends the dataset that is available to the network forensics researcher. Surfnet not only logs data in a variety of formats but also processes captured malcode and runs it against virus scanners

and sandboxes. It also allows aggregation of nepenthes data and external data from sandbox outputs from Anubis and CWSandbox for further analysis.

Table 1 is some example raw logged data from a nepenthes honeypot. The following data is extracted from the logged_downloads file that records when a piece of malware has been downloaded by nepenthes.

```
[2007-05-15T23:45:58] 203.129.220.98 -> 203.161.117.122 link://203.129.220.98:39770/RhhlQg==
[2007-05-15T23:45:58] 203.129.220.98 -> 203.161.117.122 link://203.129.220.98:39770/RhhlQg==
[2007-05-16T03:51:22] 203.136.73.89 -> 203.161.117.122 link://203.136.73.89:40058/+Hcs+A==
[2007-05-16T04:31:52] 203.161.114.141 -> 203.161.117.122 tftp://0.0.0.0/WinEUM.exe
[2007-05-16T04:46:29] 203.145.168.45 -> 203.161.117.122 tftp://0.0.0.0/wbemstest.exe
[2007-05-16T06:34:38] 203.161.114.141 -> 203.161.117.122 tftp://0.0.0.0/WinEUM.exe
[2007-05-16T07:42:32] 203.88.202.98 -> 203.161.117.122 link://203.88.202.98:50004/OOci+A==
[2007-05-16T12:52:57] 203.124.171.18 -> 203.161.117.122 link://203.124.171.18:18907/OMkq+A==
```

Table 1 – Raw data from nepenthes

Its basic format is Date & Time expressed as YYYY-MM-DD (T) HH:MM:SS, Source IP, Destination IP, Protocol (Serving IP:Port)/PayloadName. This richness of record enables a wide amount of alternate lenses for analyzing data using the Graphviz engine for analysis. Table 2 is a partial representation of combinations as there are 24 possible combinations from this log file format excluding dates, inclusive of dates there would be 120.

Source IP	**Destination IP**	**Payload**
203.161.114.141	203.161.117.122	WinEUM.exe
Source IP	**Destination IP**	**Protocol**
203.161.114.141	203.161.117.122	Tftp
Source IP	**Destination IP**	**PayloadIP**
203.161.114.141	203.161.117.122	0.0.0.0
Source IP	**PayloadIP**	**Destination IP**
203.161.114.141	0.0.0.0	203.161.117.122
Payload	**Source IP**	**Destination IP**
WinEUM.exe	203.161.114.141	203.161.117.122
Protocol	**Source IP**	**PayloadIP**
Tftp	203.161.114.141	0.0.0.0

Table 2 – Some combinations from logfile data

Some of the combinations would clearly not be useful for analysis and other would produce very similar graphical representations of the data. The remaining meaningful combinations in turn do give a different view of the data in visualized form when compared to static graphics and textual outputs of conventional analysis tools. A selection of these are displayed on the following figures.

It can be clearly seen in Figures 1 through 4 that the same data set can have many different visualizations or lenses for analysis of the dataset. Each of the visualizations tells a slightly different story by the lense or representation that it provides the investigator into the raw data set. The ability to use different lenses allows an investigator to view data rapidly and from different viewpoints that can significantly aid in interpretation, detection and potential amelioration of issues or events. This potential for rapid interpretation that a graphical data view provides is simply not readily achievable through text based analysis. It is easy to see different holistic patterns in the presented data even though in each figure they are drawn from the same dataset, they just have a different articulation or causal linkage when they have been interpreted and represented by the Graphviz utilities.

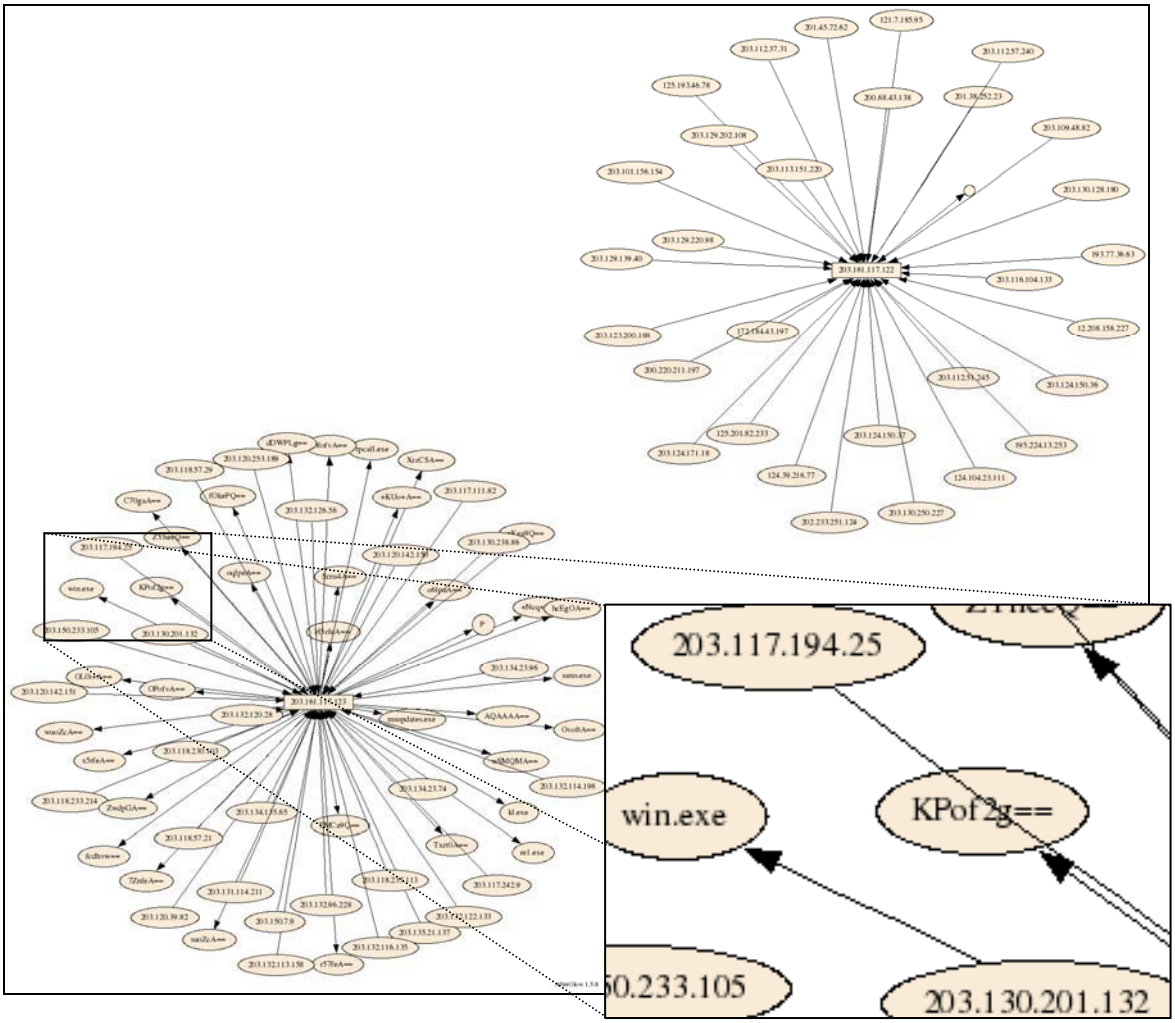


Figure 1. SIP PIP Payload

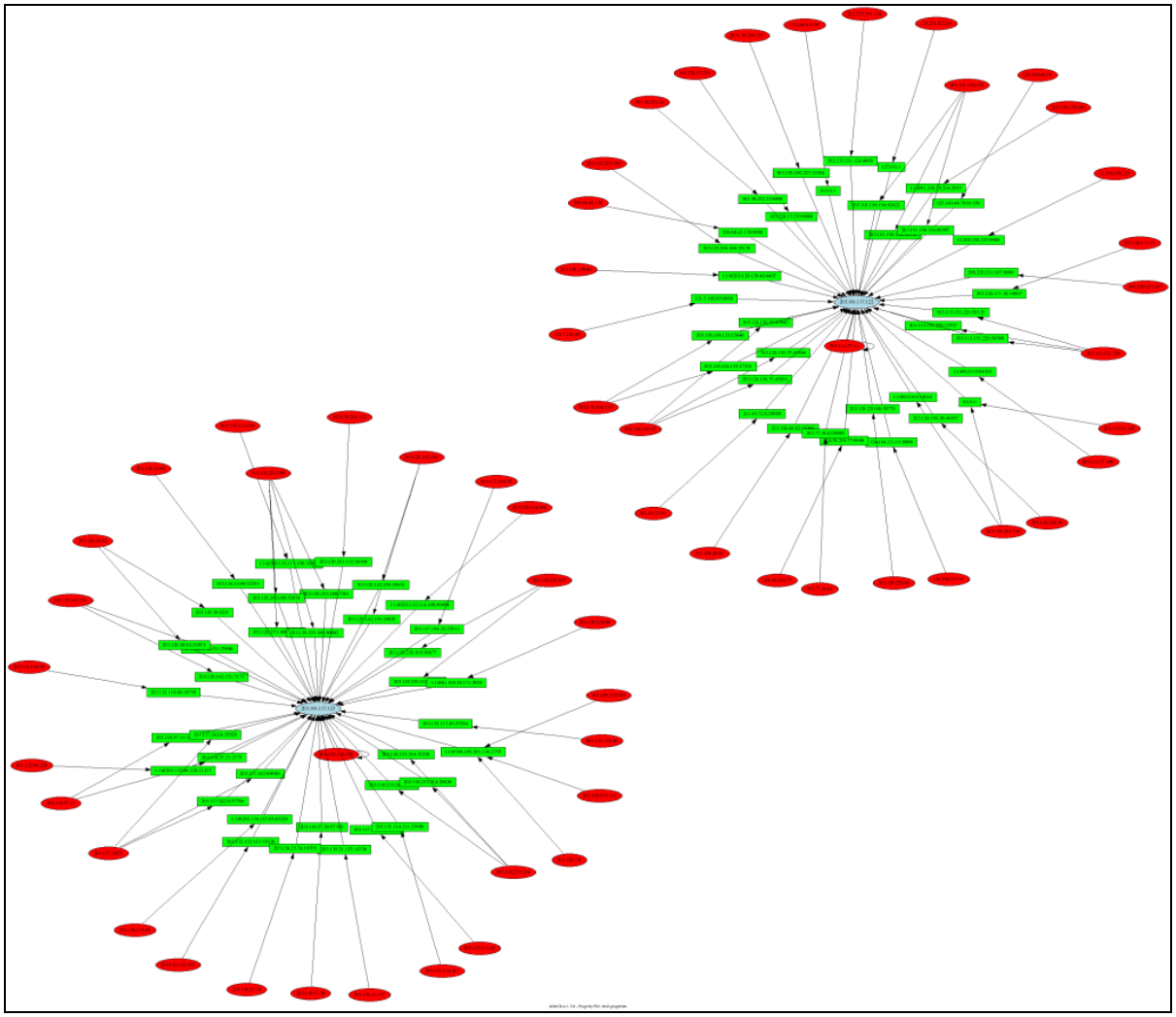


Figure 2. SIP PIP DIP

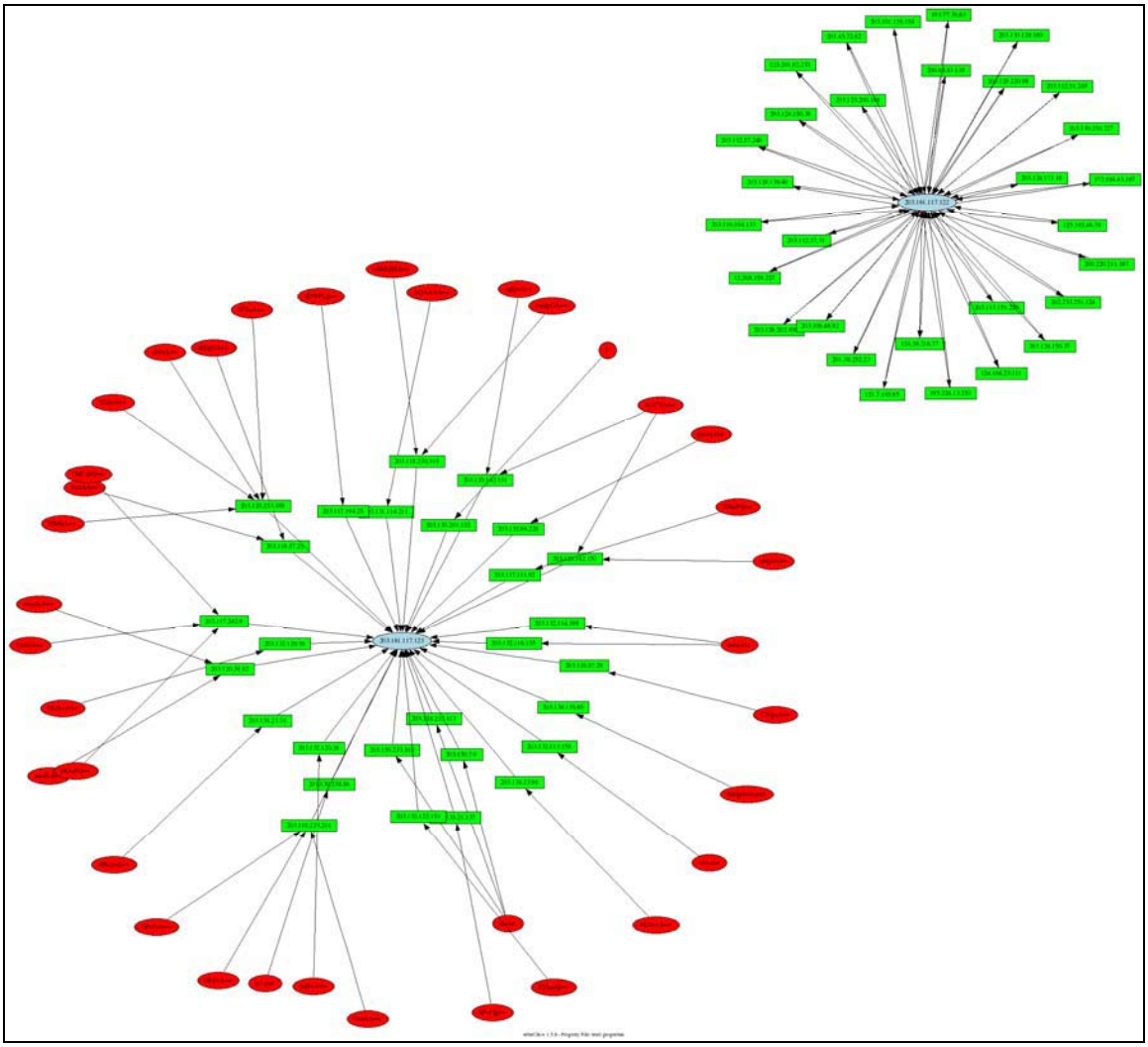


Figure 3. Payload SIP DIP

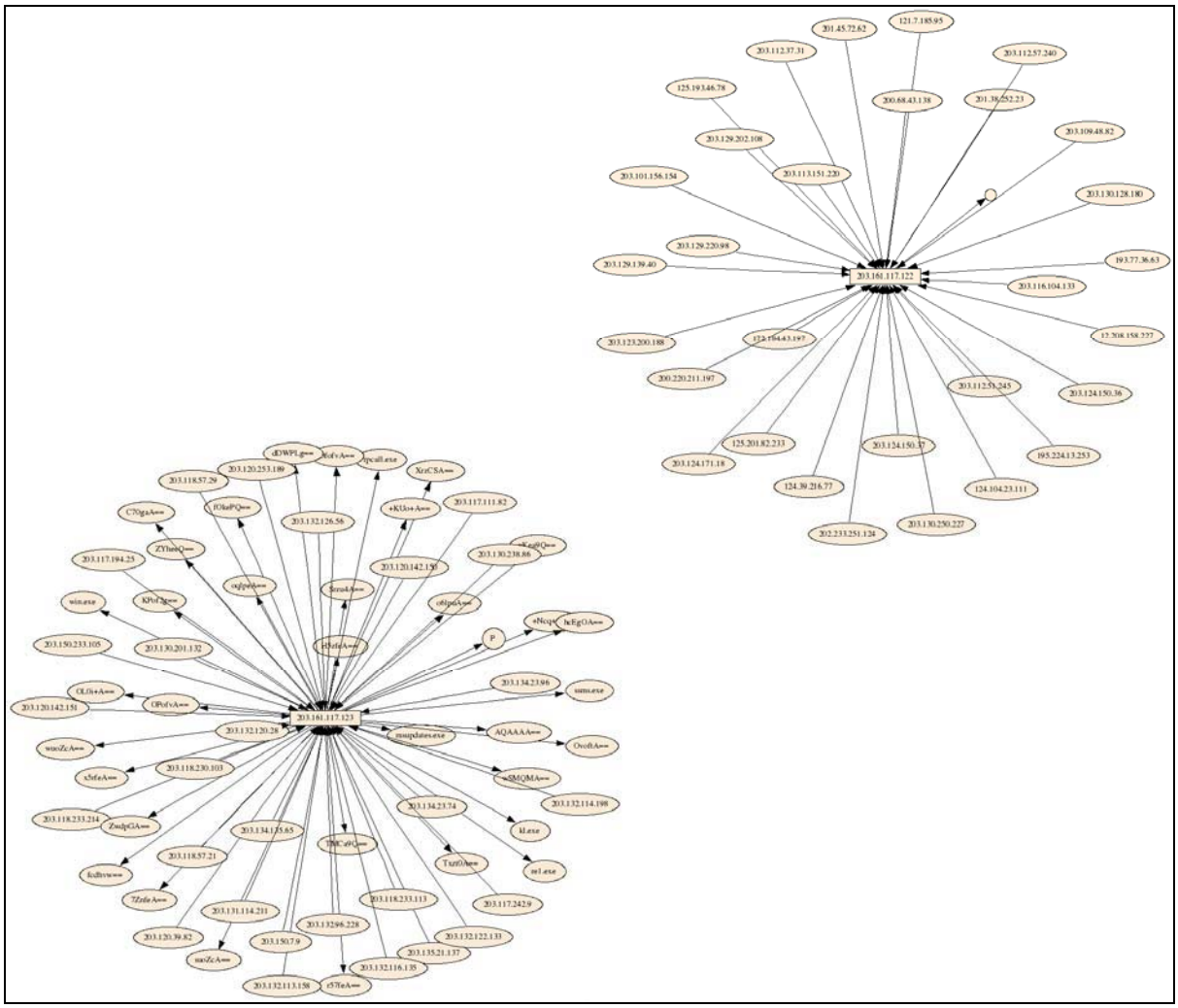


Figure 4. SIP PIP Payload

The graphs represented here are by no means the largest generated in this initial research. One graph that used the complete set of data generated a 35000 x 33000 pixel bitmap file. Even with this large file it was clear to see from a holistic perspective the epicenters or areas of high network activity. Figure 5 is a very reduced snapshot of that file.

The graphical nature of the method as clearly demonstrated in Figure 5 demonstrates it is relatively easy to pick out areas of concentration of attack. This method elucidates how also this type of graphical representation performs clustering and subsequent visualization of like events experienced on honeypot systems. These groupings unlike conventional methods of interpretation do not suffer from the same temporal separation that other log file analysis tools can have for instance if these attacks were perpetrated across a range of weeks. The focus for the generation of the graphics is the entity, with the size of the ring or circle in this case indicating magnitude and depth of attack. It is far easier to track extended attacks using these methods of analysis and presentation than conventional log file analysis techniques due to the visual clustering of the data representations.

The graphical nature of these types of outputs provides a medium for the rapid interpretation of patterns and emergent trends. The use of DOT graphs generated from Graphviz provides a method for making a particular entity or entities the focal point of the analysis in a consistent visual manner for the human analyst.

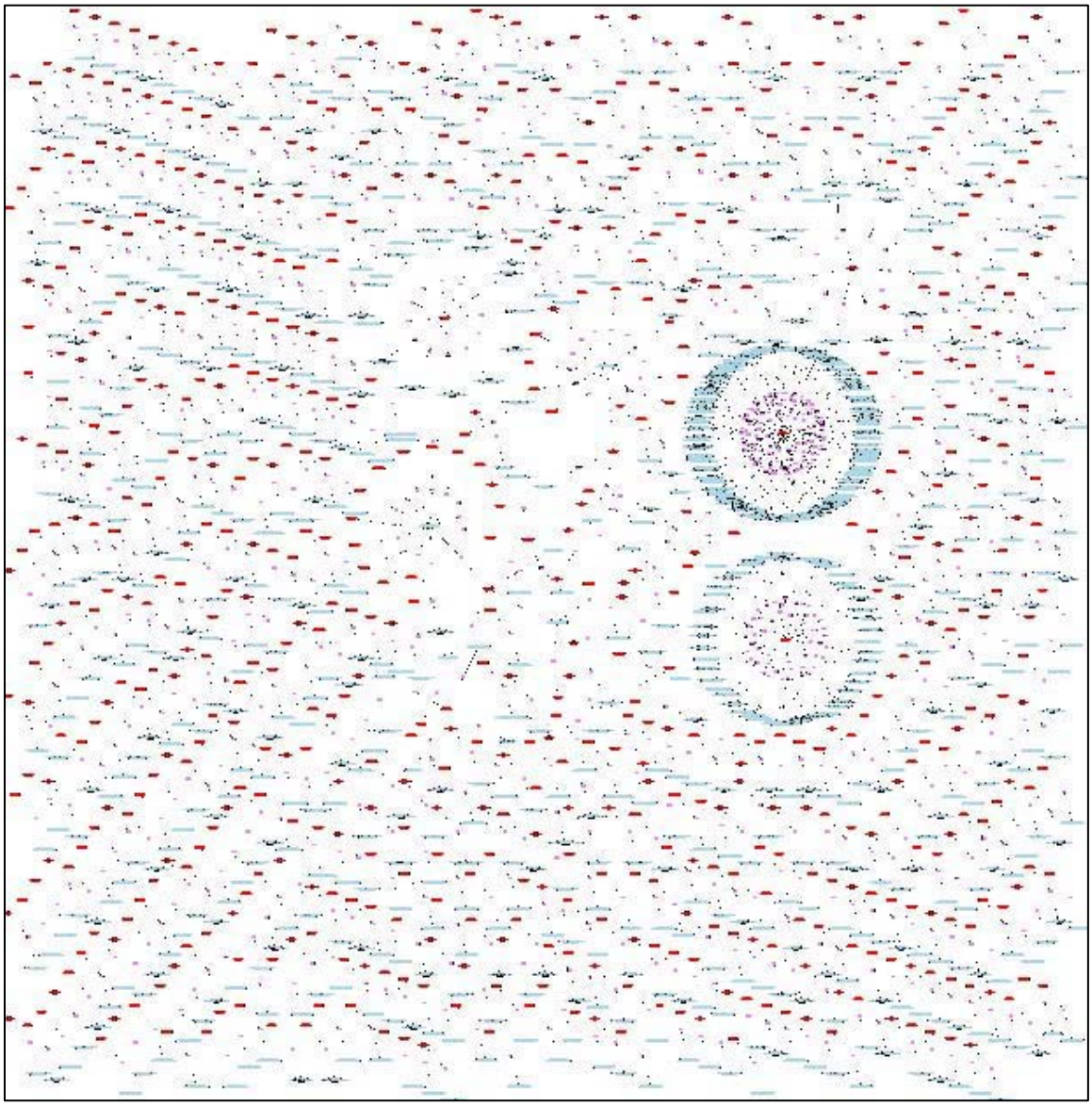


Figure 5. Reduced snapshot of large visualisation

4. CURRENT DRAWBACKS AND LIMITATIONS

Currently the generation of bitmap Graphviz dot graphs is a processor intensive task. The author used a 64 bit Debian based Dual Core – Dual Opteron 2.8 GHz machine with 6GB of memory with a U320 SCSI hard disk subsystem during the development of graphs for this research. To generate even the relatively simple sample graphs contained in Figures 1 to 4, considerable runtime at 100% of CPU was experienced for several minutes on the experimental computer.

This current use of computing resource has significant impact on the ability to use this type of technology currently for generation of near real time graphical data from a honeypot, IDS, firewall or similar traditional network countermeasure. The amount of data processed in the initial trials was comparatively small compared to the level of incident that would need to be rendered on a larger network pipe or dataset.

These particular problems of computational shortfall maybe overcome by utilization of current GPU technology to render and also compute the graphics (Ellson and Gansner 2008)**.** Similar use of GPUs has been recently successfully exploited for rapid visualization of graphical data where combining 4 high end GPU cards has produced graphical computational performance equivalent to 320 2.4Ghz CPU cores (ASTRA 2008). The application of GPU technology to certain mathematical functions and graphical rendering tasks delivers teraflops of performance at a relatively low cost. This type of power yield is dependent on the types of mathematical functions involved in the processing, however it would appear that Graphviz and its utilities may significantly benefit from the utilization of this technology.

The incorporation of a SQL system to store and process data extracted from honeypot data streams and instead of reprocessing log files will be a significant leap forward in analysis turnaround. Interfacing with the Surfnet IDS system that incorporates nepenthes data into an existing SQL structure is an easy candidate system for supporting SQL capabilities with a Graphviz/afterglow engine for generation of graphics. This extension of Surfnet would enable not only near real time generation of graphical data but also provide a method for longitudinal analysis of patterns and trends relating to honeypot activity. As mentioned previously this type of system reduces some of the temporal and spatial issues that are found in conventional textual query and analysis engines. The system will also allow replaying of items of interest by extracting it from the database. Temporary storage within a database structure would also allow for some buffering and load balancing to occur, by smoothing out spikes in the transmissions received.

5. CONCLUSION

This is research in progress, however, there are clear benefits for the analyst in using graphical approaches to perform analysis of honeypot data. Particularly as virus scanners, firewalls and intrusion detection systems are starting to fail as a response and countermeasure and honeypots are now trapping malcodes that bypass these with relative impunity.

Currently there is a bottleneck with regards to fast processing of data that would be arriving on a network interface in a honeypot system. The use of modern GPU technology and high-speed storage technologies may see many of these impediments removed. Also there are methods of the vector based rendering of the images which maybe better suited to interpretation of live data. These will be explored in the next phase of this research.

The ability to use graphical representations to view data from a variety of viewpoints and lenses for near real-time forensics of network incidents is a compelling reason for furthering this type of research and investigation. The graphical methods utilized in this exploratory work demonstrate an ability to represent magnitude and intensity of attack not afforded to text based analysis of data. By using graphical techniques such as these it may yield significant benefit in detecting and ameliorating threats

that are now posed by network borne malcodes which are increasingly becoming a problem in the network security arena.

6. REFERENCES

ASTRA. (2008). "Belgian researchers develop desktop supercomputer.", from http://fastra.ua.ac.be/en/index.html.

Baecher, P., M. Koetter, et al. (2009). "nepenthes." from http://nepenthes.carnivore.it/.

Ellson, J. and E. Gansner (2008). Graphviz - Graph Visualization Software, AT&T.

Marty, R. (2007). "Afterglow." from http://afterglow.sourceforge.net.

Microsoft. (2003). "Microsoft Security Bulletin MS03-026 - Buffer Overrun In RPC Interface Could Allow Code Execution (823980)." Retrieved 6th Feb, 2006, from http://www.microsoft.com/technet/security/bulletin/MS03-026.mspx.

Graduate Accounting Students' Perception of IT Forensics: A Multi-Dimensional Analysis

Grover S. Kearns, Ph.D., CPA, CFE
College of Business
140 7th Avenue South
University of South Florida St. Petersburg
St. Petersburg, FL 33701
Phone: 727-553-4085
Email: gkearns@stpt.usf.edu

ABSTRACT

Forensics and information technology (IT) have become increasingly important to accountants and auditors. Undergraduate accounting students are introduced to general IT topics but discussion of forensic knowledge is limited. A few schools have introduced an undergraduate major in forensic accounting. Some graduate schools offer accounting students an emphasis in forensic or fraud accounting that includes instruction in forensics and information technology. When students do not view the IT topics as being equally important to their careers as traditional accounting topics, these attitudes may reduce the quality of the course. In an effort to assess student attitudes, a survey of 46 graduate accounting students was conducted to measure two dimensions − knowledge and skills and interest and enjoyment − along nine common topics found in a forensics IT course. The association of the two dimensions was then measured. Also, the relationship between IT attitudes and the nine topics was measured along both dimensions. Fifteen hypotheses are presented and tested. Results are discussed to posit what instructors can do in order to increase the quality of the class and the positive perception of IT for accounting students.

Keywords: Forensics, Accounting education, Information technology forensics, IT auditing.

1. INTRODUCTION

Forensics is playing a more prominent role in various disciplines and is becoming increasingly important to the field of accounting. Accountants, and to a greater extent auditors, are reliant on computer-based analytical software skills and an understanding of information technology (IT) for assurance services, assessment of corporate efficiencies, assessment of internal controls and the investigation of possible fraud. The AICPA and PCAOB have issued statements regarding auditors' increased responsibility for IT knowledge, fraud recognition and the importance of evaluating IT controls during a financial audit. The Sarbanes-Oxley Act of 2002, increased evidence of business fraud and advancements in IT networks and systems should lead organizations to a higher expectation of auditors' IT skills (Sumners and Soileau, 2008).

Because these emerging requirements are often the focus of various fields including information technology, law enforcement and criminal justice, it is imperative that accountants be prepared to communicate knowledgeably with others. As the demand for accountants with forensic IT knowledge increases, academic institutions must provide the basic knowledge and skills about forensic IT topics.

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

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

Possible reasons are the lack of qualified accounting instructors with a forensic IT background and an underlying feeling that forensic IT issues are less important for accounting student's traditional career paths. Thus, many accounting students are now graduating without the requisite forensic IT foundation knowledge. Another problem is that some students are less likely to perceive forensic IT knowledge and skills as important to their careers. Without proper motivation, the students may not learn and retain sufficient forensic IT knowledge to provide the future capabilities that they will need. Accounting programs may not reflect recent and significant changes in the business environment. As a result, students are not equipped with the knowledge and skills they will actually need in practice (Gabbin 2002).

The purpose of this study is to determine how graduate accounting students perceive the forensic IT topics and to examine the relationship between their level of *interest and enjoyment* in IT topics and their level of *knowledge and skills* with these topics.

2. IMPORTANCE OF FORENSIC IT TO ACCOUNTANTS

2.1 IMPORTANCE OF FORENSIC IT KNOWLEDGE AND SKILLS

The Wells report showed that occupational, or business, fraud was increasing and usually involved asset misappropriation, corruption, and fraudulent financial statements (Hunton et al., 2004). Well known companies such as Marriott, Choicepoint, and Bank of America have suffered the loss of confidential customer information and data breaches (Websense, 2006). Because occupational fraud and computer related crimes are expected to grow, many universities are offering courses targeted to forensic IT (Busing et al., 2005/2006).

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

Information technology specialization has been identified as one of the most difficult areas to staff for internal auditors (Sumners and Soileau, 2008). An analysis of 595 job listings for IT auditors found that a large percentage specifically mentioned technical skills/abilities in IT controls (31%), networking (19%), security (18%), operating systems (23%), and CAATs (15%) (Merhout and Buchman, 2007). Additionally, 85 percent mentioned advanced certifications. Another survey revealed that internal auditors need to enhance their knowledge and skills of computerized information systems and increase understanding of systems development and acquisition activities (Abu-Musa, 2008).

Lack of forensic IT knowledge and skills can impair an auditor's ability. Continuous auditing (CA) is an increasingly important audit tool that corrects many of the deficiencies of other IT audit approaches. CA, however, often relies upon an underlying knowledge of IT and related technologies such as computer-aided auditing systems and Generalized Audit Software. Auditors who lack forensic IT knowledge will experience greater difficulty in integrating computer-aided auditing systems with their professional audit knowledge. This will impair the auditors' ability to independently and continuously perform tests in the CA environment (Li, Huang and Lin, 2007). Accountants must also understand forensic protocol and how to handle digital information that might be used in legal proceedings. Information assurance and proper authentication is vital in order that auditors protect the integrity of digital evidence (Duerr et al., 2004).

Studies have also affirmed that deficient IT controls weaken the organization's overall internal control structure and its ability to protect information assets (Davis et al., 2007). Because the role of IT in financial reporting systems is escalating, it is important that auditors be able to identify IT problems that affect financial reporting, evaluate the extent and nature of the problems and be familiar with steps to correct these weaknesses (Grant et al., 2008). IT control deficiencies lead to accounting and financial reporting errors (Alaali, Grant, and Miller, 2008). Companies report IT security and end-user computing controls as the major IT control problems. These include deficiencies in "segregation of IT duties, and IT policies, procedures, and documentation" (Alaali, Grant and Miller, 2008).

Many internal auditors regard technology issues as the exclusive domain of IT auditors. This perspective has created a knowledge gap between internal auditors' knowledge of a process under review and the systems that support that process. To overcome this knowledge gap, companies need to develop integrated or hybrid auditors who understand both IT general controls and application controls (Cascarino, 2007). This will require educating auditors about the IT environment, including organizational and administrative activities, infrastructure and environmental controls over how systems are linked, and physical security over IT assets and physical and logical access (Chaney and Kim, 2007). Accountants also need to understand the importance of advanced certifications. Students who are primarily interested in traditional accounting careers may wish to pursue a CPA (Certified Public Accountant), CIA (Certified Internal Auditor) or CMA (Certified Management Auditor) while those who enjoy IT topics may consider an IT Auditor certification such as CISA (Certified Information Systems Auditor). Such certifications have been shown to advance IT auditors in their careers more quickly (Wier, Hunton et al., 2000).

2.2 IMPORTANCE OF INTEREST AND ENJOYMENT OF IT TO LEARNING

A requisite for success may be the enjoyment that accounting students experience with forensic IT topics. Hunton et al. (2004) posit that enjoyment of computers and technology is important to success for IT auditors. They go on to explain that IT audit engagement success is predicated on the ability to work with people in various disciplines and effectively communicate their ideas and conclusions to several levels of management. Without the foundation knowledge and skills, accountants and auditors cannot interact or communicate effectively with other informed personnel. Developing a positive attitude towards IT topics is important for educators of IT auditors (Cangemi, 2000). Merhout and Buchman (2007) state that the education of IT auditors requires a blending of skills and educators "should strive to cultivate such a positive attitude in their students, and they should also make their students aware of the potential opportunities in the challenging IT audit career path."

2.3 FORENSIC INFORMATION SYSTEMS COURSE OBJECTIVES

Important forensic IT knowledge and skills have been identified in literature. These include the general IT knowledge including the assessment, implementation, operation and control of computer resources (Hall and Singleton, 2005) and an understanding of computer-based analytical software (Hunton, Bryant and Bagranoff, 2004).

A major shift in emphasis over the past ten years has been to devote more attention to both IT threats

and controls and the use of CAATs for forensic investigation. This is most likely the product of increased attention by the profession on fraud and the growing importance of IT controls on the reliability of financial data. Course objectives should include fraud and forensic techniques used by accountants to detect anomalies in the organization's financial data. CAAT software, such as ACL, represents an important analytical tool that should be part of the students' skills set. The textbooks cited typically devoted separate chapters to general IT subjects, IT security, IT controls, IT governance, and forensic and fraud auditing subjects (Whitman and Mattord, 2009).

At the undergraduate level, most programs offer a class in Accounting Information Systems that provides an introduction to information technology for accountants. It is not sufficiently broad, however, to cover forensic techniques. A second AIS class may be offered as an elective in which students are exposed to some foundation forensic knowledge and skills such as the use of computer assisted tools and techniques software (CAATs). Some graduate accounting programs offer an emphasis in forensic accounting. This normally consists of three or four courses that cover forensic and fraud accounting subjects. Few programs, however, offer more than one class in which course objectives include forensic based IT instruction.

An online investigation of both graduate and undergraduate classes that provided forensic IT course instruction revealed a common set of course objectives and textbooks normally devoted at least one chapter to each (Nelson et al., 2008; Cascarino, 2007). All of the Colleges of Business were AACSB accredited. The topics may be separated into three categories. First, the ones that represent traditional accounting topics with which students have already been introduced. Two of these are:

- Financial Auditing
- Financial and Accounting Analytical Methods

Other topics represented those that are fairly new to the students and with which they have the least amount of familiarity. Three of these are:

- Information Technology
- IT Security
- Forensic and Fraud Auditing Techniques

Another set of topics represent a blend and extend the students' accounting knowledge with IT knowledge. Four of these are:

- IT Auditing
- Computer Based Analytical Methods
- IT Governance Frameworks
- Internal Control

Students were expected to possess advanced knowledge in the two strictly accounting topics. The three strictly IT topics represented areas of knowledge the student may have acquired but for which no expectation existed. The four blended areas represented topics in which the student was expected to

have some knowledge that was later expanded with instruction in IT. For example, all accounting students are introduced to the concepts of internal control at the undergraduate level. However, they lack detailed knowledge of IT controls.

3. HYPOTHESES AND METHODOLOGY

3.1 HYPOTHESES

Fifteen hypotheses are presented. First, it was theorized that accounting students would perceive their level of *knowledge and skills* to be lower in both accounting and forensic IT topics than their *interest and enjoyment*. This is important because the students' objective is the pursuit of knowledge and skills in these areas. It is assumed that an adequate level of interest and enjoyment must exist to motivate students to pursue advanced studies. Thus,

> **H1:** Graduate accounting students perceived level of *knowledge and skills* in accounting and forensic IT topics will be lower than their level of *interest and enjoyment* in these same topics.

Next, it was expected that students who possessed higher *knowledge and skills* might also be expected to display greater *interest and enjoyment* of forensic IT topics. Similarly, students who were interested in and enjoyed forensic IT topics were more likely to increase their knowledge and skills more highly. Other forces may motivate students to acquire knowledge and skills. In particular, those students who are highly motivated to make good grades and those who see mastery of forensic IT topics as essential to success will acquire the skills and knowledge even though their level of interest and enjoyment is not high. However, as one becomes more proficient in a subject, that subject becomes more interesting and relevant. Thus,

> **H2:** Graduate accounting students who have a higher perceived level of *knowledge and skills* in accounting and forensic IT topics will rate their level of *interest and enjoyment* in these topics more highly.

Although H2 implies a causal relationship, it cannot be determined whether higher values for *knowledge and skills* leads to higher *interest and enjoyment* or vice-versa. For this reason, a separate hypothesis is not provided. It is possible that both work at the same time: as students' knowledge and skills increase, the interest and enjoyment increases which results in greater dedication and an ensuing increase in knowledge and skills. This circular relationship is depicted in Figure 1.

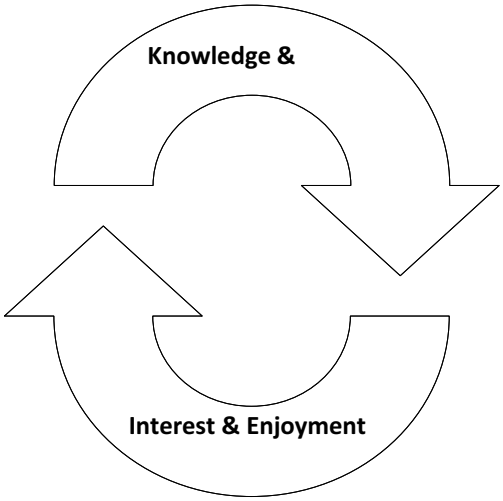


Figure 1: Circular Relationship between Knowledge & Skills and Interest & Enjoyment

Students who are highly dedicated to accounting or are highly focused on traditional accounting subjects might not wish to make a major commitment to another discipline such as IT and forensic analysis (Dunn and Grabski, 1998; Ravel, 1991). Therefore, it was hypothesized that students who scored highly on the two accounting topics would show lower interest in strictly forensic IT topics. Thus,

> **H3**: Graduate accounting students who have a higher perceived level of *knowledge and skills* in strictly accounting topics will display a higher perceived level of *interest and enjoyment* in strictly accounting topics.

Conversely, students who scored lower on commitment to strictly accounting topics might be interested in pursuing a related discipline and would then show a higher interest in strictly forensic IT topics. Students may find the forensic IT topics to be more interesting than traditional accounting topics or may find that they excel in these topics as compared to accounting and that forensic IT offers a more attractive alternative to the more traditional accounting topics. Thus,

> **H4:** Graduate accounting students who have a higher perceived level of *knowledge and skills* in strictly forensic IT topics will display a higher perceived level of *interest and enjoyment* in strictly forensic IT topics.

Similarly, for the same reasons as above, students who scored higher on commitment to a blend of accounting and forensic IT topics should be interested in pursuing both the accounting and IT disciplines and would be expected to show a higher interest in pursuing either advanced certification. Thus,

> **H5:** Graduate accounting students who have a higher perceived level of *knowledge and skills* in a blend of accounting and forensic IT topics will

display a higher perceived level of *interest and enjoyment* in both accounting and forensic IT topics.

Students whose attitude towards the forensic IT class is highly positive can be expected to rate both their *knowledge and skills* and their *interest and enjoyment* with forensic IT topics more highly. Knowledge and skills will shape positive attitudes because the objectives will appear more easily attainable. Interest and enjoyment may be a product of positive attitudes or actually create those attitudes. However, from H1, it is assumed that students will rate their K&S lower than their I&E particularly for forensic IT topics for which they may be expected to possess a low level of familiarity. Because students' attitudes are shaped by viewing the class as means towards obtaining an emphasis in Fraud Accounting, a positive *attitude* for the forensic IT topics should be positively associated with their *interest and enjoyment* for these topics even where their knowledge and skills are low. Thus,

H6: Graduate accounting students' Forensic IT Attitude will display a higher association with the perceived level of *interest and enjoyment* in strictly forensic IT topics than with the perceived level of *knowledge and skills* in strictly forensic IT topics.

Similarly, and for the same reasons, higher *attitudes* for forensic IT topics could be expected to be more highly associated with *interest and enjoyment* in either strictly accounting topics or a blend of accounting and forensic IT topics than for their perceived level of *knowledge and skills*. Although some topics have been identified as strictly accounting, students are likely to perceive their attainment of knowledge as lower then their interest. The reverse notion would not make sense. If an acceptable level of knowledge and skills had already been attained then interest in pursuing greater knowledge would be lower and students would not be motivated to pursue further education. Thus,

H7: Graduate accounting students' Forensic IT Attitude will display a higher association with the perceived level of *interest and enjoyment* in the accounting and blend topics than with the perceived level of *knowledge and skills* in accounting and blend topics.

The next set of hypotheses concern students' intent to pursue advanced certifications, either as a traditional CPA or as an IT Auditor.

First, because of the increased benefits of a CPA certification, it was expected that many of the graduate accounting students would intend to pursue a CPA. These students would most likely possess a lower attitude towards forensic IT topics and perceive the class as a means to acquiring knowledge that would be useful on the CPA exam. Because pursuit of a CPA requires a great investment of time, these students would be less likely to value the forensic IT topics as important to their career. For these reasons, an inverse relationship between intent to pursue a CPA and the level of forensic IT attitude could be anticipated. Thus,

H8: Graduate accounting students' Forensic IT Attitude will be negatively associated with intent to pursue a CPA certification.

Conversely, some students might believe that pursuit of a CPA required too high an investment and that the CPA might be devalued because of its popularity. Other students might be bored with strictly accounting topics and excited about the possibility of learning a new discipline that would make them

more unique and marketable. Particularly bright students might pursue both certifications. An increased attitude towards these topics would then be associated with the intent to pursue an IT Auditor certification. Thus,

H9: Graduate accounting students' Forensic IT Attitude will be positively associated with the intent to pursue an IT Auditor certification.

Students who scored highly on the two accounting topics along both dimensions would show lower interest in strictly forensic IT topics and a higher interest in pursuing the CPA certification. Accountants with CPAs have been shown to advance more rapidly and earn a higher salary but studying for the exam requires a commitment of time that might preclude pursuit of other knowledge. Thus,

H10: Graduate accounting students who have a higher perceived level of *knowledge and skills* in strictly accounting topics will display a higher interest in pursuing a CPA certification.

H 11: Graduate accounting students who have a higher perceived level of *interest and enjoyment* in strictly accounting topics will display a higher interest in pursuing a CPA certification.

Conversely, students who scored lower on commitment to strictly accounting topics might be interested in pursuing a related discipline and would then show a higher interest in strictly forensic IT topics and a higher interest in pursuing an IT Auditor certification. It might also be possible that students perceive the CPA certification as too difficult to pursue whereas a certification as an IT Auditor might be more easily attained. Studies have also shown that both internal and financial auditors who possess a certification as IT Auditor have higher average salaries and advance more quickly. Thus,

H12: Graduate accounting students who have a higher perceived level of *knowledge and skills* in strictly forensic IT topics will display a higher interest in pursuing an IT Auditor certification.

H13: Graduate accounting students who have a higher perceived level of *interest and enjoyment* in strictly forensic IT topics will display a higher interest in pursuing an IT Auditor certification.

H14: Graduate accounting students who have a higher perceived level of *knowledge and skills* in a blend of accounting and forensic IT topics will display a higher interest in pursuing an IT Auditor certification.

H15: Graduate accounting students who have a higher perceived level of *interest and enjoyment* in a blend of accounting and forensic IT topics will display a higher interest in pursuing an IT Auditor certification.

3.2 DATA COLLECTION

An anonymous survey of 46 students was used to evaluate graduate accounting students' attitude towards specific forensic IT topics and to determine if there was a relationship between their perceived level of *knowledge and skills* and their perceived level of *interest and enjoyment* of these topics. Other questions were asked to determine (1) if their attitude towards forensic IT in general affected the perceived levels of knowledge and skills and interest and enjoyment and (2) if their attitude towards forensic IT in general affected their commitment to pursue advanced certifications as a CPA or IT Auditor or both.

3.2.1 The Survey Group: Data was provided by an in-class survey, administered by the professor, to graduate students enrolled in a Masters of Business Administration program at a university in a metropolitan area in the southeastern United States. The university attracts many commuter students who are already placed in accounting careers and are completing their education. The school is accredited by the Association to Advance Collegiate Schools of Business (AACSB) which is the highest level of accreditation that can be achieved by a college of business. The class is one of four classes offered in a Fraud Accounting emphasis. The course objectives include materials that are appropriate for graduate level students who have only limited background in forensic IT and investigative techniques. Data were collected from two sections of the same class over two semesters. The classes were of approximately equal size. All of the students had undergraduate majors in accounting. The professor explained to the students that the responses were totally anonymous and could not affect their grade in any way. Because no written responses were elicited, it was not possible for the professor to discover the identity of a particular respondent. All surveys were given to an administrative assistant for data entry and the results were checked by another assistant.

Because exact age information is not available, the average age can only be approximated. Most of the MBA students are between 25 and 35 but at least 20 percent were over 35. There was an equal distribution of male and female and nearly all commuted within a radius of 25 miles. Of the 46 students, 37 worked at least 20 hours each week and 27 worked full-time.

3.2.2 The Research Instrument: The complete research instrument is shown in Appendix A. Students ranked their levels on a seven-point Likert-type scale for nine different topics representing broad topic areas. Two topics referred to strictly accounting topics: Auditing and Financial & Accounting Analytical Methods. Three topics were strictly related to forensic IT topics: Information Technology, IT Security, and Forensic and Fraud Accounting. Four topics were a blend of both: IT Auditing, IT Governance Frameworks, Computer Based Analytical Methods (CAATs such as ACL), and Internal Control. These topics were interspersed on the survey in order that the student did not unconsciously provide answers that reflected the separate groupings.

Seven topics were IT related and reflected knowledge and skills that were the principal course objectives determined annually during a needs assessment. The two others represented traditional accounting topics used for comparative purposes. It was thought that the students' knowledge and skills in these two topics would be higher and that they would also display a higher level of interest and enjoyment because they were important parts of the accounting discipline.

The blended topics represented areas that combined existing accounting knowledge and new IT knowledge. For example, the MBA students were already familiar with auditing concepts so that IT auditing was an extension of existing knowledge. Blended knowledge is normally easier to attain as it builds upon a foundation with which the student is already comfortable. Thus, it was expected that these four topics would be viewed differently by the student.

Because the survey was administered at the end of the semester, students were expected to understand the knowledge represented by each of the topics. During the course, these had been referred to frequently. Forensic and fraud auditing was the emphasis for the program and represented more sophisticated knowledge than the traditional accounting topics. IT governance represented frameworks such as COSO (Committee of Sponsoring Organizations) and CobIT (Control Objectives for IT) which were covered in the class.

The research instrument was further reviewed by two faculty members familiar with the course objectives. Comments and suggestions were used to craft the final instrument.

4. ANALYSIS OF DATA AND RESULTS

4.1 CORRELATION ANALYSIS

Pearson product-moment correlation analysis was performed on the 46 observations of students in the graduate class over two separate semesters. The 24 questions measured four separate phenomena: Knowledge and Skills; Interest and Enjoyment; Forensic IT Attitudes; and, Intent to Pursue Advanced Certifications. Knowledge and Skills and Interest and Enjoyment were further investigated by separating the nine topics into three components of student interest: Accounting (measured by two questions), Forensic IT (measured by three questions), and Blend (measured by four questions). The survey instrument is shown in the appendix.

4.2 METRICS AND THE KNOWLEDGE INDEX

Metrics for the two dimensions of *knowledge and skills* (K&S) and *interest and enjoyment* (I&E) are shown in Table 1. A Knowledge Index has been calculated by dividing the K&S mean by the I&E mean. This can be interpreted as follows. If students possess equal amounts of *knowledge and skills* and *interest and enjoyment*, then the index will be 1.0. If their *interest and enjoyment* exceeds their *knowledge and skills* then the index will be less than 1.0 while the index will exceed 1.0 when their *knowledge and skills* exceeds their *interest and enjoyment*. In most cases, we would expect that I&E would exceed K&S and the index would be less than 1.0. Except for IT Governance Frameworks, this is the case. The greatest disparity is for Forensic and Fraud Accounting. The I&E mean is high (5.65) and the index is .69. Large disparities will also affect correlation coefficients because high levels of perceived I&E could be associated with both low and high perceived levels of K&S. The index provides a measure of the relationship between I&E and K&S but not the levels of each which is provided by the means.

To test the null hypothesis that the sample means for knowledge and skills were not statistically different than the sample means for interest and enjoyment, paired two-tailed t-tests were calculated using a 95% confidence level. A low p-value for this test (less than 0.05) means that there is evidence that the difference in the two means are statistically significant (Hair et al., 1998). Results, in Table 1, show that the p-values for four of the topics to be less than 0.05. In all cases, there were large differences between the I&E and K&S means. None of these four topics represent topics that are strictly accounting related. Thus, for IT related topics, there is further evidence that students' perceive *interest and enjoyment* as different from their *knowledge and skills*.

Table 1: Metrics for Two Dimensions : Students' Perceived Levels of Knowledge & Skills and Interest & Enjoyment (Based on 46 responses for each item.)

Metric \ Topic	Financial Auditing	IT Auditing	Information Technology	IT Security	IT Governance Frameworks	Financial & Accounting Analysis	Computer Based Analytical Methods	Internal Control	Forensic and Fraud Auditing
K & S Mean	4.35	2.52	3.22	3.04	3.13	3.43	3.08	4.43	3.91
Std Dev	1.62	1.03	1.43	1.41	1.41	1.63	1.36	1.57	1.43
I & E Mean	4.57	3.39	3.43	3.87	2.96	4.82	4.17	4.70	5.65
Std Dev	1.46	1.73	1.81	1.67	1.56	1.69	2.04	1.47	1.32
Knowledge Index	0.95	0.74	0.94	0.79	1.06	0.71	0.74	0.94	0.69
t-Test	0.16	0.00	0.47	0.01	0.41	0.23	0.00	0.24	0.00

Mean is based on a scale of 1 to 7 where 1 represents the lowest level and 7 represents highest level.
Knowledge index is calculated as K&S Mean / I&E Mean.
t-Test is two-tailed paired samples with 95% confidence level.

Because eight of the nine knowledge indexes were less than 1.0, H1 was supported and we can state that the perceived level of *knowledge and skills* in accounting and forensic IT topics will be lower than their level of *interest and enjoyment* in these same topics.

4.3 Association of Knowledge and Skills with Interest and Enjoyment

Results for correlation of students' *knowledge and skills* with *interest and enjoyment* are shown in Table 2. Coefficients of correlation, or correlates, along the diagonal represent the strength of relationship between the *knowledge and skills* and *the interest and enjoyment* for each of the nine topics (Hair et al., 1998). Of the 81 coefficients, 71 are positive. Statistical significance is indicated for each correlate in the table. On the diagonal, there is a strong relationship for financial auditing (r = .69) while IT security has the lowest relationship (r = .02). Six of the nine topics on the diagonal display a positive and medium to strong relationship which supports H2 and we can state that accounting students who have a higher perceived level of knowledge and skills in accounting and forensic IT topics will rate their level of interest and enjoyment in these topics more highly (and vice-versa by Figure 1). Surprisingly, correlates for forensic and fraud auditing were low although the mean value for I&E was high. This indicates that high interest in this topic is related to both high and low K&S levels for other topics. As expected, I&E for financial auditing relates well with the K&S for financial auditing (r = .69), IT auditing (r = .51), IT governance frameworks (r = .33), and internal control (r = .51). [1]

[1] The correlation coefficient, *r*, varies in magnitude according to the strength of the relationship between the data observations. In social research, ranges of 0 to .3 are viewed as weak, .3 to .5 as moderate and above .5 as strong. Statistical significance measures the probability that the relationship is by random chance. Values greater than .10 are generally not accepted as being significant. In this study, the cutoff was .05 indicating that the likelihood of random chance was five percent or lower.

Table 2: Correlation of Students' Knowledge & Skills with Interest & Enjoyment

Level of Interest and Enjoyment	Level of Knowledge and Skills: ACCTG – Financial Auditing	BLEND – IT Auditing	FOR IT – Information Technology	FOR IT – IT Security	BLEND – IT Governance Frameworks	ACCTG – Financial & Accounting Analysis	BLEND – Computer Based Analytical Methods	BLEND – Internal Control	FOR IT – Forensic and Fraud Auditing
Financial Auditing	0.69 ***	0.51 ***	0.13	0.14	0.33 *	0.23	0.20	0.51 ***	0.13
IT Auditing	0.35 **	0.38 **	0.38 **	0.08	0.47 **	0.21	0.36 **	0.39 **	0.05
Information Technology	0.07	0.31 *	0.53 ***	0.24	0.19	0.06	0.42 **	0.21	0.08
IT Security	-0.06	0.17	0.25	0.02	0.33 *	0.12	0.18	0.06	0.07
IT Governance Frameworks	0.11	0.43 **	0.20	0.06	0.29 *	0.13	0.38 **	0.30 *	-0.18
Financial & Accounting Analysis	0.19	-0.20	-0.10	0.04	-0.15	0.13	0.12	0.32 *	-0.27
Computer Based Analytical Methods	0.13	0.04	0.11	0.34 *	-0.13	0.37 **	0.41 **	0.27	-0.01
Internal Control	0.31 *	0.14	0.07	0.07	0.28	0.37 **	0.26	0.52 ***	-0.08
Forensic & Fraud Auditing	0.00	-0.19	0.21	0.03	0.05	0.07	0.07	0.07	0.05

*, **, *** represent significance levels of .05, .01, and .001respectively

4.4 Students with Preference for Accounting Topics

The next set of hypotheses compared students' *knowledge and skills* and *interest and enjoyment* with preferences for accounting, forensic IT, and blended topics. The correlates for these relationships are shown in Table 3. Two topics represented knowledge that students were expected to possess prior to taking the class. These were strictly accounting topics. One was Financial Auditing and covered the audit knowledge that all undergraduate accounting majors are expected to master in a minimum of two

classes. The second was Financial & Accounting Analytical Methods that represented the quantitative and analytical skills that accountants might be expected to possess based on a minimum of four classes in Financial, Managerial and Intermediate Accounting. *Knowledge and skills* and *interest and enjoyment* were moderately related for strictly accounting topics (r = .44). Thus, H3 is supported and we can say that accounting students who have a higher perceived level of knowledge and skills in strictly accounting topics will display a higher perceived level of interest and enjoyment in strictly accounting topics (and vice-versa).

4.5 Students with Preference for Forensic IT Topics

Three topics comprised the component for strictly forensics and IT topics: Information Technology, IT Security, Forensic and Fraud Auditing. From Table 3, there was a moderate relationship (r = .35) between *knowledge and skills* and *interest and enjoyment* for the forensic IT component. Thus, H4 is supported and we can say that accounting students who have a higher perceived level of knowledge and skills in strictly forensic IT topics will display a higher perceived level of interest and enjoyment in strictly forensic IT topics (and vice-versa).

Table 3: Correlation of Students' *Knowledge & Skills* and *Interest & Enjoyment* with Preferences for Accounting, Forensic IT, and Blended Topics

		Knowledge & Skills			Interest & Enjoyment		
		FOR IT	ACCTG	BLEND	FOR. IT	ACCTG	BLEND
Knowledge & Skills	FORENSIC IT	1.00					
	ACCOUNTING	0.31 *	1.00				
	BLEND	0.70 ***	0.65 ***	1.00			
Interest & Enjoyment	FORENSIC IT	0.35 *	0.05	0.37 *	1.00		
	ACCOUNTING	-0.06	0.44 **	0.26	0.26	1.00	
	BLEND	0.25	0.38 *	0.61 ***	0.68 ***	0.56 ***	1.00

*,**,*** represent significance levels of .05, .01, and .001respectively

4.6 STUDENTS WITH PREFERENCE FOR A BLEND OF TOPICS

Four topics represented the component that is a blend of accounting and forensic IT: IT Auditing, IT Governance Frameworks, Computer Based Analytical Methods (CAATs such as ACL), and Internal Control. There was a strong relationship (r = .61) between *knowledge and skills* and *interest and enjoyment* for the four topics that comprised an interest in a blend of accounting and forensic IT topics. Thus, H5 is supported and we can say that accounting students who have a higher perceived level of *knowledge and skills* in both accounting and forensic IT topics will display a higher perceived level of *interest and enjoyment* in a blend of accounting and forensic IT topics (and vice-versa).

4.7 Forensic IT Attitudes

These hypotheses test the relationship between IT attitudes, as determined by students' interest in the forensic IT class as a means towards career goals and their perception that forensic IT skills are important to accountants, with their level of *knowledge and skills* and *interest and enjoyment* in strictly

forensic IT topics. These attitudes were tested using six questions that were then correlated with the responses for the nine questions for the accounting, forensic IT and blend topics.[2] The three topics that represent strictly forensic IT are: Information Technology, IT Security, and Forensic and Fraud Auditing. The results for *knowledge and skills* are shown in Table 4 and for *interest and enjoyment* in Table 5. The strictly forensic IT topics are shaded in both tables and discussion is limited to these topics only.

For the *knowledge and skills* dimension, in Table 4, results are mixed. Only 5 of the 18 correlates representing the three forensic IT topics are significant and two of these are negative. Information technology is positively associated with the belief that the class has increased overall interest in IT and forensic techniques (r = .34) but negatively associated with the belief that the class will help performance on the CPA exam (r = -.36). IT security was positively associated with the belief that the class has increased overall knowledge (r = .29) and would help career performance (r = .30) and negatively associated with the belief that the class will help performance on the CPA exam (r = -.46). Internal control was strongly associated with the belief that the class has increased overall knowledge (r = .61) and associated with the belief that the class will help performance on the CPA exam (r = .37). These correlates could indicate that the students rated their *knowledge and skills* as low to medium regardless of their attitudes. Table 1 showed that the K&S means were lower than the I&E means in all but one case.

For the *interest and enjoyment* dimension, in Table 5, results are highly consistent and 14 of the 18 correlates are significant and all are positive. Information technology is positively associated with the belief that the class has increased overall interest in IT and forensic techniques (r = .28) associated with the belief that the class will help performance on the CPA exam (r = .26). It also shows strong associations with the attitude that the class has increased overall interest in IT and forensic auditing (r = .55), that IT and forensic knowledge and skills are important for accountants (r = .34), the class will help students perform well in their career (r = .55) and interest in a second forensic techniques class (r = .57).

Interest and enjoyment in IT security was positively associated with the belief that the class has increased overall knowledge (r = .23) and the belief that the class will help performance on the CPA exam (r = .38). IT security shows strong associations with the attitude that the class has increased overall interest in IT and forensic audit techniques (r = .51), that IT and forensic knowledge and skills are important for accountants (r = .29), the class will help students perform well in their career (r = .53) and interest in a second forensic techniques class (r = .65).

Interest and enjoyment in forensic and fraud auditing was positively associated with the belief that the class has increased overall knowledge (r = .39) but weakly with the belief that the class will help performance on the CPA exam (r = .25). Forensic and fraud auditing shows moderate to strong associations with the attitude that the class has increased overall interest in IT and forensic audit techniques (r = .36), the class will help students perform well in their career (r = .48) and interest in a second forensic techniques class (r = .41). It was not related to the notion that IT and forensic knowledge and skills are important for accountants (r = .11). The large number of positive and significant correlates supports H12 and we can state that students' forensic IT attitude will be

[2] Students ranked nine topics representing broad areas of knowledge. These were Financial Auditing, IT Auditing, Information Technology, IT Security, IT Governance Frameworks, Financial & Accounting Analytical Methods, Computer Based Analytical Methods, Internal Control, and Forensic & Fraud Accounting. Six questions were ranked to measure IT Attitude. These were: This class has increased my overall knowledge of IT forensics? This class has increased my overall interest in IT forensics? I believe that IT forensic skills are important for accountants? I believe this class will help me perform well in my career? I believe this class will help me perform well on the CPA exam? I would be interested in taking an IT Auditing class?

associated with *interest and enjoyment* of strictly forensic IT topics.

Because interest and enjoyment had a larger number of significant and positive associations with forensic IT attitudes, H6 is accepted and we can state that students' level of forensic IT attitudes will have a higher level of association with *interest and enjoyment* rather than *knowledge and skills* of forensic IT topics.

A similar approach is taken for the blend of accounting and forensic IT topics. Table 4 shows that 12 of the 24 correlates for the *knowledge and skills* dimension are significantly associated with forensic IT attitudes while Table 5 shows that 18 of the 24 correlates for the *interest and enjoyment* dimension exhibit significant associations. In Table 4, students with a high level of *interest and enjoyment* in IT auditing feel that the class has increased their overall knowledge (r = .35), has increased their overall interest (r = .55), will help them in their career (r = .42), and would be interested in taking a second class (r = .76). Those with a high level of interest and enjoyment in IT governance frameworks feel that the class has increased their overall knowledge (r = .34), has increased their overall interest (r = .58), topics are important for accountants (r = .35), will help them in their career (r = .66), will help them on the CPA exam (r = .52), and would be interested in taking a second class (r = .82). Interest and enjoyment is also highly related to attitudes for computer based analytical methods and internal control. Because there are more significant and stronger correlates for interest and enjoyment we conclude that H7 is supported and we can state students' level of forensic IT attitudes will have a higher level of association with *interest and enjoyment* rather than *knowledge and skills* of a blend of accounting and forensic IT topics.

Table 4: Correlation of Students' IT Attitude with Knowledge and Skills

	Level of Knowledge & Skills									
	ACCTG	BLEND	FOR IT	FOR IT	BLEND	ACCTG	BLEND	BLEND	FOR IT
IT / Forensics Attitude	Financial Auditing	IT Auditing	Information Technology	IT Security	IT Governance Frameworks	Financial & Accounting Analysis	Computer Based Analytical Methods	Internal Control	Forensic and Fraud Auditing
The class has increased my overall knowledge of IT and forensic audit techniques	0.48 ***	-0.03	0.19	0.29 *	0.32 *	0.56 ***	0.33 *	0.61 ***	0.19
The class has increased my overall interest in IT and forensic audit techniques	-0.04	0.34 *	0.34 *	0.22	0.41 *	0.28	0.33 *	0.14	-0.11
I believe that IT and forensic knowledge and skills are important for accountants	-0.22	0.07	-0.04	0.15	0.03	0.12	0.10	0.07	-0.06
I believe the class will help me perform well in my career	-0.19	0.26	0.17	0.30 *	0.05	0.22	0.39 **	0.14	0.09
I believe the class will help me perform well on the CPA exam	0.29 *	-0.32 *	-0.36 *	-0.46 **	-0.20	0.09	-0.23	0.37 *	-0.02
I would be interested in taking a second class in forensic techniques as an elective	0.14	0.31 *	0.24	0.08	0.47 **	0.23	0.34 *	0.28	-0.17

*, **, *** represent significance levels of .05, .01, and .001respectively

Table 5: Correlation of Students' IT Attitude with Interest & Enjoyment

IT / Forensics Attitude	Level of Interest & Enjoyment									
	ACCTG	BLEND	FOR IT	FOR IT	BLEND	ACCTG	BLEND	BLEND	FOR IT
	Financial Auditing	IT Auditing	Information Technology	IT Security	IT Governance Frameworks	Financial & Accounting Analysis	Computer Based Analytical Methods	Internal Control	Forensic and Fraud Auditing
The class has increased my overall knowledge of IT and forensic audit techniques	0.33 *	0.35 *	0.28 *	0.23	0.34 *	0.46 **	0.57 ***	0.56 ***	0.39 **
The class has increased my overall interest in IT and forensic audit techniques	-0.09	0.55 ***	0.55 ***	0.51 ***	0.58 ***	0.11	0.20	0.50 **	0.36 *
I believe that IT and forensic knowledge and skills are important for accountants	0.08	0.16	0.34 *	0.29 *	0.35 *	-0.07	0.39 **	0.24	0.11
I believe the class will help me perform well in my career	0.06	0.42 **	0.55 ***	0.53 ***	0.66 ***	0.39 **	0.57 ***	0.37 *	0.48 **
I believe the class will help me perform well on the CPA exam	0.19	0.07	0.26	0.38 *	0.52 ***	0.27	0.23	0.25	0.24
I would be interested in taking a second class in forensic techniques as an elective	0.19	0.76 ***	0.57 ***	0.65 ***	0.82 ***	0.43 **	0.35 *	0.56 ***	0.41 **

*, **, *** represent significance levels of .05, .01, and .001respectively

4.8 INTENT TO PURSUE ADVANCED CERTIFICATION

Students' forensic IT attitude was correlated with intent to pursue an advance certification and results are shown in the upper half of Table 6. It was expected that students who were primarily interested in pursuing a CPA would be less committed to the forensic IT topics because they represented an investment of time that students perceived as having limited benefit. Therefore, a negative association was anticipated. As expected, five of the six attitudes were negatively related with intent to pursue a

CPA and four of these were significant. There was a negative association with the belief that IT and forensic knowledge and skills are important to accountants (r = -.34). There is a positive but not significant association with the belief that the class would help them perform well on the CPA exam (r = .23). The negative correlates support H8 and we can state that the students' will display a negative association between their forensic IT attitudes and intent to pursue a CPA.

Conversely, students who value the forensic IT topics are expected to display a positive association with intent to pursue an IT auditor certification. There were three positive and significant correlates that supported this notion. These students feel that the class has increased their overall in IT and forensic audit techniques (r = .48), that this knowledge is important to accountants (r = .47), the class will help them perform well in their career (r = .22), and would be interested in taking a second course (r = .47). Because only three of the six correlates are significant, H9 is only partially supported and we can state there is some evidence that students' will display a postive association between their forensic IT attitudes and intent to pursue an IT auditor certification.

The three components were also correlated with students' intent to pursue an advanced certification along each dimension. Results are shown in the lower half of Table 6. None of the components for either dimension was significantly related to intent to pursue a CPA and four of the six correlates were negative. Therefore, both H10 and H11 are rejected and we can state that students' will not display any significant relationship between either *knowledge and skills* or *interest and enjoyment* with intent to pursue a CPA.

For strictly forensic IT topics, there is no significant relationship for the knowledge and skills dimension with intent to pursue an IT auditor certification (r = -.14) but a strong and significant association for the interest and enjoyment dimension (r = .53). Therefore, H13, but not H12, is supported and we can state that students' who display higher *interest and enjoyment* in forensic IT topics will be more likely to pursue an IT auditor certification but those who display higher *knowledge and skills* in forensic IT topics will not be more likely to pursue an IT auditor certification.

For a blend of accounting and forensic IT topics, there is no significant relationship for the knowledge and skills dimension with intent to pursue an IT auditor certification (r = .00) but a strong and significant association for the interest and enjoyment dimension (r = .42). Therefore, H15, but not H14, is supported and we can state that students' who display higher *interest and enjoyment* in a blend of accounting and forensic IT topics will be more likely to pursue an IT auditor certification but those who display higher *knowledge and skills* in a blend of accounting and forensic IT topics will not be more likely to pursue an IT auditor certification.

Table 6: Correlation of Students' Attitude with Interest in Pursuing Advanced Certification

IT / Forensics Attitude		Intent to Pursue CPA	Intent to Pursue IT Auditor
The class has increased my overall knowledge of IT and forensic audit techniques		-0.37 **	0.19
The class has increased my overall interest in IT and forensic audit techniques		-0.39 **	0.48 **
I believe that IT and forensic knowledge and skills are important for accountants		-0.34 *	0.47 **
I believe the class will help me perform well in my career		-0.19	0.22
I believe the class will help me perform well on the CPA exam		0.23	-0.06
I would be interested in taking a second class in forensic IT techniques as an elective		-0.29	0.47 **
Knowledge & Skills	IT	-0.16	-0.14
	ACCTG	-0.07	-0.07
	BLEND	-0.27	0.00
Interest & Enjoyment	IT	0.08	0.53 ***
	ACCTG	0.01	-0.03
	BLEND	-0.25	0.42 **

*,**,*** represent significance levels of .05, .01, and .001respectively

5. DISCUSSION

This study has three primary contributions. First, it provides information that will assist instructors of graduate accounting students in a forensic techniques class to improve students' attainment of forensic IT knowledge of skills. Second, it provides important information about the relationship between two dimensions of learning: *knowledge and skills* and *interest and enjoyment*. Third, it provides a tested instrument for use by researchers and practitioners.

Instructors of forensic IT techniques need to know what motivates their students to learn and how they perceive the topics that typically support course objectives. Student responses can be expected to vary along different dimensions including their existing level of IT knowledge and skills when they enter the class. Some students who are highly interested in the traditional accounting topics and are intent on pursuing a CPA may view the class as a required subject that provides important knowledge and skills but one that does not provide interest and enjoyment or they may view it as less important to their

career objectives than other classes in their major. In any case, information that assists the instructor to improve the students' overall performance is important. Unfortunately, there has been scant attention to learning motivators at this level.

Nine topics, representing forensic IT and accounting topics, were used to determine if there existed a relationship between perceived IT *knowledge and skills* and *interest and enjoyment*. Fifteen hypotheses explored the relationships between the two dimensions and attempted to determine if particular topics – accounting, forensic IT, a blend of accounting and forensic IT – were more likely to appeal to students and be associated with the intent to pursue advanced certifications. Students' attitudes towards the class were measured to determine the association with these topics and along the two dimensions. Ten of the hypotheses (H1, H2, H3, H4, H5, H6, H7, H8, H10, H11 and H12) were supported, one was partially supported (H9), and four were rejected (H10, H11, H12, and H14).

Results showed a positive and strong relationship between the two dimensions. Whether *knowledge and skills* leads to higher levels of *interest and enjoyment* or vice-versa or both cannot be proven. We can only state that a causal relationship appears to exist based upon study data. The purpose of the forensic IT class is to increase knowledge and skills. However, it is reasonable to assume that the level of interest and enjoyment will make it easier and be an important motivator for the students to acquire such knowledge. It will also motivate them to seek further knowledge in the future. Knowing that the relationship is strong should prompt instructors to increase efforts to improve students' interest and enjoyment. These facilitators have not been addressed in this paper and could be the subject of future research. Furthermore, while it was shown that the relationship between the two dimensions was stronger when measured *within* the components (accounting, IT, blend), the correlates were mainly medium to strong *between* components. This leads us to believe that even those students who are primarily interested in either accounting or forensic IT will be spurred to learn other topics if they have sufficient interest and enjoyment. Thus, the level of interest and enjoyment is an important condition to the creation of knowledge and skills.

Data also revealed that neither *knowledge and skills* or *interest and enjoyment* for the three components had significant associations with intent to pursue a CPA. However, the intention to pursue a CPA certification was very high: out of the 46 responses, 36 students stated that they did intend to pursue the certification and 10 stated that they may pursue it. Thus, the response was high regardless of the value of the component and the resulting correlates could be expected to be low.

The association of both *knowledge and skills* and *interest and enjoyment* for forensic IT and blend components with intent to pursue an IT Auditor certification are particularly strong. This evidence is important because the demand for IT auditors is increasing and provides accounting students an attractive alternative to a traditional accounting career path.

The relatively high means for the interest and enjoyment of the blend topics, as shown in Table 1, also highlights the importance of blending existing with new knowledge. Students appear to accept knowledge more readily if it is more comprehensible. If the knowledge set described by the strictly IT topics could be blended with existing accounting knowledge, it might be accepted more readily by the students.

IT Attitude was found to be an important predictor of *interest and enjoyment* for all three forensic IT topics but not as much for the *knowledge and skills* dimension. The forensic IT topics represent the more challenging topics for accounting students. Because there were six questions measuring IT attitude, there were a total of 36 correlates (six questions, three topics, and two dimensions as shown in Tables 4 and 5). Since 19 of the 36 relationships were significant and many in the moderate to strong range, the evidence shows how important attitude is to learning. For this reason, instructors should find opportunities to stress the importance of the topics to accounting careers. The various ways in which this can be done would be an important topic for future research. Overall, IT Attitude had stronger associations with all nine topics for *interest and enjoyment* than for *knowledge and skills*.

A positive attitude is clearly associated with interest and enjoyment.

5.1.1 Benefits for Researchers and Practitioners: Both researchers and practitioners can benefit from the study results. Researchers can replicate the study results to support the reliability and validity of the instrument. They can also examine ways in which instructors of forensic IT topics can increase the students' interest and enjoyment of the topics. Practitioners can use the instrument to determine the level of their own students' interest and enjoyment and to assess their IT attitudes. They can also benefit from knowing that increasing the interest and enjoyment is likely to increase the knowledge and skills outcomes. Furthermore, instructors can provide those students who are interested in pursuing IT auditor certifications with information on how to advance their IT knowledge and direct them to informational resources.

Students' attitudes were inversely related to intent to pursue a CPA. This could signify that students who are dedicated to the traditional accounting topics and are intent on pursuing a CPA will possess lower attitudes and view the class as less important than other classes that are more directed towards achieving traditional knowledge and skills and supporting CPA topics. Thus, the instructor may wish to spend more time early in the class educating students about how the forensic IT topics can help accountants in traditional jobs and how the course content is directly related to portions of the CPA exam.

Accounting departments at AACSB accredited schools regularly perform a needs analysis to systematically identify course objectives for each class. Course objectives reflect the knowledge and skills that students must attain to successfully complete the class. These are determined by input from various sources: syllabi from similar courses at other universities, discussions with instructors at conferences and workshops, an evaluation of market demand, the advisory board, and prerequisites for subsequent courses. Assessment of success includes exams, homework assignments and term papers. While vital to the overall program goals, needs analysis and subsequent assessment are outside the scope of this paper. However, this study addresses the students' perceived level of knowledge and skills in specific topics that reflect important course objectives. This information is important to assessing the success of the course in achieving its goals and in identifying gaps in knowledge.

A list of suggested approaches for improving outcomes and closing the knowledge gap is shown in Table 7. Instructors of both undergraduate and graduate classes may benefit although some are specific to the MBA class in IT Forensics Investigations for Accountants.

Finally, for future studies, the research instrument should be altered to reflect any changes in course objectives. Properly crafted, it can be an important tool for assessing needs and improving course outcomes.

Table 7: Suggestions to Improve Forensic IT Course Outcomes for Graduate Accounting Students

1	Identify the most important knowledge areas from the course objectives.
2	Measure students perceived level of *knowledge and skills* and *interest and enjoyment* for each knowledge area at the beginning and end of the course. This aids in needs assessment by identifying gaps of knowledge and areas where interest is low.
3	For strictly IT topics, motivate students' interest and enjoyment in various ways. Use current interest articles, outside speakers and videos.
4	Educate students about the value of IT knowledge to their careers. Identify career paths that depend on forensic IT knowledge.
5	Educate students about the value of a certification as IT auditor. Discuss the increase in average salary and alternative career opportunities.
6	Where possible, blend forensic IT knowledge with accounting knowledge that they already possess in order to speed their learning and create more interest.
7	Explain to students that a mastery of the forensic IT material is useful for a portion of the CPA examination.
8	Most importantly, realize that, in most cases, *interest and enjoyment* must be present before *knowledge and skills* will be acquired.

5.1.2 Study Limitations: There are several limitations to the study. First, the sample size of 46 students is small. A larger number of observations would provide higher levels of reliability and validity. However, sample size was sufficient to conduct tests of significance. Second, results represent the students' self-reported perceived level for each question. Often, with self-reporting, respondents will supply socially acceptable answers. However, anonymity was expected to adequately address this problem. Self-reporting can also lead to variability induced by individual egos. Over a large sample size, this is assumed to average out. One of the purposes of the survey was to gauge students' interest and enjoyment for each of the topics and this can only be accomplished by self-reporting. Third, the nine topics were broad without subsets. Internal control, for example, encompasses a broad knowledge set that is addressed in a number of accounting classes. However, the survey was administered two weeks prior to the end of the semester and students had developed an understanding of what was represented by each of the topics and could be expected to evaluate their levels of knowledge and skills accordingly.

6. CONCLUSIONS

Instructors of forensic IT for graduate accountants might improve attainment of course objectives by encouraging students' interest and enjoyment of course topics. They can also assess attitudes and guide students towards furthering their IT education where interest exists. Communicating the importance of course topics to students will be particularly meaningful to those students who do not perceive the importance of the topics to their own career or to pursuing advanced certification. Some students will continue to see the course as necessary to furthering a traditional accounting career and will not develop a high level of interest in forensic IT topics. On the other hand many students will embrace the forensic IT topics and be interested in pursuing an advanced certification as IT auditor. Because the course objectives are broad and the technology can be highly sophisticated, students may not perceive that they have attained even a moderate level of knowledge and skills in certain topics. In this study, students ranked their knowledge and skills for several IT topics as low (IT auditing mean = 2.52, Table 1). This may reflect a deficiency in attaining these objectives or simply a problem with perception. For an introductory forensic IT course, objectives are rather broad and sufficient time for in-depth discussions of all topics is not available. For this reason, it is imperative that the syllabus objectives adequately reflect the most appropriate mix of topics and that class time be allotted in the most optimal manner. Where students are also required to advance their skills in analytical software,

such as ACL, this task is imposing.

By creating interest and enjoyment in the forensic IT topics and stressing the importance to the students' careers, instructors are most likely to achieve success with the course objectives and improve students' learning outcomes.

APPENDIX

IT INVESTIGATIVE TECHNIQUES FOR AUDITORS:
MBA ACCOUNTING STUDENT IT INTEREST SURVEY

STRICTLY CONFIDENTIAL – DO NOT PLACE YOUR NAME ON THIS SURVEY

Instructions: The purpose of this survey is to gather information that will allow your professor and the Program of Accountancy to make informed choices about the materials included in the curriculum. Your answers are confidential and only summary information will be reviewed and evaluated. Under no circumstances will an individual student's responses be identified or used as part of his or her grade. Please answer all questions as accurately as possible.

Do you plan to pursue advanced certifications such as a CPA, CMA, or CIA?
YES MAYBE NO

Would you consider pursuing a certification as an IT Auditor?
YES MAYBE NO

Circle the number that best describes your level where 1 = Low, 4 = Neutral, 7 = High		
	Rank your level of *knowledge and skills* with …	
1	Auditing (Financial / Internal / Operations)	1 2 3 4 5 6 7
2	IT Auditing	1 2 3 4 5 6 7
3	Information Technology	1 2 3 4 5 6 7
4	IT Security	1 2 3 4 5 6 7
5	IT Governance Frameworks (COSO/COBIT)	1 2 3 4 5 6 7
6	Financial & Accounting Analytical Methods	1 2 3 4 5 6 7
7	Computer Based Analytical Methods (CAATs such as ACL)	1 2 3 4 5 6 7
8	Internal Control	1 2 3 4 5 6 7
9	Forensic & Fraud Accounting	1 2 3 4 5 6 7
	Rank your level of *interest and enjoyment* with …	
10	Auditing (Financial / Internal / Operations)	1 2 3 4 5 6 7
11	IT Auditing	1 2 3 4 5 6 7
12	Information Technology	1 2 3 4 5 6 7
13	IT Security	1 2 3 4 5 6 7
14	IT Governance Frameworks (COSO/COBIT)	1 2 3 4 5 6 7
15	Financial & Accounting Analytical Methods	1 2 3 4 5 6 7
16	Computer Based Analytical Methods (CAATs such as ACL)	1 2 3 4 5 6 7
17	Internal Control	1 2 3 4 5 6 7
18	Forensic & Fraud Accounting	1 2 3 4 5 6 7

Circle the number that best describes your *level of agreement* where 1 = Highly Disagree, 7 = Highly Agree		
19	This class has increased my overall knowledge of IT forensics	1 2 3 4 5 6 7
20	This class has increased my overall interest in IT forensics	1 2 3 4 5 6 7
21	I believe that IT forensic skills are important for accountants	1 2 3 4 5 6 7
22	I believe this class will help me perform well in my career	1 2 3 4 5 6 7
23	I believe this class will help me perform well on the CPA exam	1 2 3 4 5 6 7
24	I would be interested in taking an IT Auditing class	1 2 3 4 5 6 7

7. REFERENCES

Abu-Musa, A. (2008) Information Technology and Its Implications for Internal Auditing; An Empirical Study of Saudi Organizations. *Managerial Auditing Journal*, Vol. 23 (5), pp. 438-466.

Ahmed, A. (2003) The Level of IT/IS Skills in Accounting Programmes in British Universitities. *Management Research News*, Vol. 26 (12), pp. 20-58.

Alali, F., Grant, G. H. and Miller, K. C. (2008) IT Control Deficiencies that Impact Financial Reporting. *Internal Auditing*, Vol. 23 (4), pp. 28-37.

Buckoff, T.A. and Schrader, R.W. (2000) The Teaching of Forensic Accounting. *Journal of Forensic Accounting*, Vol. 1 (1), pp. 135-146.

Busing, M.E., Null, J.D. and Forcht, K.A. (Winter 2005/2006) Computer Forensics: The Modern Crime Fighting Tool. *The Journal of Computer Information Systems*, Vol. 46 (2), pp. 115-119.

Cangemi, M. P. (2000) (ed.) What Recruiters and Staffing Agencies Say about Trends in IS Auditing, *Information Systems Control Journal*, Vol. 5, pp. 43-45.

Cascarino, R. (2007) *Auditor's Guide to Information Systems Auditing*. John Wiley & Sons, Inc., Hoboken, NJ.

Cerullo, M. and Cerullo, M. J. (2005) How the New Standards and Regulations Affect an Auditor's Assessment of Compliance With Internal Controls. www,isaca.org/jonline. February 10, 2009.

Chaney, C. and Kim, G. (2007) The Integrated Auditor. *The Internal Auditor*, Vol. 64 (4) pp. 46-52.

Crumbley, D.L, Heitger, L.E. and Smith, G.S. (2007) *Forensic and Investigative Accounting,* 3rd ed. CCH Incorporated, Chicago, IL.

Davis, C. Schiller, M. and Wheeler, K. (2007) *IT Auditing: Using Controls to Protect Information Assets*. McGraw-Hill, New York.

Duerr, T.E., Beser, N.D. and Staisiunas, G.P. (2004) Information Assurance Applied to Authentication of Digital Evidence. *Forensic Science Communications*, Vol. 6 (4).

Dunn, C. and Grabski, S. (1998) The Effect of Field Dependence on Conceptual Modeling Performance. *Advances in Accounting Information Systems*, Vol. 6, pp. 65-77.

Gabbin, A.L. (2002) The Crisis in Accounting Education. *Journal of Accountancy*, Vol. 193 (4), pp. 81-86.

Grant, G. H., Miller, K. C. and Alali, F. (2008) The Effect of IT Controls on Financial Reporting. *Managerial Auditing Journa*l, Vol. 23, (8), pp. 803-823.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998) *Multivariate Data Analysis with Readings,* 5th ed. Prentice Hall, Englewood Cliffs, New Jersey.

Hall, J. A. and Singleton, T. (2005) *Information Technology Auditing and Assurance*, 2nd ed., South-Western College Publishing, Mason, OH.

Hunton, J. E., Bryant, S. M. and Bagranoff, N. A. (2004) *Core Concepts of Information Technology Auditing*, John Wiley & Sons, Inc., Hoboken, NJ.

Hoffman, T. (2004) IT Auditors Coveted, Hard to Find. *Computerworld*, Vol. 38 (18), pp. 1-16.

ITGI (2004) *IT Control Objectives for Sarbanes-Oxley*. www.isaca.org/sox. January 8, 2009.

Kearns, G. S. (2006) A Curriculum for Teaching Information Technology Investigative Techniques for Auditors. *Journal of Digital Forensics, Security and Law*, Vol. 1 (4), pp. 9-28.

Li, S., Huang, S. and Lin, Y. G. (2007) Developing a Continuous Auditing Assistance System Based

on Information Process Models. *The Journal of Computer Information Systems*, Vol. 48 (1), pp. 2-13.

Merhout, J. W. and Buchman, S. E. (2007) Requisite Skills and Knowledge for Entry-level IT Auditors. *Journal of Information Systems Education*, Vol. 18 (4), pp. 469-477.

Nelson, B., Phillips, A., Enfinger, F. and Steuart, C. (2008) *Guide to Computer Forensics and Investigations*, 3rd ed. Thomson Course Technology, Boston, MA.

Ravel, V. (1991) Perspectives on Students' Teaching Evaluations of AIS Courses. *The Journal of Information Systems*, Vol. 5 (2), pp. 62-72.

Sumners, G. E. and Soileau, J. S. (2008) Addressing Internal Audit Staffing Challenges. *Information Systems Management*, Vol. 25 (2), pp. 3-11.

Van Grembergen, W., De Haes, S, and Moons, J. (2005) Linking Business Goals to IT Goals and COBIT Processes. *Information Systems Control Journal*, Vol. 4, pp. 18-21

Websense (2006) Security Trends Report, Second Half 2005. www.websense.com/global/en/. December 11, 2008.

Whitman, M. E. and Mattord, H. J. (2009) *Principles of Information Security*, 3rd ed. Thomson Course Technology, Boston, MA.

Wier, B., Hunton, J. E. and Beeler, J. D. (2000) The Impact of Higher Education and Professional Certification on the Careers of Information Systems and Non-Information Systems Auditors. *Information Systems Control Journal*, Vol. 5, pp. 38-41.

Presentation: Pedagogical Issues in Digital Forensics: A Case Study

Anil Aggarwal
Accounting and MIS Department
aaggarwal@ubalt.edu
University of Baltimore
Baltimore, MD

Veena Adlakha
Management Department
vadlakha@ubalt.edu
University of Baltimore
Baltimore, MD

With worsening economy, more and more cases of fraud, theft and other legal violations are emerging. Collapse of Enron to Madoff Ponzi scheme has created an urgent need for systematic approach to identify and investigate these violations. Forensics science can assist in diagnosing such crimes. According to Wikipedia, "*forensics* encompasses the accepted scholarly or scientific methodology and norms under which the facts regarding an event, or an artifact, or some other physical item (such as a corpse, or cadaver, for example) are to the broader notion of authentication whereby an interest outside of a legal form exists in determining whether an object is in fact what it purports to be, or is alleged as being." The above issues make computer forensics a necessary topic for information systems students. Thus it is important for MIS students to understand this emerging area of technology. Computer forensics requires skills from many disciplines, mainly information systems, computer science, psychology, sociology and law. This inter-disciplinary nature of computer forensics makes it very challenging topic to teach. Some effort has already started in this direction. many universities are offering full scale programs and even degrees in digital forensics. Overill (2009) discusses digital forensics modules for computer and forensics science students. This paper goes a step further and discusses teaching digital forensics as part of an introductory information systems course (Haag and Cummins, 2008) and the pedagogical issues related to teaching digital forensics "without" the digital forensics lab.

Paper describes our experience and discusses alternate ways of teaching digital forensics and will discuss results of such experiments. Digital forensics is becoming important as economy moves into recession and more and more computer based fraud are committed. Forensics methodology usage is widespread in all business fields including accounting, finance, production, economics etc. Many functional areas have come under scrutiny for fraud, embezzlement of funds, information gaps etc. and many are committed using information technologies. As internet grows so does internet based fraud. What is interesting is that fraud is spreading into once reliable supply chain area. The latest Kroll Global Fraud Report highlights the growing danger of supply chain fraud to businesses worldwide. Large companies have increasingly become "extended enterprises" as they have globalized, outsourced, re-engineered their business processes, brought business partners and vendors closer and specialised their functions. The result makes them more complex and leaves them vulnerable to an array of frauds ranging from simple theft, to the misrepresentation of inventory to fool investors, through to counterfeiting, grey market diversion and piracy. (Kroll, 2008). According to Richard Abbey of Kroll, "Fraud thrives on complexity and companies are facing fraud from the very beginning, on every single factor: raw materials, production, and delivery," . This implies digital forensics is not confined to information systems but it can be used to detect fraud in supply chain

management (SCM), inventory management, outsourcing and similar complex processes. The investigation process may be little different in SCM and may involve digital forensics experts, accountants and production specialists.

We must prepare students to recognize and prevent fraud irrespective to the business area. Also we must teach them how to investigate cases of fraud using digital forensics. This paper is an attempt in that direction.

REFERENCES:

Haag, S and Cummings, S (2008). Management Information Systems for the Information Age, McGraw-Hill Publisher.

Kroll's report (2008) available at http://www.kroll.com/news/releases/index.aspx?id=19541

Overill, R. (2009). Development of master modules in computer forensics and cybercrime for computer science and forensic science students, International Journal of Electronic Security and Digital Forensics 2009 - Vol. 2, No.2 pp. 132 - 140

The Impact of Hard Disk Firmware Steganography on Computer Forensics

Iain Sutherland
Faculty of Advanced Technology
University of Glamorgan
CF37 1DL
+44(0)1443 654085
isutherl @glam.ac.uk

Gareth Davies
Faculty of Advanced Technology
University of Glamorgan
CF37 1DL
+44(0)1443 654085
gddavies@glam.ac.uk

Nick Pringle
Faculty of Advanced Technology
University of Glamorgan
CF37 1DL
+44(0)1443 654085
npringle@glam.ac.uk

Andrew Blyth
Faculty of Advanced Technology
University of Glamorgan
CF37 1DL
+44(0)1443 654085
ajcblyth@glam.ac.uk

ABSTRACT

The hard disk drive is probably the predominant form of storage media and is a primary data source in a forensic investigation. The majority of available software tools and literature relating to the investigation of the structure and content contained within a hard disk drive concerns the extraction and analysis of evidence from the various file systems which can reside in the user accessible area of the disk. It is known that there are other areas of the hard disk drive which could be used to conceal information, such as the Host Protected Area and the Device Configuration Overlay. There are recommended methods for the detection and forensic analysis of these areas using appropriate tools and techniques. However, there are additional areas of a disk that have currently been overlooked. The Service Area or Platter Resident Firmware Area is used to store code and control structures responsible for the functionality of the drive and for logging failing or failed sectors.

This paper provides an introduction into initial research into the investigation and identification of issues relating to the analysis of the Platter Resident Firmware Area. In particular, the possibility that the Platter Resident Firmware Area could be manipulated and exploited to facilitate a form of steganography, enabling information to be concealed by a user and potentially from a digital forensic

investigator.

Keywords: Digital Forensics, Hard Disk Drive, Firmware, Steganography.

INTRODUCTION

One of the main forms of storage media and therefore sources of data for forensic analysis is the hard disk drive. Current forensics practice for the most part deals with the analysis of hard disk drives found in server, desktop and laptop systems. It is also now becoming common for these drives to be embedded in certain other devices such as CCTV systems, games consoles and entertainment systems. The majority of the literature available on the forensic investigation of these devices concerns the analysis and extraction of evidence from the numerous forms of file systems which may reside in the user accessible area of the disk. It is known that there are other manufacturer areas of the drive which could be used to conceal information. However there are further areas of the drive that could be manipulated to enable data to be concealed from a user and potentially from a forensic investigator.

The focus of this paper is to highlight the potential problems that could arise from the manipulation of disk firmware, enabling the possible concealment of information on hard disk drives. Currently this technology is available mainly in data recovery labs that have specialised hardware to deal with the problem of faulty hard disk drives. However the technology is becoming available to the end user and as such should be of concern to the forensic investigator.

THE HARD DISK

The most common form factors are the 3.5 inch and 2.5 inch disks found in desktop and laptop systems respectively, although smaller versions can be found in other devices such as older versions of the Apple iPod. The disk may be housed either within the system unit or in a removable caddy. The main types are ATA and SCSI (Small Computer Systems Interface).

A hard disk drive is a complex device and can be viewed as a small computer system in itself. The hard disk drive is composed of platters, voice coils, read / write heads, casing, mountings, a motor, and a controller board. The data area consists of a stack of metal, ceramic or glass platters coated with a magnetic film. One rotation of the disk at a particular radius is known as a track. For sets of surfaces, a set of tracks at the same radius is known as a cylinder. A separate armature and head assembly is present for each disk surface (one surface may be used to control the position of the read/write heads). The sector is the smallest addressable unit. A specific sector address can be found using the cylinder address (C) the Head (H) and the Sector (S). At a higher level of abstraction the Logical Block Address (LBA) method assigns a sequential address to each sector, but which may not relate to the block's physical location.

USER ADDRESSABLE SPACE

Hard disk maximum sizes are currently in the region of 1TB although speculations on higher capacities suggest the possibility of 4TB disk in the near future. The continuously increasing size of the hard disk drive poses a problem for the potential forensics examiner as it provides an ever increasing search space to examine as part of an investigation. Also with the increasing capacity and the decreasing cost of this form of storage device it is now not uncommon to find two or possibly more drives in a single home system, further increasing the problem for the forensic examiner. In the case of most drives, once the drive has been formatted and a file system is put in place the typical amount of storage is slightly less than the stated capacity. The forensic investigator has to be very careful to ensure they are aware of the correct storage capacity of the drive and have accessed all of the available information present on the storage media. This may include areas not addressable by the average user.

NON-USER ADDRESSABLE SPACE

Not all areas of the disk are addressable by the host computers operating system. In addition to the user addressable space there are areas of the drive that are used for the manufacturer to record data and perform diagnostics. These include the Host Protected Area (HPA) and Device Configuration Overlay (DCO), either or both of which can exist on a hard disk [1], [2] and in addition to this the firmware areas, which exist in the lower range of the address space, sometimes referenced by negative numbers (see Fig.1 below).

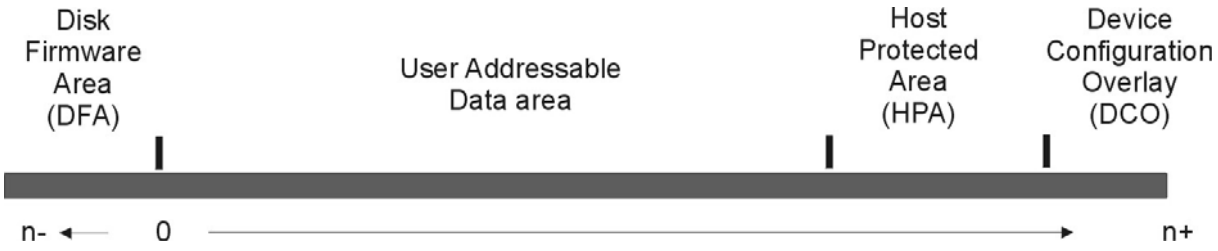


Figure 1. Overview of Disk Areas

Host Protected Area

The Host Protected Area (HPA) is used for holding diagnostics and other utilities required by the PC manufacturer (Gupta 2006). The presence of an HPA can be identified by commands READ_NATIVE_MAX_ ADDRESS which provides the total number of sectors on the disk and IDENTIFY_DEVICE which provides total sectors a user can identify. Any difference between these two values indicates an HPA is present on the device [2].

Device Configuration Overlay

A Device Configuration Overlay (DCO) is similar to the HPA, but is used by manufacturers to configure drive sizes [3] and may exist at the same time. If a DCO is on the device this can be detected by the difference in values returned from the following two commands; READ_NATIVE_MAX_ADDRESS and DEVICE_ CONFIGURATION_IDENTIFY. An excellent overview of HPA and DCO are provided in Carrier [2].

Firmware

In addition to the HPA and DCO there are other areas of the drive which are not addressable by the average user, but are vital to the disks correct operation. The firmware area is essential for correct drive operation. The firmware is composed of a series of modules, examples are; SECU (Security System Module), P-List, G-List, T-List, SMART Attributes, and U-List (Firmware Zone Translator). A portion of the disk firmware is resident on the drive platters, this is loaded by code located on the controller board of the hard-drive. Therefore these modules are located on either a) the hard-drive in a zone that is not normally accessible by the user or operating system, and b) into flash memory located/embedded on the disk controller. The firmware controls all aspects of the internal hard drive operation from system startup:

- On start-up, when a hard drive is powered on, the controller board loads the firmware modules from the disk platters into memory ready for execution. The firmware that is responsible for correctly configuring the hard drive and putting it in a ready state; providing all elements of the disk are working correctly the disk then presents itself as ready and allowing the host PC

to load the Operating System.

- During the operation of the hard drive it is the firmware that ensures the correct operation of the hard drive, allowing it to correctly interact with other components on the system.
- When the hard drive is powered down, a shutdown sequence is executed by the firmware that ensures the hard drive powers down correctly so that it will operate successfully the next time it is powered on.

Various companies such as IBM, Hitachi and Western Digital provide software that can be used to up-date the firmware located on a hard-drive, sometimes allowing for more efficient operation.

FIRMWARE OPERATIONS

Firmware performs a number of key functions; one of these is SMART (Self-Monitoring, Analysis, and Reporting Technology) logs. As part of the ATA 3 standard there are a number of criteria which are monitored and logged as "threshold not exceeded" or "threshold exceeded" Attributes include read error, seek error, temperature, drive operation time etc. These Self-Monitoring, Analysis, and Reporting Technology (SMART) logs are aimed at predicting drive failure. The SMART attributes monitored depends on the manufacturer, for this reason they are usually of little forensic use to the examiner due to the differing implementations of the SMART log criteria in the different disks.

The firmware is also responsible for monitoring defect control, no disk is manufactured without some flaws and there will be some sectors on the drive which cannot be used. This process is transparently handled by the hard-disk and occurs 'beneath' the operating system level via the two lists, P and G [4]. Flaws identified on the drive during production are recorded in the disk firmware as the 'P' (permanent / primary / production) list. As the disk ages and through wear & tear other sectors may fail; this is recorded in the 'G' (growth) list. Reads and writes are automatically redirected (remapped) to spare sectors. P-list and G-list sectors are automatically bypassed by the drive electronics, and P/G-list sectors do not slow down drive access. By adding or removing a sector from the P-List and/or G-List we have the ability to hide/make-visible data on the hard-drive. Tools such as HDD Bad Sector Report [5] support a limited set of commands to modify/zero the P/G-lists.

TOOLS AND TECHNIQUES

To date the tools required to perform significant modifications of firmware are fairly expensive and found mostly in data recovery laboratories. However the authors are aware of two major commercial products available for this type of analysis and modification/repair of hard disk drives. Both sets of equipment were examined and comprise a combination of hardware and software tools. One coming from Russia and costing in the region of $3000 for a UDMA set of tools. The full suite which includes the ability to extract data and work with some solid state devices and SCSI disks is in the region of $15,000. A cheaper and more readily available device is offered from China and can be obtained via resellers in Europe for around $300 per disk type (manufacturer). There are also a number of free/share tools that claim to read portions of the firmware, most commonly the disk serial number [6].

A SIMPLE STEGANOGRAPHY EXPERIMENT

This paper focuses on the possibility of concealing information on the disk via the manipulation of the firmware and examines the issues and ease of use of some of the tools and techniques available. The authors are aware of a limited amount of material on this problem [7] although the potential for this particular form of drive behaviour was highlighted a number of years ago [8]. The ability to manipulate the disk firmware raises a number of potential issues which are demonstrated by the following experiment:

A 3.5" Fujitsu Hard Disk Drive (Model MPA3035AT) was forensically wiped, overwriting the

contents with zeros. The disk was then reformatted with a partition and populated with a NTFS files system and a range of typical files; .doc, .txt, .jpgs etc. Although the additional files are not directly related to the experiment they demonstrate the type of digital environment in which this type of steganography might be performed.

One of the text files (.txt) present on the drive was chosen at random and edited to include a distinctive keyword (a combination of the author names). This keyword, under normal circumstances residing within a document contained in the file system, would be easily located during a forensics analysis using the search facilities included in a number of commercial forensic software tools.

One of the firmware recovery tools outlined above was used to view the drive contents and also to locate the physical sector in which the selected text file was located. The particular model of the Fujitsu disk selected for this experiment supports two error lists in the firmware; one firmware list relating to production defects and another list relating to failing tracks on the drive. Any modification to the production list appears to require a reformatting of the drive. The P-list was not used in this experiment as our goal was to attempt to prove this form of steganography can be achieved on a 'live' system. Therefore the firmware error list relating to defective tracks (T-list) was modified to include an additional entry relating to the physical location of the modified text file.

The disk was rebooted and mounted. The firmware modification tool was then used to view the drive contents in an attempt to access the target file. The drive could no longer access the physical location (hidden data area) nor the data residing at that location. This was also confirmed via external hex editors, what is more the keywords were not present in any searches performed on the drive. The data is inaccessible by the disk drive and the computer operating system.

The firmware recovery tool was used to edit the error list returning it to its original state removing the previously added entry. The data area and text file containing the keyword was accessible on the drive.

DISCUSSION ON FORENSIC IMPACT

The findings of this initial experiment suggest that the use of this technique would permit a form of steganography enabling data to be concealed on the disk. The amount of space referenced by firmware error lists could be substantial and span thousands of sectors storing a significant amount of information. This is in effect a more sophisticated version of marking a sector as bad on a floppy disk, as the operating system cannot access this portion of the disk.

This has the potential for a significant impact on computer forensic practice. It appears that tools commercially available for $300 can be used to conceal information from a forensic investigator in a process that is relatively easy to accomplish, but difficult to detect. Standard forensic software cannot access the areas of the disk marked as 'bad' by a disk firmware error list. The manipulation of firmware is also not limited to the error lists [4]. In addition to the areas highlighted above disk firmware also contains the instructions for performing the LBA=CHS mapping, converting the LBA to the actual CHS locations on the disk. Standard forensic tools such as 'Encase' and 'AccessData' rely on the firmware translator operating correctly. They simply read all of the LBA's provided by the translator and create a corresponding disk image. The impact for forensics is that if the mapping has been tampered with or altered in some way then the data will not be present in the forensic image.

These problems suggest that in certain cases a standard forensic image may no longer be sufficient to fully analyse a hard disk drive and it may be essential to perform additional analysis on the original media. The analysis of firmware could be problematic for a number of reasons compounded by the fact that manufacturers implement the firmware in different ways on different models. This then poses a problem in the analysis of a disk if the validity of the firmware locations is to be assessed. The investigator will require a firmware tool to attempt a number of options:

Obtaining an identical model of disk will permit the validity of certain areas of the firmware on the

suspect's disk to be checked. This can also be accomplished by comparing the suspect disk to a database of known firmware. This will detect the modification of some portions of the firmware such as the code that translates an LBA to a physical disk location [4]. However, the usual forensic practice of hash comparison would necessitate an MD5 hash for each and every firmware component.

Some of these components, in particular the error lists the target of this form of steganography are unique to each drive from the point of manufacture and evolving as the disk wears during normal operation. It is therefore impossible to validate the error list by comparison to another drive. Removal of the error list also presents problems as this may render the drive unusable.

The potential impact of the malicious functionality of the firmware both as a result of correct or malicious operation can also impact on information security. When wiping the disk the sectors in the error lists are not seen by the operating system so data may be left on these bad sectors. Disk disposal has been documented as a significant issue [9], [10], [11] and traditional disk disposal tools may not be aware of this feature of a disk drive. The firmware in these disks may fail. The G-list may become full on some disk models and as a result the disk may stop working. An error in the firmware can prevent the disk being accessed while still physically healthy. Users with access to the appropriate tools and technology can now recover the data with relative ease. These users may be forensic investigators or those with less honourable intentions; potentially commercial competitors, foreign states engaged in commercial espionage although the technology is now becoming available to private individuals.

FUTURE ISSUES FOR CONSIDERATION

It is probable that the tools used to manipulate firmware and accomplish this form of steganography will become more readily available. If this is the case then future work should be focused in a number of key areas: It is suggested that work should also be undertaken to determine forensic best practice and procedure in this area. In some cases it may not be sufficient to work on an image of the disk drive. In particular it is of concern that this may be exploited as a possible route to introduce malware into a computer system. There are also a number of potential avenues that should be explored for solutions including the standardisation of some aspects in the way which disk firmware is structured or the generation of a library of trusted and true firmware which a forensic investigator can use to compare a suspects disk, although this would need to be a substantial library to cover all of the available models of hard disk drives.

SUMMARY AND CONCLUSIONS

This paper has discussed the implications of the malicious modification of firmware. It has suggested this raises a number of issues in the area of forensics and information security. In particular this route for steganography does not appear to be detectable by current forensic tools. In terms of information security working disks should make use of the ATA Secure erase function to ensure the secure removal of information from a disk. Disks that are faulty and which fail to be recognised by the operating system may not be truly 'dead'. This suggests disks considered dead or faulty should be put through some form of physical disposal system

ACKNOWLEDGEMENT

The authors would like to thank the members of the Information Security Research Group (IRSG) and in particular Mr. Konstantinos Xynos and Mr. Simon Harries.

AUTHOR BIOGRAPHIES

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

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

Mr. Nick Pringle is a part-time Ph.D. Student at the Faculty of Advanced Technology in the University of Glamorgan. The main focus of his research is forensic analysis of large data sets. He has been involved in a variety of other research projects in the area of Computer Forensics, notably hard disk recovery. Mr Pringle has also acted as a consultant and assistant investigator on disk recovery technology cases in the University's Computing Forensics Laboratory.

Professor Andrew Blyth is Head of the Information Security Research Group at Faculty of Advanced Technology. His main interests lie in the area of Computer Network Management and Computer Network Defence and Computer Forensics. He has acted as an Expert Witness for National Police and Government.

REFERENCES

[1] Vidström A., (2005) *Computer Forensics and the ATA Interface*, Technical report Swedish Defence Research Agency, FOI-R--1638—SE, February 2005, 1650-1942

[2] Carrier B, (2005) *Forensic File System Analysis,* Addison Wesley.

[3] Gupta M.R., Hoeschele, M.D., Marcus K. Rogers M.K., (2006) *Hidden Disk Areas: HPA and DCO.* International Journal of Digital Evidence, Fall 2006, Volume 5, Issue 1

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

[5] Badtrk (ADM) Documentation (Accessed 11/3/09)

http://docsrv.caldera.com:507/en/man/html.ADM/badtrk.ADM.html

[6] HDD Firmware Serial Number Source Code 1.01 Free Download (Accessed 11/3/09)

http://www.softlow.com/windows/development-tools/debugging/shareware/hdd-firmware-serial-number-source-code.html

[7] Davies G. & Sutherland I. (2009), *Forensic Implications of the modification of Hard Disk Firmware*, Proceedings of the Fourth Research Student Workshop, University of Glamorgan, 12th March 2009.

[8] Gutmann .p (1996) *Secure Deletion of Data from Magnetic and Solid-State Memory.* Proceedings of The Sixth USENIX Security Symposium, July 22–25, 1996, San Jose, California, USA

[9] Jones A., Valli C., Sutherland I. (2006) *An Analysis of Information Remaining on Disks offered for sale on the second hand market.* Journal of Digital Security, Forensics & Law. Volume 1, Issue 3.

[10] Jones A., Dardick G., Sutherland I, Valli C., (2009) *The 2007 Analysis of Information Remaining on Disks offered for sale on the second hand market.* Int. J. Liability and Scientific Enquiry. Vol.2 (1), pp.53–68

[11] Sutherland I, & Mee V. (2006) *Data Disposal: How educated are your Schools?,* 6th European Conference on Information Warfare and Security, June 2006.

Analysis of the 'Db' Windows Registry Data Structure

Damir Kahvedžić
Centre for Cyber Crime Investigation,
University College Dublin, Ireland,
Tel: +353 1 716 2485
Email: damir.kahvedzic@ucd.ie

Tahar Kechadi
Centre for Cyber Crime Investigation,
University College Dublin, Ireland,
Tel: +353 1 716 2478
Email: tahar.kechadi@ucd.ie

ABSTRACT

The Windows Registry stores a wide variety of data representing a host of different user properties, settings and program information. The data structures used by the registry are designed to be adaptable to store these differences in a simple format. In this paper we will highlight the existence of a rare data structure that is used to store a large amount of data within the registry hives. We analyse the manner in which this data structure stores its data and the implications that it may have on evidence retrieval and digital investigation. In particular, we reveal that the three of the most popular digital investigation suites fail to recognise this structure and do not allow the investigator to view the contents of the structure.

Keywords: Windows Registry, Data Structure

1. INTRODUCTION

The Windows registry, a very extensive database in the Windows Operating system, is designed to hold a wide variety of information used by the operating system and installed software to configure settings and enhance user experience. It is designed to be extensible and scalable with respect to the amount and type of data that it holds. Numerous data structures provide this functionality, the exact specification of which has recently been published (B. D. (2009), Morgan, T. D. (2008)). Implications of these structures with respect to data hiding and forensics have also only recently been explored in Morgan, T. D. (2008).

The db data structure is a registry data structure that has not been described in these publications. It is used by the Windows NT registry to store a value of a key that exceeds a certain (very large) length and as such is rarely found in the registry. Because of its rarity and lack of specification, the structure seems not to be supported by any of the large registry analysis software vendors in the Digital Forensics field. Programs such as Forensic Toolkit (FTK), Encase and X-Ways do not support it and fail to show the data accurately. This report will document the db data structure, how it is used in the registry and what relevance it has to digital forensic investigations.

2. THE BASIC STRUCTURE OF THE REGISTRY

Each registry hive is organised into a set of specific data structures each storing specific parts of the registry. The biggest and most general one of these are called hbin blocks. They are usually 4096 bytes, 4kb, (but can be a multiple of it) and store all the other smaller data structures. The most important structure is the nk structure which stores all necessary information to define a key. Amongst other things it stores pointers to a set of vk data structures. This structure holds particular values for a

key. Therefore if a key called 'Microsoft' has a value 'Version : REG_SZ : 1.1' then the value would be stored in a vk. 'Version' would be the name of the value and 'REG_SZ' the datatype and '1.1' the data.

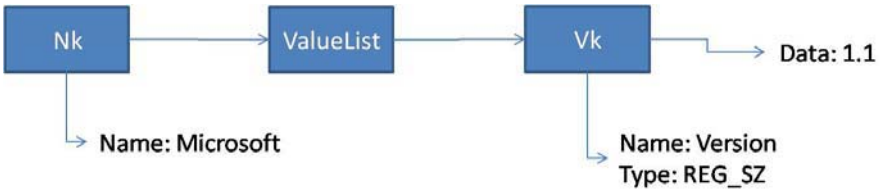


Figure 1: Typical Use of the Vk Structure

The situation is summarised in Figure 1. Further detailed specifications on the various data structures of the registries are found in (Morgan, T. D. (2008)).

The vk structure does not store the data itself but rather points to another location in the file where the actual data is stored. However, if the value is big enough the offset does not point to the actual data but to another data structure called 'db'. It is this data structure that acts as an intermediary between the vk node and the large amount of data that is stored in it.

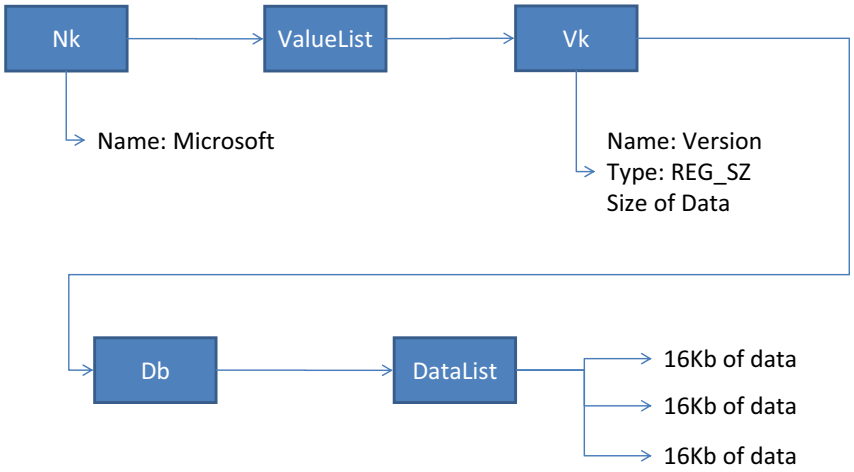


Figure 2: The Use of the Db Structure

3. THE STRUCTURE OF DB

The db data structure stores two offsets; one points to the set of data blocks where the data is stored, the other is unknown and appears to be useless. The first pointer points to a set of offsets that point to the actual data. The data is broken up into 16kb chunks. One large hbin block is allocated to each of these chunks. Therefore, if the value of the key goes above 16,344 bytes a db block is put into place where the actual data should be as show in Figure 2. The db data structure contains offsets that point to a DataList structure. The DataList is a data structure that does not contain an identifier and contains offsets to each of the 16kb chunks storing the value's data. The structure of the db and the DataList are shown in the Tables 1 and 2.

Structure Of Db		
Offset	Size	Description
0	dword	Size of Block
4	word	ID of Block (db)
6	dword	Number of 16kb data blocks
8	dword	Offset of DataList
12	dword	Unknown

Table 1: Configuration of the db Data Structure

Structure Of DataList		
Offset	Size	Description
0	dword	Size of Block
4+	word	Offsets of 16kb Data Blocks

Table 2: Configuration of the DataList Data Structure

4. THE DB SLACK

The total size of the data distributed to the various 16kb data blocks is stored in the vk data structure. The blocks are allocated to the data regardless if the data does not fully fill the 16kb space. A marker "0x00 00 00" identifies where the data ends. Anything after this marker is not read. The data found after this marker is not part of the windows registry hierarchy and may contain remnants of previous data structures or data. If for instance, the data is shortened, the end marker is changed to reflect the new end point of the data. The excess data is not deleted and remains as *slack*. Previously stored information can be found in this place. The slack is illustrated in Figure 3.

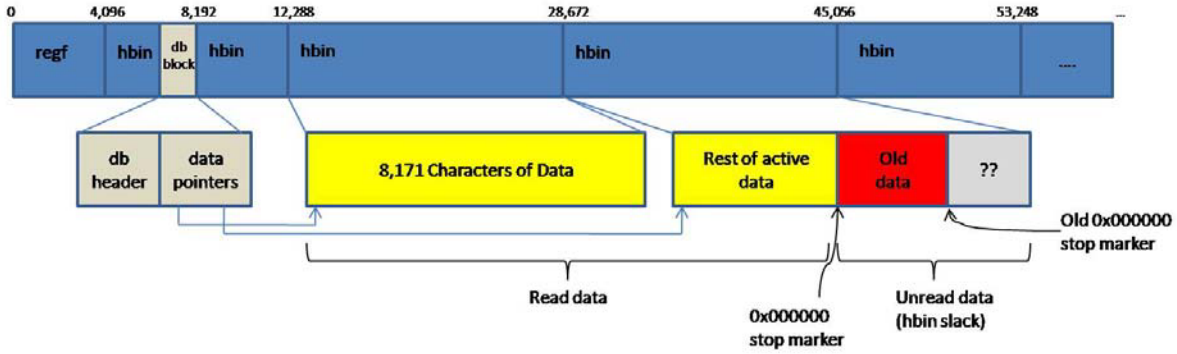


Figure 3: Db Slack

5. PROBLEMS IN CURRENT FORENSIC SOFTWARE

The current leading forensic software is not aware of the db structure and fails to parse it properly. The db data structure is processed correctly in the RegEdit32 program in Windows but offline examination of values using the db data structure is not possible with the forensic tools. As such users can take advantage of this flaw and store large (possibly encrypted) files within the registry in an attempt to hide the data. As a demonstration, a large string value was placed into a registry key. The RegEdit32 window in Figure 4 shows what the data should store. This section will review the leading forensic software and how they behave with respect to this db structure.

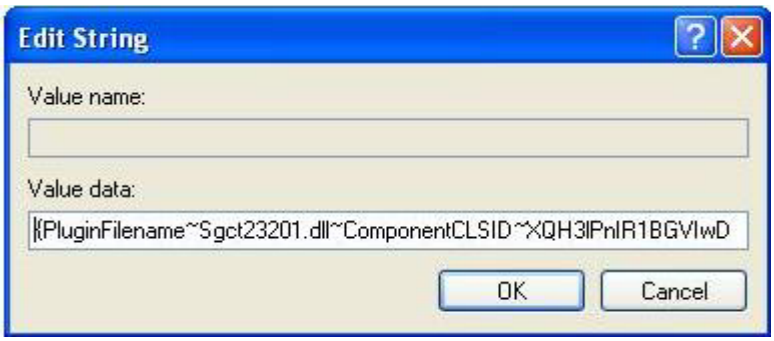


Figure 4: Normal Vk Data in RegEdit32

5.1 EnCase

The leading forensic software, EnCase, treats registry hive files as extension of the file system. To view a hive, the investigator must mount it as a drive and explore it like a file system. Data is show in Hex as well as ASCII. It doesn't recognise the special circumstances that create the Db data structure. In this case, Encase attempts to grab the data found at the vk address regardless if there is a db or not. As a consequence it does not follow the db offsets and reads in whatever data happens to be at the offset. As a result, it appears as if the key holds nothing but random data, thus giving a false impression. The situation is illustrated using EnCase Enterprise v.6.8 (EnCase (2009)) and is shown in Figure 5.

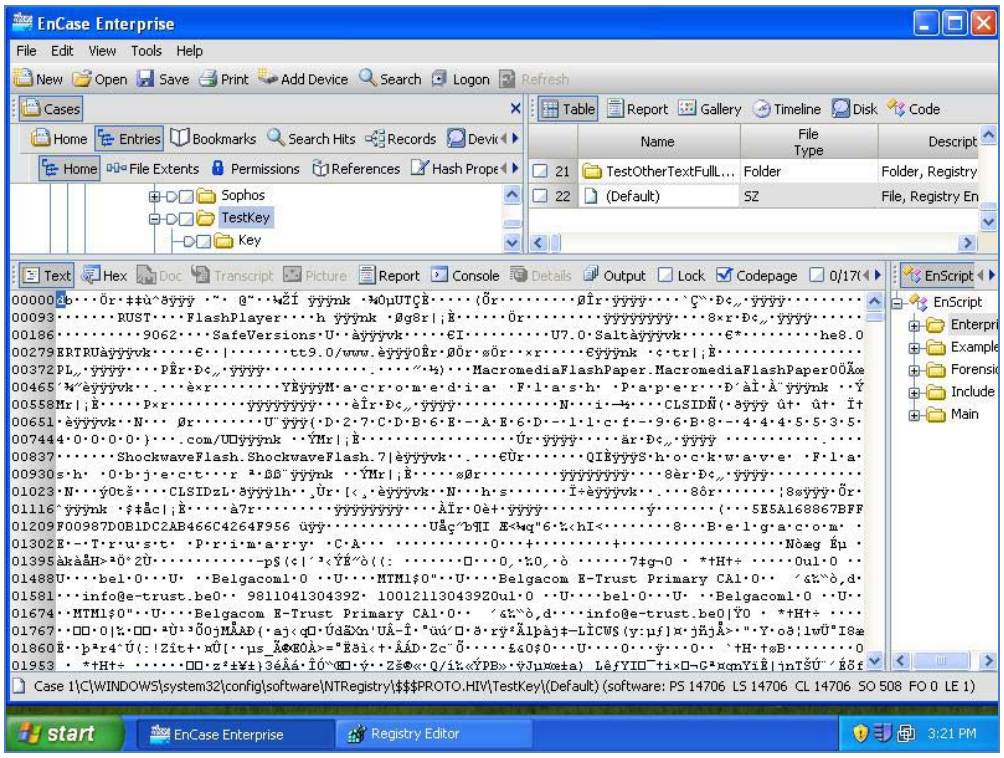


Figure 5: Encase Displaying Incorrect Data

5.2 X-Ways Forensics

The X-Ways Forensic v14.0 (X-Ways (2009)) program includes a separate registry viewer to view the hive files in a similar manner to RegEdit32. Similarly to EnCase above, if a registry key with the db data structure is found the data is read at the db offset. The structure is not parsed properly and attempted to be read as Unicode string values. Since the values describing the db are binary the

displayed data is nonsensical. The program throws an exception and crashes shortly after viewing the data.

5.3 Forensic Toolkit

Access Data's FTK Registry Viewer v1.5.0.14 (Forensic Toolkit (2009)) program fails to find any data at the db data structure and displays nothing. It does not crash but fails to display the correct data.

6. SUMMARY AND DISCUSSION

We have presented a specification for a rare registry data structure, db, used by the Windows registry to manage a key with a huge amount of data. We have described the function of this structure and analysed it with from a forensic point of view. As such we have described the structure's slack that may store remnants of past data.

The Db data structure is exceedingly rare but is created by at least one major multimedia software provider, such as Real Player (RealNetworks (2009)). As mentioned above, the db data structure is accessed as normal using the RegEdit32 program. Data being stored in it can be readily accessed using this Windows built in application. However investigators may fail to examine it if viewed offline using any of the aforementioned investigation suites. Although the data referenced may be found by brute force string searching algorithms the data would not be attributable to any active part of the registry and may bring doubt to any assertions made based on that evidence.

REFERENCES

B. D. (2009), 'Registry File Format', http://home.eunet.no/pnordahl/ntpasswd/WinReg.txt, visited Feb 2009.

Encase (2009), 'Guidance Software Digital Investigations', http://www.guidancesoftware.com/, visited Feb 2009.

Morgan, T. D. (2008) "Recovering Deleted Data From the Windows Registry", Digital Forensic Research Workshop, Aug 2008, Baltimore, USA.

RealNetworks (2009), 'Real Player 11 Basic', http://europe.real.com/player/win/, visited Jan 2009.

Forensic Toolkit (2009), 'Access Data', http://www.accessdata.com/, visited Feb 2009.

X-Ways (2009), 'X-Ways Software Technology AG', http://www.x-ways.net/, visited Feb 2009.

ACKNOWLEDGEMENTS

The authors would like to acknowledge "Higher Education Authority" (HEA) and "Irish Research Council for Science, Engineering and Technology" for providing funding that made this research possible.

CORRELATING ORPHANED WINDOWS REGISTRY DATA STRUCTURES

Damir Kahvedžić
Centre for Cyber Crime Investigation,
University College Dublin, Ireland,
Tel: +353 1 716 2485
Email: damir.kahvedzic@ucd.ie

Tahar Kechadi
Centre for Cyber Crime Investigation,
University College Dublin, Ireland,
Tel: +353 1 716 2478
Email: tahar.kechadi@ucd.ie

ABSTRACT

Recently, it has been shown that deleted entries of the Microsoft Windows registry (keys) may still reside in the system files once the entries have been deleted from the active database. Investigating the complete keys in context may be extremely important from both a Forensic Investigation point of view and a legal point of view where a lack of context can bring doubt to an argument. In this paper we formalise the registry behaviour and show how a retrieved value may not maintain a relation to the part of the registry it belonged to and hence lose that context. We define registry orphans and elaborate on how they can be created inadvertently during software uninstallation and other system processes. We analyse the orphans and attempt to reconstruct them automatically. We adopt a data mining approach and introduce a set of attributes that can be applied by the forensic investigator to match values to their parents. The heuristics are encoded in a Decision Tree that can discriminate between keys and select those which most likely owned a particular orphan value.

Keywords: Windows Registry, Data Structures, Retrieval, Orphans, Correlation

1.INTRODUCTION

The Windows Registry is a hierarchical database that stores information about the system software, hardware, its users and their preferences. Investigators tend to concentrate on the active data already present in the hives (Carvey, H. (2005), Farmer, D.J. and Burlington V. (2009), Registry Hives. (2008), Kahvedžić, D. and Kechadi, T. (2008), Kahvedžić, D. and Kechadi, T. (2008)ii, Wong, L. W. (2009)). After it has been deleted however, this information may still reside in the system files of the registry (B. D. (2009), Morgan, T. D. (2008)). The space of the deleted keys is marked as free and can be reallocated for new keys. If the space is not yet overwritten, the deleted keys can be retrieved. The keys are found by simply parsing deallocated space and following any links to their values and data. The links between the data structures therefore are used to correlate one data structure to another.

However, the structures found in a deleted state may be in a corrupt state and may not preserve the full information that it contains while it was active (Kahvedžić, D. and Kechadi, T. (2009)). The links in particular may not be preserved. Some data structures cannot be reattached to the registry tree and cannot be viewed in context. Registry key values are particularly important, since they store the actual key data and do not store links to their parents. We call all data structures that cannot be reattached to the registry hierarchy orphans.

Software uninstallers and registry cleaners (JavaCoolSoftware (2009)) usually delete many values and keys from the registry. Many links between the deleted values and the points in the registry that they were deleted from can be easily retrieved. However some links may be lost and require specific

processing to retrieve them. In this paper we will illustrate how many orphans are created during an uninstallation. We will then formalise the problem of reassembling these keys using a finite state machine model and describe a methodology for reattaching them by exploiting a number of observations on how the registry keys are managed. We consider these observations as attributes and take a data mining approach to reattach the orphaned values to the most likely owner keys.

Section 2 defines the formal model. Section 3 discusses the data mining approach undertaken to solve this problem and the experimental setup that is used to test the technique and validate the results. Section 4 discusses the various patterns identified in the way that the registry stores its data structures. Section 5 combines the various attributes to construct the Decision Tree classifier. Section 6 and 7 summarise the paper and describe future work.

2. FORMAL MODEL

We concentrate on the three major registry data structures; the key, the value and the security key (B. D. (2009)). In this section, we formalise the operation of the registry with respect to these structures. We define and describe how orphans are created and formalise the problem of connecting them to the most likely parent keys.

2.1 Concepts

Let $\mathbb{R}, \mathbb{V}$ and $\mathbb{S}$ be the set of all keys, values and security keys respectively in the registry. Let $\mathbb{U}$ be the union of all of them. Any registry entry (key) r can therefore be defined as a triplet consisting of a set of subkeys, values and a single security key:

$$r_i(k_i, v_i, s_i), \text{ where } k_i \subseteq \mathbb{R}, v_i \subseteq \mathbb{V}, s_i \subseteq \mathbb{S}$$

$r(0,0,0)$ is a key that does not have any values, subkeys or security keys associated to it. The maximum number of structures a key can have is $|\mathbb{R}| \quad 1$ subkeys, $|\mathbb{V}|$ values and a single security key, where $|\mathbb{R}|$ and $|\mathbb{V}|$ denote the cardinality of $\mathbb{R}$ and $\mathbb{V}$, respectively.

In addition, the keys can be in one of the two states; Active ($\mathcal{A}$) and Deleted ($\mathcal{D}$). All keys must have been active at some point in the registry. At certain time *t*, we can have some active keys and deleted keys. So we can write:

$$\mathbb{R} = \{\mathbb{R}_a \subseteq \mathcal{A} \cup \mathbb{R}_d \subseteq \mathcal{D}\}$$
$$\mathbb{V} = \{\mathbb{V}_a \subseteq \mathcal{A} \cup \mathbb{V}_d \subseteq \mathcal{D}\}$$
$$\mathbb{S} = \{\mathbb{S}_a \subseteq \mathcal{A} \cup \mathbb{S}_d \subseteq \mathcal{D}\}$$

Therefore the set of all deleted structures is $\mathbb{U}_d = \mathbb{R}_d \cup \mathbb{V}_d \cup \mathbb{S}_d$. In normal operation of the registry, the user cannot access the deleted structures or retrieve deleted data. All of the keys available to the user are in the Active state and are denoted by $\mathbb{U}_a$.

We can extend the definition of a key r by adding a number of constraints based on its state:

$$r_d \equiv r(k_d, v_d) \text{ or}$$
$$r_a = r(k_a, v_a, s_a)$$

Therefore, all deleted keys must reference deleted structures. A deleted key is not restricted to reference a deleted security key. Security keys can be used by multiple registry entries, ensuring a similar permissions policy (B. D. (2009)).

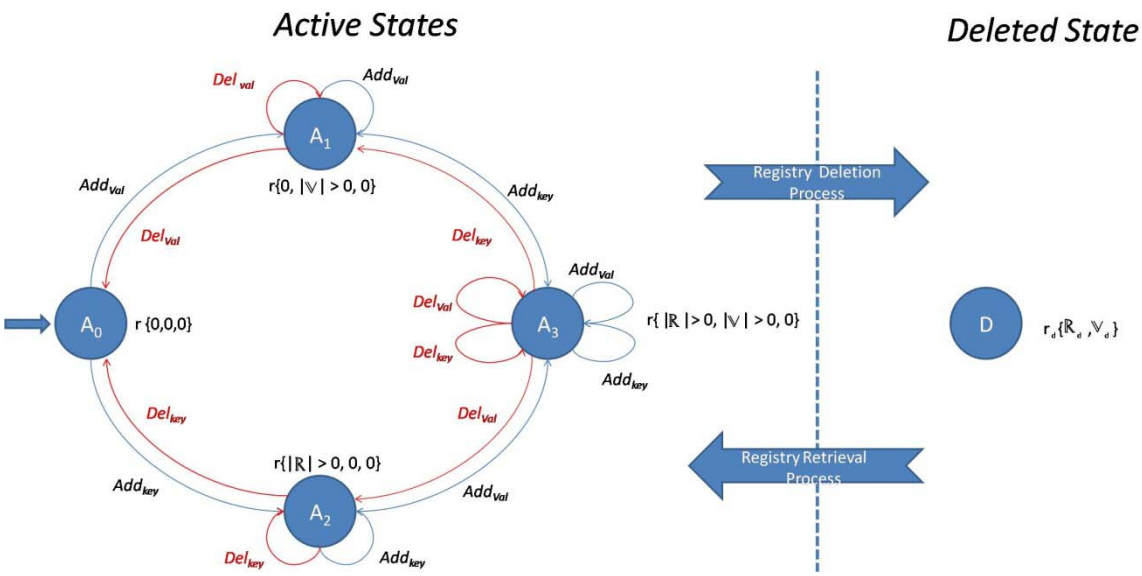


Figure 6:Deletion and Retrieval of Keys

Finite State Machine

A finite state machine model is adopted to illustrate the different actions that are possible on the registry key during its lifetime. Formally, a finite state model (FSM) for this problem can be defined as a tuple of three elements $T = (Q, I, \varphi)$ where:

- Q: A finite set of all possible states.
- I: A finite set of all possible events.
- $\varphi : I \times Q \rightarrow Q$: A *transition* function that maps input symbols of the current state to the next state.

The resulting FSM can be illustrated with a transition graph. Figure 1 illustrates the states, transitions and events between them. A single key can contain many peripheral data structures to store values and keys more efficiently. The starting state of the key is an empty state where it contains no subkeys or values. During its lifetime a key can have multiple values or keys and conversely can have these subkeys or values removed. The events are termed $Add_{Val}, Add_{Key}, Del_{Val}, Del_{Key}$ and have important distinctions.

The transitions Add_{Val} and Add_{Key} associate a new value or new subkey to a key. The space for the structures is allocated and a link from the key to the structure is created. In the case of a new subkey, a link from the subkey to its parent is also created.

The transition, Del_{Key}, converts $r_a \in \mathbb{R}$ to r_d. All of the structures referenced by the key are deleted, $r(k_a, v_a, s_a)$ to $r(k_d, v_d)$. The links between them are preserved and the associations can be retrieved. The key data structure itself contains a link from the structure to the parent key in the hierarchy. These are not modified when a key is deleted and can be retrieved and used to reattach the deleted keys to the tree hierarchy.

The transition, Del_{Val}, deletes a single value and converts it from $v_a \in \mathbb{V}$ to v_d. A value data structure does not retain a link to its parent key. Therefore the value, once deleted, is disconnected from the tree hierarchy. They cannot be easily reattached to the registry hierarchy.

The series of transitions from one state to another are termed *computations*. A finite computation is a sequence of steps $c_f = (t_f, q_f)$ where each step is made up of an event, $t_f \in I$, and a state $q_f \in Q$.

The set of computations is denoted by $\mathcal{C}_T$. The computation, $\mathcal{C}_T$, is defined by a finite number of steps, $\mathcal{C}_T(t)$, ranging from key creation to key deletion.

Orphans are defined as any data structure that has lost its link to the registry tree and are represented in the set $\mathcal{O}$, where $\mathcal{O} \subset \mathcal{D}$. Orphans are denoted $\mathbb{U}_o$ and represent both orphan keys $\mathbb{R}_o \in \mathbb{U}_o$ and values $\mathbb{V}_o \in \mathbb{U}_o$.

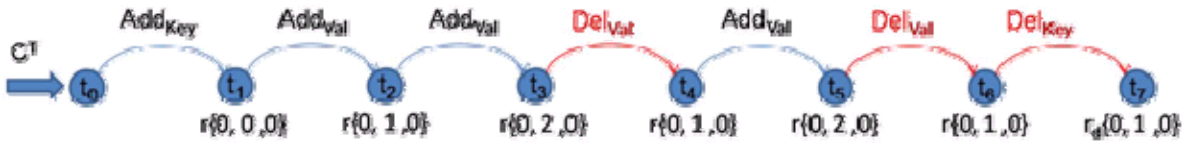


Figure 7: Sample Computation

2.3 Problem Statement

Figure 2 illustrates one possible computation in the model T for a registry key r. There are eight states starting with the key creation and ending with its deletion. During the process three values were added (at t_1, t_2 and t_4) and two were deleted (at t_3 and t_5). Current registry recovery programs can retrieve the key r_n and its values, but cannot associate the deleted values to the key. These orphans seem unrelated to the key. Therefore the number of data structures associated with a deleted key depends on its state when it was deleted.

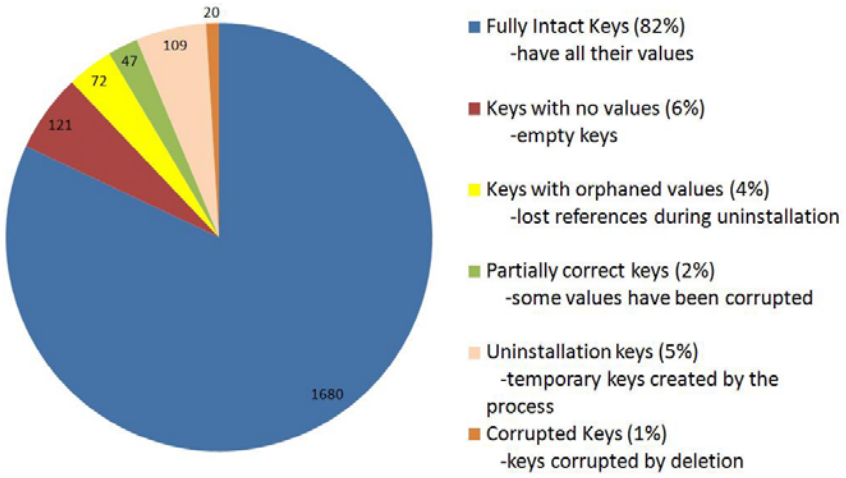


Figure 8: Different states of the retrieved keys (2049 total)

Let $P(v,r)$, where $v \subset \mathcal{O}$ and $r \subset \mathbb{R}$, be the probability of an orphan value v belonging to the registry key r. Our aim is to maximise P with respect to all orphans $v \in \mathcal{O}$.

The further in the past a value is deleted the less likely it can be retrieved (Kahvedžić, D. and Kechadi, T. (2009)). This is due to the reallocation of deallocated space by the registry manager (Russinovich. M. (2009)). However a large number of orphans can be created during the uninstallation of software or if a large number of values were recently deleted.

As an illustration, registry keys belonging to the media platform Real Player (RealNetworks. (2009)) were retrieved after the software's uninstaller was performed. To minimise possible orphan creation as a normal operation of the software, Real Player was uninstalled directly after it was installed. Even in this best case scenario, where the time range t of the computation is kept deliberately very low, 20%

of the keys were modified and 15% of the values-key relations were lost in the process. The values could still be retrieved but the link to the keys that they belonged to was removed.

Figure 3 illustrates the different states of the retrieved keys for the Real Player uninstallation process. Of particular interest are the 72 keys that have had their values deleted prior to the key itself. The values of these keys can be found in the deallocated space but could not be associated to the keys. The rest of this paper will detail our methodology in extracting these associations and linking the orphans to the most likely key that may have contained them it.

3. APPROACH AND VALIDATION TEST SETUP

We use a data mining approach to classify the orphan values to their most likely keys. We are able to process the values and other structures that are retrieved from the unallocated spaces with high accuracy. The stages of the mining process are described in Han, J. and Kamber, M. (2006). From this point of view, the registry retrieval processes is the first step; Data Acquisition. The next sections describe subsequent stages, namely Attribute Selection and Key Associations.

The attributes are evaluated and validated using sample test systems. A number of different registries were extracted from a variety of systems to test out the attributes. All registries were from Windows XP SP2 Operating System computers. They ranged from registries of three months usage to registries taken from systems that were in use for a number of years. The Software and the `ntuser.dat' hives are the most active hives and contain the most useful information for a forensic investigation. As such we will concentrate on them in our validation tests.

The summary of the various registries and the number of keys and values that they contain are summarized in Table 1. Security keys, although important for forensics, account for less than 0.05% of the total data structures. As such we will concentrate on the key and value structures only for the rest of the paper.

Details of Hives Used		
Hives	**# of Keys**	**# of Values**
Software Hives	285378	426929
System Hives	33226	95832
Security Hives	625	622
SAM Hives	234	260
ntuser.dat Hives (38 users)	53723	200130

Table 3: Summary of the Hives and their Combined Keys and Values

4. ATTRIBUTE SELECTION

In this section we describe some of the features of the way keys and values are stored in the registry. These features will define attributes describing the value and key data structures. We use classifiers to predict which orphan values belong to which key. The attributes are based on a number of observations in the functioning of the registry. The observations are formalised as attributes and validated through experimentation in the following sections.

4.1 Value-Key Position Relation

The position of the value data structure with respect to its key depends on the allocation strategy of the registry management system; the Configuration Manager (CM). Microsoft Windows documentation describes the allocation strategy of the CM when new keys are created (Russinovich. M. (2009)). The CM would first find the parent key data structure and then search for a block that is big enough to store the new key. If there is no block that satisfies the request then a new space is allocated at the end

of the hive file. It has been found that the hive files contain much more smaller free fragments than large ones (Kahvedžić, D. and Kechadi, T. (2009)) and that some key values are created as soon as the key is created (Morgan, T. D. (2008)). Therefore it is likely that when a new key is created the CM would not find a big enough fragment and allocate new space. At least some values therefore are found close to, if not directly after, its key data structure.

Formally, we consider a registry hive as a list of data structures with the physical position of the data structure within the hive denoted as $Pos(u)$, where $u \in \mathbb{U}$. We also define a binary operator, $Pos(u) \twoheadrightarrow Pos(u_1)$, which states that the data structure $Pos(u_1)$ is stored later in the hive file than $Pos(u)$. The operator $\twoheadleftarrow$ is the opposite. This attribute states that $P(r,v) \succ P(r,v_1)$ where $Pos(r) \twoheadrightarrow Pos(v) \twoheadrightarrow Pos(v_1)$. Namely, the closer in the hive file a value is found after a key, the more likely that the value belongs to that key. In this case the value v is more likely to belong to r than v_1.

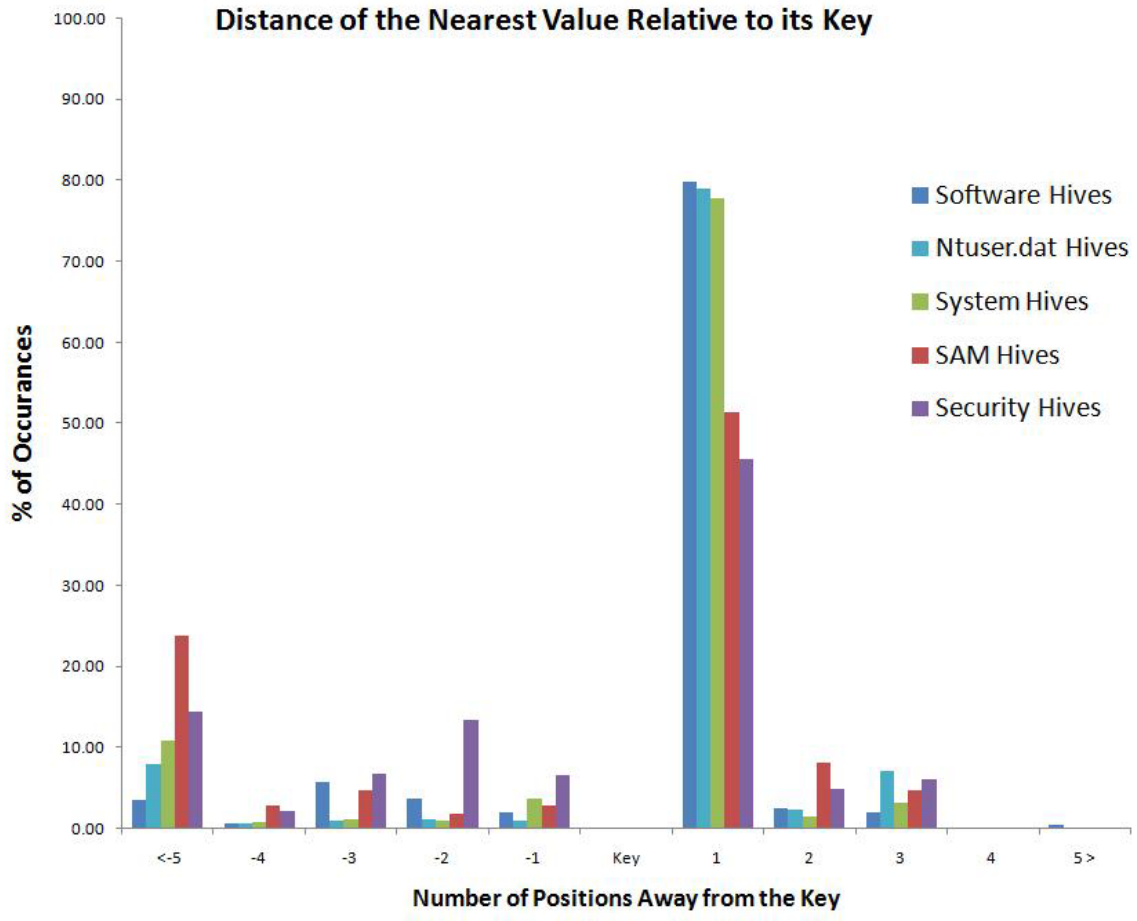


Figure 9: Relative Positions of Values to the Keys they Belong to

Figure 4 shows the respective distances of a key's closest value data structure relative to the position of the key in the hive in our test corpus. It is clear that in the majority of cases, the first value encountered after the key in the hive belongs to that key. Hives with a high amount of activity, such as the user hive (`ntuser.dat') or the `SOFTWARE' hive, tend to have more keys that follow the above heuristic. While hives with a lower activity, such as `SAM' and `SECURITY', tend not to follow the

heuristic as much as the active ones.

Value-Value Position Relation

Similar to the key-value position relation, if one key has a large number of values, the CM, if possible, would allocate space for all of them in the same place in the hive. MRU lists for example often have a large number of values with their data structures found one after another in the hive. Similarly, if an application creates a number of values at installation, the space is allocated as contiguously as possible and as such the values can be found grouped together. In conjunction with the first attribute, this attribute states that if the first key is found then the orphans found in its group are likely to belong to that key as well. Figure 5 illustrates this case.

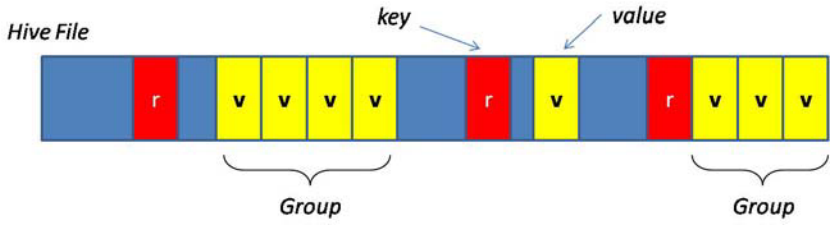


Figure 10: Group Allocation Strategy

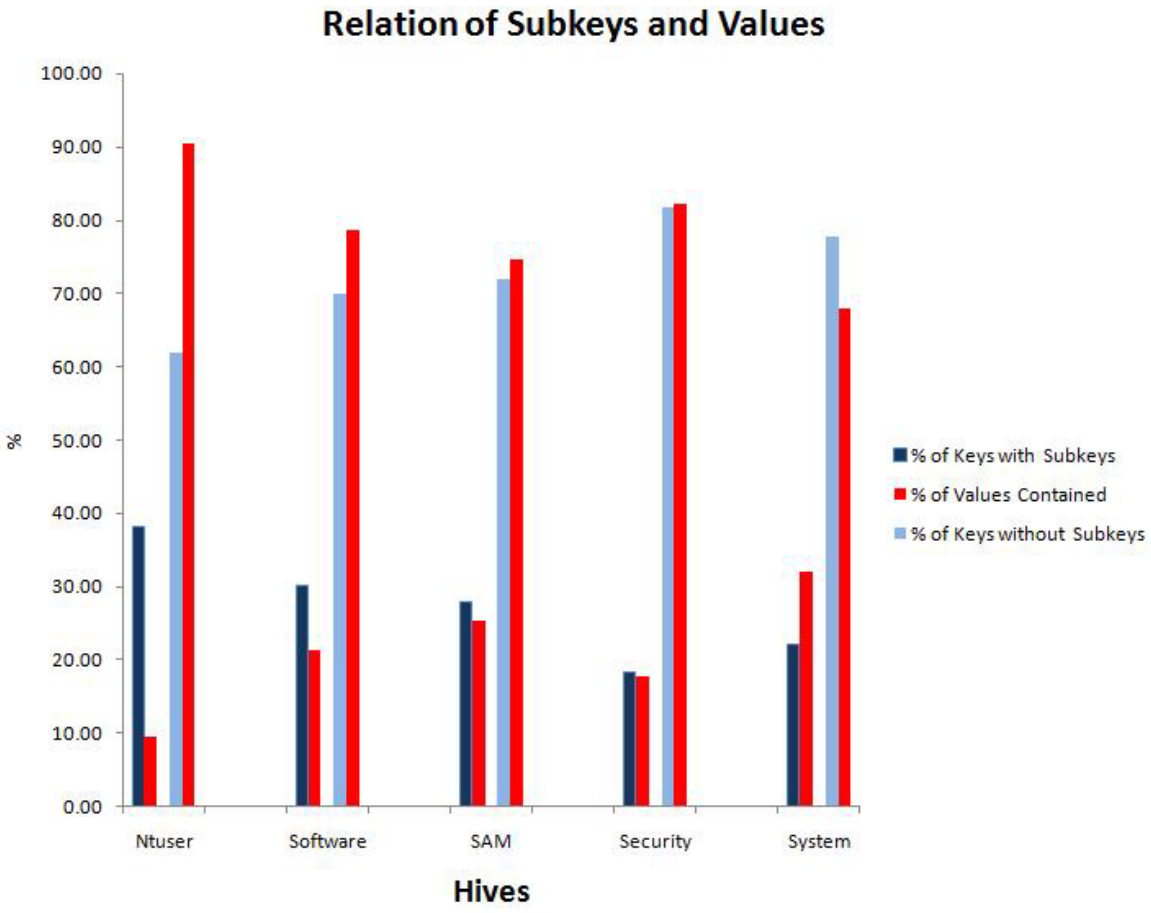


Figure 11: Internal/Leaf Key to Value Relations

Formally, we define a group $g \in G$ to be a set of orphans $g = \{v_0, \ldots, v_n\}$ where $v_i \in \mathbb{O}$ and there does not exist a situation where $Pos(v_0) \leftarrow Pos(r) \leftarrow Pos(v_n)$. Expansion of the group is also

stopped if a value with an identical name is found.

Similarly, $P(r,g)$ is the probability of all orphans in g belonging to the key r. Through experimentation, we have found that a group with two members is not likely to belong to the same key than any group with a higher number.

4.3 Subkeys Number Relation

In the tree like organisation of the registry, we use the term *leaf* to define those keys that do not have any subkeys, ie. $r(|R| = (0,v,s)$. These keys, found at the bottom of the registry hierarchy tend to store more values than those keys found higher up in the registry. There is not an enforced policy on how programs using the registry are meant to store its registry values but Microsoft recommends software vendors to follow the "*HKCU\Software\Vendor\Program\Version"* organisation for storing keys in the registry (Honeycutt, J. (2002)). This results in the leaf keys storing meaningful information such as values while the internal keys are used for organisational purposes.

In the formal model, the pattern can be described as, $P(r,v) > P(r_1,v)$ where $r(|k|) < r1(|k|)$. Namely, the probability of an orphan value belonging to key *r* is increased if the key contains a low number of subkeys. For simplicity, in our classifier we let $r(|k|) = 0$ and assume that if a key has *any* subkeys then they do not have any values.

Figure 6 illustrates the relationship between the number of values and the position of keys in the registry hierarchy of our test corpus. In the majority of cases, a key that does not have values would have subkeys. The user hives in particular contain the majority of values in the leaf keys rather than the internal keys.

4.4 Value List Presence Relation

Once a single value is identified, the value list of that key can also be found. A value list is a peripheral data structure and stores references to all the values of that key. The structure does not have an identifier and cannot be easily recognised when retrieved. However all the unidentified structures can be searched to see if any contain the address for the orphan value. Any unidentified structure that contains this address is likely to be a value list. The value list can be as small as 8 bytes. Therefore, the list, in contrast to values, can be found far away from its key.

Amount of Value List Slack						
Hive	**# of Keys**	**With Slack**	**Without Slack**	**Slack (bytes)**	**# of Pointers**	**Pointers to Hidden Keys**
Software	43,212	7,818	30,529	33,060	304	7
System	9,140	5,572	2,617	24,156	255	0
Security	208	1	206	8	1	1
SAM	75	8	66	32	0	0
ntuser.dat	533	148	179	592	7	0

Table 4: Amount of Value List slack

In addition, the Value List continuously adds and deletes offsets of created and deleted values (Morgan, T. D. (2008)). Under certain conditions, offsets to deleted values are still present at the end of the list data structure in an area called Value List slack. Table 2 shows the amount of slack in a sample of hives in the test corpus. Note that some keys do not contain any value lists at all. As seen in the table, many keys do have value list slack and do contain many offsets that appear to be valid. Most

of the offsets, however, are copies of active keys and do not point to hidden keys. Although the presence of any useful pointers is not common, the slack should not be disregarded in the process for correlating orphans.

4.5 Value size

Although the values in the registry have a variable structure, some valuable forensic keys have a predictable size that can be used to identify them if they are orphans. In particular, features of the Windows Operating System registry keys are known and can be used to identify them if they are found as orphans (Rubenking. N. J. (2009)). The "HKLM\Software\Microsoft\Windows\CurrentVersion\App Management\ARPCache*AppName*" key, for example, stores information of each application displayed by the control panel of Windows. Each application has a '*SlowInfoCache*' value which includes important information on the application's usage frequency, install size and last used time. The values have a constant size of 552 bytes, the specific structure of which is known (Rubenking. N. J. (2009)). As well as the aforementioned data, the value also contains the name of the program it references and can be used to relate the value to the relevant key.

4.6 Specific Value Content

Similar to using the value size as an attribute, some important keys contain values that store a predictable type of content. The most important of these keys are the MRU (Most Recently Used) lists which store records of most recently opened files. MRUs are widespread in the registry and can provide vital clues to user activity (Farmer, D.J. and Burlington V. (2009)). Windows does not enforce a standard way of storing values in all MRU lists. Therefore it is possible to extract and identify the relation between an orphan MRU entry and its parent key. The identifying features of these values include the information type (binary or ASCII), the naming conventions, the type of data and the recording policy. A few important MRUs and the identifying features of their values are documented in Table 3.

Identifying Features of MRUs			
1: \Windows\CurrentVersion\Explorer\ComDlg32\OpenSaveMRU			
Type: Last Saved File	Name: Letters (a-j)	Scheme: ASCII	Data: Filepath
2: \Windows\CurrentVersion\Explorer\ComDlg32\OpenSaveMRU\'extension'			
Type: Last Saved File by Extension	Name: Letters (a-j)	Scheme: ASCII	Data: Filepath with specific extension
3: \Windows\CurrentVersion\Explorer\ComDlg32\LastVisitedMRU			
Type: Last Opened File	Name: Letters (a-y)	Scheme: Binary	Data: Filepath
4: \Windows\CurrentVersion\Applets\Paint\Recent			
Type: MSPaint Recent Files	Name: 'File(n)' where n>0	Scheme: ASCII	Data: Filepath of 'Image' file
5: \MediaPlayer\Player\RecentFileList			
Type: Media Player Recent Files List	Name: 'File(n)' where n>0	Scheme: ASCII	Data: Filepath of 'Multimedia' file
6: \SearchAssistant\ACMru\5603			
Type: XP Searches	Name: 3 Digit Number (001)	Scheme: ASCII	Data: User search query strings
7: \Windows\CurrentVersion\Explorer\RecentDocs			
XP Recently Opened Docs	Name: 1 Digit Number (0-9)	Scheme: Binary	Data: Filepath of opened docs
8: HKCU\Software\Windows\Microsoft\Office\12.0\Word\File MRU			
Word Recently Opened Docs	Name: 'Item (n)' where n>0	Scheme: ASCII	Data: '[F0000000][17characters]Filepath'
9: HKCU\Software\Windows\Microsoft\Office\11.0\PowerPoint\Recent File List			
Powerpoint Recently Opened Docs	Name: 'File(n)' where n>0	Scheme: ASCII	Data: Filepath
10: HKCU\Software\Windows\Microsoft\Office\12.0\PowerPoint\File MRU			
Powerpoint Recently Opened Docs	Name: 'Item (n)' where n>0	Scheme: ASCII	Data: '[F0000000][17characters]Filepath'

Table 5: Identifying features of MRU Values

5. DATA MINING STAGE

In this section we will use the attributes described above to create a decision tree classifier for deleted values. We use Decision Trees as an initial mining technique. Other algorithms such as association rules, clustering algorithms and neural networks will be applied in future work. The situation

described in Section 2.3 revealed that 72 deleted keys had all their values orphaned prior to being deleted. In this section we will use a decision tree classifier, illustrated in Figure 7, to re-attach the orphans to their keys.

The decision tree classifier described here represents our initial work in associating orphan values to keys. The classifier is based on structural attributes described above and exploits the manner in which the Configuration Manager allocates new structures. It does not take into account any of the content that the various data structures may hold and does not attempt to do any content analysis or similarity matching as described in the latter parts of Section 4. A clustering approach would be more suited for that stage of mining and is left for future work.

5.1 Results

As previously stated, 72 keys in the Real Player uninstallation had their values orphaned, totalling 146 values. 143 (98%) of these values were retrieved from deallocated space while the remaining 3 values were overwritten and could not be retrieved. The aim of the decision tree is to associate the 143 values back to their keys. The total number of orphans retrieved was 360. The decision tree should avoid these orphans and minimise false positives.

The classifier first attempts to find the first value data structure after the key's data structure in the hive. The classifier associated 68 orphans to 68 keys in this way. However, it couldn't associate values to 4 keys. Of the 68 values, 63 of them were accurate and did indeed belong to the key prior to the uninstallation. 5 of the orphans were incorrectly associated. The precision rate is therefore 93%.

The classifier extends the association criteria by finding any groups that the first orphan belongs to. The groups are filtered out if they only contain two elements as we have found that $P(r, g)$ is maximised in larger groups. The classifier correctly associated 77 values to their corresponding keys and misclassified 11. The precision rate therefore is reduced to 86%.

Before the orphans are associated to the keys, the classifier attempts to find the value list storing the offset to the value. If the value list is found, all of the values that it references will be associated to the key. As described in Section 4.4, the value list may not be found even if it did at one point list the orphan. Of the 68 values found, 32 of them found value lists, 36 of them did not. 8 of the found value lists were incorrect. The precision rate is therefore 78% for value list searches. The found value lists classified 2 new orphans correctly and 2 incorrectly. As described above, the vast majority of value lists 97% only contained a single offset.

In total 79 values were correctly associated by using the closest value and value list search heuristics, 13 were misclassified. Therefore, 55% of the orphans were associated correctly with a reliability of 84%.

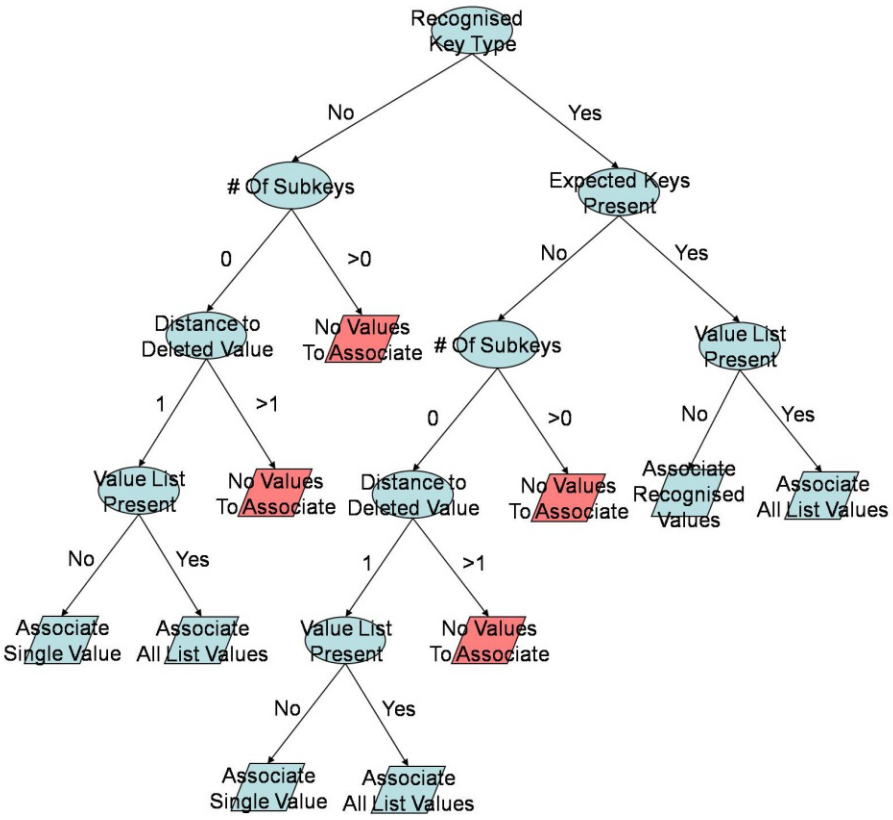


Figure 12: Decision Tree Classifier

6. SUMMARY AND CONCLUSION

In this paper we have illustrated how values, once deleted from the registry, can be orphaned from their parent keys and cannot be easily re-associated to them. We have explored the effects of the Real Player uninstallation process on the registry and demonstrated how the uninstallation created orphans. The resulting orphans can make it difficult for the investigator to know the context of the retrieved registry entry and lay doubt in legal arguments. We have laid the formal theory and the ground work for the application of data mining techniques in retrieving and correlating these deleted data structures to the points in the registry that they were found.

The formal theory is presented in the form of Finite State Machines. Using this model we have demonstrated that the amount of information retrieved from a deleted key depends on the state when it was deleted. However, values that once belonged to the key but that were deleted prior to the key can also be retrieved. The link between the key and its former value is not preserved and the association cannot be made trivially. The unlinked data structures are termed orphans and require special processing to reattach them to their parent keys.

Presented in this paper are a preliminary set of patterns in the organisation of the registry data structures. The patterns are encoded as attributes that can be used by classifiers to relate orphaned values to the most likely owner keys. Tests were carried out on sets of real world registries to analyse the validity of the attributes presented. A decision tree classifier was described and used as a preliminary Data Mining technique to associate orphan values to the most likely parent keys.

7. FUTURE WORK

The features presented here constitute the preliminary attributes discovered while manually reconstructing the registry keys. A more formal and in-depth feature extraction methodology will be developed to extract ever more accurate features to discriminate between keys. Future work will entail

expanding on the data mining stages to further refine the accuracy of the classifier. Clustering of the values, based on the value's name and content, will be used to further associate the orphans with more accuracy. Data pre-processing stages will be expanded to clean and filter out any outliers and data structures that have been corrupted, for example if they are partially overwritten.

The features presented in the classifier in Figure 7 are binary. A more accurate classifier would discriminate between continuous attributes. For example, in the subkey-number relation, in Section 4.3, if a key is found to contain subkeys it is assumed not to have any values. This feature does not take into account the position of the key in the global tree hierarchy. A key may be deep in the tree hierarchy and may contain both subkeys and values. The position of the key may dictate the probability of the key holding values even if it has subkeys.

Content features are not used in the above decision classifier and is also left for future work. The content of the deleted keys can be used to find similar active keys and predict which values are missing from the deleted key based on the properties of the similar active one. This type of mining process is more suited to clustering. Outlier analysis in particular can be used to determine not which values belong to keys, but which values do not belong to them. The removal of these outliers can eliminate clutter to the investigator.

Following hidden pointers to deleted data structure is prone to error. It is possible that the structure is a newer data structure belonging to some other key which was subsequently deleted. Although unlikely it cannot be disregarded, extra processing and checks are required to disprove this possibility.

The Windows Vista operating system subtly changes the manner in which registry keys are stored (SWGDE (2008), Hargreaves, C. et al. (2008)). The MRU keys in particular are store in binary more frequently and stored more information that the corresponding keys in the XP operating system. Future work on the classifier will include a method for identifying these MRU keys in addition to their XP counterparts.

REFERENCES

Carvey, H. (2005) “The Windows Registry as a forensic resource”, Digital Investigation, Vol 2 (Issue 3) p201–205, 2005.

B. D. (2009) ‘Registry File Format’, http://home.eunet.no/pnordahl/ntpasswd/WinReg.txt. visited Feb 2009.

Farmer, D.J. and Burlington V. (2009) ‘A Forensic Analysis of the Windows Registry’, http://eptuners.com/forensics/Registry_Forensics.pdf, visited Jan 2009.

Han, J. and Kamber, M. (2006) ‘Data Mining: Concepts and Techniques’, Morgan Kaufmann, 2nd edition, 2006.

Hargreaves, C. et al. (2008), “Windows Vista and Digital Investigations”, Digital Investigation, Vol 5 (Issue 1), p34 – 48, 2008.

Honeycutt, J. (2002) ‘Microsoft Windows XP Registry Guide’, Microsoft Press, 2002.

JavaCoolSoftware. (2009) ‘MRU-Blaster’, http://www.javacoolsoftware.com/mrudownload.html, visited Jan 2009.

Kahvedžić, D. and Kechadi, T. (2008). ‘Extraction of User Activity through Comparison of Windows

Restore Points', SECAU08, 6th Australian Digital Forensics Conference, Dec 2008, Perth, Australia.

Kahvedžić, D. and Kechadi, T. (2008)ii. "Extraction and Catagorisation of User Activity from Windows Restore Points", JDFSL: Journal of Digital Forensics, Security and Law, Vol4 (Issue4) (to be published).

Kahvedžić, D. and Kechadi, T. (2009). 'On the Persistence of Deleted Windows Registry Data Structures', 24th Annual ACM Symposium on Applied Computing, March 2009, Hawaii, USA.

Morgan, T. D. (2008) 'Recovering Deleted Data From the Windows Registry', Digital Forensic Research Workshop, Aug 2008, Baltimore, USA.

RealNetworks. (2009). 'Real Player 11 Basic', http://europe.real.com/player/win/. visited Jan 2009.

Registry Hives. (2008) 'Forensicmatter.com: Registry Hives', http://www.forensicsmatter.com/registry hives.php, visited Feb 2009.

Rubenking. N. J. (2009) 'Unclean 2', http://www.pcmag.com/article2/0,1759,1159867,00.asp. visited Jan 2009.

Russinovich. M. (2009) 'Inside the registry', http://technet.microsoft.com/en-gb/library/cc750583.aspx. visited Jan 2009.

SWGDE (2009), 'Technical Notes on Microsoft Vista (submitted for review)', Scientific Working Group on Digital Evidence, http://www.swgde.org/documents.html, visited Feb 2009.

Wong, L. W. (2009) 'Forensic Analysis of the Windows Registry', http://www.forensicfocus.com/forensic-analysis-windows-registry, visited Jan 2009.

ACKNOWLEDGEMENTS

The authors would like to acknowledge "Higher Education Authority" (HEA) and "Irish Research Council for Science, Engineering and Technology" for providing funding that made this research possible.

Don't Touch That! and Other E-Discovery Issues

Linda Volonino
R. J. Wehle School of Business, Dept. of Information Systems
Canisius College, Buffalo, NY
volonino@canisius.edu

ABSTRACT

The ability to preserve and access electronically stored information (ESI) took on greater urgency when amendments to the Federal Rules of Civil Procedure went into effect in December 2006. These amendments, referred to as the *electronic discovery (e-discovery) amendments*, focus on the discovery phase of civil litigation, audits, or investigations. Discovery is the investigative phase of a legal case when opponents learn what evidence is available and how accessible it is. When ESI is the subject of discovery, it is called e-discovery. Recognizing that most business and personal records and communications are electronic, Judge Shira A. Scheindlin stated, "We used to say there's e-discovery as if it was a subset of all discovery. But now there's no other discovery." Computer forensics experts, given their expertise in identifying, acquiring, preserving, and searching ESI, can play a key role throughout the e-discovery process, if they choose to do so. They can also assist in the drafting of the e-discovery request, in preparing the response to such a request, and initiating a legal hold for evidence preservation. The objective of this paper is to provide an overview of the e-discovery amendments and case law, their impact on the duty to preserve and produce ESI, and the computer forensic work that can support the e-discovery process.

Keywords: Electronic discovery, litigation, preservation, Federal Rules of Civil Procedure

1. INTRODUCTION

In April 2006, the U.S. Supreme Court approved changes to the Federal Rules of Civil Procedure (FRCP) to bring the law into alignment with the most common type of evidence—electronic evidence (e-evidence). After Congress approved, the amended FRCP became law on December 1, 2006. These amended rules all aim at one issue—the discovery of *electronically stored information* (ESI). ESI used as evidence is known as electronic evidence, or e-evidence. What this means for companies and individuals is that e-discovery imposes an inescapable obligation to be ready and able produce all relevant ESI on demand.

Litigants and their lawyers can expect to face harsh consequences when requested ESI has been destroyed or made inaccessible. Destruction of evidence, which is called *spoliation*, is arguably the most damaging position a party can be in because a court may find it to be an obstruction of justice. To appreciate the risk, consider that obstruction of justice charges were the reasons for the demise of the major accounting firm Arthur Andersen and jail time for Martha Stewart.

2. MOTIVATION FOR THE E-DISCOVERY AMENDMENTS

The *Judicial Conference Committee on Rules of Practice and Pro*cedure (2005) identified three reasons for implementing the e-discovery amendments:

1. The volume of ESI created discovery issues that had not existed when legal cases dealt with only paper documents.
2. Unlike information memorialized on paper, ESI can be deleted or overwritten with or without the user's knowledge.

3. Unlike paper documents, ESI sometimes can be unintelligible if separated from the system in which it is created and stored.

Basically, the amendments acknowledge that the law had to change in order to keep up with technology. Few trial or corporate lawyers were prepared for this new job function, which largely remains true. As such, they rely on the expertise of those who understand ESI, search methods, and e-evidence investigation procedure.

3. PRESERVATION: TOUCH OR DON'T TOUCH?

A company's ESI has a dual nature in that it is both fragile and persistent. It is easily altered or destroyed when backup tapes are overwritten or corrupted. Yet it can also persist on employees' hard drives and digital devices. The mistake that many companies and their employees make is believing that they can "game" (e.g., outsmart or play dumb) the e-discovery process. That tactic is equivalent to playing the lethal game of Russian roulette.

Figure 1 contrasts differences in how paper and ESI are destroyed or altered and how they are preserved. Because ESI exists only on storage media that may be overwritten, corrupted, or otherwise be unreadable, proactive procedures are needed to preserve it. Without deliberate action to preserve ESI, the expectation is that it will be destroyed or altered. Courts have recognized the fragility/persistence paradox and the need for companies to take affirmative steps to preserve ESI, as shown in Figure 1. Judges do not tolerate ignorance of computer technology or improper handling of ESI--or attempts to use those excuses to defend the destruction of e-evidence.

	Affirmative Steps	Passivity
Paper	Destroy, alter	Preserve
ESI	Preserve	Destroy, alter

Figure 1. Differences in the preservation of paper and ESI.

Jeff Rothenberg, a senior computer scientist at RAND, captured the paradox by pointing out humorously that "digital information lasts forever, or five years – whichever comes first." What is not humorous is how employees' react when they are informed of the need to preserve their e-mail, documents, or memos related to anticipated or current litigation. Their immediate reaction is to delete such ESI despite the futility of those efforts and the risk of spoliation sanctions they create. Too many companies have relied, in effect, on directives such as "don't touch" to the employees or other data custodians. Computer forensics experts are often needed to help companies preserve ESI in a legally defensive manner.

4. E-DISCOVERY STEAMROLLS THE LITIGATION LANDSCAPE

The United States' FRCP govern the conduct of all lawsuits and other civil actions brought in Federal district courts (LII, 2006). Their e-discovery amendments dramatically increased the number of cases that involve ESI and its preservation. In effect, e-discovery rules have steamrolled the litigation landscape. Typically, lawyers and litigants are unprepared to comply with this type and volume of discovery and all its complexities. Two reasons account for most of this lack of preparedness.

1. Lawyers are not IT people. The huge majority of lawyers never had a course in IT or e-discovery in their law schools. E-evidence lives on in many places and forms that are tough to find, collect, store, and interpret without tech skills.
2. E-discovery must be addressed when a lawsuit is filed. That is, when litigation initiates so does e-discovery.

Comparing Figure 2 to Figure 3 shows how the discovery phase of litigation has changed. Prior to December 2006, discovery was an afterthought because most cases did not get to trial. As a result, cases were ending before discovery got started.

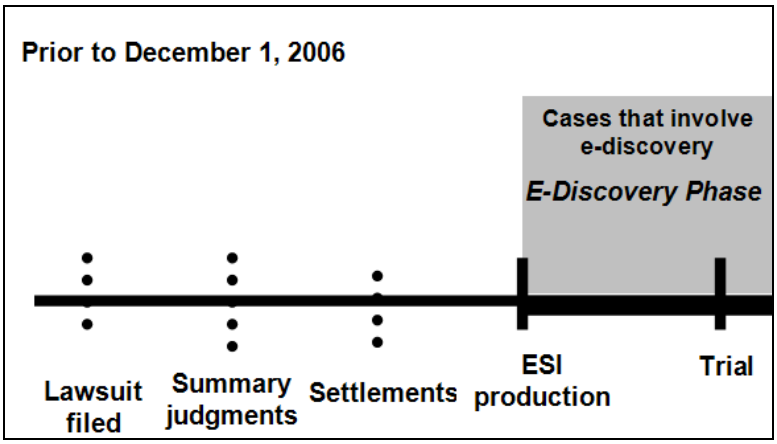


Figure 2. Cases involving e-discovery and ESI production prior to amended FRCP

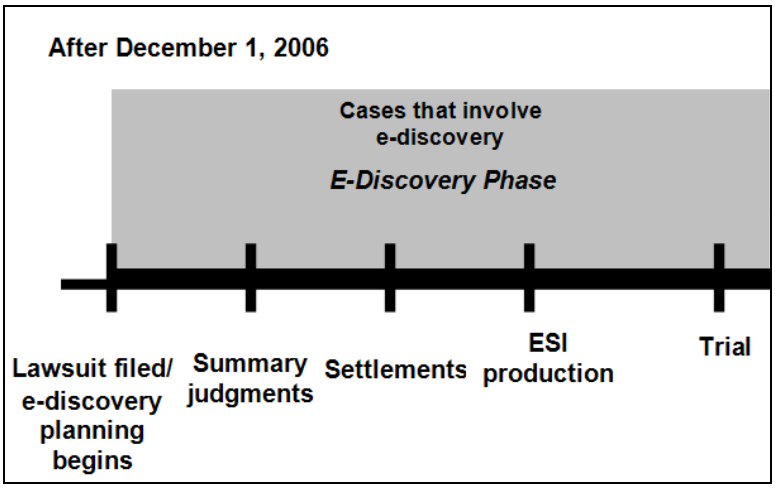


Figure 3. Cases involving e-discovery and ESI production prior to amended FRCP

Three other factors add to the magnitude of e-discovery and the increasing volume of potentially relevant ESI:

1. The rules apply to every type of litigation. Class action lawsuits, complex corporate fraud, and employment cases (e.g., discrimination, wrongful termination, and harassment) involve e-discovery. Government investigations of fraud or improper conduct invariably dig into e-mail, instant messages, and appointment calendars.
2. Companies with at least $1 billion in annual revenues are involved in an average of 147 lawsuits at any one time, while the corresponding number for companies with under $1 billion in revenues is thirty-seven.
3. Everything from terabyte-sized databases to text messages and tweets may be *discoverable* (subject to discovery).

All computer systems, digital devices, and anything with a flash drive used by businesses, government agencies, health care and education institutions, and individuals store electronic documents (word processing, spreadsheets, calendars, presentations, etc.) and other forms of ESI. Contact lists on iPhones, instant messages on Blackberries, posts on MySpace, and GPS and EZ-Pass records may be part of the ESI universe.

5. E-DISCOVERY RULES & TIMELINE

Additions and revisions were made to Federal Rules 16, 26, 33, 34, 37, and 45 and to Form 35. Form 35 standardizes discovery agreements to avoid delays and motion practice around discovery later on.

The amendments actually introduced the term "electronically stored information" in Rules 26(a)(1), 33, and 34, to acknowledge that ESI is discoverable. This phrase is intended to be broad enough to cover all current types of computer-based information, and flexible enough to encompass future changes and technological developments.

Amended rules, their requirements, and considerations that highlight the importance of managing ESI throughout its lifecycle—from creation to retention or destruction—are listed in Table 1.

TABLE 1. Amended Federal Rules Related to the Management of ESI
source: U.S. Courts, *http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf*

Federal Rule	Requirements	Questions and Considerations
Rule 26(a)(l)	Requires an exhaustive search for all ESI, including e-mail that is "in the possession, custody, or control of the party." It must be disclosed "without awaiting a discovery request." The only exception to the disclosure rule is privileged information. The phrase "in the possession, custody, or control of the party" has not been interpreted by the courts. Requires presenting a copy or description by category and location of all ESI that the disclosing party may use to support its claims or defences.	Can an employee's laptop or BlackBerry device be considered *under the control of the company*, even if it is in a remote location? Companies should consider keeping a centralized copy or backup of everything, including e-mail that might be stored on a remote device.
Rule 26(b)(2)(B)	Even if one party identifies information "as not reasonably accessible because of undue burden or cost," its description, category, and location must be disclosed. This means that the information must be identified, even if it is difficult to retrieve.	Delay in producing requested or subpoenaed ESI is not an option. Nothing about the ESI can be left out and opposing counsel can challenge the claim that the ESI is not reasonably accessible.
Rule 26(f)(3)	It is expected that most documents will need to be produced in their original form, although the companies can discuss the form in which data is to be produced.	If requested, ESI must be submitted in readable electronic form and metadata must be preserved to facilitate searching potential e-evidence.
Rule 16(b)	The search for relevant ESI must be done at the beginning of a legal case and no later than the first pre-trial discovery-related meeting, which is required to be within 99 days of the filing of the legal action.	Extra caution must be taken with any information that could be used as evidence. A best practice is to place a "litigation hold" on documents and e-mail relevant to a case.

Rule 34(a)	Specifies that ESI is subject to discovery. This rule sets forth a clear duty to preserve and produce relevant electronic documents, databases, and communication once a company has notice of impending litigation.	When faced with pending or impending litigation or a regulatory investigation, a company must have a response plan to find and produce pertinent ESI.
Rule 37(e) *"Safe harbor" rule*	Provides that courts may not sanction parties for information "lost as a result of routine, good faith operation of an electronic information system." To come within the protection of Rule 37(e), a company would have to show that: (1) the information was lost due to the routine operation of an information system (IS), and (2) the routine operation of the IS was operated in good faith.	Rule 37(e) does not provide any protection if the information is lost outside the routine operation of an IS. Even if good faith does exist, a court may find that "exceptional circumstances" trump the responding party's good faith such that the imposition of some sanction may be justified.

The rules are mapped onto a timeline in Figure 4. While their purpose is to provide early structure, uniformity and predictability to the litigation process, the reality is that from Day 1 of a lawsuit, a party must be ready to start evaluating with IT, legal, and perhaps computer forensics experts where it stands in terms of its ESI.

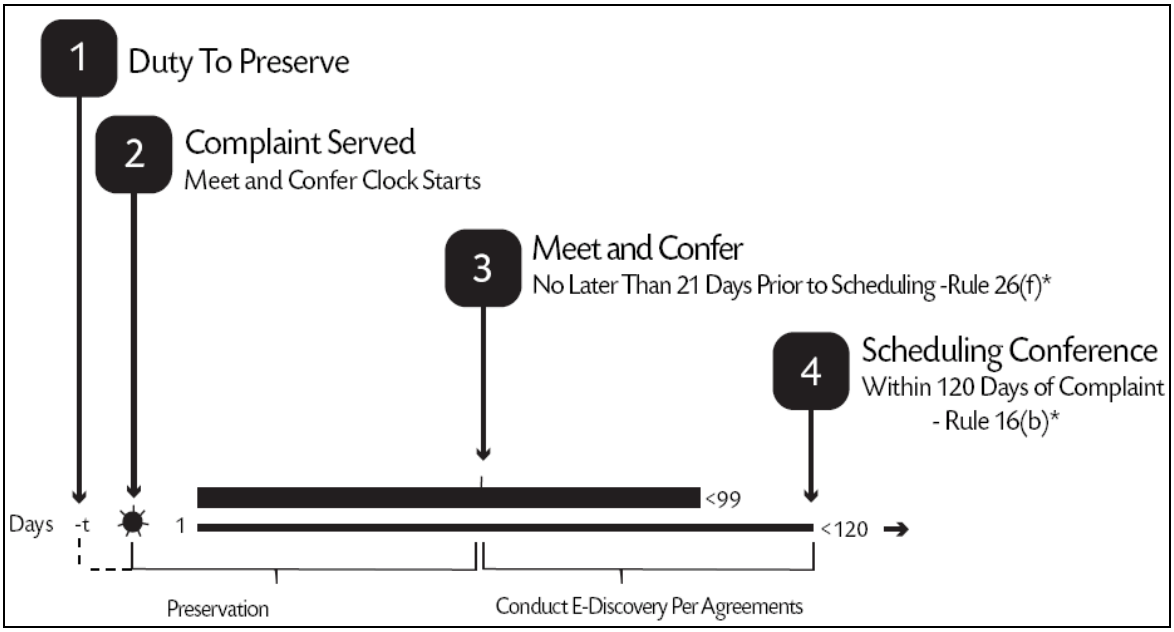


Figure 4. Timeline of the litigation process created by the e-discovery rules.

Time minus zero: **Duty to preserve.** You need to take active and timely measures to prevent the destruction or alteration of what might be relevant e-evidence. This duty generally begins when a legal action is reasonably anticipated. That's a tough duty to comply with because the scope of what needs to be preserved and as of when are not clear. Regardless, the courts consistently require counsel to be aware of these issues, and to have guided their clients appropriately in regard to the duty to preserve ESI.

Day 1**: Complaint served.** When the lawsuit is filed and complaint is served on the defendant, it starts a clock that counts off days.

By Day 99: **Meet and Confer conference.** The meet and confer conference (more simply referred to

as the *meet and confer*) is also a duty. Litigants must participate in a meet and confer conference to negotiate an e-discovery plan. The list of topics to negotiate include the following:

- Any issues relating to preserving discoverable ESI.
- Any issues relating to search, disclosure, or discovery of ESI.
- Format in which ESI should be produced.
- Scope of ESI holdings
- Estimated costs in terms of difficulty, risk, time, and money of producing the ESI.

Agreements made at the meet and confer and that are listed in Form 35 need to be conducted. Form 35 was amended by the new FRCP to include a report to the court about any agreements that the parties have reached.

By Day 120: **Scheduling conference.** A scheduling conference is a hearing attended by the prosecuting attorneys, defendants, defendant's attorneys, and the judge to schedule certain dates and deadlines for the case. This event is generally the first time the litigants and their attorneys come before the Court.

By forcing these events early on in a case, by way of the FRCP amendments and case law, parties really have no choice but to be ready to move forward with e-discovery at the start of a case. An alarming example of the potential magnitude of e-discovery and the consequences of not fulfilling e-discovery duties is the ongoing case of *AMD vs Intel,* which is discussed in §6.

6. E-DISCOVERY CASE: AMD V INTEL (2005 - 2010)

In July 2005, Advanced Micro Devices (AMD) brought a lawsuit against its arch-rival Intel for alleged anticompetitive practices in the chip-maker market. In charges filed in federal court, AMD says Intel used its huge size to coerce customers into shunning AMD's chips. Intel had about 80 percent of the market for PC processors while AMD had about 20 percent. The long-running *AMD v. Intel* case is scheduled for trial in February 2010. Both parties recognized that they faced the largest e-discovery ever. Estimates of production were roughly "a pile 137 miles high."

Intel, the world's largest chipmaker with 99,000 employees worldwide and an e-mail load of 3 million messages per day, was ordered by the court in March 2007 to recover 1,000 lost e-mails that it was required to keep. The court gave Intel 30 days. When the company was unable to find them, it pleaded with the court for an extension. The court granted a 10-day extension to come up with a report to AMD on how the e-mail search was progressing or whether the corporation will be able to produce the e-mails at all.

Intel's e-mail system running on Microsoft Exchange servers is automated to expunge e-mail sent or received by employees every 35 days. For senior executives, e-mail is purged every 60 days. Some of the e-mails may be recoverable from backup tapes or by employee-initiated backup. However, Intel used non-indexed backup tapes designed for disaster recovery, but not e-discovery. Trying to find all e-mail messages with specific keywords is tedious and requires a staggering amount of time. One problem is that individual backup tapes have to be mounted one at a time, and then have their contents restored to get them into shape to be examined.

With an understanding of the e-discovery rules and timeline--and a look at e-discovery chaos, we examine the widely used and respected reference model for e-discovery.

7. ELECTRONIC DISCOVERY REFERENCE MODEL (EDRM)

EDRM is an organization composed of numerous working groups that develop guidelines and standards for e-discovery; and help reduce the cost, time and manual work associated with e-discovery. Their widely used Electronic Discovery Reference Model (EDRM), downloadable from http://www.edrm.net/, is shown in Figure 5. For excellent and the latest details on each stage in the

EDRM, visit their interactive website.

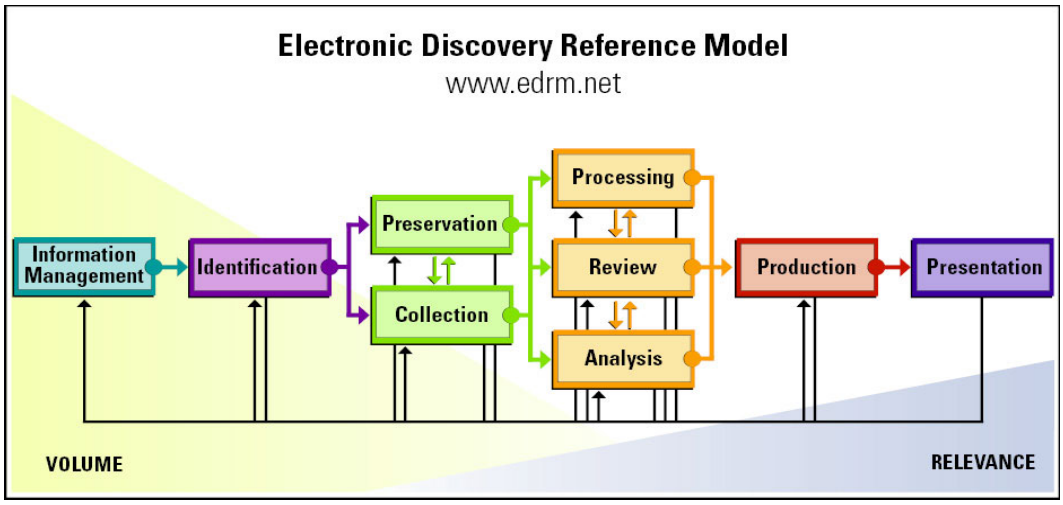


Figure 5. The Electronic Discovery Reference Model (EDRM). *Source:* http://www.edrm.net/

Figure 6 integrates the EDRM with the e-discovery timeline to create the EDRM document production model.

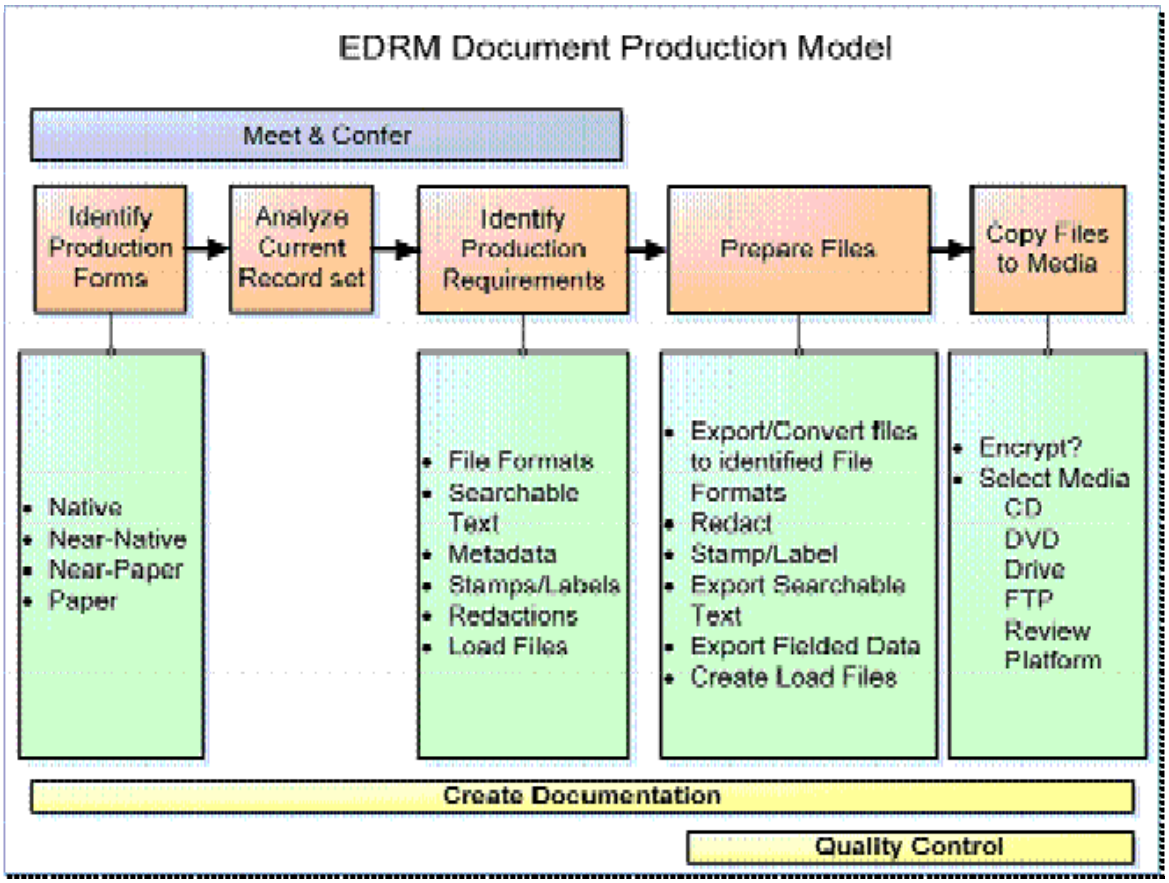


Figure 6. EDRM Document Production Model. Source: http://edrm.net/blog/archives/146

The EDRM Document Production Model shows areas where computer forensics experts can help with e-discovery duties and procedures. One crucial function is to keep the IT department and anyone else from touching data on servers and PCs unless and until the ESI has been preserved and/or the corporate legal department confirms that it is safe to do so. As a member of an e-discovery response team, computer forensics professionals can guide decisions regarding native or non-native formats, preservation, effective keyword searches, and the production of ESI.

8. AUTHOR INFORMATION

Linda Volonino, Ph.D., CISSP. ACFE is a professor of Information Systems at Canisius College. She has published six books (Prentice-Hall and Wiley publishers) on IT, information security, and computer forensics. Currently, she's writing *E-Discovery For Dummies* (Nov. 2009) and is a computer forensics investigator with Robson Forensic.

9. REFERENCES

Amended Federal Rule of Civil Procedure (2008). http://www.uscourts.gov/rules/CV2008.pdf

EDRM. http://edrm.net/

EDRM News. http://edrm.net/blog/archives/146

*Judicial Conference Committee On Rules of Practice and Pro*cedure (2005). Committee on Rules of Practice & Procedure of the Judicial Conference of the United States, Summary of the Report of the Judicial Conference.

Legal Information Institute (LII, 2006), Cornell Law School, *http://www.law.cornell.edu/rules/frcp*

U.S. Courts, Federal Rulemaking, http://www.uscourts.gov/rules/index.html

Why are we not getting better at Data Disposal?

Prof. Andy Jones [1 2]
[1]Head of Information Security Research,
Centre for Information & Security Systems Research, BT
[2]Adjunct Professor, Edith Cowan University

ABSTRACT

This paper describes two sets of research, the first of which has been carried out over a period of four years into the levels and types of information that can be found on computer hard disks that are offered for sale on the second hand market. The second research project examined a number of second-hand hand held devices including PDAs, mobile (cell) phones and RIM Blackberry devices. The primary purpose of this research was to gain an understanding of the reasons for the failure to effectively remove potentially sensitive information from the disks and handheld devices. Other objectives included determining whether there were regional variations in the results and whether there had been any changes in the results detected over time. From the research it was possible to develop advice on the measures that could be adopted to reduce the level of data being inadvertently released into the public domain.

Keywords: Disk Study, mobile device, security data destruction, disk erasure

1. BACKGROUND

In the last few years there has been an ever increasing level of media attention on high profile data losses such as the Feb 2009 Kaiser Permanente[3] loss of a data file containing details of names, addresses, dates of birth and Social Security numbers that resulted in 30,000 California employees having to be notified of the release of personal information, the Feb 2009 Parkland Memorial Hospital[4] loss of a laptop computer that may have contained the names, birthdates and Social Security numbers of 9,300 employees and the UK Ministry of Defence (MoD) [5] loss of a portable computer drive containing the names, addresses, passport numbers, dates of birth and driving licence details of around 100,000 serving personnel across the Army, Royal Navy and RAF, plus their next-of-kin details and the details of 600,000 potential services applicants and the names of their referees.

These incidents have served to highlight the damage that can be caused as a result of a data security breach, whether malicious or accidental. Unfortunately, the high profile nature of these incidents has, in some ways, diverted attention from a number of the underlying issues. At the same time that these high profile data breaches are occurring, a huge number of less prominent or significant losses are not highlighted in the press. Some of the reasons for this are that the losses go undetected or unreported and also that the individual cases are not of themselves, newsworthy.

It is increasingly clear that the levels and types of information that are given away on a daily basis when equipment that contains digital storage media such as computers, PDAs, mobile phones etc. is

[3] Michael Barkoviak, Daily Tech, Kaiser: Employee Data Breached, Identity Theft Reported, 8 Feb 2009, http://www.dailytech.com/Kaiser+Employee+Data+Breached+Identity+Theft+Reported/article14196.htm
[4] Sherry Jacobson, Dallas News Laptop theft at Parkland Memorial Hospital could imperil employee information, 09 Feb 2009, http://www.dallasnews.com/sharedcontent/dws/news/localnews/stories/021009dnmetparkland.3574199.html
[5] Rosamond Hutt, The Independent, MoD stunned by massive data loss**, 10 0ct 2008,** http://www.independent.co.uk/news/uk/home-news/mod-stunned-by-massive-data-loss-957099.html

disposed of at the end of its useful life is one of the primary causes of the ongoing issues.

2. FINDINGS

Research has been carried out during the last four years to determine the levels and types of information that individual home users and organisations have inadvertently made available to the public when they have disposed of computers and hand held devices. The hand held devices looked at in the research included devices such as mobile (cell) phones, RIM Blackberries and PDAs that contain either disks or solid state storage. This research has been undertaken by a partnership between industry and academia that has been led by British Telecommunications with academic partners at Edith Cowan University in Perth, Australia, the University of Glamorgan in Wales and Longwood University in Virginia, USA.

Over the four year period of the study more than 1000 disks have now been examined together with 160 hand-held devices. The results of the research have provided an insight into the consistently poor level of protection that organisations and individuals give to the information that is contained within them when they dispose of these types of equipment.

The purpose of the research was to obtain a better understanding of the volumes and types of information that were left, in an easily recoverable form, on magnetic media that was offered for sale on the second hand market.

Prior to this research, there had only been limited journalistic reporting on the problems that had been associated with the incorrect disposal of data. One of the earliest known reports was from 1993 when there was an article in the Canadian Globe[6] relating to the discovery of a computer hard disk that was reported to contain information on the employees of a small company. Another article in 2000 in the UK Daily Express reported the discovery of banking information regarding Sir Paul McCartney[7].

There had also been limited academic research on the subject, including a paper by Garfinkel and Shelat in 2003[8]. However, there had been no long-term scientific investigation to determine whether the issues relating to the problems regarding the safe disposal of equipment were changing in response to the developing technical and regulatory environments. The disk studies over the last four years have focussed on the levels and types of information that have been found on computer disks that were obtained on the second hand market in a number of countries. The hand held device research took place for the first time in 2008. This was in response to the realisation that the increasing processing power and storage of these devices means that there was a possibility that they were subject to the same issues of unsafe data removal as computers.

The results of the research have been widely reported and publicised, but indicate that little has changed with regard to the level and type of information that is being found on the media and devices. During the research the only tools that were used to gain access to the data on the disks were those that are commonly available and could be used by any competent computer user. For the mobile devices, the tools used were the software that was available from the device manufacturers and free tools.

Throughout the research period, the information that has been found came from a wide range of sources, including government, the financial sector, the legal profession, academia, healthcare, the automotive, agrochemical and other industries, the leisure sector, the retail sector and individual's personal computers and devices.

The results of the study revealed that information from which the organizational that had previously

[6] Canadian Globe and Mail (1993), Disk Slipped Into Wrong Hands, *Canadian Globe and Mail*, 2nd August 1993

[7].Calvert, J, Warren, P (2000), Secrets of McCartney Bank Cash Are Leaked, *Daily Express*, 9 February 2000, pp 1–2.

[8].Garfinkel S.L, Shelat A, (2003), Remembrance of Data Passed: A Study of Disk Sanitization Practices. *IEEE Security & Privacy*, Vol. 1, No. 1, 2003.

owned the disks could be identified was recoverable from 52 percent of the disks and that information from which an individual could be identified was recoverable from 51 percent of the disks (some of the disks contained both personal and organization information). Only 31 percent of the disks in the study had had the data deleted to a standard where it could not easily be recovered.

In the study into the data that could be recovered from hand held devices the results showed that while no data could be easily recovered from 51 percent of the devices, 23 percent of the devices that were working and could be accessed contained organizational information and 19 percent of the devices contained personal information.

One example of the type of data that was recovered was patients' records of individuals that were being treated for cancer that had been left on the disk from a computer that originated at a healthcare organisation. The psychological damage that the publication of this type of information may have had on the individuals concerned could have been significant.

Another example that came from a disk that was acquired in France that appears to have originated from the Embassy of another European Country. This disk came from a Linux system and contained corporate information relating to 'Cidale' (*Centre for Information and Documentation*) belonging to the Embassy which is located in Paris. This disk contains a wide range of material including details of the network configuration and security data, internal IP addresses, security logs, a domain key for Dotnetflux and the minutes of internal meetings.

Yet another example was disk that was obtained in the USA contained documents relating to Test Launch Procedures from a Government Contactor. Also on the disk were a number of design documents, documents relating to the subcontractor, security policies, blueprints of facilities and personal information on employees and SSN.

A disk obtained in the USA contained information relating to $50 Billion currency exchange proposals that referred to a 15% transaction fees. The currency exchange was between US dollar and Spanish Euros and there were also million dollar bank transfer documents. The disk also contained information including bank account numbers and details of business dealings in a number of countries including Venezuela, Tunisia, and Nigeria. Amongst the information recovered was correspondence from an individual that appeared to be a member of the US Federal Reserve Board which suggested that the acquisition of a Bank Guarantee might not be forthcoming because of some 'questionable' circumstances.

The potential damage that would be caused by this type of information being available to anyone who cares to look for it, without them having to invest any significant effort or specialised tools, could be serious for any organisation or an individual. For a business, the exposure of information such as their current business plans to a competitor could have a detrimental effect on the business's potential profitability or future. For an individual it could lead to identity theft, embarrassment and exposure to potential blackmail attempts.

3. ENVIRONMENTAL FACTORS

There are a number of factors that contribute to the failure to destroy or effectively remove of data from computer hard disks and hand held devices. One is that the storage capacity of computer disks has continued to increase over time at a rate that is close to exponential (exponential growth is described in Moore's Law). Evidence of this has been observed over the period of the research. The storage capacity of the disks purchased over the period has increased from an average of between 20-40Gb in the first year to between 200-300Gb in the past year. Another is the changing use of laptop computers and handheld devices, where there has been an increasing demand for devices that support an increasingly mobile population. A third factor has been the greater availability of high quality and the increasing speed of mobile communications which have developed to support the demands for

computing capability on the move.

These have all contributed to increasing volumes of data being transmitted and stored on an ever wider range of devices. One effect of the increasing availability of storage capacity has been that people both in their employment and in private use have been less likely to destroy data that is no longer required in order to maintain storage space on the media.

4. LEGISLATIVE CHANGES

As the computing and networking technologies have developed to provide new and enhanced capabilities, legislation such as the California state law on disclosure and the UK Data Protection Act has been introduced to meet the changing environment. An effect of the new legislation is that any organisation that holds information from which a person can be identified is now required to have put measures in place to adequately protect the information. Organisations in many sectors such as Government, finance and healthcare also have sector specific regulations, such as the Basel II accord for the financial sector and HIPPA for the healthcare sector. Other regulations such as the Sarbanes Oxley Act have been introduced to improve corporate accountability, but these also have a beneficial effect in the protection of sensitive information. These legislations and regulations have been developed to ensure that the measures that are put in place for the protection of information are adequate and that suitable audit measures are used to ensure that the measures are being followed.

5. CAUSES OF FAILURES

The research and follow up activity with organisations and individuals that could be identified from the recovered disks and devices identified a wide range of reasons for the data being released to people not intended to have access to it. Amongst the most common causes that were isolated were the theft of the device, accidental losses, failures in procedures and negligence. The main cause of the failure to properly dispose of the information in most organisations was attributed to poorly worded and managed third party arrangements where a disposal or recycling company had been contracted to dispose of the equipment and remove the data. In all of the cases that were investigated, there were arrangements in place by the organisation for the third party to destroy the data. In a small number of cases the third party had failed to take any action to remove the data. However, in most of the cases, the third party had fulfilled this requirement to destroy the data from the devices by the use of the Windows format command or a tool that had a similar effect. The underlying causes of this problem were twofold: First, the wording of the contract with the third party did not specify the standard to which the data should be destroyed; Secondly, the organisation did not have in place measures to test that the destruction of the data had been effectively carried out to the required standard.

Any competent information security professional and most competent computer users would know that the use of the Windows format command does not actually destroy the data, it only removes the file structure which is normally used to access it. In an attempt to support users who make mistakes, Microsoft also created an Unformat command which allows the file structure to be recreated with relative ease. While it could not be proven during the research and subsequent follow up investigation, it is clear that the actions taken by some recycling companies have clearly met the contractual requirements of the disposing organisation (destruction of the data to an unspecified standard). The fact that the data could subsequently be recovered appears to be a result of them carrying out their contract in the most cost effective manner (in terms of the cost of specific tools and manpower. It is known that for central government and a number of other organisations, the contracts with the third party recycling organisations mandate the use of specific tools and processes for the destruction of data on media and devices and include the right to inspect the processes and test their effectiveness.

For disks and devices from private individuals, the main reasons for the failure to adequately destroy

the data was that of ignorance of the potential value of the information that was contained in them or the potential impact and a lack of technical knowledge. The majority of private users do not have easy access to the knowledge, skills and tools that could be used to adequately destroy the data.

RECOMMENDATIONS

From the research there were a number of measures identified that can be taken to ensure that information is destroyed effectively and does not end up in the public domain. For computer disks, the measures include:

- Education of users - Public awareness campaigns by Government, academia, the media and within organisations.

- Best Practice - The development of best practice within sectors and its adoption by organizations to ensure that computer hard disks are disposed of in an appropriate manner.

- Risk Assessments – Organisations need to carry out risk assessments to determine the sensitivity of the information on computer disks and determine the measures that need to be taken for its effective removal.

- Tools– The development of, and access to, data erasure tools such as Blancco[9] and access to facilities to enable individuals to effectively remove the information from their computers.

- The use of Encryption - The full or partial encryption of hard disks to protect sensitive information and to ensure that, in the event of the disks being released into the public domain, information could not be easily recovered. The types of tools that can be used to achieve this include software such as TrueCrypt[10] or PGP whole disk encryption[11] or hardware encryption devices such as the Secure Data Vault[12].

- Asset Tracking - organisations could improve the effectiveness of the security of their data if asset tracking is conducted at the hard disk level. This would require that asset tags are placed on individual disks rather than the computer system unit to ensure appropriate disposal.

- Allocation of responsibilities – Responsibility should be assigned to all of those involved in the process of the disposal of hard disks, including those that are damaged or have failed. Disks that are not working or faulty should have the same disposal practices applied to them as disks that are working correctly.

- Physical Destruction - Where appropriate, if the sensitivity of the data demands it, the physical destruction of the disks using services such as the Ultratec Secure Data Erasure service[13] or that offered by DataTerminators[14] should be considered.

[9] Blancco - http://www.blancco.com/en/frontpage/

[10] TrueCrypt - http://www.truecrypt.org/downloads.php

[11] PGP Corporation, Whole disk ensryption - http://www.pgp.com/products/wholediskencryption/index.html

[12] Secure Systems Secure Data Vault - http://www.securesystems.com.au/

[13] Ultratec Limited - http://www.ultratec.co.uk/

[14] DataTerminators - http://www.data-terminators.co.uk/

For mobile devices such as PDAs or mobile (cell) phones, the measures that should be considered include:

- Education of users - Education and awareness training should be developed and delivered to improve user awareness.
- Development of Best Practice – The development within organizations of best practice for the appropriate disposal of the information on mobile devices.
- Data Erasure Tools – Ensure that access is available to the tools and instructions such as model specific data removal information[15] for the appropriate removal of data from hand held devices.
- Contracts – Ensure that the chosen recycler[16] or organisations[17] that accept donated hand held devices guarantee that the devices are data cleansed before they are sold on and have procedures in place to ensure that they are carried out.

In isolation, it is unlikely that these measures will reduce the level of risk of the potential exposure of sensitive information to the individual or an organization. It is only when the appropriate measures are used together that any significant change is likely to occur.

FUTURE WORK

It is planned that this research will continue into both the computer disks and hand held devices. The research into computer hard disks will continue unchanged, but future research into hand held devices will be concentrated on 2.5 and 3G devices, RIM Blackberries and PDAs. The reason for this change is that the initial research has determined that the risk of information loss from 2G devices low due to their limited functionality and storage capacity. In addition, these devices will be progressively replaced by the more function rich 3G type devices.

[15] Recellular Free Data Erasure tools - http://www.recellular.com/recycling/data_eraser/default.asp

[16] PHS Datashred - http://www.recyclemycomputer.co.uk/recycle-mobile-phones.htm

[17] Birmingham Focus on Blindness - http://www.birminghamfocus.org.uk/html/display.php/id/419

The Computer Fraud and Abuse Act and the Law of Unintended Consequences

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

Vicki Luoma
Minnesota State University
145 Morris Hall
Mankato, Minnesota 56001
507 389-1916
507 389-5420
Vicki.Luoma@mnsu.edu

ABSTRACT

One of the most unanticipated results of the Computer Fraud and Abuse Act arose from the law of unintended consequences. The CFAA was originally enacted in 1984 to protect federal government computers from intrusions and damage caused by hackers, identity thieves, and other cyber criminals. The law was later amended to extend the scope of its application to financial institutions', business's and consumers' computers. To aid in the pursuit of cyber criminals, one of the subsequent revisions to the law included provision "G" that gave the right to private parties to seek compensation for damages in a civil action for unauthorized computer intrusions. This amendment to the law has had the unintended consequence of bolstering, or in some cases supplanting, claims against employees and former employees for claims such as trade secret violations, intellectual property violations, and violations of covenants not to compete. This amendment has also aided employers in their defense of employee claims of sexual harassment, wrongful termination, and other claims by facilitating counterclaims against employees and former employees for computer misuse. This paper examines these developments in the law and likely unintended consequences of the original amendments to the Computer Fraud and Abuse Act.

Keywords: computer, fraud, intellectual property, law

1. INTRODUCTION

In response to a substantial increase in cybercrimes, the United States Congress passed the Computer Fraud and Abuse Act (CFAA) in 1984. The first version of the act dealt with illegal acts performed against government computers and government financial records. As soon as the act was passed it was clear that the law was not adequate to deal with the ever increasing frequency of cybercrimes and the increasing interdependence of computers. (Luoma, 2008) As a result, the act was amended several times and in 1996 the act was changed to include provision "G" that allows civil actions and civil penalties. (Computer Fraud and Abuse Act, 1984)

Section G reads as follows:

> Any person who suffers damage or loss by reason of a violation of this section may maintain a civil action against the violator to obtain compensatory damages and injunctive relief or other cases involving equitable relief. A civil action for a violation of this section may be brought only if the conduct involves 1 of the factors set forth in clause (i) (ii), (iii), (iv), or (v) of subsection (a) (5)(B). Damages for a violation involving only conduct described in subsection (a) (5) (B) (i) are limited to economic damages. No action may be brought under this subsection unless such action is begun within 2 years of the date of the act complained of or the date of the discovery of the damage. No action may be brought under this subsection for the negligent design or manufacture of computer hardware, computer software, or firmware." (Computer Fraud and Abuse Act, 1984)

The inclusion of a civil section in a criminal statute is unprecedented in federal criminal statutes, and it can only be assumed that in an effort to combat increased crime, the government was willing to empower civil litigants to fight cybercrime, too. However, the section was not used by any civil litigants until the attorneys in the Shurgard case argued that the CFAA should be applicable to cases in which an employee forwards company information to his new employer by using his company's computer to send the information over the Internet. (Shurgard v Safeguard)

Warren Rheaume and Roanne Spiegel, attorneys representing Shurgard, sued the defendants, former Shurgard employee Eric Leland and his new employer, Safeguard, with a variety of causes of action including violations of the Computer Fraud and Abuse Act. (Luoma, 2008) The defendants moved the court to summarily dismiss the alleged cause of action under the CFAA because defendant-employee Eric Leland had permission to use his company computer so long as he was still an employee of Shurgard. The court ultimately ruled that even though Shurgard gave its employees permission to use company computers, the employee loses computer authorization as soon as the employee's actions are disloyal. (Shurgard v. Safeguard) Therefore, in the Shurgard case, as soon as Leland sent his new employer proprietary information, copied proprietary data, or did any other disloyal act using the computer, he no longer had the right to use his employer's computer, and hence, the CFAA provisions were applicable to his misdeeds. (Shurgard v Safeguard)

2. THE IMPORTANCE OF THE SHURGARD CASE

In any cause of action alleging tort, the plaintiff has the burden of proving four points. First, the plaintiff must first prove that the defendant owed a duty to the plaintiff. Second, the plaintiff must establish that a breach of that duty occurred. Third, the plaintiff must show by a preponderance of the evidence that the breach of the duty was the proximate cause, or legal cause, of the damages that flowed from the breach. Finally, the plaintiff has to prove the nature and amount of the damages. (Restatement 2d of Torts) Proof of the existence of a duty is generally straightforward in virtually all tort cases. For some complex torts, proof that the breach of duty was the proximate cause – that the breach of duty was cause in fact and was foreseeable – can be difficult.

In civil litigation against employees or former employees that involves any computer misuse, the CFAA has become the allegation of choice because the only requirement to is to prove, first, that the defendant misused the plaintiff's computer, and second, that the plaintiff suffered at least $5,000 in damages. At least one court has held that the hiring of a computer forensics expert to determine whether there have been damages is, in fact, part of the damages. (EF Cultural Travel BV v. Explorica, Inc., 2003) If $5,000 is spent on forensics experts, then the CFAA applies.

Legal actions under CFAA are beginning to replace or to bolster litigation that involved solely accusations of violations of trade secrets, employee misappropriation of information, restraint of trade, violations of covenants not to compete, and other torts. This portends a trend in litigation where allegations of violations of the CFAA will become routine. One of the primary benefits of including this cause of action in a lawsuit is that it is relatively easy to prove compared to other common causes

of action in business litigation. A brief review of some of these causes of action will illustrate this point.

2.1 Trade Secret Litigation

Trade secrets are business processes not protected under trademarks, patents or copyrights, but are still considered to be an important part of what makes the business unique and successful. (Ellis, 2005) It can includes items like customer lists, pricing, marketing plans, business plans, store locations, business floor plans, and secret recipes. (Restatement 2d of Torts) One of the most significant advantages of trade secrets over other forms of intellectual property protection is the fact that the protection is perpetual so long as the secret is kept intact. The right to exclusive use does not expire after some statutory time period. (Restatement 2d of Torts)

There are strict laws against the theft of trade secrets, including the Uniform Trade Secret Act, which has been passed in part by thirty states, and the Economic Espionage Act, which makes it a federal crime to steal trade secrets. However, with all of this legal support the plaintiff must make a prima facia case to the court of each and every element of the cause of action. (Restatement Torts 757)

The difficulty in proving a violation of the trade secret law is that the plaintiff has the legal burden to establish that the information was stolen, that the information in fact was legally protected, and that the plaintiff was damaged. In, Coco v. A.N. Clark, the court set the standard that plaintiffs must prove to win a trade secret case as follows:

- the information itself must have the necessary quality of confidence about it;
- that information must have been imparted in circumstances imparting an obligation of confidence;
- there must be an unauthorized use of that information to the detriment of the party communicating it. (Coco v. A.N. Clark Engineers Ltd, 1969)

In the Restatement Second of Torts, comment b of the first Restatement lists six factors to be used to determine whether something is a trade secret of a particular person:

- the extent to which the information is known outside of his business;
- the extent to which it is known by employees and others involved in his business;
- the extent of measures taken by him to guard the secrecy of the information;
- the value of the information to him and to his competitors;
- the amount of effort or money expended by him in developing the information;
- the ease or difficulty with which the information could be properly acquired or duplicated by others. (Restatement Second Torts)

In addition, the plaintiff must prove how it was harmed. Harm can be a very difficult element to prove in the court. Attempting to prove these elements can be extremely costly and time consuming. Alternatively, bringing the action under the theory of a violation of the CFAA is much easier to prove – "Has the employee misused the computer use agreement – yes or no?" The only remaining question is how much are the damages?

2.2 Covenants not to Compete

Often employers require employees to sign a "covenant not to compete" as a condition of employment. Yet violations of these covenants not to compete are also difficult to pursue legally because courts often find them to be a restraint of trade and they are reluctant to rule in favor of the plaintiff. In a typical case alleging that defendants violated a company's covenant not to compete, Spiegel v. Thomas, the court found that before the court would consider the covenant enforceable, the court must consider whether there was adequate consideration for the covenant, whether there was a threatened danger to the employer in the absence of the covenant, whether economic hardship would

be imposed on the employee, and whether the covenant is against the public policy. In addition, the employer must prove that the covenants are reasonable in time and geography before the court will uphold these agreements. (Spiegel v. Thomas, Mann & Smith, P.C., 1991) Proving all of these points can represent a very difficult burden for a plaintiff. Even if a plaintiff does have convincing evidence of a breach of the covenant, courts are reluctant to prevent a person from being gainfully employed or limiting competition that violates free market principles. Again, if the employee or former employee can be shown to have violated a computer use policy in some manner, damages can be recovered from the employee.

In another case, P.C. Yonkers, the plaintiff, claimed that former employees not only started a competing store within their former employer's sales district but accessed their computers more than one hundred times to gather marketing, sales and other information that they used to compete with the plaintiff's business. Alleging and proving computer misuse was much easier and less costly than the requirements to prove that the plaintiff provided adequate consideration to each of these defendants for the signed covenant not to compete and that the actions of the defendant starting a competing business in their area was a danger to the plaintiff.

In addition, the plaintiff would have to prove explicitly how these actions caused economic loss to the plaintiff. Proof would require more than the plaintiff's revenue decreased. It would require proving specific customers of the plaintiff went to the defendant based on defendants' illegal action to entice them to change companies. If the customers claimed they independently changed suppliers, the plaintiff's case would evaporate.

If the plaintiff chose to pursue the defendant based on trade secret violations the plaintiff would have the burden to prove that the defendant actually established that each business process in question was in fact stolen. The plaintiff would have to prove that each of the alleged stolen trade secrets – floor plan, customer lists, marketing plan or other business process – was in fact legally protected. Defendants could argue that any given process was not unique, was common industry practice, or independently designed. Then, as with the covenant not to compete, trade secret damages must be actually proved. (P.C. Yonkers v. Celebrations) Again, courts generally require very convincing evidence in order to find such a violation because to do so operates against free market principles.

2.3 Intellectual Property Litigation

Another action in which litigants explored the use of CFAA occurred in an international intellectual property dispute. Laws and treaties that cover international intellectual property rights and trade secret violations exist, but they carry with them stringent evidence requirements. For example, in Facebook, Inc. v. Studivz, Ltd. a German company created a website almost identical to Facebook's Internet social network site. Facebook was created by Mark Zuckerberg, Dustin Moskovitz and Chris Hughes, Harvard Students, as a social network and it quickly developed into a multi-billion dollar business. (Facebook v Studivz, 2008) On the Facebook site members and guests are required to click on a button indicating that they have read the "Terms of Use." The Terms of Use agreement states in part that

> You understand that the website is available for your personal non-commercial use only. You agree that no materials of any kind submitted through your account will violate or infringe upon the rights of any third party, including copyright, trademark, privacy or other personal rights or contain libelous, defamatory or otherwise unlawful material. (Facebook, Inc. v Studivz LTD, 2008)

In fact, Studivz admitted it copied the Facebook site and the only changes made on the Studivz site were to change the language to German and the background color to red. Facebook sued Studivz under CFAA. Facebook argued that anyone who uses or visits the Facebook website must signify that they have read and agreed to be bound by Facebook's "Terms of Use" whether or not they are a

registered member of Facebook. (Facebook, Inc. v Studivz LTD, 2008) This case has not yet been resolved but it shows a movement toward suing under the easier requirements of the CFAA.

The other important aspect of this case is the definition of computer misuse is also expanding. In an earlier case, EF Cultural Travel BV v. Explorica, Inc., former employees of the travel agent used a scraper program to gather data information obtain pricing information from their former employer's travel provider's website. Anyone with a scraper program could have obtained the unprotected information. Further, the website did not require users to agree not to gather the information. (Luoma, 2008) In this case the court found that the defendants violated the CFAA *even though the information gathered could have been gathered legally by any other person in the world other than the employees.* (Emphasis added) The court found that it was a violation for employees because the employees knew how to use the information scraped from the website and use it in competition with their former employer. In addition, the court ruled that a company "can easily spell out explicitly what is forbidden." (E.F. Cultural Travel v. Explorica, 2003)

Companies such as Facebook heed the court's ruling in E.F. Cultural by setting explicit terms of use. Studivz, Ltd creators obtained access to the Facebook site as a guest. The complaint alleged that Studixz creators accessed the Facebook site on servers in California from Germany and from other worldwide locations. This allegation establishes the jurisdiction of the United States court to hear and consider the suit brought by Facebook against the German company. (Facebook v Studixz)

3. DAMAGES

The second requirement for an action under CFFA is that the plaintiff must be able to prove damages sustained in the amount of at least $5,000. The Computer Fraud and Abuse Act defines 'damage' as "impairment to the integrity or availability of data, a program, a system, or information." 18 U.S.C. Sec 1030 (e)(8). Likewise, the CFAA defines 'loss' as "any reasonable cost to any victim, including the cost of responding to an offense, conducting a damage assessment, and restoring the data, program, system, or information to its condition prior to the offense, and any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service. (Ackerman, 2005) The United States ("US") Computer Fraud and Abuse Act (CFAA), 18 U.S.C. Sec 1030 et. seq.,

The proof of damages under the CFFA is not so difficult as proving damages under trade secrets or intellectual property torts. In the E.F. Cultural case found that the cost of hiring a computer forensic expert to prove the defendants had scraped their site was sufficient to meet the damages requirement. Subsequent cases have found that merely hiring a computer forensic expert to find the breach is sufficient to meet the $5,000 damages requirement even if the breach itself does not meet the $5000 threshold. (Akerman, 2004)

In Charles Schwab & Co., Inc v. Brian D. Carter, Acorn Advisory Management, Schwab either transferred or terminated all employees in a division of Schwab, Soundview Capital Markets' Investment Analytics Division (IA) by November 1, 2004, for which defendant Carter worked. Carter resigned on October 22, 2004 and began working with Acorn. IA received analytical research from various companies including Acorn. When IA announced it was closing this division, Acorn offered to purchase the company. IA turned down the offer. IA then made job offers to several employees including Carter. IA claims that Acorn induced Carter to copy computer information to Acorn. IA alleged it incurred costs of at least $5,000 in damages over a year. The court found that the plaintiff's could pursue a CFAA case and this allegation of damages over a year period is sufficient. (Charles Schwab & Co., Inc v Brian D. Carter, Acorn Advisory Management, LLC and Acorn Advisory Capital, L.P., 2005)

However, in 2008 in American Family Mutual Insurance Co. v. Rickman an employee had accessed his former employer's computer without permission and copied files; however, the court found that the employer had to prove damages to the computer system or interruption of a computer service.

(American Family Mutual Insurance Co. v. Rickman) In Cohen v. Gulfstream Training Academy the court required the same definition of damages as the American Family Mutual Insurance Case; however, most cases have found a much broader definition of damage. (Cohen v Gulfstream Training Academy) In Creative v. Getloaded LLC the court found that CFAA does not require the $5,000 damages to be from a single act. (Creative v. Getloaded LLC , 2004) The court ruled that "reaching the damage amount could include conducting a damage assessment, restoring the data, program, system or information or other consequential damages or even upgrading the computer system to prevent future violations." (GetLoaded) Another definition of loss was cited Four Seasons where the court determined that damages could also be revenue loss. (Four Seasons)

In a 2009, Kalow & Springnut, LLP v. Commence Corporation the federal district court in New Jersey found that plaintiff's allegation that defendant's software product contained a "time-bomb" causing the software to stop working after a period of time met the CFAA's standard of intent to cause harm. The defendant argued that the plaintiff's claim relied on faulty logic "which fails to consider other possible explanations, such as a programming error in the software." The court turned down the motion for summary judgment using the standard set in Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007) that "simply calls for enough facts to raise a reasonable expectation that discovery will reveal evidence of the necessary element."

While there does seem to be inconsistent interpretations of what may be included in determining the $5,000 jurisdictional amount, the trend seems to be with a broader interpretation of that provision.

Use of CFAA as the Basis of a Counterclaim

Perhaps an even more important potential use of the CFAA could be as a counterclaim to a suit brought by an employee. For example, suppose an employee or a former employee brings an action for sexual harassment or wrongful discharge. Rather than being entirely on the defense, the employer could use the CFAA as a counterclaim to the plaintiff's claim if the employee engaged in any misuse of his or her computer during employment. Misuses of company computers as benign as surfing the Internet or otherwise engaging in personal business during working hours could lead to a claim against the employee. If the misuse involved illegal activities and very serious breaches of company loyalty during employment, the plaintiff-employee may well decide that his or her original claim may not be as strong as originally thought.

Since the federal rules of civil procedure require a litigation hold if the employer has reason to believe that an employee may file a claim in litigation, it is incumbent upon employers to immediately secure and quarantine any computer used by such an employee and then obtain a forensic examination of it. Such an examination will be vital to determining whether the employee engaged in any misuse of that computer during his or her employment. Such an examination will preserve evidence that can be used to substantial advantage in any litigation that may subsequently occur.

4. CONCLUSION

In conclusion, the unanticipated results of the Computer Fraud and Abuse Act arise from the ambiguous nature of the law that only leads to the possibility of future unintended but creative uses of the act. The use of civil remedies of section "G" of CFAA have been a creative and efficient method to bring a private cause of action for misappropriation of confidential information or trade secrets against current and former employees. Even though there is an epidemic of stolen identities, hacked computers and a variety of other cyber crimes, the additional uses of the Computer Fraud and Abuse Act have been positive additional unanticipated uses. Whether employees are aware of this possible application is not as important as the fact that it is a useful addition to the prevention of computer misuse. Employees must comply with computer use policies of their companies and not conduct criminal or other unlawful activities. Future employers of these departing employees must seriously review what information their employees bring with them and where and how they obtained that information. The CFAA continues to be an effective method for employers to stop disloyal employees

from committing tortious acts against the employer's interests. However, with the recent split of authority regarding "unauthorized access," employers are advised to draft clauses into their confidentiality agreements with employees that clearly define what access is authorized and what access is unauthorized, including the precise time when authorized access becomes unauthorized. Finally, it would be appropriate to inform employees of the potential application of the Computer Fraud and Abuse Act in the event of computer misuse.

REFERENCES

Ackerman, N. (2005). CFAA as a Civil Remedy. *National Journal*, 12.

Akerman, N. (2004). CFAA's $5,000 Threshold. *National Law Journal*, 19-21.

American Family Mutual Insurance Co. v. Rickman .

Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007)

Burke, E. (2001). The Expanding Importance of the Computer Fraud and Abuse Act. *Giglaw*.

Charles Schwab & Co., Inc v Brian D. Carter, Acorn Advisory Management, LLC and Acorn Advisory capital, L.P., Case No. 04C7071 (United States District Court for the Northern District of Illinois Eastern Division 005 U.S. W. Feb. 11, 2005).

Coco v. A.N. Clark Engineers Ltd, (1969) R.P.C. 41 at 47

Computer Fraud and Abuse Act, 18. U.S.C.1030 (Federal 1984).

Creative v. Getloaded LLC , No. 02-35856 (9th Circuit October 15, 2004).

Ellis, E. (2005). Trade Secrets. *The Computer and Internet Lawyer*, 7-29.

Facebook, Inc. v Studivz LTD, 5:2008cv03468 (California 2008).

Four Seasons Hotels & Resorts BV v. Consorcio Barr, SA, 267 F. Supp.2d 1323-1324 (SD Fla. 2003)

Heath Cohen v. Gulfstream Training Academy, Inc. and Gulfstream International Airlines, Inc.

Case No. 07-60331-Civ-Cohn/Seltzer (S.D. Fla., April 9, 2008)

In E.F. Cultural Travel BV v. Explorica, Inc. (U.S. Court of Appeals for the First Circuit 2003).

Kalow & Springnut, LLP v. Commence Corporation No. 07-3442, 2009 WL 44748 (D.N.J. Jan. 6, 2009).

Luoma, M. & Luoma, V. The Computer Fraud and Abuse Act: An Effective Tool for Prosecuting Criminal and Civil Actions in Cyberspace. *Forum on Public Policy.*(2008)

Nexans Wires S.A. v. Sark – USA Inc., 319 F. Supp. 2d 468, 469 (Southern District of New York).

P.C. of Yonkers, Inc. v. Celebrations! The Party and Seasonal Superstore, L.L.C., 2007 U.S. Dist. LEXIS 15216 (D.N.J. 2007)

Shurgard Storage Centers, Inc. v Safeguard Self Storage, Inc., 119 F. Supp. 2d 1121 (Washington 2000).

Spiegel v. Thomas, Mann & Smith, P.C., 811 S.W. 2d 528, 529–30 (Tenn. 1991

Uniform Trade Secret Acts Section I(4).

Concerning File slack

Stephen P Larson
VCU School of Business, Richmond, VA
larsonsp@vcu.edu

ABSTRACT

In this paper we discuss the phenomena known as file slack. File slack is created each time a file is created on a hard disk, and can contain private or confidential data. Unfortunately, the methods used by Microsoft Windows operating systems to organize and save files require file slack, and users have no control over what data is saved in file slack. This document will help create awareness about the security issue of file slack and discuss research results concerning file slack.

Keywords : Computer Forensics, File Slack, Ram Slack, Disk Slack

1. INTRODUCTION

In this digital age, keeping personal or confidential data private is quite difficult. Regrettably, this problem is exacerbated by the very technology we use to create the digital data. It has already been established that a problem exists with users and companies selling off old hard disks that still contain commercial or personal data, even if the hard disk has been formatted (Garfinkel and Shelat, 2003; Jones, Valli, Sutherland, and Thomas, 2006; Jones, Valli, Dardick, and Sutherland, 2008). What has yet to be established is the extent of commercial, private, or personal data that can be transmitted via file slack. "File slack" can contain data dumped randomly from the computer's memory, data from previously deleted files, etc., and can potentially reveal prior uses of the computer such as fragments of email messages, network or internet site logon names and passwords, etc (Volonino, Anzaldua, and Godwin, 2006).

For this paper, we will limit our discussion to Microsoft Windows OS because "unlike Windows … file systems, UNIX does not have file slack space. When UNIX creates a new file, it writes the remainder of the block with zeros and sets them as unallocated. Therefore it is not possible to recover deleted data from slack space on UNIX systems" (Casey 2004). According to NetApplications, roughly 90% of operating systems on PCs are some version of Microsoft Windows (NetApplications 2008). Additionally, Steve Ballmer of Microsoft stated that "forty percent of servers run Windows" (Niccolai 2008).

"File slack is a data storage area most users are unaware of" (Vacca, 2002). "It is a source of significant *security leakage* and consists of raw memory dumps that occur during the work session as files are closed" (Vacca, 2005).

This paper will introduce how data is saved on hard disk drives, give definitions of file slack, ram slack, and disk slack, explain how file fragmentation affects file slack, and explain how file slack on file servers is shared among users, and determine whether file slack is "portable." We will also discuss future research needs for file slack.

2. HOW DATA IS SAVED ON HARD DISK DRIVES

During its manufacture, a low level format of the hard disk drive is done by the manufacturer to ready it to be partitioned into one or more logical partitions or volumes. During the low-level format tracks and sectors are created. A track is a concentric ring around the platter containing information. Hard drives typically contain several platters, with the tracks on each platter lining up. The tracks are then divided up into sectors. A sector is the smallest unit of the hard drive and is 512 bytes in size. A cluster contains one or more sectors. "All Microsoft operating systems read and write in blocks of data called *clusters*" (Volonino, et al, 2006).

More simply put (see Figure 1):

- A platter contains concentric tracks
- A track contains 512 byte sectors
- A cluster contains one or more sectors
- A cluster is the smallest unit on disk for storing a file

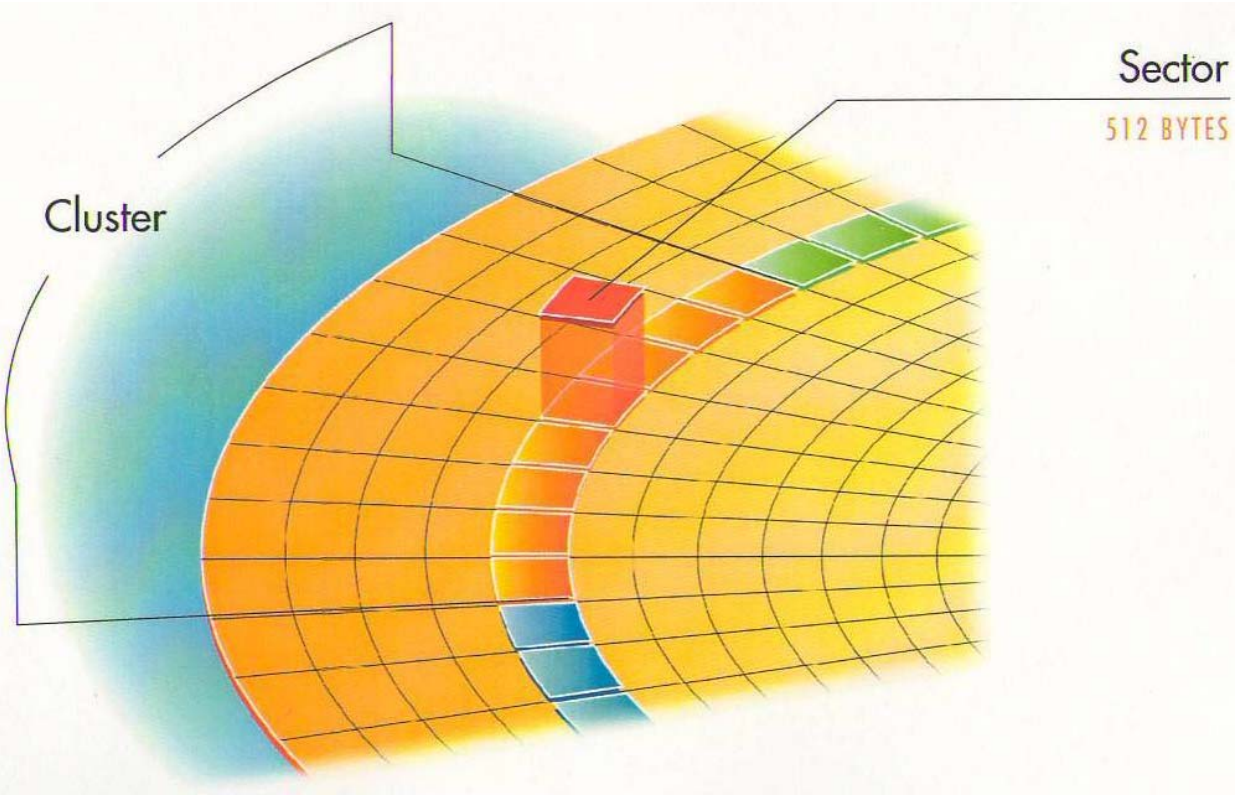


Figure 1. Sectors and Clusters (New Technologies, Inc. 2001b)

In Microsoft Windows operating systems, the default cluster sizes are shown in Table 1. As most PCs running a Windows OS have hard disk drives with partitions or volumes greater than 2 GB but less than 2 TB in size, the most common cluster size is 4 KB; each cluster contains 8 sectors that are 512 bytes in size.

Volume or partition size	NTFS cluster size
7 MB – 512 MB	512 bytes
513 MB – 1,024 MB (1 GB)	1 KB
1,025 MB – 2 GB	2 KB
2 GB – 2 TB	4 KB

Table 1. Default cluster sizes for the NT File System in Microsoft Windows (Microsoft 2007).

3. FILE SLACK, RAM SLACK, AND DRIVE SLACK

How does cluster size affect the size of files on your hard disk? A file on the hard disk must be the same size as a default cluster size; currently the most common cluster size is 4 KB, or multiple of 4 KB. But file sizes rarely exactly match the size of the clusters. The space that exists from the end of the file to the end of the last cluster assigned to the file is called "file slack". Larger cluster sizes mean more file slack and also the waste of storage space (Reyes & Wiles 2007). A cluster size of 4 KB means there is a potential for 3.9 KB of space wasted for a file; this also means a potential for 3.9 KB of unwanted file slack being attached to the end of a file.

3.1 File Slack

Let's create a text file with the word "hello" in it and save it on the desktop as a file named 5byte. Then we will view the properties of the file by right-clicking on the file and choosing "properties." The file's properties are shown in Figure 2.

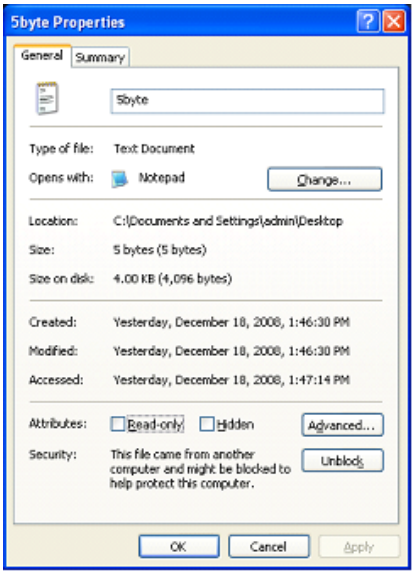


Figure 2. File properties

Notice the size of the file is 5 bytes, but the size on disk is 4 KB. This tells us that there is 4,091 bytes of file slack attached on the end of this file. What are the contents of file slack? "Ram slack" plus "disk slack."

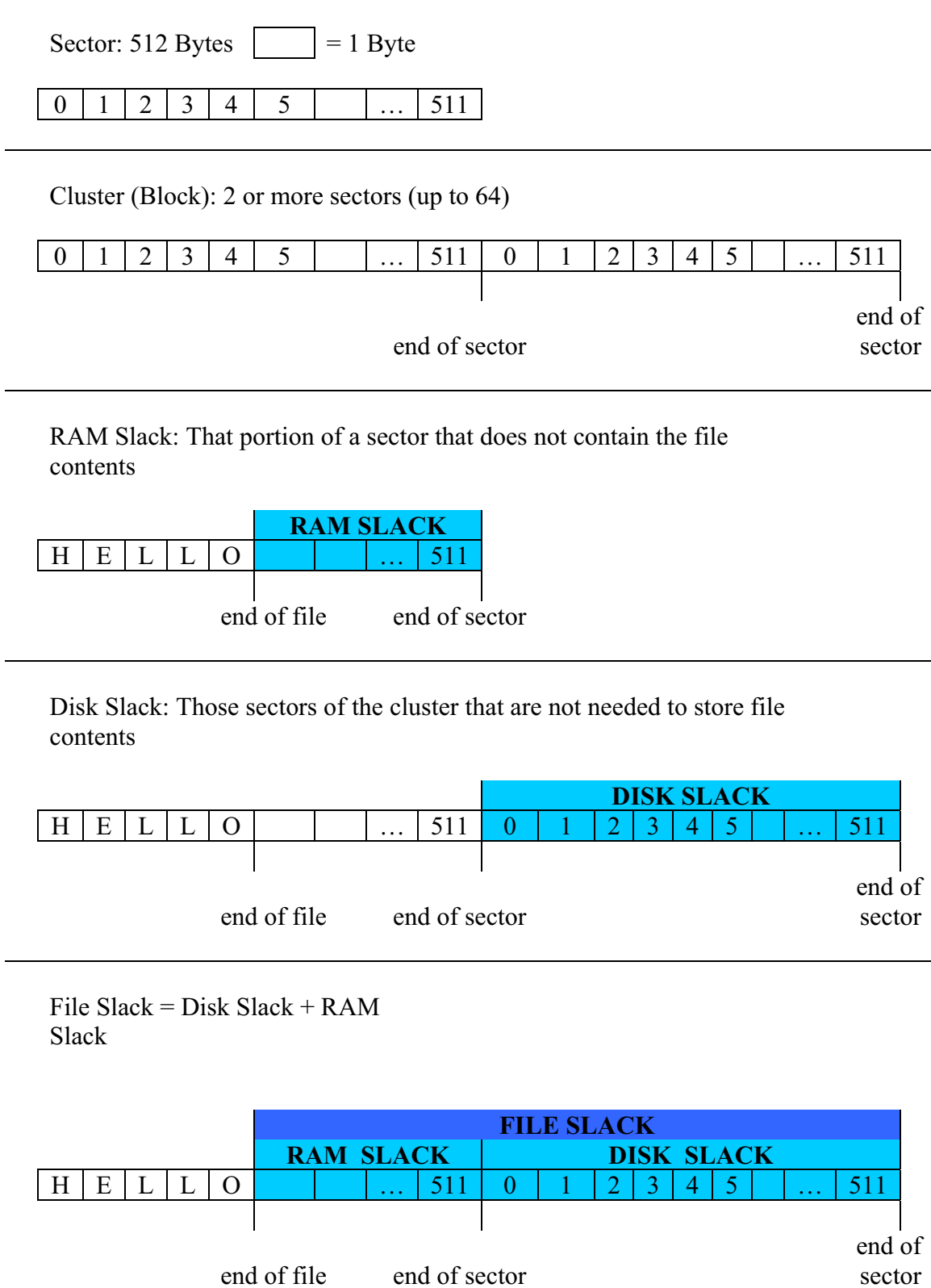
3.2 Ram Slack and Disk Slack

Using our example file, the 4,091 bytes of "file slack" contains "ram slack" and "disk slack."

The ram slack is the data required to fill in the space from the end of the file to the end of the sector. As previously discussed, a sector is 512 bytes. The first 5 bytes of the file are used by the text of the file. The next 507 bytes are filled with "ram slack". Why is this called ram slack? The Windows operating system used to fill in the space from the end of the file to the end of the sector with randomly selected data pulled in from RAM (New Technologies, Inc. 2001a), but now will fill in the space with zeros.

That leaves 3.5 KB, or 7 sectors of 512 bytes each, of space to fill in. Disk slack is the space from the end of the sector that contains the end of the file to the end of the cluster. This can contain one or more sectors. Because the file does not have a need for this space, but the file system needs to fill in the space with something, the data that was previously in the sectors is used and not overwritten. A graphical representation of file slack (containing ram slack and disk slack), using a cluster size of 2 sectors or 1KB, is shown in Figure 3.

Figure 3. Sectors, Clusters, RAM slack, Disk slack, and File Slack (adapted from Dampier 2008).



3.3 File Slack and Disk Fragmentation

Any time a file is deleted, the clusters on the hard disk which were used by the file is marked free or unallocated by the file system. These clusters either get overwritten by another file or become disk slack. As you use your hard disk, clusters belonging to a file can become scattered all over the hard disk. This phenomenon is called fragmentation. To illustrate, let's consider an example. You start off with a newly formatted hard disk using 4 KB clusters. If you save a file that is 8 KB in size, the file system will use clusters 1 and 2 for that file. The next file you create is 12 KB in size. Because clusters 1 and 2 are in use, the file system will start the new file in cluster 3 and end in cluster 5. All clusters in a file contain information about the physical location of each other on the hard disk. Next you delete the first file you created, and create another file that is 3 KB in size. The file system will see that the first available cluster is cluster 1, and will save the file in cluster 1. However, because the file is only 3 KB and the cluster is 4 KB in size, on KB of file slack, which contains data from the deleted file, is "attached" to the new file. You then create a fourth file that is 8 KB in size. The file system sees that the first available cluster is cluster 2, and starts saving the file in that cluster. When cluster 2 is filled, the file system finds the next available cluster, cluster 6, and saves the rest of the file there. File number 4 is fragmented. After normal use for even a few months, a hard disk can become heavily fragmented. The computer on which this paper was written has been in use for 5 months, and has 4279 fragmented files and 23892 excess fragments, with an average of 1.48 fragments per file. Anytime a file gets deleted, its clusters, which could be scattered all over the hard disk, become available for use by another file. Recent versions of MS Windows can be configured to run disk defragmentation automatically and continuously to minimize fragmentation. This process also moves parts of files around the disk. Thus, file slack could contain any type of data.

3.4 File slack on file servers

Until now, our discussion has focused on file slack on a personal computer. The file slack issue is compounded on a file server. File servers using a Windows operating system store information in clusters (Volonino, et al. 2006). Consequently, when a file is deleted, its clusters are marked as unallocated and can end up as file slack in another user's file. Thus, if all departments use the same file server to store their files, confidential Human Resources or Finance information, such as social security numbers or salaries, can become part of the file slack in another user's file.

4. DISCUSSION

Due to the nature of how Windows OS stores data in clusters, there is a potential for private, personal, or commercially sensitive data to be included with a file in the file slack area. File slack is a primary source of electronic evidence because the clusters that made up deleted files are released by the operating system and are unallocated until overwritten by new file content.

4.1 Questions about file slack

Exploring file slack brought to mind certain questions:

- Does file slack accompany a file when it is emailed?
- Does file slack accompany a file when it is copied to another disk or media (hard disk, USB, CDROM, etc.)?
- Does file slack accompany a file when FTP is used to copy the file to another location?
- Does file slack accompany a file when the file is saved under a different name?
- Does file slack from a file on a file server accompany the file when it is sent to a different location?
- Does the recipient's OS add file slack from its own hard disk or file server's hard disk?

To answer these questions, I started with a USB disk that had been prepared by formatting it, creating a file with easily viewable text on it, then erased that file, and created the file mentioned above (5byte.txt). Using a blank USB disk instead of a hard disk drive ensured the file wouldn't be created in the MFT of the Windows XP operating system's Master File Table. "If the data in the file is small (typically a few hundred bytes), then this data can be completely contained within the Master File Table (MFT) record of the file" (Microsoft, 2004). A file saved in the MFT will not exhibit the same file slack as a file saved elsewhere.

After creating the file, I then proceeded to run tests to answer the above questions. I used ProDiscover and FTK to verify the following results:

Question	Result
Does file slack accompany a file when it is emailed?	No (file contents copied, slack from recipient disk)
Does file slack accompany a file when it is copied to another disk or media (hard disk, USB, CDROM, etc.)?	No (file contents copied, slack from recipient disk)
Does file slack accompany a file when FTP is used to copy the file to another location?	No*
Does file slack accompany a file when the file is saved under a different name?	No (file contents copied, slack from new portion of disk)
Does file slack from a file on a file server accompany the file when it is sent to a different location?	Untested**
When copying a file from a file server to a PC, does the recipient's OS add file slack from the recipient's hard disk or file server's hard disk?	Recipient's hard disk
Does file slack occur on a CDRW?	Yes (contents from previously deleted files were found in file slack)

*An secure FTP client was not used. ** Examining files and file slack on a source file server and destination file server to compare the contents of file slack was beyond the scope of this investigation.

I invite readers to verify these findings and provide feedback.

5. FUTURE RESEARCH

Clearly this study is only the beginning of necessary research on file slack and its security implications. During a literature search, an article mentioned that versions of Linux using the ext2 file system can have file slack: "if the file is removed by /bin/rm, its content still remains on disk, unless overwritten by other files," and mentions the use of an obscure tool called bmap that can insert data into the slack space of files. (Chuvakin, 2002).

Other areas that need closer inspection is MAC's HFS (Hierarchical File System) and HFS+, the various "flavors" of Unix, transferring files with a secure ftp client, file slack in the MFT, etc.

REFERENCES

Casey, E (2004) Digital Evidence and Computer Crime, Second Edition, Academic Press, p 301.

Chuvakin A. (2002) LinuxSecurity.com, available from www.linuxsecurity.com/content/view/117638/49/, 2002, (accessed 15 April 2009).

Dampier, D. (2008) Introduction to Cyber Crime and Computer Forensics, at www.cse.msstate.edu/~dampier/CSE6273/Slides/CSE6273-Intro-2.ppt, (accessed 18 December 2008).

Garfinkel, S., and Shelat, A (2003) Remembrance of Data Passed: A Study of Disk Sanitization Practices, *IEEE Xplore*, at http://computer.org/security, (downloaded 4 December 2008).

Jones, A., Valli, C., Sutherland, I., and Thomas, P (2006) The 2006 Analysis of Information Remaining on Disks Offered for Sale on the Second Hand Market, *Journal of Digital Forensics, Security and Law*, Vol. 1:3.

Jones, A., Valli, C., Dardick, G., and Sutherland, I. (2008) The 2007 Analysis of Information Remaining on Disks Offered for Sale on the Second Hand Market, *Journal of Digital Forensics, Security and Law*, Vol. 3:1.

Microsoft WinHec 2004. Local File Systems for Windows. Available from http://download.microsoft.com/download/5/b/5/5b5bec17-ea71-4653-9539-204a672f11cf/LocFileSys.doc, (accessed 15 April 2009).

Microsoft Support Article ID: 140365 Revision: 5.3 (2007) Default cluster size for FAT and NTFS, at http://support.microsoft.com/kb/140365, August 22, 2007, (accessed 15 December 2008).

NetApplications (2008) Top Operating System Share Trend, http://marketshare.hitslink.com/os-market-share.aspx?qprid=9, (accessed 17 December 2008).

New Technologies, Inc. (2001a) File Slack Defined, at www.forensics-intl.com/def6.html, (accessed 17 November 2008).

New Technologies, Inc. (2001b) Sectors, handout received during forensics training 2003.

Niccolai, J (2008) Ballmer Still Searching for an Answer to Google, *IDG News Service*, at http://www.pcworld.com/businesscenter/article/151568/ballmer_still_searching_for_an_answer_to_google.html, September 26, 2008, (accessed 15 December 2008).

Reyes, A. and Wiles, J. (2007) Best Damn Cybercrime and Digital Forensics Book Period, Syngress, p 495.

RAIDS.co.uk (2008) RAID 5, at http://www.raids.co.uk/raid_5.htm, (accessed 21 December 2008).

Vacca, J. (2002) The Essential Guide for Storage Area Networks, Prentice Hall.

Vacca, J. (2005) Computer Forensics: Computer Crime Scene Investigation, Charles River Media, p 244.

Volonino, L., Anzaldua, R., and Godwin, J (2006) Computer Forensics Principles and Practices, Pearson.

Presentation: Data Hiding Tools for Digital Forensics Experts

Abbas Cheddad
School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster at Magee, BT48 7JL
Northern Ireland, United Kingdom
Emails: cheddad-a@email.ulster.ac.uk

Joan Condell
School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster at Magee, BT48 7JL
Northern Ireland, United Kingdom

Kevin Curran
School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster at Magee, BT48 7JL
Northern Ireland, United Kingdom

Paul Mc Kevitt
School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster at Magee, BT48 7JL
Northern Ireland, United Kingdom

ABSTRACT

Much research has been done in the area of steganography which is the science of concealing data in a transmission medium in such a way that it would not draw the attention of eavesdroppers. Steganography has various useful applications such as for human rights organizations (since encryption is prohibited in some countries), smart IDs where individuals' details are embedded in their photographs (content authentication), data integrity by embedding checksum, medical imaging and secure transmission of medical data to name a few. For decades people strove to develop innovative methods for secret communication. The majority of existing techniques suffer from intolerance to any kind of geometric distortion applied to the stego-image. For instance, if rotation or translation occurs, all of the hidden data will be lost. A remedy to this problem could be achieved through incorporating computer vision into the process as proposed in this short briefing paper.

Keywords: Digital Image Steganography; Self-embedding; Dithering; Security; CCTV; Forensics

1. BACKGROUND

The standard and concept of "What You See Is What You Get (WYSIWYG)" which we encounter sometimes while printing images or other materials, is no longer precise and would not fool a Steganographer as it does not always hold true. Images can be more than what we see with our Human Visual System (HVS). Historically, the forgery of a document occurred mechanically, however, since the recent boost in communication technology, the massive increase in database storage and the

introduction of the concept of e-Government; documents are more often being stored digitally. This goes hand in hand with the aim of the paperless workspace, but it does come at the expense of security breaches especially if the document is transmitted over a network. Forgery is a worry for a range of organisations such as Governments, Universities, Hospitals and Banks. To summarise, the recent digital revolution has facilitated communication, data portability and on-the-fly manipulation. Unfortunately, this has brought along some critical security vulnerabilities that put digital documents at risk.

2. PROBLEMS

This short briefing paper unveils two novel systems: one to combat digital document forgery and the other a secure storage system for confidential data.

1) **Forged Documents**: In July 2005 it was discovered that a number of Second World War files held at the National Archives contained forged documents. An internal investigation found that the forgery took place during or after the year 2000. Also recorded CCTV video frames will not stand up in court as reliable evidence since they are prone to tampering.

2) **Loss of Confidential Data**: During 2008 there have been large scale losses of personal sensitive data, e.g. the loss of 25 million child benefit records after HMRC sent two discs to the National Audit Office.

3. SOLUTIONS

The concept behind our research stems from advanced research into the strengthening of digital steganography in digital imaging (Cheddad et al. 2009). Steganography is defined as the science of hiding or embedding

"data" in a transmission medium. Our proposed system will use steganography techniques for frames self-embedding and will also include additional bytes (metadata) that will be useful for query purposes such as: unique reference, date and time stamp, officer name, officer number, location, operation details.

1) **Combating Forgery using Self-Embedding:** We propose an information hiding approach to scanned document forgery detection and correction which is secure, efficient and robust to various image attacks. It is a novel method that allows documents to remain unchanged after any forgery attack. The payload, which is a dithered version of the cover, has a low bit rate while capturing the main image characteristics needed for reconstruction. This payload is encrypted using a key to generate a balanced bit version which provides a balanced visual effect.

2) **Securing Confidential Data**: Using the aforementioned method we can embed unrelated data in the cover, i.e., confidential data, so as to hide its very presence. If the feature is visible the point of attack is evident. To cope with the limited space, frames in a video can be used for embedding.

4. CONCLUSION

Research and development is currently being carried out to produce a workable prototype for secure ID cards. The aim is to increase the security level for data authentication with the least possible cost. Another application that we are investigating is a secure system for CCTV camera image frame authentication (see Figure 1). Meetings with international companies working in security are ongoing with positive feedback being given. Currently two patent applications have been filed.

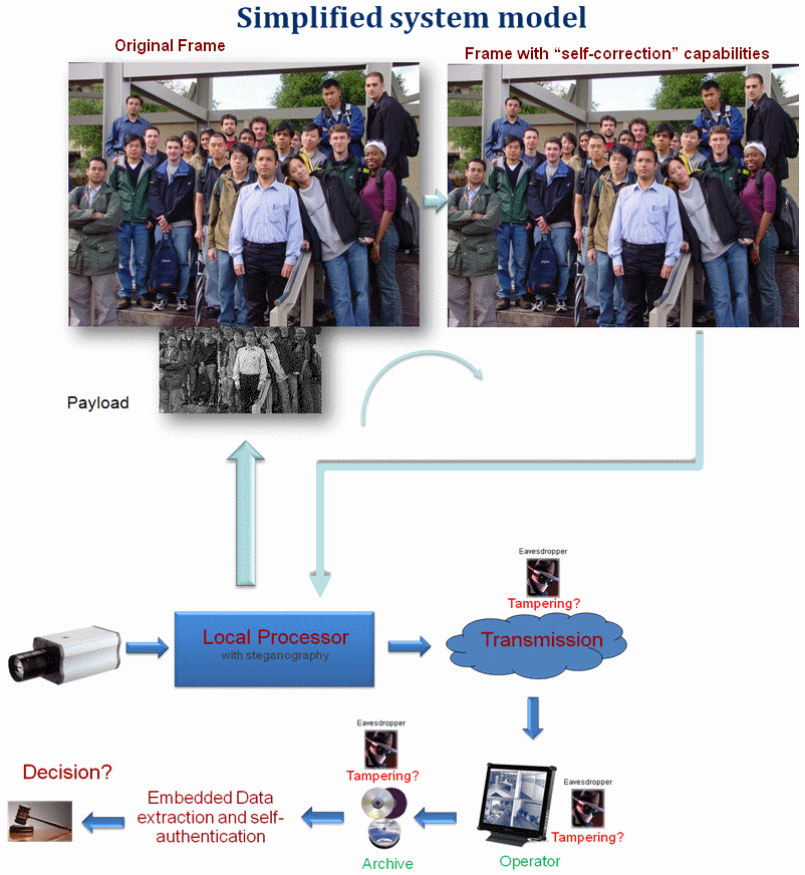


Fig. 1. Authentication of CCTV image frames.

Cheddad, A., Condell, J.V., Curran, K.J. and McKevitt, P. (2009), "A secure and improved self-embedding algorithm to combat digital document forgery," In press, Signal Processing, doi:10.1016/j.sigpro.2009.02.001.

Bluetooth Hacking: A Case Study

Dennis Browning
Champlain College Center for Digital Investigation
Burlington, Vermont
dennisbrowning@gmail.com

Gary C. Kessler
+1 802-865-6460
Champlain College Center for Digital Investigation
Burlington, Vermont
Edith Cowan University
Perth, Western Australia
gary.kessler@champlain.edu

ABSTRACT

This paper describes a student project examining mechanisms with which to attack Bluetooth-enabled devices. The paper briefly describes the protocol architecture of Bluetooth and the Java interface that programmers can use to connect to Bluetooth communication services. Several types of attacks are described, along with a detailed example of two attack tools, Bloover II and BT Info.

Keywords: Bluetooth hacking, mobile phone hacking, wireless hacking

1. INTRODUCTION

Bluetooth (BT) is one of the newer wireless technologies in use today. The name derives from that of Harald Blaatand, a tenth-century king of Denmark and Norway who united many independent Scandinavian tribes into a single kingdom. Bluetooth wireless communication technology is meant to be a universal, standard communications protocol for short-range communications, intended to replace the cables connecting portable and fixed electronic devices (Bluetooth SIG, 2008a). Operating in the 2.4 GHz range, Bluetooth is designed to allow wire-free communication over a range of short-haul distances in three power classes, namely, short range (10-100 cm), ordinary range (10 m), and long range (100 m) (Sridhar, 2008). Cell phones, personal digital assistants (PDAs), and smart phones are a few of the devices that commonly use Bluetooth for synchronizing email, sending messages, or connecting to a remote headset (Mahmoud, 2003a). What are less well known to users of Bluetooth devices are the risks that they incur due to various vulnerabilities of the technology. Bluehacking, bluejacking, marphing, bluesniping, and bluesnafting are just a few of the names given to the act of hacking a device via Bluetooth (Laurie, Holtmann, & Herfurt, 2006). In this paper, we will discuss the technology needed to hack a cell phone, some of the tools, and precautions that users can take to help protect their Bluetooth devices.

2. TECHNOLOGY

Figure 1 shows a diagram of the Bluetooth protocol stack in order to show the various attack vectors. The protocol layers of particular interest in this paper are:

- Logical Link Control and Adaptation Protocol (L2CAP): Provides the data interface between higher layer data protocols and applications, and the lower layers of the device; multiplexes multiple data streams; and adapts between different packet sizes (Hole, 2008a, 2008d; Sridhar, 2008).

- Radio Frequency Communications Protocol (RFCOMM): Emulates the functions of a serial communications interface (e.g., EIA-RS-232) on a computer. As Figure 1 shows, RFCOMM can be accessed by a variety of higher layer schemes, including AT commands, the Wireless Application Protocol (WAP) over the Transmission Control Protocol/Internet Protocol (TCP/IP) stack, or the Object Exchange (OBEX) protocol (Hole, 2008a, 2008e; Sridhar, 2008).

- Object Exchange protocol: A vendor-independent protocol allowing devices to exchange standard file objects, such as data files, business cards (e.g., vCard files), and calendar information (e.g., vCal files). OBEX is a higher layer application and runs over different operating systems (e.g., PalmOS and Windows CE) and different communications protocols (e.g., Bluetooth and IrDA) (Gusev, n.d.).

Most of the tools that are being used to hack Bluetooth phones use the Java programming language. In order for the software to work, the phone that is used to initiate the attack needs to support JSR-82, which is the official Java Bluetooth Application Programming Interface (API) (JCP, 2009). If the attacker's phone does not support JSR-82, that phone cannot be used to attack other phones. This is an important note because although Bluetooth is widely available on cell phones, Java and JSR-82 support may not be.

JSR-82 consists of two packages, namely, javax.bluetooth, which is the core Bluetooth API, and javax.obex, which is independent of the Bluetooth stack and provides APIs to other protocols, such as OBEX. The capabilities of JSR-82 include the ability to (Hole, 2007; Mahmoud, 2003b):

- Register services
- Discover devices and services
- Establish L2CAP, RFCOMM, and OBEX connections between devices, using those connections to send and receive data (voice communication is not supported)
- Manage and control the communication connections
- Provide security for these activities

Hole (2008a, 2008f) and Mahmoud (2003b) provide good overviews of how this code functions.

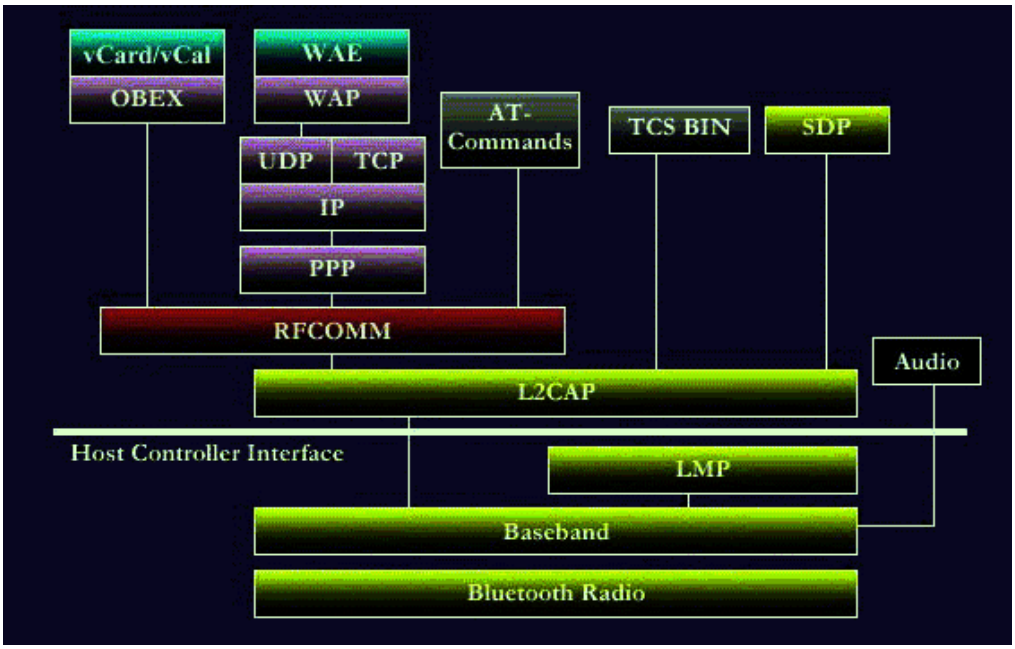


Figure 1: Bluetooth protocol stack (Source: *Tutorial-Reports.com, n.d.*)

3. BLUETOOTH SECURITY

Bluetooth defines three security modes. Security Mode 1 provides no security enforcement, meaning that the device is effectively taking no steps to protect itself. Security Mode 2 enforces security at the service level. In this mode, a particular application might be relatively safe but no additional device protection has been added. Security Mode 3 is the highest level of security, employing link level enforced security mechanisms. Security Mode 3 protects the device from certain intrusions and, therefore, all services and applications (Bluetooth SIG, 2008b; Hole, 2008b; Laurie et al., 2006).

All Bluetooth services have a default set level of security. Within the service level security, there are also three levels of security. Some services that require authorization and authentication in order to be used, some require authentication only, and some are open to all devices (Bluetooth SIG, 2008b). Bluetooth devices themselves have two levels of security when describing other devices, namely trusted devices and untrusted devices.

4. TYPES OF ATTACKS

There are a variety of attacks that can be employed against Bluetooth devices, many with colorful names such as bluebugging, bluebumping, bluedumping, bluejacking, bluesmacking, bluesnarfing, bluespooofing [sic], bluestabbing, bluetoothing, and car whisperer. All take advantage of weaknesses in Bluetooth that allow an attacker unauthorized access to a victim's phone. It is imperative to note that while Bluetooth is commonly associated with networks limited in scope to 100 m, attacks on Bluetooth devices have been documented at ranges in excess of 1,500 m. using Bluetooone [sic] (Laurie, 2006).

One common approach to hacking Bluetooth devices is to employ malformed objects, which are legal files exchanged between BT devices that contain invalid information, thus causing unexpected results. When a Bluetooth device receives a malformed object, such as a vCard or vCal file, the device may become unstable or fail completely. Alternatively, an attacker might also use a vCard or vCal file to

inject commands allowing the attacker to take control of the device. This kind of an attack can be very harmful to a phone (E-Stealth, 2008; Laurie et al., 2006).

Some of the common attacks on Bluetooth devices include:

- *Bluebugging:* An extraordinarily powerful attack mechanism, bluebugging allows an attacker to take control of a victim's phone using the AT command parser. Bluebug allows an attacker to access a victim's phone in order to make phone calls, send short message service (SMS) messages, read SMS messages stored on the phone, read and write contact list entries, alter phone service parameters, connect to the Internet, set call forwarding, and more (Bluebugging, n.d.; Laurie et al., 2006).

- *Bluejacking:* The sending of unsolicited messages to open Bluetooth devices by sending a vCard with a message in the name field and exploiting the OBEX protocol (Bluejacking, 2009).

- *Bluesmack:* A Bluetooth analog of the Ping-of-Death denial-of-service attack. This is a buffer overflow attack using L2CAP echo messages (Bluesmack, n.d.; Laurie, 2006).

- *Bluesnarf and Bluesnarf++:* Attacks allowing for the theft of information from a Bluetooth device using the OBEX Push Profile. The attacker needs only find a phone that has Bluetooth in discoverable mode. Bluesnarf works by a connection to most of the Object Push Profile services and the attacker retrieves the file names of known files from the Infrared Mobile Communications (IrMC) list instead of sending vCard information as expected. With these attacks the hacker can retrieve items such as the phonebook, calendar, and other personal information. With Bluesnarf++, the attacker has full read and write access to the file system of the phone. When an attacker is connected via the OBEX Push Profile, he/she has full access to the victim's phone without having to pair the two devices. The biggest risk with this function is that an attacker can delete crucial file system files, rendering the victim's device useless. In addition, the attacker can access any memory cards that are attached to the device (BlueSnarf, n.d.; Bluesnarfing, n.d.; Laurie et al., 2006).

- *Helomoto:* Helomoto is functionally similar to the Bluebug attack but takes advantage of poor implementations of "trusted device" handling on some phones. As in bluebug attacks, the attacker pretends to send a vCard to an unauthenticated OBEX Push Profile on the victim's phone. Once started, the attacker interrupts the transfer process and the victim then lists the attacker's phone as a trusted device. The attacker can then connect to the victim's phone and take control of the device by issuing AT commands. This attack is so-named because it was first discovered on Motorola phones (Helomoto, n.d.; Laurie et al., 2006).

These attacks are only a few that can be launched against Bluetooth interfaces in phones, laptops, and even automobiles. E-Stealth (2008) and Laurie et al. (2006) offer information about a wide range of attacks that can be launched via Bluetooth vulnerabilities.

5. TOOLS FOR ATTACK

There are many options that a user can choose from when looking to attack a Bluetooth phone. Web sites such as E-Stealth (http://www.e-stealth.com/) and FlexiSPY (http://www.flexispy.com/) offer commercial products to allow one party to eavesdrop or attack another party's Bluetooth device, ostensibly to trap an unfaithful spouse, catch an unscrupulous employee, or monitor a teenage child. These are merely commercial versions of hacker tools that include Bloover, Bloover II, BT Info, BT_File_Explorer, ISeeYourFiles, MiyuX, and STMBlueS (D3scene, 2008; E-Stealth, 2008; Getjar,

2008; Laurie et al., 2006; SE-NSE, 2006). Many of these programs (like so many hacker tools such as Back Orifice and SubSeven), are distributed as "management tools" but what differentiates them from bona fide management tools is that the managed party may not be aware that the program is running. And, like any "management" tool, these programs are often platform-dependent so that they work best on certain brands of devices and may not work on all devices; MiyuX, for example, works best on Sony Ericsson phones. A nice collection of all of these tools in one package can be found at tradebit (http://www.tradebit.com/filedetail.php/5006527-basic-bluetooth-spy-software).

5.1 Testing the Software

The first author experimented with the feasibility of actually using this software in a real environment, employing Bloover II (which allows an attacker to obtain information from a victim's phone) and BT Info (which allows an attacker to control the victim's phone). Both were part of the Ultimate Bluetooth Mobile Phone Spy Software New Edition 2008 available from E-Stealth (http://www.e-stealth.com/).

It is worth noting that this software claims to be useable on any Bluetooth phone to hack any other Bluetooth phone but, like so many software claims, this one was overstated. Initial attempts to use the software on a Sanyo SCP-7050 failed because the software could not be installed. Later, the first author purchased a BlackBerry Curve. Although the software user guide provided instructions on how to install the software on a BlackBerry, the install failed when an error stated that the phone did not support the correct Java API.

The phones that were used successfully for testing throughout this project were United Kingdom versions of a Sony Ericsson W550i and a W800i. Theses phone both support JSR-82 enabling them to run the software. In order to actually use the phones, a Subscriber Identity Module (SIM) card was needed for each phone. The SIM card does not actually need to be active if the attacker is only going to be probing and manipulating the target phone and not making calls. Throughout the testing for this project both phones used inactive SIM cards.

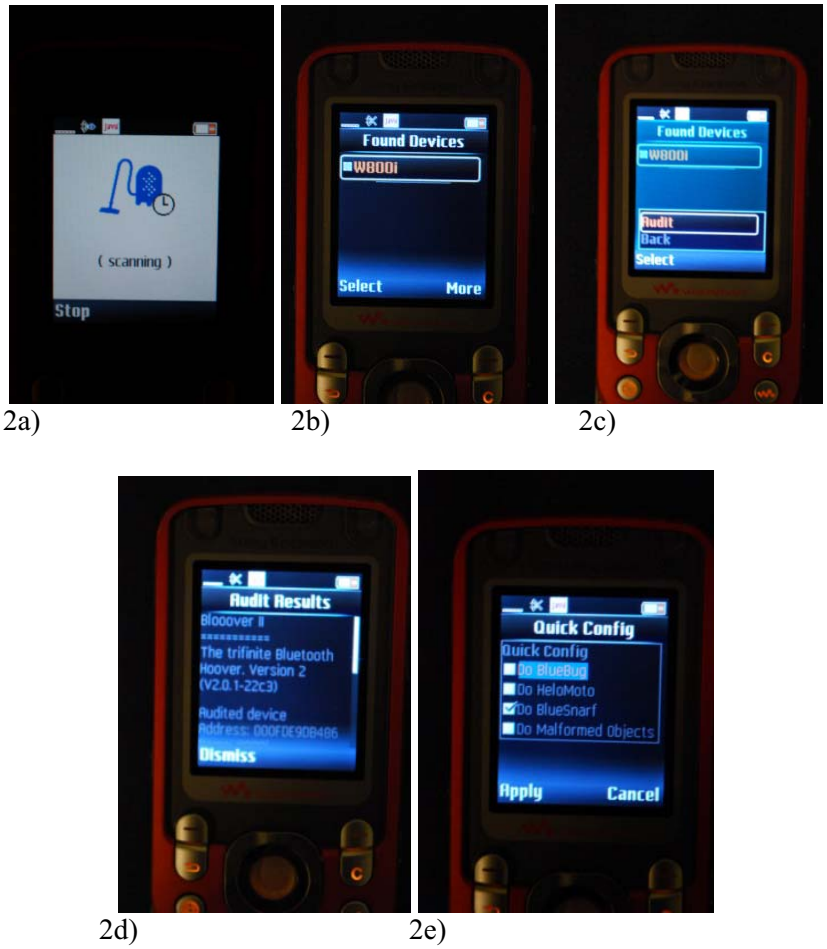
5.2 Bloover II

Bloover (also known as Blooover), standing for Bluetooth Wireless Technology Hoover, is a proof-of-concept application. Bloover II is a second-generation version of a program that consists of several different types of attacks, including Bluebug, Bluesnarf, Helomoto, and the use of malformed objects. Breeder is a related program that propagates Bloover II clients (Laurie et al., 2006).

The attack software package that was purchased included a program called Bloover II. Once a JSR-82 enabled phone was found, the program installed easily. As for running the program, it seemed to always halt on one of the processes. One of the processes that the software kept halting on was when the program was running the "HeloMoto" attack. During this attack, the hacking phone tries to "plant" an entry into the victim's phonebook. Within the options of the Bloover II program, the hacker can chose which attacks they would like to use on the victim's phone. When going through and trying each attack by itself, the software would always halt on some process. The only operation that could be conducted was the initial audit of the phone to get basic information about the phone.

Figure 2 shows a series of screen shots using Bloover II from a W550i phone to access a W800i phone. Figure 2a shows the attacker's phone scanning for another Bluetooth phone; in Figure 2b, a device named W800i is found. The audit feature of Bloover is initiated (Figure 2c) and results (Figure 2d) include the target device's address, communications channel for communication with the headset and other functional profiles, the RFCOMM channel, and phone contact information. A specific attack

type (Bluebug in this case) is selected from the Quick Config menu (Figure 2d).



2a) 2b) 2c)

2d) 2e)

Figure 2. Bloover II screen shots.

5.3 BT Info

Because of increased functionality, a larger amount of time was spent using a program called BT Info. With this program, the attacker can completely control the target device *if* the attacker can become paired with the target. Once the Bluetooth pairing takes places, the attacker can perform a broad set of functions on the target phone, ranging from placing a phone call or sending an SMS message to turning the phone off or performing a master reset. The hardest part for the attacker, in fact, is finding a device with an open Bluetooth connection or tricking someone into pairing his or her phone.

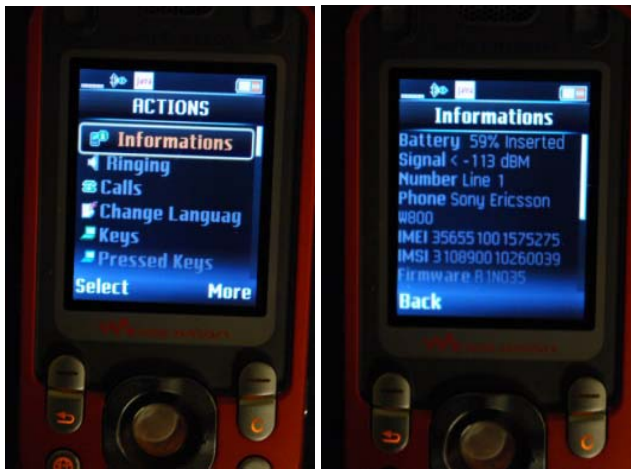
3a) 3b)

3c) 3d)

Figure 3. BT Info screen shots (device pairing).

Figure 3 shows a series of screen shots of an attacker's phone (W550i) pairing up with a target phone (W800i). Once pairing has been successfully accomplished, BT Info displays a menu of possible actions (Figure 4a). The Informations screen (Figure 4b) allows the attacker to retrieve basic information about the target phone, such as the phone manufacturer and model, firmware version, battery level, signal level, International Mobile Equipment Identity (IMEI), and International Mobile Subscriber Identity (IMSI).

The Ringing screen (Figure 4c) allows the attacker to control the ringing on the target phone. This option allows the attacker to force the target phone to start ringing and not stop until the target phone is turned off or the attacker issues the *Stop* command. Within the Ringing option, the attacker is able to select the type of ringtone to start.

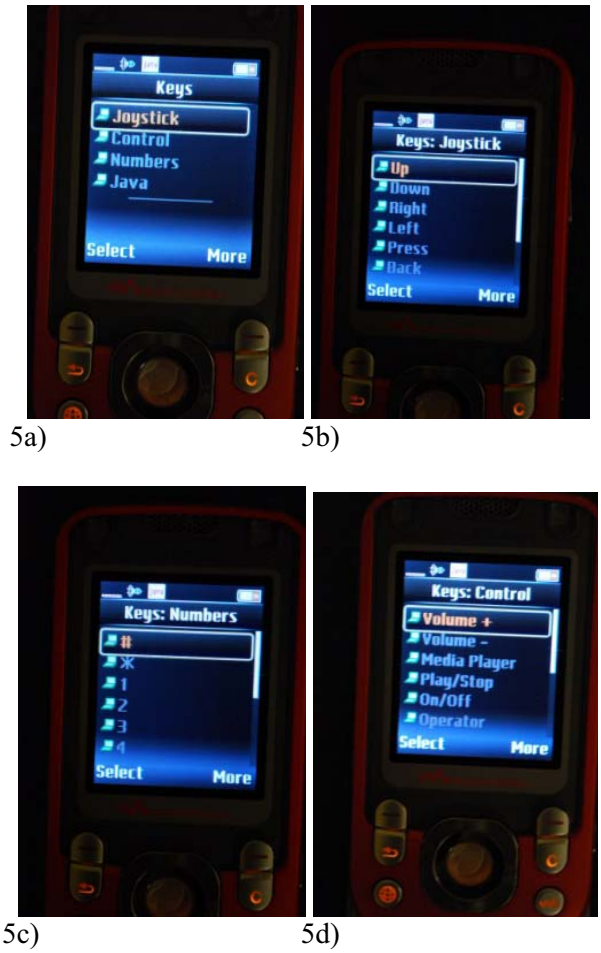


4a) 4b)

4c) 4d)

Figure 4. BT Info screen shots (initial menu functions).

The Calling menu (Figure 4d) offers four options, allowing the attacker to dial any number, hang up a call, place a current call on hold, or redial the last number. An attacker can use the Calling option, for example, to call a second phone owned by the attacker in order to listen in on the victim's conversations. If the target phone has a speaker function that operates when the phone is closed, the attacker can still be able to establish a call and listen in. From the main Actions menu, the attacker can also change the display language that the phone uses.



5a) 5b)

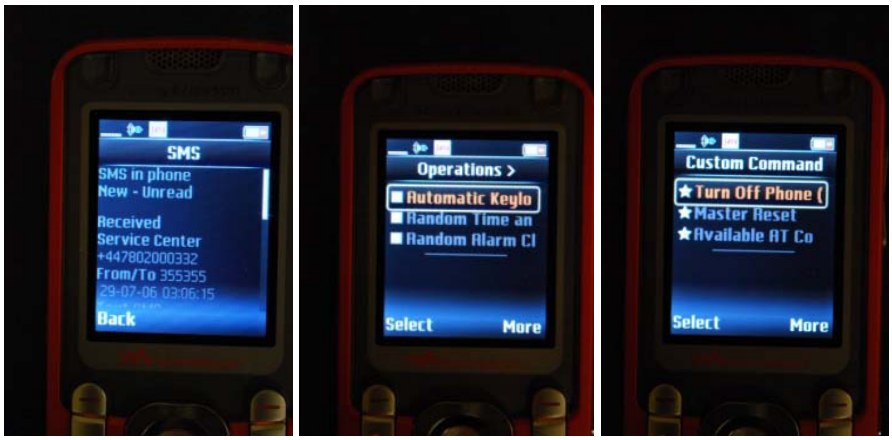
5c) 5d)

Figure 5. BT Info screen shots (Keys functions).

The Keys function (Figure 5a) is a feature of BT Info that allows an attacker to watch the keys that the victim pushes as they are being pushed or allows an attacker to remotely press keys on the victim's phone. For the latter function, the attacker can access the target phone's "joystick" keys (Figure 5b) or individual keypad keys (Figure 5c). The control function of BT Info (Figure 5d) allows the attacker to remotely access the target's control keys, including volume control, media player, and camera.

BT Info also gives an attacker access to the target phone's text messages. The SMS action (Figure 6a), for example, allows the attacker to select a mailbox on the victim's phone and retrieve the complete contents of all SMS messages. Some of the other actions are simply informational, including the temperature of the phone, what Bluetooth devices are trusted on the victim's phone, what sound, if any, the phone makes when a button is pressed, the memory status, and what action forces a keylock.

The Operations action (Figure 6b) has several options. Automatic Keylock gives an attacker the ability to automatically lock the victim's when it is unlocked; i.e., when the victim unlocks the phone, it will automatically relock itself. The Random Time and Date Change option randomly changes the date and time on the victim's roughly a hundred times per minute. Similarly, the Random Alarm option randomly sets the victim phone's alarm settings.



6a) 6b) 6c)

Figure 6. BT Info screen shots (miscellaneous).

The Custom Command function (Figure 6c) allows an attacker to power down or force a master reset on a victim's phone. This function can also be used to execute whatever AT commands are available on the target phone. BT Info also has a Phonebook function that allows an attacker to read the victim's phonebook and recent call history.

BT Info was tested using several different Bluetooth phones and was employed most successfully between the two Sony Ericsson phones mentioned above. The first author was able to use one of the Sony Ericsson phones to connect with a Motorola Razr, although the functionality of BT Info was somewhat limited, only allowing call initiation and access to SMS messages. Functionality of BT Info will vary by the model of both attacker and target phone (E-Stealth, 2008).

A video of the first author using BT Info between the two Sony Ericsson phones can be found at http://c3di.champlain.edu/TR/BTInfo_Browning.m4v (11 minutes, 350 MB).

6. PRECAUTIONS

As with so many aspects of security, user awareness and vigilance is the best defense against the kinds of attacks described here. The best way to protect a device, obviously, is to simply turn Bluetooth off. A device cannot be hacked via a Bluetooth attack vector if other Bluetooth devices cannot see it. Some devices come with Bluetooth turned on by default so users need to check this setting.

If Bluetooth must be enabled, the user can set the device to be hidden (analogous to not broadcasting the network name on a wireless network). Setting a device to be invisible will still allow Bluetooth communications to function but will only allow connections to trusted devices that have been previously configured. This protection is not perfect, however; if an attacker finds out that a particular device is trusted, they can use their own Bluetooth device to masquerade as the trusted device and will then be able to connect to the target phone (this is a common spoofing attack).

If a user must use Bluetooth, they should also only turn it on as needed. In addition, users should change their Bluetooth personal identification number (PIN) every month or so. Changing the PIN requires that any Bluetooth devices that the user regularly employs will need to be re-paired, but it also makes it a bit harder for attackers. Attacks succeed because many users will balk at constantly turning their Bluetooth port on and off, or changing the PIN, but at the very least users should change the default PIN when they first get their Bluetooth enabled device (Jansen & Scarfone, 2008).

7. CONCLUSION

The intent of this project was to determine how real the threat is of attacks to Bluetooth-enabled devices and how easy such attacks are to launch. After spending a relatively short amount of time and a few dollars, it is clear how vulnerable Bluetooth technology really is. The idea that someone could listen to all conversations a victim is having without them even knowing, or have their text messages read, are key examples of the dangers of Bluetooth. Even worse, an attacker can initiate a call to someone or text someone without the victim ever knowing. The only way a user would be able to catch this activity is if they were to look through their call log or look at the sent messages on their phone. Even that might be insufficient, as the attacker can delete the records of their nefarious activity and the victim would never know until their bill comes out. The victim would only know about unusual behavior if they carefully look at their bill, which is increasingly problematic since many people do not even look at their detailed call records. And even if someone complains that they "did not make a call on this date and time," the mobile service carrier has proof that the call was made from this device because, indeed, it was.

Users need to be made aware of the vulnerabilities of these devices so that they can employ them more effectively, safely, and confidently.

8. ACKNOWLEDGEMENTS

This work was 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 State Department of Justice.

9. AUTHOR INFORMATION

Dennis Browning received his B.S. degree in Computer & Digital Forensics from Champlain College in May 2009 and currently works in the Information Technology Department at Fletcher Allen Health Care in Burlington, Vermont.
Gary C. Kessler, Ed.S., CCE, CISSP, is an Associate Professor, director of the M.S. in Digital Investigation Management program, and principle investigator at the Center for Digital Investigation at Champlain College. He is also an adjunct associate professor at Edith Cowan University in Perth, Western Australia.

10. REFERENCES

Bluebugging. (n.d.). trifinite.stuff Web site. Retrieved January 27, 2009, from http://trifinite.org /trifinite_stuff_bluebug.html

Bluejacking. (2009, January 6). Wikipedia. Retrieved January 27, 20909, from http://en.wikipedia.org/wiki/Bluejacking

Bluesmack. (n.d.). trifinite.stuff Web site. Retrieved January 27, 2009, from http://trifinite.org /trifinite_stuff_bluesmack.html

BlueSnarf. (n.d.). trifinite.stuff Web site. Retrieved January 27, 2009, from http://trifinite.org /trifinite_stuff_bluesnarf.html

Bluesnarfing. (n.d.). Bluejacking Tools: The Biggest Collection of Bluetooth Tools on the Internet Web site. Retrieved January 27, 2009, from http://www.bluejackingtools.com /bluesnarfing/

Bluetooth SIG. (2008a). How Bluetooth Technology Works. Bluetooth.com Web site. Retrieved January 6, 2009, from http://www.bluetooth.com/Bluetooth/Technology/Works/

Bluetooth SIG. (2008b). Security. Bluetooth.com Web site. Retrieved January 6, 2009, from http://www.bluetooth.com/Bluetooth/Technology/Works/Security/

D3scene. (2008, April 30). BTInfo. Retrieved January 29, 2009, from http://www.d3scene.com /forum/general-mp/13279-btinfo.html

E-Stealth.com. (2008). *Ultimate Bluetooth Mobile Phone Spy Software User Manual*. Retrieved January 29, 2009, from http://www.jamsa.us/inventory/UltimateMobilePhoneSpyManual.pdf

Getjar. (2008, March 10). STM Bluetooth Software and Tools. Retrieved January 29, 2009, from http://www.getjar.com/products/8042/STMBlueS

Gusev, A. (n.d.). Object Exchange (OBEX) Protocol Primer. Developer.com Web site. Retrieved January 29, 2009, from http://www.developer.com/ws/article.php/3573636

Helomoto. (n.d.). trifinite.stuff Web site. Retrieved January 27, 2009, from http://trifinite.org /trifinite_stuff_helomoto.html

Hole, K.J. (2007, March 2). Bluetooth -- Part 3: Link Controller and JSR-82 API Architecture. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF /Bluetooth3alt.pdf

Hole, K.J. (2008a, February 24). Bluetooth -- Part 1: Overview. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF/Bluetooth1alt.pdf

Hole, K.J. (2008b, March 8). Bluetooth -- Part 10: Introduction to Wireless Security. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF/Bluetooth10alt.pdf

Hole, K.J. (2008c, March 8). Bluetooth -- Part 4: Link Manager and J2ME Programming. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF /Bluetooth4alt.pdf

Hole, K.J. (2008d, March 11). Bluetooth -- Part 6: Logical Link Control and Adaptation Protocol. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF /Bluetooth6alt.pdf

Hole, K.J. (2008e, March 23). Bluetooth -- Part 7: RFCOMM. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF/Bluetooth7alt.pdf

Hole, K.J. (2008f, March 29). Bluetooth -- Part 8: The JSR-82 API for Device Discovery. Retrieved January 29, 2009, from http://www.kjhole.com/Standards/BT/BT-PDF /Bluetooth8alt.pdf

Java Community Process (JCP). (2009). JSR 82: Java APIs for Bluetooth. Community Development of Java Technology Specifications Web site. Retrieved January 27, 2009, from http://jcp.org/en/jsr/detail?id=82

Jansen, W., & Scarfone, K. (2008, October). *Guidelines on Cell Phone and PDA Security*. National Institute of Standards and Technology Special Publication 800-124. Retrieved February 24, 2009, from http://csrc.nist.gov/publications/nistpubs/800-124/SP800-124.pdf

Laurie, A., Holtmann, M., & Herfurt, M. (2006, March 30). Bluetooth Hacking. WEBSEC 2006, London, U.K. Retrieved January 27, 2009, from http://trifinite.org/Downloads /trifinite.presentation_websec2006.pdf

Mahmoud, Q.H. (2003a, February). Wireless Application Programming With J2ME and Bluetooth. Sun Developer Network (SDN) Web site. Retrieved January 27, 2009, from http://developers.sun.com/mobility/midp/articles/bluetooth1/

Mahmoud, Q.H. (2003b, April). Part II: The Java APIs for Bluetooth Wireless Technology. Sun

Developer Network (SDN) Web site. Retrieved January 7, 2009, from http://developers.sun.com /mobility/midp/articles/bluetooth2/

SE-NSE. (2006, November 5). MiyuX. se-nse v5 Web site. Retrieved January 29, 2009, from http://forums.se-nse.net/index.php?showtopic=5653

Sridhar, T. (2008, December). Wi-Fi, Bluetooth, and WiMAX. *The IP Journal*, *11*(4), 2-17.

Tutorial-Reports.com. (n.d.). Bluetooth Tutorial: Protocol Stack. Retrieved January 28, 2009, from http://www.tutorial-reports.com/wireless/bluetooth/protocolstack.php

Cybercrime and the 2012 London Olympics

Denis Edgar-Nevill
Canterbury Christ Church University
North Holmes Rd, Canterbury CT1 1QU United Kingdom
Tel +44 (0) 1227 782089
Email denis.edgar-nevill@canterbury.ac.uk

ABSTRACT

The London 2012 Olympics is just three years away and the clock is ticking to put in place plans get it right. The potential for cybercrime to cause harm during this event is very great; harm to national reputation, harm to the reputation to the Olympic movement, and harm to individuals competing, watching or officiating. This paper considers the need to address these risks by taking a look at what has happened in the past at sporting events and the rising wave of electronic security threats and fraud facilitated by computers at recent Olympics. The problems for law enforcement are discussed surrounding the need to capture and preserve computer forensics data from such a complex live system. The paper concludes by considering the remaining imponderable factors which remain for groups being established by the UK Government to consider.

Keywords Cybercrime, Forensics, Olympics, London, 2012

1. INTRODUCTION

We now live in a world where every large event, bringing the focus of attention of a worldwide audience of hundreds of millions of people, is a target. Any matters surrounding the preparation, planning, resourcing, logistics, financing, development, building, staffing, organisation and delivery becomes a matter for press and public scrutiny. The Olympic Games is a global phenomenon and at a greater risk than many events. Very high in everyone's consciousness are the consequences of terrorist actions seeking "the oxygen of publicity" (Thatcher 1985) with demonstrations of their ability to cause disruption and harm. The attack on the Twin Towers in New York remains as one of the turning-points in history; summarised for all time by the reference to "9/11".

Sports events have been prone to political demonstrations in the past. In 1913 Miss Emily Wilding Davison, a Suffragist, ran on to the course during the race for the Derby and was knocked down by the King's horse. She died from the effects of the injuries she received in Epsom Cottage Hospital a few days later (Morning Post 2009).

The Olympic Games have not been immune to political demonstrations and attacks leading to loss of life.

2. DISRUPTION AT PREVIOUS OLYMPIC GAMES

Mexico 1968

On October 16, 1968, at the Mexico City Olympics, two African-American sprinters Tommie Smith and John Carlos, (gold and bronze medalists, men's 200 metres), raised black-gloved fists on the podium for the medal ceremony as the US national anthem was played. They were both members of the Olympic Project for Human Rights. The "Black Power" salute led to them both being suspended from the US Olympic team and banned from the Olympic Village. ***It was many years before their athletic achievements were finally honored by the US (Slot 2005).***

Munich 1972

In 1972 at the summer games of the XX Olympiad held in Munich, the terrorist group "Black

September" held hostage members of the Israeli Olympic team. As the direct result of a failed rescue attempt when the terrorists attempted to move the hostages at the end of the siege, eleven Israeli's and one German police officer were killed.

Atlanta 1996

On July 27th 1996 the Games of the XXVI Olympiad in Atlanta, Eric Robert Rudolph, former explosives expert for the United States Army, planted the largest pipe bomb in US history in the Centennial Olympic Park. When it exploded at 01.20 am, two people died and 111 were injured. On August 22, 2005, Rudolph was sentenced to three concurrent terms of life imprisonment without parole for the Georgia incidents; on top of a life-sentence for an earlier bomb attack. Rudolph made a statement when sentenced in which he stated that he was angry at the US Government and hoped the Olympics would be cancelled (Gross 2005).

Beijing 2008

The Summer Games of the XXIX Olympiad took place Beijing, China. This venue was always destined to attract demonstration because of the comparatively closed nature of China up until the last decade and international campaigns against perceived breaches of human rights in that country. The run up to the event was marked by a number of human rights demonstrations focused on the repression of the people of Tibet (Wilson et al. 2009) following the invasion of Tibet by China 50 years ago. Less visible than the protests, but potentially every bit as embarrassing, was the dramatic rise in attempts to attack the IT infrastructure of the Beijing Games.

There have been instances of direct electronic attacks on the websites of Olympic Games for at least 20 years with large numbers of security alerts being recorded. The growth in direct attacks during each day of the Beijing Olympics was staggering. One company reported that during each of the days of the Beijing Games there were, on average, in excess of 12 million security alerts. A better way of visualising this is there were around 140 per second, every minute of every hour of every day the Olympic Games in Beijing were open.

Professor Rongsheng Xu (Network Security Group, Institute of High Energy Physics, Chinese Academy of Sciences, China) led the group providing protection to the Beijing website infrastructure and reported that none of these attacks were successful to any significant degree (Xu 2008).

More important perhaps were the crimes reported relating to ticketing fraud. Despite the Chinese government passing laws to make the resale of tickets for Olympic venues illegal tickets were changing hands on online auctioning sites in their thousands. For example, you could have secured a seat at the Opening Ceremony of the Beijing Olympics for around $26,000 (a little more than 40 times the original face value). That is just the sale of tickets which were once legal. Dozens of fake ticket selling sites existed selling unsuspecting athletics fans fake tickets which they only discovered as fake after having travelled to the events in China. Reports of ticketing fraud exceeded $1.5 billion in Europe alone.

3. OPPORTUNITIES FOR CYBERCIME AT THE LONDON 2012 OLYMPICS

Any large event bringing together millions of people is an opportunity for crime. As well as the potential for people to gain notoriety to further a political cause, cybercrime at a distance, with the potential to reach over a billion possible victims, is fast becoming the preferred method of operations for confidence tricksters and thieves.

In September 2008 the British Government announced the creation of a new Police Central e-Crime Unit based within the Metropolitan Police (PCeU 2009). Part of this new units function is to coordinate computer security measures being put into place for the London 2012 Games. This unit has already begun arranging contracts for commercial organisations to provide electronic infrastructure

and security such as Atos Origin (Atos Origin 2009).

Already we have seen the beginnings of cybercrime relating to the 2012 Olympics. The most widespread type email scam which has been received by hundreds of thousands of people is a variation of the lottery scam (Figure 1). Don't expect to claim your prize anytime soon unless you wish to have your bank details stolen!

Figure 1 - 2012 Olympic Lottery Email Scam

In December 2008 the British Computer Society held the inaugural meeting of a new national specialist group in Cybercrime Forensics (Computer Weekly 2009). The BCS Cybercrime Forensics SG (BCS 2009) will advise the society and the professional community on aspects of crime relating to the 2012 Olympics.

4. CYBERCRIME FORENSICS CHALLENGES FOR LONDON 2012

What will be foremost in law enforcement in the run up to and during the 2012 London Olympics will be the need to ensure information vital to computer forensic investigations is not lost:

- **Forensic Computing Data Acquisition and Storage in Real-Time** – capturing data from a live system involving more than a thousand servers and ten thousand PCs spread over 74 venues. On top of that we must also note the Wifi coverage across venues and the potential interactions with hundreds of different types of mobile devices by sportsmen and women, building contractors, officials and spectators;
- **Acting on Forensic Computing Data Acquired in Real-Time** – analysing, fixing flaws and thwarting attacks which might cause damage and might lead to disruption if left unaddressed which constitute the more visible outward signs of attacks;
- **Data Mining Forensic Computing Data** – after it's all over, going back through information gathered to analyse it and trace criminals back to source for future prosecutions.

Each one of these represents a major challenge for law enforcement.

5. ASSESSING THE RISK

Table 1 gives an assessment of the risk of some of the many hundreds of cybercrime types it might be possible to perpetrate against the 2012 London Olympics. The Games are of such high prestige and importance to host countries that the kernel operations of the event itself is very likely to be unaffected by attempts to disrupt it. Perhaps we should 'never say never' but all of the evidence from recent Olympics suggests that core functions will be well protected.

Of growing concern are the attacks which effect individuals rather than the Games themselves where the risks remain very high. Problems such as how do you ensure a fake ticketing site does not domain squat on a name which looks very like the official ticketing site. Olympic organisers already accept that trying to register very variation of domain name which might be used by criminals is an impossible task.

	RISK ITEM	LIKELYHOOD OF ATTEMPTS	LIKELYHOOD OF SUCCESSFUL ATTEMPTS
MAJOR DISRUPTION DURING THE GAMES	Loss-of-life due to electronic infrastructure misuse	Low?	Very Low
	Suspension of the Games because of systemic electronic infrastructure failures caused by deliberate actions	Very High	Very Low
	Suspension of individual events because of electronic infrastructure failures caused by deliberate actions	Very High	Low
	Loss of coverage for TV Feeds due to misuse or attacks	Medium	Low
	Illegal entry to venues	High	Medium
	Attacks on the Games websites	Very High	Medium
DISRUPTION IN PREPARATION FOR THE GAMES	Fraudulent ticketing using false websites	Very High	Very High
	Online auction merchandising scams using Olympic name	Very High	Very High
	Email scams using the Olympic name	Very High	Very High
LONGER TERM DAMAGE TO REPUTATION	Identity theft using false websites for further fraud	Very High	Very High
	Paedophiles using Olympic blogs, Facebook and other Web 2.0 resources created	Very High	Very High

Table 1: Risks of Cybercrimes at the 2012 Olympics

6. REMAINING IMPONDERABLES

The Games in London are three years away. With reference to Moore's Law, we can reasonably expect the continuing evolution of technology itself to present us with new challenges. Devices will be faster and larger as well as more mobile with greater functionality and connectivity. Developers are playing a continual game of catch up with criminal finding and exploiting weaknesses and flaws in such systems.

What remains true is that, even with the 200,000 hours systems stress-testing the Olympic IT infrastructure being planned by firms such as Altos Origin, it is inconceivable that everything will be perfect.

AUTHOR BIOGRAPHY

Denis Edgar-Nevill holds the post of Head of Department of Computing at Canterbury Christ Church University in the UK. His research includes more than 160 publications and UK and European Union research projects. He is a Fellow of the British Computer Society and member of the BCS Elite group. In 2002 he developed an MSc in Cybercrime Forensics validated with the UK NPIA (National Policing Improvement agency). He chairs the annual CFET (Cybercrime Forensics Education and Training) international conferences and was elected as the founding Chair of the national British Computer Society Cybercrime Forensics Specialist Group in 2008.

REFERENCES

Atos Origin (2009), 'Atos Origin Company website for 2012 Olympics',

http://www.atosorigin.com/en-us/olympic_games/past_future_games/london_2012/default.htm, 18th February 2009

BCS (2009), 'BCS Cybercrime Forensics SG website',

http://www.bcs.org/server.php?show=conWebDoc.23570, 18th February 2009

Computer Weekly (2009) 'BCS Think Tank to Help Protect the 2012 Olympics',

http://www.computerweekly.com/Home/tags/cybercrime-forensics.htm, 18th February 2009

Gross, D. (2005), 'Eric Rudolph lays out the arguments that fueled his two-year bomb attacks', Associated Press, SignonSanDiego.com by the Union-Tribune; April 14, 2005

Morning Post (2009) 'Derby Day Suffragist Incident – Death of Miss Davison, The Morning Post June 9th

1913',http://freepages.genealogy.rootsweb.ancestry.com/~thelamp/sufferage/THE%20MORNING%20POST%20JUNE%209%201913.htm, 18th February 2009

PCeU (2009), 'Police Central e-Crime Unit', http://www.met.police.uk/pceu/index.htm, 18th February 2009

Slot (2005), 'America finally honours rebels as clenched fist becomes salute'. The Sunday Times, UK

Thatcher (1985), Margaret Thatcher – British Prime Minister 1985 Margaret Thatcher Speech

Wilson, S. and Pathitis, N. (2009), 'Tibet Protests Mar Beijing Olympic Plans', ABC News, http://abcnews.go.com/International/WireStory?id=4510954&page=1, 18th February 2009

Xu, R (2008), 'Digital Forensics Research in China', Proceedings of the 2nd International Conference on Cybercrime Forensics Education and Training, Canterbury UK, 1st& 2nd September 2008, ISBN 1899253-19x

Methodology for Investigating Individuals Online Social Networking Persona

Jonathan T. Rajewski
Champlain College
Burlington, Vermont 05403
Jonathan.Rajewski@champlain.edu
jtrajewski@gmail.com

ABSTRACT

When investigators from either the private or public sector review digital data surrounding a case for evidentiary value, they typically conduct a systematic categorization process to identify the relevant digital devices. Armed with the proper methodology to accomplish this task, investigators can quickly recognize the appropriate digital devices for forensic processing and review. This paper purposes a methodology for investigating an individual's online social networking persona.

Keywords: Social Networking, Web 2.0, Internet Investigations, Online Social Networking Community

1. INTRODUCTION TO THE ONLINE SOCIAL NETWORKING COMMUNITY

Online Social Networking Communities (OSNC) are utilized by nearly 45 percent of all active Internet users (Nielsen//NetRatings, 2006), which equates to approximately four hundred sixty million people (Internet World Stats, 2007). The top ten online social networking sites grew nearly 47 percent over the past few years. This trend has been on the rise and doesn't show any signs of declination (Nielsen/NetRatings, 2006).

So what are Online Social Networking Communities? The concept of "Online Social Networking" is not new. When the notion of "Online Social Networking" was combined with "Community" it evolved to what we now know as the concept of "Web 2.0".

Web 2.0 is concept that explains the evolution of how people use the Internet. *Table* 1 contains a comparative example of websites which are considered Web 1.0 and Web 2.0.

Table 1. Web 1.0 vs. Web 2.0 Websites

Web 1.0		Web 2.0
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

(O'Reilly, 2005)

To further describe OSNCs, or its synonymous term Web 2.0's, conceptualize how the Internet revolutionized the world – online banking, search engines, email, instant communication methods, and the list goes on. Now, imagine traditional everyday social interaction – such as: greeting your spouse or co-worker in the morning. Combine the two, and you have the phenomenon known as Web 2.0 or Online Social Networking Communities.

Traditional Online Social Interaction + Community = Web 2.0

2. DIFFERENT TYPES OF ONLINE SOCIAL NETWORKING COMMUNITIES

By design, there are numerous types of OSNCs. Each OSNC provides unique features and opportunities for its users. Below are several popular examples which investigators will typically encounter during an investigation:

1) Blog – Users have the ability to publically (can also be privately) publish their thoughts on a particular topic. Unlike the traditional news article published on physical paper, Blogs are posted on the Internet for everyone or a specific user group to see (Kazakoff, 2009). In the past, Internet users may have created a website to achieve an equivalent goal, but with open source software solutions such as Wordpress (WordPress, 2008) being introduced to the market, users now have a scalable and easily manageable solution to communicate to the masses.
2) Digital Photograph Hosting – Users have the ability to save photos to Internet based repositories for "everyone" or a select group to view. Websites such as Flicker.com (Flickr, 2008)and Photobucket.com (Photobucket, 2008)have created a seamless process for users to share their photos with the world.
3) Video Hosting – Users have the ability to take videos in the real world and save them in an online storage area to ultimately share them with the world. Websites such as YouTube! (YouTube, 2008) and Google Video (http://video.google.com/, 2008)have pioneered the industry and made this process a very scalable and easy function for the end user.
4) Online Collaboration – These websites foster online communication with users across the world. Essentially, users have a unique persona and are encouraged to collaborate with the community. Some examples of these websites are web based message boards, Facebook (Facebook, 2008)and Myspace (Myspace, 2008)

There are a plethora of OSNCs available to explore. The examples listed above have both communal and exclusive features that should be explored in every investigation.

3. HOW TO UTILIZE ONLINE SOCIAL NETWORKING COMMUNITIES

When people decide to join an OSNC, they typically embark on an OSNC selection process. This procedure is typically carried out by either querying an Internet search engine or by learning about a specific OSNC from another person. To describe this further, if an accountant in the real world sought out to join a club or professional organization relating to their industry, they might read a trade magazine or ask a colleague which organizations they belong to. The same is true in the selection process of online communities. If one wishes to join an OSNC that discusses "Microsoft Windows Vista", they might use an internet search engine or a collogues advice to find one.

Typically, upon selecting an OSNC, a user must "register". This process allows users participating in OSNCs to uniquely identify and authenticate themselves. In order for users to successfully complete the registration process, they typically have to provide at least two things to the OSNC:

1) A distinctive username or screen name – This is the "unique identifier" will help differentiate users on in the OSNC;
2) An email address – This is often the "communication" method the OSNC will use with the user. An email address is typically used to authorize the registration process.

In the real world, communities can put more of a focus on credentials or who is authorized to access a particular social group, same is true for OSNCs. That said, once the registration process is complete, a user is permitted to start online collaboration. Also, just as in the real world, when a participant is no longer welcome in a social community, they can be removed just as easily as they were introduced.

4. CONSIDERING ONLINE SOCIAL NETWORKING COMMUNITIES AS DATA SOURCES IN INVESTIGATIONS

Investigations are typically geared to the incident presented to those conducting the investigation. Oftentimes there is a framework for investigations that stays uniform (How Should We Conduct Investigations?, 2007), yet when investigations into OSNCs are conducted, one must take due care even when exigency is presented. For example, if exigency is a factor during an investigation, such as a person's life in immanent danger, the typical process of conducting an investigation is modified based on the needs of the investigation.

In the recent investigation into the death of 12-year-old Vermonter, Brook Bennett, police reported that MySpace, a widely used OSNC, may have been used to arrange a meeting with someone she had been communicating with (Slota, 2008). This example demonstrates that when OSNCs are considered in investigations, additional relevant data can be added to the investigation, which may not have been available via traditional information gathering techniques.

The Brooke Bennett example and others alike are reasons enough to arm investigators with a proven methodology to investigate an individual's online social networking persona.

5. INVESTIGATING ONLINE SOCIAL NETWORKING COMMUNITIES

The Online Social Networking Community is a very large and oftentimes overlooked source for evidence in an investigation. Thousands of OSNCs are available on the Internet to choose from. Investigators must be armed with a systematic and intelligent approach when investigating such data sources.

In reality, a new OSNC can be created by anyone who has the means and access to the internet. Staying current with every new community created would be an unrealistic task for investigators. Therefore, a high level understanding on how these communities work is an essential part of the process.

Due to the nature of how Electronically Stored Information (ESI) is stored on a social networking or online community website, it's critical for investigators to be equipped with the proper knowledge and tools necessary to quickly locate and interrogate the digital data residing on them.

A conceptual understanding of the data available to investigators is key to realizing the vast amount of information available from an OSNC. Typically, there are at least two data sources to consider when investigating an OSNC:

1) The OSNC facing the internet – This is normally what investigators will find after visiting a suspects OSNC. The information found on these pages are active and can be changed at any time.

2) The OSNC subscriber records – This data usually consists of successful user authentication access times with Internet Protocol (IP) addresses and registration email addresses. This data is typically stored by the OSNC and cannot be accessed without court authorization.

6. IDENTIFY IF A SUSPECT ACCESSED/UTILIZED A SOCIAL NETWORKING AND/OR ONLINE COMMUNITY

There are five guidelines that should be followed when investigating an individual's online social networking persona. In essence, these steps will help mitigate the risks associated with an online investigation and identify which steps to take in what order. Some of the steps can be added or subtracted based on the nature of the investigation, but the following are three guidelines that should be followed:

1) Recognize the risks of searching for a personal on an OSNC
2) Ensure that the investigator appears anonymous to the internet
3) Determine which OSNCs to search
4) Probe OSNCs for potential evidence
5) Collect potential evidence

Step 1 – Recognize the risks of searching for a persona on an OSNC

It's imperative, that investigators take the same due care when investigating OSNC's as they do when investigating a suspect in the real world. Keep in mind, that a single visit to an OSNC can compromise the entire investigation and it's critical to ensure anonymity. A gross example of this would be if Law Enforcement drove a marked police cruiser to a drug house to conduct an undercover drug buy. Not

only would the drug dealers refuse to sell the drugs, but also they would most likely move to another house.

Another example for investigators to be conscious of is services that claim to uncover ones online persona. It's critical that the investigator always conduct independent tests prior to using such services. A specific example that, at the time of this publications writing, is the web service called "Yo Name" (http://www.yoname.com/) which advertises the ability to "search across social networks, blogs and more". This service does what it purports to do (*Figure 1*), but it also will notify the person being searched via email.

Figure 1. www.Yoname.com search

Step 2- Ensure that the investigator appears anonymous to the internet

When conducting an investigation into an OSNC, investigators must appear anonymous to the Internet. Just because one is using a work computer and work Internet connection with a fictitious OSNC alias, this does not ensure anonymity.

To ensure anonymity, investigators should seek out the following:

1) Dedicated computer[1] with a normal system configuration[2]
2) Dedicated undercover internet connection[3]

In *figure 2*, you will find a typical network topology diagram that satisfies Step 2.

[1] A "dedicated computer" is a computer that is only used for a predefined task

[2] A "normal system configuration" includes commonly available software and hardware.

[3] A "Dedicated undercover internet connection" is one that was purchased with an undercover identity. Typically these internet connections are Cable/DSL connections. It's becoming more prevalent to purchase mobile carrier "Air Cards" due to their portable nature.

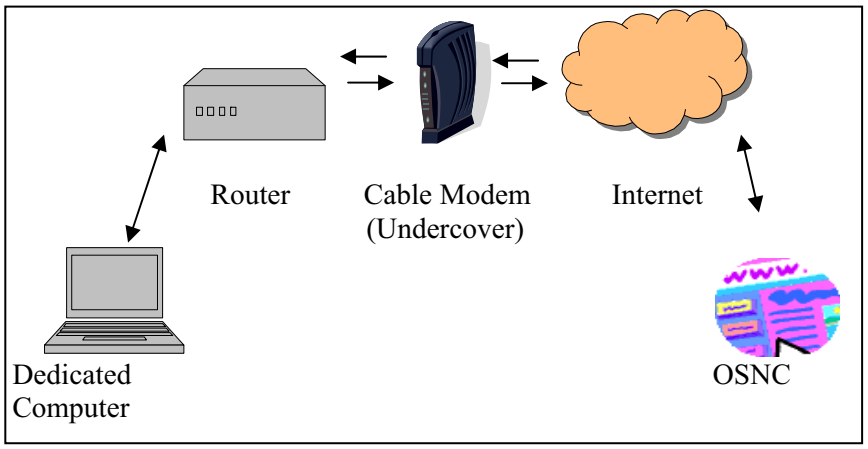


Figure 2 – Ideal network topology

Step 3 – Determine which OSNCs to search

After completing Steps One and Two, investigators will need to begin the process to determine if a suspect has an online persona within am OSNC. Due to the anonymous nature of the Internet, without analyzing the suspect's digital device or knowing their specific persona, this can be a very challenging if not an impossible task. With that being said, it's in the best interest of an investigator to use an educated approach when trying to track down the suspect.

The Top 10 social networking websites would be a good first choice for an investigator to explore (Table 2). The ability for a social networking site to retain its users should also influence your investigative methods (Table 3).

Table 2. Top Social Networking Sites for April 2006

Site	# Unique Visitors April 2006
MySpace	38,359,000
Blogger	18,508,000
Classmates Online	12,865,000
YouTube	12,505,000
MSN Groups	10,570,000
AOL Hometown	9,590,000
Yahoo! Groups	9,165,000
MSN Spaces	7,165,000
Six Apart TypePad	6,711,000
Xanga.com	6,631,000

Source: (Nielsen/NetRatings, 2006)

Table 3. Top 5 Social Networking Sites ranked according to Retention Rate, April 2006

Brand	Retention Rate (%)
MySpace	67.04
MSN Groups	57.62
Facebook	51.73
Xanga.com	48.92
MSN Spaces	47.33

Source: (Nielsen/NetRatings, 2006)

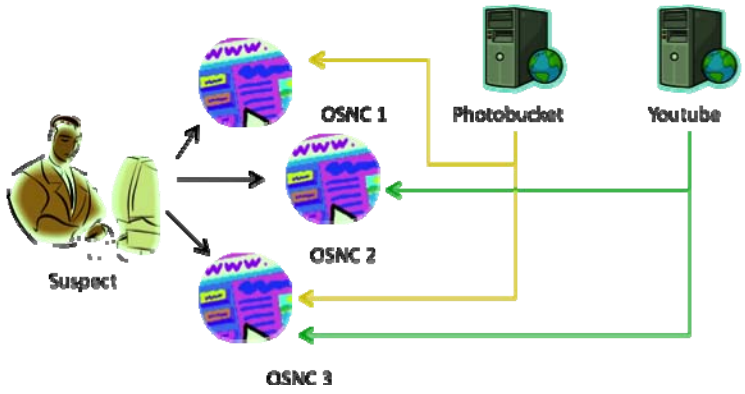
Due to the ever-evolving/changing OSNCs, it's critical to stay mindful of current and up-and-coming social networks. A simple query of international news sites, intelligence groups and/or high school lunchrooms could reveal real time trends and data sources to consider.

Step 4 - Probe OSNCs for potential evidence

Now that the investigator identified the statistically leading websites where suspects may be partaking in online social networking, gathering known unique information about the suspect is critical to this process. In other words, in order to find a particular straw in the haystack, you need to know something about it. For example, if an investigation uncovers the that suspect "Johna Doeet" had a dog named "fluffy", an email address – fluffy1992fluffy@aol.com and a mobile phone number of 212-244-9089 you would have several great keyword combinations to use. For example:

	Keyword
1	Johna
2	Doeet
3	Johna+doeet
4	Fluffy
5	Fluffy1992
6	Fluffy1992fluffy
7	Fluffy1992fluffy@aol.com
8	212-244-9089

Another technique to utilize is when the investigator learns of one OSNC persona and is trying to locate another. Using data from the known persona can aid in the search. For example, leaning where one stores their photographs or videos can help the investigator discover an unknown persona. Online photo and video storage websites allow users to store a large amount of data, which oftentimes equates to over one gigabyte in size. This presents the suspect with the convenient opportunity to only use one photo sharing website, of which the investigator can exploit to link unknown personas back to an individual. See *Figure 3* for an example. In Figure 3, the suspect has three OSNCs. The investigator identified the Photobucket account using traditional search techniques and used the data collected to search OSNC1 and OSNC3 to help identify the hidden persona. Also, after reviewing OSNC3, the investigator located an unknown Youtube account which later linked was linked to OSNC2.



Typically, every online community and social networking website will have a search function. Some are more powerful than others, but please be advised that some communities permit only registered users to utilize the search feature. Please note the necessary steps to ensure anonymity still apply to these searches.

Another feature that is a great resource with which to conduct investigations into online personas is Google. Correctly implementing Google's powerful indexing/search engine to your investigative arsenal is key to an investigation. Two features which will be discussed in this paper are "Google Alerts" and Google's Advanced Search features.

A Google Alert is a feature of Google that will send an email to the requesting person when a specific text string is indexed by Google (Google, 2007). This feature alone is very powerful, but will only alert the requestor from the time they configure the alert forward, not retroactively.

The second Google feature that is very helpful in an online investigation is Google's advanced search features. One in particular is the ability to search an entire website from the Google website by using the following string:

apple site:www.cnn.com

The above search term will search the website www.cnn.com for the term "apple". As you can see this is a very powerful feature to be aware of.

To access Google's advanced search features, please click the "Advanced Search" link by the search bar.[4]

Aside from the educated statistical approach, if the investigator has access to the digital devices suspected of being utilized to access OSNCs, the next immediate step would be to contact a competent digital forensic examiner to review the digital devices. Such digital devices of interest include but are not limited to: laptops, desktops, mobile telephone devices, media players. Due to the nature of how technology is ever changing, it may be appropriate on a case-by-case basis, to consult with a technology specialist (or an expert of the like) to ensure that you exhausted your search of available digital devices.

A knowledgeable digital forensic examiner should be able to determine which if any OSNC was utilized and persona used on each.

[4] There are several books and publications available on leveraging the power of the Google search engine.

Step 5: Collect potential evidence

Once the investigator identifies potential evidence it's critical to collect the information and preserve it as soon as possible. Due to the OSNC's typically storing active content on their websites, as soon as a change is made, its immediately made and there are no backups available.

There are two recommended steps to be followed by investigators when collecting potential evidence from OSNCs. The first of which should be done under most instances and the second should be done when needed.

1) Use a tool to collect/save/copy/print the content from the OSNC. One example is a tool called Camtasia[5] that enables an investigator to save a video capture of the data presented on the computer monitor. Another tool noteworthy of mentioning is HTTrack[6]. According to the vendors website, the software allows investigators to "download a World Wide Web site from the Internet to a local directory, building recursively all directories, getting HTML, images, and other files from the server to your computer. HTTrack arranges the original site's relative link-structure. Simply open a page of the "mirrored" website in your browser, and you can browse the site from link to link, as if you were viewing it online. HTTrack can also update an existing mirrored site, and resume interrupted downloads. HTTrack is fully configurable, and has an integrated help system."

2) Investigators can obtain court authorization to acquire the content and logs associated with an OSNC persona. The results of such a request can yield information not available from the "internet facing" OSNC. For example, the investigator will be typically presented with registration email addresses, Internet Protocol (IP) logs with dates/times of persona activity.

7. CONCLUSIONS AND OBSERVATIONS

In both private and public sector investigations, OSNC's can provide a great amount of evidentiary information. Arming an investigator with the purposed methodology detailed in this paper, initiates the process of potentially discovering more information about a suspect or targeted individual.

REFERENCES

Facebook. (2008, January 28). *Facebook*. Retrieved January 28, 2008, from Facebook: http://www.facebook.com/

Flickr. (2008, January 28). *Flickrr*. Retrieved January 28, 2008, from Flickr: http://www.flickr.com/

Google. (2007, 11 28). *Google Alerts*. Retrieved 11 2008, 2007, from Google.com: www.google.com/alerts

Heckers, J. (2007, 11 19). *Three areas management must handle with delicacy.* Retrieved 11 28, 2007, from bizjournals: http://www.bizjournals.com/business_resources/hr_careers/business_advice/employment/2007/11/19/column9.html

How Should We Conduct Investigations? (2007, December). *Directorship* , 10-11.

http://video.google.com/. (2008, January 28). *http://video.google.com/*. Retrieved January 28, 2008, from http://video.google.com/: http://video.google.com/

[5] Camtasia is a product of TechSmith and can be reviewed at http://www.techsmith.com/camtasia.asp

[6] HTTrack can be reviewed at http://www.httrack.com/

Internet World Stats. (2007, 11 11). *Internet Growth Statistics*. Retrieved 11 28, 2007, from www.internetworldstats.com: http://www.internetworldstats.com/emarketing.htm

Kazakoff, L. (2009, Spring). Care and feeding of blogs. *The Masthead* , p. 3.

Myspace. (2008, January 28). *Myspace*. Retrieved January 28, 2008, from Myspace: http://www.myspace.com/

Nielsen/NetRatings. (2006, 5 11). *www.nielsen-netratings.com.* Retrieved 11 28, 2007, from Nielsen/NetRatings: http://www.nielsen-netratings.com/pr/pr_060511.pdf

O'Reilly, T. (2005, 09 30). *What Is Web 2.0*. Retrieved 03 05, 2008, from Oreilly Net: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Photobucket. (2008, January 28). *Photobucket*. Retrieved January 28, 2008, from Photobucket: http://photobucket.com/

WordPress. (2008, January 28). Retrieved January 28, 2008, from WordPress: http://wordpress.org/

YouTube. (2008, January 28). *YouTube*. Retrieved January 28, 2008, from YouTube: http://youtube.com/

ACKNOWLEDGEMENTS

This project was partially supported by Grant No. 2004-MU-MU-K001 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.

AUTHOR BIOGRAPHY

Jonathan T. Rajewski is a Computer & Digital Forensics instructor at Champlain College and a Computer Forensic Examiner with the Vermont Internet Crimes Task Force in Burlington, Vermont. He has experience with both civil and criminal digital forensic investigations and in providing expert written and oral digital forensic testimony. He has served many high profile confidential clients and has worked alongside both international and local, state/federal governmental entities. Jonathan holds a B.S in Economic Crime Investigation and the following professional certifications: Certified Computer Examiner (CCE), EnCase Certified Examiner (EnCe), Certified Fraud Examiner (CFE) and Certified Information Systems Security Professional (CISSP).

Subscription Information

The Proceedings of the Conference on Digital Forensics, Security and Law is a publication of the Association of Digital Forensics, Security and Law (ADFSL). The proceedings are published on a non-profit basis.

The proceedings are published in both print and electronic form under the following ISSN's:

ISSN: 1931-7379 (print)

ISSN: 1931-7387 (online)

Subscription rates for the proceedings are as follows:

Institutional	- Print & Online:	$120	(1 issue)
Institutional	- Online:	$95	(1 issue)
Individual	- Print:	$25	(1 issue)
Individual	- Online:	$25	(1 issue)

Subscription requests may be made to the ADFSL.

The offices of the Association of Digital Forensics, Security and Law (ADFSL) are at the following address:

Association of Digital Forensics, Security and Law
1642 Horsepen Hills Road
Maidens, Virginia 23102
Tel: 804-402-9239
Fax: 804-680-3038
E-mail: editor@jdfsl.org
Website: http://www.adfsl.org

Contents