Conference on Digital Forensics, Security and Law

Proceedings of the
Conference on
Digital Forensics,
Security, and Law
2011

Richmond, Virginia
May 25-27

Conference on Digital Forensics, Security and Law

Richmond, Virginia
May 25-27, 2011

Conference Chair

Glenn S. Dardick
Longwood University
Virginia, USA

ADFSL

Association of Digital Forensics, Security and Law

ISSN 1931-7379

Sponsor

LONGWOOD
UNIVERSITY

Contents

Conference Committee

The 2011 ADFSL Conference on Digital Forensics, Security and Law is pleased to have the following members of the conference committee.

Glenn Dardick
gdardick@dardick.net
General chair
Longwood University
Virginia
USA

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

Diane Barrett
Dbarrett@uat.edu
University of Advanced Technology, Arizona
USA

Mohamed Chawki
chawki@cybercrime-fr.org
University of Aix-Marseille III
France

Gareth Davies
gddavies@glam.ac.uk
University of Glamorgan
UK

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

Kevin Harris
KevinL.Harris@nscc.edu
Nashville State Community College
Tennessee
USA

Andy Jones
andrew.jones@kustar.ac.ae
Khalifa University
UAE

Grover Kearns
gkearns@mail.usf.edu
Univ. of South Florida, St. Petersburg
St. Petersburg, FL
USA

Gary Kessler
gck@garykessler.net
Gary Kessler Associates
Vermont
USA

Stephen Larson
larsonsp@vcu.edu
Virginia Commonwealth University
Virginia
USA

Jigang Liu
Jigang.Liu@metrostate.edu
Metropolitan State University
Minnesota
USA

Milt Luoma
Milt.Luoma@metrostate.edu
Metropolitan State University
Minnesota
USA

Vicki Luoma
vicki.luoma@mnsu.edu
Minnesota State University Mankato
Minnesota
USA

Angela Orebaugh
angela_orebaugh@yahoo.com
George Mason University
Virginia
USA

John Riley
jriley@bloomu.edu
Bloomsburg University
Pennsylvania
USA

Marcus Rogers
rogersmk@purdue.edu
Purdue University
Indiana
USA

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

Joseph J. Schwerha IV
schwerha@calu.edu
Owner, TraceEvidence, LLC
California U of Pennsylvania
USA

Craig Valli
c.valli@ecu.edu.au
Edith Cowan University
Western Australia
Australia

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

Bob Zeidman
bob@zeidmanconsulting.com
Zeidman Consulting
Cupertino, California
USA

Schedule

Wednesday, May 25

- 08:00 AM CONTINENTAL BREAKFAST
- 08:00 AM On-site Registration
- 08:45 AM Introductions
  - *Glenn S. Dardick, Conference Chair and Director of the ADFSL*
- 09:00 AM Welcome
  - *Patrick Finnegan, President of Longwood University*
- 09:15 AM Papers/Presentation session
  - *Milton Luoma and Vicki Luoma: Sampling: Making Electronic Discovery More Cost Effective*
  - *Karon N. Murff, Hugh E. Gardenier, III and Martha L. Gardenier: Digital Forensics and the Law*
- 10:45 AM BREAK
- 11:00AM Papers/Presentation session
  - *Josiah Dykstra and Alan T. Sherman: Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies*
  - *Ashley L. Podhradsky, Rob D'Ovidio and Cindy Casey: A Practitioners Guide to the Forensic Investigation of Xbox 360 Gaming Consoles*
- 12:30 PM LUNCH (provided)
- 01:30 PM Keynote
  - *Brian Snow, Consultant and Former Technical Director at the NSA*
- 02:00 PM Papers/Presentation session
  - *P. Vinod Bhattathiripad and Lt. Dr. S. Santhosh Baboo: Software Piracy Forensics: Impact and Implications of Post-Piracy Modifications*
  - *David Primeaux, Robert Dahlberg, Kamnab Keo, Stephen Larson, B. Pennell and K. Sherman: MAC OS X Forensics: Password Discovery*
- 03:30 PM BREAK
- 03:45 PM Papers/Presentation session
  - *Grover Kearns and Katherine J. Barker: Forensic Implications of a Continuous Auditing Model*
  - *Satoshi Kai and Tetsutaro Uehara: Development of A Distributed Print-Out Monitoring System for Efficient Forensic Investigation*
- 05:15 PM Conference Close for Day

Thursday, May 26

- 08:00 AM CONTINENTAL BREAKFAST
- 08:00 AM On-site Registration
- 08:45 AM Papers/Presentation session
  - *Craig Valli, Andrew Woodward and Peter Hannay: Assessing Backtrack in the Outback - A Report on Cyber Security Evaluation of Organisations in Western Australia*
  - *Kam Woods, Christopher A. Lee, Simson Garfinkel, David Dittrich, Adam Russell and Kris Kearton: Creating Realistic Corpora for Forensic and Security Education*

Schedule

Thursday, May 26 (continued)

- 10:15AM BREAK
- 10:30 AM Papers/Presentation session
  - *David Biros: National Information Assurance and Forensics Education Repository (NAIFER)*
  - *David Biros: DC3 Center of Digital Forensics Academic Excellence Program*
- 12:00 Noon LUNCH (provided)
- 01:00 PM Keynote Speech
  - *Ricky Windsor: National Centers of Digital Forensics Academic Excellence Program*
- 01:30 PM Papers/Presentation session
  - *Bob Johnston: Digital Forensics Investigation in A Collegiate Environment*
- 02:15 PM Panel
  - *Defining the Definition: What is the Future of Legal Obligations Relating to Evidence Contained Mobile Devices?*
- 03:15 PM BREAK
- 03:30 PM Papers/Presentation session
  - *Linda Lau and Cheryl Davis: AACSB-Accredited Schools' Adoption of Information Security Curriculum*
- 04:15 PM Panel
  - *Cyber Forensics Curriculum*
- 05:15 PM Conference Close for Day

Friday, May 27

- 08:00 AM CONTINENTAL BREAKFAST
- 08:30 AM Papers/Presentations session
  - *Peter Hannay: Kindle Forensics: Acquisition & Analysis*
- 09:15 AM Workshop
  - *Diane Barrett: Physical Image Analysis in Mobile Devices*
- 10:30 AM BREAK
- 10:45 PM Papers/Presentations session
  - *Felix Freiling, Michael Spreitzenbarth and Sven Schmitt:: Forensic Analysis of Smartphones: The Android Data Extractor Lite (ADEL)*
  - *Rita M Barrios, Michael Lehrfeld: Forensicating iOS Mobile Devices*
  - *Keyun Ruan, Ibrahim Baggili, Joe Carthy and Tahar Kechadi: Survey on Cloud Forensics and Critical Criteria for Cloud Forensic Capability: A Preliminary Analysis*
- 12:45 PM Conference Close

SAMPLING: MAKING ELECTRONIC DISCOVERY MORE COST EFFECTIVE

Milton Luoma
Metropolitan State University
700 East Seventh Street
St. Paul, Minnesota 55337
651 793-1246 (fax)
651 793-1481
Milt.Luoma@metrostate.edu

Vicki Luoma
Minnesota State University
145 Morris Hall
Mankato, Minnesota 56001
507 389-5497
507 389-1916
Vicki.Luoma@mnsu.edu

ABSTRACT

With the huge volumes of electronic data subject to discovery in virtually every instance of litigation, time and costs of conducting discovery have become exceedingly important when litigants plan their discovery strategies. Rather than incurring the costs of having lawyers review every document produced in response to a discovery request in search of relevant evidence, a cost effective strategy for document review planning is to use statistical sampling of the database of documents to determine the likelihood of finding relevant evidence by reviewing additional documents. This paper reviews and discusses how sampling can be used to make document review more cost effective by considering issues such as an appropriate sample size, how to develop a sampling strategy, and taking into account the potential value of the litigation in relation to the costs of additional discovery efforts.

Keywords: sampling, statistical sampling, electronic discovery

1. INTRODUCTION

Litigation has always been about the adversarial relationship and zealous representation of one's clients and their interests, but with the rapid expansion in the volume of electronically stored information (ESI) lawyers have found themselves having to become non-adversarial in the discovery phase of litigation now that electronic discovery is the norm. With the relatively low cost of data storage and the seemingly limitless amount of ESI to search, the new Federal Rules of Civil Procedure were amended to require lawyers and the parties to fully cooperate in the management of the discovery process.

At the onset of litigation, a party must comply with Federal 26(a)(1)(B), which requires full disclosure of a great deal of basic information as described below.

> Rule 26. Duty to Disclose; General Provisions Governing Discovery
>
> (a) Required Disclosures.
>
> (1) Initial Disclosures.

> (A) *In General*. Except as exempted by Rule 26(a)(1)(B) or as otherwise stipulated or ordered by the court, a party must, without awaiting a discovery request, provide to the other parties:
>
> > (i) the name and, if known, the address and telephone number of each individual likely to have discoverable information — along with the subjects of that information — that the disclosing party may use to support its claims or defenses, unless the use would be solely for impeachment;
> >
> > (ii) a copy — or a description by category and location — of all documents, electronically stored information, and tangible things that the disclosing party has in its possession, custody, or control and may use to support its claims or defenses, unless the use would be solely for impeachment; (Federal Rules of Civil Procedure, 2007)

The rule requires adversaries to exchange either a copy of or a description of all electronically stored information by category and location that may be used in their legal claim or defense against the claim. This requirement is tantamount to asking a poker player to show his or her hand before bets are placed. However, it really is not as simple as showing your hand in a poker game because most of the time the party has no idea what they have, where it is located and how to produce it.

2. HOW MUCH TRUTH CAN YOU AFFORD?

There is simply too much information to produce all of one's ESI or even to list everything one has or even to know what one has. In the oft cited case of Zubukake v. UBS Warburg, Judge Shira Scheindlin wrote: "Discovery is not just about uncovering the truth, but also about how much of the truth the parties can afford to disinter." (Zubulake v. UBS Warburg LLC, 2003)

Laura Zubulake sued her former employer UBS Warburg over gender discrimination. The case became a catalyst for development of the new discovery rules and procedures. Zubulake requested documents stored or produced in electronic format, which were primarily emails. UBS Warburg claimed either the data could not be found or it had been lost. In a series of five pre-trial rulings the judge examined cost shifting, discovery obligations, and responsibilities of maintaining and retrieving data. Judge Scheindlin ultimately found that the defendant had a duty to preserve data that it knew or should have known were relevant to the litigation. To determine the issue of cost shifting the judge ordered the defendant to restore and review information from five backup tapes out of a total of 94 available tapes. The court allowed Zubulake to select five tapes out of the 94 for sampling. Defendant Warburg was ordered to submit an affidavit with the results of the sampling along with costs. (Zubulake, 2004)

Zubulake chose five tapes with emails from her former supervisor. After the five sample tapes were restored, the defendant revealed there were 6,203 unique emails contained in the sample data. In the next step in the recovery process keyword searches were used to find emails that made reference to Zubulake, reducing the messages to 1,075 unique messages and claimed that of those 1,075 only 600 were subject to Zubulake's document request. This process cost Warbug over $19.000.00. Warburg estimated the cost to restore and produce the remaining tapes to be approximately $273,649.39. (Zubulake v. UBS Warburg LLC, 2003). There are numerous important rulings in this case but what is remarkable is the use of sampling as a method of reducing costs and narrowing search requirements. This case used sampling to determine whether more searching should be conducted and whether costs should be shifted to the party seeking the information. (Zubulake, 2004)

In 2007 the new Federal Rules of Civil Procedure were adopted and Rule 34 included a provision for sampling. Rule 34 reads as follows:

Rule 34. ***Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes***

(a) In General

A party may serve on any other party a request within the scope of Rule 26 (b)

(1) to produce and permit the requesting party or its representative to inspect, copy, test, or sample the following items in the responding party's possession, custody, or control:

(A) any designated documents or electronically stored information — including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations — stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form; or

(B) any designated tangible things; or

(2) to permit entry onto designated land or other property possessed or controlled by the responding party, so that the requesting party may inspect, measure, survey, photograph, test, or sample the property or any designated object or operation on it. (FRCP 26(b)

The important point is that a party may serve a request to sample data. Sampling should be used routinely in cases with large amounts of electronically stored information to find the data needed whether in producing the data or in defending a search methodology. It also permits a lawyer to be both cooperative and adversarial at the same time. This procedure places a greater responsibility on the requesting party to apply the reasonableness standard to determine what should be sampled. On the other hand, sampling permits the party providing the data to verify to the court that he or she has made a reasonable effort to comply with discovery by checking the results.

3. RECENT CASES FOLLOWING THE ZUBULAKE GUIDELINES

In 2010 in Makrakis v. Demelis, the plaintiff sought damages from the defendant nurse Demelis and her employer, Brigham & Women’s Hospital, for damages when the nurse improperly administered a toxic dose of a drug to the plaintiff. The plaintiff asked the court for an order requiring the hospital to restore all electronic backup tapes containing emails originating from thirteen employees or former employees of the hospital from 1987 to 2010. The plaintiffs sought an order requiring the hospital to hire a third-party vendor to search the restored email archives using the keywords “Makrakis,” “DeMelis,” “pancuronium,” and “Pavulon.” Further, plaintiffs sought a court order compelling production of all emails sent or received by DeMelis at any time. The defendants opposed the request on the grounds that the search would be unduly burdensome, prohibitively expensive and not add anything relevant to the information they already had. The court, citing Zubulake, ordered the defendants to sample a small number of backup tapes, at the expense of the requesting party. (Makrakis v. Demelis, 2010)

In another 2010 case the court ruled that a phased approach to ESI discovery is appropriate and reasonable approach. In this case through sampling the discovery costs were reduced from the estimated $60,000 to $13,000 (Barrera v. Boughton, 2009)

In a 2009 case the court found that, that “sampling to test both the cost and the yield is now part of the mainstream approach to electronic discovery.” (S.E.C v. Collins & Aikman Corp, 2009)

Courts now require litigants to be even more responsible for the success and accuracy of the discovery process by requiring them to defend their chosen search methods and to show how they have verified or validated their results. In other words, what tests were done to establish that the methodologies used were efficacious?

Judge Grimm found:

> Additionally, the defendants do not assert that any sampling was done of the text searchable ESI files that were determined not to contain privileged information on the basis of the keyword search to see if the search results were reliable. Common sense suggests that even a properly designed and executed keyword search may prove to be over-inclusive or under-inclusive, resulting in the identification of documents as privileged which are not, and non-privileged which, in fact, are. The only prudent way to test the reliability of the keyword search is to perform some appropriate sampling of the documents determined to be privileged and those determined not to be in order to arrive at a comfort level that the categories are neither over-inclusive nor under-inclusive. There is no evidence on the record that the Defendants did so in this case. (Victor Stanley, Inc. v. Creative Pipe, Inc., 2010)

Based on these and other cases with similar rulings, sampling must be considered by all parties to litigation in order to reduce discovery costs.

4. SAMPLING – HOW IT WORKS

Sampling can assist one both in finding the data required to strengthen one's case, but it can also be used to certify one's ESI discovery results. In sampling one must be able to show precision, confidence, and the expected deviations. The Electronic Discovery Reference Model (EDRM) Search Group outlines a strategy for using sampling. (The Electronic Discovery Reference Model, 2005) As the EDRM Search Guide states, sampling can only be done by the one who has the data, which may not always be in line with the requesting parties' demands. The Sedona Conference, Working Group Commentary, Achieving Quality in the E-Discovery Process discusses several sampling methods and their purposes. (Working Group 1, 2009)

Essentially, sampling a set of electronic documents is a tradeoff between obtaining every possible relevant document, which will invariably result in a very high cost, versus reviewing a smaller set of the documents at a lower cost, but running the risk of missing relevant documents that may be critical to the case. For large scale litigation or in cases where limited resources may be available for the discovery process, sampling is an intelligent alternative to attempting to review every possible document that is available.

There are several types of sampling that can be used in a sampling procedure. For example, the Sedona Conference identified five quality measures including judgment sampling as very helpful (p12) even though one cannot make generalized statements about the entire population of documents. This form of sampling can be used in a quality control context where a small sample of documents can be selected from a set of documents that have been reviewed by junior counsel to determine whether or not the document reviewer has exercised proper judgment regarding how the document was classified, that is, as relevant or not. However, not just any sample will do. The very best kind of sample is one that is representative of the entire population of electronic documents.

While several other sampling methods exist, but the most important of these is statistical sampling that permits one to generalize about the entire population of documents based on a random sample of documents. The question that must be answered for anyone designing a sampling procedure is how large must the sample be? The answer to that question depends on how confident one wants to be that the sample size is truly representative of the population and what range of the estimate of the proportion of relevant documents is required.

To determine the sample size when one wishes to determine the proportion of documents in the population of documents that are relevant for discovery purposes, one must determine or estimate five items: 1) the desired interval range within which the population proportion is expected, 2) the confidence level for estimating the interval within which to expect the population proportion, 3) the standard error of the proportion, 4) an estimate of the proportion of the population which contains relevant documents, and 5) calculate the sample size.

First, the desired interval range within which the population proportion is expected is a wholly subjective decision. For example, if one wants the resulting interval range to be within 10 percent of the population's true proportion of relevant documents, then this figure will be plus or minus 0.10. If one wants a tighter limit on the interval range, such as five percent, then this figure will be plus or minus 0.05. So, if one wants to be able to say the population of electronic documents contains X% relevant documents plus or minus 10%, then the sample size will be determined with this requirement in mind as shown below.

Second, the confidence level desired for the final estimate of the population range is a subjective choice where the calculations are based on the Normal distribution, or classical bell curve, and incorporates values based on the standard deviation or standard error of the Normal distribution. For example, a common choice for confidence level is 90% or 95%, so ultimately one will be able to say something like, "I am 95% certain that the population of electronic documents contains 70% + or - 10% documents relevant to the litigation at hand."

Third, one must estimate the standard error of the proportion of relevant documents. This figure is obtained by dividing the result of step 1 by 1.65 if one desires a confidence level of 90% or dividing the result of step 1 by 1.96 if one desires a confidence level of 95%; or dividing the result of step 1 by 3.00 if one desires a confidence level of 99%. The following table shows the results of using interval ranges of 10%, 5%, and 1% and confidence levels of 90%, 95%, and 99%.

Estimate of the Standard Error of the Proportion of Relevant Documents			
Proportion of Relevant Documents + or - %	90% confidence	95% confidence	99% confidence
10%	0.06061	0.05102	0.03333
5%	0.03030	0.02551	0.01667
1%	0.00606	0.00255	0.001667

Fourth, to determine sample size one first needs to estimate the proportion of the documents in the population that are relevant. Since that is not generally known beforehand, one must estimate that proportion before calculating the sample size. The best way to estimate that proportion is to complete some preliminary testing or pilot sampling by randomly selecting several documents and determining the proportion of this sample that contains relevant documents. Usually, about 30 documents per pilot sample are sufficient. This preliminary testing or pilot sampling can be repeated several times. If the selection of documents for each pilot sample is random, then the average proportion of relevant documents contained in the samples should be close to the population's proportion of relevant documents. The product of the proportion of relevant documents multiplied by the proportion of non-relevant documents is referred to as the dispersion of the sample.

Finally, the sample size is calculated by dividing the sample dispersion by the estimate of the standard error of the proportion multiplied by itself. For example, suppose we wish to be 95% confident that the proportion of relevant documents in the population is 20% plus or minus 5% and the proportion of relevant documents in the pilot sampling procedure was 20%, then our sample size is (0.20)(0.80) / 0.02551 =245.866 rounded off to 246, which is a reasonable number to review.

On the other hand, consider the situation if one desires to be 99% confident that the proportion of relevant documents in the population is 20% plus or minus 1% and the proportion of relevant documents in the pilot sampling procedure was 20%, then our sample size is (0.20)(0.80) / 0.001667 = 57,577! Clearly, the tighter the interval and the higher the confidence level desired increases the sample size – in some cases quite dramatically.

5. CONCLUSION

In conclusion, it is clear that ESI sampling has become an important aspect of electronic discovery. It has been used as a means of validating search methodologies as well as a means of containing discovery costs and maintaining quality control over the discovery process. While several sampling methods are available, statistical sampling can be an effective way of describing the characteristics of an entire population of ESI documents based on a relatively small sample of documents randomly selected from the population. It further permits one to establish the confidence level of the sampling results and the range of accuracy of the results. It therefore behooves lawyers to educate themselves on the procedures involved in the development of statistical sampling methodologies, which may at the very least satisfy the safe harbor provisions of the Federal Rules of Civil Procedure.

REFERENCES

Barrera v. Boughton, 256 F.R.D. 403, 418 (S.D.N.Y. 2009).

Cooper, D. R. and P.S. Schindler (2003), Business Research Methods, McGraw-Hill, Boston.

Federal Rules of Civil Procedure, 26(a)(1)(B) (December 2007).

Makrakis v. Demelis, 2010 WL 3004337 (09-706-C July 13, 2010).

S.E.C v. Collins & Aikman Corp, 256 F.R.D. 403, 418 (S.D.N.Y. 2009).

The Electronic Discovery Reference Model. (2005). Retrieved December 16, 2010, from The Electronic Discovery Reference Model: http://www.law.com/jsp/legaltechnology/eDiscoveryRoadmap.jsp

Victor Stanley, Inc. v. Creative Pipe, Inc., 2010 WL 3530097 (D.MD 2010).

Working Group 1. (2009). "Achieving Quality in the E-Discovery Process.". *Commentary by the Working Group 1 of The Sedona Conference®*. Sedona: Sedona Conference.

Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 312 (S.D.N.Y 2003).

DIGITAL FORENSICS AND THE LAW

Karon N. Murff
Sam Houston State University
Department of Computer Science
USA
knm002@shsu.edu

Hugh E. Gardenier, III
Sam Houston State University
Department of Computer Science
USA
heg003@shsu.edu

Martha L. Gardenier
Sam Houston State University
Department of Computer Science
USA
mlh022@shsu.edu

ABSTRACT

As computers and digital devices become more entrenched in our way of life, they become tools for both good and nefarious purposes. When the digital world collides with the legal world, a vast chasm is created. This paper will reflect how the legal community is failing to meet its obligation to provide adequate representation due to a lack of education about digital (computer) forensics. Whether in a civil litigation setting or a criminal setting, attorneys, prosecutors and judges have inadequate knowledge when it comes to the important questions they need to ask regarding digital evidence. Reliance on expert witnesses is not enough when the attorney cannot discern whether the opinion presented by the expert (even their own expert) is accurate, factual, or even plausible. The results of a survey distributed to attorneys, prosecutors and judges throughout the United States bear this out in a startling manner.

Keywords: attorneys, lawyers, computer forensics, digital forensics, CLE

1. INTRODUCTION

In 2002, Scott C. Williams, a supervisory special agent for the FBI's computer analysis and response team in Kansas City was quoted by writer David Hayes in the Kansas City Star newspaper, saying that over fifty percent of crimes investigated involved a computer. From January 1 through December 31, 2009, the FBI Internet Crime Complaint Center data reflected 336,655 complaint submissions, which represented a 22.3 percent increase in computer related crimes over 2008 (http://crimeinamerica.net/2010/03/16/computer-crime-reports-increase-22-percent-in-2009.html, March 16, 2010). These are just the crimes reported to the FBI. How many crimes involving computers are never actually reported or are investigated by local agencies?

Once law enforcement has investigated these crimes, prosecutors, defense attorneys and judges take over. The final outcome, be it an acquittal, plea bargain, or guilty verdict, is dependent on the quality of the evidence and the ability of the prosecutor or the defense attorney to convey the story in the most understandable manner to the judge and jury. The public depends on the prosecutor to represent the good of the people in an honest manner and to understand the evidence. A client depends on his or her attorney to be knowledgeable about the evidence in order to provide an adequate defense. This paper demonstrates the gap which exists between expectation and reality.

1.1 Background

"Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." Frye v. United States, (1923).

Associate Judge Van Orsdel wrote this in his opinion denying the appeal of a man convicted of murder. James Alphonso Frye was convicted of second degree murder and appealed his conviction based on the trial court ruling that his expert witness, who conducted a polygraph test on Mr. Frye, could not testify on his behalf. Frye v. United States (1923) became the standard in jurisdictions across the United States with regard to scientific evidence. As such, the validity of methodologies and techniques used in gathering and processing evidence has gone through rigorous scrutiny to gain acceptance in the judicial system.

In 1975, the Federal Rules of Evidence went into effect. Up to this point Frye v. United States (1923) remained the yardstick and was widely accepted and followed by the courts. That the legislative history of the Federal Rules never addressed Frye v. United States (1923) or the issue of admittance of scientific evidence or use of expert witnesses, kept the 1923 opinion at the forefront in the making of judicial decisions. This finally changed in 1993 when the U.S. Supreme Court decided the first of the Daubert Trilogy. In Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993), 589, the Court ruled that scientific expert testimony should be admitted based on the following:

> **Judge is gatekeeper:** ". . . under the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable." (Daubert 589)
>
> **Relevance and reliability:** The trial judge must ensure that the expert's testimony is "relevant to the task at hand" and rests "on a reliable foundation". (Daubert 584-587)
>
> **Scientific knowledge:** "The Rule's requirement that the testimony "assist the trier of fact to understand the evidence or to determine a fact in issue" goes primarily to relevance by demanding a valid scientific connection to the pertinent inquiry as a precondition to admissibility (Daubert, 1993).
>
> **Factors relevant:** The Court defined "scientific methodology" as the process of formulating hypotheses and then conducting experiments to prove or falsify the hypothesis, and provided a nondispositive, nonexclusive, "flexible" test for establishing its "validity" (Daubert, 1993):
>
> 1. Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested.
> 2. Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication.
> 3. Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.
> 4. Finally, "general acceptance" can yet have a bearing on the inquiry.

1.2 No Algorithms Allowed

How does the legal community deal with the requirements set out by the Supreme Court? Not very well, as seen by the results of our survey, research and the results of case law. And how does all of this relate to a survey of attorneys regarding their knowledge of digital forensics? There exists a general lack of foundation with regard to digital forensics (computer forensics). Many in law do not recognize digital forensics as a "forensic science," and others just glaze over at the thought of having to learn anything about the topic. Countless attorneys and law students will admit they chose law school over other graduate programs to avoid math or science courses. In fact, statistics show that the arts and humanities and business administration comprise the vast majority of law school feeder degrees. Law schools have perpetuated this trend by not emphasizing the application of science and math to legal concepts; this, despite the growing necessity to provide education in all of the forensic sciences.

The widespread belief among attorneys is that the expert witness will take care of the issue. However, the attorneys, prosecutors and judges must know the correct questions to ask the expert in order to determine the validity, pertinence, and admissibility of the evidence.

1.3 What Would Perry Mason Say?

One of the problems confronted in the courtroom is the CSI effect. Television and movies dramatize the collection of forensic evidence, including digital evidence. The evidence is always clear and convincing, and the case is solved in sixty minutes with no worries about warrants or research time. This is one of the preconceptions which jurors bring with them. Unfortunately, what is shown on CSI or NCIS is not representative of sound evidence collection techniques, nor in some cases do the televised techniques even exist.

Jessica D. Gabel, in her Summer 2010 article, "Forensiphilia: Is Public Fascination with Forensic Science a Love Affair or Fatal Attraction?" posed the question which plagues many in the legal and scientific community nowadays. Gabel posits that the CSI effect has caused a bias in juries which affects verdicts. In cases in which no forensic evidence is produced, jurors may have a tendency to decide in favor of the defense; however, when forensic evidence is presented by the prosecution, then jurors may make the connection to CSI, and assume that if it is good science on television, then it is good science in the courtroom. Gabel feels there is a larger issue: "bad science is slipping through the cracks, creating a glut of bad decisions and wrongful convictions." (Gabel, 2010, p.5)

2. THE SURVEY

The purpose of this study is to measure the understanding of practicing attorneys in the United States with respect to the field of digital forensics (aka computer forensics) and the application of digital evidence in the courtroom environment. In order to accomplish this, a four-step process was used to collect and evaluate data. This methodology consisted of:

1. Defining a problem for evaluation,
2. Collecting data to evaluate the problem,
3. Summarizing data collected in a suitable manner for analysis, and
4. Data analysis, interpretation of results, and communication of those results.
   (Longnecker and Ott, 2010, p. xi)

2.1 Defining a Problem for Evaluation

The Texas Disciplinary Rules of Conduct (http://www.texasbar.com/AM/Template.cfm?Section= Grievance_Info_and_Ethics_Helpline&Template=/CM/ContentDisplay.cfm&ContentFileID=96) for attorneys states that, "in all professional functions, a lawyer should zealously pursue clients' interests within the bounds of the law. In doing so, a lawyer should be competent, prompt and diligent." According to the American Legal Ethics Library at Cornell University Law School, "Competent" or "Competence" denotes possession or the ability to timely acquire the legal knowledge, skill, and training reasonably necessary for the representation of the client. Professional rules of conduct in all states require a similar application of professional skill, knowledge, and conduct. Based upon application of the Texas Disciplinary Rules of Conduct the problem for evaluation in this study is:

- Do attorneys have sufficient knowledge and training with respect to digital forensics to reasonably and competently represent their clients?

2.2 Case Law as an Index of Knowledge

Defining what is sufficient knowledge and training with respect to digital forensics so that an attorney has the tools necessary to adequately represent their client is, of course, subjective. Criminal defense work typically requires a strategic use of resources to achieve a verdict which in the minds of the jurors is "beyond a reasonable doubt," while civil litigation is directed to verdicts based on the "preponderance of evidence." The stakes are different, available resources are markedly dissimilar, and the weight of digital forensics evidence is often insurmountable for the criminal defense attorney. In many instances, such as sex crime cases, the perception of guilt is so great that the most valued attribute of the attorney is their ability to plea bargain a sentence that will eventually result in the release of their client from prison before the end of their natural life. This, of course, calls for a different skill set and does not result in appealable convictions.

Competence is a touchy area with practicing attorneys, and it requires conclusions that are judgmental rather than analytical. The kiss of death for trial counsel is to be judged to provide ineffective assistance of counsel. *Black's Law Dictionary* defines this as "a representation in which the defendant is deprived of a fair trial because the lawyer handles the case unreasonably, usually either by performing incompetently or by not devoting full effort to the defendant . . ." *Black's* relates ineffective counsel to a defendant being deprived of his Sixth Amendment right to a fair trial.

This argument implies that a defendant in a criminal case could have their Sixth Amendment rights contravened if their attorney does not have sufficient knowledge and training with respect to digital forensics to reasonably and competently represent their client. A baseline for measuring this was obtained by reviewing Westlaw citations for Federal and state cases appealed during the last ten years using the search term "computer forensics" in conjunction with "ineffective assistance of counsel." This combination appears in thirteen Federal cases, and twenty-one state cases since 2001. Review of these cases revealed that seventeen of the state cases involved issues related to the identification and retrieval of evidence from digital devices, and that such evidence was used at trial.

Westlaw Search Term	**Cites in Federal and State Courts**	**Additional Search Term "computer forensics"**
"inadequate defense"	139	No citations
"ineffective assistance of counsel"	> 10,000	34
"ineffective counsel"	3,721	No citations

In each state case one or more assignments of error were raised on appeal by appellants, which involved computer forensics evidence and alleged ineffective preparation of legal counsel with respect to such evidence. In order to determine the substance of these allegations and to identify common weaknesses in the presentation of computer forensics evidence and testimony in court, the seventeen state cases were examined in detail. While all of these cases were selected from the ten-year period (2001-2010), in actuality they were heavily-weighted to the period 2008 to 2010 which represented

76.5% of the cases reviewed. This was consistent with: (1) evolving digital technology; (2) increased spending on computer investigative services in conjunction with increased funding for Homeland Security programs; (3) the evolution of joint federal/state and federal/international child pornography and human trafficking task forces that effectively identified and provided evidence and assistance for the indictment of individual child pornographers; and (4) the evolution of a digital information-based culture in much of the world. More importantly, appeals court activity during this later time period was indicative of an evolving legal culture in the United States that was being forced to leave traditional measures of evidence in the realm of the observable and tangible, and cope with rapidly evolving digital evidence that was understandable only after technically-skilled experts massaged the storage devices and tapped a virtual jackpot of evidence. This, in many respects, changed the traditional role of defense attorneys as advocates for their clients, and created a deer-in-the-headlights effect for many practitioners as it became increasingly difficult to refute a new source of forensic evidence.

3. COLLECTING DATA TO EVALUATE THE PROBLEM

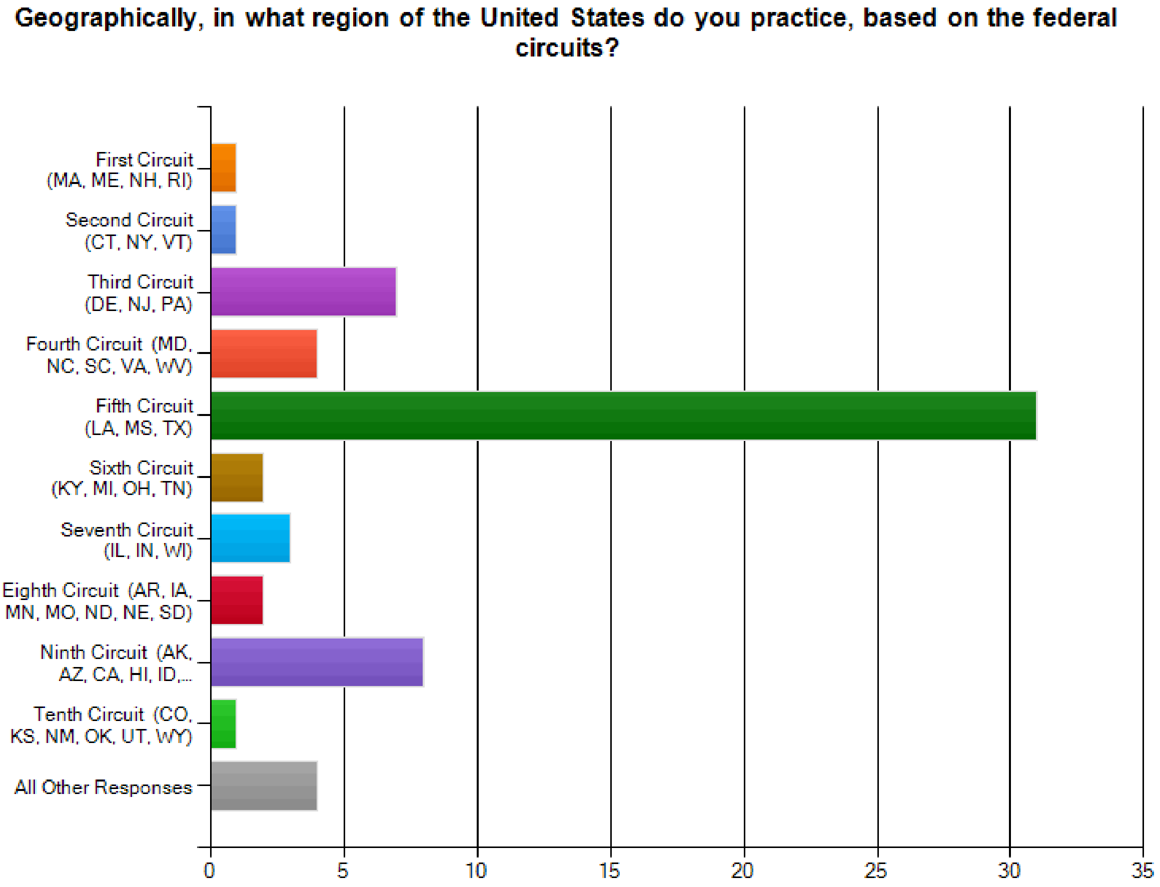
In order to properly evaluate our problem beyond subjective case law analysis, a survey was developed consisting of thirty-nine questions designed to provide answers about respondents' professional background, technical knowledge, and use of digital forensics evidence in the courtroom. This survey was only made available to attorneys licensed in the United States. Specific questions solicited information about participant attitudes, knowledge and experience with digital forensics, legal education, practice specializations, geographic practice regions by Federal Circuit, the ability of participants to identify knowledgeable digital forensic experts, and willingness of participants to take CLE courses in digital forensics. The survey was designed using the resources of a subscription service, SurveyMonkey.com and was available to participants by clicking a URL address provided to participants on the Internet.

3.1 Survey Participants

Survey participants were originally selected on a judgment basis based upon email listings obtained from professional journals, web site listings, telephone directory advertisements, court documents, and prior business dealings with the law firms. This circularization was done in August and September 2010, and consisted of approximately 1,100 direct email and fax survey solicitation requests directed to attorneys in all eleven Federal Circuits. Emails and faxes were personalized in order to avoid identification of survey participation requests as spam. The response rate from participants using these survey solicitation methods was poor. Due to the poor response rate, solicitation of responses was then encouraged by listings on business oriented, legal profession networking websites on the Internet (LinkedIn), consisting of law school alumni, legal practice areas (for example family law, corporate law, prosecutors, and criminal law), and special-interest areas directed toward attorneys. The identity of the respondents was anonymous to ensure candid answers.

The estimated number of attorneys per each of the fifteen LinkedIn groups selected was determined by reviewing membership listings by profession and determining the number of attorneys from sample pages selected on a judgment basis. The potential population was estimated to be in excess of 15,600. Using these circularization methods seventy-nine responses were received by November 11, 2010. Of these responses, sixty-six participants completed all thirty-nine questions.

Responses were received from each of the Federal Circuits; however, survey results were geographically biased based upon participant responses which were heavily weighted to the Fifth Circuit (Louisiana, Texas and Mississippi). This was attributable to many of the respondents being attorneys on legal list servers in the Greater Houston area, and a significant number of attorneys responding who were alumni of South Texas College of Law.



3.2 Summarizing Data Collected for Analysis

In order to determine the weight of responses provided by participants, and therefore to determine the significance of survey answers to our problem, survey questions were divided into eleven distinct categories (Table 2 – Response Rank Based on Category). Category weight was then determined by the ratio of questions by category to the number of total questions. Using the average number of responses per question, a response rank per category of (1 = most responses per question, to 11 = least responses per question) was assigned to each category for the purpose of determining the completeness of answers. The average number of responses for all questions was 53.67.

Table 2: Response Rank Based on Category

Question Category and Question Numbers	Number of Questions in Category	Category Weight	Total Responses for all Questions in Category	Response Rank Based on Responses	Average Number of Responses Per Question
Education (1 thru 3)	3	7.69%	89	11	29.7
Continuing Education (4 thru 7)	4	10.26%	197	7	49.3

Courtroom Experience (8 thru 12)	5	12.82%	268	6	53.6
Discovery (13 thru 17)	5	12.82%	294	5	58.8
Expert Testimony (18 thru 21)	4	10.26%	171	10	42.8
Admissibility of Evidence (22 thru 23)	2	5.13%	96	8	48.0
Expert Credentials (24 thru 25)	2	5.13%	89	9	44.5
Attorney Subject Knowledge (26 thru 30)	5	12.82%	325	2	65.0
Professional Specialization (31 thru 34)	4	10.26%	236	4	59.0
Geographic Location (35 thru 36)	2	5.13%	130	3	65.0
Experience (37 thru 39)	3	7.68%	198	1	66.0

In order to identify questions that reflected a response rate representative of a significant statistical variance from the expected mean, the standard deviation of the population of 39 questions was calculated. The standard deviation was determined to be 18.33, thereby providing the expectation that approximately 68% of all responses in a normal distribution would be between 35.34 and 72.00. From this ten questions were identified as having response rates which were more than one standard deviation from the population mean of 53.67. Answers to these questions were isolated and further analyzed in order to determine if responses were possibly invalid based upon survey design or population bias, or if answers were reflective of an evolving trend or different knowledge base. Review of answers to these ten questions indicated that responses were consistent with expectations, the purpose of the survey, and the definition of the problem being reviewed.

Question #	Question	Responses	Reason for Variance
1.	Did you have any courses in law school which dealt in whole or part with digital forensics (computer forensics, cell phone forensics, e-discovery, etc.)?	79	Initial question in survey. All respondents answered.
2.	If the answer to question 1 was yes, were these topics: (a) In courses dedicated to the topic (i.e. "Digital Forensics and the Law), (b) Topics within another course (i.e. Evidence), (c) Both	5	Five respondents answered this question. Only 6.33% of the attorneys answering this survey had any courses in law school that addressed digital forensics issues. This was explained by Question 39 – "How long ago did you graduate from law school?" Of sixty-six respondents only nine (13.6%) indicated that they had graduated

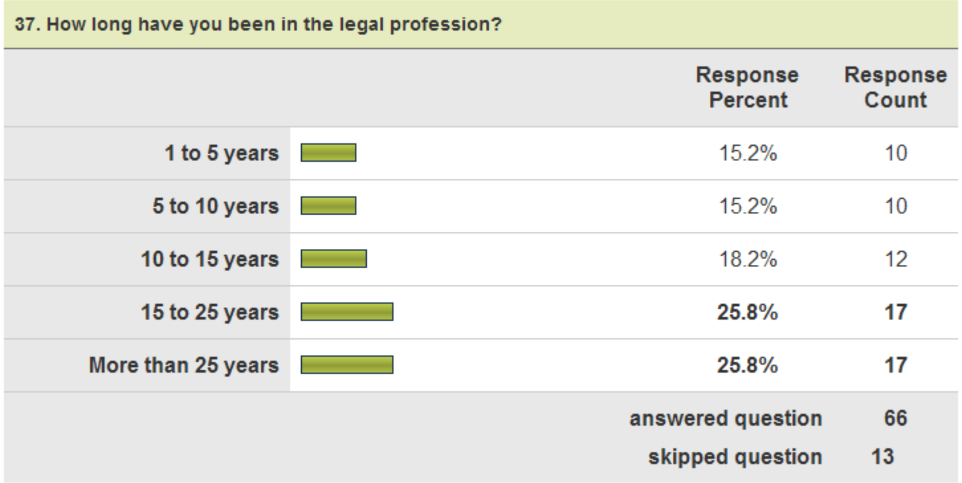
			within the last five years. This was consistent with the case law analysis earlier in this paper which indicated that 47.1% of the cases reviewed “reflected a clear misunderstanding of, or serious lack of knowledge with respect to the acquisition of computer forensics evidence and testimony provided to explain that evidence.”
3.	If your answer to question 1 was yes, did you feel the attention to the topic of digital forensics was adequate?	5	Of the five responses, only one respondent felt that the topic was adequately addressed. This represents only 1.27% of the survey responses.
4.	Have you taken any CLE courses on the topic of digital forensics (including e-discovery)?	78	Responses on this question were almost evenly split with forty respondents (51.3%) saying that they had taken CLE courses on digital forensics, and thirty-eight (48.7%) saying they hadn’t. This response was consistent with the interpretation of the case law analysis.
14.	If the answer to question 13 was yes, how knowledgeable do you feel the attorneys were with regard to their client’s e-discovery issues?	20	Only twenty respondents of sixty-nine answering question 13 had participated in a Rule 26(f) conference regarding e-discovery. This represented 28.99% of the attorneys responding to this question. Of this number only 10.00% were considered to be very knowledgeable. This represented 2.90% of all attorneys responding to question 13.
19.	If you have engaged a digital forensics expert, what services did they perform? (may choose more than one answer)	34	Thirty-four of sixty-nine respondents answered this question (49.28%). This represented a significant level of reliance on expert witnesses in this area. This response did not correspond to the analysis of cases where only three defense computer forensics expert witnesses were used in seventeen cases (17.6%), however, it closely correlated with the responses to Question 31 where 21.6% of respondents indicated that they were a judge, prosecutor, or defense attorney.
20.	If you have participated in litigation in which a digital forensics expert was used, do you feel they were effective?	34	Twenty-five of the thirty-nine respondents (73.5%) felt that a digital forensics expert was effective in litigation.

21.	If you have participated in litigation in which a digital forensics expert was used, was the information they provided understandable to the attorneys, the judge and the jury, if applicable?	34	Twenty-five of the thirty-four respondents (73.5%) felt that the information provided was "not at all understandable," or was "somewhat understandable." Only nine of the sixty-nine respondents completing this part of the survey (13.04%) felt that digital forensics information provided at trial was "very understandable."
23.	If you have participated in litigation in which a digital forensics expert was used, did the information provided by the expert play a role in the outcome of the case?	33	Of thirty-three respondents, thirteen (39.4%) felt that a digital forensics expert played a large role in the outcome of a case. Evaluated in conjunction with responses to Question 21 above it appears that responding attorneys felt that it was not necessary to understand digital forensics information presented at trial in order for it be highly effective in the outcome of a case. When this response is evaluated in light of the conclusions drawn from the case law analysis earlier in this paper it becomes apparent that on occasion computer forensics evidence is obfuscated at trial in an attempt to achieve a desired verdict. This conclusion is particularly disturbing because traditional gatekeepers in the form of professional training and education appear to be lacking.
24.	If you have engaged a digital forensics expert, what was their background? (may choose more than one)	31	Eleven professional groups were represented as possible answers for this question. No profession got more than 20% of total responses (CCE – Certified Computer Examiner), and all professions represented got at least one response. Consistent with the Obstacles to the Engagement of Computer Forensics Experts section of this paper, private investigators received the fifth highest response rate.

3.3 Review of Questions by Category Weight

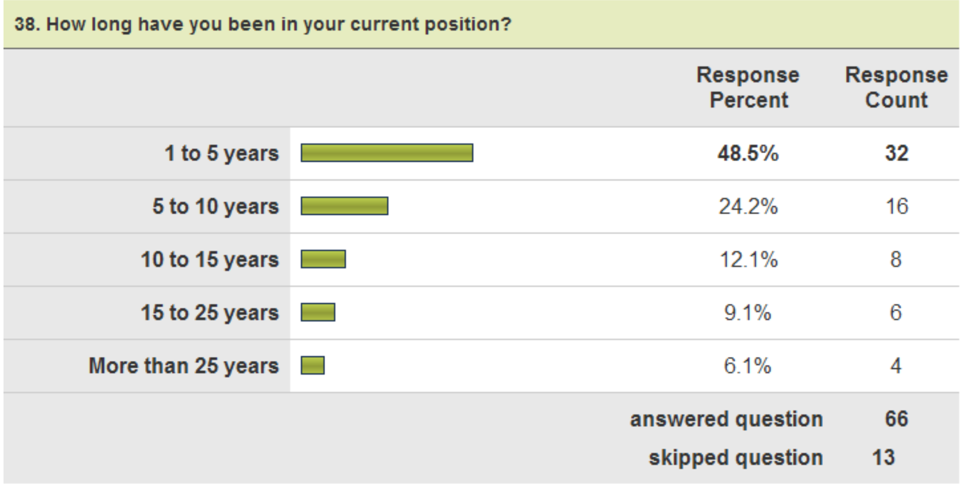
Questions 37 through 39 (EXPERIENCE) reflected the greatest category weight with 69.7% of respondents having been in the legal profession more than ten years. Over half of all respondents (51.6%) had been in the profession fifteen or more years which corresponded with more traditional law school educations (Question 37). Career mobility was also evident with almost half (48.5%) of those answering this question having been in their present position for less than five years (Question 38). Graduation from law school was also consistent with the number of years that respondents had been practicing law, with 54.5% of those answering the question indicating that they had graduated from law school fifteen or more years ago.

Taken as a whole, answers to the EXPERIENCE category were reflective of a mature, upwardly mobile sample of attorneys who were advancing in their careers, but had been, in all likelihood based upon their age, educated in a traditional law school environment.



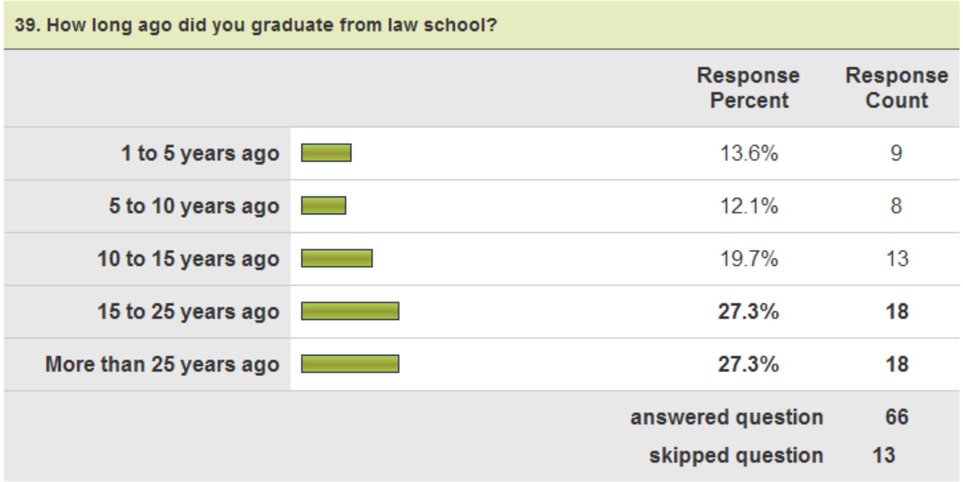
37. How long have you been in the legal profession?

	Response Percent	Response Count
1 to 5 years	15.2%	10
5 to 10 years	15.2%	10
10 to 15 years	18.2%	12
15 to 25 years	**25.8%**	**17**
More than 25 years	**25.8%**	**17**
	answered question	66
	skipped question	13



38. How long have you been in your current position?

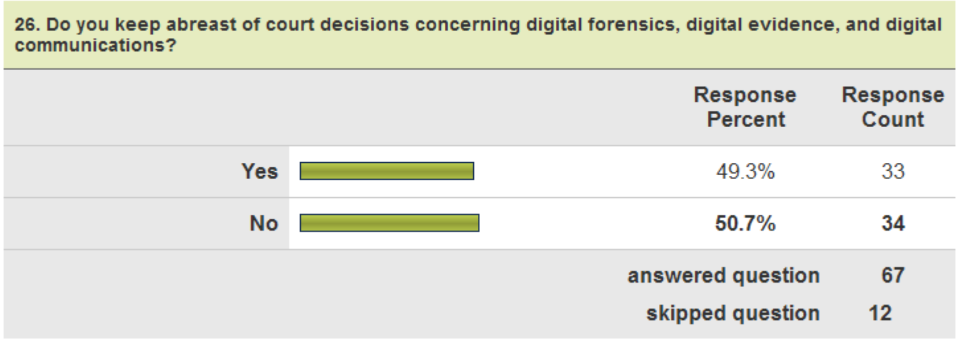
	Response Percent	Response Count
1 to 5 years	**48.5%**	**32**
5 to 10 years	24.2%	16
10 to 15 years	12.1%	8
15 to 25 years	9.1%	6
More than 25 years	6.1%	4
	answered question	66
	skipped question	13



39. How long ago did you graduate from law school?

	Response Percent	Response Count
1 to 5 years ago	13.6%	9
5 to 10 years ago	12.1%	8
10 to 15 years ago	19.7%	13
15 to 25 years ago	**27.3%**	**18**
More than 25 years ago	**27.3%**	**18**
	answered question	66
	skipped question	13

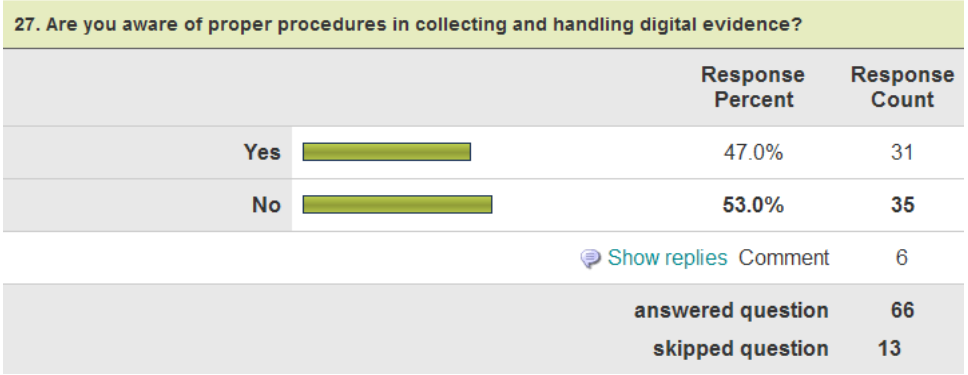
Questions 26 through 30 (ATTORNEY SUBJECT KNOWLEDGE) reflected the second-highest category weight with almost half of all attorneys answering this question (49.3%) indicating that they stay current with court decisions concerning digital forensics, digital evidence, and digital communications (Question 26).



26. Do you keep abreast of court decisions concerning digital forensics, digital evidence, and digital communications?

	Response Percent	Response Count
Yes	49.3%	33
No	**50.7%**	**34**
	answered question	67
	skipped question	12

Question 27 was more indicative, however, of the actual level of technical knowledge that attorneys responding had with respect to proper procedures in the collection and handling of digital evidence. 47.0% indicated that they were knowledgeable, but none of the additional responses left by six of the sixty-six were representative of a great degree of individual knowledge or confidence.



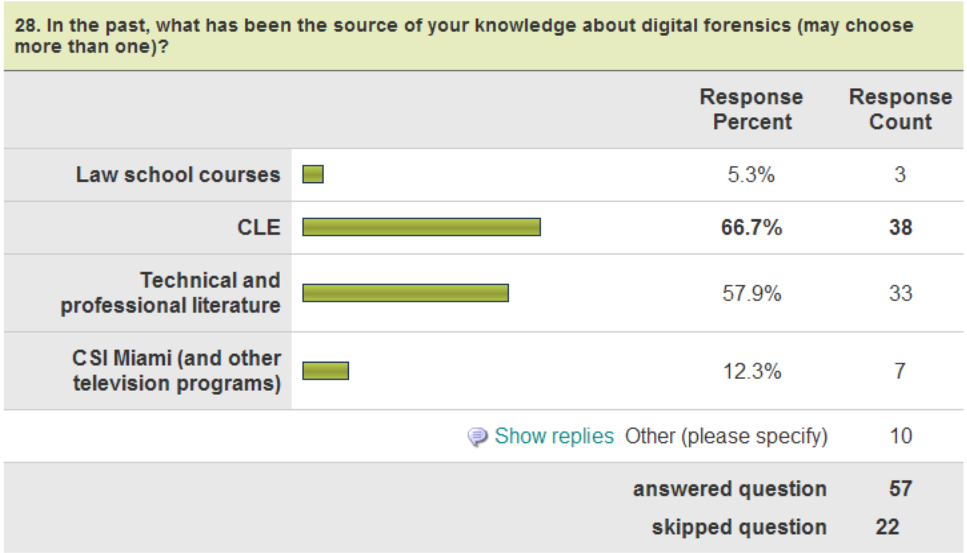
27. Are you aware of proper procedures in collecting and handling digital evidence?

	Response Percent	Response Count
Yes	47.0%	31
No	**53.0%**	**35**
	Show replies Comment	6
	answered question	**66**
	skipped question	**13**

Responses to Question 27 –

1. "Somewhat, at least aware of how to research case law and seminar materials to find the procedures if the issue may be relevant in a case."
2. "Somewhat."
3. "I would have checked "somewhat" if that had been an option."
4. "I'm not at all oblivious to the problem posed, but I don't claim to know what the proper procedures are."
5. "Not sure what is meant by "proper procedures." We have internal procedures to retain and collect digital information."
6. "I am not aware of all of the specifics, but I have access to individuals and experts for consultation, if necessary."

Question 28, which was answered by fifty-seven people, provided a measure of where attorneys surveyed are getting information about digital forensics. Personal responses were varied and indicative of a small group of the attorneys having nontraditional career backgrounds and educations before they entered law school. This was as compared to traditional undergraduate educations in liberal arts, business, and political science, which have been the normal foundation. (It should be noted that on some questions that respondents could select more than one answer. Due to this the Response Percent totals to more than 100%.)



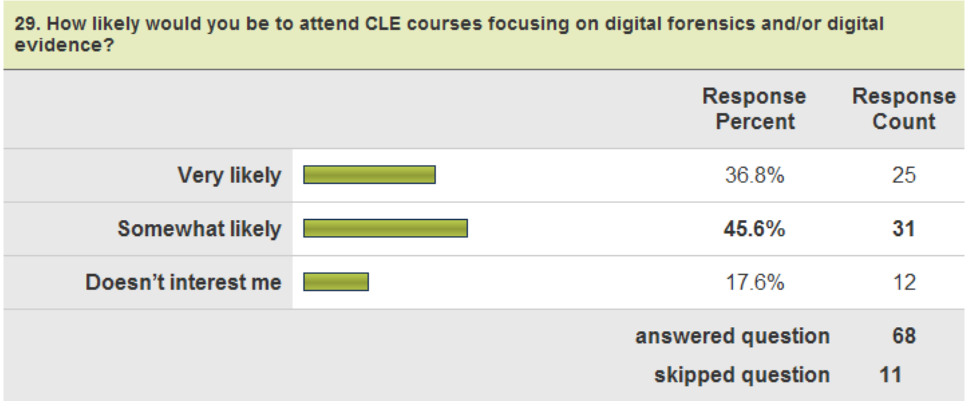
28. In the past, what has been the source of your knowledge about digital forensics (may choose more than one)?

	Response Percent	Response Count
Law school courses	5.3%	3
CLE	**66.7%**	**38**
Technical and professional literature	57.9%	33
CSI Miami (and other television programs)	12.3%	7
Show replies Other (please specify)		10
answered question		57
skipped question		22

Responses to Question 28 –

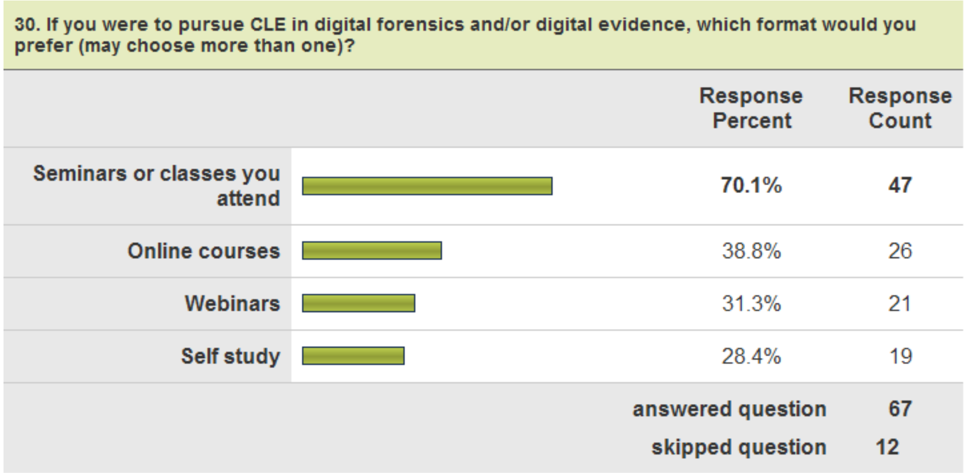
1. “My home was one of the very first adopters of personal computers. My mother was a computer analyst. My undergraduate major was in computer science.”
2. “Aaron Hughes.”
3. “I ask my tech guy when I have a question.”
4. “Interest in computers.”
5. “The problem is that what I’ve seen or read or heard has been limited, so far.”
6. “Personal, professional experience as a digital forensic examiner. Daily contact with digital forensic examiners.”
7. “I am an Electrical Engineer and Computer Engineer who spent 12 years as a R&D engineer for a major computer company before attending law school.”
8. “On the job.”
9. “CSI is not a source of knowledge.”
10. “Discussions with IT professional.”

Questions 29 and 30 address the receptiveness and interest of practicing attorneys in taking CLE courses focused on digital forensics and digital evidence. Participants were very receptive to this subject area with 82.3% of all respondents either being “Somewhat likely” or “Very likely” to attend a CLE course on these objects. The favored delivery method was seminars or classes.



29. How likely would you be to attend CLE courses focusing on digital forensics and/or digital evidence?

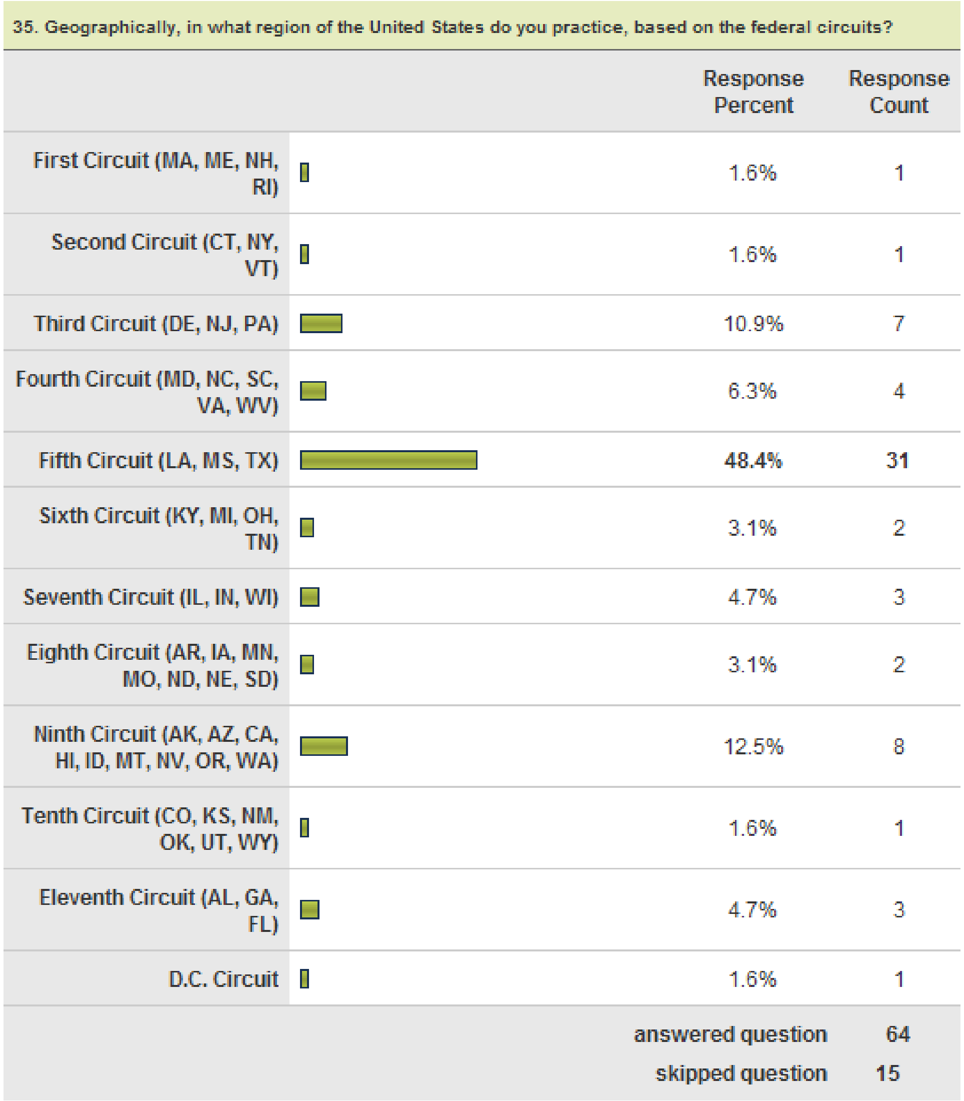
	Response Percent	Response Count
Very likely	36.8%	25
Somewhat likely	**45.6%**	**31**
Doesn't interest me	17.6%	12
answered question		68
skipped question		11



30. If you were to pursue CLE in digital forensics and/or digital evidence, which format would you prefer (may choose more than one)?

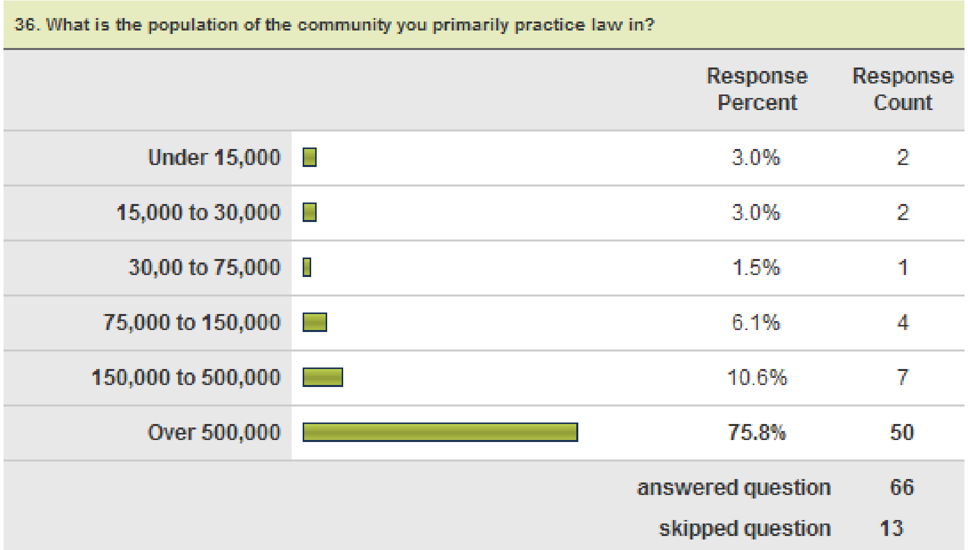
	Response Percent	Response Count
Seminars or classes you attend	**70.1%**	**47**
Online courses	38.8%	26
Webinars	31.3%	21
Self study	28.4%	19
answered question		67
skipped question		12

Questions 35 and 36 (GEOGRAPHIC LOCATION) represented the third highest category weight. As explained in the Survey Participants section of this paper, participant sample selection was biased based upon the large number of participants (48.4%) practicing law in the Fifth U.S. Circuit (Texas, Louisiana, and Mississippi). Sample participants, however, were largely homogeneous with 75.8% of all responses being from attorneys that practice in urban regions of 500,000 or more people. This implies that the majority of practices might be more similar than dissimilar.



35. Geographically, in what region of the United States do you practice, based on the federal circuits?

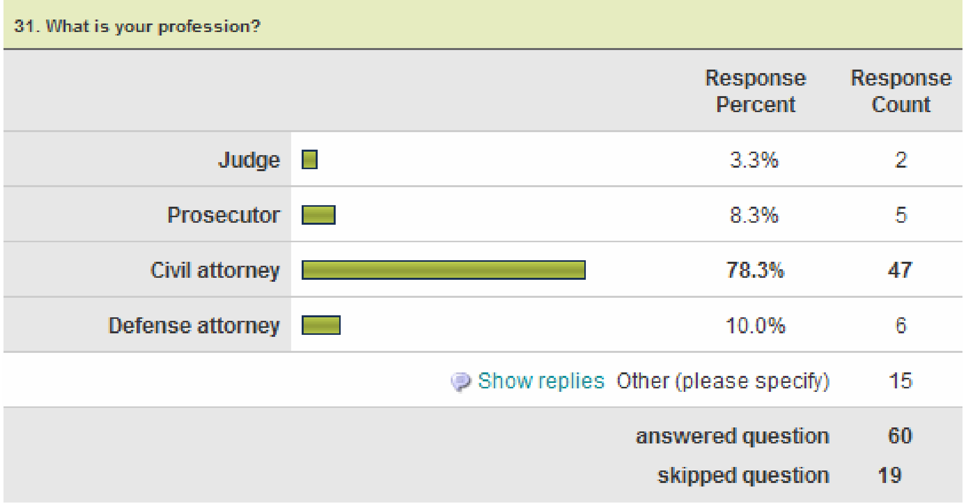
	Response Percent	Response Count
First Circuit (MA, ME, NH, RI)	1.6%	1
Second Circuit (CT, NY, VT)	1.6%	1
Third Circuit (DE, NJ, PA)	10.9%	7
Fourth Circuit (MD, NC, SC, VA, WV)	6.3%	4
Fifth Circuit (LA, MS, TX)	**48.4%**	**31**
Sixth Circuit (KY, MI, OH, TN)	3.1%	2
Seventh Circuit (IL, IN, WI)	4.7%	3
Eighth Circuit (AR, IA, MN, MO, ND, NE, SD)	3.1%	2
Ninth Circuit (AK, AZ, CA, HI, ID, MT, NV, OR, WA)	12.5%	8
Tenth Circuit (CO, KS, NM, OK, UT, WY)	1.6%	1
Eleventh Circuit (AL, GA, FL)	4.7%	3
D.C. Circuit	1.6%	1
	answered question	64
	skipped question	15



36. What is the population of the community you primarily practice law in?

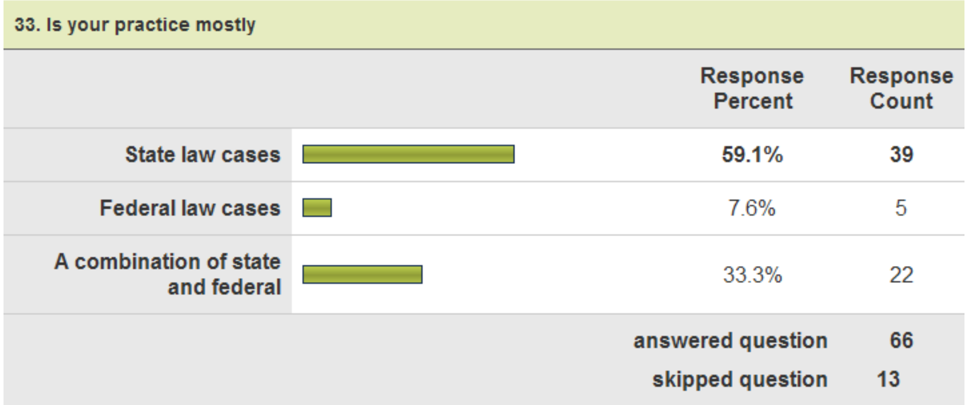
	Response Percent	Response Count
Under 15,000	3.0%	2
15,000 to 30,000	3.0%	2
30,00 to 75,000	1.5%	1
75,000 to 150,000	6.1%	4
150,000 to 500,000	10.6%	7
Over 500,000	**75.8%**	**50**
	answered question	66
	skipped question	13

PROFESSIONAL SPECIALIZATION (Questions 31 through 34) was the fourth highest ranked category based on the number of responses. Judges, prosecutors and defense attorneys were in the minority constituting 21.7% of total responses.



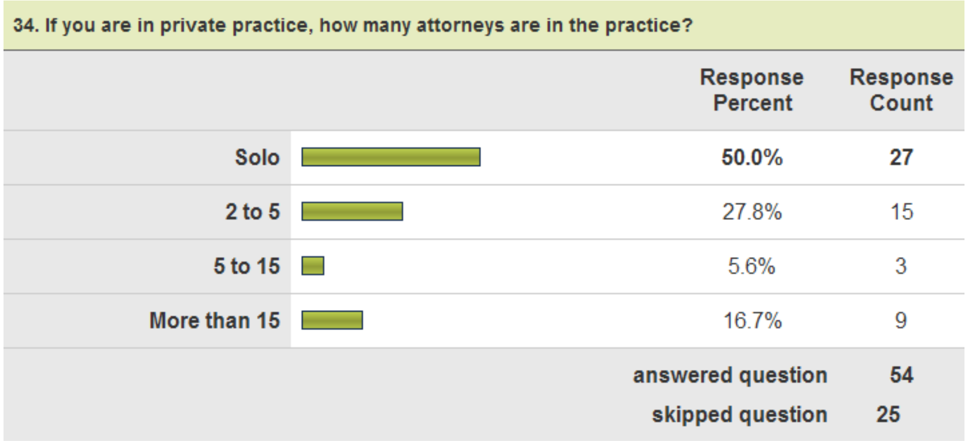
31. What is your profession?

	Response Percent	Response Count
Judge	3.3%	2
Prosecutor	8.3%	5
Civil attorney	**78.3%**	**47**
Defense attorney	10.0%	6
Show replies Other (please specify)		15
answered question		60
skipped question		19



33. Is your practice mostly

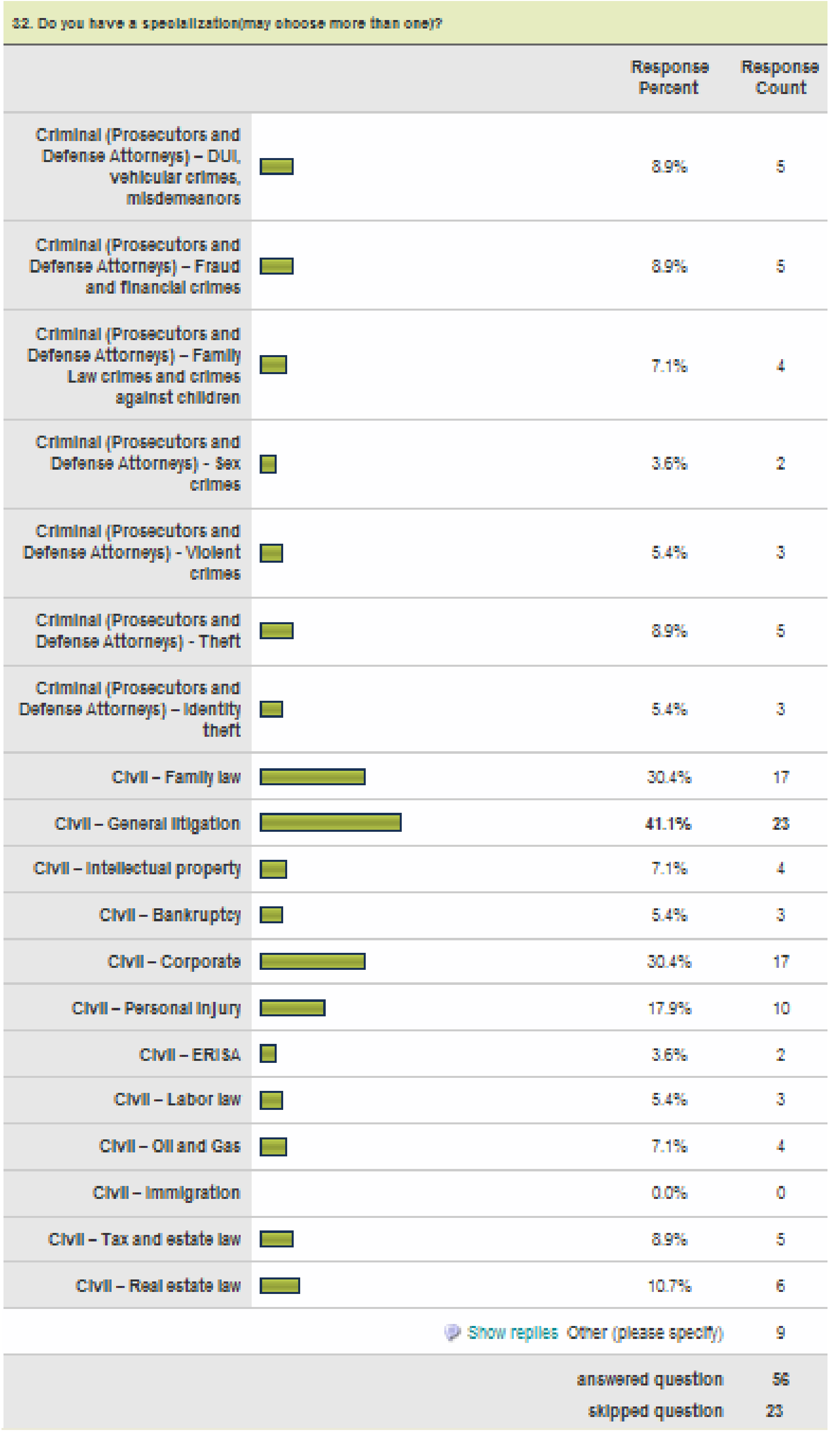
	Response Percent	Response Count
State law cases	**59.1%**	**39**
Federal law cases	7.6%	5
A combination of state and federal	33.3%	22
answered question		66
skipped question		13



34. If you are in private practice, how many attorneys are in the practice?

	Response Percent	Response Count
Solo	**50.0%**	**27**
2 to 5	27.8%	15
5 to 15	5.6%	3
More than 15	16.7%	9
answered question		54
skipped question		25

All practice specializations (Question 32) except “Civil-Immigration” had two or more responses. The most significant practice areas were “Civil-General litigation” with 19.0% of total responses, “Civil-Family law” with 14.0%, and “Civil-Corporate” with 14.0%. All criminal categories represented 22.3% of all responses, with criminal categories that are most indicative of using digital forensics evidence (fraud and financial crimes, family law and crimes against children, sex crimes, violent crimes, and identity theft), representing 14.0% of all responses to this question. Reponses to survey Questions 33 and 34 indicated that attorneys in practice, in most instances, were solo practitioners or were in practice units that consisted of less than five attorneys (77.8%). This was further reflective of respondents having to wear “multiple hats,” being driven to “case-driven pragmatic” solutions, and eschewing “elegant solutions” that would be prevalent in an academic-driven or theoretical environment. This is a sign of a profession being driven from “billable hours” to “fixed-fee-contracts,” and the difficulty of collecting professional fees, and in some instances the fees of expert witnesses, from clients that do not receive a favorable outcome at trial.

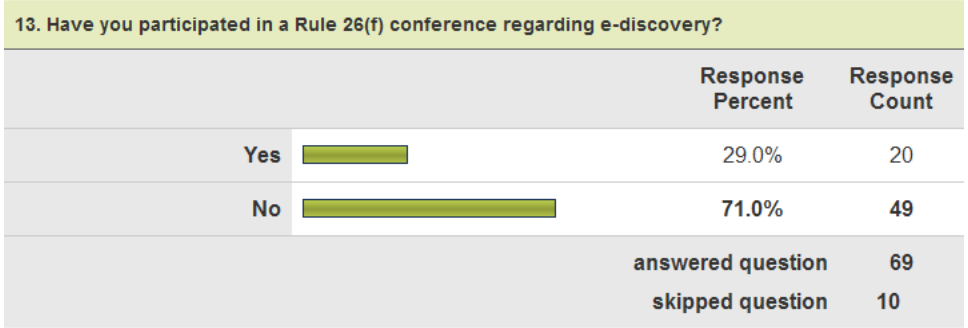


32. Do you have a specialization(may choose more than one)?

	Response Percent	Response Count
Criminal (Prosecutors and Defense Attorneys) – DUI, vehicular crimes, misdemeanors	8.9%	5
Criminal (Prosecutors and Defense Attorneys) – Fraud and financial crimes	8.9%	5
Criminal (Prosecutors and Defense Attorneys) – Family Law crimes and crimes against children	7.1%	4
Criminal (Prosecutors and Defense Attorneys) - Sex crimes	3.6%	2
Criminal (Prosecutors and Defense Attorneys) - Violent crimes	5.4%	3
Criminal (Prosecutors and Defense Attorneys) - Theft	8.9%	5
Criminal (Prosecutors and Defense Attorneys) – Identity theft	5.4%	3
Civil – Family law	30.4%	17
Civil – General litigation	41.1%	23
Civil – Intellectual property	7.1%	4
Civil – Bankruptcy	5.4%	3
Civil – Corporate	30.4%	17
Civil – Personal injury	17.9%	10
Civil – ERISA	3.6%	2
Civil – Labor law	5.4%	3
Civil – Oil and Gas	7.1%	4
Civil – Immigration	0.0%	0
Civil – Tax and estate law	8.9%	5
Civil – Real estate law	10.7%	6
Show replies Other (please specify)		9
answered question		56
skipped question		23

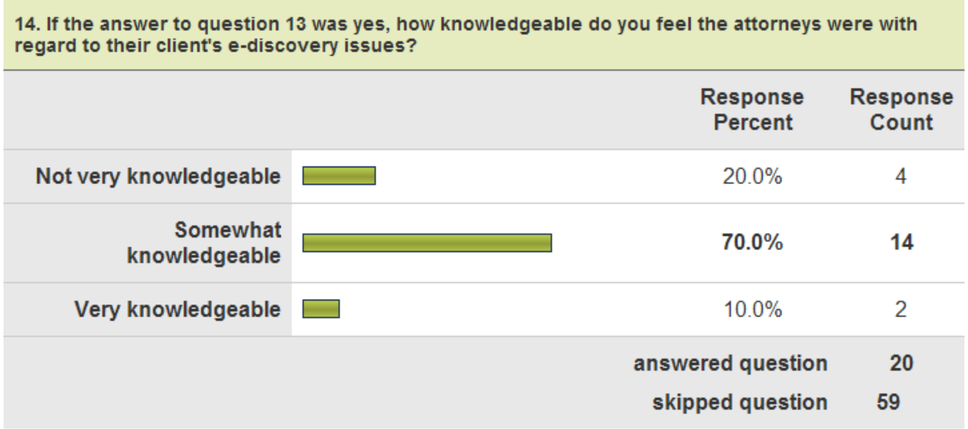
DISCOVERY (Questions 13 through 17) is central to all litigation, but questions in this category were weighted in fifth place. This relatively low ranking in relationship to the importance of this area reflects that responding attorneys did not have very much experience with e-discovery (Questions 13 and 14), did not routinely use preservation letters detailing digital evidence to be retained (Question 16), and received preservation letters infrequently (Question 17). The ability of responding attorneys to correctly identify sources of digital evidence (Question 15) was very good on an overall basis, but based on the earlier analysis of the answers to the ATTORNEY SUBJECT KNOWLEDGE questions establishment of a "link" between knowing where digital evidence can be found, and requesting that information in discovery is not very strong. In short, as reflected in the EXPERIENCE category questions attorneys responding to this survey were primarily trained in a classical law school environment that did not place emphasis on forensic sciences.

Responses in this question area reinforced observations from the Conclusions from Case Law section of this paper – "that trial tactics used by the defense, statements made by the state, or rulings of the trial court or the appeals court reflected a clear misunderstanding of, or serious lack of knowledge with respect to the acquisition of computer forensics evidence and testimony provided to explain that evidence."



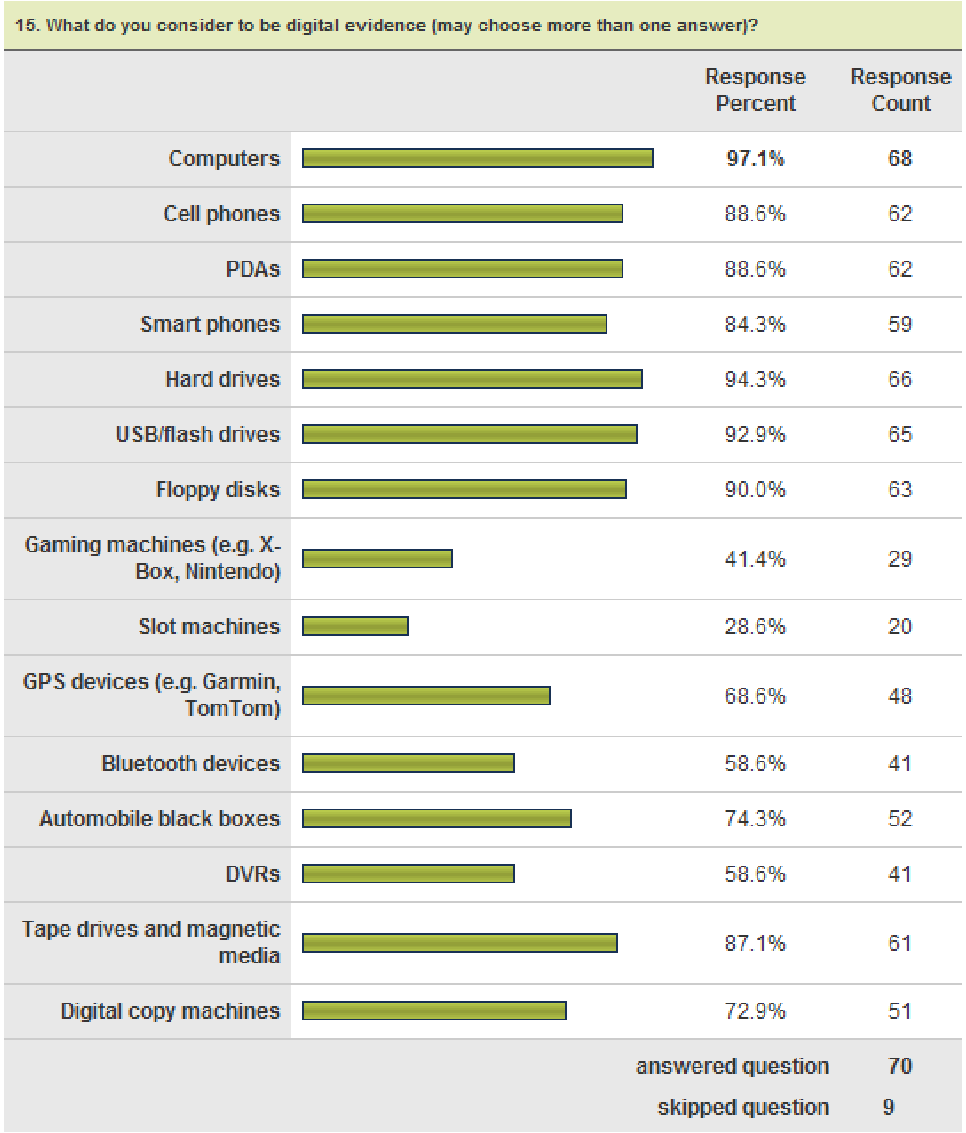
13. Have you participated in a Rule 26(f) conference regarding e-discovery?

	Response Percent	Response Count
Yes	29.0%	20
No	**71.0%**	**49**
answered question		69
skipped question		10



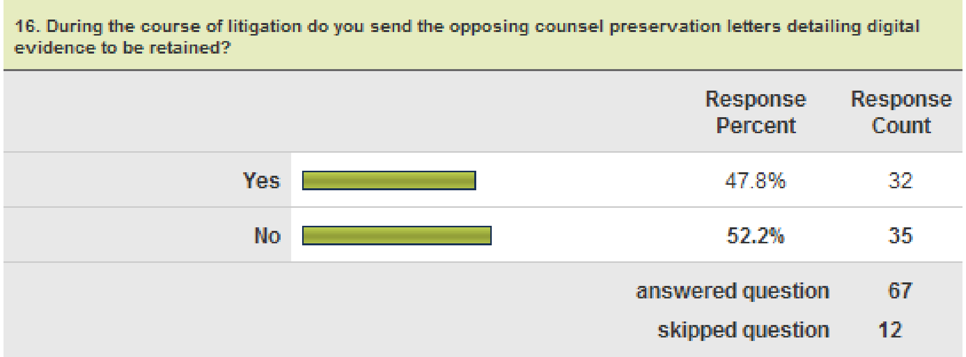
14. If the answer to question 13 was yes, how knowledgeable do you feel the attorneys were with regard to their client's e-discovery issues?

	Response Percent	Response Count
Not very knowledgeable	20.0%	4
Somewhat knowledgeable	**70.0%**	**14**
Very knowledgeable	10.0%	2
answered question		20
skipped question		59



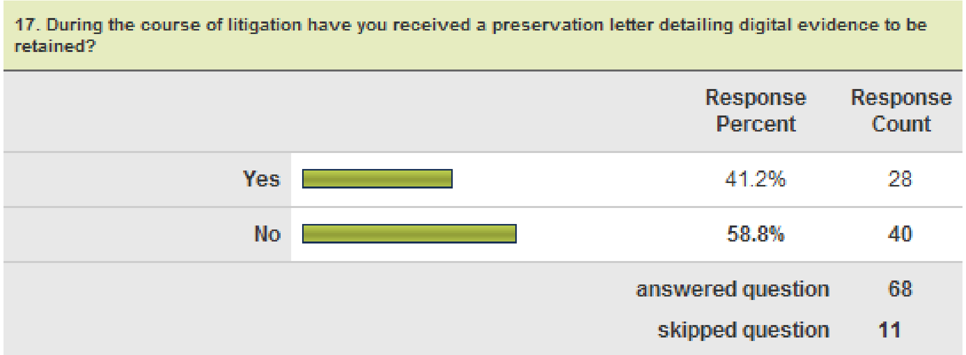
15. What do you consider to be digital evidence (may choose more than one answer)?

	Response Percent	Response Count
Computers	**97.1%**	**68**
Cell phones	88.6%	62
PDAs	88.6%	62
Smart phones	84.3%	59
Hard drives	94.3%	66
USB/flash drives	92.9%	65
Floppy disks	90.0%	63
Gaming machines (e.g. X-Box, Nintendo)	41.4%	29
Slot machines	28.6%	20
GPS devices (e.g. Garmin, TomTom)	68.6%	48
Bluetooth devices	58.6%	41
Automobile black boxes	74.3%	52
DVRs	58.6%	41
Tape drives and magnetic media	87.1%	61
Digital copy machines	72.9%	51
	answered question	70
	skipped question	9



16. During the course of litigation do you send the opposing counsel preservation letters detailing digital evidence to be retained?

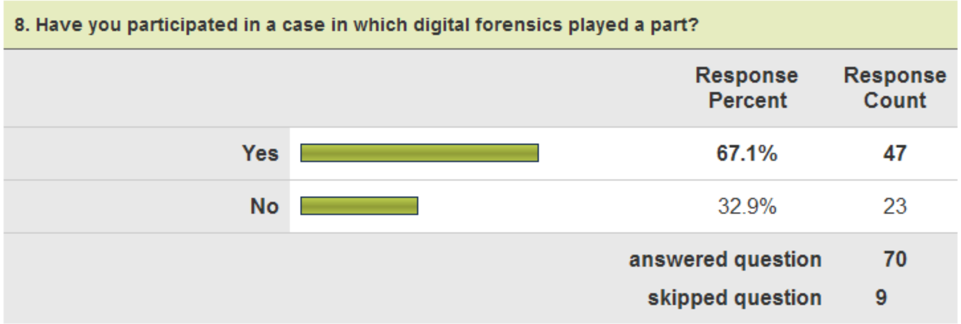
	Response Percent	Response Count
Yes	47.8%	32
No	52.2%	35
	answered question	67
	skipped question	12



17. During the course of litigation have you received a preservation letter detailing digital evidence to be retained?

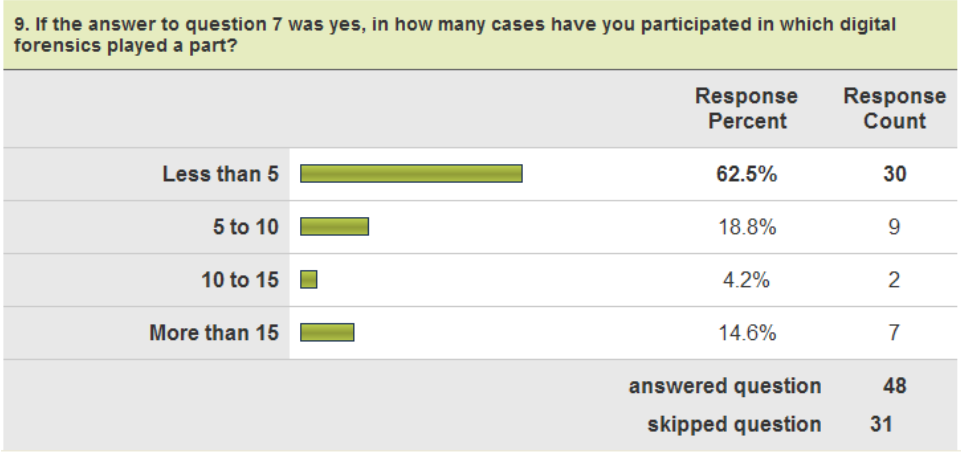
	Response Percent	Response Count
Yes	41.2%	28
No	58.8%	40
	answered question	68
	skipped question	11

Questions 8 through 12 addressed the COURTROOM EXPERIENCE of attorneys. Questions in this category were weighted in the sixth position according to response rate. Forty-seven of the seventy respondents (67.1%) who answered Question 8 – ("Have you participated in a case in which digital forensics played a part?") responded in the affirmative. Based upon responses to other sections of the survey this appears to be an unexpectedly high percentage, and taken in combination with responses in the EXPERIENCE, ATTORNEY SUBJECT KNOWLEDGE, and DISCOVERY question sections the matter has to be more carefully reviewed because the "courtroom skill level" of individual practitioners may be overstated based on self-assessment versus trial outcomes. Since this was a blind survey there is no way to reconcile individual responses with cases, verdicts, resources used, and jurisdictional prejudices. The conclusions from Case Law section of this paper also suggest that an overstatement of trial skills may be possible.



8. Have you participated in a case in which digital forensics played a part?

	Response Percent	Response Count
Yes	**67.1%**	**47**
No	32.9%	23
	answered question	**70**
	skipped question	**9**

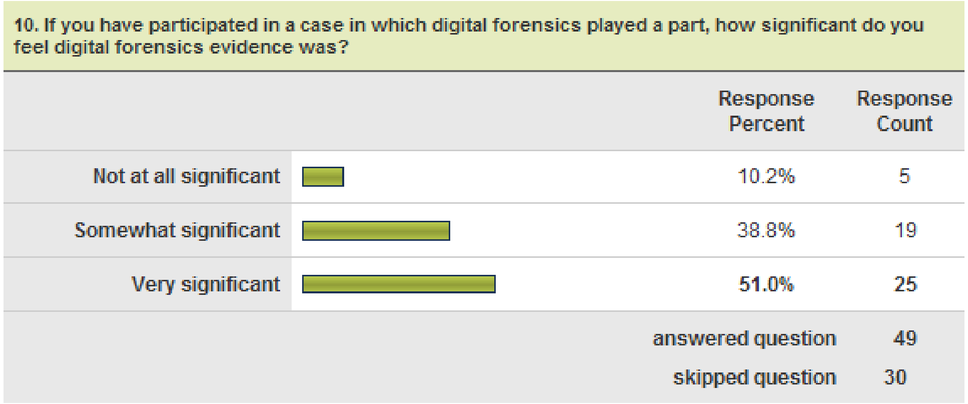


9. If the answer to question 7 was yes, in how many cases have you participated in which digital forensics played a part?

	Response Percent	Response Count
Less than 5	**62.5%**	**30**
5 to 10	18.8%	9
10 to 15	4.2%	2
More than 15	14.6%	7
	answered question	**48**
	skipped question	**31**

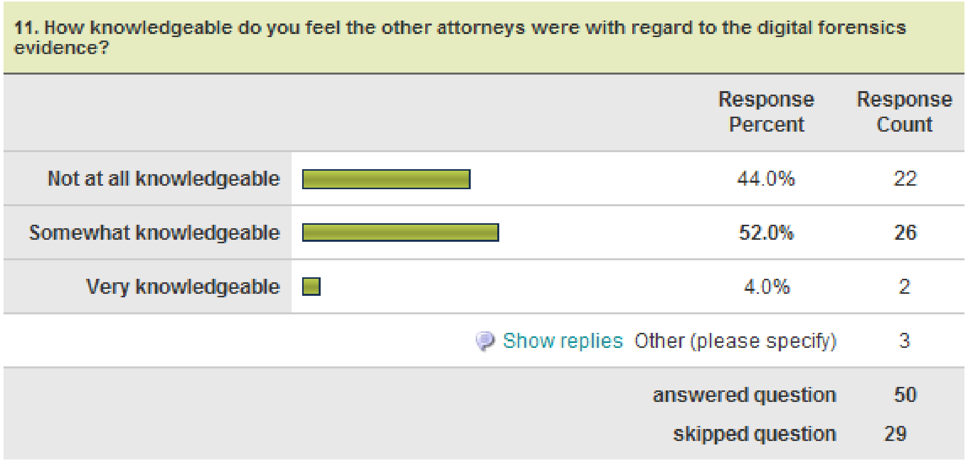
Questions 10 through 12 were particularly revealing because they provided an assessment of how responding attorneys felt about the significance of digital evidence, the knowledge base of opposing counsel, and the knowledge of judges hearing the cases. The later assessment was easily the most disturbing answer in the entire survey, with only two of fifty-one respondents (3.9%) answering that they felt the judges were very knowledgeable with regards to digital forensics evidence in their cases.

This response, of course, raises the question of: "If only one in twenty-five judges are rated as being very knowledgeable with regards to digital forensics evidence presented in cases in their courts, how are defendants' rights being protected with respect to the Sixth Amendment?" More importantly, does this support the theory that ineffective assistance of counsel is highly likely in many criminal cases rich in digital evidence, but that no one who could challenge the digital evidence knows enough to do it? That answer is beyond the scope of this paper, but it is a fertile ground for further inquiry.



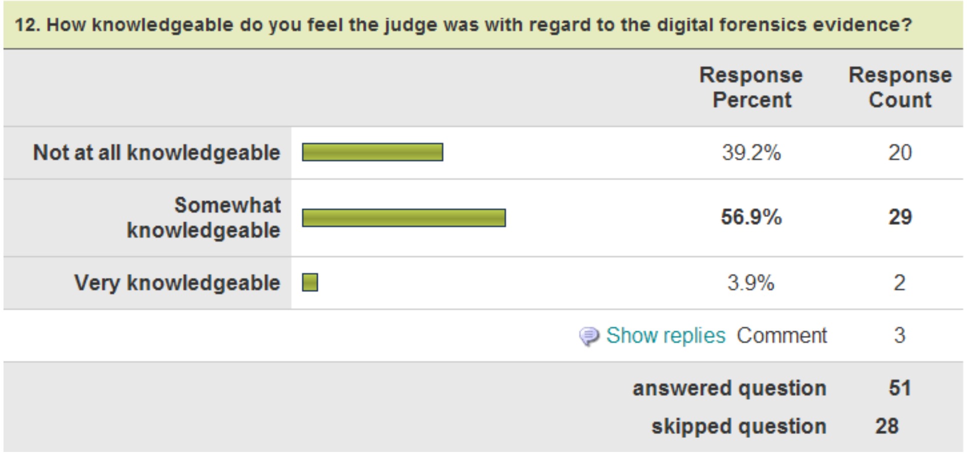
10. If you have participated in a case in which digital forensics played a part, how significant do you feel digital forensics evidence was?

	Response Percent	Response Count
Not at all significant	10.2%	5
Somewhat significant	38.8%	19
Very significant	**51.0%**	**25**
answered question		49
skipped question		30



11. How knowledgeable do you feel the other attorneys were with regard to the digital forensics evidence?

	Response Percent	Response Count
Not at all knowledgeable	44.0%	22
Somewhat knowledgeable	**52.0%**	**26**
Very knowledgeable	4.0%	2
Show replies Other (please specify)		3
answered question		50
skipped question		29

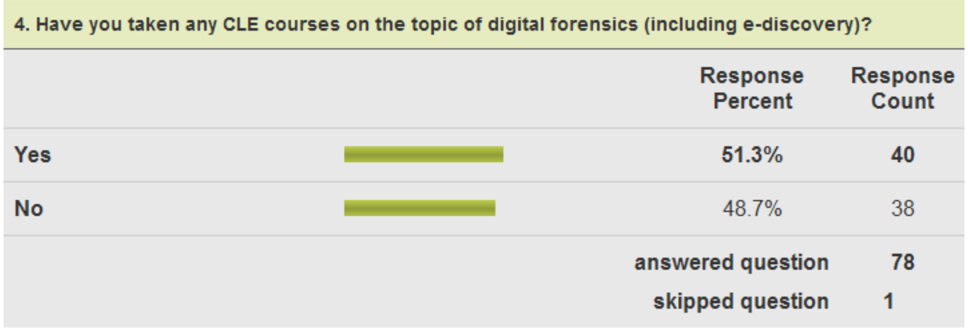


12. How knowledgeable do you feel the judge was with regard to the digital forensics evidence?

	Response Percent	Response Count
Not at all knowledgeable	39.2%	20
Somewhat knowledgeable	**56.9%**	**29**
Very knowledgeable	3.9%	2
Show replies Comment		3
answered question		51
skipped question		28

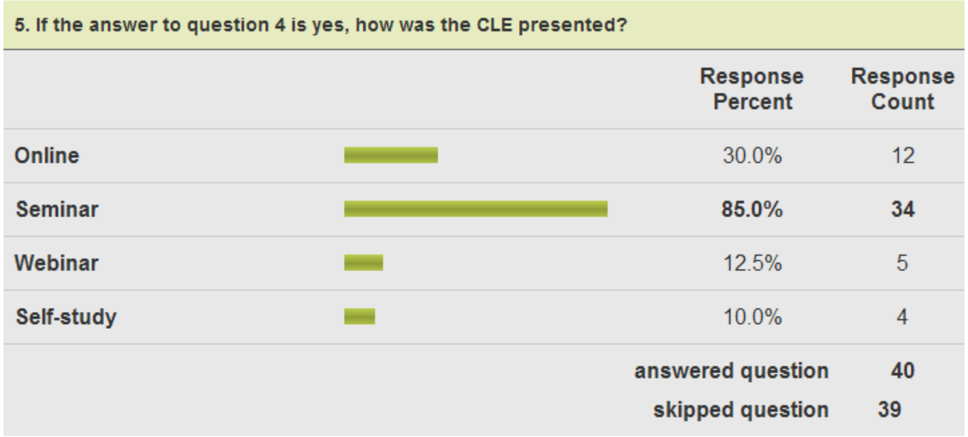
CONTINUING EDUCATION (Questions 4 through 7) was in the seventh position based on response rate. Question 4 reflected an almost even split between attorneys who have taken CLE courses that addressed digital forensics (51.3%) and attorneys who haven't (48.7%). To provide the proper context to these questions it is necessary to understand the position of CLE courses and the legal profession. Attorneys in Texas are required, as a condition for maintaining their license to practice law in the state, to take a minimum of fifteen mandatory hours of CLE per year. CLE is not mandatory in all states, and states that require it range from three hours per year (Alaska) to sixteen hours per year (New York) for new attorneys.

Question 4 responses indicate that forty participants who responded to this survey question have taken CLE courses which discussed digital forensics and/or e-discovery. Of this number, twenty-four (30.8%) of the original seventy-eight participants responding to Question 4 considered topics to be adequately covered.



4. Have you taken any CLE courses on the topic of digital forensics (including e-discovery)?

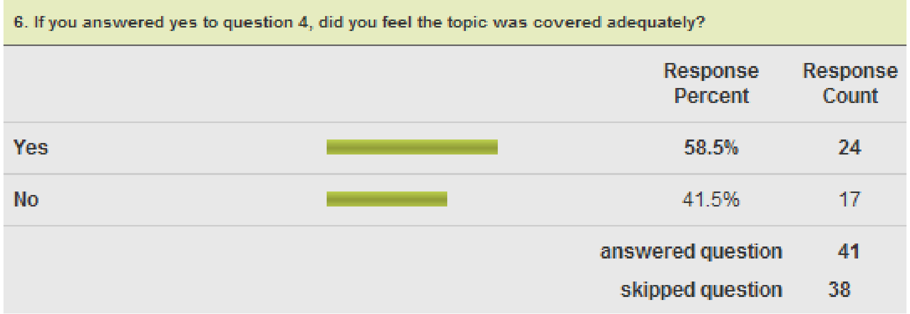
	Response Percent	Response Count
Yes	**51.3%**	**40**
No	48.7%	38
	answered question	**78**
	skipped question	**1**



5. If the answer to question 4 is yes, how was the CLE presented?

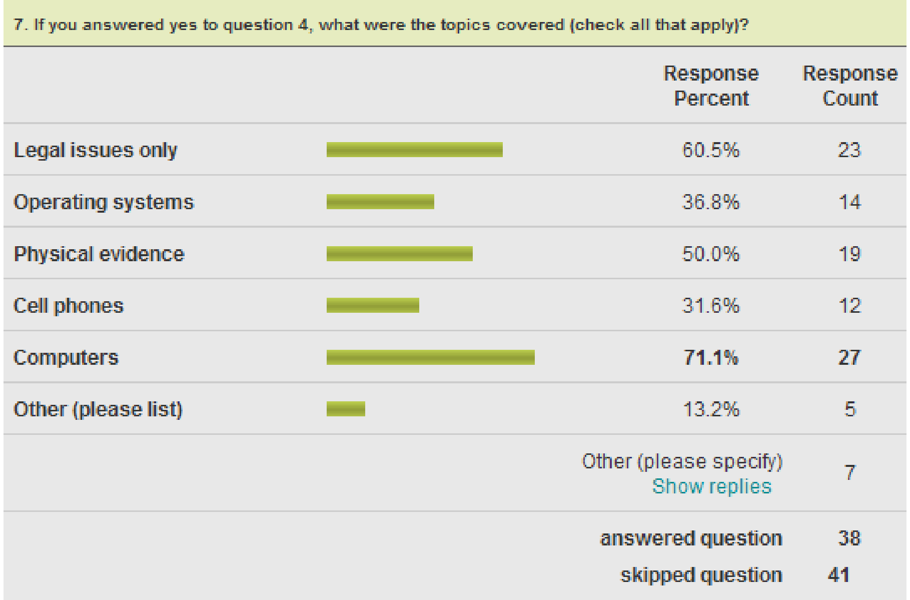
	Response Percent	Response Count
Online	30.0%	12
Seminar	**85.0%**	**34**
Webinar	12.5%	5
Self-study	10.0%	4
	answered question	**40**
	skipped question	**39**

Seven replies were left in the comments section for Question 7 by respondents. These responses provide a greater understanding of professional responsibilities and computer forensics knowledge, and provide context to information covered in CLE courses.



6. If you answered yes to question 4, did you feel the topic was covered adequately?

	Response Percent	Response Count
Yes	**58.5%**	**24**
No	41.5%	17
	answered question	41
	skipped question	38



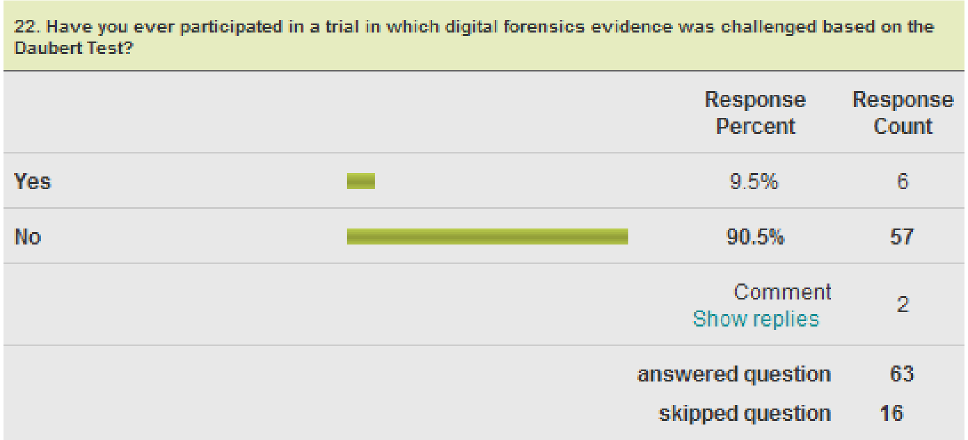
7. If you answered yes to question 4, what were the topics covered (check all that apply)?

	Response Percent	Response Count
Legal issues only	60.5%	23
Operating systems	36.8%	14
Physical evidence	50.0%	19
Cell phones	31.6%	12
Computers	**71.1%**	**27**
Other (please list)	13.2%	5
	Other (please specify) Show replies	7
	answered question	38
	skipped question	41

Responses to Question 7 –

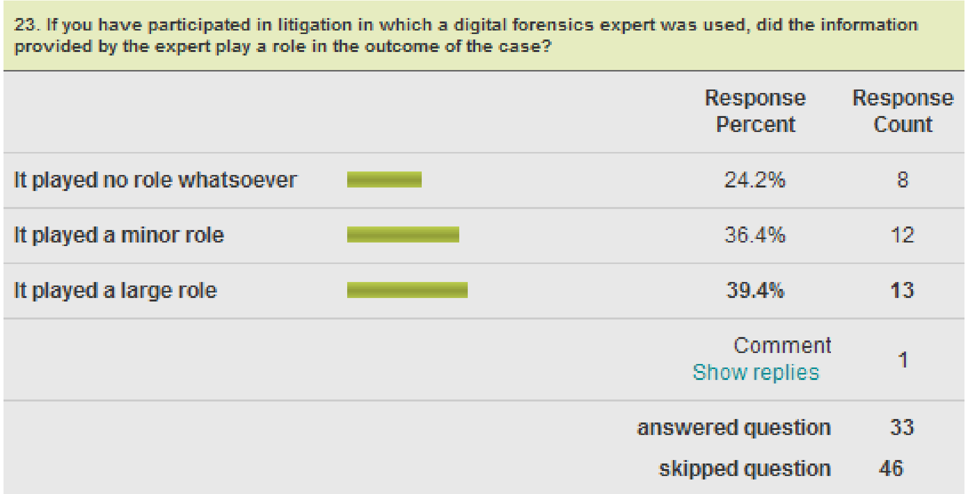
1. “e-mail.”
2. “More precisely: ‘beige boxing’.”
3. “While I haven’t taken any “courses,” I am – of necessity – well schooled in e-discovery legal issues (having managed complex, multi-party, corporate cases involving e-discovery), the vulnerabilities of operating systems, computers generally, wireless security, security vulnerability/evidence value/potential for anonymity of cell phones, cryptography, IP/TCP, etc.”
4. “Legal issues also, not legal issues only.”
5. “Possible sanctions for non-compliance; importance of litigation holds for electronic documents and information.”
6. “Social media and other forms of data that could (and likely is) relevant to a case.”
7. “Covered specifics minimally – recommendation is usually to engage an expensive forensic computer expert, which is not cost-effective or available in lower-value cases.”

Answers to Questions 22 and 23 (ADMISSIBILITY OF EVIDENCE) are indicative of a lack of overall experience on the part of survey respondents with respect to the application of Daubert (1993) as it applies to computer forensics evidence and expert witness testimony. Only six participants in this survey responded that they had "ever participated in a trial in which digital forensics evidence was challenged based on the Daubert Test." This represents only 7.6% of the participants who started this survey on Question 1, and when considered in conjunction with the COURTROOM EXPERIENCE questions, in particular Question 8, suggests courtroom "dust-ups" with respect to computer forensics evidence have been minimal. This may be because of: (1) the types of cases and subject matter, (2) resources available to trial counsel, (3) application of the principles of Daubert under some other theory of case law, (4) failure to see the Daubert Test as applying to digital evidence, (5) lack of experience, or in the worst case, (6) insufficiency of the judiciary. Dependent upon the, case these factors may collectively testify to ineffective assistance of counsel.



22. Have you ever participated in a trial in which digital forensics evidence was challenged based on the Daubert Test?

	Response Percent	Response Count
Yes	9.5%	6
No	90.5%	57
	Comment Show replies	2
	answered question	63
	skipped question	16



23. If you have participated in litigation in which a digital forensics expert was used, did the information provided by the expert play a role in the outcome of the case?

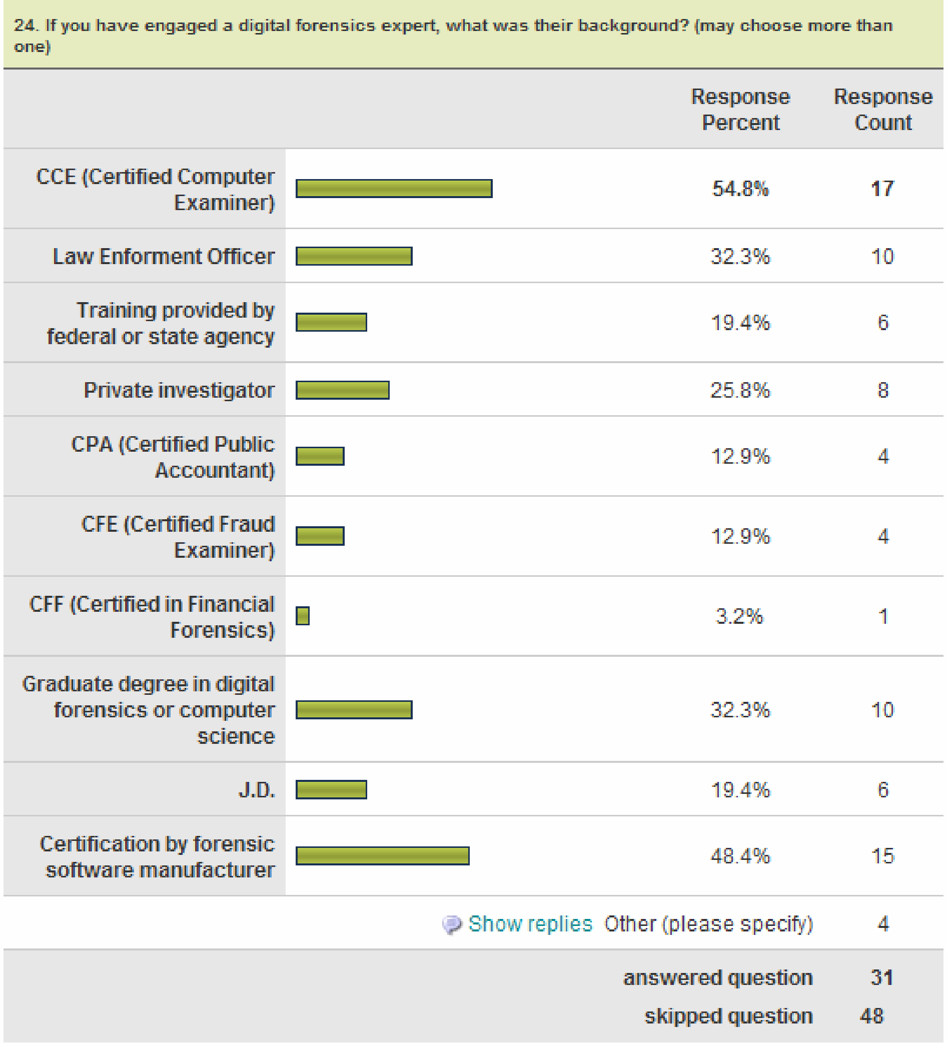
	Response Percent	Response Count
It played no role whatsoever	24.2%	8
It played a minor role	36.4%	12
It played a large role	39.4%	13
	Comment Show replies	1
	answered question	33
	skipped question	46

Responses to Question 22 –

1. "What is the daubert test?"
2. "Our state courts still apply Frye."

Digital forensics evidence as presented by expert witnesses was seen as very significant, however, and almost forty percent of responses to Question 23 indicated that it played a large role in case outcome.

EXPERT CREDENTIALS (Questions 24 and 25), which was the ninth ranked category, reflected a lack of consensus with respect to the professional qualifications of experts who have provided expert testimony for responding attorneys, and a responding affirmation of who should be providing digital forensics expert testimony in the future.

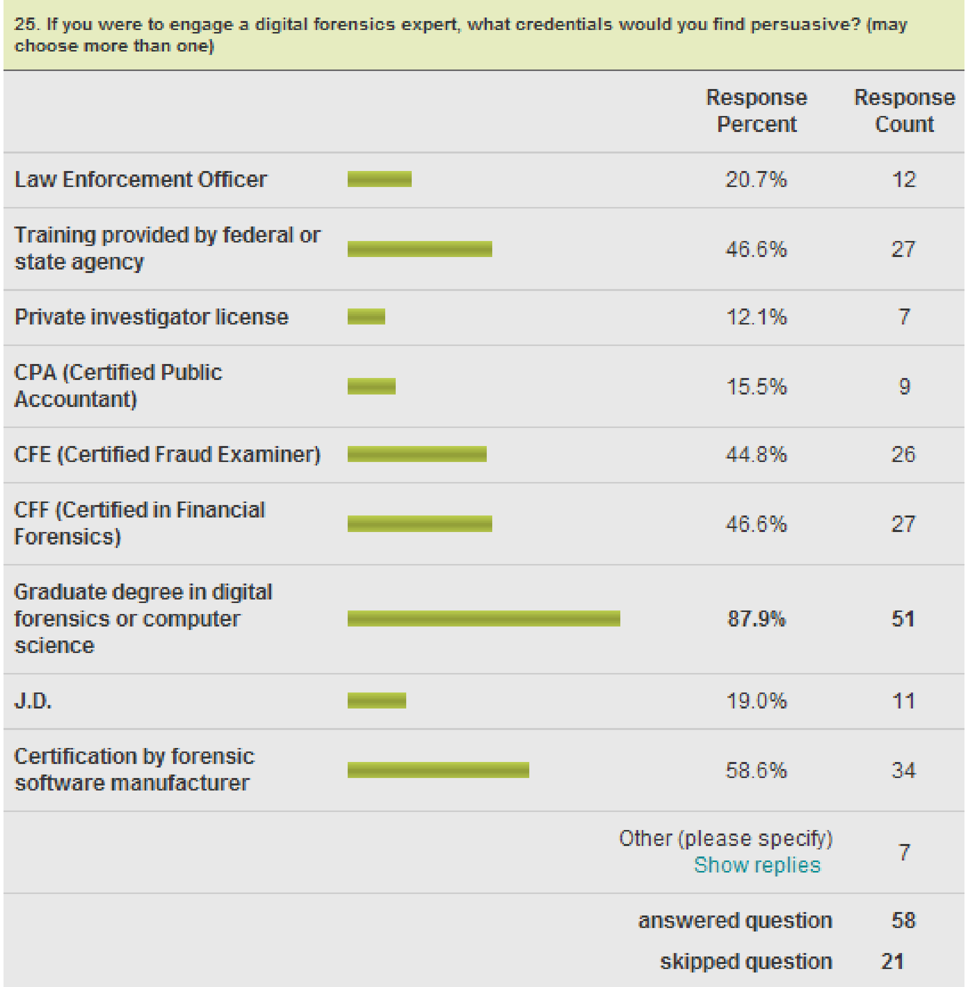


24. If you have engaged a digital forensics expert, what was their background? (may choose more than one)

	Response Percent	Response Count
CCE (Certified Computer Examiner)	54.8%	**17**
Law Enforment Officer	32.3%	10
Training provided by federal or state agency	19.4%	6
Private investigator	25.8%	8
CPA (Certified Public Accountant)	12.9%	4
CFE (Certified Fraud Examiner)	12.9%	4
CFF (Certified in Financial Forensics)	3.2%	1
Graduate degree in digital forensics or computer science	32.3%	10
J.D.	19.4%	6
Certification by forensic software manufacturer	48.4%	15
Show replies	Other (please specify)	4
	answered question	**31**
	skipped question	**48**

Responses to Question 24 –

1. "For one of the experts I don't recall his specific credential, but it was related to digital forensics/data recovery."
2. "My computer guy, flashed the hard drive, and then examined the results."
3. "Don't know."
4. "IT consultant."



25. If you were to engage a digital forensics expert, what credentials would you find persuasive? (may choose more than one)

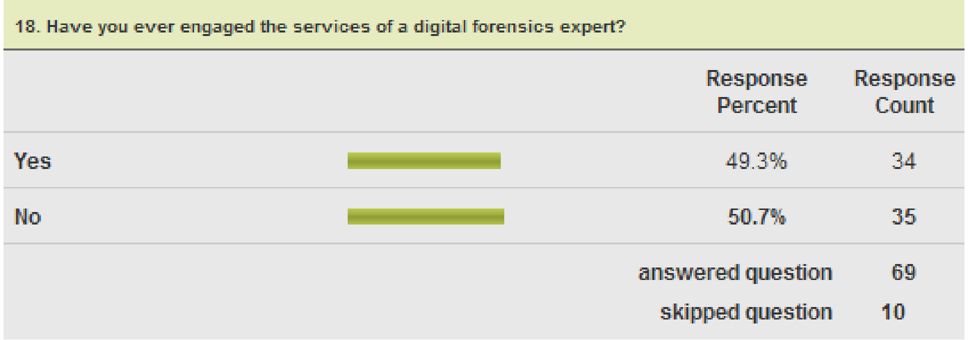
	Response Percent	Response Count
Law Enforcement Officer	20.7%	12
Training provided by federal or state agency	46.6%	27
Private investigator license	12.1%	7
CPA (Certified Public Accountant)	15.5%	9
CFE (Certified Fraud Examiner)	44.8%	26
CFF (Certified in Financial Forensics)	46.6%	27
Graduate degree in digital forensics or computer science	**87.9%**	**51**
J.D.	19.0%	11
Certification by forensic software manufacturer	58.6%	34
	Other (please specify) Show replies	7
	answered question	58
	skipped question	21

Responses to Question 25 –

1. “Not sure certification is that important.”
2. “By “training provided by federal or state agency,” I limit my answer to the FBI (particularly counterintelligence) and the intelligence community.”
3. “I’m a judge. How persuasive any of these credentials would be is unknown to me. Some are likely going to establish enough expertise for the witness to qualify as an expert, but other, e.g., CPA or training provided by a federal or state agency, or certification by forensic software manufacturer, I’d want to know what that’s all about.”
4. “CFCE.”
5. “Not sure.”
6. “E-discovery expert as certified by ACEDS or another organization.”
7. “Recommendation based on prior performance.”

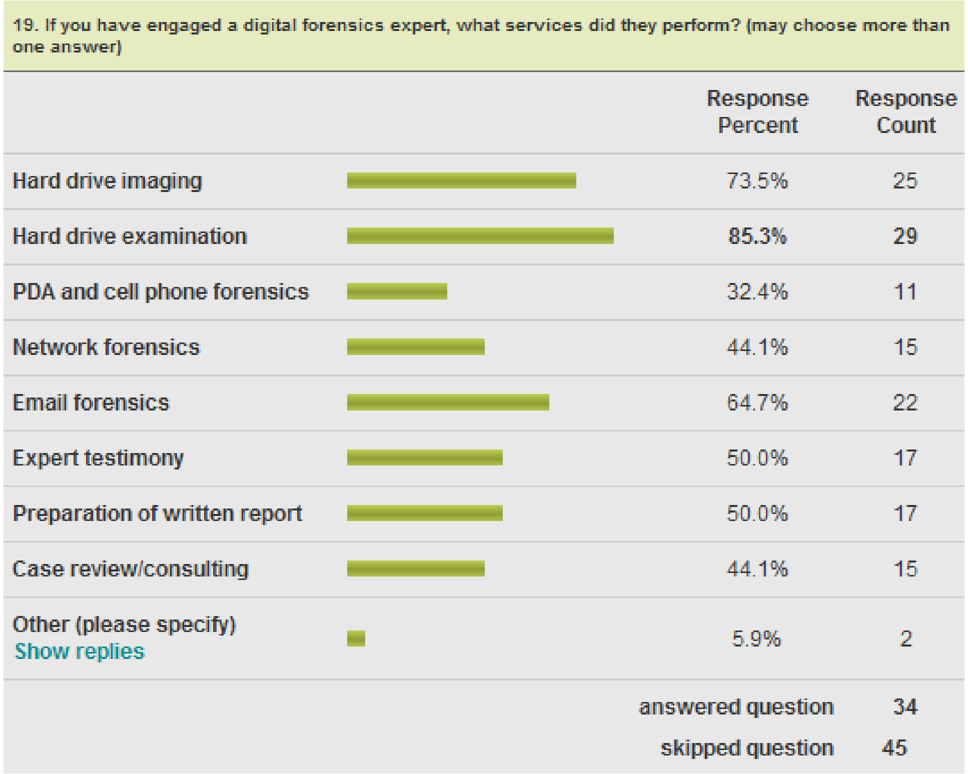
Expert witnesses who have earned graduate degrees in digital forensics or computer science were favored over the other eight professions and were considered persuasive in 25.0% of the total responses. Private investigators were found to be the least persuasive of all professions with only 3.4% of responses indicating that they were persuasive. This was less than one seventh of the preference rate for expert witnesses with graduate degrees in digital forensics or computer science.

The next to last category, EXPERT TESTIMONY (Questions 18 through 21) indicated no reluctance on the part of attorneys to hire digital forensics experts, but did reflect fundamental issues with respect to communications, usefulness of information, understandability of testimony, and comprehension of digital evidence in the courtroom. Particularly strong reactions were registered by a few of the respondents who had apparently had bad experiences with "computer experts" who were felt to have created distressing results during discovery. With respect to using the services of digital forensics experts, attorneys responding were more inclined to use them for "traditional services," such as hard drive imaging and examination (35.3%), rather than expert testimony (11.1%).



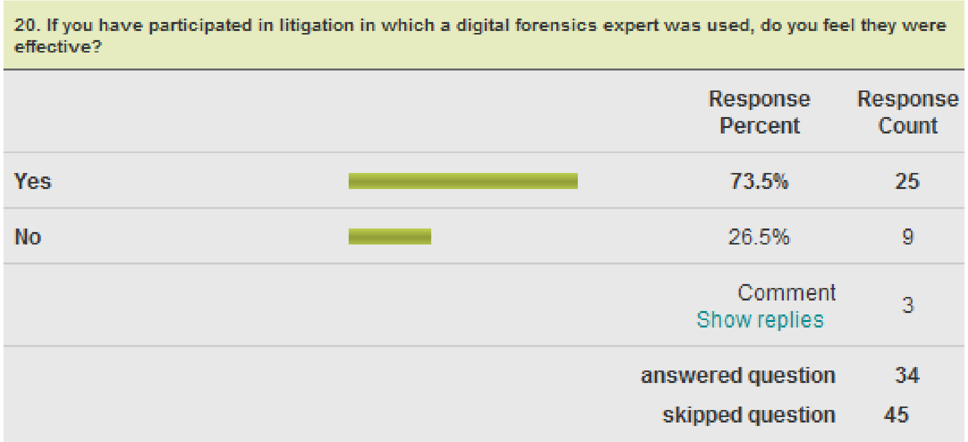
18. Have you ever engaged the services of a digital forensics expert?

	Response Percent	Response Count
Yes	49.3%	34
No	50.7%	35
	answered question	69
	skipped question	10



19. If you have engaged a digital forensics expert, what services did they perform? (may choose more than one answer)

	Response Percent	Response Count
Hard drive imaging	73.5%	25
Hard drive examination	85.3%	29
PDA and cell phone forensics	32.4%	11
Network forensics	44.1%	15
Email forensics	64.7%	22
Expert testimony	50.0%	17
Preparation of written report	50.0%	17
Case review/consulting	44.1%	15
Other (please specify) Show replies	5.9%	2
	answered question	34
	skipped question	45

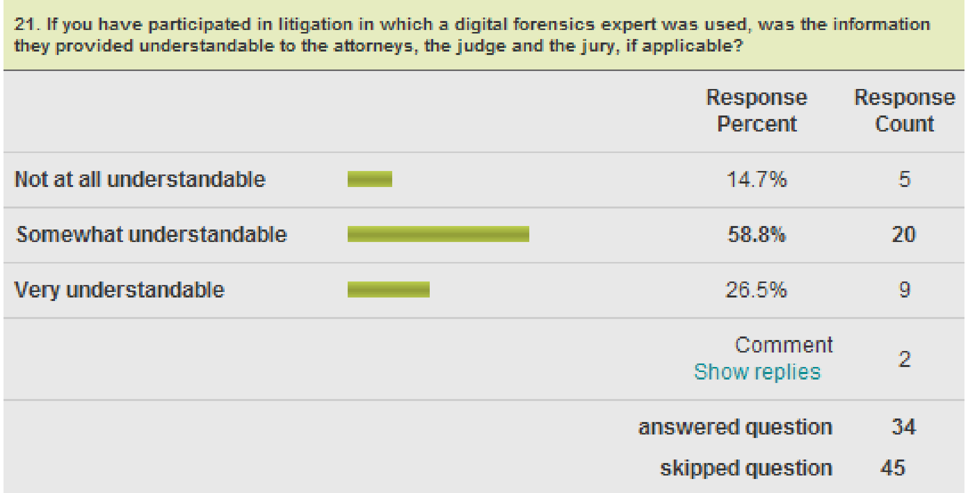


20. If you have participated in litigation in which a digital forensics expert was used, do you feel they were effective?

	Response Percent	Response Count
Yes	73.5%	25
No	26.5%	9
	Comment Show replies	3
	answered question	34
	skipped question	45

Responses to Question 20 –

1. "Opposing party hired a "computer expert" who probably fouled up the evidence; subsequently they decided "not" to use the expert."
2. "Poor communications skills."
3. "Helpful in getting our information searched and transmitted properly; not intended for testimony; only used to get information produced."



21. If you have participated in litigation in which a digital forensics expert was used, was the information they provided understandable to the attorneys, the judge and the jury, if applicable?

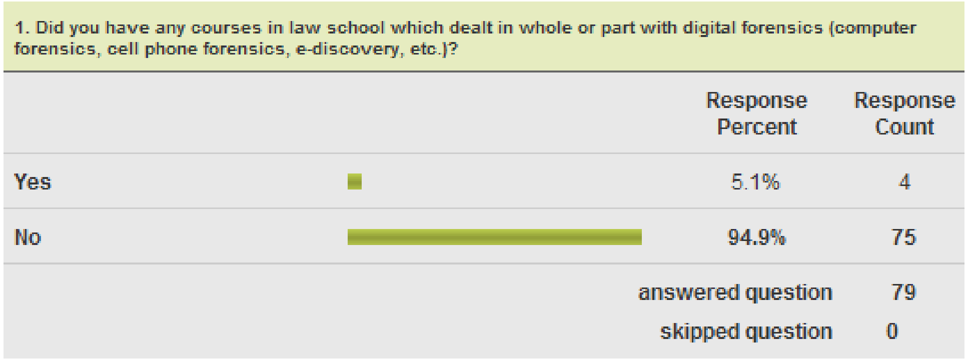
	Response Percent	Response Count
Not at all understandable	14.7%	5
Somewhat understandable	58.8%	20
Very understandable	26.5%	9
	Comment Show replies	2
	answered question	34
	skipped question	45

Responses to Question 21 –

1. "Never came to that; opposing party's expert rendered such evidence unusable."
2. "The answers above do not cover everything. Yes, some of what they had to say – much of it – was incomprehensible, but some was understandable."

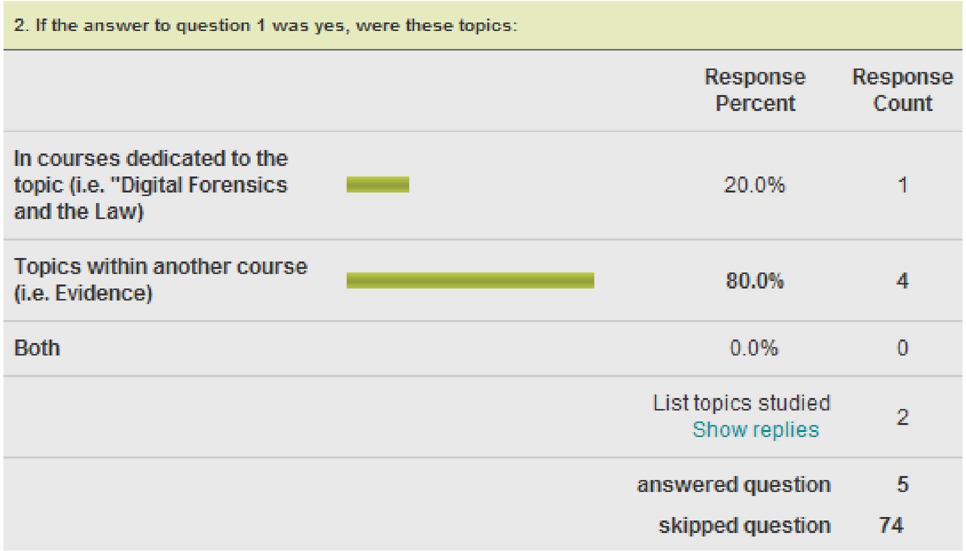
When responses to these questions are analyzed as a whole, there appears to be little reluctance to use digital forensics expert witnesses to isolate, identify, and report on digital evidence; but significant communications issues exist between counsel and experts, which further exposes the gulf between the training and education of attorneys, and the background of commonly accepted expert witnesses in digital forensics.

As explained earlier in this paper the response rate of answers to (Questions 1 thru 3), EDUCATION, represented a significant statistical variance from the expected mean. This was attributable to all survey participants answering Question 1, and only five participants answering Questions 2 and 3.



1. Did you have any courses in law school which dealt in whole or part with digital forensics (computer forensics, cell phone forensics, e-discovery, etc.)?

	Response Percent	Response Count
Yes	5.1%	4
No	94.9%	75
answered question		79
skipped question		0



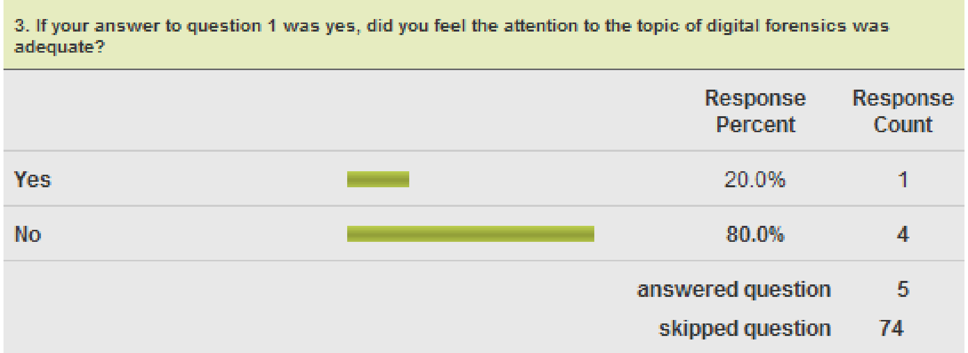
2. If the answer to question 1 was yes, were these topics:

	Response Percent	Response Count
In courses dedicated to the topic (i.e. "Digital Forensics and the Law)	20.0%	1
Topics within another course (i.e. Evidence)	80.0%	4
Both	0.0%	0
List topics studied Show replies		2
answered question		5
skipped question		74

Responses to Question 2 –

1. "Evidence, Criminal Procedure, Civil Procedure."
2. "How to use AccessData. Imaging using old school technology. Maintaining a chain of custody. Creating reports. Working with all OS."

The final question in this series addressed the adequacy of digital forensics education provided in law school to the five participants that responded. All but one of the attorneys who answered this question considered that education to be inadequate. The one attorney, out of seventy-nine, that initially responded to this survey represented 1.3% of the total. This is an ominous warning when consideration is given to an exploding digital age where Moore's Law predicts a continuation of exponential growth in computer and digital device capabilities.



3. If your answer to question 1 was yes, did you feel the attention to the topic of digital forensics was adequate?

		Response Percent	Response Count
Yes		20.0%	1
No		80.0%	4
		answered question	5
		skipped question	74

4. CONCLUSION

"New technologies create interesting challenges to long established legal concepts." (United States v. Maxwell, 45 M.J., 1996 p. 410).

Law schools have not caught up to the digital age. According to Gabel, the bar must be raised in educating young lawyers (Gabel, 2010). In his blog, "What do you call someone who gets the lowest passing grade on the Bar exam?" (EDD Update, 2010), Craig Ball, a noted Austin, Texas attorney and digital forensics expert, relates a conversation he had with a third-year law student at the University of Texas in Austin following a lecture he gave in an e-discovery class. The student balked at having to learn about digital forensics. Ball reminded the student that the penalty of not knowing, and being accused of gross negligence was severe. In response, the student asked, "What's the least I need to know?" (Ball, 2010)

Taking this as a whole, what is to be done? First and foremost, a system of continuing education, more extensive than is currently obtainable, should be made available to judges, prosecutors and practicing attorneys. Programs such as the Cybercrime Initiative at the National Center for Justice and the Rule of Law at the University of Mississippi School of Law, provide two to four day seminars to judges and prosecutors only, mostly in the area of child pornography. In fact, most programs offered are only for judges and/or prosecutors, the thought being that such knowledge should not be given to the "dark side." This sets a dangerous, and unethical, precedent as it steps on the Sixth Amendment rights of a defendant.

Law schools must step up to the plate and take responsibility. Course curriculums must be increased to include more than e-discovery. Digital forensics procedures and analysis should be taught as a part of evidence courses. As an example, currently the University of Memphis uses a multi-discipline method, combining the resources of the law school, the business school, and the colleges of engineering, criminal justice and computer science to form the Center for Information Assurance, which also spearheads the efforts of The U of M as a Center of Excellence in Information Assurance Education. Perhaps this should be used as a model for other universities which have law schools or affiliations with law schools.

REFERENCES

Ball, C., (2010), 'What do you call someone who gets the lowest passing grade on the Bar exam?" *EDD Update [online]* http://www.eddupdate.com/2010/02/what-do-you- call-someone- who-gets-the-lowest-passing-grade-on-the-bar-exam.html, October 20, 2010.

Daubert v. Merrill Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Frye v. United States, 293 F. 1013 (DC Cir. 1923).

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Gabel, J. D. (2010), "Forensiphilia: Is Public Fascination with Forensic Science a Love Affair or Fatal Attraction?" New England Journal on Criminal and Civil Confinement, Summer: 233.

Greenbaum, M. (2010), "No more room at the bench," Los Angeles Times, http://articles.latimes.com/2010/jan/08/opinion/la-oe-greenbaum8-2010jan08, January 8, 2010.

Hayes, D. (2002), " KC to join high-tech fight against high-tech crimes: FBI to open $2 million center here," Kansas City Star, page A1, April 26, 2002.

Johns, A. (2009), " Computer Science and the Reference Manual for Scientific Evidence: Defining the Judge's Role as a Firewall," Intellectual Property Law Bulletin, 14: 23.

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Longnecker, M., & Ott, R. L. (2010), An Introduction to Statistical Methods and Data Analysis (6th ed.), Brooks/Cole, Belmont, CA

UNDERSTANDING ISSUES IN CLOUD FORENSICS: TWO HYPOTHETICAL CASE STUDIES

Josiah Dykstra and Alan T. Sherman
Cyber Defense Lab, Department of CSEE
University of Maryland, Baltimore County (UMBC)
1000 Hilltop Circle, Baltimore, MD 21250
{dykstra, sherman}@umbc.edu

ABSTRACT

The inevitable vulnerabilities and criminal targeting of cloud environments demand an understanding of how digital forensic investigations of the cloud can be accomplished. We present two hypothetical case studies of cloud crimes; child pornography being hosted in the cloud, and a compromised cloud-based website. Our cases highlight shortcomings of current forensic practices and laws. We describe significant challenges with cloud forensics, including forensic acquisition, evidence preservation and chain of custody, and open problems for continued research.

Keywords: Cloud computing, cloud forensics, digital forensics, case studies

1. INTRODUCTION

Crime committed using cloud computing resources and against cloud infrastructures is inevitable. Though real incidents have already taken place against cloud providers including Google, an absence of documentation indicates that no crimes using the cloud or targeting it directly have been publicized nor litigated thus far. Forensic investigators must understand that current tools and techniques are inadequate in the cloud environment where acquisition, examination and analysis will be in practice executed very differently than is done today. To illustrate these issues, we fabricate two hypothetical crimes and deconstruct the forensic investigation against them.

Companies are embracing cloud technology to offload some of the cost, upkeep, and growth of equipment that they would otherwise have purchased themselves. Cloud infrastructure, with exceptional bandwidth, storage and computing power, offers an attractive prize for hackers. While many people have lamented how the users of the cloud and their data are protected, few of these discussions have considered the difficulty of responding to security breaches, including forensics and criminal prosecution.

In this article, we consider the investigative response and forensic process of two hypothetical, but plausible, case studies of crimes tied to cloud computing. In Section 2, we present previous and related work. In Section 3 we discuss the applicability of forensic frameworks. Section 4 contains our case studies. The first explores a case of child pornography in the cloud, and the trouble with both acquiring and analyzing data. The second case study deals with the cloud as the target of a crime, and the complex issues of chain of custody and trust. We examine issues of attribution, forensic integrity and chain of custody in Section 5, and we conclude in Section 6.

2. PREVIOUS WORK

Despite significant research in digital forensics, little has been written about the applicability of forensics to cloud computing environments. Furthermore, no case law exists on which to extrapolate the desire of the courts on the matter. Garfinkel recently suggested that "cloud computing in particular may make it impossible to perform basic forensic steps of data preservation and isolation on systems of forensic interest" (Garfinkel 2010). In one of the only published books on cloud forensics, the subject is approached as a matter of network forensics combined with remote disk forensics (Lillard 2010). While legal complications are introduced, including cloud-based evidence admissibility, no solutions are presented. Wolthusen identified some research challenges, including “discovery of computation structure,” “attribution of data,” “stability of evidence,” and “presentation and visualization of evidence” (Wolthusen 2009). In 2009, researchers at UC San Diego demonstrated that it was possible to locate a particular virtual machine (VM) in Amazon Elastic Compute Cloud (EC2) and mount side-channel attacks by co-locating a new VM with the target (Ristenpart 2009).

In 2009, Google and 34 other companies were hacked and infected with data-stealing malware. While the attack at Google involved Gmail, a cloud-based email service, the vulnerabilities and exploits were end-user based and not an attack on the cloud (Symantec 2010). Using Amazon EC2, researchers recently demonstrated how to crack passwords quickly and cheaply, a potentially criminal activity (Bagh 2011). In 2010, presenters at the DEFCON Conference used EC2 to launch a demonstration denial of service against a small network (Lemos 2010). In the investigation of individual users, cloud providers have begun to offer services that aid law enforcement. For example, Facebook gives a user the option to download their entire personal profile and history (Facebook 2011). However promising this may be for an investigator, these data cannot be said to be forensically sound. Guidance Software, the maker of EnCase, has produced a training video showing how to recover and analyze Facebook chat artifacts from a local hard drive (Guidance 2009).

Lawyers and computer scientists alike have expressed views about remote forensics, a field that shares an important similarity to the cloud. Schwerha and Inch (Schwerha and Inch 2008) list remote forensic software and survey legal analysis and case law. They undertook no application to cloud computing. Law professor Orin Kerr has written extensively on the applicability of the Fourth Amendment to electronic evidence and the Internet (Kerr 2009). His suggestions on search warrant language for shared resources are apropos to cloud forensic research. In Australia, lawmakers are already being made aware that current law enforcement is not equipped to investigate attacks on cloud services (Choo 2009).

3. FRAMEWORKS

To frame the approach of forensic investigation of any environment, including the cloud, it is helpful to have a procedure that guides the activity. The cloud environment does not affect the need for a framework, and does not inherently demand a new one. Frameworks for the digital forensic investigation are plentiful: at least 14 have been published since 1995 (Selamat *et al*. 2008). Digital forensic labs often choose a combination of approaches, or develop their own process that considers their particular personnel, workload, and budget. The generality of many investigative frameworks makes them applicable under many circumstances and irrespective of technology. While there is hardly a generic computer forensic case that would lend itself to routine and standardized steps, in practice the general forensic process for a particular type of crime tends to look similar each time. For example, the examination of digital artifacts to find evidence of child pornography almost always involves taking a bit-for-bit hard drive image and searching common file system locations and slack space for contraband images.

Consider the “Guide to Integrating Forensic Technique into Incident Response” published by NIST (Kent *et al*. 2006). The NIST process, like many others, can be roughly summarized as follows:

- Collection
- Examination
- Analysis
- Reporting.

Collection involves the process of physical acquisition of data. Examination is the process of combing through the data for items of interest. Analysis is the application of the interesting items to the investigative question at hand, and whether it supports or refutes that question. Reporting describes the output of analysis, including the analysis steps taken.

4. CASE STUDIES

We have developed two hypothetical case studies to reason about the state of digital forensics for cloud-related crimes. While fictional, they describe computer crimes that are not uncommon today. Case Study 1 uses the cloud as an accessory to a crime. Case Study 2 targets the crime against the cloud. These crimes require a reinterpretation when set in a cloud computing environment. In both scenarios, the following themes emerge that differentiate these investigations from traditional digital forensics:

- Acquisition of forensic data is more difficult.
- Cooperation from cloud providers is paramount.
- Current forensic tools appear unsuited to process cloud data.
- Cloud data may lack key forensic metadata.
- Chain of custody is more complex.

We will return to address these issues in more detail in Section 5.

4.1 Case Study 1

Polly is a criminal who traffics in child pornography. He has set up a service in the cloud to store a large collection of contraband images and video. The website allows users to upload and download this content anonymously. He pays for his cloud services with a pre-paid credit card purchased with cash. Polly encrypts his data in cloud storage, and he reverts his virtual webserver to a clean state daily. Law enforcement is tipped off to the website and wishes both to terminate the service and prosecute the criminal.

This is a case where the computer is incidental to the offense. Let us assume that the cloud model used in this case is Infrastructure as a Service, such as Amazon EC2. In this service model, the provider has responsibility and access to only the physical hardware, storage, servers and network components. In the public interest, law enforcement first contacts the cloud provider with a temporary restraining order to suspend the offending service and account, and a preservation letter to preserve evidence pending a warrant.[1] Tracking down the user is the more difficult task. The onus in this case is on the forensic examiner to piece together a circumstantial case based on the data available.

The examiner has no way to image the virtual machine remotely since the cloud provider does not expose that functionality, and in doing so would alter the state of the machine anyway. Deploying a remote forensic agent, such as EnCase Enterprise, would require the suspect's credentials, and functionality of this remote technique within the cloud is unknown. Today the forensic examiner, with no case law or standard methodology on the matter, may be tempted to attempt standard practices in digital evidence collection. Namely, with proper recording and documentation, the examiner accesses the offending website and takes snapshots or videotaping the collection of the evidence, and saving the web pages locally. Simply viewing the target website is enough to confirm that the content is illegal, but it tells us nothing about who put it there. Additionally, no guarantee can yet be made that the target webserver has not been compromised by an attacker, or that the examiner's request to the web server was not the victim of DNS poisoning, man-in-the-middle, or some other alteration in transit.

Consider other possible sources of digital evidence in this case: credit card payment information, cloud subscriber information, cloud provider access logs, cloud provider NetFlow logs, the web server virtual machine, and cloud storage data. Law enforcement can issue a search warrant to the cloud provider, which is adequate to compel the provider to provide any of this information that they

[1] 18 U.S.C. §2703(f)(1) ("A provider of wire or electronic communication services or a remote computing service, upon the request of a governmental entity, shall take all necessary steps to preserve records and other evidence in its possession pending the issuance of a court order or other process.")

possess. Law enforcement need not execute or witness the search.[2] The warrant specifies that the data returned be an "exact duplicate," the forensic term that has historically meant a bit-for-bit duplication of a drive. Since child pornography is a federal offense, the provider must comply with the order. A technician at the provider executes the search order from his or her workstation, copying data from the provider's infrastructure and verifying data integrity with hashes of the files. Files may have been distributed across many physical machines, but they are reassembled automatically as the technician accesses them. Though the prosecution may call the technician to testify, we have no implicit guarantees of trust in the technician to collect the complete data, in the cloud infrastructure to produce the true data, nor in the technician's computer or tools used to collect the information correctly. Nonetheless, the provider completes the request, and delivers the data to law enforcement.

Let us say that Polly had two terabytes of stored data.[3] To transfer that quantity of data, the provider saves it to an external hard drive and delivers it to law enforcement by mail. In addition, the provider is able to produce: account information, 10MB of access logs, 100MB of NetFlow records, and a 20GB virtual machine snapshot. After validating the integrity of the data, the forensic examiner is now charged with analysis.

We would expect the forensic expert to identify the following that would aid in prosecution:

- Understand how the web service works, especially how it encrypts/decrypts data from storage
- Find keys to decrypt storage data, and use them to decrypt the data
- Confirm the presence of child pornography
- Analyze logs to identify possible IP addresses of the criminal.

It is not unreasonable to expect that this activity may take many man hours to analyze. According to performance testing from the manufacturer, AccessData found that their Forensic Toolkit (FTK) product took 5.5 hours to process a 120GB hard drive fully on a top-of-the-line workstation, and as long as 38.25 hours on a low-end workstation (AccessData 2010). At that rate, 2TB of data could take 85 hours of processing time. The examiner is likely to dive in first to the data store. The provider may have returned individual files or large files containing "blobs" of binary data. In either case, it will become quickly evident that the data are encrypted. Tools like EnCase and Forensic Toolkit can analyze VMware data files but not snapshots which include suspended memory. The human analyst will have to fix-up and run the VM snapshot in order to understand the website source and observe how encryption is used. Once the keys are uncovered, and data are decrypted, 2TB of data must be analyzed for evidence. We were already aware of illegal content, but not aware of the data owner. Timestamps or file metadata may prove useful, provided they are available and accurate. Evidence of the owner may be gleaned from NetFlow, timestamp, and potentially in the coding style of the website. We can safely assume that an IP can be found that points to Polly. All of the forensic analysis is documented and presented to counsel.

In the absence of legal precedent, existing case law must be considered in the forensic process used. In 2007, the 100-page opinion by Judge Grimm in Lorraine v. Markel issued guidance about the admissibility of original or duplicates of original evidence, as legislated in Rules 1001-1008 of the Federal Rules of Evidence (Lorraine 2007). As mentioned above, service providers are already empowered to conduct searches on behalf of law enforcement. Several important issues regarding the issuance of a warrant were omitted above.

[2] 18 U.S.C. §2703(g)("... the presence of an officer shall not be required for service or execution of a search warrant issued in accordance with this chapter requiring disclosure by a provider of electronic communications service or remote computing service of the contents of communications or records or other information pertaining to a subscriber to or customer of such service.")

[3] Interestingly, 18 U.S.C. §2703(b) allows a cloud provider to disclose the contents of an account used for remote storage without a warrant, and without notifying the customer or subscriber. Kerr suggested that this is unconstitutional (Kerr 2009).

- Search warrants must specify the search of a person or location for evidence of a crime. With cloud computing, a problem emerges because the data may not be location-specific, other than a known public-facing URL or the cloud provider hosting the data. A search warrant must describe the physical place to be searched with particularity.[4] This becomes further complicated if cloud resources are distributed across state or international boundaries.
- The Fourth Amendment presents a preposterous assumption about search preceding seizure,[5] which the courts may be compelled to reinterpret. As Kerr has explored extensively, traditional digital evidence collection is the reverse process of seizure then search (Kerr 2005). Further, digital evidence, and especially cloud evidence, is never "seized" in the sense that it ceases to exist in one place, but the data are the target of the seizure, which are copied and the original remains.

Given the procedure undertaken above, consider the issues which the defense may raise to introduce doubt in the examination:

- Since raw bit-for-bit copies of hard drives were not provided, how do we know that the cloud provider provided a complete and authentic forensic copy of the data? Can the authenticity and integrity of the data be trusted? Can the cloud technician, his/her workstation and tools be verifiably trusted?
- Were the data located on one drive, or distributed over many? Where were the drives containing the data physically located? Who had access to the data, and how was access control enforced? Were the data co-mingled with other users' data?
- If data came from multiple systems, are the timestamps of these systems internally consistent? Can the date and time stamps be trusted, and compared with confidence?
- Does the virtual machine have a static IP address? How can the prosecution tie the malicious activity on the virtual machine to Polly?
- What jurisdiction governs the data in question? If the cloud provider's jurisdiction, then which of their geographic locations or datacenters?

Some of the digital evidence collection from the cloud mirrors traditional collection. In other respects the process is new, such as data dispersed over many storage systems and virtual machine use. Current tools are ill-equipped to process the data in this case easily. The case in almost every respect hinges on how the cloud provider cooperated. Without greater transparency into how the provider operates, it is difficult or impossible to counter the above objections from the defense.

Finally, we note that cloud providers have a legal obligation to purge child pornography from their systems. Many providers keep duplicate copies of stored data, which here requires that they know where all copies are located and how to verifiably delete the contraband. Microsoft and Amazon declined to comment about their compliance abilities in this situation.

4.2 Case Study 2

Mallory is a hacker who intends to exploit victims by placing a malicious webpage in the cloud. She uses a vulnerability to exploit the cloud presence of Buzz Coffee, a legitimate company. From there, she installs a rootkit that injects a malicious payload into web pages displayed, and hides her malicious activity from the operating system. She then redirects victims to the website, which infects

[4] Search warrants for online webmail have traditionally specified only the email address as the "place to be searched." See the search warrant for a Gmail mail account at http://docs.justia.com/cases/federal/district-courts/michigan/miedce/2:2009mc50275/237762/2/

[5] "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

them with malware. Users complain to the legitimate company that they are being infected, so the company seeks to fix the problem and investigate the crime.

This example is a different type of computer crime, one where the target is the computer. Let us assume that Buzz Coffee uses a Software as a Service provider, such as RackSpace. In this service model, the provider has responsibility and access to the hardware, the operating system, and the hosting platform. Buzz wishes to make an example of this hacker, and hires a lawyer to prosecute the attacker. The attorney contracts a forensic specialist to conduct the digital investigation. Using experience as a guide, the investigator constructs a plan to access the cloud provider remotely over a secure channel using Buzz Coffee's credentials and retrieve the website source files. However, when the data are returned, nothing malicious is found since Mallory's rootkit hid the files from the host operating system and the provider's APIs. The forensic investigator determines that the following are additional possible sources of data: cloud provider access logs, cloud provider NetFlow logs, and the web server virtual machine.

The prosecutor approaches the cloud provider with a subpoena and requests all of this data, including a forensic copy of the virtual machine.[6] The provider is willing to conduct an internal investigation; however, it is reluctant to produce the raw data citing confidential and proprietary information. In fact, the Service Level Agreement lacks any language requiring compliance with intrusion response or remediation. The attorney is able to convince a judge that there is likely evidence of a crime inside the cloud, and a search warrant is issued to the provider.[7] Even in this case, the provider complies to the extent that its legal counsel feels is appropriate, which in this case includes: NetFlow logs, web access logs, and files from the virtual machine that comprise Buzz Coffee's website. Any further data from the operating system or hosting platform, they claim, would threaten their business and competitive advantage.

A technician at the provider executes the court order from his workstation, copying data from the provider's infrastructure and verifying integrity with MD5 hashes. This information is burned to DVD, and contains 2 MB of NetFlow logs, 100 MB of web access logs and 1 MB of web source code. Using this information, we wish our investigator to uncover the following:

- A chronology that shows when the web pages have been viewed and modified/accessed/created
- Determine the malicious webpage and how the system was compromised
- Analyze the scope of the intrusion, and possible spread to other systems
- Identify the origin of the malicious activity.

Comparing the original website files created by Buzz Coffee to the data returned from the cloud provider would be a constructive first step. Here the technique employed during collection becomes paramount. If the host operating system was used to retrieve the files, Mallory's rootkit would have hidden the malicious files. If files were acquired by reading the physical disk, bypassing the operating system, the complete collection of files will be accurate. Constructing a timeline is a common practice for forensic examiners, and one important in determining when Mallory's files were created. Unfortunately, the procedure employed by the provider again determines whether the investigator receives useful metadata, such as file creation timestamps.

Web access logs are likely the most definitive evidence of the original intrusion, corroborated by NetFlow records. The suspected attacker IP is identified in the logs, which is presented alongside the

[6] Unlike warrants, subpoenas do not require probable cause and can be issued by prosecutors without judicial approval, as long as they are not unreasonably burdensome. See William J. Stuntz, Commentary, *O.J. Simpson, Bill Clinton, and the Transsubstantive Fourth Amendment*, 114 HARV. L. REV. 842, 857-58 (2001).

[7] See examples in NIJ's *Investigations Involving the Internet and Computer Networks*, http://www.ncjrs.gov/pdffiles1/nij/210798.pdf

complete analysis in the subsequent forensic report. Prudent readers might also approach this problem by analyzing the malware installed after visiting the now-hacked webpage, and trying to determine who wrote it or to where it beacons back, but that is not considered here.

Taken to court, the following are questions that could be raised by the defense to discredit the forensic process used in this case:

- Was the chain of custody preserved throughout the process?
- Can the malicious page be definitively attributed to Mallory? Who else had access to create/modify this page? Were other clients hosted on the same infrastructure that could have had access?
- What process did the cloud provider use to copy and produce the webpages? Can they make any claims about the forensic integrity of this process? Are timestamps across the different evidence (NetFlow, web logs, etc.) synchronized enough to create an accurate timeline?
- What was the physical location of the virtual machine that is run by the hosting website? By what laws/regulations is it governed?
- What detection and protection mechanisms are employed by the provider to keep their infrastructure secure and to identify intrusions?
- Since the provider refused to provide operating system evidence, can the prosecution have enough evidence to prove that a compromise actually occurred?

In this case the closed nature of the provider was the primary hindrance to a routine investigation. The provider has an incentive to keep as much of its infrastructure private as possible, since it may give them a competitive advantage. Unfortunately, this decision hinders the investigative process and may discredit the legal proceedings that follow.

5. ANALYSIS

Whether in the cloud or not, forensic investigation can be an intensive process. Exams are almost always limited by time and budget, since clients are unwilling or unable to support them indefinitely. Cloud computing, for better or worse, gives customers an ability to terminate virtual machines or revert them to a saved state almost instantaneously. Providers and investigators may also benefit from easy data duplication, system copying/imaging, and extensive business logging. Investigators must recognize the extreme fragility of the evidence. These attributes are indeed positive and contribute towards well-rounded security preparation for incident response. The hindrances seen in these case studies illustrate areas for continued research and development. Consider how we might address the five issues presented at the beginning of Section 4.

First, in our case studies, acquisition was accomplished using legal vehicles of subpoena and search warrant. While somewhat cumbersome given the complex legal system, if a forensic investigation is to support a potential criminal proceeding, this approach is necessary. More efficient mechanisms for the secure transfer of data from providers and law enforcement would be ideal.

Second, cloud consumers will need to negotiate or lobby providers for an appropriate level of cooperation and transparency about how their infrastructure works, the amount of support available during incident response, and forensically-sound practices for assisting law enforcement. One potential approach is a forensic service level agreement (SLA) appended to the existing SLA signed by providers and subscribers. This legal backing would give customers assurance about the support available to them from their provider during an investigation, a quantitative measure by which to compare providers.

Third, it is clear that remote forensic tools applied to cloud computing are prone to scrutiny, and local processing tools of cloud-stored data are not designed to handle the format or scope of the data. In the case of Infrastructure as a Service, analysis will certainly include the investigation of a virtual machine. Forensic analysts need a tool for parsing, searching and extracting information from virtual machine snapshots, including suspended memory state.

Fourth, the lack of forensic metadata may be addressed in several ways. One proposal is to introduce data provenance in order to track the history and access of cloud objects. In 2007, a report from the Department of Justice recommended asking "what is the chronology of the access to or changes in the data?" of persons providing digital evidence (National Institute of Justice 2007). Another proposal is to introduce preemptive forensics in the cloud, the forensically-sound logging of information at all times without evidence of a crime in order to specifically support forensic investigations after a crime takes place. For example, keeping regular virtual machine snapshots would create a forensic record back in time once an event arises. This computer-generated evidence may benefits from being protected against hearsay arguments, a viewpoint now recognized by some courts.

Finally, chain of custody remains complex given the number of people that may have access to the evidence, and the third-party collection as discussed above. In traditional digital forensics, a chain of custody exists for both physical evidence (*e.g.* the computer) and its associated data. In the cloud case, data are the only evidence. As such, pristine copies of the data, and associated integrity information like MD5 checksums, must be carefully handled. Since chain of custody is the legal equivalent of secure provenance, transfers of custodianship could be documented by a digital provenance system.

Note that we have not addressed the issue of responsibility and fault in either case study. In Case Study 1, we have not established what liability the cloud provider has for hosting the illegal content. In all likelihood, the cloud provider demonstrated no negligence, and is simply a data custodian unaware of the activity. Nonetheless, the law demands they identify and remove all illegal content. In Case Study 2, can users who were infected sue the legitimate company or the cloud provider for negligence? Could Buzz coffee sue the hosting provider if they failed to secure their infrastructure, or to notice the intrusion? These questions may be answerable using an interpretation of current laws. Additionally, we have not explored the investigative complexity of cloud service resellers who themselves offer services that utilize cloud technology. The layering of providers may further complicate the preservation and acquisition of evidence.

Finally, both case studies assume trust in the provider, its employees and infrastructure. Providers have their business reputation and customer base to lose if trust is lost in their ability to provide secure and reliable service. However, if an adversary or corrupt insider gains control over the cloud infrastructure—particularly the hypervisor—no data or computational results in the hosted virtual machines can be trusted.

6. CONCLUSIONS

Cloud security is a much discussed topic, but planning about incident response and forensics needs to happen in parallel. The move of data and services to the cloud is already underway, and research and development in the forensic research community must keep pace. These two case studies illustrate larger issues that exist beyond the scope of our specific examples. Forensic acquisition is a renewed challenge, one unsuited for today's tools, which will possibly be addressed by a combination of technological and legal approaches. We have begun to evaluate the ability of popular forensic tools to obtain evidence from a cloud environment. Cooperation with providers will empower consumers to understand their risks and give them leverage to prosecute crimes. The preservation and availability of forensically-relevant metadata remains an open problem.

We have highlighted the issues of common crimes that vary from today only in their use of the cloud. This technology alone introduces peculiarities and open problems that demand immediate attention. As we have shown, deficiencies in both law and technology can be addressed with proper advances.

7. ACKNOWLEDGMENTS

We thank Simson Garfinkel and George Steele for helpful comments on early drafts. Sherman is supported in part by the Department of Defense under IASP Grant H98230-10-1-0359.

8. AUTHOR BIOGRAPHIES

Josiah Dykstra received the B.A. degree in computer science from Hope College in 2002 and the M.S. degree in information assurance from Iowa State University in 2004. He is pursuing the Ph.D. degree in computer science at the University of Maryland, Baltimore County. He is a network analyst at the U.S. Department of Defense. His research interests include computer security, intrusion detection, malware analysis, digital forensics, and cloud computing. Dykstra is a member of ACM and IEEE Computer Society.

Alan T. Sherman earned the PhD degree in computer science at MIT studying under Ronald L. Rivest, the SM degree in electrical engineering and computer science from MIT, and the ScB degree in mathematics, magna cum laude, from Brown University. He is an associate professor of computer science at the University of Maryland, Baltimore County (UMBC) in the CSEE Dept. and Director of UMBC's Center for Information Security and Assurance. His main research interest is high-security voting systems. Sherman has carried out research in election systems, algorithm design, cryptanalysis, theoretical foundations for cryptography, and applications of cryptography. http://www.csee.umbc.edu/~sherman

9. REFERENCES

AccessData (2010), "FTK Performance Testing," http://www.accessdata.com/downloads/media/FTK Performance Testing.pdf, accessed December 10, 2010.

Bagh, C. (2011), "Amazon EC2 helps researcher to crack Wi-Fi password in 20 minutes," http://www.ibtimes.com/articles/100314/20110112/amazon-ec2-password-wi-hacking-cracking-brute-force-attack-wpa-psk-encryption-cloud-computing-iaa.htm, accessed January 12, 2011.

Choo, K.-K. R. (2009), "Cloud computing: Challenges and future directions," Trends & Issues in Crime and Criminal Justice, no. 400. Canberra, ACT, Australia, October, 2009.

Facebook (2011), "Help Center: How can I download my information from Facebook?" http://www.facebook.com/help/?page=18830, accessed January 4, 2011.

Garfinkel, S. L. (2010), "Digital forensics research: The next 10 years," Proceedings of the Tenth Annual DFRWS Conference, August 2 – 4, 2010, Portland, OR.

Guidance Software (2009), "Facebook Chat Examinations," http://www.encaseondemand.com/EnCast/EnCastVideos/tabid/1383/ProductID/99/CategoryID/129/List/1/Level/1/Default.aspx, accessed January 6, 2011.

Kent, K., Chevalier, S., Grance, T. and Dang, H. (2006), "Guide to Integrating Forensic Techniques into Incident Response," http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf, accessed September 8, 2010.

Kerr, O. S. (2009), "Applying the Fourth Amendment to the Internet," Stanford Law Review, vol. 62, no. 4, 2009, pp. 1005-1049.

Kerr, O. S. (2005), "Search Warrants in an Era of Digital Evidence," Mississippi Law Journal, vol. 75, 2005, pp. 85-135.

Lemos, R. (2010), "Cloud-Based Denial Of Service Attacks Looming, Researchers Say," http://www.darkreading.com/smb-security/167901073/security/perimeter-security/226500300/index.html, accessed August 4, 2010.

Lillard, T. V. (2010), Digital Forensics for Network, Internet and Cloud Computing: A Forensic Evidence Guide for Moving Targets and Data, Syngress, Rockland, MA.

"Lorraine v. Markel American Insurance Company," 241 F.R.D 534 (D.Md. May 4, 2007).

National Institute of Justice (2007), "Digital Evidence in the Courtroom: A Guide for Law

Enforcement & Prosecutors," http://ncjrs.gov/pd_les1/nij/211314.pdf, accessed September 7, 2010.

Ristenpart, T., Tromer, E., Shacham, H. and Savage, S. (2009), "Hey, you, get off my cloud: Exploring information leakage in third-party compute clouds," Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09), New York, NY, pp. 199-212.

Schwerha, J. J. and Inch, S. (2008), "Remote Forensics May Bring the Next Sea Change in E-discovery:

Are All Networked Computers Now Readily Accessible Under the Revised Federal Rules of Civil

Procedure?" Journal of Digital Forensics, Security and Law, vol. 3, no. 3, 2008, pp. 5-28.

Selamat, S., Yusof, R. and Sahib, S. (2008), "Mapping Process of Digital Forensic Investigation Framework," International Journal of Computer Science and Network Security, vol. 8, no. 10, 2008.

Symantec (2011), "The Trojan.Hydraq Incident: Analysis of the Aurora 0-Day Exploit," http://www.symantec.com/connect/blogs/trojanhydraq-incident-analysis-aurora-0-day-exploit, accessed January 21, 2011.

Wolthusen, S. D. (2009), "Overcast: Forensic Discovery in Cloud Environments," Proceedings of the 2009 Fifth International Conference on IT Security Incident Management and IT Forensics (IMF '09), pp. 3-9, September 15-17, 2009, Stuttgart, Germany.

SURVEY ON CLOUD FORENSICS AND CRITICAL CRITERIA FOR CLOUD FORENSIC CAPABILITY: A PRELIMINARY ANALYSIS

Keyun Ruan
University College Dublin
keyun.ruan@ucd.ie

Ibrahim Baggili (PhD)
Zayed University
ibrahim.baggili@zu.ac.ae

Prof Joe Carthy
University College Dublin
joe.carthy@ucd.ie

Prof Tahar Kechadi
University College Dublin
tahar.kechadi@ucd.ie

ABSTRACT

In this paper we present the current results and analysis of the survey “Cloud forensics and critical criteria for cloud forensic capability” carried out towards digital forensic experts and practitioners. This survey was created in order to gain a better understanding on some of the key questions of the new field - cloud forensics - before further research and development. We aim to understand concepts such as its definition, the most challenging issues, most valuable research directions, and the critical criteria for cloud forensic capability.

Keywords: Cloud Forensics, Cloud Computing, Digital Forensics, Survey, Cloud Forensic Capability

1. INTRODUCTION

Cloud computing has the potential to become one of the most transformative developments in the history of computing, following the footsteps of mainframes, minicomputers, PCs (Personal Computers), smart phones, and so on (Perry et al.,2009). It is radically changing how information technology services are created, delivered, accessed and managed.

Gartner estimates by 2015, 20% of non-IT Global 500 companies will be cloud service providers (Gartner, 2010). However, the rapid growth and adoption of cloud computing as a non-standard system (Beebe, 2009), is bringing digital forensics deeper into the crisis it is facing (Garfinkel, 2010). Encryption, proliferation of endpoints, multi-jurisdiction, loss of data control, to name a few, are all challenges exacerbated in cloud environments for forensic investigations due to a general lack of tools and expertise. Cloud organizations, including CSPs (Cloud Service Provider) and cloud customers, have to establish a cloud forensic capability, otherwise, they will face tremendous difficulties in carrying out investigations on critical incidents in a cloud architecture such as criminal intrusions and major policy violations in order to restore operations, data and services. They will also face difficulties when collaborating with law enforcement in cases of resource confiscation, etc., due to lack of forensic knowledge and preparation.

Ruan et al. (2011) first gave an overview of cloud forensics, introduced the cloud forensics three-dimensional model, and analyzed some of the major challenges and opportunities of cloud forensics. In order to validate the key areas covered in Ruan et al. (2011) and to study the critical criteria for cloud forensic capability, the researchers carried out this survey towards digital forensic experts and practitioners around the world on some key questions of cloud forensics, such as the definition of cloud forensics, the most significant challenges and opportunities of cloud forensics, the most valuable research direction for cloud forensics, etc. The survey was opened on 13th Feb 2011 and was widely circulated.

2. LIMITATIONS

Until 23rd Mar 2011, the survey has received 156 responses. The major limitation of the survey is the limited sample size. Only a limited number of experts (around 80) who responded to the survey have completed all the questions. According to the feedback, the reason for this can be the fact that cloud forensics is relatively a new topic. However, it is the first and only survey carried out towards the digital forensics community that is explicitly focused on cloud forensics, thus the researchers decided to share a preliminary analysis of the current survey results in this paper.

3. METHODOLOGY

In this research digital forensics experts and practitioners are surveyed on the definitions of cloud computing and cloud forensics, cloud forensics research and techniques, and critical criteria for cloud forensic capability.

The survey is hosted by Zayed University, United Arab Emirates (UAE). All participants are required to agree to a consent form, which contains key terms on the voluntary nature of participation and confidentiality of the survey results, before starting filling out the survey. Demographic data of participants is collected at the beginning of the survey.

The main body of the survey is divided into three sections:

- Part I Background
- Part II Cloud Forensics Research and Techniques
- Part III Critical Criteria for Forensic Capability

In "Part I Background", the researchers designed the following questions:

(1) What is cloud computing: as cloud computing is becoming mainstream, it still remains a confusing and evolving term in the industry. All the studies and research around cloud computing have to be based on a consensus on its definition. In this question, participants are presented with several definitions from respected organizations, such as NIST, Gartner, Oracle, Cloud Security Alliance (CSA) without the names of these organizations shown in the survey, as well as several popular views on the definition of cloud computing.

(2) Cloud computing as a trend: cloud computing has attracted massive investment and is seeing rapid adoption in both businesses and governments worldwide (INPUT, 2009). Gartner (2009A) forecasted that the worldwide cloud service market is expected to reach $150.1 billion in 2013. According Merrill Lynch (2008), the volume of the cloud computing market opportunity will amount to $160 billion by 2011. According to an October 2008 forecast by IDC (International Data Corporation)(Gens, 2008), spending on cloud services is growing at five times the rate of traditional on-premises IT. What is the underlying reason for cloud computing as a trend? Is it because of the top advantage of cloud computing, i.e., cost-effectiveness? (CSA, 2009), or it is a new phase of the evolution of computing since the 1960s

towards utility computing (Buyya et al., 2008)? By understanding better what is cloud computing as a trend, cloud forensics can be better placed in the big picture.

(3) What is cloud forensics: cloud forensics is a new area, a new way to call old techniques, or a mixture of both? By asking this question, the researchers aim to get opinions from the industry experts on the how to define cloud forensics.

(4) How significant is cloud forensics: is cloud forensics a component of cloud security, or an independent segment in parallel to cloud security with the same importance? This question is designed in order to understand the significance of cloud forensics in the cloud architecture.

(5) What is the impact of cloud computing on forensics: some say cloud computing makes forensics harder (Sawyer, 2009), while others say cloud computing makes forensics easier (Morrill, 2008). The researchers want to survey the digital forensic experts on their opinions on the impact of cloud computing on forensics: whether it is making forensic harder or easier, or both?

(6) What are the dimensions of cloud forensics: the emerging cloud computing, with its worldwide availability and resource sharing environments, has introduced much complexity into digital forensics, which is traditionally a technical discipline. The legal concerns have been further strengthened due to the default multi-jurisdiction setting. The organizational paradigm has become much more complex, when collaborations on all levels are needed among CSPs, cloud customers and law enforcement, compare to a single organization coping with its own on-premise networks, thus cloud forensics is a multi-dimensional discipline. Ruan et al. (2011) defined the three-dimensional model for cloud forensics, i.e., technical dimension, organizational dimension and legal dimension. In this question, the researchers aim to validate the three-dimension model from the opinions from the experts.

(7) What are the uses of cloud forensics: this question is important in order to attract more funding and investment on cloud forensic research and development. It is crucial to make both CSP and cloud customer understand the various uses of cloud forensics and how it can benefit their service availability and overall robustness of operations.

In "Part II Cloud Forensics Research and Techniques", the researchers designed the following questions:

(1) What are the challenges of cloud forensics: this question is valuable for understanding what are the most challenging issues regarding cloud forensics.

(2) What are the opportunities of cloud forensics: this question is valuable for understanding what are the biggest opportunities for cloud forensics.

(3) Valuable research directions of cloud forensics: this question is valuable for designing a research agenda for cloud forensics so that researchers and developers can focus on the most valuable research directions.

(4) What are the parties involved in a cloud investigation: this question is valuable for reaching a consensus on who should be involved in a cloud investigation.

In "Part III Critical Criteria for Forensics Capability", the researchers designed the following questions:

(1) Who should be assessed for the cloud forensic capability: this question is valuable for reaching a consensus on who should be assessed for cloud forensic capability.

(2) Importance of procedures and toolkits: in the technical dimension of cloud forensics (Ruan et al., 2011), a set of tools and procedures need to be developed to address the need for forensic investigations in the Cloud. From this question we can understand what are the most important tools and procedures and direct efforts to developing them.

(3) Staffing importance: this question is valuable for researching a consensus on the staffing structure in the organizational dimension (Ruan et al., 2011) of cloud forensics.

(4) Policy importance: this question is valuable for understanding what policies are more important than the others within cloud organizations to facilitate cloud forensic investigations.
(5) Agreement importance: this question is valuable for understanding what legal agreements are more important than the others among all parties involved in cloud forensic investigations.
(6) Guideline importance: this question is valuable for understanding what guidelines are most needed internally or externally for cloud organizations in the organizational dimension of cloud forensics.

4. RESULTS

4.1 Demographics

133 respondents answered the question of age. 36% of them are above 40, 34% between 31 and 40, 17% between 25 and 30, 5% between 19 and 24. 121 respondents answered the question of gender. 83% of them are male, 17% female. 126 respondents answered the question of education. 44% of them have obtained a master degree, 23% have obtained a PhD, and 28% have obtained a bachelor degree or a diploma. 124 respondents answered the question "years of experience in computer forensics field". 46% of them have more than 5 years experience in computer forensic field, 16% have 3 to 4 years experience, and 17% have 1-2 year experience. 127 respondents answered the question "how familiar are you with digital forensic tools". 76% of them are "very familiar" or "familiar" with digital forensic tools. According to the demographics results, the researchers believe that the respondents of the survey have good knowledge and sufficient experience in digital forensics.

4.2 Cloud Computing and Cloud Forensics

4.2.1 Definition of cloud computing

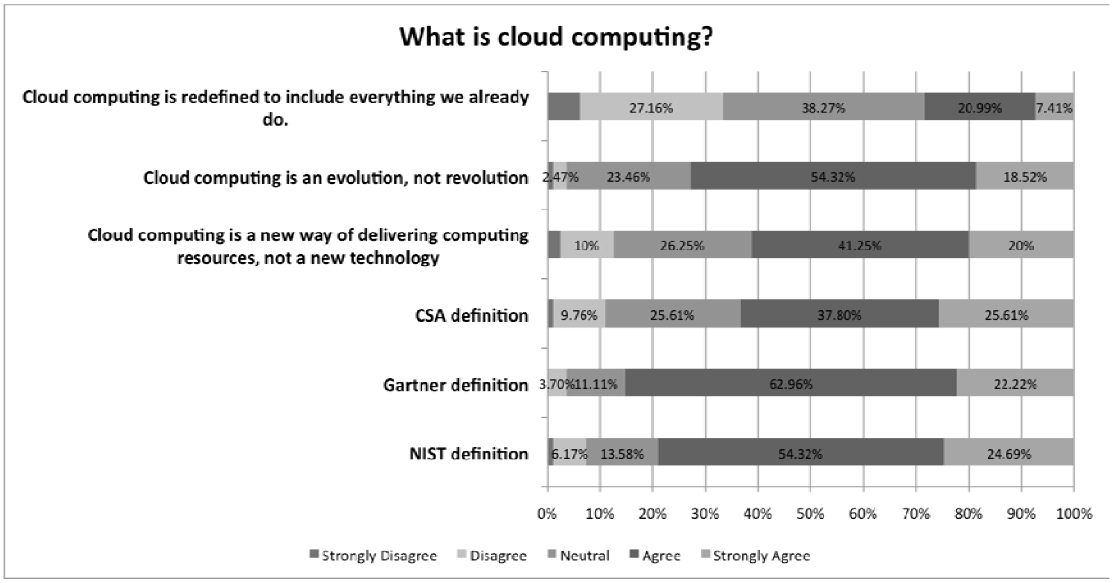


Fig 1. What is cloud computing?

On the definition of cloud computing, 82 respondents answered the question. As shown in Fig 1 above, 85.18% of them agree or strongly agree with the definition from Gartner (2009B):

> *"Cloud computing is a style of computer where scalable and elastic IT-related capabilities are provided 'as a service' to multiple external customers using Internet Technologies".*

79.01% of them agree or strongly agree with the 15th version of NIST definition of cloud computing (Mell and Grance, 2009):

> *"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".*

63.41% of them agree or strongly agree with the definition from Cloud Security Alliance (CSA, 2010):

> *"Cloud computing is an evolving term that describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from underlying infrastructure, and the mechanisms used to deliver them. Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing"*

72.84% of them agree or strongly agree that "cloud computing is an evolution, not revolution." 61.25% of them agree or strongly agree that "cloud computing is a new way of delivering computing resources, not a new technology". Only 28.4% of them agree or strongly agree with Oracle's CEO's famous remark "cloud computing is redefined to include everything we already do" (Farber, 2008), while 38.27% remain neutral.

4.2.2 Cloud computing as a trend

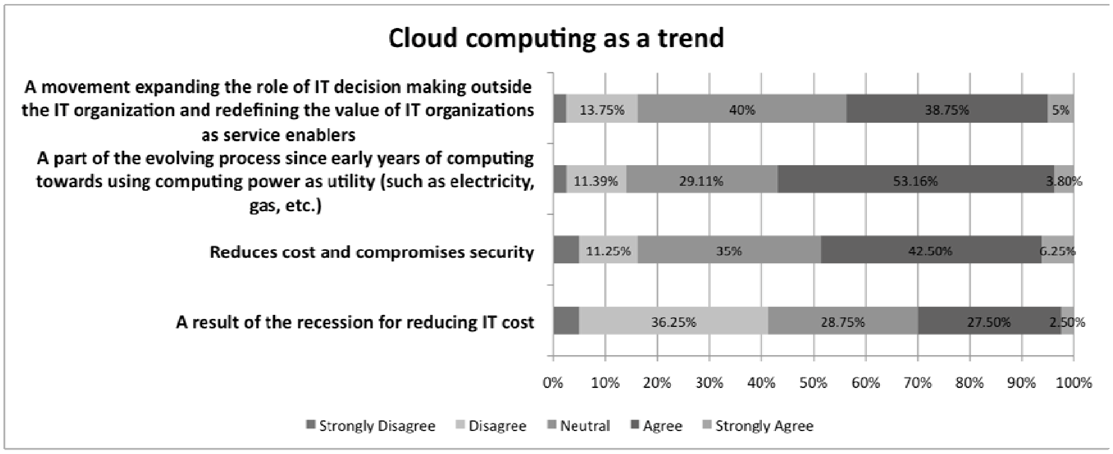


Fig 2. Cloud computing as a trend

82 respondents answered the question "cloud computing as a trend". As shown in Fig 2 above, 56.96% of them agree or strongly agree that cloud computing as a trend is "a part of the evolving process since early years of computing towards using computing power as utility (such as electricity, gas, etc.)". 48.75% of them agree or strongly agree that cloud computing as a trend "reduces cost and compromises security". 43.75% agree or strongly agree with the Gartner statement (Gartner, 2010) that cloud computing as a trend is "a movement expanding the role of IT decision making outside the IT organization and redefining the value of IT organization as service enablers", while 40% remain neutral. Only 30% agree or strongly agree that cloud computing as a trend is "a result of the recession for reducing IT cost".

4.2.3 Definition of cloud forensics

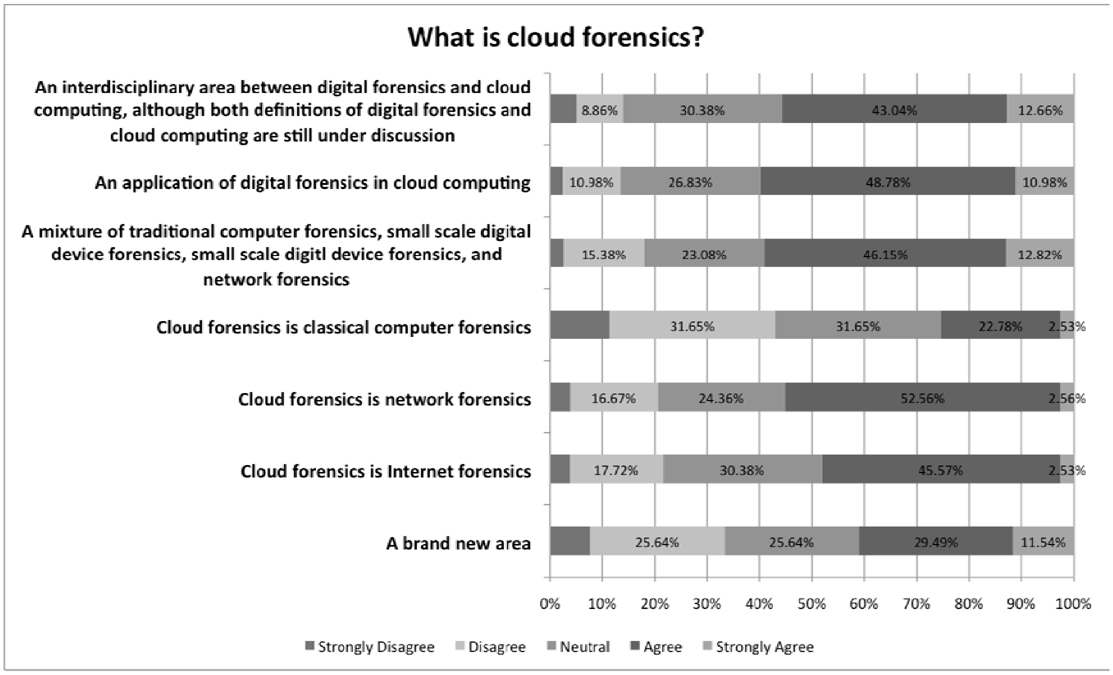


Fig 3. What is cloud forensics?

82 respondents answered the question “what is cloud forensics”. 59.76% of them agree or strongly agree that cloud forensics is “an application of digital forensics in cloud computing”. 58.97% agree or strongly agree that cloud forensics is “a mixture of traditional computer forensics, small scale digital device forensics, and network forensics". 55.7% agree or strongly agree that cloud forensics is “an interdisciplinary area between digital forensics and cloud computing, although both definitions of digital forensics and cloud computing are still under discussion" (Ruan et al., 2011). 55.12% agree or strongly agree that “cloud forensics is network forensics”. 48.1% agree or strongly agree that“cloud forensics is Internet forensics”. 41.03% agree or strongly agree that cloud forensics is “a brand new area”. 25.31% agree or strongly agree that “cloud forensics is classical computer forensics”.

4.2.4 Significance of cloud forensics

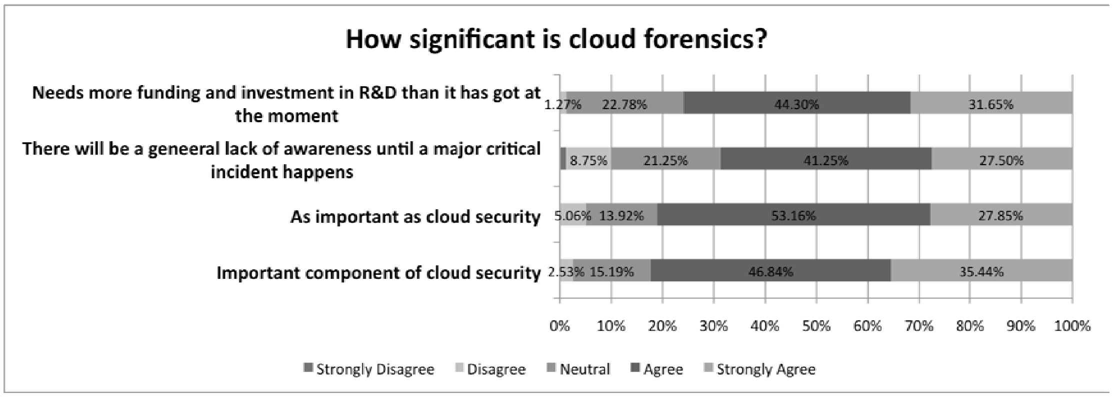


Fig 4. How significant is cloud forensics?

82 respondents answered the question on the significance of cloud forensics. 82.28% of them agree or strongly agree that cloud forensics is "an important component of cloud security". 81.01% agree or strongly agree that cloud forensics is "as important as cloud security". 75.95% agree or strongly agree that cloud forensics "needs more funding and investment in R&D than it has got at the moment." 68.75% agree or strongly agree "there will be a general lack of awareness until a major critical incident happens".

4.2.5 Impact of cloud computing on digital forensics

81 respondents answered the question on the impact of cloud computing on forensics, 50% of them agree that "cloud computing makes forensics harder", while 42% agree that "cloud computing makes forensics easier".

When asked why “cloud computing makes forensic harder”, comments from the participants are heavily focused on following issues:

- Loss of data control
- No access to physical infrastructure
- Legal issues of multi-jurisdiction, multi-tenancy and multiple ownership
- Lack of tools for large-scale distributed and virtualized systems

Other issues mentioned in the comments are

- No standard interfaces
- Data ownership
- No provider cooperation
- Difficulties in producing forensically sound and admissible evidence in court

When asked why “cloud computing makes forensics easier”, comments from the participants mentioned the following aspects

- More computing resources and processing power can be used for forensic investigation
- Cloud resources and computing power can be used for forensic research and development
- Rapidly scalable auditing, reporting, and testing analysis can be used for larger datasets and distributed applications
- Forensic implementations and activities can be centrally administered and managed
- Investigations can be provided as a service by the CSP
- Running forensic applications in the cloud may reduces cost

4.2.6 Dimensions of cloud forensics

88 respondents answered the question on the dimensions of cloud forensics, 84% of them agree there is a technical dimension for cloud forensics, 84% agree there is a legal dimension for cloud forensics, 75% agree there is an organizational/administrative dimension for cloud forensics, 42% agree there is a social dimension, and one respondent also added the ‘political’ dimension.

4.2.7 Uses of cloud forensics

88 respondents answered the question on the uses of cloud forensics, 83% of them agree that cloud forensics can be used for "investigations on digital crimes, civil cases, policy violations, etc.", 51% agree that it can be used for "regulatory compliance", 43% agree that it can be used for "due diligence", 43% agree that it can be used for “data and system recovery”, 36% agree that it can be used for "log monitoring", 26% agree that it can be used for "troubleshooting", comments were made to add

"security policy feedback" and "presentation of legal matters in legal venues" to the uses of cloud forensics.

4.3 Cloud Forensics Research Techniques

4.3.1 Challenges for cloud forensics

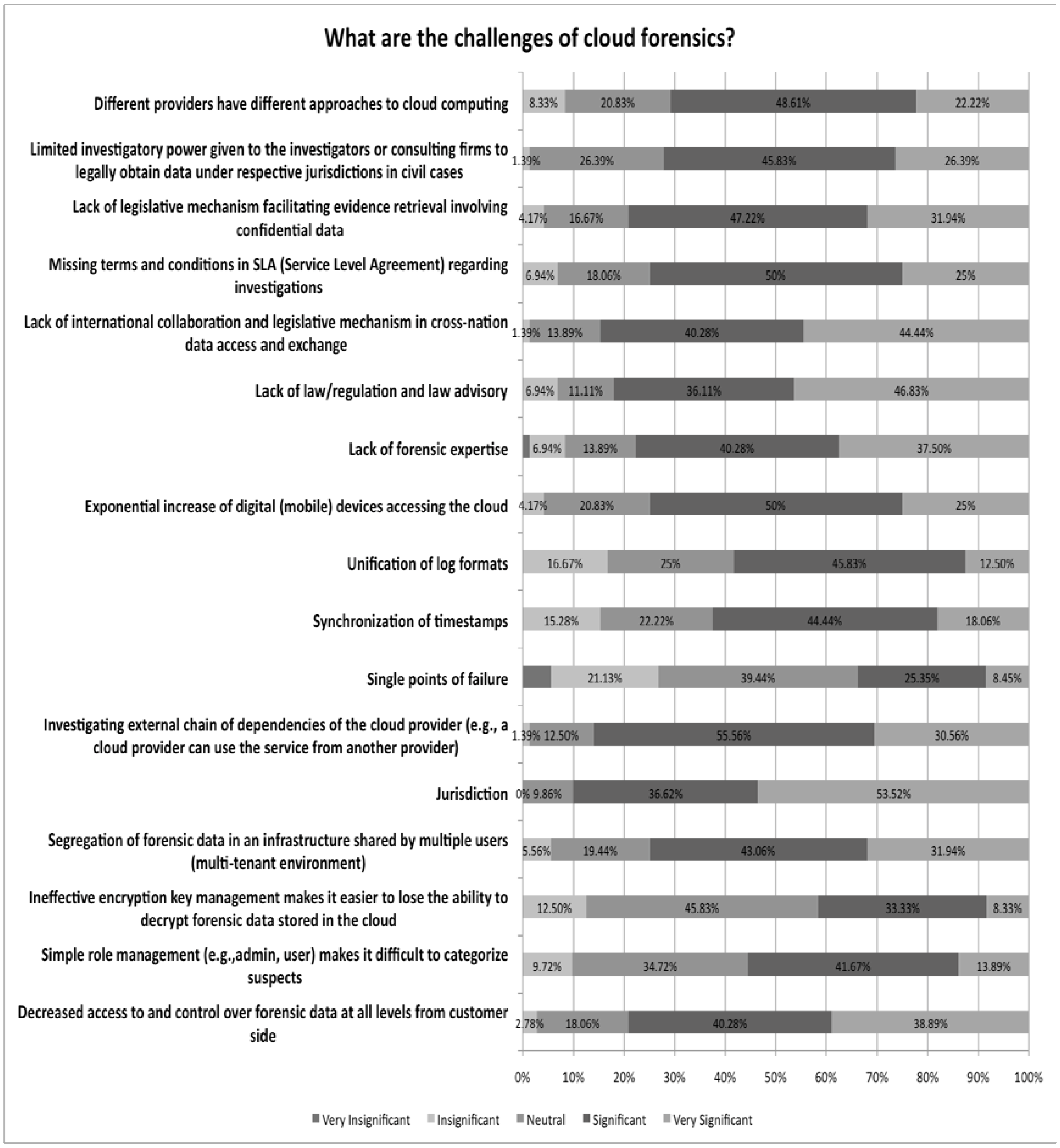


Fig 5. What are the challenges for cloud forensics?

72 respondents answer the question on the challenges for cloud forensics. As we can see from the survey results in Figure 5 above, the top 5 challenges for cloud forensics are:

(1) Jurisdiction (90.14% agree or strongly agree, 53.52% strongly agree)
(2) Investigating external chain of dependencies of the cloud provider (e.g., a cloud provider can use the service from another provider) (86.12% agree or strongly agree)

(3) Lack of international collaboration and legislative mechanism in cross-nation data access and exchange (84.72% agree or strongly agree)
(4) Lack of law/regulation and law advisory (82.94% agree or strongly agree)
(5) Decreased access to and control over forensic data at all levels from customer side (79.17% agree or strongly agree)

4.3.2 Opportunities for cloud forensics

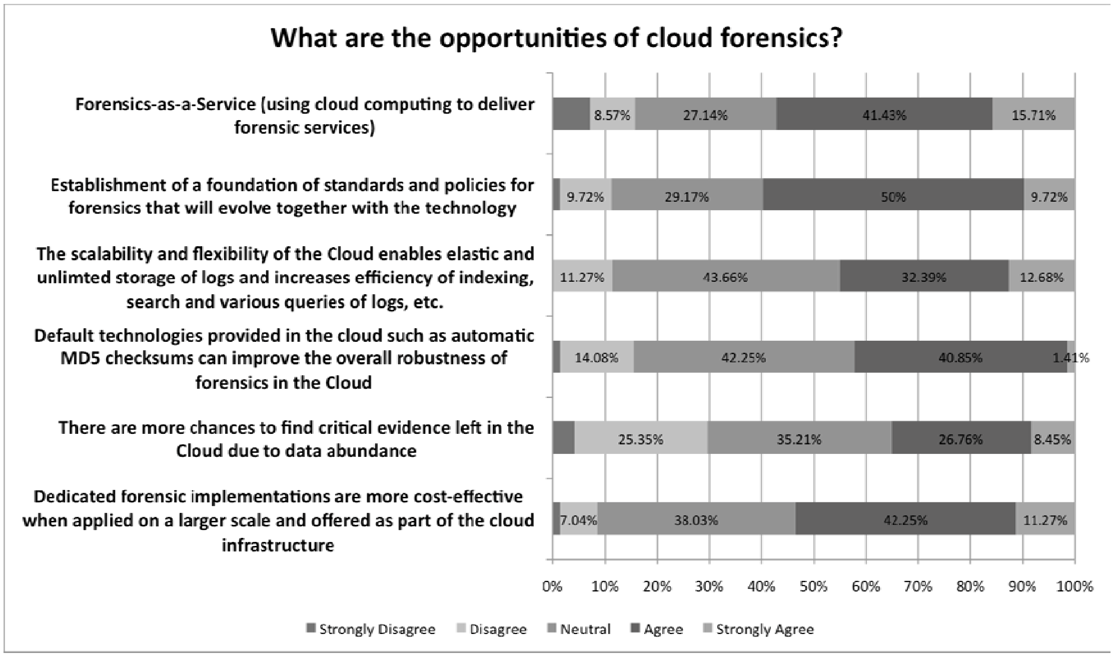


Fig 6. What are the opportunities of cloud forensics?

Compared to the challenges, more respondents chose to remain neutral towards the opportunities of cloud forensics. 72 respondents answered this question. As shown in Fig 6 above, 64.79% of them disagree, strongly disagree or remain neutral towards "there are more chances to find critical evidence left in the Cloud due to data abundance". 57.74% disagree, strongly disagree or remain neutral towards "default technologies provided in the Cloud such as automatic MD5 checksums can improve the overall robustness of forensics in the Cloud". 54.93% disagree, strongly disagree or remain neutral towards "the scalability and flexibility of the Cloud enables elastic and unlimited storage of logs and increases efficiency of indexing, searching and various queries of logs, etc.". However, 59.72% and 57.14% of them agree or strongly agree that the "establishment of a foundation of standards and policies for forensics that will evolve together with the technology" and "Forensics-as-a-Service (using cloud computing to deliver forensic services)" are opportunities for cloud forensics.

4.3.3 Valuable research directions for cloud forensics

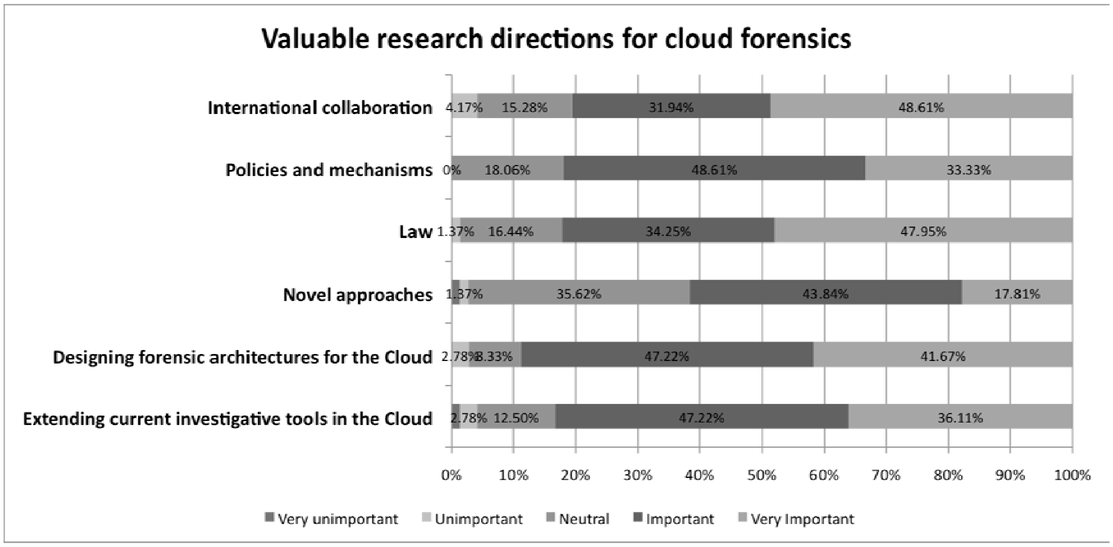


Fig 7. Valuable research directions for cloud forensics

As we can see from the survey results shown in Fig 7 above, the top 3 most important research directions are

(1) Designing forensic architectures for the Cloud (88.57% agree it is important or very important)
(2) Extending current investigative tools into the Cloud (82.86% agree it is important or very important)
(3) Law (82.2% agree or strongly agree, 47.95% strongly agree).

73 respondents answered this question.

4.4 Critical Criteria for Forensic Capability

4.4.1 Parties to be assessed for cloud forensic capability

74 respondents answered the question on who should be assessed for cloud forensic capability. 78% of them think the CSP should be assessed. 54% of them think the cloud customer should be assessed. 38% of them think the Internet service provider should be assessed. 32% of them think the cloud end user should be assessed. Several comments were made to add that the investigators also need to be assessed.

4.4.2 Importance of procedure and toolkits

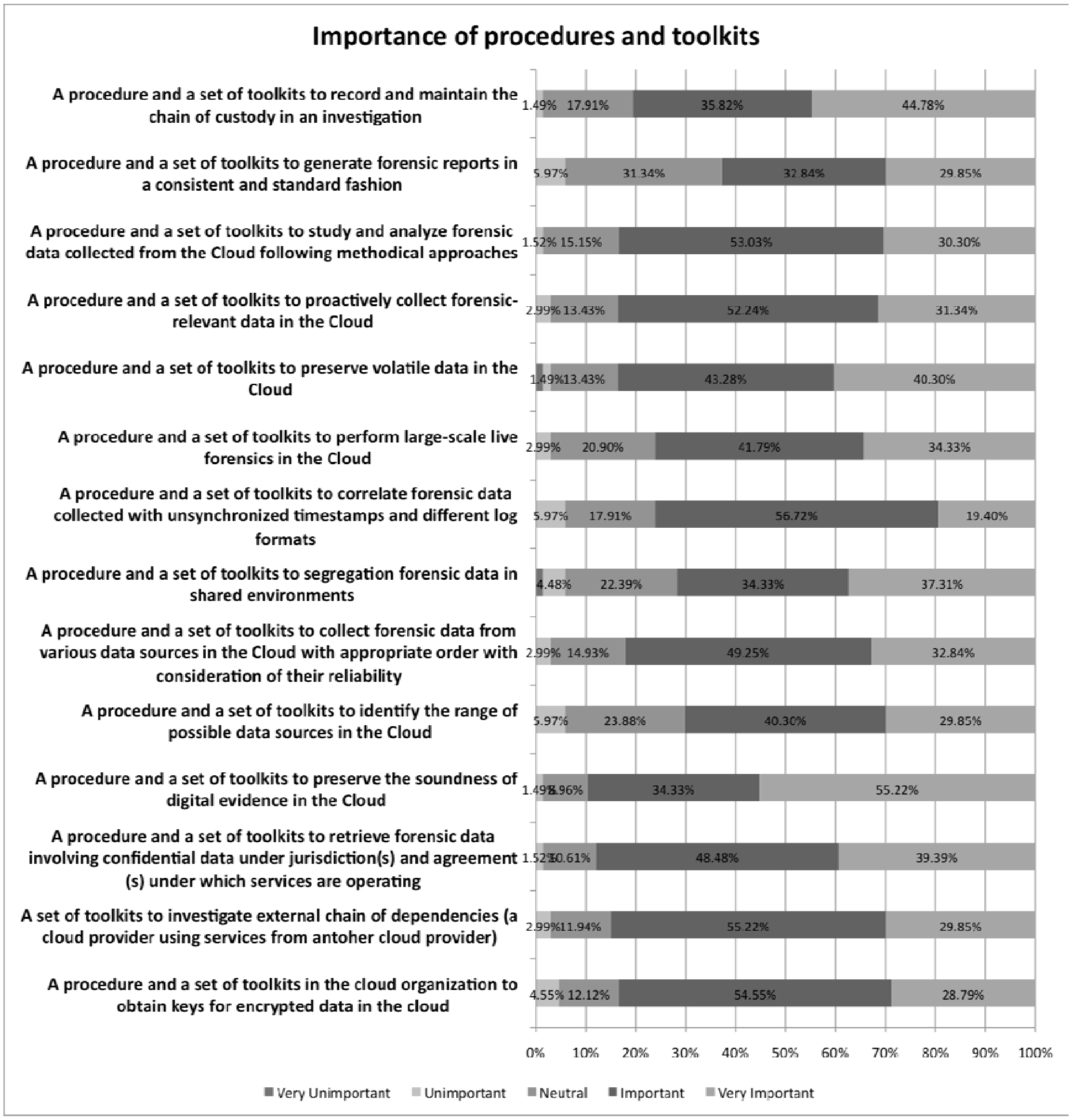


Fig 8. Importance of procedures and toolkits

Despite the close results, according to the survey results shown in Fig 8 above, the most needed tools and procedures for cloud forensics are:

(1) A procedure and a set of toolkits to preserve the soundness of digital evidence in the Cloud (89.55% think it is important or very important, 55.22% think it is very important)
(2) A procedure and a set of toolkits to retrieve forensic data involving confidential data under jurisdiction(s) and agreement(s) under which services are operating (87.87% think it is important or very important)
(3) A set of toolkits to investigate external chain of dependencies (a cloud provider using services from another cloud provider) (85.07% think it is important or very important)
(4) A procedure and a set of toolkits to preserve volatile data in the Cloud (83.58% think it is important or very important, 40.30% think it is very important)

(5) A procedure and a set of toolkits to proactively collect forensic data in the Cloud (83.58% think it is important or very important, 31.34% think it is very important)

67 respondents answered this question.

4.4.3 Staff importance

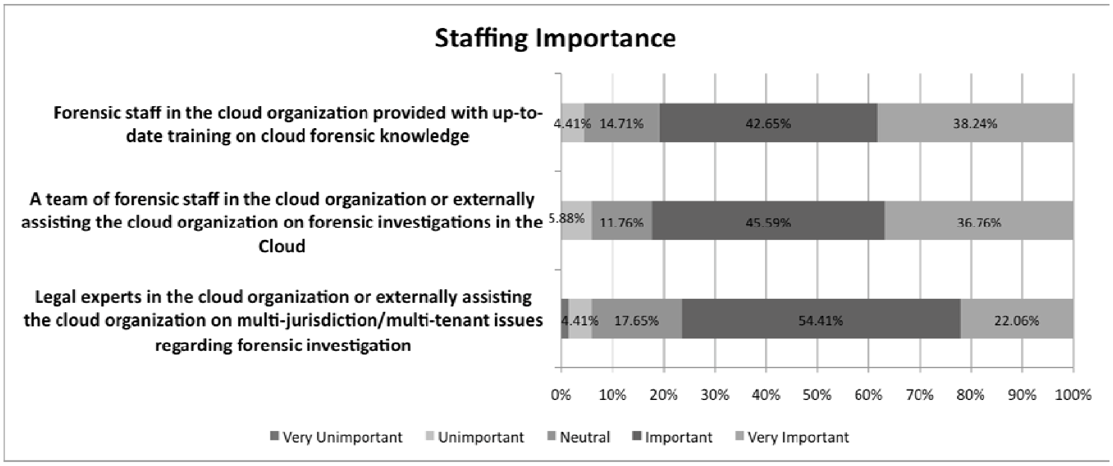


Fig 9. Staffing importance

69 respondents answered the question on staffing importance and they have reached majority consensus as for cloud forensic staffing, as shown from the results in Fig 9 above. 82.35% of them agree that to have “a team of forensic staff in the cloud organization or externally assisting the cloud organization on forensic investigations in the Cloud” is important or very important. 80.89% agree that to have “forensic staff in the cloud organization provided with up-to-date training on cloud forensic knowledge” is important or very important. 76.47% agree that to have “legal experts in the cloud organization or externally assisting the cloud organization on multi-jurisdiction/multi-tenant issues regarding forensic investigation” is important or very important.

4.4.4 Policy importance

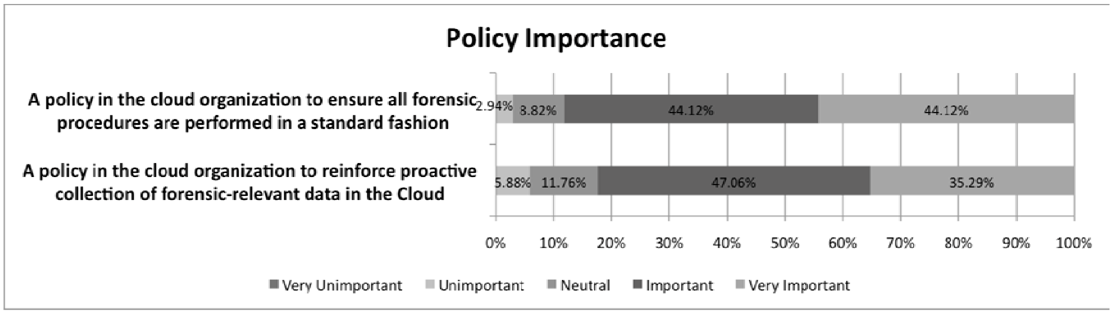


Fig 10. Policy importance

69 respondents answered the question on policy importance, and they have also reached majority consensus, as shown in the survey results in Fig 10 above, 88.34% of them agree that to have “a policy in the cloud organization to ensure all forensic procedures are performed in a standard fashion” is important or very important, and 82.35% agree that “a policy in the cloud organization to reinforce proactive collection of forensic-relevant data in the Cloud” is important or very important as for forensic policies within the cloud organization.

4.4.5 Agreement importance

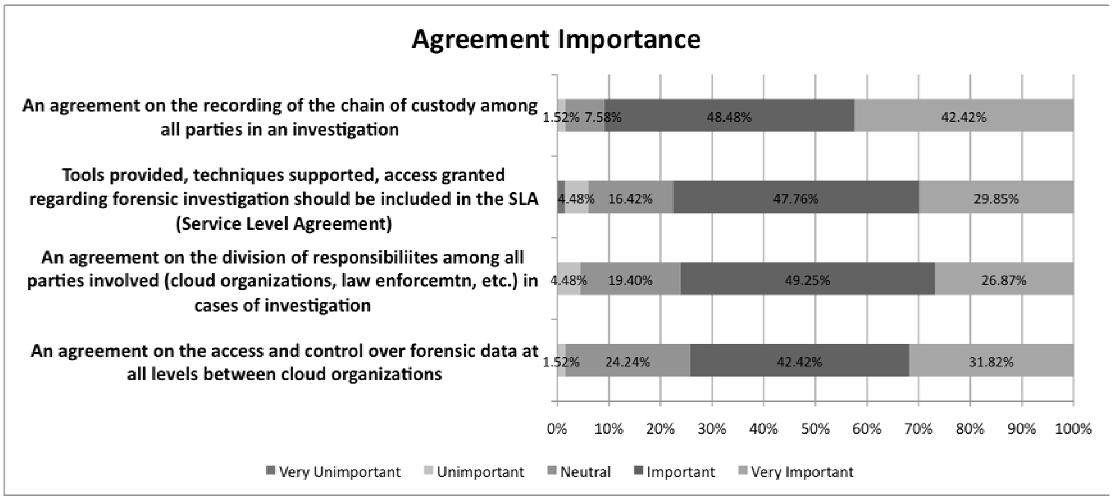


Fig 11. Agreement importance

As for agreements between various parties regarding cloud forensics, as shown in Fig 11 above, a mass majority of 90.9% of the respondents agrees "an agreement on the recording of the chain of custody among all parties in an investigation" is important or very important, and 42.42% think it is very important. 77.61% of the respondents agree that "tools provided, techniques supported, access granted regarding forensic investigation should be included in the SLA (Service Level Agreement)" is important or very important. 76.12% of the respondents agree "an agreement on the division of responsibilities among all parties involved (cloud organizations, law enforcement, etc.) in cases of investigation" is important or very important. And 74.24% of the respondents think that "an agreement on the access and control over forensic data at all levels between cloud organizations" is important or very important. 67 respondents answered this question.

4.4.6 Guideline importance

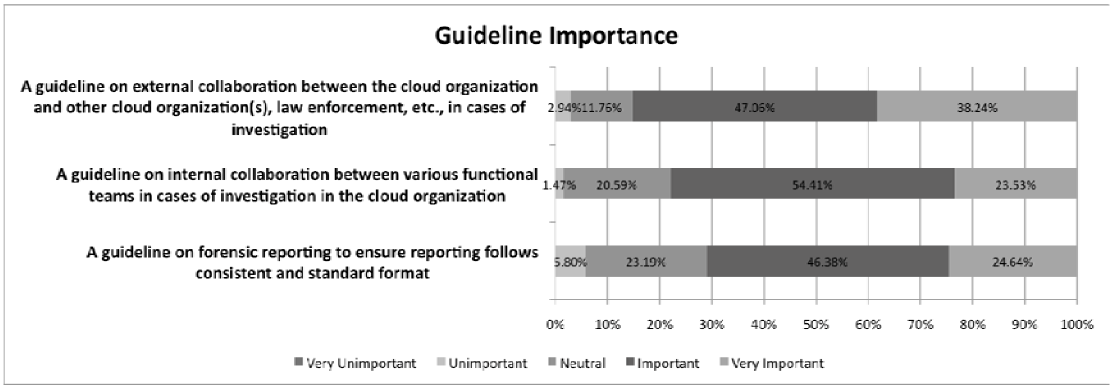


Fig 12. Guideline importance

Lastly, 68 respondents who answered the question on guideline importance have reached majority consensus, as shown in Fig 12 above. 85.3% of the respondents agree that "a guideline on external collaboration between the cloud organization and other cloud organization(s), law enforcement, etc. in cases of investigation" is important or very important. 77.94% of the respondents agree that "a guideline on forensic reporting to ensure reporting follows consistent and standard format" is important or very important. 71.02% of the respondents agree that "a guideline on internal collaboration between various functional teams in cases of investigation in the cloud organization" is important or very important.

5. CONCLUSION

In this paper, we presented the results and a preliminary analysis on the survey 'cloud forensics and critical criteria for cloud forensic capability' towards a group of digital forensic experts and practitioners who have good knowledge and sufficient experience in the field of digital forensics. From the survey results, we found out the majority of our respondents agree that cloud forensics is an application of digital forensics in cloud computing and is a mixture of traditional computer forensics, small-scale digital device forensics, and network forensics. The respondents are more concerned about the challenges for cloud forensics, such as jurisdiction issues, and the lack of international collaboration, than optimistic about the opportunities of cloud forensics. As a result, a forensic architecture needs to be developed for cloud computing environments. Furthermore, the respondents have reached a consensus on what kind of tools, procedures, staffing, agreements, policies and guidelines are required for a cloud forensic capability.

6. FUTURE WORK

We will continue running this survey for a longer period of time in order to get more responses so that analysis can be made in depth. Base on the survey results we will start working on a framework of critical criteria for cloud forensic capability to suggest to the cloud computing industry as the next step.

7. REFERENCES

Buyya, R., Chee Shin Yeo, Venugopal, S (2008) 'Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilites' in Proceedings of 10th IEEE International Conference on High Performance Computing and Communications

Cloud Security Alliance [CSA] 2009 Security Guidance for Critical Areas of Focus in Cloud Computing V2.1.

Farber, D. (2008) Oracle's Ellison nails cloud computing. CNET September 26

Garfinkel, S.L. (2010) 'Digital forensics research: The next 10 years' Digital Investigation 7: pp64-73

Gartner (2009A) Worldwide Cloud service revenue will grow 21.3 percent in 2009.

Gartner (2009B) Gartner Highlights Five Attributes of Cloud Computing. Gartner Press Releases June 23

Gartner (2010) Gartner's Top Predictions for IT Organizations and Users, 2011 and Beyond: IT's Growing Transparency

Gens, F. (2008) IT Cloud services forecast – 2008 to 2012: A key driver of new growth. IDC

INPUT (2009) Evolution of the Cloud: The future of cloud computing in government.

Merrill Lynch (2008) The Cloud wars: $100+ billion at stake.

Morrill, D. (2008) Cloud Computing Making Forensics Easier, CloudAve September 22

Perry, R., Hatcher, E., Mahowald, R.P., Hendrick, S.D. (2009) Force.com Cloud platform drives huge time to market and cost savings. IDC

Ruan, K., Carthy, J.,Kechadi, T., Crosbie, M. (2011) 'Cloud forensics: An overview' Advances in Digital Forensics VII

Sawyer, J.H. (2009) Hazy Forecast for Cloud Computing Forensics, Darkreading March 9

SOFTWARE PIRACY FORENSICS: IMPACT AND IMPLICATIONS OF POST-PIRACY MODIFICATIONS

P. Vinod Bhattathiripad
Cyber Forensic Consultant
Polpaya Mana
Thiruthiyad, Calicut-673004
Kerala, India
Telephone: +91-495-2720522, +91-94470-60066 (m)
E-mail: vinodpolpaya@gmail.com; vinodpolpaya@yahoo.co.in

Lt. Dr. S. Santhosh Baboo
Reader
P G & Research
Dept of Computer Science
D.G.Vasihanv College
Chennai, India.
E-mail : santhos2001@sify.com

ABSTRACT

Piracy is potentially possible at any stage of the lifetime of the software. In a post-piracy situation, however, the growth of the respective versions of the software (both the original and pirated) is expected to be in different directions as a result of expectedly different implementation strategies. This paper shows how such post-piracy modifications are of special interest to a cyber crime expert investigating software piracy and suggests that the present software piracy forensic (or software copyright infringement investigation) approaches require amendments to take in such modifications. For this purpose, the paper also presents a format that is jargon-free, so as to present the findings in a more intelligible form to the judicial authorities.

Keywords: Piracy, post-piracy modifications, software piracy, source code, copyright, software copyright infringement, software piracy forensics, database forensics, MIS forensics, AFC, SCAP, technical expert, substantial similarity test, CDAC

1. INTRODUCTION

Piracy is potentially possible at any stage in the lifetime of the software. If and when that happens, the original and pirated versions of the software will continue to be used contemporaneously. This being so, in the post-piracy period, the profile of the pirated[1] could well be in a different pattern to that of the original[2] as both the original developer as well as the pirate may modify the respective versions in their own ways. Because of this, although the original and the pirated software are prone to grow functionally in almost the same direction (because the post-piracy life time of the pirated software is in the same functional area of expertise as that of the original), the growth is expected to be with different modification strategies (because both are handled by different persons). This phenomenon of different patterns of growth is a very valuable and useful dimension of study for the expert in cyber forensics. A

[1] Throughout this article, pirated means the allegedly pirated software

[2] Throughout this article, original means the version of the software that the complainant submits to the law enforcement agency for software piracy forensics. This article presupposes that the law enforcement agency has satisfactorily verified the legal aspects of the documentary evidence of copyright produced by the complainant and is convinced that the complainant is the copyright holder of this version of the alleged software.

proper study of post-piracy modification of the pirated will contribute substantially to the reliability of software piracy forensic investigation. This article attempts to discuss the impact and implications of post-piracy modifications in software piracy forensics (or software copyright infringement investigation) and to suggest that proper amendments be made in the existing forensic approaches / techniques so that evidence concerning post-piracy modifications gets proper consideration and treatment.

Software piracy forensic investigation often requires comparison of the original with the pirated by juxtaposing the two. In order to perform the task of comparing two software packages, several software tools are used and these tools are based mostly on academically accepted mathematical techniques and theoretical frameworks like Discourse Analysis (Van der Ejik, 1994), SMAT (Yamamoto et al, 2004), and MOSS (Lancaster and Culwin, 2004). A recently (November-2009) edited work "Handbook of Research on Computational Forensics, Digital Crime, and Investigation: Methods and Solutions" (Chang-Tsun Li, 2010) prescribes SCAP (Frantzeskou, 2007) for comparison of two software packages. An exception to all these (because of the judicial acceptance in the US) is the theoretical frame work called AFC (Abstraction-Filtration-Comparison) (Walker, 1996) which has been professionally implemented by a French software firm European Software Analysis Laboratory in the form of the product namely SIMILE Workshop (ESALab, 2007). All these techniques, theoretical frameworks and tools are capable in their respective areas of software piracy investigation. Even so, none of them properly and adequately deal with matters related to post-piracy modifications in the pirated and all these need to be made sensitive to the implications of post-piracy modifications.

Post piracy modifications need also to be incorporated into many other theoretical proposals and studies in software piracy forensics. For instance, the Ginger Myles (2006, p.69), proposes watermark (as a weaker evidence) to indicate that "one program is likely to be a copy of the other", needs further explanation on extending watermarking to post-piracy modifications in the pirated. Anthony Reyes (2007) beautifully discusses areas of difficulties, misconceptions and flaws in the cyber investigative methodology by explaining techniques for preparing for prosecution, testifying, and incidence response, solving legal issues, conducting seizure procedure, performing data analysis, and preventing of cyber crimes, including software piracy, but this work requires further considerations on analysis of post-piracy schematic changes. Another example is the work on software forensics by Robert M. Slade (2004), where he explains from his experience, several ways of collecting evidence of software piracy and presents overviews of forensics programming, plagiarism detection, code analysis, source code recovery and even forensics linguistics. Even so, the book does not cover matters related to post-piracy modifications in the pirated.

2. ESTABLISHING THE CRIME

As mentioned above, the starting point of the investigation into piracy is the juxtaposed comparison of the original and the alleged pirated versions. This delicate and demanding situation of comparing two software packages arises usually when one party lodges a complaint of software piracy or copyright infringement against the other. A full-fledged forensic investigation of the pirated software has to be done to establish piracy. As software piracy investigation involves technical comparison, the judge usually appoints an uninvolved cyber forensic expert for the task. Given the source codes from two different software systems, the technical expert concentrates on digging out the pieces of potential evidence of copyright infringement by evaluating the similarities and commonalities that form the basis for validating or invalidating the alleged crime. The duty of the cyber forensic expert is to establish possible piracy through a rigorous formulation of *statistical occurrences* of the data structures, variables, data base tables, fields, modules, procedures, logic, remark, error and blunders in the allegedly pirated software and arrive at several values, preferably in percentages, to indicate the strength of piracy (Author, 2009, p.54), all of which require comparing the original and the pirated source codes, database schemas and procedures. Alternatively, the cyber forensic expert can abstract the original as well as pirated, filter out globally common elements from them and then compare the remaining two kernels in order to establish copyright infringement (Walker, 1996). Either way, the procedure needs to take into account evidence concerning post-piracy modifications.

3. PROCEDURE

As a prelude to comparing the two software systems, the cyber forensic expert can ask the original developer (the complainant) to make available their pre-modified version of the source code, the embedded images and finger prints, the database procedures and the database schemas that were prevailing at the time of piracy (see footnote 2 above). At the same time, the source code, the embedded images and finger prints, the database procedures and the database schemas of the pirated are generally made available for comparison by the police or judiciary through a seizure procedure (Authors, 2009, p.177) or through a disclosure procedure wherein the technical expert (or sometimes the plaintiff too) has direct access to the defendant's source code (Hollaar, 2002, p103). (This procedure might vary from country to country). It is highly unlikely that this (or thus made available) version is the pre-modified version of the pirated software. The seizure procedure usually ends up seizing some as-and-when available version, mostly a modified / customized version of the pirated software, leaving the cyber forensic expert with this version of the pirated software to compare with the original.

4. THE IDENTIFICATION OF POST-PIRACY MODIFICATIONS

The modified / customized version of the pirated software that is made available through the seizure would most certainly have gone through a few, if not quite a lot of, modifications. This being so, before listing out the similarities (and commonalities) and making an expert judgment from their statistical representation and profile, the cyber forensic expert has to first generate, ideally, the originally pirated pre-modified version of the software out of the seized version by identifying and filtering out the post-piracy modifications, if any, from it. These post-piracy modifications can be found in various parts of the pirated software, namely, the source code, object files, embedded fingerprints and images, database procedures, and/or the database schemas and so, the cyber forensic expert has to necessarily identify and filter out all identifiable post-piracy modifications from all these parts, one by one. The possibility of their tainting the statistical rigor of the results of the comparison will thus be eliminated or at least minimized. The expert has to first convert the seized, pirated into its pre-modified infant form. The objective of and hence the emphasis on this process is not the detection or confirmation of piracy but the identification and filtering out of all post-piracy modifications by a more rigorous scrutiny of differences between the two codes. The output of this initial process will be something closest to the pre-modified version, which forms the basis for a reliable eventual comparison with the original.

5. POST-PIRACY MODIFICATIONS – THE WHY AND THE HOW

The above facts demand a detailed study on post-piracy modifications and ways to incorporate their role, effect and impact in the report to the court. Different techniques to analyze post-piracy modifications are required because post piracy changes can happen along a variety of parameters, in a variety of ways, and for a variety of reasons not all of which may be visible, noticeable and reliable initially during the cyber forensic investigation. However, on detailed investigation, they all can be seen to become relevant and can largely influence the cyber forensic report.

Difference is not exculpation from piracy: While similarity between two sets of software is most certainly indicative of piracy, difference is not exculpation from piracy either. In fact it is the differences that trigger the need for careful observation since they may be the result of post-piracy modifications. Such modifications may be motivated by a number of factors. For instance, one motivation for modification could be a customer demanding an additional feature in the software.

Another could be a government-directive to be incorporated in business. In both cases the software will have to be subsequently modified accordingly. In order to incorporate a customer request or government directive into the software, the pirate may modify, say, the structure of the table by introducing one or more new fields into it. While a government directive can bring about a modification in both original and the pirated (with different implementation patterns), the implementation of a customer request by the pirate brings about a change only in the pirated. Such a modification would induce difference between the pirated and the original data base tables and it is the duty of the cyber forensic expert to consider and properly question these differences during software piracy investigation. In addition to customer requirements and government directives, the pirate himself / herself may introduce intentional changes in the database schema in order to escape copyright violation litigations in the future and such intentional changes also cause questionable differences. It needs to be stressed that customer requests, government directives, and intentional changes are only some of the potential reasons and motivations that can cause questionable difference in the pirated from the original, and database is only one of the areas where such questionable differences can be found in the pirated.

Differences exist in various forms: During the post-piracy lifetime of the software, the pirate might modify database schema (which formally defines the tables in each database, the fields in each table, and the relationships between fields and tables), by adding, removing or editing a few fields, as part of either the post implementation tuning up or of the customization of the pirated software. Thus, any difference found in the schema of the pirated can either be in the form of the presence of one or more additional fields (that are absent in the original) or absence of one or more fields (that are found in the original) or modifications in names and/or other properties of any of the fields already existing intact in both original as well as pirated.

An example of suspected post-piracy modifications: These forms of post-piracy modifications and the resulting forensic challenges / difficulties[3] can be better explained with an example of a database-related situation. Table-1, which was extracted from a software comparison report (Author, 2002, p.14), gives a sample of database table-level comparison. The first part of the table corresponds to the original and the second, to the pirated. Fifteen out of sixteen fields in the original are found exactly in the same sequence in the pirated also (93% similarity or 93 % of the fields in the original can be mapped to at least one field in the pirated) and 15 out of 21 fields in the pirated are found in original also exactly in the same sequence (71% similarity). While these two percentages are fair enough to give some clue to possible piracy, there is still scope for the expert to further analyze the two database schemas with the intention of improving on the above two percentages, suggestive of suspected piracy. A further analysis of the remaining 1 field in the original and 6 fields in the pirated results in more reliable percentages (of nomenclature level piracy) than the above two. This is better explained here with the one field namely USERNAME in the original which can be seen to have some correspondence with two fields, namely, CREATEDUSER and MODIFIEDUSER in the pirated. Just as the field name USERNAME is one of the globally used variable names to save the name of the author of the transaction, the two field names, viz., CREATEDUSER and MODIFIEDUSER in the pirated are also globally used to save the names of the authors of the transaction and thus, a correspondence can be attributed between them. While the field, USERNAME, has some degree of nomenclature level similarity with CREATEDUSER and MODIFIEDUSER, the degree of dissimilarity can be possibly because of a post-piracy development, in which the ‘pirate’ himself may have replaced USERNAME with two fields, namely, MODIFIEDUSER and CREATEDUSER. This

[3] These challenges / difficulties are usually explained using theoretical situations involving source codes but the situations used in this article are live and data base related. Live situations are often more valuable than theoretical ones. Moreover, any post-piracy modification in the database would generally subsume the corresponding change in the respective source code too.

possibility is further strengthened by the similarity[4] in the other properties of these fields. For instance, all these three fields are of type CHAR, and are of length six (see Table-1). Thus, this strong possibility of post-piracy modification demands either re-calculation of the above two percentages or incorporating this possibility separately in the cyber forensic report. By mapping USERNAME in the original to both the CREATEDUSER and MODIFIEDUSER in the pirated, it can be seen that each of the 16 fields in the original can be mapped to at least one field in the pirated, and thus the above given 93% similarity in effect becomes 100%. Similarly, by mapping both the CREATEDUSER and MODIFIEDUSER in the pirated to USERNAME in the original, it can be seen that 17 out of 21 fields in the pirated can be mapped to at least one field in the original, which increases the above given 71% to 81%. As 17 out of 21 fields in the pirated could successfully be mapped to at least one field in the original, the remaining 4 fields, namely ACCTRANSCHEQUEDATE, ACCTRANSISSUEBANK, INTERNALENTRY, and VOUCHERTYPE, also require further attention and analysis, as these 4 fields also can possibly be post-piracy fields. Thus, an analysis of the modifications happened in the schemas of the pirated does shed further light on the suspected piracy. This example illustrates how a post-piracy modification can happen along database fields (one of the above listed parameters of a database) and shows a way for the expert to overcome the resulting forensic challenges / difficulties by using his / her expertise, intuition and common sense in identifying post-piracy modifications. The test used in the above example is “Substantial similarity “ test (Davis, 1992).

[4] Further, such strong similarity can also happen if the software was originally made by the respondent but later pirated and then unethically copyrighted by the complainant. The law enforcement agency needs to collect evidence from the respondent and from other sources and prepare the case accordingly.

Table -1: Comparison of the structures of two database tables (Author, 2002, p.14)

Original software's database table structure	
TABLE NAME: ACCOUNTTRANSACTIONS	
Field names and properties	
1	ACCOUNTHEAD CHAR(8) NOT NULL
2	FINYEAR CHAR(4) NOT NULL
3	ACCTRANSVOUCHERNUMBER CHAR(8) NOT NULL
4	ACCTRANSBILLNUMBER CHAR(11)
5	ACCTRANSCHEQUENUMBER CHAR(10)
6	ACCTRANSCREDIT NUMERIC(12,2) DEFAULT 0.00
7	ACCTRANSDATE DATE
8	ACCTRANSDEBIT NUMERIC(12,2) DEFAULT 0.00
9	ACCTRANSDESCRIPTION CHAR(300)
10	ACCTRANSRECDATE DATE
11	ACCTRANSRECONCILE CHAR(1) DEFAULT 'N'
12	ACCTRANSTYPE CHAR(2)
13	COSTCENTRE CHAR(2)
14	DIVISION CHAR(2)
15	USERNAME CHAR(6) NOT NULL
16	MACHINEID CHAR(10) NOT NULL
Seized (allegedly pirated) software's database table structure	
TABLE NAME: ACCOUNTTRANSACTIONS	
Field names and properties	
1	ACCOUNTHEAD ACCOUNTHEAD_DM /*CHAR(8) CHARACTER SET NONE*/ NOT NULL
2	FINYEAR /*RDB$914*/ CHAR(4) CHARACTER SET NONE NOT NULL
3	ACCTRANSVOUCHERNUMBER /*RDB$915*/ CHAR(8) CHARACTER SET NONE NOT NULL
4	ACCTRANSBILLNUMBER /*RDB$916 */ CHAR (11) CHARACTER SET NONE
5	ACCTRANSCHEQUENUMBER /*RDB$917 */ CHAR(10) CHARACTER SET NONE
6	ACCTRANSCREDIT AMOUNT_DM /*NUMERIC(15,2 */ DEFAULT 0.00
7	ACCTRANSDATE /* RDB$918 */ DATE
8	ACCTRANSDEBIT AMOUNT_DM /*NUMERIC(15,2) */ DEFAULT 0.00
9	ACCTRANSDESCRIPTION /*RDB$919 */ CHAR(300) CHARACTER SET NONE
10	ACCTRANSCHEQUEDATE /*RDB$920 */ DATE
11	ACCTRANSISSUEBANK NAME_DM /*CHAR(50) CHARACTER SET NONE*/
12	ACCTRANSRECDATE /*RDB$921 */ DATE
13	ACCTRANSRECONCILE BVALUE_DM /*CHAR(1) CHARACTER SET NONE */ DEFAULT 'N'
14	ACCTRANSTYPE BOOKCODE_DM /*CHAR(2) CHARACTER SET NONE */
15	COSTCENTRE /*RDB$922*/ CHAR(2) CHARACTER SET NONE
16	DIVISION /*RDB$923*/ CHAR(2) CHARACTER SET NONE
17	INTERNALENTRY BVALUE_DM /*CHAR(1) CHARACTER SET NONE */ DEFAULT 'N'
18	VOUCHERTYPE /*RDB$924*/ CHAR(1) CHARACTER SET NONE*/ DEFAULT 'E'
19	MODIFIEDUSER USERID_DM /*CHAR(6) CHARACTER SET NONE */
20	CREATEDUSER USERID_DM /*CHAR(6) CHARACTER SET NONE */ NOT NULL
21	MACHINEID MACHINEID_DM /*CHAR(10) CHARACTER SET NONE */

Post-piracy modifications and AFC test: If the above investigation is done using the AFC test (the recognized test in the US judiciary for software copyright infringement investigation), most of the data base fields mentioned above can be filtered out (from the original as well as pirated) during the filtration stage of test (Hollaar, 2002, p89) and in such case these fields will not be available for final comparison. This is because, most of these data base fields carry names which are globally not uncommon and thus these fields may be treated under "widely accepted programming practices within the computer industry". This sort of filtration of AFC can seriously impair the evaluation of the evidence concerning post-piracy modifications in the pirated and so defeat the purpose of the software copyright infringement investigation. Thus, this sort of filtration is tantamount to an act of discarding valuable digital evidence of post-piracy modifications in the databases. (What is required here is to re-design the filtration stage of AFC so as to avoid filtering out the possible evidence of post-piracy modifications from the pirated.)

Factors encouraging post-piracy modifications: Two factors that encourage the possibility and extent of post-piracy modifications are; (a) the time lapse between the actual act of piracy and the complaint from the original developer; and (b) the market of the pirated version. It is commonsense to believe that there is a direct though not systematic connection between the extent of post-piracy modifications and the time available to do it on the one hand and the nature and extent of the consumer (of the pirated version) on the other. A good illustration of the importance of the time factor is the suit Sesame Software Solutions Vs. Perfect Software, filed in 2007, the final verdict of which is still pending in a court in India. The complainant in the case had alleged that four of his former employees had appropriated his software product and were marketing it as their own since the time they left his employment six years earlier. The alleged software seized on court order through a raid (Author, 2007) was the latest (as-is where-is) version and very probably a modified version. The court appointed the cyber forensic division of Centre for Development of Advanced Computing (CDAC, Thiruvananthapuram, India) as the expert to investigate the piracy. In a case like this, since there is every possibility that the software would have been modified drastically during the six years, no attempt by the CDAC to compare the original with the pirated would yield realistic results if the post-piracy modifications are ignored, particularly in view of the long period of six years involved. The incidence of such suits may well be on the rise globally.

Choice of Tool matters: In the situation of software piracy forensics, often what determines the credibility of the result is not just a matter of the professional status of the expert, or the dexterity of the analysis but also of the appropriate choice of tools and approaches used in the comparison of software. For instance, the file comparison utility software used by CDAC in the above case was not versatile enough to identify and filter out the post-piracy modifications. The weakness arising from non-use of the required tool is illustrated in Table 2, extracted from pages 71 and 119 of the cyber forensic analysis report performed by CDAC (2008) in the above mentioned suit, Sesame Software Solutions Vs. Perfect Software. This table provides an instance of table level comparison, in which the file comparison utility software (whose name is not mentioned in the report) used by CDAC found that some of the data base fields in the pirated do not prima facie appear to be fully similar to those in the original (see the last part of table-2). Nowhere in table-2 (and also in the report) is there any mention about the possible post-piracy modifications that could have happened during the six-years of post-piracy life of the pirated. This laxity can be because of the lack of skills of the file comparison utility used in this case by CDAC. A supplementing and in-depth manual comparison (with the intention of identifying the post-piracy modifications) would have easily revealed that the fields AcgCode and ACGrCode differ only by the character 'r', AcsCode and AcSuCode, by the 'u' and AcsName and AcSuName by the 'u'. In other words, the first three fields in the pirated are different from the respective three fields of the original only by a single character each and this difference can possibly be a result of post-piracy modifications with an explicit intention of obfuscating similarities. This possibility has not been explained properly in the report. The cyber forensic expert could have 'properly' reported to the court that the three "relatively similar fields" (see Table-2) differ only by a single character and that this minor difference (favouring the alleged culprit) could possibly have been brought about by an intentional act of obfuscation. In other words, a thorough expert would question and further explore manually the minor, nominal degree of dissimilarity in the three "relatively similar fields" to weed out the possibility of a deliberate act of obfuscation and judiciously report the findings to the court. Once the questionable nature of such factors has been pre-supposed, that would then logically form a legitimate precedent for a similar investigation of the remaining two fields namely UsrCode and UsrEnteredOn, which too may well be suspected as a post-piracy add-on. Almost all other table level comparison results in this report (CDAC, 2008) are incomplete in this manner. Quite a lot of such superficial differences, thus, need further manual supplementary analysis to establish their legitimacy and such manual analysis draws upon clear insight, commonsense, hands-on experience, and intuitive skill of the expert.

Table-2: Results of comparison of the two database table structures (CDAC, 2008, p.71, p.119)

Original software's database table structure	
TABLE NAME: AcSubGroup	
Field names and properties	
1	[ACGrCode] [int] NULL
2	[AcSuCode] [int] NULL
3	[AcSuName] [varchar] (30) NULL
Seized (allegedly pirated) software's database table structure	
TABLE NAME: AcSubGroup	
Field names and properties	
1	[AcgCode] [tinyint] NOT NULL
2	[AcsCode] [tinyint] NOT NULL
3	[AcsName] [varchar] (40) NOT NULL
4	[UsrCode] [varchar] (5) NOT NULL
5	[UsrEnteredOn] [datetime] NOT NULL
Results of comparison of the above two table	
	Number of fields in the allegedly pirated: 5
	Number of fields in the original : 3
	Same fields : 0
	Relatively similar fields : 3
	Not similar fields : 2 fields in the allegedly pirated and 0 fields in the original

A further point to look into while choosing the right tool is the tool's ability to analyze the positioning or placement of the post-piracy add-ons in the software. During the post-piracy modifications, the pirate may add additional fields at the end of the table structure or in between two existing fields in the table structure. Even if several modern data base management systems (DBMSs) provide techniques to introduce the new field logically in between two fields, say, between 4th & 5th fields, programmers usually tend to add the new field at the end of the table. Some DBMSs provide necessary software facility to add a new field without letting the user be bothered about the position of the new field in the database table and these tools usually place the new field at the end of the respective database table.

Often programmers either opt to add the additional field at the end of the table or simply don't care about inserting the additional field in the proper logical position. Some of them just use the software facility to add a new field and simply don't bother about where the software facility places the new field in the table. What is more important for the programmers is not the positioning of insertion of the additional field but the establishing of proper relationship. Irrespective of where the additional field is added (physically positioned), programmers can easily establish proper relationships or use proper SQL statements to display (or use) the additional field in any report generated by the software, logically, and in proper places. All these mean that even though post-piracy fields can be seen anywhere in the table, there are greater chances of finding them at the end of the table[5]. This also means that any difference found among the ending fields of the respective tables of the original and the pirated (or any successfully-unmapped fields at the end of the pirated table) can possibly be due to post-piracy modifications and so, special analysis of ending fields (with extra effort to identify and filter out post-piracy modifications) can yield reliable forensic result. For instance, Table-2 contains two unmapped fields UsrCode and UsrEnteredOn. . These two successfully-unmapped fields (see explanation on table-2 above) appear at the end of the database table in the pirated and so, the presence of these two fields (in the pirated) are to be further analysed.

External evidence of post-piracy modifications can exist: In some cases, in order to prove that a particular difference found was caused by post-piracy modifications and that the pre-modified version of the pirated had greater similarity with the original, the expert may require external supporting evidence, such as official documents. Log books (or documents for the software modifications done) and government directives (or documents initiating modifications in the software) belonging to post-piracy period are pieces of potential evidence acceptable to the court and the dates appear in these documents can be taken as evidence for post-piracy modification. In addition to log books and government directives, any official document that carries the date on which a particular new facility (say, a new MIS report) has been put to use by the client of the pirate, may be of help to the cyber forensic expert to argue unequivocally that the difference in the pirated is attributable to a post-piracy modification.

Summary of the discussion: In short, just as similarities need not always indicate piracy (Authors, 2009, p.176), differences need not always indicate non-piracy either. If, by establishing a schematically tangible patterning or mapping, the expert can identify the differences as attributable to post-piracy development, particularly suggestive of having been 'contrived', then the whole software comparison process may require closer attention along several parameters in different ways before rejecting or confirming piracy.

Need for judiciary-friendly presentation of results: The results of the analysis, when presented in transparent tabular form (to the judge) might provide more convincingly effective forensic evidence. It is hoped that Table-3 below, which is derived from Authors (2009, p.180), but specifically fine-tuned for post-piracy modifications, provides a seminal illustration for such a tabular presentation of, for instance, a database piracy forensics result[6]. Finally, it is needless to say that the result of analysis should be presented by cyber forensic expert as his/her views, strictly in an un-interpretive manner, because the right to interpretation solely rests with the court (Slade, 2004).

Further scope of this research: Needless to say, questionable differences between the pirated and the original can be found not only in data bases but also in other parts of the pirated, and for each part, along a variety of parameters. Some of the identifiable parts of software are source codes, databases, embedded images, fingerprints and so on. Again, one part can encode differences along several parameters. In case of source code, for instance, questionable differences can occur along parameters

[5] A statistical study to enumerate this chance is beyond the scope of this article.

[6] The test used in this example is "Substantial similarity" test (Davis, 1992). For other tests / approaches (for example, AFC), similar judiciary-friendly reports need to be arrived at.

like program variables, loop variables, names of functions, procedure calls, algorithms and so on. In the case of database, some of the parameters are field name, field type, field length and so on. In any case, post-piracy modifications along these parameters can make the software piracy investigation, delicate and demanding. This offers further scope in this research.

Table -3: The proposed format for presenting the result of comparison of two database tables (post-piracy modifications are also considered)

i. **Similarity in the table names**: ___% commonality.

ii. **Length of the ‘original’ table**: a

iii. **Length of the ‘pirated’ table**: b

iv. **Percentage of similarity in lengths**: (a/b)*100

v. **Field count of the ‘original’**: c

vi. **Field count of the ‘pirated’**: d

vii. **Percentage of similarity in field count**: (c/d)*100

viii. **Perfect commonality in the names of fields**: __ out of __ names of fields in the ‘original’ are found in ‘pirated’ also. So, ___% commonality

ix. **Perfect commonality in name and data type among fields**: __ out of ___ fields have the common name and data types. So, ___% commonality in name and data type.

x. **Perfect commonality in name, data type and length among fields**: ___out of ___ fields have same names, data types and length. So, ___% commonality

xi. **Perfect commonality in name, data type, length and the default values set in the fields**: ___out of ___ fields have same names, data types, length and default values. So, ___% commonality.

xii. **Perfect commonality in sequence of the fields with same name**: ___out of ___ fields with same name, do occur in the same sequence. So, ___% commonality.

xiii. **Perfect commonality in sequence of the fields with same name, data type, length and default values**: ___out of ___ fields (with same name, data type, length and default values) do occur in the same sequence. So, ___% commonality.

xiv. **Count of comparable (mappable) fields including suspected-post-piracy modified / created fields**: __ out of __ fields in the ‘pirated’ can be perfectly or approximately mapped (in terms of names) to at least one field in the ‘original’. So, _____ % comparable fields in the ‘pirated’.

xv. **Count of non-mappable but suspected-post-piracy fields, including ending fields**: __ out of __ fields in the ‘pirated’ could not be properly mapped to any of the fields in the ‘original’ but they can be suspected to be post-piracy modifications. So, _____ % incomparable but suspected fields in the ‘pirated’

xvi. **Count of non-mappable, non-suspected fields**: __ out of __ fields in the ‘pirated’ could not be properly mapped to any of the fields in the ‘original’ and do not provide any clue to be suspected as post piracy modification. So, _____ % incomparable, non-suspected fields in the ‘pirated’

xvii. **Inference:** Piracy is confirmed / largely suspected / loosely suspected / not suspected.

7. CONCLUSION

To sum up, one can conclude that observed surface differences between the original and pirated does not necessarily provide automatic grounds for exculpation from piracy in that many times much of the observed differences could be a direct result of post-piracy modifications both in the original and the pirated. In fact it is the differences that trigger the need for careful observation since they may be the result of post-piracy modifications. Such modifications may be motivated by a number of factors. A proper study of post-piracy modification, using the most appropriate tools both automatic and manual especially in the data base schemas of the pirated will unearth the differences that are invariantly attributable to post-piracy modifications, and thus will contribute substantially to the reliability of cyber forensic investigation. Some of elements discussed in this paper, like the positioning of the post-piracy fields and dates of post-piracy modification, are often incorrectly discounted as not too reliable; but under clever and careful handling, they can provide valuable supporting evidence. Ideally, in the interests of justice, a technical expert should be able to identify and put to use some or all the techniques of identifying post-piracy modifications to supplement the digital evidence (put forward by the automated tools, established judiciary approaches etc.) and other physical evidence.

8. REFERENCES

Authors, (2009), Software Piracy Forensics: Exploiting Nonautomated and Judiciary-Friendly Technique, Journal of Digital Forensic Practice, 2:4, 177 — 179

Author., (2002) Expert Commissioner Report submitted to the honourable court of Judicial I class magistrate, Kozhikode, Kerala, India, case number CMP 10371 / 2002, Software Associates vs. Together Infotech

Author, (2007), Seizure report submitted to the honourable District Court, Kozhikode, Kerala, India, on case number OS 2/2007, Sesame Software Solutions Vs. Perfect Software Solutions, p.2

Author, (2009), Judiciary-friendly computer forensics, Kerala Law Times, India, Part 13 & Index, 29th June, 2009, p.54

CDAC, (2008) Software Analysis Report number CDAC/RCCF/2007-20AR/Jan/2008 of the Resource Centre for Cyber Forensics, Centre for Development of Advanced Computing (CDAC), Government of India, Thiruvananthapuram – 695 033, Kerala, India, on the suit number OS 2/2007, Sesame Software Solutions Vs. Perfect Software Solutions, in the honourable District Court, Kozhikode, Kerala, India

Chang-Tsun Li, (2010) Handbook of Research on Computational Forensics, Digital Crime, and Investigation: Methods and Solutions, chapter XX, Information Science Reference, www.info-sci-ref.com,

ESALab (2007), The “SIMILE Workshop”: Automating the detection of counterfeit software, available at www.esalab.com,

Frantzeskou, G., Stamatatos, E., Gritzalis, S., Chaski, C. E., and Howald, B. S., (2007) Identifying Authorship by Byte-Level N-Grams: The Source Code Author Profile (SCAP) Method, International Journal of Digital Evidence, 6, 1,

Hollaar, L. A., (2002), Legal Protection of Digital Information, BNA Books

Lancaster, T., and Culwin, F., (2004) A Comparison of Source Code Plagiarism Detection Engines, Computer Science Education, from http://www.informaworld.com/

Myles, G., (2006), Software Theft Detection through Program Identification, Ph. D. thesis, University of Arizona, Department of Computer Science, available at http://sandmark.cs.arizona.edu/ginger_pubs_talks/defense_3_06.pdf

Reyes, A., (2007), Cyber Crime Investigations: Bridging The Gaps Between Security Professionals, Law Enforcement, and Prosecutors, Massachusetts, Syngress Publishing, Inc.

Slade, R. M., (2004), Software Forensics: Collecting Evidence From The Scene Of A Digital Crime, New York, The McGraw-Hill Companies, Inc.

van der Ejik P. (1994), Comparative Discourse Analysis of Parallel texts, eprint arXiv:cmp-lg/9407022, Digital Equipment Corporation, Ratelaar 38, 3434 EW, Nieuwegein, The Netherlands, CMP-lg/ 9407022,

Walker, J., (1996), Protectable 'Nuggests': Drawing the Line Between Idea and Expression in computer Program Copyright Protection, 44, Journal of the Copyright Society of USA, Vol 44, Issue 79

Yamamoto, T., Matsushita, M., Kamiya, T., and Inoue, K., (2004) Measuring Similarity of Large Software Systems Based on Source Code Correspondence, IEEE Transactions on Software Engineering, XX, Y (Proposed Draft found on http://www.google.com/url?sa=t&source=web&cd=3&ved=0CCwQFjAC&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.19.7035%26rep%3Drep1%26type%3Dpdf&rct=j&q=Measuring%20Similarity%20of%20Large%20Software%20Systems%20Based%20on%20Source%20Code%20Correspondence&ei=1SbTTZ-fEoLL0AGBjMn7Cw&usg=AFQjCNFc1Rip5lw3igaRe0o1yzI5pqDIlA&sig2=TMFVFzP6RfLRTbs4E42jSg&cad=rja

MAC OS X FORENSICS: PASSWORD DISCOVERY[1]

David Primeaux (dprimeau@vcu.edu),2
Robert Dahlberg (dahlbergra@vcu.edu),2
Kamnab Keo (kkeo@vcu.edu),3
Stephen Larson (larsonsp@vcu.edu),3
B. Pennell (pennellbd@vcu.edu),2
K. Sherman (shermankc@vcu.edu)2

Virginia Commonwealth University
401 West Main Street
P.O. Box 843019
Richmond, Virginia 23284-3019

ABSTRACT

OS X provides a password-rich environment in which passwords protect OS X resources and perhaps many other resources accessed through OS X. Every password an investigator discovers in an OS X environment has the potential for use in discovering other such passwords, and any discovered passwords may also be useful in other aspects of an investigation, not directly related to the OS X environment. This research advises the use of multiple attack vectors in approaching the password problem in an OS X system, including the more generally applicable non-OS X-specific techniques such as social engineering or well-known password cracking techniques such as *John the Ripper* or other versions of dictionary attacks and Rainbow table attacks. In some successful approaches the components of the attack vector will use more OS X specific techniques such as those described here: application-provided password revealing functions, a Javascript attack, an "Evil Website" attack, system file scavenging, exploitation of the keychain, and an OS X install disk attack.

Keywords: OS X, password, password discovery, social engineering, sleepimage, keychain,

1. BACKGROUND

Passwords are of forensic value because while they may be helpful in protecting the interests of computing users who have benign intent, passwords can also obstruct a forensic investigation. A solution to this *password problem*, from the point of view of the forensic investigator, is the discovery of the password (or set of passwords) obstructing an investigation. Password discovery involves locating the password and, as necessary, decrypting that password.

There is no simple, direct solution to all instances of the password problem in OS X. However, in a forensic investigation of an OS X system, the investigator may benefit from the facts that he or she can make use of standard (that is, non-OS X-specific) password attack techniques, and that the typical user may have little or no knowledge of, or control over, the location of some passwords used in the system. Furthermore, when the investigator cannot readily locate a potentially useful password, or when such a password is located but encrypted or obscured (often by means of a password hash function), the investigator may benefit from understanding typical human behaviors that often affect

[1] This research is partially funded by ManTech.

[2] School of Engineering, Computer Science Department.

[3] School of Business, Information Systems Department.

[3] School of Business, Information Systems Department.

password use.

Independent research conducted by Sophos Labs revealed that 33% of the respondents used the same password for every website and 48% used a few different passwords for every web site [1]. Therefore, if an investigator is able to discover one password, it is highly likely that the investigator will see this password again, re-employed by the user for some other purpose. Another survey conducted by Sophos targeted 500 business PC users and revealed that 72% of the respondents used weak (easily discovered) passwords. The respondents had a tendency to use passwords such as their girlfriend's name, favorite football team, or pet's name [2].

Some salient information (such as a girlfriend's name) related to passwords formed in this manner may be known through other aspects of an investigation such as the questioning of witnesses, or through use of known social engineering techniques; and, with increasing likelihood, the user may have published this information on public social networking sites such as *Facebook* and *MySpace*.

We also expect that several passwords for the same user will show a tendency to have similar characteristics with regard to meaning-to-the-user, length and complexity. Knowing this type of behavior is useful because it gives the investigator insight into trends related to the user's password management. For example, if the investigator discovers two or three passwords that are derived from animal names, when mounting a dictionary attack on an encrypted password that investigator may benefit from loading a dictionary list that contained variations of animal names.

Additional information is available regarding the impact of human and social behaviors on password formation. Recently, for example, phpbb.com was hacked and the passwords of 20,000 users were published. Robert Graham wrote an application to analyze these passwords and found the following trends [3]:

16% of passwords matched a person's first name.
14% of passwords were patterns on the keyboard, such as 1234 or qwerty or asdf
4% were variations of the word password such as passw0rd or password1
5% were passwords referenced to pop culture, such as pokemon, ironman
4% appear to be references of things nearby, such as Samsung, Packard or apple
3% were swear words, the F-word was very popular
3% were "don't care" words like whatever or blahblah
1% were sports related, team names or sports

Table 1 gives the rate of use for the top 20 passwords for these 20,000 users.

Percentage of use followed by password			
3.03% "123456"	0.59% "12345678"	0.36% "trustno1"	0.30% "hello"
2.13% "password"	0.58% "letmein"	0.33% "dragon"	0.30% "monkey"
1.45% "phpbb"	0.53% "1234"	0.31% "abc123"	0.28% "master"
0.91% "qwerty"	0.50% "test"	0.31% "123456789"	0.22% "killer"
0.82% "12345"	0.43% "123"	0.31% "111111"	0.22% "123123"

Table 1. Percentage of shared password use in 20,000 phpbb.com users.

Note that for each password shown, between 660 users (3.03%) and 44 users (.22%) used the same password.

The passwords shown in Table 1 may demonstrate the potential value of a social engineering perspective in the search for user passwords. Using a social engineering perspective, the investigator would take into account a particular work environment or hobby or other sort of personal interest known to be associated with the subject of the investigation. In the case of these users, for example, some of the commonly used passwords may be indicative of the sorts of users attracted to phpbb.com. Among such passwords are: “trustno1”, “letmein”, “dragon”, “master”, and “killer”. For other sorts of investigation, on the other hand, it may be interesting to use a “reverse” social engineering attack: if the passwords can be hacked for a given population of users, some passwords may directly or indirectly reveal information useful to the investigation.

A social engineering password attack may not provide a directly usable password, but may still be helpful to the investigator with access to some other means of attack. In this regard, information stored in a user’s browser history or bookmarks may be helpful as a means of advancing a social engineering attack, as may discoveries related to the user’s behaviors on his or her frequently visited websites. For instance, if a user has visited a gaming site and has a character on that site – the name of that character may provide some insight into the patterns the user has employed in constructing passwords. For example, if a social engineering attack were to reveal that the user often uses a pattern of password structure that contains, in part, the name of an animal (*lionxyz*, for example, or *yrtmonkey*), this knowledge may prove useful in decreasing the time required to mount a successful dictionary attack, described below.

In some investigations, it is possible that known secure applications or remote systems, accessed by the user, have password strength restrictions such as a requirement that a password contain a combination of characters and numbers. An investigator mounting a social engineering attack in this environment should look for numbers that may be significant to the user, such as birthdays, graduation years, anniversaries, and so forth. Or a password constraint may require the password be changed every 30 days. In these cases, it is likely that the user’s password will follow some pattern such as some combination of a base structure modified by some addition reflecting a month (for example, Betty01 for January’s password, Betty02 for February’s password).

A dictionary attack is a non-OS X-specific password discovery method applicable, as are other standard techniques to OS X password discovery. A dictionary attack is a technique for defeating a cipher or authentication mechanism by trying to determine its decryption key or passphrase by searching likely possibilities. A dictionary attack uses a modified brute-force technique of successively trying all the strings in an extensive list of strings. We say the technique is modified brute-force, because a true brute force attack (not further discussed in this paper) would search all possible combinations of characters, whereas a dictionary attack typically limits its list to variations of strings that have some meaning. Such attacks often succeed because many people have a tendency to choose passwords that are short, single words found in a dictionary or simple words appended with a numerical character. In the example provided in the paragraph above, the examiner might benefit from trying a dictionary loaded with strings that represent variations of animal names.

Another standard password-related attack is a Rainbow table attack. A Rainbow table is a lookup table that may sometimes prove useful to the investigator trying to recover a cleartext password from a password hash. A hash is a mathematical technique that takes one input and maps it to another in such a way that it is often very difficult to determine the initial input from the result of applying the hash function. A password hash is the result of inputting a password into a hash function. As we would expect, for security reasons, OS X stores only the password hash (and not the cleartext password itself). We would also expect the selection of an appropriate hash function so that determining the password itself from the password hash would be challenging. The Rainbow attack is similar to a dictionary attack, except the list it uses contains the hashes (or portions of the hashes) generated by introducing possible password into a known hash function, such as OS X’s password hash function. If the resulting hash is found to match a password hash, the Rainbow attack can use a simple lookup table to find the password that generated the hash.

2. PASSWORDS

In this section, we outline what can be password protected in OS X, and what are its possible locations and forms for stored passwords.

In addition to providing a means by which a user can password protect individual files and folders, OS X employs password protection (or permits the use of password protection) within several contexts including user login, privileged access as root, *FileVault*, keychain (a means of creating and managing passwords, typically accessed through the *Keychain Access* utility), networking and webpage logins, and other applications.

Password-related information is found in locations associated with utilities such as *Keychain Access* and with applications such as the Firefox browser, as well as in the OS X user-specific password hash file /var/db/shadow/hash (the user-specificity of the file in OS X contrasts with the more usual situation in other UNIX-like or UNIX-related systems in which a single such file holds password hashes for all users). Generally, the actual location of a password is controlled by the operating system, by an operating system related utility, or by a third party application. However, it may be useful for the investigator to be aware that requirements for construction of more complex passwords increases the likelihood of the user having recorded clear text copies of passwords (or of mnemonics associated with them); these user copies of passwords may be stored on the system being investigated (for example on a *Stickies* note), or elsewhere (such as a scrap of paper attached to the system monitor). In any case, as will be discussed below, some passwords in an OS X system will be stored in encrypted form, while others may be stored in a non-encrypted ("clear text") form.

2.1 Additional technologies to assist in OS X password discovery

Given the password characteristics just described, a successful methodology for password discovery is likely both to employ multiple techniques and to leverage any resulting positive results. Several available technologies can be useful in the process of password discovery. These include password-cracking utilities (such as the well-known *John the Ripper*), the use of native password revealing functions in applications that store passwords, the OS X password reset tool, and java scripts. The generally applicable password cracking approaches that use techniques such as dictionary attacks and Rainbow Table attacks should be well known and are very briefly described in Appendix 5; other technologies, perhaps not so widely known, are discussed here. These include: application-provided password revealing functions, a Javascript attack, an "Evil Website" attack, system file scavenging, *Keychain Access* attack, and an OS X install disk attack.

Application-provided password revealing functions. Because human maintenance of passwords is becoming increasingly burdensome, it is expected that many users may elect to have passwords remembered for them. The *Keychain Access* utility in OS X has a *show password function* that has the capability of showing many system-remembered user passwords (and sometimes also relevant, of showing many user usernames). Although use of the password revealing function in *Keychain Access* does require access to the user's keychain password, if that password can be discovered it will permit the investigator to access to all passwords stored in *Keychain* (An "Always allow" authorization option existing in the password revealing function may give the impression that the user could unintentionally provide access to this function for all time; however, our tests show that this is not the case). *Keychain Access* is further discussed later in this section.

Passwords may be more readily available to the investigator through the password revealing function available in at least some versions of the Firefox browser (such as version 3.6). In order to have clear text access to any passwords the user has asked to be stored for browser use, the investigator need only select the following sequence of options once Firefox has started: <Firefox>, <Preferences>, <Security>, <Saved Passwords>, <Show Passwords>. The investigator will be asked whether to confirm the <Show Passwords> selection and will be permitted to do so without further authentication. A user may prevent the <Show Passwords> option from working as described, in the following manner. For increased security, Firefox permits the user to set up a master password within the <Security> option. If this is done, authentication with that master password will be required to use the <Show Passwords> function. We hypothesize that a high percentage of Firefox users are unaware of this security precaution and do not implement a master password for Firefox. This exploit, and the possible use of a master password to thwart it, is available in some form on at least some other Mozilla 5.0 derived browsers such as SeaMonkey 2.0.4.

Java script attack. While this Java script attack has been tested on FireFox, Internet Explorer, Safari, and Chrome, it has not been tested in all browsers; nor, of course, is it feasible to test its usefulness for every website. This attack reveals the password used as credentials in a website [4]. When this exploit is effective, pasting the following java script into the address bar of the webpage will cause the password to be revealed:

```
(javascript:(function(){var s,F,j,f,i; s = ""; F = document.forms; for(j=0;
j<F.length; ++j) { f = F[j]; for (i=0; i<f.length; ++i) { if
(f[i].type.toLowerCase() == "password") s += f[i].value + "\n"; } } if (s)
alert("Passwords in forms on this page:\n\n" + s); else alert("There are no
passwords in forms on this page.");})();”
```

Evil Website attack. This investigative attack is a modification of the “black hat” technique called the “Man in the Middle Attack.” The Evil Website attack is only outlined here, and requires these three elements:

- A forensic copy of the suspect’s hardrive that the investigator will use to boot a system and access webpages. For simplification we will call this the *suspect’s system*.
- The suspect’s system has passwords stored for browser use either through OS X keychain or through the browser’s built in password management tool.
- A web server hosting the Evil Website which is actually a dummy forensic website to which the suspect’s browser will be made to send passwords.

The basis of this attack is that a browser running on the suspect’s system will pass credentials, otherwise hidden from the investigator, in clear text to the Evil Website. The Evil Website web server will be used to capture the user’s username and password.

System file scavenging. As determined in related research [5], it is possible to force the content of *active* (and possibly of *inactive*) physical RAM to an OS X *sleepimage* file. Furthermore, information in both *active* and *inactive* virtual memory is stored in the OS X *swapfiles*. The investigator can use a standard hex editor on a forensic copy of these files to scavenge for passwords, either encrypted or in clear text. Our preliminary tests show, for example, that the user system password can be found in clear text in the *sleepimage* file. This specific fact may, however, simply be an artifact of undesirable memory management, but more exploration of this and related issues is required. Nevertheless, this result supports the notion that passwords will be stored at least occasionally in the OS X *sleepimage* and *swapfiles.*

Keychain Access attack. The forensic investigator should become very familiar with the OS X *Keychain Access* utility. Many password-requiring applications that run in OS X are designed to be "keychain aware." This means that such applications can use *Keychain Access* to manage their credentials. As mentioned above in discussion of password revealing functions, *Keychain Access* does require the user to authenticate in order to reveal in plain text stored passwords. But by default OS X sets the user's required *Keychain Access* password to his or her system password. This design decision may be a direct reflection of Apple's desire to have Macs and the OS X operating system perceived as easy to use and requiring little user intervention to perform daily computing task. But, as seen here, it also can be a cause of concern with respect to their security, especially given the power provided by knowledge of the *Keychain Access* password; the forensic investigator should seek to exploit that security concern to the fullest extent possible. In particular, if the investigator can acquire a suspect's system password, it is highly likely that he or she will be able to use this utility will unlock many additional doors.

As expected, the OS X user does have the more secure option of changing the *Keychain Access* password but seems to be discouraged from doing so by information provided through the *Keychain Access* <Help>:

> You can change the password for your keychain at any time. However, if you want your default keychain to be unlocked automatically when you log in, make sure your keychain password is the same as your Mac OS X login password for your account.
>
> If your Mac OS X login password is not the same as your default keychain password, you'll be asked for the password whenever an application needs access to your keychain and your keychain is locked [6].

For a number of reasons, including: the user not wanting to be asked for a password each time a "keychain aware" applications needs access to the keychain; the default setting for the *Keychain Access* password as the user's system password; the perception that Apple provides a more secure computing environment than Windows does; and the expected relative obscurity of *Keychain Access* to many users --- we are confident that many OS X users will keep their *Keychain Access* password the same as their login password.

OS X install disk attack. The OS X install DVD comes with a built in utility that permits a user to reset nearly any password on the system. Resetting a user's OS X password will not reset the user's *Keychain Access* password or FileVault password. At this time we have not discovered an alternative means of resetting the *Keychain Access* password except through direct use of the *Keychain Access* utility itself, or an alternative means of resetting the FileVault password except through direct use of the OS X *System Preferences* utility. Future research will be required to determine whether some other method, such as the use of a command line interface like that provided by the OS X security to directly access the keychain, might be useful in further exploiting the keychain or breaking into a FileVault-protected system.

In OS X, "root", as in other UNIX-like or UNIX-related systems, is the name the most highly privileged system user, capable of reading, writing, deleting, moving, or otherwise accessing any file or folder in any account on the system. This makes "running as root" a particularly dangerous way to work within such a system. Because of the dangers of running as root, OS X makes access to root privileges a little less obvious than do some other UNIX-like or UNIX-related systems. In particular, by default, the root user is not enabled in OS X. However, by using the OS X install DVD the investigator can enable the root user and/or change the root user's password. This may prove significant in an investigation because it may also lead to information about the password for a suspect's account: in OS X, when root attempts to change another account's password, the password hint associated with that account is revealed. This can give the investigator insight as to what the account's password might be. If the password hint should be something like "high school attended" or "favorite pet's name," then it might be relatively easy to determine that information.

3. CONCLUSIONS AND FUTURE RESEARCH

OS X provides a password-rich environment in which passwords protect OS X resources and perhaps many other resources accessed through OS X. Every password an investigator discovers in an OS X environment has the potential for use in discovering other such passwords. Additionally, the investigator will be aware that these discovered passwords may also be useful in other aspects of an investigation, not directly related to the OS X environment.

There is no direct, unique solution to the password problem in OS X. An effective approach to the problem may require the use of multiple attack vectors. In some successful approaches the components of the attack vector will include the more generally applicable techniques such as social engineering or well-known password cracking techniques such as *John the Ripper* or other versions of dictionary attacks and Rainbow table attacks. And in some successful approaches the components of the attack vector will use more OS X specific techniques. These include: application-provided password revealing functions, a Javascript attack, an "Evil Website" attack, system file scavenging, exploitation of the keychain, and an OS X install disk attack

Our future password discovery research will focus in three directions. We will explore the command line OS X security interface to determine the extent to which this interface can used to provide additional information about the keychain, as well as to provide methods of manipulating the keychain with the potential of permit additional system access. We will also continue our exploration of the OS X *swapfiles* and *sleepimage* file with the specific goals of determining which if any passwords may be routinely stored in these files and whether the location of any such passwords can be predicted. Finally, we intend to explore the impact of enabling Secure Virtual Memory on OS X password discovery.

REFERENCES

1. Cluley, Graham (2010) Graham Cluley's Blog : "Do you use the same password for every website?", http://www.sophos.com/blogs/gc/g/2009/03/10/password-website/, accessed 26 March 2010.
2. Sophos (2010) "Employee password choices put business data at risk, Sophos poll reveals." http://www.sophos.com/pressoffice/news/articles/2006/04/passpoll06.html, accessed 26 March 2010.
3. Graham, Robert (2009) "PHPBB Password Analyses", http://www.darkreading.com/blog/archives/2009/02/phpbb_password.html, accessed 26 March 2010.
4. Davis, Sam (2008) "Reveal a Password in Internet Explorer or Firefox", http://www.blastedthing.com/misc/mag-reveal-a-password-in-internet-explorer-or-firefox/ , accessed 19 April 2010.
5. Primeaux, D.; Dahlberg, R.; Keo, K., Larson, S.; Pennell, B.; and Sherman, K. (2010) "MAC OS X Forensics: Volatile Memory Acquisition." Technical Report VCU-CISS-TR-10-1, Virginia Commonwealth University, Computer Science Department.
6. Apple Inc. (2010) "Mac OS X 10.4 Help, Changing your keychain password" http://docs.info.apple.com/article.html?path=Mac/10.4/en/mh463.html, accessed 23 September 2010.

DEVELOPING A FORENSIC CONTINUOUS AUDIT MODEL

Grover S. Kearns, Ph.D., CPA, CFE (Contact Author)
Gregory, Sharer & Stuart Term Professor in Forensic Accounting
Phone: 727-553-4085
gkearns@stpt.usf.edu

Katherine J. Barker, CPA, CFE
Assistant Professor in Accounting

College of Business
140 7th Avenue South
University of South Florida St. Petersburg
St. Petersburg, FL 33701

ABSTRACT

Despite increased attention to internal controls and risk assessment, traditional audit approaches do not seem to be highly effective in uncovering the majority of frauds. Less than 20 percent of all occupational frauds are uncovered by auditors. Forensic accounting has recognized the need for automated approaches to fraud analysis yet research has not examined the benefits of forensic continuous auditing as a method to detect and deter corporate fraud. The purpose of this paper is to show how such an approach is possible. A model is presented that supports the acceptance of forensic continuous auditing by auditors and management as an effective tool to support the audit function, meet management's regulatory objectives, and to combat fraud. An approach to developing such a system is presented.

1. INTRODUCTION

Over the past decade, businesses have faced increased regulatory oversight and reporting requirements combined with global competition and increased costs of raw materials and labor. As a result, management seeks an efficient but effective approach to governance which satisfies compliance requirements but also protects the organization from fraud at an affordable cost.

With organizations routinely processing terabytes of information daily achieving important audit objectives has become a daunting task. Traditional audit approaches and sampling methods cannot be expected to uncover the majority of transactional errors or occupational fraud (Wells, 2011; Oringel and Aldhizer, 2009). Technology offers opportunities to detect and deter fraud more efficiently and effectively. Statement on Audit Standards No. 99 (SAS 99), *Consideration of Fraud in a Financial Statement Audit*, codifies many fraud detection procedures and encourages their use by auditors to detect client fraud risk and identify transactions to be tested (AICPA 2002, AU 316.52, AU 316.61; Lanza and Gilbert, 2007). Technological skills, however, often exceed the competency of auditors causing them to resort to less effective manual approaches.

The regulation that has had the most profound impact on management and auditors in the past decade, the Sarbanes-Oxley Act of 2002 (SOX02), requires that CEOs and CFOs assess and attest to the effectiveness of the organization's internal control structure. It also imposes increased penalties for financial statement fraud. Both SOX02 and SAS 99 encourage management and external auditors to employ technological approaches and embedded audit modules to audit financial transactions and internal controls (Roth and Espersen, 2003). SOX02 Section 409 accelerates the SEC filings for Form 10-Q and annual report Form 10-K. The new rules will eventually require public companies to file annual reports within sixty days of their year-end and quarterly reports within thirty-five days of the end of the quarter. The FTC's red flag rules, effective December 31, 2010 for financial institutions and certain other firms under FTC jurisdiction including CPA firms, require companies to check for and report specific violations. These rules are expected to increase compliance costs. Automating the audit process will enhance the company's ability to comply with these reporting requirements and lower overall governance costs. Although increased regulatory pressure mandates more attention to internal controls, these pressures could actually increase fraud opportunities by overwhelming management and auditors with reporting requirements.

Despite increased attention to internal controls and risk assessment, traditional audit approaches lack effectiveness in uncovering occupational fraud. In its *2010 Report to the Nations*, the ACFE noted that most of the frauds were uncovered by anonymous tips and less than 20 percent are uncovered by either internal or external auditors. This is partly because external auditors focus on the organization's financial statements only once a year and most auditing concentrates on small sample sets of selected transactions over fixed periods of time. A more effective approach would be to audit all or a large part of the transactions continuously.

Continuous auditing, which has been the focus of much research and has notable successful implementations (Alles and Vasarhelyi, 2008), still eludes many companies (Alles et al., 2008). The major barriers are technical – the lack of embedded audit modules (EAMs) and auditor's lack of the requisite technical skills (Li et al., 2007). Once operable, however, continuous auditing systems require less technical expertise and offer auditors a wealth of information that can increase audit quality while reducing the overall workload.

Forensic accountants have recognized the need for automated approaches to fraud analysis yet research has not examined the benefits of continuous auditing as a method to detect and deter corporate fraud. The purpose of this paper is to show how such an approach is possible. Contributions are twofold. First, cogent arguments are presented, in the form of five propositions that support the necessity for a system of forensic continuous auditing. Second, the paper presents an approach to forensic continuous auditing that is scalable and can be phased-in to accommodate the needs of management, auditing and information technology.

2. DEVELOPING THE FORENSIC CONTINUOUS AUDIT MODEL

2.1 Impact of Regulation

Management concerns about fraud have been heightened in the post-SOX02 environment due to increased penalties for financial statement fraud and governance requirements for a costly internal control framework. Requirements for auditors have increased dramatically and are costly. Compliance with SOX02 Sec. 404, that requires management to evaluate and attest to the internal control structure within ninety days of the audit report date, cost Fortune 100 companies about $7.8 million in 2005 of which audit fees were $1.9 million (Nondorf et al. , 2011).

Public Company Accounting Oversight Board (PCAOB) Auditing Statement 2, *An Audit of Internal Control over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*, states that it is management's responsibility to design and implement a program of controls to prevent, detect and deter fraud.

According to the Association of Certified Fraud Examiners (ACFE), estimated fraud losses in the United States for 2008 were $994 billion. The highly publicized frauds of the past decade have led to increased emphasis on internal controls. Adoption of the *Committee of Sponsoring Organizations* (COSO) framework and Statement on Auditing Standard No. 78, *Consideration of Internal Control in a Financial Statement Audit*, place greater demands on external auditors. The more detailed information technology (IT) controls, such as those found in the *Control Objectives for Information Technology* (COBIT) framework, have made IT audits standard for larger companies. Lack of technical expertise to conduct such audits has caused many audit firms to seek out and depend upon more expensive third-party support.

In SAS 99, the AICPA basically mirrored the tenets of SOX02 and increased the auditor's due diligence responsibility for recognition of fraud. It also recommended extended use of technology for substantive testing and audit of controls. Auditors recognize that traditional audit practices that rely heavily on sampling small sets of transactions on a limited basis are not sufficient for evaluating internal controls or for detecting and deterring fraud. Also, financial audits that are based primarily on substantive testing and neglect detailed analysis of transactions or auditing through the computer cannot provide high levels of assurance.

2.2 Auditing for Fraud

Traditional audit techniques are not sufficient and do not provide continuous assurance. Nor are they likely to uncover the most risky frauds – those perpetrated by managers who can override controls and alter ledger and journal entries. In order to audit *through* the computer, a process is necessary that allows for testing of a significant number of transactions on a real-time basis and throughout the year rather than brief discrete intervals. The process should focus on areas of high risk, areas of concern by key stakeholders, and risks that are significant – those that may be unlikely but where an adverse incident could threaten the life of the enterprise. Through control frameworks such as COSO and COBIT, companies monitor and assess activities to detect incidents of errors, misuse and fraud and respond in a timely manner.

To determine the likelihood that financial statements contain material misstatements, auditors conduct tests of transactions and substantive tests. Tests of transactions determine whether erroneous or falsified data have been processed. Substantive tests examine balances such as accounts receivable and accounts payable, inventories, liabilities and depreciation to provide assurance that financial statements are free from material misstatements (Rezaee et. al, 2001). Normally, if tests of transactions do not reveal irregularities then less reliance is required on substantive testing. However, if tests of transactions reveal abnormalities then substantive testing must be expanded. In a continuous auditing environment, tests of transactions is an ongoing process and evidence is collected on a larger set of transactions and over a wider time-frame that with traditional methods. This lessens the need for substantive testing and reduces the role of the external auditors resulting in savings for the client firm.

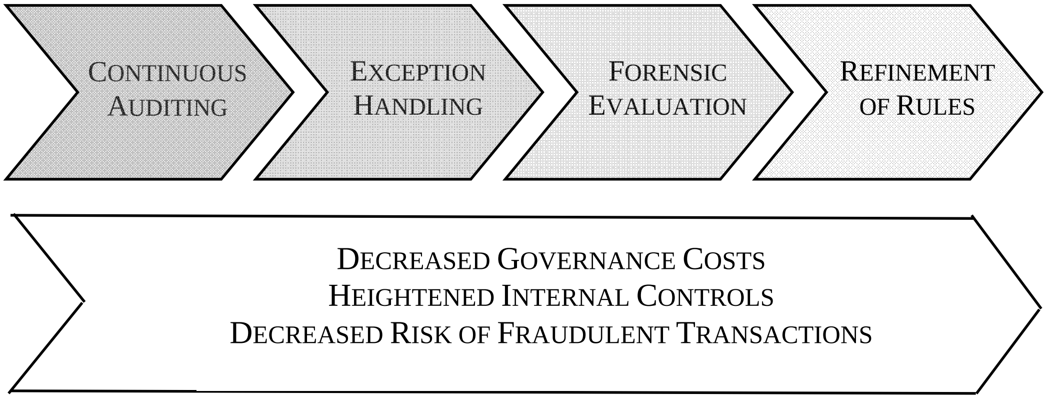
As a result of new regulatory requirements for compliance and emphasis on IT governance, auditors with forensic IT skills have been in increased demand (Hoffman, 2004). Because IT control deficiencies lead to accounting and financial reporting errors (Alaali, Grant, and Miller, 2008), it is important 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). The Forensic Continuous Audit Model is shown in Figure 1. The first requirement is continuous auditing.

3. CONTINUOUS AUDITING

3.1 Advantages of Continuous Auditing

According to the American Institute of Certified Public Accountants (AICPA) and Canadian Institute of Chartered Accountants (CICA) "A continuous audit is a methodology that enables independent auditors to provide written assurance on a subject matter, for which an entity's management is responsible, using a series of auditors' reports issued virtually simultaneously with, or a short period of time after, the occurrence of events underlying the subject matter." (AICPA/CICA Research Study on Continuous Auditing, 1999).

Figure 1: Forensic Continuous Audit Model


Continuous Auditing
Exception Handling
Forensic Evaluation
Refinement of Rules
Decreased Governance Costs
Heightened Internal Controls
Decreased Risk of Fraudulent Transactions

Because auditors often lack technological skills, a large percentage of companies rely primarily upon manual methods to evaluate internal controls. Consequently, these companies cannot determine how effective their control processes are on a daily basis despite large investments in governance (KPMG, 2010). In a 2009 survey by the Institute of Internal Auditors, only 32 percent of 305 companies reported that they performed continuous auditing. By providing for automatic analysis of transactions, continuous auditing would relieve the auditors of the burdensome strain and allow greater focus on the analysis of suspicious transactions.

3.2 Impact on Auditors and Governance

Continuous auditing offers several advantages for auditors. Because it tests more transactions over a wider time-frame, it provides more comprehensive and timely assurance. Also, it is scalable allowing the magnitude and timing of tests to be performed based upon the assessed risk of the targeted transactions. It can reduce the amount of substantive testing performed during financial audits and allow greater focus on more important investigative matters. It can reduce audit risk and increase management confidence in financial reports. It supports compliance reporting and reduces both errors and fraud. While continuous auditing assumes that all transactions are monitored in real-time, judicious application of the cost/benefit rule would schedule tests based upon the likelihood and severity of the risk. Performing the analytical procedures on a routine basis would lessen the work of the independent auditors and reduce their time on-site thus avoiding costly tests and unnecessary distractions during the workday. Continuous auditing can result in substantial savings by reducing the amount of external auditor fees (Hermanson et al., 2006). Thus, it reduces overall governance costs while reducing the opportunity for errors or fraud.

Tests of transactions using analytical procedures plus confirmation of account balances and events are the most common work product of financial audits. Confirmations may be either positive or negative. The negative confirmation is expected to be responded to only if the balance is not accurate. Research shows, however, that negative confirmations may not signify correctness as recipients may ignore them or they may be lost by mishandling (Aldhizer and Cashell, 2006; Caster and Sriram, 1996). When limited to small sample sets, tests of transactions may not be representative and cannot be expected to detect a large percent of errors or fraudulent activities. Given the increased transaction processing for most firms and increased regulatory pressures, the traditional approaches appear inadequate and require increased substantive testing.

KPMG's 2008 publication, *Continuous Auditing/Continuous Monitoring: Using Technology to Drive*

Value by Managing Risk and Improving Performance, comments that: "As business risks of all kinds continue to proliferate, management and internal audit departments are actively seeking new ways to quickly gain access to valuable information to manage risk and improve performance. Such efforts increasingly include continuous auditing and continuous monitoring of organizational processes, systems, and controls."

3.3 Forensic Continuous Auditing

Forensic continuous auditing (FCA) differs in the respect that more focus is placed upon the evaluation of sophisticated audit rules and examination of trends and anomalies that may reflect underlying errors or fraudulent commissions. FCA places more emphasis on the analysis of sensitive data sets and less emphasis on transactions for which detection risk is low. It also provides for a greater range of analysis and emphasizes improvement of the audit rules over time.

Regulatory standards encourage the use of computer assisted audit tools and techniques (CAATs) for accessing and analyzing data files and suggest that risk assessment reflect the client IT standards (AICPA 2001, 2006). Recent research, however, indicates that only a minority of firms use CAATs for substantive testing because of the high level of complexity (Janvrin et al., 2009). Continuous auditing can provide much of the substantive testing in a routine manner and allow auditors to concentrate on the forensic analysis of data.

3.4 Developing an Approach to Continuous Forensic Auditing

There are various approaches to continuous auditing. The embedded audit module (EAM) approach depends upon audit specific software that resides in the targeted application (Alles, 2002). It allows auditors to determine which transactions are to be tested and at what frequency. Results are collected and reported real-time. Enterprise resource planning (ERP) systems often contain EAM functionality (Groomer and Murthy, 1989). Surveys show, however, that companies that use enterprise resource planning (ERP) systems often do not activate the EAM because of the significant resource requirements which can slow overall processing dramatically (Kuhn and Sutton, 2010; Debreceny et al., 2005).

The technical nature of EAMs require that auditors acquire a higher levels of technical skills to implement these tools effectively and may hamper their adoption (Debreceny et al., 2005). Some researchers state that auditors cannot effectively administer continuous auditing because of low technical proficiency and inability to communicate with IT personnel (Li et al., 2007).

An alternative approach is the monitoring control layer (MCL) which uses an external software module linked to the target applications and databases (Vasarhelyi et al., 2004).

Ghosting allows the EAM or MCL to be used outside the production version of the application and avoid system performance problems. System ghosting creates a copy of an entire system on separate hardware and eliminates any risk associated with processing live transactions.

In the FCA Model, the second requirement is exception handling which applies the audit rules in order to uncover errors and suspicious transactions.

Table 1 presents the steps for developing a FCA system.

Table 1: Developing a Forensic Continuous Audit System

1. Examine internal controls for adequacy to mitigate risks.
2. Determine which risks are most likely or could cause the most harm to the organization. These risks should be continuously audited.
3. Examine each risk to determine the appropriate audit rules to be applied.
4. Examine each risk to determine the appropriate number of transactions to be tested – this will vary depending upon perceived risk and management objectives.
5. Examine each risk to determine the appropriate frequency of auditing – continuously, hourly, daily, weekly etc.
6. Identify target applications and databases for the associated transactions and events.
7. Establish a protocol for reviewing and handling the selected transactions.
8. Build the link between the CAAT and the data file to automate the continuous audit cycle. Ghost the application to an audit server.
9. Maintain an audit trail of the selected transactions and examine trends and anomalies.
10. Refine the audit rules making modifications based on experience.
11. Report results to management, the audit committee and external auditors.
12. Set alarms for suspicious transactions or events that require immediate action.

4. EXCEPTION HANDLING

4.1 Handling of Selected Transactions

Exception handling is critical to the efficacy of continuous auditing. By performing a large number of tests over a much higher percentage of transactions, continuous auditing expands the testing of details to a large percentage of the overall data and can reduce reliance upon analytical procedures (Alles et al., 2008). It will also result in a large number of selected transactions that have failed the audit tests. FCA takes the process one important step further: it adds analytical tools to examine the selected transactions for possible errors or acts of fraud.

Transactions that trigger exceptions or alarms must be responded to in a timely manner by qualified individuals with forensic knowledge and skills. Exceptions could be handled by internal auditing. Hermanson et al. (2006) suggests that software be coded to categorize incidents (selected transactions or events) into three categories: errors, misuse, and fraud. By responding to errors immediately, the source department may be able to take corrective action that eliminates future errors. System misuse could lead to increased employee training and awareness. It could also indicate the need for adjusting policies.

Selected transactions that require the most scrutiny and careful response are those that indicate the possibility of fraud. In this case, auditors must rely upon established protocols for response. In any event, managers should be alerted and action taken to prevent or isolate any further occurrence of the event. As Smith (2005) points out, as the time lag increases between the suspicion of fraud and the recovery of forensic data, evidence becomes less valuable. Larger companies may have an incidence response team. If so, they will probably require an analysis of the situation that could be performed by the internal auditors. Internal auditors should play an important role in fraud investigation. Using the FCA process, data sets can be examined to uncover and document fraudulent commissions. The FCA process may be the first line of defense in proactively identifying possible fraud.

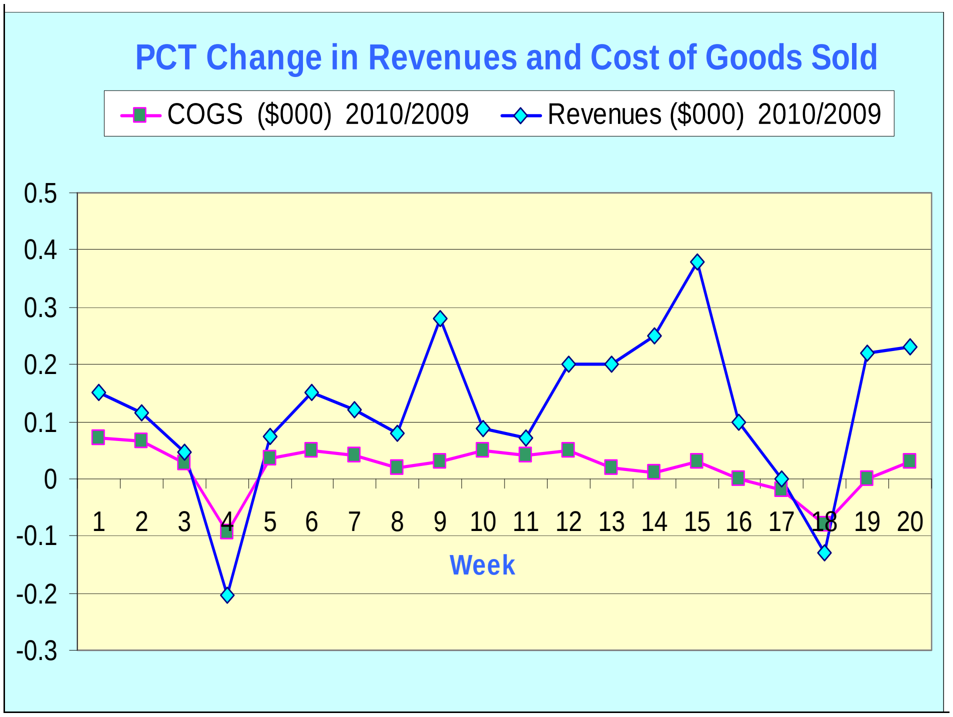
Auditors may also play an investigative role in the development and maintenance of forensic evidence.

This might require the auditor to perform read-only searches, preserve time-stamps, secure data and maintain a proper chain of custody (Smith, 2005).

4.2 Audit Rules

Regulators and the public expect auditors to uncover fraud. Research, however, does not support the ability of either external or internal auditors to uncover significant amounts of fraud (Albrecht et al., 2001). Thus, auditors must be trained to seek out specific types of fraud when analyzing the selected transactions. Special attention should be given to revenue manipulation and income-increasing manipulation because these are the most frequently occurring items in financial statement fraud (Johnson and Ireland, 2007). Transactions that fail the audit rules or highlight anomalies would be selected for forensic evaluation by internal auditors. For example, in Figure 2, the relationship between revenues and cost-of-goods sold is tracked over time. Revenues may be expected to vary but the relationship between revenues and cost-of-goods should exhibit a low variance and remain fairly smooth. In Figure 2, anomalies are readily apparent. These discrepancies require investigation and could reveal a fraudulent misstatement of revenues.

Figure 2: Trending Revenues with Cost of Goods Sold



The third requirement of the FCA Model is forensic evaluation of the selected transactions to determine what actions should be taken.

5. FORENSIC EVALUATION

Selected transactions might also uncover control weaknesses. For example, monitoring of access rights might identify instances of employee attempting to access unauthorized files or incompatible sets of files. The event would allow supervisors to take immediate action and correct the problem and modify the control.

Continuous auditing and monitoring can be expected to increase the likelihood that all fraud, including financial statement fraud, is prevented or detected in a timely manner. A large percentage of transactions are investigated and some results are presented in a graphical format. Transactions that fail audit rules would be written to a selected transactions file. Forensic evaluation using extended analytical procedures applied to the selected transactions allows proper and timely scrutiny. The forensic evaluation should examine the relationships between financial data within a period and over periods to detect anomalies that require investigation.

Financial statement fraud can have a devastating impact on a firm's stock price causing shares to drop as much as 1,000 times the fraud amount (Albrecht et al., 2001). Financial statement fraud in the United States accounted for 68 percent of reported fraud losses in 2009 (ACFE, 2010).

FCA systems could monitor 100 percent of an organization's financial transactions and business activities in real-time. Automating the analysis and testing reduces the cost of SOX02 compliance and reduces the risk of loss beyond what could be expected of periodic testing of small transaction sets.

5.1 Analytical Tools

Although many firms have adopted ERP solutions and have access to embedded audit routines, there are valid arguments for examining a modified approach to FCA. Research has shown limited support for the use of embedded audit modules in ERP systems (Debreceny et. al, 2005). Firms may have multiple ERP systems and each would require auditors to master the internal EAM. EAMs that operate internally can cause significant reductions in performance. An alternative is to use generalized audit software and apply established audit rules to transaction files in order to uncover erroneous or fraudulent transactions.

Audit Command Language (ACL) and Interactive Data Extraction and Analysis (IDEA) are well known audit software that can be used for developing CAATs. Both can be learned without extensive training and have a high level of vendor support. ACL, for example, offers the ability to conduct continuous auditing over several ERP systems and other applications. In addition to supporting a large number of analytical functions these CAATs are capable of extracting data from a large number of file formats. Using ACL analytics such as Benford Analysis, one investigative audit by Forensic Strategic Solutions uncovered more than $70 million of fraudulent expenditures (ACL, 2011). Events such as these could be detected routinely using established fraud audit criteria to test transactions and controls. Controls could allow selected transactions to be further inspected using other computer tools that support forensic analysis. For example, Benford analysis uses a *z*-statistic to measure the probability that a group of data falls outside the expected distribution. For certain data sets, transactions for which the first digit was outside a *z*-statistic of 2.0 could be triggered for further examination. The trigger points could be adjusted based upon experience. Correlation and time-series analysis can also be used to detect errors and fraud in the selected transactions (Nigrini, 2006). The final stage of the FCA Model is refinement of the rules.

6. REFINEMENT OF RULES

6.1 Designing Audit Rules

Examples of possible audit tests are shown in Table 2. Such tests are commonly performed manually on smaller sets of transactions and at distinct time intervals. The audit rules can test for errors, fraud, and the strength and presence of internal controls while also performing some substantive tests. Results can be used to create compliance reports. Refining the rules will require the judgment of experienced internal auditors based upon the performance of the fraud audit model. Rules can be modified based upon perceived risks and management objectives.

Table 2: Examples of Fraud Audit Tests

Fraud Objective	Fraud Flag	Fraud Audit Tests
Fraudulent vendors	Vendor address P.O box, Vendor address matches employee address, Multiple vendor addresses	Check validity of vendor numbers, Check for P.O. boxes as addresses, Match vendor addresses to employee addresses, Flag large changes in vendor activity, Extract vendors having no tax ID no.
Ghost employees	Employees with same address	Check employee addresses for matches, Invalid Social Security numbers, Compare number of employees over years to insure changes match new minus terminated employees, Flag employees who have not used benefits
Unauthorized file access	Employees accessing unauthorized files or incompatible files	Compare log-ins to access rights and privileges
Inventory loss	Inventory adjustments	Flag all adjustments exceeding a set percentage, Check deliveries outside of regular hours, Check employee access to restricted areas during irregular hours
Vendor kickbacks to managers who order at high levels	Inventory levels exceed established peaks, Average inventories too high	Examine average inventory levels and high volume purchases, Establish an economic order quantity and require a signed override by inventory manager for larger amounts
Copying sensitive data files (intellectual property or personally identifiable information)	Employees accessing unauthorized files, Copy attempts on protected files	Compare copy attempts to rights and privileges
Financial statement fraud	Senior management making fraudulent entries	Flag all journal and ledger entries by executive management, Flag all entries that boost revenues over a certain percentage, Flag significant transactions with related party, Review sales recorded by Corporate Headquarters
Invalid earnings	Adjustments to estimates such as bad debt allowance, amortization of intangibles, insurance claims, etc.	Flag changes that exceed a set percent or ones made by executive management (should be made by lower level accountants)
Cash larceny	High differences between sales and cash receipts, High refunds, voids, A/R write-offs	Summarize information by employee and flag all large differences

Audit rules drive the analysis of transactions and events. These analytical criteria are created to flag transactions that violate policy or could indicate a fraudulent act. If the criteria are too stringent a large number of alarms (called alarm flooding) will be produced. To prevent the enormous number of false positives audit rules must be properly calibrated (Kuhn and Sutton, 2010; Alles et al., 2008).

If too loosely set the tests could fail to detect a large percentage of erroneous and fraudulent transactions. A major benefit of such a process is that the audit rules can be expanded and periodically evaluated for efficacy and adjusted based upon performance. Over time, experience-based adjustment of the audit rules can make them more efficient and effective. Fraud audit tests should be designed around objectives.

6.2 Forensic Analysis of Selected Transactions

Selected transactions can provide information to proactively detect impending frauds. By examining trends in certain data series, anomalies can be inspected for possible defalcations. Chen and Sennetti (2005) demonstrated seventeen financial and non-financial variables useful in predicting fraud. Most important were, relative to sales, lower research and development costs, lower marketing costs, and lower changes in free cash flows.

Special attention should be paid to financial statement fraud which is the most costly and often requires an override of internal controls. Financial statement fraud and earnings mismanagement can be detected through the judicious application of a set of quantitative and qualitative red flags (Grove and Cook, 2004). An overstatement of revenues would be a possible indicator (Johnson and Ireland, 2007). Two examples of quantitative red flags would be irrational ratio analysis of Gross Margin Index and Sales Growth Index in order to determine if they fell outside of the industry norm. Horizontal analysis of the ratios could also point out trends and anomalies. Two examples of qualitative red flags would be significant insider sell-off of shares and opaque financial reporting and disclosures designed to confuse and mislead investors (Grove and Cook, 2004).

Figures 3 and 4 illustrate the possible application of forensic analysis. In Figure 3, Benford Analysis is used to analyze employee expenses. If the distribution of first-digits follows Benford's Law, then the resulting z-statistic would be low. The auditor might have a rule such as: do not investigate unless a digit has a *z*-statistic greater than 2. Such a rule can easily be altered over time.

Figure 3: Applying Benford Analysis to Employee Expenses

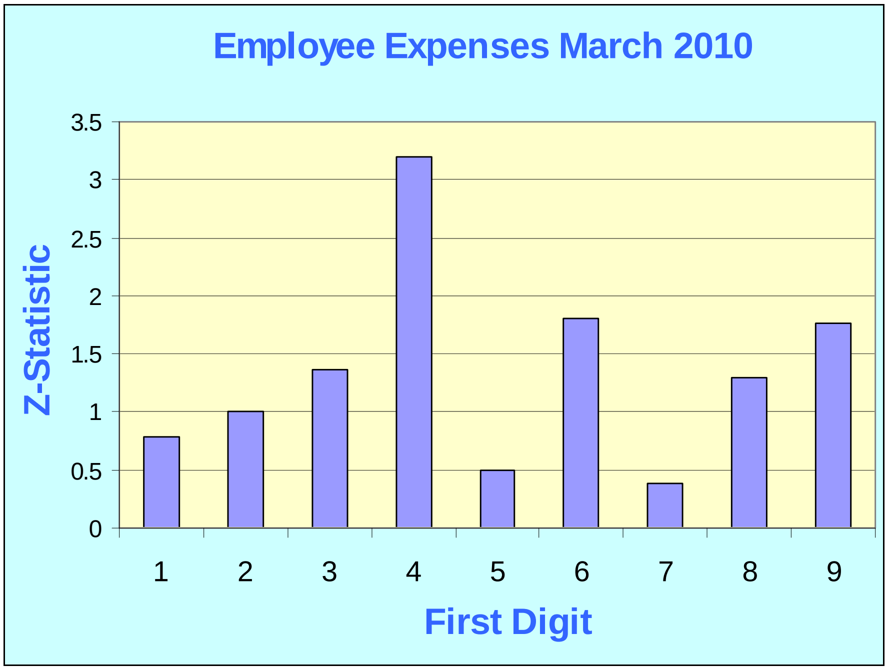
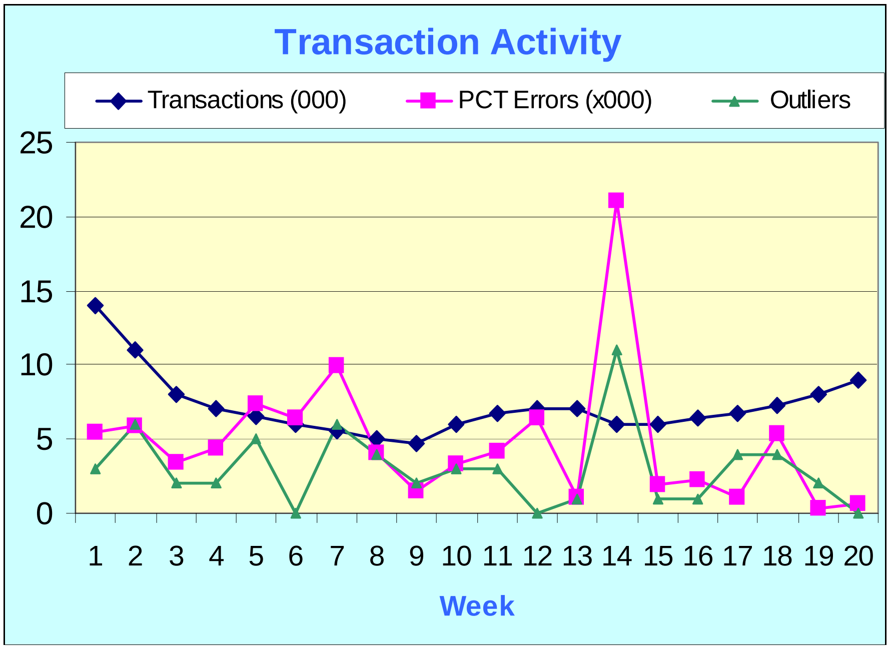


Figure 4: History of Transactions with Error Rate and Outliers



In Figure 4, the number of transactions is compared to the percentage of known errors and outliers (for example, values that exceed the average by greater than 3 standard errors). Again, anomalies become quickly apparent allowing the auditor to focus the investigation on areas that are most likely to indicate a problem.

7. PROPOSITIONS

The following five propositions support the use of the FCA model as an effective method for deterring and detecting corporate fraud. They are rooted in practical realities that are likely to persist and place undue burdens on management, auditors and key stakeholders unless a technological solution is adopted.

With continuous auditing, auditors can design audit rules that test a large set of transactions (perhaps 100%) at determined time intervals. With FCA, the rules can test for errors, fraud, and the strength and presence of internal controls, while also performing some substantive tests. Results can be used to create compliance reports. Over time, experience-based adjustment of the audit rules can make them more efficient and effective. Anomalies and outliers can quickly indicate the presence of potential problems. Thus,

Proposition 1: Forensic continuous auditing will add efficiencies to the financial audit process.

Section 404 of SOX02 has elevated the need for extensive tests of IT internal controls that may require the expensive services of a third-party firm. Thus, the need for more comprehensive yet cost-effective approaches is recognized by external auditors. By allowing the client to perform extensive testing of controls through continuous auditing procedures, the external auditor can avoid expanding the time-consuming and expensive substantive testing. Regulations require that certain substantive tests be performed. Auditing Standard AU 319.80, 81 states that "regardless of the assessed level of control risk, the auditor should perform substantive tests for significant account balances and transaction classes.'' By having access to increased data sets and allowing the client's internal auditors to perform more of the transaction testing through continuous auditing, the external auditor will be able to focus on more important activities that are more likely to lower risk. Thus,

Proposition 2: External auditors will perceive forensic continuous auditing positively.

PCAOB 5, *An Audit of Internal Control over Financial Reporting that is Integrated with an Audit of Financial Statements,* has increased the reliance that external auditors can place on evidence generated by internal audit departments in an effort to reduce duplication of efforts and lower audit costs. FCA combined with CAATs are capable of monitoring internal controls for SOX02 compliance reporting and uncover areas of higher audit risk. As external auditors rely more on the internal audit and client's automated controls and governance testing mechanisms, less time will be required of external auditors or IT auditors. Additionally, fewer requests for ad hoc data sets will be made of the IT department. By having an established process in which audit rules can be increased and modified over time to improve the quality of the results, the internal auditors will play a higher role in the assurance process and be viewed more favorably by the audit committee and by management.

Proposition 3: Internal auditors will perceive forensic continuous auditing positively.

Management can be expected to view a system that continuously audits for fraud positively because it supports compliance in a cost effective manner. As mentioned above, it will allow more work to be subsumed by the internal auditors thus decreasing costs and the time external auditors are on the premises. Furthermore, the external auditors can access and inspect data sets and reports remotely, avoid travel expenses, and not have to import data because the documentation and proof of compliance will already exist.

Management might also take a human resources view towards forensic continuous auditing. SOX02 has made acquiring IT auditors even more difficult and the number of qualified individuals is relatively small. The number of accountants with a Certified Information Systems Auditor license is less than 50,000 globally and all companies and accounting firms compete for these individuals (Kuhn and Sutton, 2010). Reducing the necessity for IT auditors will place less strain on human resources and commensurate salary levels.

Finally, management will value the ability to phase-in FCA on an application-by-application basis and expand the number of audit tests over time. Thus,

Proposition 4: Management will perceive forensic continuous auditing positively.

SOX02 requires management to evaluate and attest to the effectiveness of an internal control system (Arrens et al., 2006). Under increased regulatory scrutiny and facing increased audit costs management will seek cost-effective approaches to the detection of transaction errors and fraudulent activities. Increased penalties for fraud and the low percentage of fraud that is uncovered by auditors will make continuous auditing attractive as a forensic tool. Fraud deterrence is recognized as an important management objective. To prevent fraud, it is imperative that internal controls be tested continuously and that audit rules are established to uncover fraudulent events. This can be accomplished by examining a large percentage of the transactions and system events.

SAS 56, *Analytical Procedures*, requires that auditors perform analytical procedures during the planning and final reporting stages of the audit (AICPA, 1988). Analytical reviews, however, may not be effective at detecting frauds. Even large embezzlements may not have a material effect on the earnings of a large corporation and may escape discovery during a regularly scheduled audit (Wells, 2011). FCA, however, provides the ability for auditors to perform a multitude of analytical procedures over all transactions and significantly increases the possibility that errors and suspicious transactions are flagged (Rezaee et al., 2002). Properly constructed systems could perform hundreds of different analytical tests on a large number of transactions daily. Each test would be intended to seek out red flags. For example, delivery dates could be examined for times when deliveries are not normally made (holidays, weekends, after hours, etc.) and selected transactions would then be reviewed.

FCA also allows for special alarms called “audit hooks.” These are audit rules that snare transactions of a suspicious nature and allow for real-time intervention. A common example is when someone travels abroad and uses a credit card outside the normal venue. An audit hook captures the first use of the card in the foreign venue and immediately alerts a representative who then decides how to handle the transaction. One response is to attempt to contact the cardholder by phone or email. The response can take less than one minute. The hooks are highly effective at detecting and deterring possible fraudulent activities (Romney and Steinbart, 2008). Thus,

Proposition 5: Management will positively perceive the forensic continuous auditing model as an effective and efficient forensic tool.

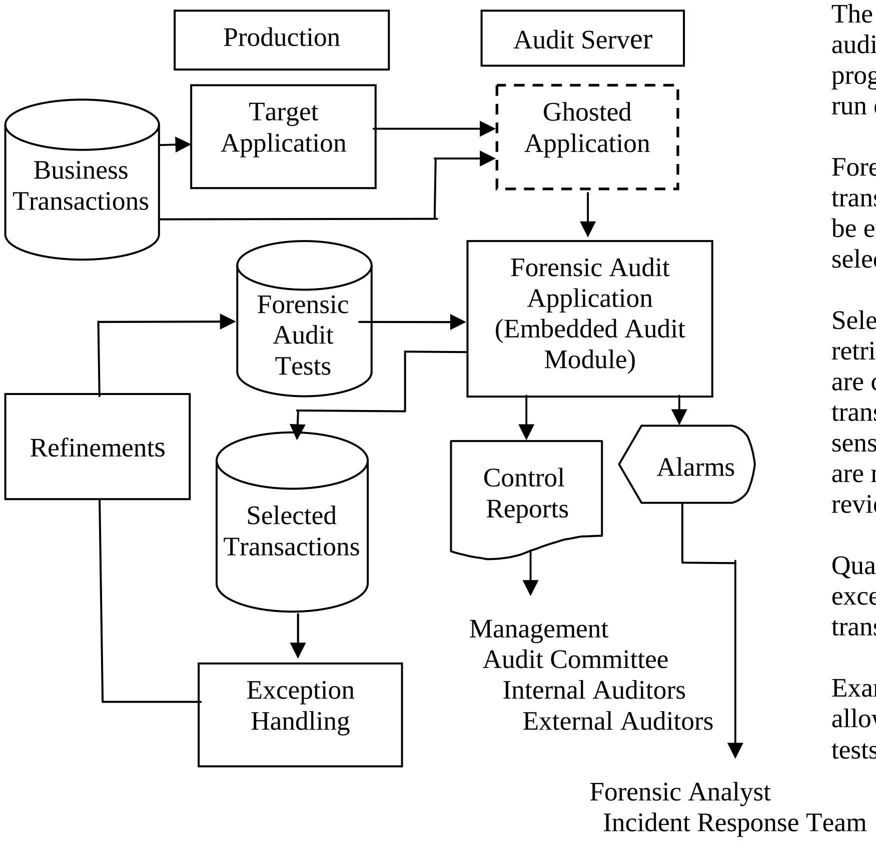
8. THE FORENSIC CONTINUOUS AUDIT SYSTEM

The FCA system is shown in Figure 5. Note that the Forensic Audit Application module functions as an embedded audit module but is outside the actual production version and can be applied to different applications eliminating the necessity of building separate embedded audit modules.

The Forensic Audit Application works with a cloned copy of the actual application using actual transactions but does not alter actual accounts or affect the performance of the system. Sensitive audit tests can trigger alarms that request immediate response. Otherwise, selected transactions are saved and reports created for scheduled reviews. A phased approach would be based on creating a tested system that could be copied for other applications. Because invoking the tests within the production version of the application could reduce performance significantly, performing the tests in the background is preferred.

Some analysis of selected transactions may indicate the need for deeper inquiry. Extended analysis could be performed using a CAAT such as ACL or IDEA.

Figure 5: Forensic Continuous Auditing System



The target application is cloned to the audit server using a ghosting program. The actual transactions are run on the audit server.

Forensic tests are applied to all transactions which may or may not be erroneous or fraudulent but are selected for review.

Selected transactions are stored for retrieval and archival. Control reports are created for review of selected transactions. Some transactions fail sensitive rules that trigger alarms and are made available for immediate review.

Qualified forensic analysts evaluate exceptions and suspicious transactions.

Examination of trends and anomalies allow auditors to refine fraud audit tests.

9. CONCLUSIONS

The role of audits is clearly important and can have a strong preventative effect on fraudulent behavior, but audits alone cannot be relied upon exclusively for fraud detection and, with the increase of transactions processed, may not be an effective mechanism for uncovering errors or misuse. Experience has shown that the traditional audit is not an effective mechanism for uncovering fraud. Auditors and managers are faced with increased pressure to tighten internal controls and reduce corporate risks. At the same time, information systems are becoming increasingly more complex and larger sets of transactions are being processed. Evidence exists that when faced with advanced technology auditors often resort to manual approaches that are less effective at detecting fraud or material misstatements. Although continuous auditing is an attractive solution, many companies have failed to embrace it because of implementation issues and lack of trained auditors. This paper presents cogent reasons for adopting a system of forensic continuous auditing. Based on five propositions, an approach is presented that is manageable and scalable and can be introduced in phases. By using the continuous auditing approach, managers can be assured of transaction integrity and auditors can be relieved of some of the burdens of repetitive testing of controls and balances, allowing auditors to focus on matters that are more likely to reduce risk.

REFERENCES

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

Albrecht, C. C., W. S. Albrecht, and J. G. Dunn. 2001. Can Auditors Detect Fraud. *Journal of Forensic Accounting*. Vol. II, pp. 1-12.

Aldhizer, G. R., and J. D. Cashell. 2006. Automating the Confirmation Process: How to Enhance Audit Effectiveness and Efficiency. *The CPA Journal*. Vol. 76 (4), pp. 28–32.

Alles, M.,, A. Kogan, and M. A. Vasarhelyi. 2008. Putting Continuous Auditing Theory into Practice: Lessons from Two Pilot Implementations. *Journal of Information Systems*, Vol. 22 (2), pp. 195–214.

Alles, M. G., A. Kogan, and M. A. Vasarhelyi. 2002. Feasibility and Economics of Continuous Assurance. *Auditing: A Journal of Practice & Theory*. Vol. 21 (1), pp. 125–138.

American Institute of Certified Public Accountants (AICPA). 2001. *The Effect of Information Technology on the Auditor's Consideration of Internal Control in a Financial Statement Audit. Statement of Auditing Standards No. 94*. New York NY: AICPA.

________. 2002. *Consideration of Fraud in Financial Statement Audit. Statement of Auditing Standards No. 99*. New York NY: AICPA.

________. 1995. *Consideration of Internal Control in a Financial Statement Audit. Statement of Auditing Standards No. 78*. New York NY: AICPA

________. 1988. *Analytical Procedures. Statement of Auditing Standards No. 56*. New York NY: AICPA

Arens, A.; Elder, R.; Beasley, M. 2006. Auditing and Assurance Services: An Integrated Approach. *Pearson Prentice Hall.*

Association of Certified Fraud Examiners (ACFE) *2010 Report to the Nation on Occupational Fraud and Abuse*.

Audit Command Language (ACL) Downloaded January 4, 2011 from: http://www.acl.com/solutions/fraud_detection.aspx

Canadian Institute of Chartered Accountants and American Institute of Certified Public Accountants (CICA/AICPA). 1999. *Continuous Auditing*. Research report. Toronto, Canada: CICA.

Caster, P. and R. Sriram. 1996. An Investigation of Accounts Receivable Confirmation Process Timing. *Auditing: A Journal of Practice & Theory*. Vol. 15 (1), pp. 135–141.

Charles Rivers & Associates. 2005. *Sarbanes-Oxley Section 404 Costs and Remediation of Deficiencies: Estimates from a Sample of Fortune 1000 Companies*. Downloaded January 5, 2011 from: http://www.sec.gov/spotlight/SOX02comp/SOX02comp-all-attach.pdf.

Chen, C. and J. T. Sennetti. 2005. Fraudulent Financial Reporting Characteristics of the Computer Industry Under a Strategic-Systems Lens. *Journal of Forensic Accounting*. Vol. VI, pp. 23-54.

Debreceny, R. S., G. L.Gray, J. J. Ng, K. S. Lee, and W. Yau. Fall 2005. Embedded Audit Modules in Enterprise Resource Planning Systems: Implementation and Functionality. *Journal of Information Systems*. Vol 19 (2), pp. 7–27.

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

Groomer, S. M., and U. S. Murthy. 1989. Continuous Auditing of Database Applications: An Embedded Audit Module Approach. *Journal of Information Systems*. Vol. 3 (1), pp. 53-69.

Grove, H. and T. Cook. 2004. Lessons for Auditors: Quantitative and Qualitative Red Flags. *Journal* *of Forensic Accounting*. Vol. V, pp. 131-146.

Hermanson, D. R. , B. Moran, C. S. Rossie and D. T. Wolfe. 2006. Continuous Monitoring of Transactions to Reduce Fraud, Misuse, and Errors. *Journal of Forensic Accounting*. Vol. VII, pp. 17-30.

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

Kuhn, J. R. Jr. and S. G. Sutton. Spring 2010. Continuous Auditing in ERP System Environments: The Current State and Future Directions. *Journal of Information Systems*. Vol. 24 (1), pp. 91-112.

Janvrin, D., D. Bierstaker and D. J. Lowe. Spring 2009. An Investigation of Factors Influencing the Use of Computer-Related Audit Procedures. *Journal of Information Systems*. Vol. 23, (1), pp. 97–118.

Johnson, C. B. and T. C. Ireland. 2007. An Empirical Examination of Manipulation in Components of the Income Statement. *Journal of Forensic Accounting*. Vol. VIII, pp. 1-28.

Lanza R. B., and S. Gilbert. 2007. A Risk-Based Approach to Journal Entry Testing. *Journal of Accountancy*. Vol. 204, pp. 32–35.

Nigrini, M. J. 2006. Monitoring Techniques Available to the Forensic Accountant. *Journal of Forensic Accounting*. Vol. VII, pp. 321-344.

Nondorf, M. E., Singer, Z. and You, H., (February 2011) A Study of Firms Surrounding the Threshold of Sarbanes-Oxley Section 404 Compliance. AAA 2008 Financial Accounting and Reporting Section (FARS) Paper. Available at SSRN: http://ssrn.com/abstract=1004965

Rezaee, Z., A. Sharbatoghlie, R. Elam, and P. L. McMickle. 2002. Continuous auditing: Building

automated audit capability. *Auditing: A Journal of Practice & Theory*. Vol. 21 (1), pp. 147–163.

Roth, J. and D. Espersen. 2003. *Internal Audit's Role in Corporate Governance: Sarbanes-Oxley Compliance*. Altamonte Springs: The Institute of Internal Auditors Research Foundation.

Li., S., S. Huang and Y. G. Lin. Fall 2007. Developing a Continuous Auditing Assistance System based on Information Process Models. *Journal of Computer Information Systems*. Vol. 48 (1), pp. 2-13.

Oringel, J. and G. R. Aldhizer. Fall 2009. Continuous Auditing and Monitoring: Enhancing the Efficiency and Effectiveness of Auditing and ERM. *Internal Auditing*. Vol. 24 (5), pp. 17-26.

Public Company Accounting Oversight Board (PCAOB). 2007. Auditing Standard No. 5: *An Audit of Internal Control Over Financial Reporting that is Integrated with an Audit of Financial Statement*.

Romney, M. B. and P. J. Steinbart. 2008. *Accounting Information Systems, 11th ed.* Prentice-Hall.

G. S. Smith. 2005. Computer Forensics: Helping to Achieve the Auditor's Fraud Mission?. *Journal of Forensic Accounting*. Vol. VI, pp. 119-134.

Vasarhelyi, M. A, M. Alles, and A. Kogan. 2004. Principles of Analytic Monitoring For Continuous Assurance. *Journal of Emerging Technologies in Accounting*. Vol. 1, pp. 1–21.

Wells, J. T. 2011. *Principles of Fraud Examination*. Hoboken, NJ: John Wiley & Sons.

DEVELOPMENT OF A DISTRIBUTED PRINT-OUT MONITORING SYSTEM FOR EFFICIENT FORENISIC INVESTIGATION

Satoshi Kai
Hitachi, Ltd., Yokohama Research Laboratory
292 Yoshida, Totsuka-ku, Yokohama, Kanagawa 244-0817, Japan
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Tetsutaro Uehara
Associate Professor
Academic Center for Computing and Media Studies, Kyoto University

ABSTRACT

If information leakage occurs, an investigator is instructed to specify what documents were leaked and who leaked them. In the present work, a distributed print-out monitoring system—which consists of a virtual printer driver and print-out policy/log management servers—was developed. For easily matching the discovered (i.e., leaked) paper document with the print-out log, the virtual printer driver acquires full-text of printed-out documents by DDI hooking technique to check the content, transforms a spool file to a picture file and creates both a thumbnail and text log for forensic investigation afterwards. The log size is as only about 0.04 times bigger than that for printed-out electronic documents, so the storage size needed for the thumbnail and text log is also small.

Keywords: Information leakage, Print-out, Digital forensics, Log, Virtual printer driver

1. INTRODUCTION

Information leakage is one of the most serious incidents facing a company or an organization. Many leakage incidents happen in the form of documents. As for documents created in an office, it was found that 93% are in electronic form and 7% are in paper form (Kevin 2000 [1]). However, 72.6% of leakage routes are known to be via paper medium (JNSA 2010 [2]). In other words, although paper documents make up a smaller percentage of the total amount of documents, they are the main cause of information leakage. Since information-communication technology (ICT) is becoming ever more common in all styles of working, these paper documents are considered to be those created in electronic form first and then printed-out in paper form. Accordingly, the security of such print-out matter is an important factor in preventing and detecting information leakage.

Once information leakage occurs, the company or organization starts incident response using digital forensics. According to Takahashi 2008 [3], this response is composed of following steps.

1. Detection
2. Initial response
3. Investigation
4. Disclosures
5. Restraint and recovery
6. Post incident

From start to finish of this incident response, digital forensics is used to determine leakage facts such as what documents were leaked and who leaked them.

In the present work, a print-out monitoring system is in place that prevents illegal print-out according to the content under usual working circumstances as well as supports digital forensics when information leakage occurs. Moreover, this system is easy to install on existing PCs and requires less storage size to accumulate the print-out logs.

2. DIGITAL FORENSIC SCENARIO CONCERNING INFORMATION LEAKAGE

2-1. Supposed information-leakage incident

Information-leakage incidents differ from one to another in terms of situation, impact, and so on. To clarify situations and motivation concerning digital forensic, an incident such as that shown in Fig. 1 is presented in this paper. This scenario is taken and modified from a report issued by the Tokyo Metropolitan Police Department in 2010 [4].

Organization profile:

Employees in a given organization create and manage documents classified as state secrets (such as materials containing international-terrorism-related data). The security administrator imposes a strict security policy and audits employee's working records four times a year.

Information-leakage incident:

One day, certain documents concerning a state secret were found in a book at a book store. When an investigator checked the book, the state secret was found to be contained in a scanned file of a printed-out document.

Digital-forensic purpose:

The investigator was instructed to specify what electronic documents were printed-out and who made the print-out. If these facts were specified, the organization would be able to make the appropriate lawful response.

Fig. 1: Supposed organization and information-leakage incident

2-2. Supposed document-management model

The organization must manage the documents properly and prevent information leakage. Typical document-management models are classified as a central-management model or a distributed model.

2-2-1. Central-document-management model

The central-document-management model (see Fig. 2) is one of client-server models. Clients are "dumb terminals," which can only handle "KVM" (keyboard, video, and mouse) operations, i.e., not storage. The servers are file servers and document-management servers. All documents created by users are stored only on the server side, and any paper documents are printed out on the shared printer. Any printed-out documents are therefore almost identical to the original one on the server side (see broken arrow in Fig. 2).

If the information leakage mentioned in section 2-1 occurs, the investigator must collect and search both the print-out logs at the shared printer and the documents on the server side (see unbroken arrow in Fig. 2). These days, search engines are used widely on the server side, so they are useful for supporting digital forensics.

The central-management model is ideal in regard to digital forensics because the investigator only has to collect and search documents on the server side.

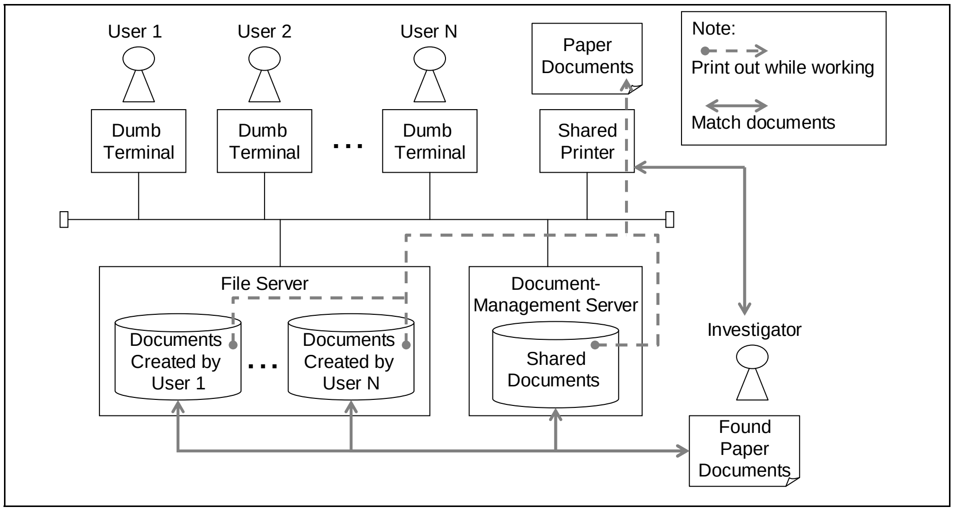


Fig. 2: Central-document-management model

2-2-2. Distributed-document-management model

The distributed-document-management model (see Fig. 3) is a client-server model in which the clients are PCs that can handle storage. The servers are the same as those in the central-document-management model. The documents created by a user are stored on both the server side and the client side. Any printed-out documents are thus almost identical to those on both the server side and the client side (see broken arrow in Fig. 3).

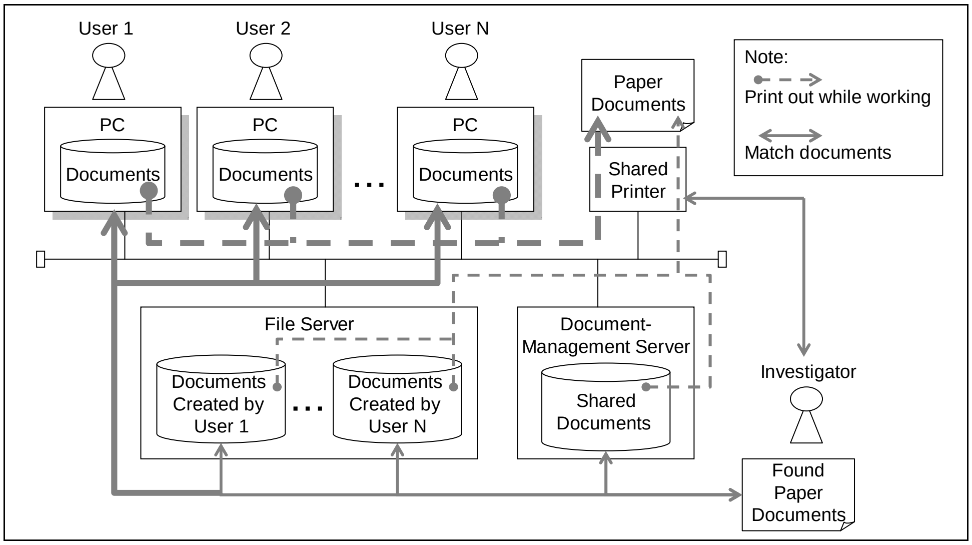


Fig. 3: Distributed-document-management model

If the information leakage mentioned in section 2-1 occurs, the investigator must collect and search the print-out logs at the shared printer, the documents on the server side, and the documents on the client side (see solid unbroken arrow in Fig. 3). In particular, the documents on the client side are sometimes hard to investigate because many more PCs may exist on the client side than on the server side and because not only complete versions of documents but also incomplete manuscripts in poor order exist. The present study focused on the distributed-document-management model (Fig. 3) and especially addresses collecting and searching the printed-out documents on the client side.

By the way, it may be considered that the documents are transported electronically to an off-site location (e.g., via flash drive or email) and then printed-out. In that case, the documents can be protected by a conventional digital rights management (DRM) function [5][6]. By using the DRM, print-out can be controlled from the central DRM server. But this DRM is only useful for delivering the documents, not creating and modifying. So DRM is out of scope of the present study.

2-3. Digital-forensic techniques for print outs

Digital forensics includes many investigation procedures. To specify what electronic documents were printed out and who did the printing out, the following procedure, shown schematically in Fig. 4 as four steps, is used for digital-forensic investigations on Windows PCs. Note that Unix PCs or Mac PCs can also be investigated using almost the same or alternative steps. However, Windows PCs are used widely, so this study addresses information leakage with Windows PCs.

Step1: Check the registry key, such as “HKEY_LOCAL_MACHINE/SYSTEM/CurrentControlSet /Control/Print/Printers.” If no printer driver is installed, the PC is judged to be not used for print outs; that is, it is not suspected of information leakage.

Step2: Check the event log, such as event ID10, ID540, and ID560 (Microsoft’s Log Audit Guide 2007 [7]). If no print-out log is recorded, the PC is judged to be not used for print outs; that is, it is not suspected of information leakage.

Step3: Check the spooler located at “C:/WINDOWS/system32/spool/PRINTERS”. If any residual spool files are left, the investigator can match the printed-out image with the found paper documents.

Step4: Check the documents listed in the print-out log so that the investigator can match the documents with the discovered paper documents and specify when and where the document was printed out and who printed it out.

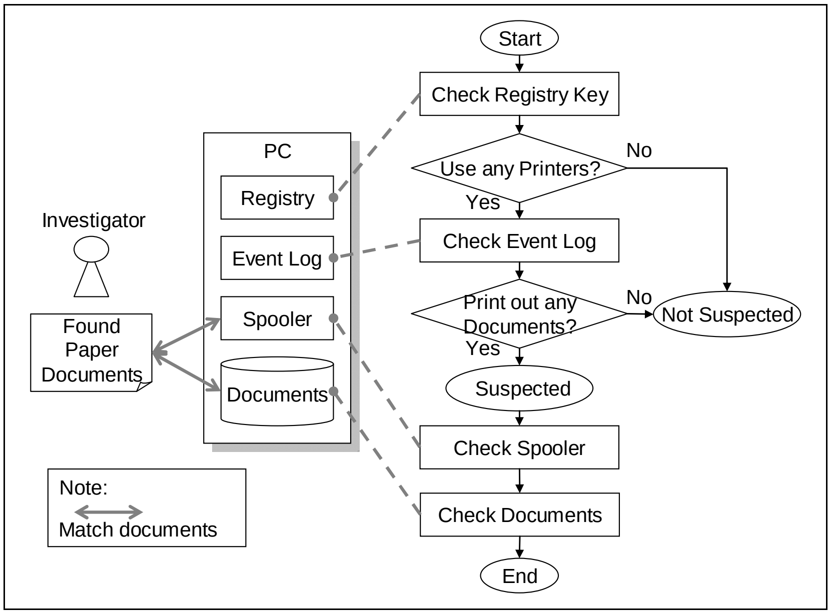


Fig. 4: Digital forensic procedure for identifying printed-out documents with found paper documents.

2-4. Problems

The above-mentioned digital-forensic procedure for investigating print outs is sometimes useful, but it suffers the following residual problems.

Problems 1: Uncertainty regarding what documents were printed out.

The Windows event log records “print job name”, which depends on each print-out application and often includes only the file name, not the file path. The investigator thus cannot always match what documents were printed-out with the discovered paper document, even if the investigator knows the print job name. Moreover, spool files are deleted after succeeding print outs and overwritten one by one. Recovering the spool files is therefore difficult.

Problems 2: It takes a lot of time to collect and confirm the registry, event log, and spool file.

The number of client PCs exceeds that of servers, and the PCs are distributed in a variety of places. Moreover, access to the registry, event log, and spool file needs an administrator privilege for each PC. Consequently, acquiring the registry, event log and spool file data takes more time to collect and confirm.

3. DESIGN OF THE DISTRIBUTED PRINT-OUT MONITORING SYSTEM

To solve the problems described in section 2-4, a distributed print-out monitoring system was designed and constructed.

3-1. Operational model

When an employee needs to print out a document, he (or she) must install the printer driver of the shared printer. The printer driver is often provided by the print server. Even if PCs are distributed in a variety of places, the printer driver can be managed by the print server. The monitoring system was designed with a focus on the printer driver. Moreover, the supposed organization has a strict security policy, so the monitoring system is also equipped with a print-out control function that benefits both users and security administrators. The design of the operational model is shown schematically in Fig. 5.

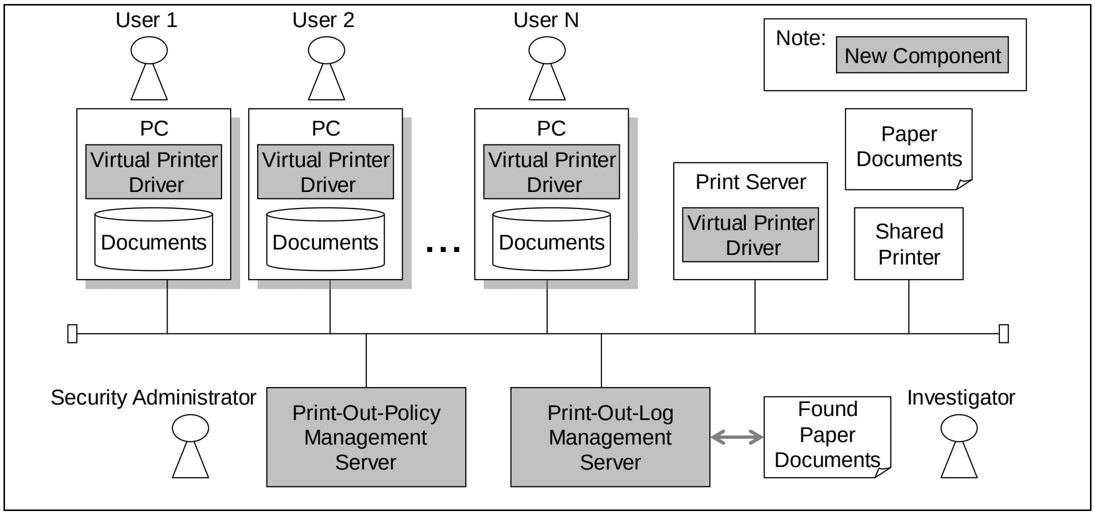


Fig. 5: Operational model of distributed print-out monitoring system

Users perform their business as following:

(1) Users install a virtual printer driver from the print server on each PC.

(2) Connected with the print-out-policy management server, the virtual printer driver checks the print-out content and controls the print jobs on each PC.

(3) The virtual printer driver acquires the print-out logs and sends them to the print-out-log management server.

If an information leakage occurs:

(4) The investigator searches the print-out logs to match a log entry with the leaked paper document.

3-2. Print-out logs

The print-out log is the key to match the log with the found paper documents. The print-out log consists of three items: (1) a spool file itself, (2) a picture file (transformed from (1)), and (3) a spool file acquired as text. These items are compared in Table 1.

The spool file, item (1), itself is sure to match the leaked paper document, but it needs to be re-printed out. The picture file, item (2), is easy to match with the leaked paper document without having to be re-printed out. However, its file size is prone to be big, and optical character recognition (OCR) is not always accurate in the case of text search. The text file, item (3), is easy to search, and the file size tends to be small. However, figures, pictures, and document layout are dropped from item (3).

To find interesting logs in a large amount of print-out logs, item (3) (text) is useful. On the other hand, to match the leaked document, (1) (spool file) or (2) (picture file) is useful. Accordingly, the print-out log was selected to be hybrid, both a picture file and text. Moreover, the picture file was selected to be thumbnails of all the pages of the printed-out document.

Table 1: Comparison of print-out-log items (1), (2), and (3)

Items	(1) Spool file	(2) Picture file	(3) Text
Examples	RAW, EMF, XPS, PS	JPG, PNG etc.	TXT
Match with leaked paper document	Easy	Easy	Not always easy (figures, pictures, and layout are dropped)
Need to re-print-out	Yes	No	No
Log file size	Prone to be big	Prone to be big	Tends to be small
Find an interesting log	Need to find by eye	OCR can be used to extract text (but prone to be incorrect)	Easy to text search.

The print-out-log format and its supposed size are listed in Table 2.

Table 2: Print-out log format

Items		Description	Supposed size
Date		Year, month, day, hour, minutes, seconds	14 bytes
User		Username	≤20 bytes
Printer		Printer name	≤32 bytes
Print job name		Print-job name (depends on print-out application)	≤255 bytes (possibly)
Page number		Number of printed-out pages	≤4 bytes
Content	Thumbnail	Thumbnails of each page	Depends on documents
	Text	Full text of all pages	Depends on documents

3-3. Print-out control

The print job is a key to control printing out documents. However, the print job itself is hard to check according to its content. Accordingly, it was decided to extract text information stored on the virtual printer driver, to check its content of text information, and to allow or prohibit the print job to send to the shared printer. Extracting text information from the text print-out log is described in section 3-2.

Checking text typically follows two strategies: (1) index search and (2) GREP search. These strategies are compared in Table 3. Index search is fast but not accurate; that is, precision and recall rate (Ricardo et al. 1999 [8]) is not always 100%. In detail, precision rate means the fraction of retrieved documents that are relevant to the search, and recall rate means the fraction of the documents that are relevant to the query that are successfully retrieved. In contrast GREP search is accurate; that is, recall rate is always 100%, but speed is low. From the viewpoint of checking text, precision below 100% is allowed but recall rate below 100% is never allowed because of the possibility of missing the interesting print-out logs. Strategy (2) (GREP search) was thus chosen for checking text.

Table 3: Comparison of strategies for checking text

Strategy	(1) Index search	(2) GREP search
Search speed	Fast	Slow
Spare resource before search	Indexing time and storage space for index files are needed.	Spare time and storage are not needed.
Precision	≤ 100%	≤ 100%
Recall	≤ 100%	Always equals 100%

Examples of the GREP search keywords are listed in Table 4. These keywords are set by the security administrator on the print-out policy-management server. Alternatively, the investigator may set them on print-out-log management server when performing GREP search of the print-out logs.

Table 4: Examples of keywords

Category	Keywords sample
Confidential	“Confidential”, “Do not print”, “Internal use only”, etc.
Customer	Customer name (depends on each organization or business), credit-card numbers (often expressed by regular expression), etc.

3-4. Digital forensic use

When the investigator uses the distributed print-out monitoring system (Fig. 5), the following procedure is followed step by step.

Step1: Extract characteristic keywords in the leaked paper document

Step2: Perform GREP search for the print-out logs containing those keywords

Step3: Check the thumbnail pictures matched by the keywords, then match the thumbnails with the leaked paper documents.

Step4: Determine when the document was printed out (according to the print-out logs) and who did the printing.

By following this procedure, even if the client PCs are distributed widely, the investigator can collect print-out logs and search them accurately and efficiently. This procedure thus solves the problems stated in section 2-4.

4. IMPLEMENTATION OF VIRTUAL PRINTER DRIVER

4-1. Basic function of printer driver

A printer driver is a program (called by a print-out application) that sends a print job to a printer (Microsoft Developer Network 2010 [9]). The process followed by the printer driver is typically classified as two processes: layout arrangement and character output. Layout arrangement determines how many pages are needed and where to arrange characters and figures, etc. in the pages. Character output determines font, size, color, and decoration of the characters. Especially, the characters included in an electronic document are used as the input of the character-output process (see Fig. 6). For example, “a” is expressed by the character code “U+0061” in an electronic document. The character-output process transforms the code “U+0061” to the shape of “a”.

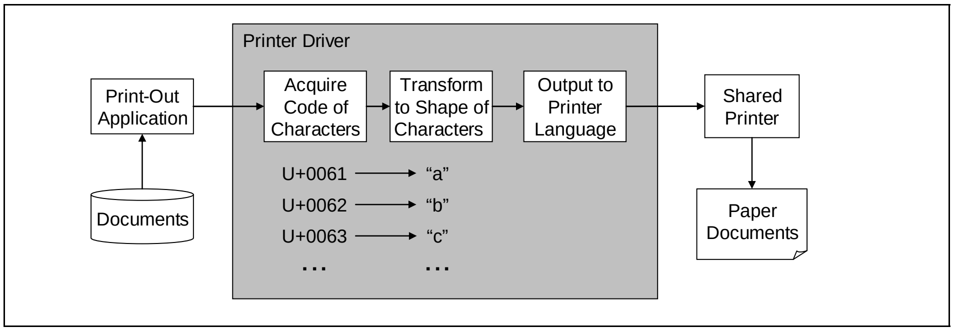


Fig. 6: Outline of character-output processing

4-2. Virtual printer driver

The virtual printer driver is a key component of the distributed print-out monitoring system. It is generally called a print-out application and sends a print job as a bitmap file, which can be printed out by a real printer driver of any kind. The architecture of the virtual printer driver is shown in Fig. 7.

When a character code is acquired, a DDI (device-driver-interface) hooking technique (Microsoft Developer Network 2010 [10]) modifies the acquisition process and transforms the characters into Unicode character code. All the characters are connected to be full-text and the full-text is then checked by the GREP search. If any NG keywords are included, the print job is deleted and the print out is stopped. If no NG keywords are included, both a text log and a thumbnail log are created and sent to the print-out-log management server.

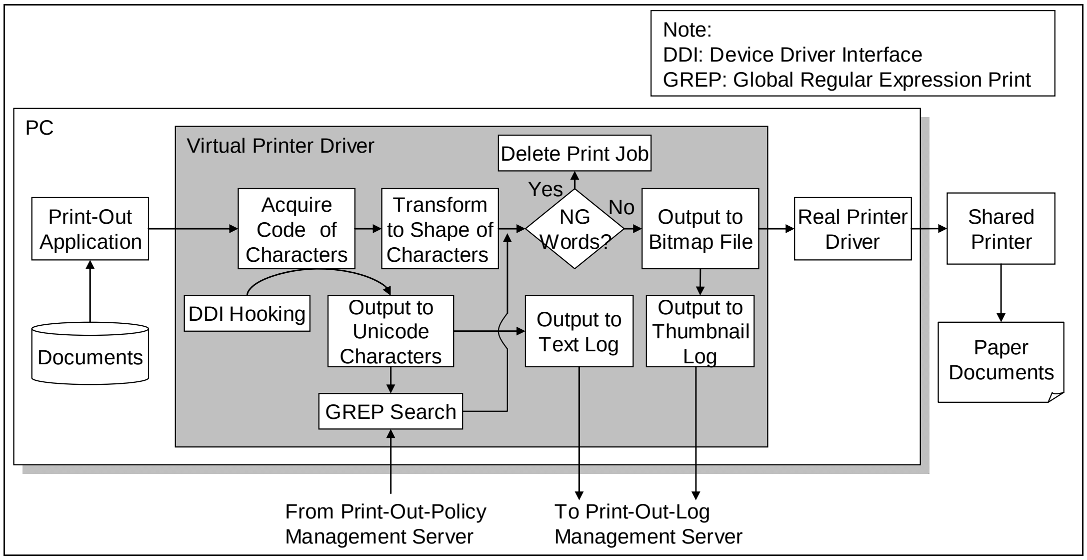
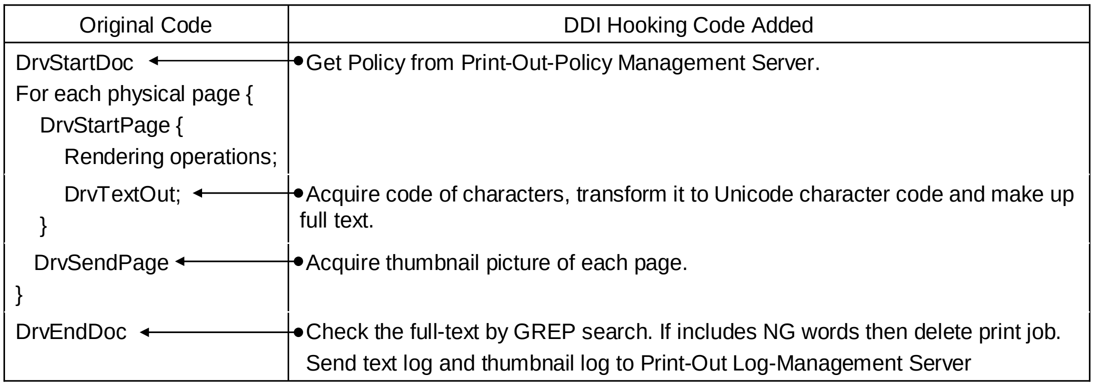

Note:
DDI: Device Driver Interface
GREP: Global Regular Expression Print
PC
Virtual Printer Driver
Delete Print Job
Print-Out Application
Acquire Code of Characters
Transform to Shape of Characters
Yes
NG Words?
No
Output to Bitmap File
Real Printer Driver
Shared Printer
Documents
DDI Hooking
Output to Unicode Characters
Output to Text Log
Output to Thumbnail Log
Paper Documents
GREP Search
From Print-Out-Policy Management Server
To Print-Out-Log Management Server

Fig. 7: Architecture of virtual printer driver

4-3. DDI hooking

The virtual printer driver was implemented on Windows XP SP3. The pseudo-code is shown in Fig. 8. The document print-out process begins with a DrvStartDoc call and ends with a DrvEndDoc call. For each physical page, the page-print-out process begins with a DrvStartPage call and ends with a DrvSendPage call. Between the DrvStartPage call and the DrvSendPage call, rendering operations and DrvTextOut are called as needed.

DDI hooking is provided by the Windows OS. By using that, the developer can refer or modify many kinds of the print-out control information. By hooking the DrvTextOut call, all characters code can be acquired. The hooking process is shown schematically in Fig. 8.



Original Code	DDI Hooking Code Added
DrvStartDoc	Get Policy from Print-Out-Policy Management Server.

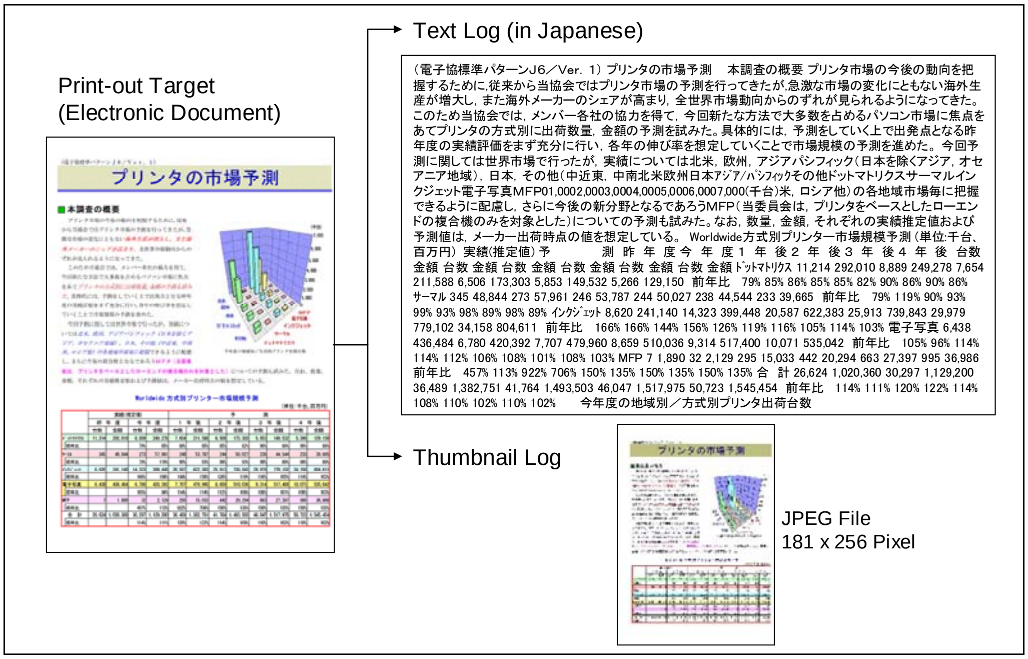
For each physical page {
  DrvStartPage {
    Rendering operations;	

DrvTextOut;
  }
	Acquire code of characters, transform it to Unicode character code and make up full text.

DrvSendPage }	Acquire thumbnail picture of each page.
DrvEndDoc	Check the full-text by GREP search. If includes NG words then delete print job. Send text log and thumbnail log to Print-Out Log-Management Server

Fig. 8: Pseudo-code with DDI hooking process added

An example of a print-out log is shown in Fig. 9. The thumbnail picture is set as a JPEG file with a size of 181 × 256 pixels because the thumbnail picture included in the XPS file has the same specification (Microsoft 2010 [11]).



```
（電子協標準パターンJ6／Ver. 1）プリンタの市場予測　本調査の概要 プリンタ市場の今後の動向を把
握するために,従来から当協会ではプリンタ市場の予測を行ってきたが,急激な市場の変化にともない海外生
産が増大し，また海外メーカーのシェアが高まり，全世界市場動向からのずれが見られるようになってきた。
このため当協会では，メンバー各社の協力を得て，今回新たな方法で大多数を占めるパソコン市場に焦点を
あてプリンタの方式別に出荷数量，金額の予測を試みた。具体的には，予測をしていく上で出発点となる昨
年度の実績評価をまず充分に行い，各年の伸び率を想定していくことで市場規模の予測を進めた。今回予
測に関しては世界市場で行ったが，実績については北米，欧州，アジアパシフィック（日本を除くアジア，オセ
アニア地域），日本，その他（中近東，中南北米欧州日本ｱｼﾞｱ/ﾊﾟｼﾌｨｯｸその他ドットマトリクスサーマルイン
クジェｯﾄ電子写真MFP01,0002,0003,0004,0005,0006,0007,000(千台)米，ロシア他）の各地域市場毎に把握
できるように配慮し，さらに今後の新分野となるであろうMFP（当委員会は，プリンタをベースとしたローエン
ドの複合機のみを対象とした）についての予測も試みた。なお，数量，金額，それぞれの実績推定値および
予測値は，メーカー出荷時点の値を想定している。 Worldwide方式別プリンター市場規模予測（単位:千台、
百万円） 実績(推定値) 予　　　　測 昨 年 度 今 年 度 1 年 後 2 年 後 3 年 後 4 年 後 台数
金額 台数 金額 台数 金額 台数 金額 台数 金額 台数 金額 ﾄﾞｯﾄﾏﾄﾘｸｽ 11,214 292,010 8,889 249,278 7,654
211,588 6,506 173,303 5,853 149,532 5,266 129,150 前年比　79% 85% 86% 85% 85% 82% 90% 86% 90% 86%
ｻｰﾏﾙ 345 48,844 273 57,961 246 53,787 244 50,027 238 44,544 233 39,665 前年比　79% 119% 90% 93%
99% 93% 98% 89% 98% 89% ｲﾝｸｼﾞｪｯﾄ 8,620 241,140 14,323 399,448 20,587 622,383 25,913 739,843 29,979
779,102 34,158 804,611 前年比　166% 166% 144% 156% 126% 119% 116% 105% 114% 103% 電子写真 6,438
436,484 6,780 420,392 7,707 479,960 8,659 510,036 9,314 517,400 10,071 535,042 前年比　105% 96% 114%
114% 112% 106% 108% 101% 108% 103% MFP 7 1,890 32 2,129 295 15,033 442 20,294 663 27,397 995 36,986
前年比　457% 113% 922% 706% 150% 135% 150% 135% 150% 135% 合　計 26,624 1,020,360 30,297 1,129,200
36,489 1,382,751 41,764 1,493,503 46,047 1,517,975 50,723 1,545,454 前年比　114% 111% 120% 122% 114%
108% 110% 102% 110% 102%　　今年度の地域別／方式別プリンタ出荷台数
```

Fig. 9: Example print-out logs

4-4. Searching Japanese text

English and Japanese differ in that English sentences have blanks between words to distinguish each word and that Japanese does not separate words with blanks. Chinese and Korean have the same characteristics as Japanese. So a GREP search is prone to be slower in the cases of Japanese, Chinese, and Korean. To distinguish every word, morphological-analysis tools [12][13][14] are known to be useful. Using both the GREP search and morphological-analysis tools is one way to search Japanese text.

By using the morphological-analysis tools, the full-text is divided into each word. Especially the noun words tend to be divided exactly. Many keywords are usually noun words, so the tools influence little on search leakage.

5. EVALUATION OF PRINT-OUT LOG SIZE

The print-out log is better if its size is smaller. The following evaluation addresses the size of the print-put log.

5-1. Precondition

Print-out log size depends on the target electronic documents. To standardize the evaluation, standard-test-patterns for printers were used (JEITA 2003 [15]). In Fig. 9, one of the test patterns is shown. These test patterns are as follows.

- File formats are Microsoft Word 97, Excel 97, Power Point 97, and so on.
- Characters, graphs, pictures, tables, figures, images, and so on are included.
- Page numbers are from 1 to 12 pages only.
- Both monochrome and color documents are included.

To compare the print-out log size of different logs, the following two kinds of logs were chosen from Table 1.

- Size of spool file
- Size of both thumbnail and text log (shown in Table 2)

5-2. Evaluation result

5-2-1. Size of spool file

The spool-file formats were RAW, EMF, XPS, and PS. The standard-test-patterns were printed-out by a RAW printer driver, an EMF printer driver, a XPS printer driver, and a PS printer driver. Average of their spool-file sizes was then calculated. The calculation results are shown in Fig. 10. The relationship between standard-test-pattern size and average spool-file size is almost proportional. The average spool-file size is as about 1.85 times bigger than that of the standard-test-patterns.

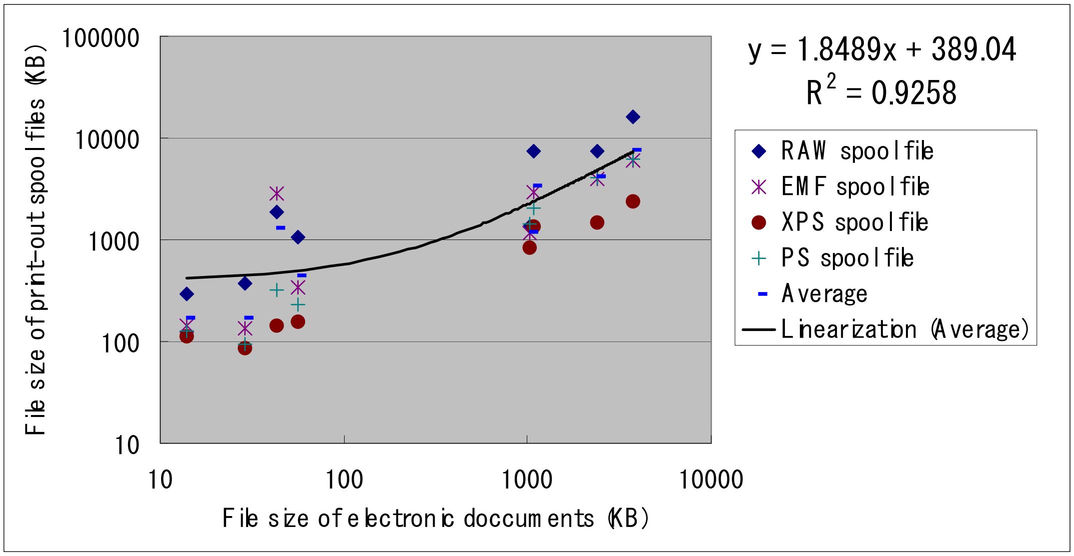


Fig. 10: Average size of spool files (RAW, EMF, XPS, and PS)

5-2-2. Size of thumbnail and text log

Total size of the log is the thumbnail log size plus the text log size. The standard-test-patterns are printed out by the virtual printer driver described in section 4.2. Two kinds of log sizes were then added. The result is shown in Fig. 11. The total size is as about 0.04 times bigger than that of the standard-test-patterns.

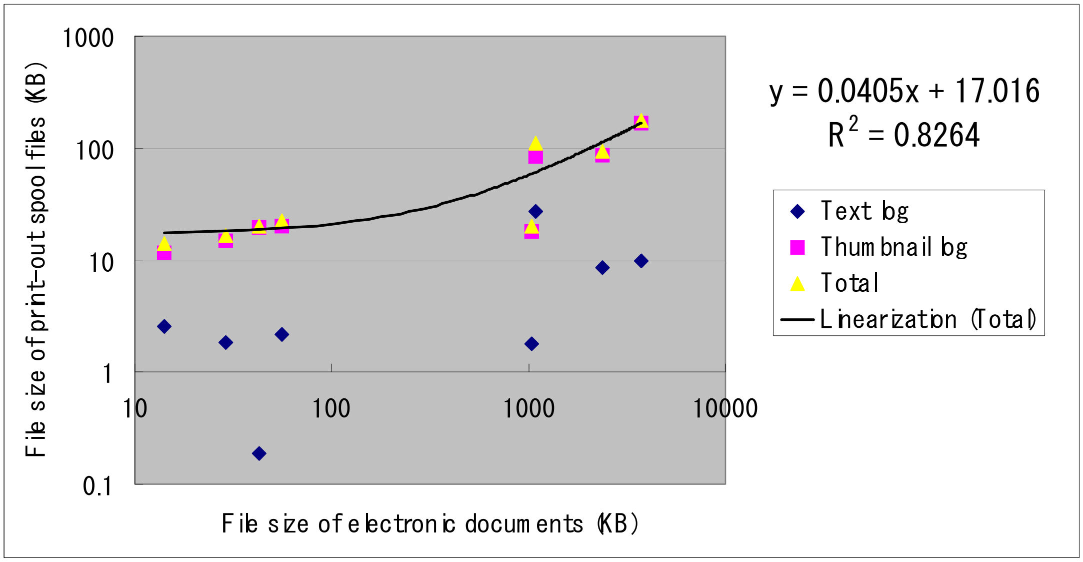


Fig. 11: Total size of thumbnail and Text Logs.

5-3. Application to typical office

The required storage size was estimated for the office supposed as follows.

- 30 employees share 1 printer
- Each employee print outs 3000 pages per year
- Average printed-out electronic document size is 1 MB

If the print-out log is a spool file, the estimated size is 201.4 GB per year. If the print-out log is a thumbnail and text log, the estimated size is 5.2 GB per year. In other words, the thumbnail and text log size decreases by 97.4% compared to the spool file. This means that only 2.6% of the storage space is needed in the case of the thumbnail and text log compared to the spool file.

6. RELATED WORKS

(1) Print-out logs

Related print-out forensic work have been done on print servers (Canon, 2008 [16] and Ricoh, 2008 [17]). The print servers acquire text information from the print jobs and put the print-out records in storage. In another research (Fujii, 2010 [18]) text information is acquired by EMF spool file. This work demonstrated a virtual printer driver that acquires text information. The virtual printer driver is faster in acquiring text information than the work on the print servers.

(2) Watermark print

Watermark print outs have also been researched (Ono, 2004 [19]). A watermark, which includes date, username, and filename, is printed out on paper documents. If a paper document was leaked, the watermark can be extracted by scanning, and the investigator can determine the date, username, and filename. A watermark print is thus useful only after an information leakage; in contrast, the distributed print-out monitoring system developed in the present work is useful not only after a leakage but also for daily control and periodical auditing.

7. CONCLUSION

A distributed print-out monitoring system—composed of a virtual printer driver and a print-out policy/log management server—was developed. The virtual printer driver acquires text information by DDI hooking, performs GREP search to check the content, and creates a thumbnail and text log. The log size is about 0.04 times bigger than as that of printed-out electronic files. That is, compared to the storage size required for retrieving a print-out log as a spool file, the required storage size for the virtual driver is 97.4% smaller. In our future work, we will address the challenge of confirming the actual usefulness of the system for forensic investigation after information leakage.

Windows, Windows XP, Microsoft Word, Excel, and PowerPoint are registered trademarks of Microsoft Corporation in the United States and/or other countries.

Unix is a registered trademark of The Open Group in the United States and other countries.

Mac is a registered trademark of Apple Computer, Inc., in the United States and other countries.

REFERENCES

[1] Kevin Craine (2000), “Designing a Document Strategy”, MC2 Books.

[2] NPO Japan Network Security Association (2010), “Year 2009 Research Report about Information Security Incidents Version 1.1” (in Japanese), http://www.jnsa.org/result/incident/2009.html, 2011-01 accessed.

[3] Ikuo Takahashi (2008), “Use of Digital Forensics and the legal problems in information leakage incident response in Japan”, Proceedings of 4th Annual IFIP WG 11.9 International Conference on Digital Forensics - Short Papers -, pp. 65-72.

[4] Tokyo Metropolitan Police Department (2010), “Report about information leakage incident of international terrorism related data on the Internet” (in Japanese), http://www.keishicho.metro.tokyo.jp/image/jian_101224.pdf, 2011-01 accessed.

[5] Adobe (2011), “Adobe LiveCycle Rights Management ES2” (in Japanese), http://www.adobe.com/jp/products/livecycle/rightsmanagement/, 2011-03 accessed.

[6] Microsoft (2011), “Information Rights Management” (in Japanese), http://www.microsoft.com/japan/office/previous/2003/business/irm/default.mspx, 2011-03 accessed.

[7] Microsoft (2007), “Log Audit Guide for Microsoft Server Product - Audit for Print Jobs” (in Japanese), http://technet.microsoft.com/ja-jp/solutionaccelerators/dd285678, 2011-01 accessed.

[8] Ricardo Baeza-Yates and Berthier Ribeiro-Neto (1999), “Modern Information Retrieval,” Addison Wesley.

[9] Microsoft Developer Network (2010), “Rendering a Print Job”, http://msdn.microsoft.com/en-us/library/ff561943(VS.85).aspx, 2011-01 accessed.

[10] Microsoft Developer Network (2010), “Non-COM-Based DDI Hook-out Functions”, http://msdn.microsoft.com/en-us/library/ff557586(VS.85).aspx, 2011-01 accessed.

[11] Microsoft (2010), “XPS Specification and License Downloads”, http://www.microsoft.com/whdc/device/print/xps/downloads.mspx, 2011-01 accessed.

[12] Basis Technology (2011), “Rosette Base Linguistics for Japanese” (in Japanese), http://www.basistech.jp/base-linguistics/japanese/, 2011-03 accessed.

[13] Taku Kudo (2009), “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”,

http://mecab.sourceforge.net/, 2011-03 accessed.

[14] Nara Institute of Science and Technology, Computational Linguistics Lab. (2007), “ChaSen - morphological analyzer”, http://chasen-legacy.sourceforge.jp/, 2011-03 accessed.

[15] Japan Electronics and Information Technology Industries Association (2003), “Standards of Printer Evaluation Pattern”, JEITA IT-3011A.

[16] Canon (2008), “imageWARE Secure Audit Manager” (in Japanese), http://cweb.canon.jp/software/output/lineup/secureaudit/index.html, 2011-01 accessed.

[17] RICOH (2008), “Ridoc IO Data Selector”, http://www.ricoh.co.jp/IPSiO/related_goods/dataselector/, 2011-01 accessed.

[18] Yusaku Fujii and Yoshinobu Horita (2010), “Confidential Document Detection by Applying Character Recognition to EMF Print Data”, Proceedings of the Society Conference of IEICE Vol.2010 5, pp. 124.

[19] Tsukasa Ono and Yuki Egawa (2004), “A Study of Digital Watermark for Printed Image (Security and Society)” (in Japanese), Transactions of Information Processing Society of Japan 45(3), pp. 880-890.

CREATING REALISTIC CORPORA FOR SECURITY AND FORENSIC EDUCATION

Kam Woods
School of Information and Library Science
University of North Carolina
Chapel Hill, NC
kamwoods@email.unc.edu

Christopher A. Lee
School of Information and Library Science
University of North Carolina
Chapel Hill, NC
callee@ils.unc.edu

Simson Garfinkel
Graduate School of Operational and Information Sciences
Department of Computer Science
Naval Postgraduate School
Monterey, CA
slgarfin@nps.edu

David Dittrich
Applied Physics Laboratory
University of Washington
Seattle, WA
dittrich@uw.edu

Adam Russell
Graduate School of Operational and Information Sciences
Department of Computer Science
Naval Postgraduate School
Monterey, CA
amrussell@nps.edu

Kris Kearton
NCTAMS LANT DET ROTA
PSC 819 BOX 64
FPO AE 09645
kris.kearton@eu.navy.mil

ABSTRACT

We present work on the design, implementation, distribution, and use of realistic forensic datasets to support digital forensics and security education. We describe in particular the "M57-Patents" scenario, a multi-modal corpus consisting of hard drive images, RAM images, network captures, and images from other devices typically found in forensics investigations such as USB drives and cellphones. Corpus creation has been performed as part of a scripted scenario; subsequently it is less "noisy" than real-world data but retains the complexity necessary to support a wide variety of forensic education activities. Realistic forensic corpora allow direct comparison of approaches and tools across classrooms and institutions, reduce the time required to prepare useful educational materials, and eliminate concerns of exposing students to privacy-sensitive or illegal digital materials. The "M57-Patents" corpus can be freely redistributed without rights-restricted materials, and is available with disk images packaged in both open (Advanced Forensic Format) and commercial (EnCase) formats.

Keywords: Forensics, Corpora, Realistic Data, Education, Security, Tool Validation

1. INTRODUCTION

Digital Forensics combines expertise and methods drawn from computer science, criminology, psychology, and other related fields. Most forensic curricula expect students not only to master existing tools, but also to build an understanding of the strengths and limitations of these tools in their application to real-world data. This understanding fosters the improvement of existing technologies and development of new leading-edge techniques drawing a diverse range of areas including cryptography, machine learning, linguistics, and visualization (Garfinkel 2010).

A fundamental issue in security and forensic education and research is that *real data* is often unsuitable for education purposes due to the presence of information that is confidential. As a result, many of those who teach digital forensics spend a significant amount of their time preparing disk images, packet dumps, memory dumps and other kinds of forensic materials for student use. But the resulting data is often *insufficiently realistic.* A related problem is that many of those who created forensic data for student use, in an attempt to mimic the real-world, inadvertently make the data sets *needlessly complex*. It is exceedingly difficult to create test data sets that are both simple enough for classroom analysis and complex enough to convey external validity.

This project seeks to overcome the paucity of existing constructed realistic corpora that mimic real data without the associated privacy and security concerns. We do this through the creation and distribution of more than 40 digital forensic images, packet dumps, and memory images. These sets are free of privacy-sensitive information, are usable without IRB approval, and are freely redistributable without concern for either privacy rights or copyright.

2. PRIMARY OBJECTIVES

Several key objectives have guided the development of this corpus:

1. *Answer Keys*
   In discussions of this project with educators, the most oft-requested feature is that each digital artifact in the corpus include an "answer key" that explains what information can be found in each artifact, where that information is located, and how the problems should be solved.

2. *Realistic Wear and Depth*
   The digital artifacts are created to contain realistic *wear patterns* and *depth*. From the perspective of the investigator, these systems appear to be normal computers that are used on a regular basis for personal communication, web browsing, application installations, and file creation and transfer.

3. *Realistic Background Data*
   One of the primary difficulties that investigators encounter in real-world cases is distinguishing data that is relevant to the case from the sea of background data. The majority of data sets previously developed for educational purposes contain the scenario data and little else. We address this problem by incorporating realistic background data.

4. *Sharing and Redistribution*
   We intend the majority of the digital artifacts we create to be freely redistributable. Materials that contain commercial or copyrighted data–such as Microsoft Windows executables–are made available in redacted form or distributed as originals to organizations that affirm that they possess the appropriate license (for example, the Microsoft Developer Academic License).

The corpus we describe here is accompanied by instructional materials that can be adapted to specific classroom needs and environments. These instructional materials provide ground truth about *who, what, where, when, and how*, regardless of what information may have been lost or is unavailable to the student analyzing the data.

3. REALISTIC FORENSIC CORPORA

Creating realistic forensic corpora that are plausible, internally consistent, and useful in a range of educational contexts is a complex task. As with any attempt to simulate a real-world system, significant planning is required prior to execution of a scenario in order to facilitate the desired outcome. A scenario plan should identify specific education and forensic objectives. For example, if we wish the student or trainees to find a sequence of files that have been transferred from one device to another and subsequently deleted from the original device, these actions must be reflected both in the files themselves (via their final locations, modification and access times, and traces from deletions) and in any logs maintained by the operating system. User activity and any logs or data stores maintained by individual applications (for example, emails sent and received) must be consistent, both in content and any associated metadata.

Scenario sequencing and goals must therefore be carefully planned and transcribed onto calendar dates ahead of time. In more complex scenarios taking place over multi-day periods–in which file systems evolve significantly with use–this is the only way to ensure consistency and limit the introduction of information from non-scenario activity. That is, the scenario calendar must reflect events that take place in *real time*; it is impractical to attempt creation of forensic data that takes place in the future or the past.

In addition to planning the scenario, it is necessary to plan what sort of problems the future forensic student will be asked to solve and how the students will solve the problems. It is also necessary to capture the information required to solve the problem. Significant advance planning is required to adjust the difficulty level of a realistic scenario to meet the needs of a particular class. Introductory classes may focus on learning the operation of tools and on solving discrete problems, while more advanced classes may use exercises designed to simulate the investigative process in more depth. Although it is possible to satisfy multiple needs with a single corpus—and we believe we have done so with M57-Patents—the goal of audience flexibility adds additional complexity, making prior planning even more critical.

Once created, a corpus that is sufficiently realistic can be used for other tasks, such as tool validation and even forensics research. We elaborate on some of the issues involved with existing corpora below, and show how they may be addressed by using a realistic data.

3.1 Training and Education

Most currently available forensic datasets are inappropriate for use in a classroom environment. Drive images with real data acquired from production environments or personal hardware (or purchased from third parties) generally contain private, sensitive, or legally encumbered material. Such images may also contain illegal content. Likewise, data sources from actual forensic investigations can generally not be used in classroom and training environments. Drive images drawn from real environments may further be complicated by the use of security or obfuscation tools–e.g. encryption or steganography–which can impede training when time is limited. Finally, synthesized or collected materials–test datasets, forensic challenges, data constructed by instructor, fake or generated data, and publicly available datasets–present additional issues which may be resolved through the use of realistic corpora.

3.2 Issues with Existing Training Data

There exist a small number of corpora in the form of test data sets and forensic challenges. In our experience these datasets are frequently developed to test a suite of tools rather than as educational aids, and they do not typically represent real-life problems or present specific goals to be accomplished. Realistic corpora can provide specific problems for students to solve while remaining sufficiently complex to exercise available tools. Meanwhile, forensic challenges–including datasets developed by the Honeynet Project1, DFRWS2, and DC33–are often too difficult for students to solve.

Another problem with existing data sets is that the solutions to many of the challenges have already been widely distributed, and as a result answer keys and walkthroughs can be found online. We address this problem by restricting access to our answer keys (through the use of encrypted documents

[1] http://www.honeynet.org/

[2] http://www.dfrws.org/

[3] http://www.dc3.mil/

made available only to instructors).

Finally, there exists a range of public datasets that contain information that seems private but which is not. Examples include Enron emails, YouTube videos, public Facebook profiles, and public chat logs. While these datasets have proven invaluable to researchers for statistical analysis and tool validation, because they are publicly available, well-researched, and frequent subjects of popular media, students may already know what must be found in order to "solve" the associated cases. Realistic datasets can incorporate features common to such datasets (email exchanges, social media interactions) in novel settings that exercise the mechanism of the investigation without the risk of prior knowledge.

3.3 Tool Validation

Tool validation is an important task in forensics operations and research (Carrier 2005; Beebe 2009). Although other datasets exist for testing tools and providing tool validation, the M57-Patents scenario provides additional datasets that can be correlated across various media, and annotations for verification procedure. A tool could validate itself across traffic and verify that specific traffic was generated by checking images of drives from workstations. A primary advantage of using the M57-Patents corpus is the ability to correlate information from various media sources and to verify that tools are performing the specific functions. Additionally, detailed annotations accompany the corpus. These annotations simplify tool validation, because known attributes are already associated with the datasets.

4. CREATING REALISTIC DATA AND SIMULATING SYSTEM WEAR

Realistic datasets must contain data that now only is consistent with the situation(s) being simulated, but also appears to have been created or manipulated by entities whose personalities, motivations, goals, and modes of interaction are consistent (or can be uncovered) within a particular timeframe. We employed personas—synthetic identities, each with their own backstory, motivation and skills—to research assistants tasked with scenario creation. The personas allowed us to create realistic data and reduced the possibility of accidentally introducing information associated with real identities.

4.1 Scenario Planning, User Roles, and Automation

We created an in-depth "game plan" to help us sequence all scenario events. This plan allowed us to ensure they occur at or near a specific time and allowed us to maintain realism. At the start of each day, the research assistants were given a set of notecards with specific numeric ordering and timing information. This out-of-band communication mechanism provided the assistants with details of which commands to execute, which URLs to visit, and tasks to perform such as sending an email message to another person. Research assistants logged the time that they completed a specific task, and these logs were combined to generate a complete timeline that is included in the corpus teaching materials. The timeline allows teachers to fine-tune in-class exercises and provides a gold standard for identification of activities within the scenario.

Storylines and day-to-day activities were developed following examination of both media accounts and the observation of actually criminal and malicious activities in real-world data. We also based the evidence that we created on the specific types of activity and data storage formats that investigators would uncover during an actual investigation.

Some of the scenario activity was automated via software scripting to provide additional depth to the data contained within the file system and support the illusion of real persons carrying out daily work and personal activities. Specifically, we wrote a program that would automatically generate web traffic according to previously fetched URLs. Careful planning of these scripts was important to ensure each persona remained "in-character" during the whole scenario–*e.g.,* visiting favorite websites repeatedly.

Scenarios with sufficient breadth and depth of planning as well as extensive user activity are valuable in a variety of contexts beyond introductory forensic education. Well-planned and executed scripts produce datasets that embody "ecological validity," can be adapted according to varying instructor needs, and–most fundamentally–reduce the burden on instructors to create their own datasets (a process that is both time-consuming and error-prone).

4.2 Secondary Data Sources

In normal computer crime situations, an incident response team will acquire many types of primary and secondary data in order to fully investigate the situation and report to law enforcement (Eoghan 2004). These data can include bit-identical copies of computer workstations and related computer systems; network packet captures showing suspect communication; central login records from authentication and authorization servers; email spool files; and DHCP lease records. Analysis and correlation of these heterogeneous data sources provides the fundamental basis for a case.

Construction of realistic data corpora allows us to enrich the data that would typically be captured in a real-world investigation with supporting materials that may be used by students to explore details of the scenario background; confirm or refute theories developed about how a particular action or event transpired; or develop experiments structured to test such theories. Supporting data can include network packet captures acquired from a scenario router, memory dumps, and snapshots of critical operating system components such as the Windows Registry. Additionally, while such data are not generally available in a real-world incident response scenario, "live" forensic data such as RAM dumps from running machines provides support for training in techniques that are not yet widespread in professional practice.

5. CORPUS CONTENTS, COLLECTION, AND METADATA

To support realistic computer forensic investigation training, we collected all of the data that would typically be gathered in a real incident response or investigation scenario. Each data component was cryptographically hashed, time stamped, and accompanied by annotations describing relationships within the data and specific criminal actions associated with particular times or data sources. Accidental deviations from the scenario (for example, a missed task) and equipment failures were logged; no attempt was made to artificially insert data into any part of the corpus after the fact.

In addition to the set of data that would typically be collected by an incident response team–and extracted from hardware in a laboratory after the fact–realistic corpora can be augmented with data collected during the execution of the scenario. The data may include disk images, RAM dumps, other device images, and network traffic collected on a day-to-day basis and at the termination of the scenario. Of course, most of this data would *not* be available in a real-world incident response scenario (Brown 2010). We include it to allow for the possibility of *student research projects.* In our experience, many students who attempt original research are overcome by the difficulty of collecting the data that they wish to analyze, and rarely get to the point of doing sophisticated analysis work by the end of a class. By collecting and providing this information, we believe that students interested in doing original research will be more likely to realize their goals.

6. THE "M57-PATENTS" SCENARIO

In the following sections we describe "M57-Patents," a realistic scenario and associated corpus designed for primary use in educational and training exercises. M57-Patents has been designed to closely replicate many of the properties of real-world data.

6.1 Scenario Details

In this scenario, "m57.biz" is a new patent search company that researches patent information for clients. The business of patent search is to generally verify the novelty of a patent before the patent is granted–or to invalidate an existing patent by finding prior art (proof that the idea existed before the patent). At the start of the scenario, the firm has four employees: The CEO and founder Pat McGoo, one IT administrator, and two patent researchers. The firm is planning to hire additional employees as new clients are booked. Since the company is looking to hire additional employees, they have an abundant amount of technology on hand that is not being used.

The role of each employee persona in the scenario was performed by an individual researcher at the Naval Postgraduate School (NPS). Basic activities performed during the scenario included checking and writing email; surfing the Internet; staging and carrying out a variety of malicious and/or "illegal" activities; and using office document creation and other software. Malicious activities appearing in the scenario include but are not limited to theft of company property; proprietary information exfiltration and extortion; use of spyware such as key loggers; and viewing illegal content. (For the purpose of the exercise the "illegal content" are non-copyrighted pictures of common house cats; they are meant to be a simulant of actual illegal content such as child pornography.)

The scenario terminates when police receive information from an individual outside of m57.biz who has purchased a desktop workstation from an advertisement on Craigslist. The purchaser found the aforementioned cat photographs. Investigators are able to trace the machine back to M57. When the police contact the CEO of M57 (Pat), Pat confirms that the hardware has been stolen, and provides a list of additional items stolen from the company inventory. Pat gives consent for the police investigators to search M57 and image all of the company computers, company phones, and removable USB drives. Pat also holds a meeting of his staff and tells them that the police are on their way.

6.2 Personas

The "M57-Patents" scenario includes four main personas representing the employees of the m57.biz company: the CEO (Pat McGoo), the IT administrator (Terry Johnson), and two patent researchers (Jo Smith and Charlie Brown). Unknown to McGoo, several of these individuals are involved in illegal activities including theft, extortion, data exfiltration, and collection and distribution of illegal explicit images.

Several other personas were created outside of the company to simulate real-world interactions. These personas represent friends, acquaintances, clients, and other individuals in contact with the M57-Patents employees. Their involvement included buying company hardware via Craigslist, purchasing exfiltrated patent information from within the company, and normal personal correspondence with the main scenario actors.

6.3 Timeline

The M57-Patents scenario took place within a 17-day period between November 16th and December 11th 2009. Within the scenario, a workday started at 9:00am and ended at 4:00pm. Each day was marked in the timeline by a small number of primary objectives to be completed by research assistants playing the company personas. In addition to these objectives, each persona performed some normal background activity–web browsing, emailing friends and co-workers, patent searches, and writing word-processing documents. Researchers used out-of-band communication to facilitate activity coordination within the lab. As previously mentioned, additional texture was provided through automated web-browsing scripts.

In addition to the scenario, a number of technical procedures were performed each day outside of the scenario. These included verifying that all objectives on the daily activity checklist had in fact been accomplished; confirming that the automation and network capture scripts were running; and making a disk image of each computer.

7. SCENARIO CONSTRUCTION

7.1 Network

The network for M57-Patents consisted of four computers connected to a single switch, which then was connected to a gateway providing a connection to the Internet. Jo required two computers in the scenario. Only one of Jo's computers was on the network at a time; the replacement was made due to Jo's original hardware "failing" a week into the scenario. (In the scenario, the computer does not actually fail, but Jo is told that it fails.) Figure 1 shows the network design used for the scenario.

7.2 Workstations and Devices

Workstations used in the M57-Patents scenario were prepared as clean environments. First, we purged the hard drives with a single pass of NULL characters over the entire hard drive of each machine. From this clean state, a single partition was created onto which the operating system was installed from original installation media. All of the hard disk images were formatted with NTFS. Once installed, the systems were updated via Windows Update.

Five other devices were used in the scenario and subsequently imaged: four USB drives and one cell phone. In the scenario, one of the USB drives is Jo's personal storage device, while the remaining three were are "work" drives belonging to M57. Control of at least one drive changed during the period of the full scenario. The cellphone was likewise used for personal purposes by one of the employees and plays a part in at least one of the criminal activities.

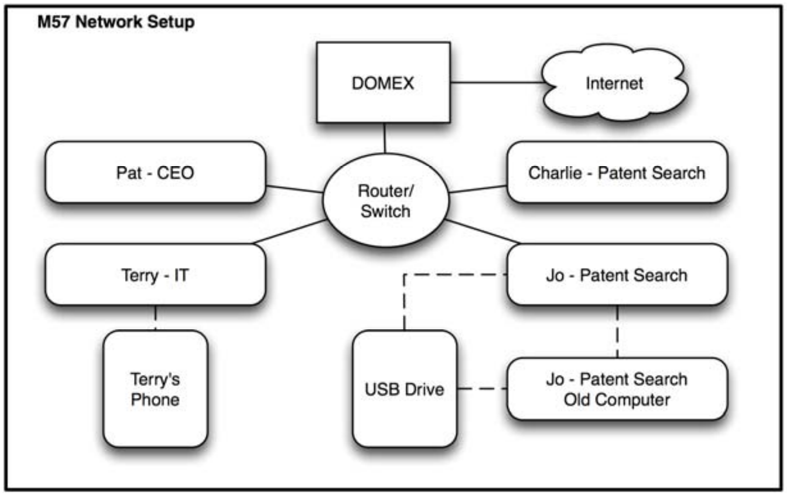


Figure 1: M57-Patents network setup, isolated through server (DOMEX). By isolating the network in this manner, it was relatively easily to tap all network traffic (at DOMEX) and to consistently route the scenario's email traffic.

7.3 Disk Images, Memory, and Network Captures

The workstation hard drive for each persona was imaged at the end of every workday (excluding weekends and holidays) using the *aimage* disk imager. At the end of the scenario the hard drives were imaged again. The disk images are stored in the Advanced Forensic Format (AFF) from which raw disk images can readily be extracted (Garfinkel et al. 2009a; Garfinkel 2009b).

RAM contents of each workstation were also captured daily, except for weekends and holidays. The contents of RAM were extracted using both *win32dd* and *mdd*. We provide both versions for download.

Four USB devices and one cell phone used during this time were imaged once at the end of the scenario. The USB devices are stored in AFF as well as RAW format. The cell phone contents were imaged via the SIM card. This method was feasible since–at the beginning of the scenario–the phone's settings were altered to store all of the non-multimedia data to the SIM card.

A network tap was placed on the gateway's interface using *tcpdump*. Data was collected every day the scenario was in operation, including weekends and any holiday that occurred during the scenario. The *tcpdump* script produced a daily *tcpdump* file with dmp file extension. The network capture dumps are currently available as single-day downloads as well as within a package containing capture data for every day of the scenario.

8. DISTRIBUTION CONSTRAINTS

The M57-Patents scenario is intended for free, public distribution. Because of this, a fundamental goal of the design and implementation was to remove instances of copyrighted material and personally identifiable data.

This section addresses copyright issues, scrubbing private information, answer key distribution, and the generation of simulated objectionable material.

8.1 Redacting Real-World Information

Separation of the scenario environment from the real world is difficult; in complex scenarios some real-world information inevitably seeps into the corpus either through user error, improperly configured services, or simply as a consequence of unforeseen issues inherent to the environment. As an example, in the M57-Patents scenario the workstations were connected to an isolated network using a local outgoing mail server (for emails between employee personas) that stamped each email with header information identifying the domain as *nps.edu*.

The personas in the M57-Patents scenario were performed by students and researchers acting out events in a pre-defined timeline. Although the prescribed events and business behavior was detailed, the actors may have accidentally engaged in activities outside of this detailed scenario. For example, at least one researcher inadvertently logged into his personal email system via a web browser. A process was put in place for any team member who introduced this type of information into the scenario to create a detailed report of the occurrence–time, site visited, and other information that would help scrub the information. At the conclusion of the scenario we scanned for a number of identifiers including the usernames and email addresses of all of the researchers. When these were found, we excised the TCP streams from the network captures and examined the hard drives to determine if the information had been recorded (it was not).

8.2 Simulated Objectionable Material

A significant asset of the M57-Patents corpus is the ability to simulate objectionable material–such as child pornography–without exposing users to actual illegal content. M57-Patents simulates such material by using images and videos of cats. The simulated material consists of 43 images and four movie files. This source material appears at various resolutions and is present in several pieces of media obfuscated with various methods. In addition, the simulated contraband is distributed in a hash database called the “Monterey Kitty” hash set. This hash set can be used with existing commercial utilities such as EnCase and FTK to automatically locate objectionable material (Guidance Software 2010; Access Data 2010).

Because a goal of the scenario design was to keep all contained information free of commercial and otherwise license-restricted media assets, the images and videos for this set were created from scratch by the researchers.

9. DISTRIBUTION AND ACCESS

9.1 Annotation, Sharing, and Publication

The M57-Patents scenario corpus provides numerous benefits to forensics research—and especially student research. It allows for the M57-Patents data to be published and publicly shared in a variety of forms, since it does not contain private or legally sensitive information. Published research using the M57-Patents corpus can be validated and reproduced, because the data is freely available. Because the data is already collected, students can spend their time developing new forensic approaches, rather than collecting data. Finally, dataset annotations distributed along with the disk images simplify familiarization with the corpus, development of classroom materials, and identifying and extracting data relevant to specific actions within the scenario.

9.2 Distribution

The M57-Patents corpus is currently available for download from the main corpus portal at digitalcorpora.org.[4] Individual workstation images (in AFF and RAW formats), RAM dumps (captured both by *mdd* and *win32dd*), and network captures can be downloaded directly from the site via a calendar link map. Because these materials are relatively large (more than 400GB for the full corpus), we have provided a peer-to-peer option for acquisition and sharing of the data between researchers and educators via BitTorrent files with permaseeds at iBiblio at the University of North Carolina, Chapel Hill. This facility allows us to create customized "views" into the raw corpus that can be downloaded as single packages depending on the needs of the organization or individual. We provide torrents for each set of workstation drive images captured during the scenario, the full set of RAM dumps, the full set of network packet captures, a "police evidence" torrent consisting of only those materials that would typically be collected during incident response, and a torrent linking the entire corpus.

9.3 Annotations, Timeline, and Answer Keys

In addition to the drive images and other raw M57-Patents data, a set of annotations, answer keys, and a full scenario timeline are available to provide background support for the scenario, detail the planning and execution of each criminal action, and provide a master reference for the events during each scenario day. The annotations include some materials to enhance the realism of the scenario and frame the process of the investigation. These include four detective reports prior to and including seizure and imaging of the M57 hardware; a search warrant and affidavit (modeled after real warrants issued in the state of California), and an informal report which can be distributed to students describing the employees of the M57 company and layout of the company's IT infrastructure.

[4] http://digitalcorpora.org/corpora/scenarios/m57-patents-scenario

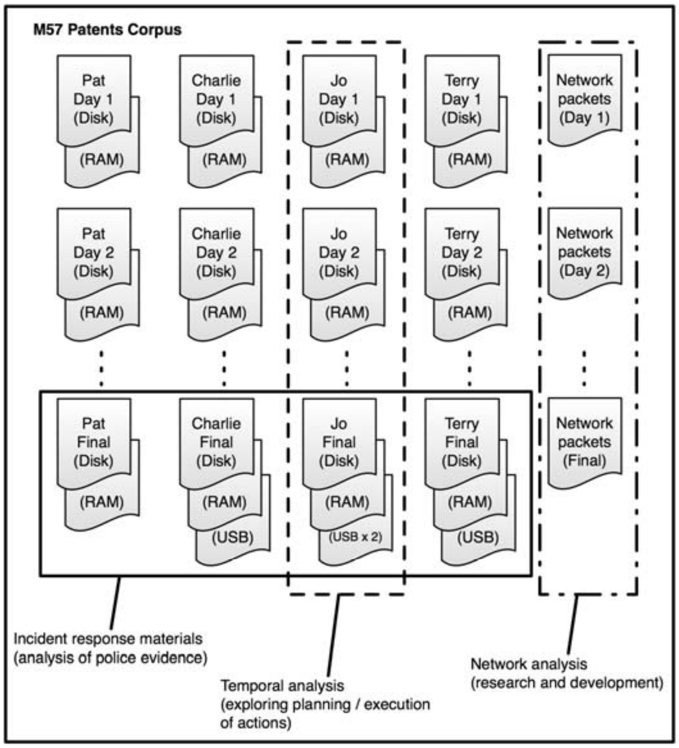


Figure 2: Overview of M57-Patents materials extracted during execution of the scenario. Various educational objectives and levels of analysis can be supported with "slices" of the scenario materials.

The full scenario timeline details (by day and time) any criminal acts carried out by employees including theft, exfiltration of data, extortion, and possession of illegal digital materials. Individual reports are available for each of these activities that further elaborate on the process, in particular providing paths within the disk images to relevant files, messages, software installations, and deleted content. Contents of the employee email accounts are provided as separate text files. Finally, a collection of the simulated illegal images is provided along with MD5, SHA1, and SHA256 hash tables to support various educational exercises.

The distribution of the answer keys is a primary concern for every forensic educator. The M57-Patents answer key is available for download only in encrypted form. The passphrase can be distributed to known professional and academic educators on request. Additionally, the educator must demonstrate that he or she is an educator, professor, or some other individual involved in the teaching of forensics material. This process will not prevent every student from obtaining an answer key (with sufficient effort), but it does introduce a reasonable barrier against cheating.

9.4 Copyright Issues

A clear concern in distributing the drive images is, "Can we legally distribute drive images that contain copyrighted files?" In particular, the M57-Patents data sets contain binaries from Microsoft Windows XP and Microsoft Windows Vista operating systems. For this corpus, Microsoft executables and libraries were disabled in the publicly available images by altering data at the start of the bitstream. While these images cannot be mounted as live workstations, this form of redaction has little to no effect on common methods of investigation using commercial or open source tools. For the end-user who has a MSDNAA license, non-redacted images can be provided upon receipt of the license. During production of the corpus, researchers were likewise careful to avoid downloading rights-restricted digital media such as music, videos, photos, or commercial software. For example, instead of using Microsoft Office, the fictional M57 company uses Open Office.

Later we plan to distribute a tool that can replace the redacted data in the disk images, allowing us to minimize the size of data that must be archived.

10. LESSONS AND FUTURE WORK

The lessons learned from creating the M57-Patents corpus–particularly in terms of handling accidental pollution of the dataset with personal identifiable information, legally encumbered data, and other sensitive materials–are informing our data creation methodologies for additional corpora. Future work will build on these lessons and will be documented to guide other researchers and educators who wish to create their own datasets.

One seemingly simple but deeply important lesson of this work concerns the day-to-day recording of scenario activity and any deviations from the planned timeline. These records provide a ground truth for the finalized timeline and reduce the likelihood of mismatches between scenario answer keys and what is actually found in the data. Understanding that human error and hardware failures are both likely in extended scenarios allows us to build a degree of flexibility into the initial scenario and plan for minor redactions (which we can reliably perform) rather than extensive manipulation of the data after the fact (a process that is error-prone and may further contaminate the data).

In addition to these issues, we are examining ways to enrich realistic corpora with additional activities and related records, including in-scenario communications with third parties or partner organizations, more complex and nuanced personas engaging in more of the kinds of everyday activities performed by real people (e.g. use of social media services), and more finely-grained records of run-time data from scenario host systems (e.g. process listings, network connection logs, and changes to Registry key settings). While some of this data is available in existing corpora, systematizing the process of its collection and organization will streamline the creation of educational materials and allow instructors to focus more efficiently on areas of interest within the data.

11. CONCLUSION

Realistic corpora provide an effective means to improve forensics education. Through careful design and implementation, corpora such as M57-Patents include data with sufficient depth and complexity to support a wide variety of classroom activities without the "noise", legal encumbrances, and privacy issues associated with real-world datasets. The mechanisms we have described here produce controlled environments that are designed to feel organic rather than contrived; can be quickly assessed with the existing timelines and answer keys; and support sharing and discussion among forensics educators.

Efficient compression and packaging of the corpus simplifies distribution and reduces storage overhead for instructors. The set of "police evidence" materials associated with M57-Patents–those materials that would be captured by an incident response team–is just over 40GB in size. Most of the scenario tasks can be investigated using just this data. Daily disk images provide further mechanisms for temporal analysis and evolution of the file systems, and the corpus includes a plethora of data that can be analyzed using memory and network analysis tools.

Realistic corpora such as M57-Patents can be used for multiple purposes at a variety of complexity and difficulty levels–in undergraduate classrooms and lab, for training exercises, and to support further research and development of digital forensics tools and techniques.

ACKNOWLEDGEMENTS

This work has been supported at the Naval Postgraduate School and at UNC-CH by "Creating Realistic Forensic Corpora for Undergraduate Education and Research" (NSF Award DUE-0919593). We would also like to recognize the research team responsible for production and recording of the M57-Patents scenario: Adam Russell, Henry Gilbert, Stephen Lavelle, Kristian Kearton, Nate Phillips, John Judd, and Joshua Gross.

REFERENCES

Access Data. 'Forensic Toolkit (FTK) Computer Forensics Software.' http://accessdata.com/products/forensic-investigation/ftk. Accessed Feb 19, 2010.

Beebe, Nicole. "Digital forensics research: the good, the bad, and the unaddressed." Fifth Annual IFIP WG 11.9 International Conference on Digital Forensics. 2009.

Brown, Christopher. "Computer Evidence: Collection and Preservation." Charles River Media. Boston, MA. 2010.

Carrier, Brian. "File System Forensic Analysis." Addison-Wesley. Upper Saddle River, NJ. 2005.

Casey, Eoghan. "Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet (2nd ed.)." Elsevier Academic Press. Amsterdam, Netherlands. 2004.

Cohen, M.I. "PyFlag – An advanced network forensic framework." Proceedings of the 2008 Digital Forensics Research Workshop (DFRWS). http://www.pyflag.net. 2008.

Cohen, M.I., S. Garfinkel and B. Schatz. "Extending the Advanced Forensic Format to Accommodate Multiple Data Sources, Logical Evidence, Arbitrary Information, and Forensic Workflow." DFRWS 2009(a).

Digital Forensics Association. 'Formal education: college education in digital forensics.' http://www.digitalforensicsassociation.org/formal-education/. 2010. Accessed February 19, 2011.

Garfinkel, S. "Digital Forensics Research: The Next 10 Years." Proceedings of the 2010 Digital Forensics Research Workshop (DFRWS). 2010.

Garfinkel, S. "Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools." The International Journal of Digital Crime and Forensics, Volume 1, Issue 1, January-March 2009(b).

Garfinkel, S., P. Farrell, V. Roussev and D. Dinolt. "Bringing Science to Digital Forensics with Standardized Forensic Corpora." Proceedings of the 2009 Digital Forensics Research Workshop (DFRWS) 2009(c).

Guidance Software, Inc. 'EnCase Forensic.' http://www.guidancesoftware.com/forensic.htm. Accessed Feb 19, 2010.

Kirschenbaum, M. G., R. Ovenden and G. Redwine. "Digital Forensics and Born-Digital Content in Cultural Heritage Collections. Council on Library and Information Resources. Washington, D.C. December 2010.

AACSB-ACCREDITED SCHOOLS' ADOPTION OF INFORMATION SECURITY CURRICULUM

Linda Lau
Longwood University
Farmville, Virginia

Cheryl Davis
Longwood University
Farmville, Virginia

ABSTRACT

The need to professionally and successfully conduct computer forensic investigations of incidents has never been greater. This has launched an increasing demand for a skilled computer security workforce (Locasto, et al., 2011). This paper examines the extent to which AACSB-accredited universities located in Virginia, Maryland and Washington, D.C. are working towards providing courses that will meet this demand. The authors conduct an online research of the information security courses and programs offered by the 27 AACSB-accredited business schools in the selected area.

The preliminary investigation revealed that eight of the 27 participating universities did not offer any courses in cybersecurity, digital forensics, and information assurance. However, nearly 70% of the participating universities have included at least one or more information security courses in their curricula and some universities have implemented more extensive information security programs. This paper will describe the research methodology and results of the study.

Keywords: digital forensics, information assurance, cybersecurity, information technology, information security, computer security

1. INTRODUCTION

Technology has redefined the process of criminal and business investigations. Investigations can involve forensics, information assurance and cybersecurity. Computers are not only part of everyday activities but are also used in criminal activities. The need to professionally and successfully conduct computer forensic investigations of incidents has never been greater. Digital information is increasingly being used as evidence in criminal and civil cases. Law enforcement and security agencies are using digital forensics not only as a tool to solve cases but to prevent them. After the tragic terrorist events that unfolded on September 11, 2001, there has been an increase in the focus on security – at airports, immigration centers, and federal and government buildings. Cybersecurity has since become a major component of that security. In November 2010, WikiLeaks exposed secrets of the inner workings of the U.S. diplomats (Rayfield, 2010). This breach of security may have put some diplomats and intelligence professionals lives at risk. These events have not only dramatically changed the way we view security, they have increased our reliance on cybersecurity and they have drastically changed the way we live.

In this study, information security will include three areas: cybersecurity, digital forensics, and information assurance. Cybersecurity refers to the protection of information and property from unwanted computer behavior with the objective of allowing the information to remain accessible and productive to its intended users (Cybersecurity, 2011). Digital forensics is defined as the process of investigating and retrieving information from a variety of electronic devices, including computer hard drives, cell phones, file servers and e-mail servers (Duerr, et al., 2004). Information assurance is the field of practice focused on managing the risks associated with storing, processing, and transmitting information (Marchant, et al., 2009).

2. LITERATURE REVIEW

Security and privacy have become the most complex and pressing subjects of information technology. From the demands of government and homeland security to the nature of the information age itself, employers -- including the government -- are faced with serious challenges of how to obtain a reasonable balance with dwindling resources. Experts agree that obtaining this balance will be found in education as information technology plays an important role in modern education (Gong, Xu, and Yu, 2004). State and local governments are showing their support for reforms through the passage of Bills. In 2006, the Virginia General Assembly passed Senate Bill 494/House Bill 1307, requiring the Governor of Virginia to develop a statewide strategic plan to address the need for reforms in workforce policy, which includes the implementation of workforce development and training initiatives (Governor Kaine's Workforce, 2011). This Bill was passed to allow Virginia to build a skilled workforce able to compete effectively in the technological 21st century.

Over the past decade, compared to the national average, fewer and fewer working-age adults in Virginia are continuing with their higher education and/or upper level training (The National Center for Public Policy, 2006). On the other hand, the Occupational Outlook Handbook predicted that the job outlook is very favorable for those in computer security (2006). However, the demand for computer security skilled professionals is much greater than the supply. The 2006 Occupational Outlook Quarterly stated that employees in the diverse field of computer security typically work very long and irregular schedules. This could be a direct result from not having enough universities offering programs to train skilled employees needed to meet the demand of employers. In February 2009, President Obama ordered a 60-day review of the federal government's various cybersecurity programs which have set the stage for a substantial overhaul of government's cybersecurity activities as well as new legislation for data protection and security breach notification (Vijayan, 2009). The Cybersecurity Enhancement Act of 2009 will provide up to $396 million in research grants over the next four years to develop best practices and standards to protect computer networks (Montalbano, 2011).

A Washington Post article highlighted the need for a dramatically different approach to cybersecurity education, outreach, as well as the hiring by the federal government (Cyber Help Wanted, 2009). This need is further complicated by the fundamental discrepancy between the users and employers' expectations, the scarce work force, and the underdeveloped educational mechanism (Locasto, et al., 2011). Because cybersecurity, digital forensics, and information assurance are constantly evolving fields, universities must offer programs that promote life-long learning in these areas.

The Association for Computing Machinery (ACM) *IS 2010 Curriculum Guidelines for Undergraduate Degree Programs in Information Systems* is a model curriculum intended to provide flexibility in designing Information Systems (IS) curricula to satisfy various local requirements. IS faculty may be affiliated with schools of business, schools of public administration, schools of information science or informatics, stand-alone schools of Information Systems, or other variations (Topi, et al., 2010). This flexibility also fuels an ongoing debate regarding the nature and identity of information systems as a discipline. The ACM guidelines suggested that universities should offer information security courses across campuses. Unfortunately, the interdisciplinary content and complexity of the information security courses require instructors to possess appropriate training in diverse contents in the field of information security (Shing, et al., 2007). A 2006 research concluded that although several entities in this country offered various certificate programs, these certifications provided limited knowledge and skills that may not be sufficient for employers (Hentea & Dhillon, 2006). The third largest reason for the high turnover of IT security employees is due to the fact that they were inadequately trained and ill-prepared for the jobs (Furnell & Clarke, 2005). A case study conducted in 2004 revealed that programs in fields such as computer science and information technology lack an emphasis on security issues in their curriculum (Bogolea & Wijekumar, 2004). A Web-based survey collected data from IS faculty members in several business colleges (Foltz & Renwick, 2010). Sixty-one instructors completed the survey, 50 of the completed surveys came from AACSB-accredited business colleges. A strong majority (73%) of the respondents indicated that IS security needs to be addressed and that the present curricula are not meeting those needs, especially in the required courses.

Current literature revealed two main concerns with the current workforce. First, there is an employer demand for a computer security skilled workforce, and this demand for computer security skilled professionals is much greater than the market can supply. Although universities play a vital role in providing this skilled workforce, there is a shortage of universities offering technology programs to meet the demand of employers. Further, there is no existing benchmark to measure the quality of the current programs. Hence, this paper will examine the information security curricula at AACSB-accredited universities located in Virginia, Maryland, and Washington, D.C.

3. RESEARCH STUDY

This section of the paper will describe the research methodology used to collect the data needed for this study. Twenty-seven universities were selected as participants for our research. The research data were collected via the Internet, summarized using Excel 2007, and the results are discussed in the Data Analysis subsection. Research limitations that may affect the validity of this research and topics for future research are also presented in this section.

3.1 Research Methodology

This research explores the information security programs offered by 27 universities located in Virginia, Maryland, and Washington, D.C. These universities are selected based on their AACSB-accredited business programs (Accredited Institutions, 2011). As of March 2011, there were 16 AACSB-accredited universities in Virginia (with two business colleges in University of Virginia), seven in Maryland, and four in Washington, D.C. The authors visited each university's Web site and performed a comprehensive search at each Web site using keywords such as cybersecurity, forensics, digital forensics, and information assurance. This online search documented pertinent information regarding the information security courses and programs, such as the field in which the courses are offered, the number of credits for each course and/or program, and the departments/schools offering courses and programs. The search results were collected, summarized, and tabulated in tables.

3.2 Data Analysis

Table 1 showed that two of the 16 Virginian universities offered at least one information security course. However, seven Virginian universities do not offer any information security courses and another seven of the Virginian universities offer some sort of information security programs. In the state of Maryland, one university does not incorporate any information security courses into its curriculum, while four universities taught at least one course in the three selected fields, and two universities have a structured information security program. Finally, the District of Columbia housed two universities that offered at least one information course and two universities have a structured information security program. Of the 27 AACSB-accredited universities surveyed, nearly one-third of the participating universities do not offer any information security courses and another one-third of them offer at least one information security course. The remaining 40% (11) have a formal structured program in this area.

Table 2 provided a more detailed description of the information security programs offered by the 11 universities: seven in Virginia, two in Maryland, and two in the District of Columbia. Of the seven Virginian universities with a more comprehensive information security agenda, four of them – James Madison, Norfolk State, Radford, and Virginia Commonwealth – have an undergraduate degree in various majors and concentrations. Three of them – George Mason, Norfolk State, and Virginia Commonwealth – have a master's degree in information security. Only two of these seven universities – Norfolk State and Virginia Commonwealth – offer both undergraduate and graduate degrees in information security. Three Virginian universities – George Mason, Longwood, and Radford – offer a minor in information security, and two of them – George Mason and Virginia Tech – offer graduate certificate programs in this area.

In Maryland, Towson University is heavily involved with the information security curricula, offering various undergraduate, graduate, and certification programs. It is also note worthy to mention that, of the 27 universities surveyed, only Towson University has established a Center of Excellence that is devoted to the education of information assurance. This sole establishment is known as the National Centers of Academic Excellence in Information Assurance Education (CAEIAE), and was approved by the National Security Agency and the Department of Homeland Security as a National Center of Academic Excellence (CAIT, 2011). University of Baltimore offers a bachelor degree in Forensic Studies. In the Capital, both George Washington University and Georgetown University offer master's degrees and certifications in the area of information security.

3.3 Research Limitations

The reliability and validity of this research depends on the accuracy of the information collected from the Internet during the research period, which is beyond the control of the authors. Further, the authors selected the participating universities based on one accreditation, AACSB. This accreditation was selected based on the authors' affiliation with teaching in an accredited business college. However, there are many other universities located in the three selected regions that offer courses and programs in information security whose curricula are approved by other types of accreditation. Unfortunately, the lack of resources delimited the number of universities that could be included in this study.

3.4 Future Research

The authors plan to continue with the current research. First, the authors plan to examine the formal structured information security programs in more details. For instance, pertinent information such as the number of credits needed for each program, the disciplinary area, the department and college offering the course, etc., will be collected, summarized, tabulated, and then analyzed further. The authors also intend to contact the participating universities to confirm the number of faculty who are teaching those courses, the number of students enrolled in those courses, as well as the date of creation of those courses. If more resources are available, the authors will increase the sample size to include AACSB-accredited universities in neighboring states such as West Virginia, Pennsylvania, Delaware, North Carolina, and South Carolina.

4. CONCLUSION

This research provided some insight into the information security curricula offered at 27 AACSB-accredited universities in Virginia, Maryland, and Washington, D.C. The conducted research supports the concerns found in the literature review, mainly: (1) There is a shortage of universities offering information security programs; and (2) There is a lack of benchmarks used to measure the quality of the current programs being offered. Only one of the 27 universities surveyed has established a Center of Excellence for information security programs. We would like to see more universities establishing their own centers of excellence and utilizing the federal and states monies set aside for the development of best practices for computer security programs.

REFERENCES

Accredited Institutions. Retrieved from http://www.aacsb.edu/ on January 3, 2011.

Bogolea, B. & Wijekumar, K. (2004). Information security curriculum creation: A case study. Kennesaw, GA, InfoSecCD Conference, October 8, 2004.

Center for Applied Information Technology (CAIT): Information Assurance Resources. (2011) Retrieved from http://www.towson.edu/outreach/cait/informationAssurance/ on March 25, 2011.

Cybersecurity. (2011). ***Wikipedia, the Free Encyclopedia***. Retrieved from http://en.wikipedia.org/w/index.php?title=Computer_security&oldid=414432823 on February 17, 2011**.**

Cyber Help Wanted: The federal government lacks a sensible hiring process – and enough good candidates – to guard computer networks. (August 1, 2009). ***Washington Post***, p. A16. Retrieved via Greenwood Library LexisNexis database on March 25, 2011.

Duerr, T., Beser, N., and Staisiunas, G. (2004). Information assurance applied to authentication of digital evidence. (Research and Technology) ***Forensic Science Communications***, October 1, 2004. Retrieved from http://www.highbeam.com/doc/1G1-137921545.html on February 21, 2011.

Foltz, C., & Renwick, J. S. (2010). Information Systems Security and Computer Crime in the IS curriculum: A detailed examination. ***Journal of Education for Business***, 86(2), 119-125.

Furnell, S. & Clarke, N. (2005). Organizational security culture: Embedding security awareness, education, and training. ***Proceedings of the IFIP TC11 WG 11.8**, **4th World Conference Information Security Education***, Moscow, Russia, 4: 213-222.

Gong, M., Xu, Y., & Yu, Y. (2004). An enhanced technology acceptance model for Web-based learning. ***Journal of Information Systems Education***, 15(4): 365-374.

Governor Kaine's Workforce Development Strategic Plan. (2011). Making connections: Virginia's new direction for workforce development. Filed on February 12, 2011 with the www.nationalskillscoalition.org. Retrieved on March 25, 2011.

Hentea, M., & Dhillon, H. (2006). Towards changes in Information Security education. ***Journal of Information Technology Education***, 5: 221-233.

Locasto, M., Ghosh, A., Jajodia, S., and Stavrou, S. (2011). Virtual Extension The ephemeral legion: Producing an expert cybersecurity work force from the air. ***Communications of the ACM***, 54(1): 129-131.

Merchant, R., Cole, R., and Chu, C. (2009). Answering the need for information assurance graduates: A case study of Pennsylvania State University's security and risk analysis major. ***Information Systems Education Journal***, 7(75): 3-11.

Montalbano, Elizabeth (February 4, 2010). Cybersecurity bill calls for research, task force. ***Information Week***. Retrieved from http://www.informationweek.com/news/government/policy/showArticle.jhtml?articleID=222601110 on March 25, 2011.

Shing, Marn-Ling, Shing, C., Chen, K., and Lee, H. (2007). Issues in information security curriculum: Collaborative learning and team teaching. ***International Journal of Innovation and Learning***, 4(5): 516-529.

The National Center for Public Policy and Higher Education. (2006). Virginia, Measuring Up 2006: The State Report Card on Higher Education, p. 1-16.

Rayfield, J. (November 2010). ***Gibbs On WikiLeaks: Stealing and Disseminating Classified Info is a Crime.*** Retrieved from http://tpmdc.talkingpointsmemo.com/wikileaks/2010/11/ on February 18, 2011.

Topi, H., Valacich, J, Wright, R., Kaiser, K., Nunamaker, J., Sipior, J., and Vreede. G. (2010). IS 2010 curriculum guidelines for undergraduate degree programs in Information Systems. Joint IS 2010 curriculum task force-Association for Computing Machinery (ACM) and Association for Information Systems (AIS).

Vijayan, J. (May 29, 2009). Obama's cybersecurity plan gets cautious praise. ***ComputerWorld***. Retrieved from http://www.computerworld.com/s/article/9133687/Obama_s_cybersecurity_plan_gets_cautious_praise on March 26, 2011.

Table 1 Information Security Courses and Programs at 27 Universities

	No information security course	Offers 1 or > information security courses	Has an information security program	TOTAL
	Virginia			
	Christopher Newport University	Virginia Military Institute (2)*	George Mason University	
	College of William and Mary	Virginia State University (1)	James Madison University	
	Old Dominion University		Longwood University	
	Shenandoah University		Norfolk State University	
	University of Richmond		Radford University	
	University of Virginia**		Virginia Commonwealth University	
	Washington and Lee University		VPI and State University	
TOTAL	7	2	7	16
	Maryland			
	Salisbury University	Frostburg State University (1)	Towson University	
		Loyola University Maryland (5)	University of Baltimore	
		Morgan State University (5)		
		University of Maryland (2)		
TOTAL	1	4	2	7
	District of Columbia			
		American University (2)	The George Washington University	
		Howard University (1)	Georgetown University	
TOTAL	0	2	2	4
	8	8	11	**27**
	29.63%	29.63%	40.74%	

* Number in parenthesis indicates the number of courses.

** University of Virginia has two business schools - Darden and McIntire.

Table 2 Information Security Programs and Certifications at 11 Universities

	Undergraduate	Graduate	Minor	Certifications
Virginia				
George Mason University		Forensic Science, MS Computer Forensic, MS ISA, MS Computer Science, BS/ISA, Accelerated MS IT, BS/ISA, Accelerated MS IT, PhD, concentration in ISA	Forensic Science	Forensics, Graduate Certificate Telecommunications Forensics and Security, Graduate Certificate Forensic Nursing, Graduate Certificate ISA, Graduate Certificate
James Madison University	Pre-Professional Health Programs/Pre-Forensic Studies in Forensic Biology, Forensic Chemistry, or Forensic Anthropology			
Longwood University			Cyber Security, Forensics, and Policy	
Norfolk State University	Computer Science-Information Assurance, BS	Computer Science-Information Assurance, MS		
Radford University	Chemistry /Concentration in Forensics, BS Anthropological Sciences/Concentration in Forensic Anthropology, BS or BA		Forensic Science	
Virginia Commonwealth University	Forensic Science, BS	Forensic Science, MS		
VPI and State University				Information Assurance Engineering, Graduate Certificate
7	4	3	3	2

Table 2 Information Security Programs and Certifications at 11 Universities (cont'd)

	Undergraduate	Graduate	Minor	Certifications
Maryland				
Towson University	Forensic Chemistry Major/General Forensic Science Track	Forensic Science, MS		ISA, Certificate
University of Baltimore	Forensic Studies, BS			
2	2	1	0	1
District of Columbia				
The George Washington University		Forensic Sciences, MS, concentrations: crime scene investigation, forensic chemistry, forensic toxicology, forensic molecular biology, high-technology crime investigation		Forensic Investigation, Graduate Certificate
Georgetown University		Professional Studies in Technology Management/ Information Security/Information Assurance Track, MS		Forensic Accounting, Certificate
2	0	2	0	2

KINDLE FORENSICS: ACQUISITION & ANALYSIS

Peter Hannay
SECAU
School of Computer and Security Science
Edith Cowan University
Perth, Australia
p.hannay@ecu.edu.au

ABSTRACT

The Amazon Kindle eBook reader supports a wide range of capabilities beyond reading books. This functionality includes an inbuilt cellular data connection known as Whispernet. The Kindle provides web browsing, an application framework, eBook delivery and other services over this connection. The historic data left by user interaction with this device may be of forensic interest. Analysis of the Amazon Kindle device has resulted in a method to reliably extract and interpret data from these devices in a forensically complete manner.

Keywords: forensics, digital forensics, kindle, mobile, embedded, ebook, ereader

1. INTRODUCTION

The Amazon Kindle eBook reader provides significant functionality aside from that of simply reading eBooks. As the Kindle is an embedded computing platform it is possible to deploy a wide range of functionality due to the use of general computing hardware (see Table 1 for details). The Kindle platform has grown to include a web browser, which utilizes an inbuilt cellular data connection, an application framework, music player, image viewer, AGPS and numerous other capabilities. The presence of this functionality leads to a situation where the ability to provide forensic analysis of these devices would be quite desirable due to the potential for nefarious use of such features.

Table 1 - Comparison of Kindle Hardware (Amazon, 2010)

Kindle Specifications					
	Kindle	Kindle 2	Kindle DX	Kindle DX 2	Kindle 3
CPU	Freescale 532 MHz, ARM-11	Freescale 532 MHz, ARM-11	Freescale 532 MHz, ARM-11	Freescale 532 MHz, ARM-11	Freescale 532 MHz, ARM-11
Flash	256MB	2GB	4GB	4GB	4GB
Comms	Cellular/3G	Cellular/3G	Cellular/3G	Cellular/3G + WiFI	Cellular/3G and/or WiFi
Kernel	Linux-2.6.26	Linux-2.6.26	Linux-2.6.26	Linux-2.6.26	Linux-2.6.26

The 2GB of flash storage is divided into four file systems (see figure 1), the last of these is mapped to act as a USB mass storage device and is the only file system that can be accessed, viewed or in any other way interacted with when the kindle is in its secure state. The other three partitions contain the root Linux file system, configuration files and a debug file system respectively.

```
$ fdisk kindle.img
Disk: kindle   geometry: 995/64/63 [4014080 sectors]
Signature: 0xAA55
         Starting       Ending
 #: id  cyl  hd sec -  cyl  hd sec [     start -       size]
--------------------------------------------------------------
----------
*1: 83    0   1   1 - 1023   3  16 [        16 -     819248]
Linux files*
 2: 83 1023   3  16 - 1023   3  16 [    819264 -      49152]
Linux files*
 3: 83 1023   3  16 - 1023   3  16 [    868416 -      16384]
Linux files*
 4: 0B 1023   3  16 - 1023   3  16 [    884800 -    3129280]
Win95 FAT-32
```

Figure 1 - Partition Structure of the Kindle

Existing digital forensics software packages have implemented limited support for Kindle devices, however there are is currently no support for examination of the flash memory other than the FAT32 partition (MacForensicsLab, 2010). In the same vein research has been performed by a number of individuals in an attempt to derive forensic methodology for the Kindle, however this research has also only focused on the FAT32 partition exposed as a USB mass storage device (Huber, 2010b; Hughes, 2010; newinforensics, 2010).

2. SECURITY

The Kindle utilizes a firmware update mechanism that allows for over the air (OTA) or manual updates. In the case of both the update file is placed in the root of the mass storage portion of the file system. The update is then applied once the user activates this functionality from the system menu of the device.

The update files themselves are essentially signed TAR archives, these are extracted and a shell script contained within executed to facilitate the update functionality. The signing mechanism relies on RSA encryption in which the update is signed with amazon’s private key and verified with amazon’s public key, which is pre-installed on the Kindle device (Hannay, 2010).

The security functionality can however be defeated as the tar archive is extracted prior to signature verification. The most commonly employed exploit to leverage this involves setting the absolute path to the public key store in the tar archive, as such prior to signature validation a new public key is added to the store. The result of this exploit is that the ability to sign arbitrary updates is gained. The jailbreak process described here is illustrated below in Figure 2.

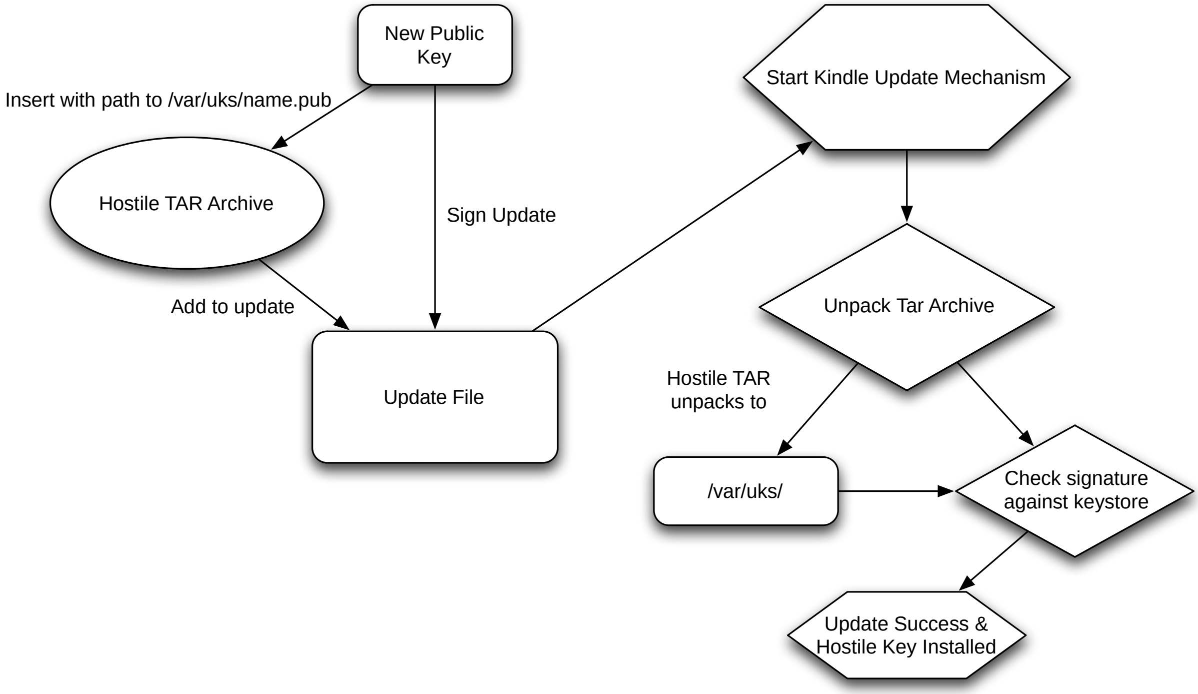

New Public Key
Insert with path to /var/uks/name.pub
Hostile TAR Archive
Sign Update
Add to update
Update File
Start Kindle Update Mechanism
Unpack Tar Archive
Hostile TAR unpacks to
/var/uks/
Check signature against keystore
Update Success & Hostile Key Installed

Figure 2 - Illustration of Jailbreak Process

3. ACQUISITION METHODOLOGY

Prior to commencement of this section it is important to note that knowledge of best practice in terms of hashing, evidence preservation and documentation are assumed and as such are out of scope of this paper. The investigator should ensure that he/she understands the impact that writing data to a device can have and the implications on forensic integrity.

In order to accomplish the acquisition and analysis of the Kindle we must first gain access to the device beyond what is available by default. This access is achieved through use of the exploit identified in the previous section, the implementation we will be using in this example is the Kindle Jailbreak (based on AVNard's earlier work), this utility includes a standard public/private key pair which is known publicly as well as an installation framework (NiLuJe, 2010). At this stage in the process we now have the ability to install custom software via the update system.

In order to gain complete access to the device it is necessary to install some form of remote access software on the device. In our case a telnet & SSH server will be installed along side scripts which allow for the USB port to be remapped as a USB Ethernet Gadget. The package commonly used to achieve this is the "USBNetwork" package, so named as it restores the USB networking functionality that was originally present in early versions of the Kindle firmware (NiLuJe, 2010). Once this has been accomplished it is possible to establish to start the USBNetwork service by issuing the ";debugOn" and "`usbNetwork" commands on the device (without quotes) as shown in Figure 3.

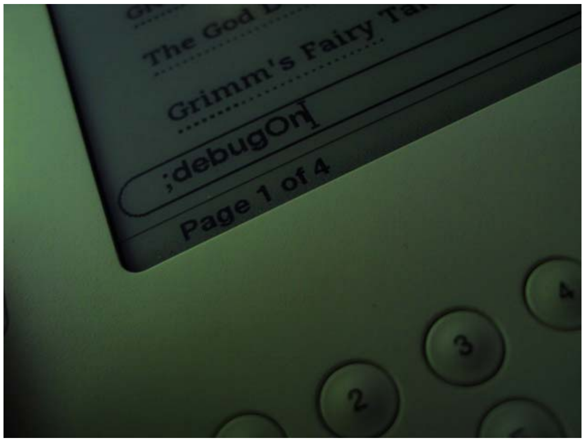


Figure 3 - The ";debugOn" command being issued

Once the USBNetworking package is installed and enabled it is possible to start acquisition. This is accomplished through the use of telnet, dd and netcat, this methodology has been commonly implemented in live system acquisitions (Burdach, 2005). In this configuration the host system is configured to listen for the data transmission, piping the output to dd. Then a telnet connection is established to the kindle and data transfer initiated, this process is shown in Figure 4 below.

```
1. Connect to kindle
$ telnet 192.168.2.2
Trying 192.168.2.2...
Connected to 192.168.2.2.
Escape character is '^]'.
 [root@kindle root]#

2. Listen for connection on host sytem
$ nc -l 55555 | dd of=kindle.img
3185454+1385484 records in
4014080+0 records out
2055208960 bytes transferred in 915.234125 secs (2245555
bytes/sec)

3. Initiate transfer of data from kindle
[root@kindle /dev]# dd if=/dev/mmcblk0 | nc 192.168.2.1 55555
4014080+0 records in
4014080+0 records out
```

Figure 4 - Acquiring image of NAND memory

Once this acquisition is complete it may be desirable to split this file into the four file systems that are contained within. The details of these can be extracted using fdisk as shown in Figure 1. Once these partition boundaries are known we can extract the individual partitions into their own files for subsequent analysis as shown in Figure 5.

```
$ dd if=kindle.img of=kindlep1.img skip=16 count=819248
819248+0 records in
819248+0 records out
419454976 bytes transferred in 23.742699 secs (17666693
bytes/sec)
$ dd if=kindle.img of=kindlep2.img skip=819264 count=49152
49152+0 records in
49152+0 records out
25165824 bytes transferred in 1.661936 secs (15142477
bytes/sec)
$ dd if=kindle.img of=kindlep3.img skip=868416 count=16384
16384+0 records in
16384+0 records out
8388608 bytes transferred in 0.315741 secs (26568018
bytes/sec)
$ dd if=kindle.img of=kindlep4.img skip=884800 count=3129280
3129280+0 records in
3129280+0 records out
1602191360 bytes transferred in 141.444850 secs (11327322
bytes/sec)
```

Figure 5 - Splitting disk image into individual partition images

The completion of this splitting leads us to the point where these images can be analysed using traditional computer forensics methodologies. The next section includes information on the various file systems and location of data that has been deemed to be of forensic interest.

4. DATA OF INTEREST

Partition 1 (root file system)

Location	Description
/opt/wan/firmware/mt-3/version.dat	Firmware version indicator
/opt/amazon/ebook/config/	Configuration files
/opt/amazon/ebook/prefs/	Preferences files
/etc/uks/	Public key store, keys other amazon’s and the key created during jailbreak may indicate tampering

Partition 2 (/var/local)

Location	Description
/audio/	Audio settings
/eink/screen_saver_last	The a reference to the last screen saver image displayed
/java/prefs/cookies	Cookies used to uniquely identify this device to amazon. These are persistent.
/java/prefs/DevicePasswordData.pw	Password data for this device
/java/prefs/browser/bookmarks	Web browser bookmarks
/java/prefs/browser/cookie.dat	Web browser cookies (no cache is present, this may provide limited historical evidence of web access
/java/prefs/browser/settings	Web browser configuration
/java/prefs/com.amazon.ebook.booklet.reader/social-clipping/social-prefs	Credentials and accounts associated with social networking services (twitter, facebook, etc) that have been set up for use with the device
/java/prefs/com.amazon.ebook.framework	User settings including: country, timezone & WAN status
/java/prefs	Details of the user, kindle name & user name
/log/	Detailed logs of users interaction with the device, including time stamps
/wan/	Network configuration

Partition 4 (User file system – available via USB mass storage)

Location	Description
/documents	Books and other publications for consumption on device
/music	Music and other audio for consumption on device
/system/Search Indexes/	History of each search conducted on the device
/system/com.amazon.ebook.booklet.reader/reader.pref	Contains details of last book read, font size selected and dictionary currently in use.

Note: On all test systems partition 3 was zero filled. Based on investigation it has been determined that this area is used for diagnostic purposes and likely will not contain information outside of the development environment.

5. CONCLUSION

eBook devices such as the Kindle are gathering increased interest from the forensic community as they become increasingly popular. The included cellular data capability of the Kindle specifically may make it a candidate for nefarious purposes, as the there is no data cost associated with the global data service (Hannay, 2010). In addition to data functionality the inclusion of an application framework and development kit in beta release will only lead to increased use of the product for purposes that were once met by the traditional computing paradigm.

The initial efforts of the forensic community have focused on acquisition of only a portion of the internal storage of the device as this area is readily accessible as a USB mass storage device (Huber, 2010a, 2010b; Hughes, 2010; MacForensicsLab, 2010; newinforensics, 2010). This paper has gone beyond the existing methodologies and provided a mechanism for the acquisition of the complete internal NAND memory and analysis of same. In order for this result to be achieved however some data must be written to the device and in doing so there is the possibility of data being overwritten. However aside from invasive hardware based acquisition there are no current known techniques that would allow for complete acquisition without this approach.

Research into small and embedded device forensics is ongoing, with increased focus on complete acquisition of all relevant data from these systems, including flash storage, memory and data stored on individual microcontrollers.

6. REFERENCES

Amazon. (2010). Kindle Wireless Reading Device, Wi-Fi, Graphite, 6" Display with New E Ink Pearl Technology. Retrieved January 7th, 2011, from http://www.amazon.com/gp/product/B002Y27P3M?ie=UTF8&tag=10inchlaptop-20&linkCode=as2&camp=1789&creative=390957&creativeASIN=B002Y27P3M

Burdach, M. (2005). Digital forensics of the physical memory. *Warsaw University*.

Hannay, P. (2010). Hooray for Reading: Hacking the Kindle. Retrieved January 3rd, 2011, from http://openduck.com/2010/11/27/hooray-for-reading-hacking-the-kindle/

Huber, E. (2010a). Additional Thoughts on Kindle Forensics Retrieved January 19th, 2011, from http://ericjhuber.blogspot.com/2010/04/additional-thoughts-on-kindle-forensics.html

Huber, E. (2010b). A Cursory Look at Kindle Forensics. Retrieved January 19th, 2011, from http://ericjhuber.blogspot.com/2010/04/cursory-look-at-kindle-forensics.html

Hughes, A. (2010). Forensics Beyond the Hard Drive: Kindle 2 Logging. Retrieved Febuary 6th, 2011, from http://inforensics.vidocrazor.com/2009/06/26/forensics-beyond-the-hard-drive-kindle-2-logging/

MacForensicsLab. (2010). Forensic Imaging of the Amazon Kindle. Retrieved January 12th, 2011, from http://www.macforensicslab.com/ProductsAndServices/index.php?main_page=document_general_info&cPath=5_18&products_id=338&zenid=be461f672b245e5f78e3800158c920e5

newinforensics. (2010). Kindle 3G Wireless Reading Device - forensically speaking. Retrieved January 9th, 2011, from http://newinforensics.blogspot.com/2010/10/kindle-3g-wireless-reading-device.html

NiLuJe. (2010). Fonts & ScreenSavers hacks for Kindles Retrieved Janurary 2nd, 2011, from http://www.mobileread.com/forums/showthread.php?t=88004

FORENSIC ANALYSIS OF SMARTPHONES: THE ANDROID DATA EXTRACTOR LITE (ADEL)

Felix Freiling
University of Erlangen-Nuremberg
Germany

Michael Spreitzenbarth
University of Mannheim
Germany

Sven Schmitt
University of Mannheim
Germany

ABSTRACT

Due to the ubiquitous use of smartphones, these devices become an increasingly important source of digital evidence in forensic investigations. Thus, the recovery of digital traces from smartphones often plays an essential role for the examination and clarification of the facts in a case. Although some tools already exist regarding the examination of smartphone data, there is still a strong demand to develop further methods and tools for forensic extraction and analysis of data that is stored on smartphones. In this paper we describe specifications of smartphones running Android. We further introduce a newly developed tool – called ADEL – that is able to forensically extract and analyze data from SQLite databases on Android devices. Finally, a detailed report containing the results of the examination is created by the tool. The whole process is fully automated and takes account of main forensic principles.

Keywords: Android, Smartphones, Mobile devices, Forensics.

1. INTRODUCTION

1.1 Why Forensic Analysis of Smartphones is Relevant

In the recent years, smartphones became a very popular medium of communication. The associated communication market is one of the world's fastest growing markets [Gre10]. Among all cellular standards, GSM (Global System for Mobile communications) is the most widely used standard with 75% market share. It is used in 200 countries and has more than 1.2 billion users in over 630 mobile networks [Sch03].

As smartphones offer more and more diversity through a growing set of features, increasing amounts of sensitive data are created and stored on such devices. Through the ubiquitous use of smartphones, an increasing amount of such devices also becomes part of forensic investigations pursued by private organizations or law enforcement. These organizations need to be able to extract and analyze data that is stored on smartphones. Thus, there is a concrete demand for methods and tools that enable the execution of the before mentioned tasks in a forensically correct way. Furthermore, the rapid development of smartphone technologies makes it necessary to frequently scrutinize and adapt existing as well as develop new methods and tools for use in small scale digital device forensics.

1.2 The Case of Android

According to Gartner [Gar10], the global distribution of smartphone operating systems is as follows: Symbian (Symbian Foundation) currently spearheads the market with about 40 percent of market share followed by Android (Google) in the second and BlackBerry (RIM) in the third place, both with about 17 percent of market share. With about 15 percent iOS (Apple) makes it to the fourth place. The market share of Android grows with between 13 and 14 percent per year. According to an additional forecast by Gartner [Gar10] Android will be at the forefront of the worldwide mobile communications market by 2015. Therefore the Android platform is relevant for research in smartphone forensics.

1.3 Challenges of Forensic Investigations

While forensic analysis of standard computer hardware – like hard disks – has developed into a stable discipline [Car05], there is still much debate on techniques to analyze non-standard hardware or transient evidence. Despite their increasing role in digital investigations, smartphones are still to be considered non-standard because of their heterogeneity. Within all investigations it is necessary to follow basic forensic principles. The two main principles are:

1. Greatest care must be taken that evidence is not manipulated or changed.
2. The course of a digital investigation must be understandable and open to scrutiny. At best, the results of the investigation must be reproducible by independent investigators.

Especially the first principle is a challenge in the setting of smartphones since they employ specific operating systems and hardware protection methods that prevent unrestricted access to the data on the system.

1.4 Contributions

In this paper we give an overview over the problems that investigators are faced with when analyzing smartphones based on Android. Furthermore, we report on a prototype tool developed by the authors to perform a forensic analysis of Android smartphones. More specifically, we make the following contributions:

- We give an insight into the Android platform focusing on the specifics from the perspective of digital forensics.
- We discuss alternatives for extraction and analysis of data stored on Android devices.
- We present an overview of the SQLite data format, a central data format used in Android.
- We report on a prototype tool that we developed for forensic analysis of digital data stored in SQLite databases on Android devices.

1.5 Related Work

The paper written by Lessard and Kessler [LK10] as well as the talk of Hoog [Hoo09] describe the forensic analysis of Android smartphones on the example of creating memory images and analyzing those with the help of well-known tools like Access Data’s Forensic Tool Kit (FTK). This proposal basically refers to data carving and data recovery techniques. However, the SQLite databases that are found in the memory are not automatically parsed and the analysis of recovered data is handed over to the investigator. Mohindra steps up a quite similar environment like we do by explicitly dumping the SQLite databases from the device [Moh08]. But in contrast to our procedure, he manually analyzes the databases. Lee et al. make use of an interesting approach within their paper [LX10]; they use a prepared SD card on which they have placed an own forensic software (similar to the applications outlined in Section 3.1) in order to analyze the smartphones' data. By using an own SD card any change of data on the device is avoided. Databases are accessed with the help of Android system calls and SQL commands. In this case, an in-depth analysis of the SQLite data structures does not take place. Instead, the authors rely on the data delivered by the system.

1.6 Roadmap

This paper is organized as follows: We give an introduction to the structures of Android in Section 2. We then discuss alternatives for forensic data extraction of smartphones in Section 3. This is followed by an introduction to the SQLite file format, an important file format used in Android (Section 4). In Section 5 we present a new software solution that allows for automated examination of data stored in SQLite databases on Android devices while taking forensic principles into account. We conclude in Section 6.

2. OVERVIEW OF ANDROID

At the beginning of this section we will illustrate the Android version history and the market share of all versions active on the market in Table 1. Afterwards we will discuss a brief overview of the Android platform from a forensic point of view.

Version	Date	Market Share	Releases and used Hardware
1.x	September 2008 - October 2009	10 %	3 major releases (1.1, 1.5, 1.6) only used on smart phones
2.x	October 2009 - now	89 %	3 major releases (2.0, 2.2, 2.3) used on smartphones and tablets
3.x	January 2011 - now	1 %	only used on tablets

Table 1: Android version history and market share [Gog11]

Regarding Figure 1 the base of the Android platform is a Linux Kernel providing the necessary hardware drivers. The Dalvik Virtual Machine (DVM) is the core of the runtime environment. If an Android application is started, it runs in its own “sandbox” and with its own DVM. Although this costs extra resources it leads to more security and availability because applications do not share common memory. The application layer of Android accesses a plurality of fixedly implemented libraries, all deployed for operating required functionalities. Android provides several programming interfaces (APIs) which allow communication between applications and between the end user and applications. The top layer of the system is represented by the collectivity of applications. Within this layer, the interaction between humans and machines and the communication between applications take place. Each application makes thereby use of the underlying programming interfaces.

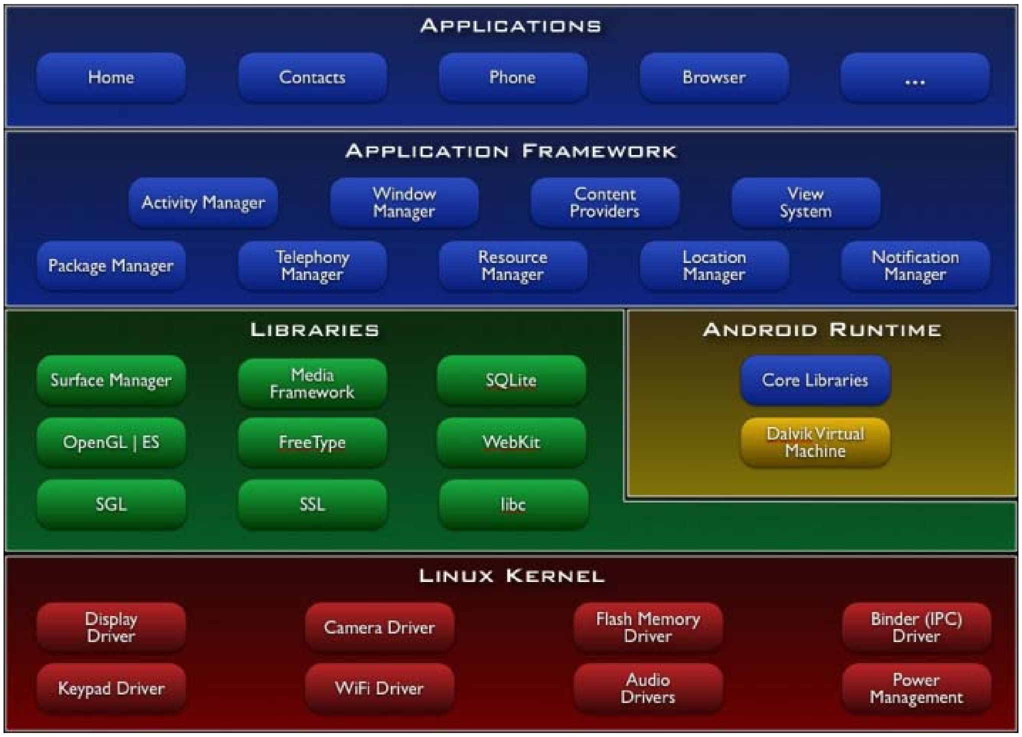


Figure 1: System architecture of Android [Gog10]

In addition to the execution in a virtual machine, applications that are run by Android are subject to several security mechanisms. They control the execution of any application and – if necessary – the access to data of other applications installed on the device. Meaning that when an application tries to interact with another database or application, the reference monitor, located in the application framework, looks at the permission labels assigned to this application and, if the target data access permission label is in that collection, allows the process to precede. If the label is not in the collection, establishment is denied. The security mechanisms consist of the following three main parts [And10]:

- At the Linux kernel level user and group IDs are assigned to an application and thus provide a kind of isolation from the rest of the file system.
- A fine-grained permission mechanism enforces restrictions regarding specific operations that a particular process is allowed to perform.
- The root access is disabled on smartphones running in production mode.

These security mechanisms hinder the forensic examination of data stored in Android devices, because adequate permissions are required to access the data. At the same time, access is only possible through the Android platform itself, which is potentially manipulated or can lead to unintentional changes in the file system. This again violates the forensic principle that examined data must not be changed. To overcome these limitations, we have re-established root access to the smartphone and directly interact with the file system through the Android Debug Bridge (adb). For a more detailed explanation please refer to Section 5.

3. ALTERNATIVES TO THE EXTRACTION OF DATA

There are two different approaches that principally exist regarding the forensic analysis of smartphone data. In the following, we will describe each while pointing out advantages and disadvantages.

3.1 Software-based Approach: Software Agents

In mobile phone forensics software agents are small programs that are installed on or copied to mobile phones to collect and analyze data locally or export data using the interface of the phone to examine it later. Two examples for forensic software agents are the “Open Source Android Forensic Agent” [San10] and “Panoptes” [Spr10]. Both agents need to be installed on the target Android device. After installation, they can be executed directly on the device and provide the investigator with CSV-files which contain the data extracted from the device in an already edited format. Through the use of “content providers” and the corresponding permissions, the sandbox of Android is broken and direct access to databases of other installed applications is granted. Due to this procedure stored and protected data from installed applications can be read.

Some advantages that result through the use of software agents are that little technical knowledge, no “rooting” of the device and no special hardware are required to read data from the device. With a software agent it is also possible to recover deleted data as long as it is still visible in the database files. However, a major disadvantage of agents is that data on the mobile phone is modified through copying or installing the agent and thereby a forensic principle is violated.

3.2 Hardware-based Approach: Desoldering Memory Chips

In the context of forensic investigations a common procedure is to remove any required memory chip from the circuit board of the device. Therefore, the memory chip is desoldered and then contacted through special hardware, such as PC-3000 Flash [Ace10]. The advantage of the hardware-based approach is that no intermediate layer potentially manipulates or prevents read access to the unaltered data on the chip. So, a high degree of forensic credibility can be assigned to the extracted data. However, one disadvantage of this method is the relatively higher effort that has to be made, compared to the software-based approach: Advanced technical equipment and knowledge are required for desoldering memory chips. When desoldering a chip from a circuit board, there also is a certain risk to damage or destroy the chip - and any potentially relevant digital trace with it. Nevertheless, this approach is a common method used in practice.

3.3 Using the Android Debug Bridge

To dump data from an Android device, it is possible to use the Android Software Development Kit (Android SDK). The Android SDK contains the Android Debug Bridge (adb) which is a client-server program that is able to connect to Android devices and execute a variety of commands on the connected device. Therefore an instance of the adb daemon must be running on the device which can be achieved by activating the option “USB debugging” on the target device.

As a further requirement, permissions on the device must grant access to the files that are to be dumped. Since permissions of Android devices in production mode deny the access to databases via the adb, one has to modify the firmware of the device or use a bootable “goldcard” in order to change the status of the phone in a way so that these security restrictions have been deactivated or can be bypassed.

Basically, this approach is a software-based approach but it is not necessary to inject new software into the smartphone. Furthermore, it requires some Android specific settings. Therefore it can be considered as an intermediate approach between software and hardware approaches. We use this approach later in this paper for our analysis tool ADEL.

4. THE SQLITE DATABASE FILE FORMAT

4.1 Why SQLite is Relevant for Smartphones

SQLite is a software library that implements a SQL database engine for embedded use. It can be integrated into other applications and – if necessary – be adapted to their specific requirements. One of the applications that make use of the SQLite software library is Android. It uses SQLite to store certain data on the underlying hardware device. This data contains information that is created by the user or by the OS, e. g. contacts, call lists, GPS data and SMS messages. Such data is of major interest within forensic examinations of mobile devices. This is why we took a closer look on how data is exactly stored by SQLite, which is defined by the SQLite database file format [SQL11]. This section will give a short introduction about important structures of the SQLite database file format. Each SQLite database is associated with a single file – the main database file – in the Android file system.

4.2 SQLite Internals

The main database file holds all of the data stored in the associated database. During the execution of database operations temporarily created files may additionally be used, e. g. to be able to rollback database operations after a power failure (rollback-journal). However, we will not discuss temporarily created files in this paper, due to the fact that most of the time all data is stored within the main database file. This file consists of one or more data blocks, called pages, with a well-defined size. The page size is a constant amount of bytes and is valid for all of the pages within the same database file. The leading 100 bytes of the first page of a database file are used to store the database header. It contains general information about the database file, e. g. the size of the database in pages. The information in the database header allows for accessing the remaining contents of the database in a structured way. For a detailed overview of each of the database header fields, refer to the official documentation [SQL11].

Each table of a database is internally represented by a single b-tree structure. Within such a b-tree structure interior pages and leaf pages may appear. While interior pages store pointers to either further interior pages or to leaf pages, the actual content of the database is exclusively stored in leaf pages. The first page of a b-tree is called the root page and may be either an interior or a leaf page. The first page of a SQLite database file represents the root page to a special b-tree structure that holds contents of the so called “sqlite_master” table. This table stores the complete database schema, including the SQL CREATE statements and pointers to the b-tree root page of each table in the database.

Each b-tree page (interior and leaf pages) is divided into different regions. In general those regions are:

- the page header,
- the cell pointer array,
- unallocated space,
- the cell content area.

There are only a few exceptions to the given order. One example is the first page of a database file that additionally stores the database header in the first place. Each page header holds general information about the layout of the page, e. g. the number of cells on this page. Immediately after the page header follows the cell pointer array. Each entry in the cell pointer array points to an offset on the same page at which the cell content is stored. For interior pages the cell content consists of two elements: the page number of the left child and a certain key value. For leaf pages each cell stores the content of a row – belonging to the corresponding table – and consists of four elements: the length of the payload in bytes, the SQLite internal ID of the row (rowID), the actual payload and an optional pointer to the first overflow page. Overflow pages are organized as linked lists and are used to store cell content that does not fit on the same page.

Database contents can thus be extracted by parsing the b-tree for each table contained in the database and extracting the contents of the cells found in any leaf page that belongs to the same b-tree (see Section 5.3).

5. ANDROID DATA EXTRACTOR LITE (ADEL)

We developed a tool named ADEL which is meant as an abbreviation of "Android Data Extractor Lite". ADEL was developed for versions 2.x of Android and is able to automatically dump selected SQLite database files from Android devices and extract the contents stored within the dumped files. In this section we describe the main tasks of ADEL and what steps the tool actually performs. However, there are conditions that must apply for ADEL to work correctly. These conditions are stated in the following sections, corresponding to the relevant tasks. A flow chart showing the structure of ADEL is depicted in Figure 2.

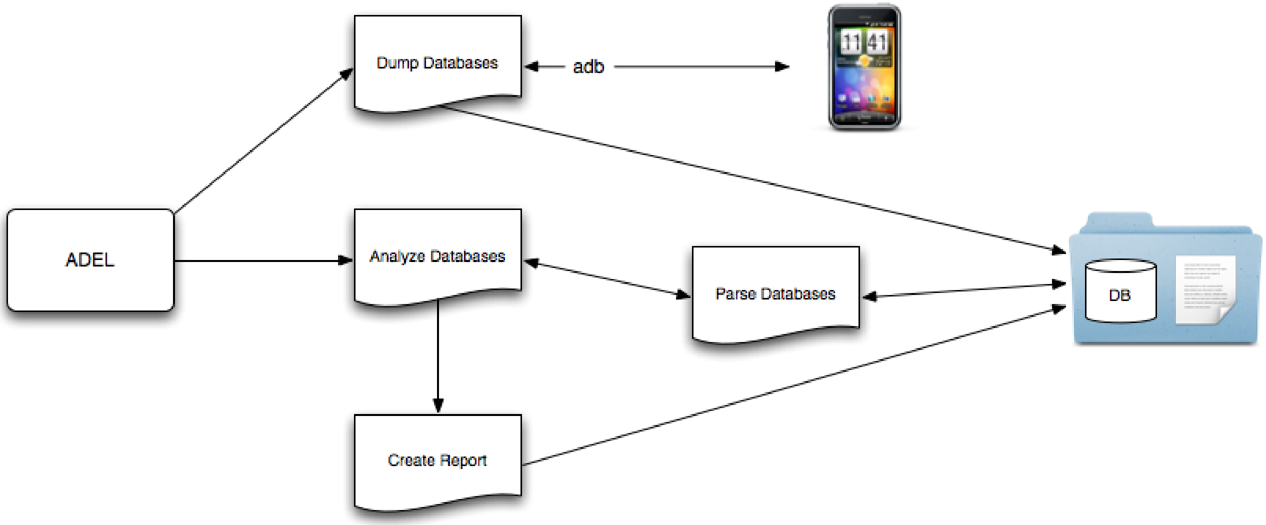


Figure 2: Structure of ADEL

5.1 Basic Development Guidelines

During the development of ADEL we primarily took into account the following design guidelines:

Forensic principles: ADEL is intended to treat data in a forensically correct way. This goal is reached by the fact that activities are not conducted directly on the phone but on a copy of the databases. This procedure assures that data does not become changed, neither by the users of ADEL nor by an uncompromised operating system. In order to proof the forensic correctness of ADEL, hash values are calculated prior and after each analysis, to guarantee that dumped data did not become changed during analysis.

Extendibility: ADEL has been modularly built and contains two separate modules: the analysis and the report module. Predefined interfaces exist between these modules and both of them can be easily amended by additional functions. The modular structure allows for dumping and analyzing further databases of smartphones without great effort and facilitates updates of the system in the future.

Usability: The use of ADEL is intended to be as simple as possible to allow its use by both qualified persons and non-experts. At best, the analysis of the mobile phone is conducted in an autonomous way so that the user does not receive any notice of internal processes. Moreover, the report module creates a detailed report in a readable form, including all of the decoded data. During the execution, ADEL optionally writes an extensive log file where all of the important steps that were executed are traced.

5.2 Data Extraction

ADEL makes use of the Android Software Development Kit (Android SDK) to dump database files to the investigator's machine (see Section 3.3).

5.3 Parsing SQLite Database Files

To extract contents contained within a SQLite database file ADEL parses the low-level data structures described in Section 4.2. After having opened the database file that is to be parsed in read-only mode, ADEL reads the database header (first 100 bytes of the file) and extracts the values for each of the header fields. Not all, but some of the values in the header fields are necessary to be able to parse the rest of the database file. An important value is the size of the pages in the database file which is required for parsing the b-tree structures (page-wise). After having read the database header fields, ADEL parses the b-tree that contains the "sqlite_master" table for which the first page of the database always is the root page. The SQL CREATE statement and the page number of the b-tree root page are extracted for each of the database tables. Additionally, the SQL CREATE statement is further analyzed to extract the name and the data type for each column of the corresponding table. Finally the complete b-tree structure is parsed for each table, beginning at the b-tree root page that was extracted from the "sqlite_master" table. Every leaf page of the b-tree is identified by following the pointers of all of the interior pages. Finally the row contents of each table are extracted from the cells found in any leaf page that belongs to the same table b-tree.

5.4 Reporting

Within this section we address the report module and its functionalities. In the current development state, the following databases are forensically treated and parsed as described in Section 5.3:

- telephone and SIM-card information (e. g. IMSI and serial number)
- telephone book and call lists,
- calendar entries,
- SMS messages.

Data retrieved this way is written to an XML-File by the report module in order to ease further use and depiction of the data. As the analysis module, it can be easily updated regarding possible changes in future Android versions or in the underlying database schemas. Therefore, we have created different tuple – e. g. [table, row, column] – to define the data that is exchanged between both modules. If the database design changes in the future, only the tuple have to be adapted. The report module automatically creates XML-files for each of the data types listed above. In addition, a report is created which contains all data extracted from the analyzed databases. With the help of a XSL-file the report will be graphically refurbished. All files created by ADEL are stored in a subfolder of the current project.

6. CONCLUSION AND FUTURE WORK

In this paper characteristics of the Android platform and the SQLite database engine have been discussed. Both aspects will become increasingly important for forensic examinations of Android mobile phones in the future.

We presented existing methods to analyze mobile phones and pointed out their advantages and disadvantages. Since these methods either violate forensic principles or necessitate advanced knowledge to perform the analysis, we have presented the tool ADEL which enables automated analysis. ADEL accesses the device via the Android Developer Interface in order to retrieve a copy of selected SQLite databases. Subsequently, the SQLite databases are parsed and data is extracted and finally transformed into a XML-report by the modularly built analysis framework. During the development of ADEL main forensic principles have been taken into consideration.

REFERENCES

[Ace10] ACE Laboratory (2010), 'Professional Data Recovery Product for SSD and Flash drives', http://www.pc-3000flash.com, 2010-10-13

[And10] Android Developers (2010), 'What is Android?', http://developer.android.com/guide/basics/what-is-android.html, 2010-10-13

[Car05] Carrier Brian (2005), 'File System Forensic Analysis', Addison-Wesley, Boston

[Gar10] Gartner (2010), 'Mobile Communications by Open Operating System, Worldwide', http://www.gartner.com, 2010-06-15

[Gog10] Google (2010), 'Android System Architecture', http://www.techflare.com.au/media/102-android%20-%20system-architecture.jpg, 2010-10-13

[Gog11] Google Developer (2011), 'Google Developer - Platform Versions', http://developer.android.com/resources/dashboard/platform-versions.html, 2011-02-06

[Gre10] Björn Greif (2010), 'Mobile telephone market increases 2009 up to 6.6 percent', http://www.zdnet.de/news/wirtschaft_unternehmen_business_bitkom_mobilfunkmarkt_waechst_2009_um_6_6_prozent_story-39001020-41000202-1.htm, 2009-02-19

[Hoo09] Hoog Andrew (2009), 'Android Forensics', Mobile Forensics World, 2009-05-26 to 2009-05-30, Chicago

[LK10] Lessard Jeff and Kessler G. C. (2010), 'Android Forensics: Simplifying Cell Phone Examinations', Small Scale Digital Device Forensics Journal, Vol. 4 No. 1

[LYC+10] Xinfang Lee, Chunghuang Yang, Shihj en Chen and Jainshing Wu (2010), 'Design and Implementation of Forensic System in Android Smart Phone', National Science Council of Taiwan

[Moh08] Mohindra Dhruv (2008), 'Android, Incident Response and Forensics', SPRING, 2008-08-08, Mannheim

[San10] SANS (2010), 'Open Source Android Digital Forensics Application', https://blogs.sans.org/computer-forensics/author/andrewhoog/, 2010-10-13

[Sch03] Schiller J.H. (2003), 'Mobile Communications', Addison-Wesley, Boston

[Spr10] Spreitzenbarth, Michael (2010), 'Panoptes: An Android Forensic Software Agent', University of Mannheim

[SQL11] SQLite (2011), 'The SQLite Database File Format', http://www.sqlite.org/fileformat2.html, 2011-02-06

IOS MOBILE DEVICE FORENSICS: INITIAL ANALYSIS

Rita M. Barrios
Assistant Professor
University of Detroit Mercy
Detroit, Mi, 48221
barriorm@udmercy.edu

Michael R. Lehrfeld
Assistant Professor
East Tennessee State University
Johnson City, TN, 37614
lehrfeld@etsu.edu

ABSTRACT

The ability to recover forensic artifacts from mobile devices is proving to be an ever-increasing challenge for investigators. Coupling this with the ubiquity of mobile devices and the increasing complexity and processing power they contain results in a reliance on them by suspects. In investigating Apple's iOS devices -- namely the iPhone and iPad -- an investigator's challenges are increased due to the closed nature of the platforms. What is left is an extremely powerful and complex mobile tool that is inexpensive, small, and can be used in suspect activities. Little is known about the internal data structures of the device or the proper method of extracting forensically sound images of them.

This article will discuss the current state of iOS mobile device forensics. An examination of what data is contained on the devices as well as what can currently be extracted from suspect device is looked at. Jailbreaking an iOS device will be evaluated against its pros and cons along with current professional and open source tools. Finally, a discourse on our continuing research into deleted file recovery and future works is presented.

Keywords: Digital Forensics, iOS, iPhone, iPad, Mobile Devices, Security, Analysis, Tools

1. INTRODUCTION

Mobile platforms have been on the horizon for many years. Tablet PCs and PDAs have made portable computing very tangible for many organizations. Lightweight laptops and net-books have furthered this trend of mobilization and have increased their immersion into the business world. Pagers and terse text messages have been replaced by full document editing and rich text emails. In 2007 Apple introduced the iPhone, and in 2009, the iPad. The uniqueness of these iOS devices and their rapid adoption into multiple domains has been propelled by their portability, usability, and processing power.

The potential uses for the iOS devices vary greatly, but there is no denying their broad adoption. By the end of 2011, there are expected to be more than 100 million iPhones and 43 million iPads in the marketplace (Chaffin, 2010; Elmer-DeWitt, 2010). To contrast this to laptop sales, BestBuy CEO Brian Dun commented that iPad sales could cut into laptop sales by as much as 50% (Yarow, 2010).

As can be expected, the devices are being used for legitimate and illegitimate purposes. These portable devices can be found in every industry whether officially supported by the institution or not. It can be expected that sensitive data will find its way onto these devices and it is ultimately the institution's responsibility to provide the information safeguards. One must consider the effect of such an event should that device be compromised. The primary questions to consider is what sensitive data may be resident and to what level would accessibility to this information exist. Current research indicates that providing security mechanisms for mobile iOS platforms is drastically different from securing traditional mobile devices such as the standard laptop and PDA (Schuessler & Ibragimova, 2009).

As the digitization of information is accelerated by governmental mandates the ease of access of the data is greatly increased. The ability to secure confidential information behind a locked door no longer applies. Couple this with powerful iOS devices that are often misplaced or stolen (Helft & Bilton, 2010) or used for malicious activities and suddenly there is a need for 1) ensuring data security; and 2) in the event of a breach, investigators need to have the ability to determine exactly what has occurred and the impact to the organization, if any, related to the potential data compromise.

In the healthcare domain, for example, the *Health Insurance Portability and Accountability Act* (HIPAA) of 1996 provides some very specific challenges for data security (HIPAA 2010). This law established defined standards for data preservation and security across differing platforms. In a 2009 study of computing habits of healthcare professionals, it was determined, that over 85% used mobile devices and connected to secure systems using a myriad of network technologies (Justice, Wu, & Walton, 2009). For example, doctors can now use their iOS devices to write electronic prescriptions (Scoop, 2010). The Justice et al (2009) survey also found that only 4% of healthcare institutions have a dedicated computer crime unit that has the ability to include investigation of mobile devices. This environment as identified by Justice et al (2009) indicates that with the increase of mobile device usage in the healthcare industry, there are exponentially more ways to facilitate a data compromise however there are less people to investigate these new environments. With this wide adoption of mobile devices and an increase of the usage of heterogeneous connectivity mechanisms, a proportional increase in the amount of security breaches related to the organizational security protocols can be expected. This increase will ultimately lead to an increase in compromised data as well as an increase in the need for forensic investigations in this environment.

By no means is the healthcare domain the only industry affected by legal standards in terms of data protection. The Sarbanes-Oxley Act of 2002 ("The Sarbanes-Oxley Act of 2002," 2010), the Family Educational Rights and Privacy Act ("Family Educational Rights and Privacy Act (FERPA)," 2010), or the various state statutes regarding identity theft ("Identity Theft State Statutes," 2010) all have one common theme – policies, procedures and controls must be in place to ensure data security. What is not so overt in these legislative documents is the mandate for an organization to perform a forensic evaluation of a data breach to determine the events that occurred in the event of a compromise. What often happens is the organization simply utilizes a security policy checklist to decide the degree of the breach. While this can net some important information, there will be no physical digital evidence produced to support the investigation.

Currently accepted forensic process models, like Palmer's model (Palmer, 2001) or Pollitt (Pollitt, 2007), do little to illuminate digital forensics in terms of the smartphone platform (Dancer & Dampier, 2010). The NIST SP800-101 recommended standard is outdated when considering the current iOS devices (NIST, 2007). Additionally, there is little documented in the literature concerning one of the most popular mobile platforms in the 21st century, namely the iOS environment, when forensic acquisition is considered. Of the limited literature available, researchers, developers and investigators acknowledge the difficulty in obtaining the breadth of information available utilizing the current toolsets that is comparable to its desktop brethren. In fact, little is published concerning forensics for the iOS v4 devices and slightly more in known about previous iOS versions (Hoog & Strzempka, 2010). This gap in knowledge may be causing loss of forensics artifacts or critical information that may prove beneficial to an investigator. As such, a methodology and toolset needs to be developed that will enable investigators to pursue potential compromises in the iOS environment. As noted above, employees will find ways to utilize consumer devices and applications in order to accomplish their business goals and objectives, even if the alternative devices are not approved by IT (Information Technology) corporate directives (Brewin, 2010). With this compromised environment the digital forensics examiner is left to find these areas of inconsistency and to determine the degree of compromise.

The research presented in the following sections will begin to bridge the knowledge gap identified above by examining the current state of the iOS environment. This examination will include enumerating the data contained within the device and as well as what information can be extracted from the iOS environment as identified in section II. An introductory overview of Jailbreaking is then presented in section III. The Zdziarski Method as well as several digital forensic software suites will be examined at a high-level in section IV. Section V presents a conversation of our on-going efforts into the research of the deleted file recovery process within the iOS environment. Additionally, section V presents our continuing efforts in developing a toolset that will aid in the investigation processes for the iOS mobile environment.

2. DATA CONTAINED ON MOBILE DEVICES

The vast array of forensic artifacts found on iOS devices is expansive and valuable. The range of data varies slightly by device, but many categories overlap between iPhones and iPads with and without a cellular radio. Physically, iOS devices are similar in makeup as other solid-state handheld device. The forensically interesting parts to date are the flash chips, GPS chip, and RAM. Dancer and Dampier (2010) compiled a list of issues when confronting smartphone device forensics as it relates to the various areas of interest. They include but are not limited to 1) the various types of memory used in the device; 2) the varying power states of the device; 3) remote wipe capabilities and other mechanism for altering data remotely; 4) proprietary information; and 5) differing ways the device can share information.

Table 1 contains a listing of forensically interesting physical parts of the iPhone 4 and iPad with and without a cellular radio ("iFixit," 2010). Table 2 addresses some of the interesting forensic artifacts that an investigator will need to conduct a thorough investigation (Hoog & Strzempka, 2010). While neither of these two tables is exhaustive in composition as well as which forensic tools can identify the specific information identified, the tables do indeed give depth of understanding just how complex the iOS environment can be. It should be noted that due to the limitations of the paper format, a discussion of mobile tools and their extraction capabilities will not be presented. The reader is directed to the 2010 study as presented by Hoog & Strzempka where a comprehensive evaluation of each tool is presented along with the tools associated data extraction capabilities.

	iPhone 4	iPad	iPad with radio
RAM	√	√	√
Flash	√	√	√
GPS	√	√	√
Cellular Radio	√	N/A	√
Wi-Fi	√	√	√
Bluetooth	√	√	√
CPU Type	A4 Processor	A4 Processor	A4 Processor

Table 1: iPhone and iPad physical components (“ifixit”, 2010)

Artifact	Definition
Call logs – Native Dialer	Determine what calls were attempted and received from the device
Call logs – VoIP Dialer (3rd Party)	“”
Voice Mail	Access deleted and stored messages
SMS – Native Application	Retrieve attempted and received SMS, including deleted SMS
SMS – 3rd Party	Gather information from installed 3rd party application
MMS – Native	“”
MMS – 3rd Party	“”
Email	Retrieve sent/received/deleted emails
Notes - Native Application	
Notes – 3rd Party	
Pictures – Native	Retrieve all pictures from device, including deleted
Pictures – 3rd Party	Access pictures from 3rd party application
Web Tracking Information	Access browser history, cookies, bookmarks
Web Tracking Info – 3rd Party	“”
Process Listing of Device	Plist
GPS Data	Access GPS waypoints
WiFi Connections	List of all access points
Songs	Recovery of all songs on device
Videos	Listing of videos contained on device or deleted

Table 2: Potential digital artifacts on iOS devices (Hoog & Strzempka, 2010)

As noted in Table 2, there are many different categories where forensic artifacts may reside. Moreover, acquisition techniques can be further broken down into the physical and logical. As in traditional computer forensics, a physical acquisition is usually the best method of acquiring evidence. Logical is usually a secondary tactic as is leaves some evidence unrecoverable. However, the ability to recover deleted files relies heavily on a physical acquisition methodology. As previously discussed, physical acquisitions of current iOS devices is difficult to obtain because of the closed architecture of Apple’s devices; thus complicating the recovery of deleted artifacts.

3. JAILBREAKING

Jailbreaking an iPhone or iPad enables the user to gain root access to the device. From this position, a physical image of the device may be obtained using various tools. The current issue with this methodology is the forensic validity of the evidence: will the evidence be accepted in court as part of an ongoing investigation or will the findings be compromised because of the acquisition method? The iDevice communicates with the computer using Apple’s Apple File Communication (AFC) protocol. This protocol enables iTunes to communicate with a sandbox on the iDevice; excluding raw access to the iDevice and a majority of the file system.

By Jailbreaking a device, the current limitations of iTunes can be subverted and root access achieved. With root access, typical Linux utilities can be loaded to the device where SSH and dd commands can be run to produce a full drive image extraction (Harrington, 2008).

Jailbreaking presents a difficult problem for law enforcement entities. According to the NIST Guideline for Mobile Phone Forensics;

- *No actions performed by investigators should change data contained on digital devices or storage media that may subsequently be relied upon in court.*
- *Individuals accessing original data must be competent to do so and have the ability to explain their actions.*
- *An audit trail or other record of applied processes, suitable for replication of the results by an independent third-party, must be created and preserved, accurately documenting each investigative step.*
- *The person in charge of the investigation has overall responsibility for ensuring the above-mentioned procedures are followed and in compliance with governing laws.*

Table 3 NIST principles for mobile phone forensics (NIST 2007)

Jailbreaking violates the first of these principles since jailbreaking circumvents the locked state by injecting processing components into the device which forces a change in the data/program composition. This injection may provide for technical issues during the legal phase for the investigator with the remaining three components of the NIST principles. As reported by Sean Morrisey in the July 2010 newsletter for Digital Forensics Magazine, Jailbreaking is a legal and acceptable method of access in the iOS environment for law enforcement agencies however; it is not legal for the civilian examiner (Morrisey, 2010). What this results in is forcing the civilian examiner to be bound by the logical data collection process which may result in incomplete evidence being reported to the judicial body. Clearly there is a need for a more forensically sound approach to obtaining a raw disk image of an iDevice while adhering to commonly accepted computer forensics processes, procedures and controls.

4. ACQUISITION METHODOLOGIES AND TOOLSETS IN THE IOS ENVIRONMENT

When considering the availability of forensic tools for the iOS device, the choices are rather limited. The commonly used toolsets of Forensic Tool Kit (FTK) as offered by Access Data and EnCase as offered by Guidance Software perform very well with standard hard disk forensics. However, both of these tools fall short when it applied to the iOS environment and the recovery of deleted files.

As noted, the most significant issues the forensic examiner is presented with in terms of toolset utilization are the ability to recover deleted files within the iOS environment. As Zarren & Baig (2010) note in their 2010 study as well as has been previously identified in table 2 of this study, a significant amount of evidence can be obtained during the deleted file recovery process. This evidence can include text messages and the contact list of the suspect device.

To begin the discussion, the following paragraphs will overview the acquisitions methods used in mobile device investigations followed by challenges often encountered during the acquisition processes. This section will conclude with a brief discussion of a few of the commonly used forensic tools for the iOS environment as well as their evaluation within the Hoog & Strzempka study (2010).

4.1 Acquisition Methods

Acquisition can be considered the most important task during the investigative process. When considering the mobile device environment, advances in the technology allow the potential to retrieve a vast amount of information. The method of acquisition employed depends largely upon the vendor of the device but also the model, condition, amount of time available and the nature of the investigation.

With these advancements, Owen, Thomas & McPhee (2010) remind the investigator that strict guidelines must be followed so that the evidence as well as the procedures presented can be considered forensically sound within the judicial setting. While there is a close relationship to traditional hard disk forensics, the current guidelines are not appropriate for the mobile environment (Owen, Thomas

& McPhee, 2010). In addition, Zareen & Baig (2010) also remind the investigator that there is no standard in place for the analysis of internal device memory. This lack of standardization becomes a barrier since the iOS device relies on flash memory rather than a hard disk.

In the iOS environment, full acquisition becomes difficult to achieve, as there is a need for the investigator to interact with several processing layers: The hardware layer, the OEM (Original Equipment Manufacturer) layer and the application layer (Owen, Thomas, & McPhee, 2010). The hardware layer includes the processor, RAM, ROM, antenna, and other input/output devices. The OEM layer maintains the boot loading, configuration files and the application layers. Finally, the application layer supports the end user applications, internet applications, remote wiping and media players (Owen, Thomas, & McPhee, 2010). Additionally, an investigator has the luxury of removing the hard drive from a standard computer system, causing it to become more static in nature in terms of evidential integrity. This is not possible with a mobile device which results in a more complex investigative process (Owen, Thomas, & McPhee, 2010).

The limited research into forensics for the iOS environment identifies six methods of acquisition. These are manual, logical, hex-dump analysis, chip-off, back-up analysis, and bit-by-bit.

Manual Acquisition is the process by which the investigator reviews the device's documentation and employs a manual browsing procedure that utilizes the keypad and display features of the device to acquire the needed evidence. This process will not net all of the needed data, especially the deleted data objects. Issues associated with this method include errors in judgment and data modification as well as the incredible amount of time needed to move methodically through all features of the device (Zareen & Baig, 2010).

Logical Acquisition is the process by which the investigator gains access to the user data via cable connected to the device and to the evidence receptacle. The investigator extracts the evidence using the AT command set as employed by commercially available toolsets. This method does support foreign languages and there is a considerable amount of knowledge and research in this area. The challenges encountered when using the logical acquisition method include the potential to have data written to the device which can be expected to be, at a minimum, changes to the log file, the requirement of many types of cables that are device dependent. While the recovery of live data can be achieved using this method, there is no access to the deleted data since the memory cards need to be directly accessed. Even with these concerns, this method is preferred over an attempt to acquire the data using a computer to which the device has been synced with (Hoog & Strezempka, 2010; Zareen & Baig, 2010)

Hex-dump analysis allows for the physical acquisition of mobile device files (Zareen & Baig, 2010). This procedure involves connecting the mobile device to an evidence receptacle or removing the SIM card and utilizing a reader then 'dumping' the contents to the receptacle. The evidence retrieved is in a raw format, which requires a data conversion. Access to the deleted files that have not been over-written can be achieved however the nature of the evidence obtained results in inconsistent reporting, is difficult to use, requires custom cables and the source code is often protected by the manufacturer (Zareen & Baig, 2010). Additionally, this method is a derivation of the hacker community that may be considered inappropriate in an investigation as is the utilization of the Jailbreaking methodology.

Chip-off is a method of acquisition where the investigator physically removes the chip from the device then proceeds to read the device using a secondary device such as another mobile device or an EEProm reader to perform the forensic analysis. This method is very expensive but is able to extract all of the data. In addition, the resulting acquisition can be difficult to interpret and convert (Zareen & Baig, 2010). It should be noted that since the drive is always encrypted in the iOS environment, this method has a low degree of success (Wright & Adler, 2010).

Back-up utilization is simply using a backup of the mobile device to perform the forensic analysis. The primary constraint when utilizing this method is that the investigator only has access to those files that have been implicitly synchronized using the device's standard protocol (Hoog & Strezempka, 2010). When considering the iPhone device, this method can serve the investigator well since there is much information in the SQLite database that is supported by the protocol. This database can be queried directly to obtain the deleted information however, to do so requires the investigator to use a Jailbreaking method, which, as has been noted, is not considered a forensically sound procedure.

Bit-by-bit method of acquisition is considered the most thorough of all acquisition methods for mobile devices (Hoog & Strezempka, 2010). This method creates a physical bit-by-by copy of the mobile device's data including the deleted files that net in the greatest amount of information. It is considered the method that is most closely related to the traditional methods of evidence acquisition. Unfortunately, in the iOS environment, this method is not possible without the use of Jailbreaking.

4.2 Challenges in Acquisition

There are many challenges when considering forensics within the iOS environment that prevent a full acquisition of the iDevice. The speed of change within the technology landscape continues to prove to be a barrier to the investigation (Owen, Thomas, & McPhee, 2010; Zarren & Baig, 2010). This causes conflicts between version of the OS as well as within the vendor's offerings.

There is also a lack of write-blocking techniques for mobile devices. Without write blocking, there is nothing to prevent the device from receiving messages such as calls and texts while performing a forensic investigation (Zarren & Baig, 2010; Zdziarski, 2010). While blocking can be prevented using a shielded lab, as Zdziarski notes (2010), it is very expensive to implement. A more economical approach may be to remove the SIM to disable reception during the investigation. However, access to the SIM, which may contain information such as encryption keys that may be associated with user authentication, will be unavailable which may in turn hinder the investigative process. If we take a different point of view from the investigative approach, it may be beneficial to maintain the incoming call reception while maintaining a block of the write activities in order to capture on-going communications. This of course is driven by the goals and objectives of the investigative body.

From a forensic process point of view, there is a lack of standardization within the manufacturing community in terms of data storage. This creates an environment where commonly known tools are rendered substandard with each release of an update to the OS.

Often times, the investigator has to work on the actual device, which affects the forensic integrity of the investigation (Owen, Thomas, & McPhee, 2010). For example, when an acquisition is taken, the device must be powered on. When this is done, the state of the device is modified. This situation forces the investigator to become acutely aware of which state the device is in at any given time and how to handle the evidence for the given state (Owen, Thomas, & McPhee, 2010). Initially, it appears as if the chip-off method would negate the need to power on the device in order to take the image. However, as noted above, in the iOS environment, the drive is always encrypted therefore the chip-off method has little degree of success (Wright & Adler, 2010).

One of the most significant challenges is that the commonly available forensic tools most often only perform logical acquisitions, which does not capture the deleted data as is done with a physical acquisition (Zareen & Baig, 2010). This is where many investigators turn to Jailbreaking as a method to perform a physical acquisition. As noted above, Jailbreaking is not considered a forensically sound procedure since in effect the investigator is altering the information contained on the device that may have an impact on the evidence presented.

Finally, although this presentation of challenges is not exhaustive by any means, there is the challenge of backward compatibility between releases of the iOS environment that needs to be addressed. One facet of our research shows that each release of the iPhone environment has a software version, a baseband version and a bootloader version which will have an impact on how one must handle the device during an investigation. Currently, it is known that the baseband updates the software version when an update occurs via iTunes. While the software version can be rolled back to its original state, the baseband cannot unless jailbreaking methods are employed. Also, the bootloader version cannot be modified as it is dependent upon the timeframe in which the device was manufactured. To negate this version dependence that is currently a factor in the investigation, one area of our research is focusing on building an external device that is platform independent. This device is expected to be attached to the iOS device which will allow the investigator to gain access to the necessary areas of the system without the need to jailbreak the device. Our future work will further address the challenges presented as well as present the findings of building the external device via the presentation of a more detailed study.

4.3 iOS Forensic Toolsets

The primary goals of any forensic toolset are to extract the evidence from the mobile device, support the reporting objectives as well as to provide for the examination functions. The level of quality that is expected of any investigation when utilizing a forensic toolset is to preserve the integrity of the acquired and extracted data at all costs. As Hoog and Strezempka (2010) state in their study, the key aspect is to avoid modification of any data components within the storage areas of the device. However, if that is not possible, all modifications must be supported by the audit trail put forth (Hoog & Strzempka, 2010).

In order to provide a complete, forensically sound acquisition, both the logical and physical acquisition must be accomplished. As Owen, Thomas and McPhee (2010) identify, with the current landscape of tools that are available to the investigator it is not possible to make a complete image of the mobile device, as these tools do not support both the physical and logical acquisition. Unfortunately, most available tool-sets provide for only the logical acquisition meaning that in order to retrieve the deleted files of the iOS device, one must also perform a physical acquisition. The reasons a second, physical acquisition must occur, as stated previously, is that the iOS device relies on flash memory instead of a traditional hard drive which renders the majority of the toolsets available inadequate (Janson, Delaitre & Moenner, 2008). It is because of this gap, that data recovery is usually carried out via the logical acquisition by utilizing one or more of the iOS supported protocols (Janson, Delaitre & Moenner, 2008).

To give an understanding of the current toolset landscape, a discussion of the current state of software tools available to the forensic investigator follows. It should be noted that this list of software tools is not exhaustive. It should also be noted that the consideration of the information obtained from a Network Service Provider, while an important part of any investigation, is beyond the scope of this research.

When considering traditional digital forensics, there is an industry focus on two primary toolsets, Encase (Guidance Software) and FTK (Access Data) (Owen, Thomas, & McPhee, 2010). With the surge of iOS devices entering the market place between 2007 and present day, these two vendors have emerged with forensic toolsets to support the iOS device.

Encase Neutrino is Guidance Software's mobile solution in forensic acquisition. It has the ability to support devices from Nokia, Motorola, Samsung, Siemens, LG, Palm, Blackberry (RIM), HTC, UTStarCom, and Sony Ericsson. (Guidance Software, 2010) The tool can collect data from unallocated space (deleted files) on select devices including the iPhone (Hoog & Strezempka, 2010). However, according to the corporate brochure, there is no mention of iPhone support (Guidance Software, 2010). Testing as presented by Hoog & Strzempka (2010) identified that the toolset missed SMS messages and photos in unallocated space (deleted files), was unable to pick up screen shots, music files, passwords, phone information, HTML files and MS Office documents. The study also identified that the tool-set fell below expectations when retrieving email (Hoog & Strzempka, 2010)

Access Data's Mobile Phone Examiner (MPE) Plus software brochure indicates that it supports more than 1200 various devices with support for 2300 devices by January 2011 however there is no indication that it supports the iOS system. Logical acquisition is supported but the vendor's website indicates that physical acquisition will be forthcoming for iPhone, iPad and Android. Following acquisition, the file must be imported to Forensic Tool Kit 3 (FTK3) as there is no backward capability to prior releases of the FTK toolset.

As recently as 2010, there are a few software toolsets and procedures to support forensics in the iOS environment. A few of the more popular software tools and methods for iOS forensics are presented.

Perhaps the most popular and receiving the most focus as of the writing of this study is Zdziarski's Method of iOS acquisition. At a high-level, the Zdziarski Method is what is termed as a "semi-Jailbreak" solution. We state this because the method uses system RAM to inject code into the space that will allow full access to a raw disk image as well as bypass security components such as user passcodes (Zdziarski, 2010). The image can be captured via a SSH protocol using a WiFi connection once access has been gained (Zdziarski, 2010). To gain a full understanding of the Zdziarski Method, the reader is encouraged to further enhance their knowledge by examining the research as presented by Zdziarski in 2010. While Zdziarski (2010) indicates that there are no Jailbreaks employed when utilizing his methods to perform a physical acquisition since the user area of RAM is left untouched, the device's system RAM is loaded with the needed imaging components to allow the iOS device to boot from memory. The modified device reverts to its original state when rebooted. By definition, this is Jailbreaking the system since RAM is modified to bypass the manufacturer's preventative measures as well as device security components. Granted, there is a lower probability that since system RAM is being modified, that critical data will be over-written. This of course assumes that the system RAM was 'clean' prior to the forensic acquisition. Zdziarski uses a tool-set that was developed in-house to perform the forensic examination and this tool-set is only available to law enforcement personal (Zdziarski, 2010). It should be noted that the Zdziarski Method was validated in draft by NIST in October 2010 (NIST, 2010). Testing showed that the methodology did acquire all supported data objects when using the Smartphone Tool Test Assertions and Test Plan with the iPhone 3G device (NIST, 2010). However, when Hoog and Stzempka (2010) performed their testing against the iPhone 3G, there were occasions where various components, such as passwords, were missed.

Another popular tool-set used for iOS forensics is the Paraben Device Seizure 4.0 tool. The software specifications indicate that 2200 devices are supported however; there is no direct indication that there is support for the iOS environment. The software specification indicates that the tool has the ability to perform both logical and physical acquisition however; the testing as perform by Hoog & Strzempka (2010) indicates that the tool uses the devices backup function to recover the deleted files. The Hoog & Strzempka (2010) testing survey indicated that the tool missed SMS messages and photos in unallocated space (deleted files). The tool also missed music files, screen shots, passwords, HTML and MS Office files as the Encase Neutrino tool did. In addition, like Encase Neutrino, the tool fell below expectations for email recovery. The tool also fell below expectation in video and voice mail recovery (Hoog & Strzempka, 2010)

There are many more commercial and open source forensic tools coming into the digital forensic landscape but continue to face the common issues as noted above (Hoog & Strzempka, 2010; Owen, Thomas, & McPhee, 2010)

As can be seen, unallocated space (deleted files) continues to be a troublesome area without the use of device modification tools and methods as demonstrated by Zdiarski's Method. The research, as will be presented in future works, will attempt to eliminate these concerns.

5. FORENSICS IN THE IOS ENVIRONMENT

As can been expected, the amount of ubiquitous information stored on mobile devices will continue to grow (Owen, Thomas, & McPhee, 2010). Zareen & Baig (2010) stress the need for the development of new forensic tools and techniques to support this non-traditional computing environment.

With this gap in mind, we are proposing the development of a forensic toolset which includes building an external device as outlined above that will support both physical and logical acquisition in the iOS environment. The software side of the toolset is expected to function in much the same way that traditional forensic toolsets perform when applied to the standard computing environment however there will be no need to first jailbreak the device prior to imaging process. We believe that enabling a toolset that does not require jailbreaking will aid the civilian examiner as noted above in regards to the legal issues that surround the jailbreaking process. Also, being able to perform both, a logical and physical acquisition in such a manner will support the integrity of the investigation.

We also are focusing on the development of this toolset in such a way as to support platform independence as well as version change independence. We believe that with an external device, the version of the software, the baseband and the bootloader of the iOS environment will no longer be a consideration when moving forward with acquisition.

Additionally, as indicated in the literature, there is not a full understanding of the ramification when using the jailbreaking methodology during the iOS investigation. As our research moves forward, we expect to develop this understanding in a well-documented study that will be presented to the research community upon its completion.

The toolset under development that will be presented to the research community is being developed based on the NIST CFTT (Computer Forensics Tool Testing) specifications. The objectives of the CFTT program is to provide measureable assurance to practitioners, researchers, and other application users that the tools used in computer forensics investigations provide accurate results (NIST, 2010).

6. CONCLUSIONS

Smartphone usage has grown considerably over the past year with the 2nd quarter of 2009 showing that these types of devices have accounted for 16% of the total mobile market (Dalrymple, 2010). This staggering surge further jumped to 23% in Q1 of 2010 (Dalrymple, 2010). iPhone and iPad devices are responsible for a considerable amount of this growth. As noted above and is presented in a study from the Nielsen organization and was presented by Dalrymple (2010), since its introduction to the market in 2007, the iPhone (28%) has more than triple market share over Android (9%). Currently, Blackberry still holds the lead at 35% (Dalrymple, 2010). iPhone and iPad are expected to continue to dominate the market place in coming years due to its user focused platform.

With this growth in mobile device usage, the primary challenges in mobile forensics, in particular the iPhone/iPad environments, continue to be rapid changes in the technology stack, a lack of standardized methods for data storage and the closeness of the OS. It is because of these reasons that there is a need for the development of new forensic tools and techniques that specifically address these unique attributes of the mobile environment.

The toolset that will be presented to the research community in future publications will address and resolve the shortcomings of obtaining a complete image (physical and logical) of the iOS device, the current usages of Jailbreaking in a forensically sound environment as well as the issues of platform and version dependence.

REFERENCES

Access Data. (2010). Mobile Forensics Examiner (product brochure). Retrieved December 27, 2010, from http://accessdata.com/products/forensic-investigation/mobile-phone-examiner

Brewin, B. (2010). VA employees tap cloud apps on their own, posing security risk. Retrieved December 24, 2010, from http://www.nextgov.com/nextgov/ng_20101222_6852.php

Chaffin, B. (2010). iSuppli Bumps 2011 iPad Forecast to 43.7 Million. Retrieved December 16, 2010, from
http://www.macobserver.com/tmo/article/isuppli_bumps_2011_ipad_forecast_to_43.7_million/

Dalrymple, J. (2010). iPhone triples Android in mobile market share. Retrieved December 27, 2010, from http://news.cnet.com/8301-13579_3-20006889-37.html

Dancer, F. C. T., & Dampier, D. A. (2010). *A Platform Independent Process Model for Smartphones Based on Invariants.* Paper presented at the IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering.

Elmer-DeWitt, P. (2010). What's driving iPhone 4 sales? Retrieved December 16, 2010, from http://tech.fortune.cnn.com/2010/06/17/whats-driving-iphone-4-sales/

Family Educational Rights and Privacy Act (FERPA). (2010). Retrieved December 16, 2010, from http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Guidance Sotftware. (2010). Encase Neutrino (software brochure). Retrieved December 27, 2010, from http://www.guidancesoftware.com/mobile-cellphone-forensics-software-neutrino.htm

Harrington, M. (2008). iPhone Forensic Examinations – A Series. Retrieved December 25, 2010, from http://mobileforensics.wordpress.com/2008/09/15/iphone-forensic-examinations-a-series/

HIPPA-1996 (2010). Retrieved December 16, 2010, from http://www.hhs.gov/ocr/privacy/index.html

Helft, M., & Bilton, N. (2010). For Apple, Lost iPhone Is a Big Deal. Retrieved December 16, 2010, from http://www.nytimes.com/2010/04/20/technology/companies/20apple.html

Hoog, A., & Strzempka, K. (2010). iPhone Forensics White Paper. Retrieved Dec 16, 2010, from http://viaforensics.com/education/white-papers/iphone-forensics/

Identity Theft State Statutes. (2010). Retrieved December 16, 2010, from http://www.ncsl.org/?tabid=12538

iFixit. (2010). Retrieved December 18, 2010, from http://www.ifixit.com/Device/iPhone_4

Janson, W., Delaitre, A., & Moenner, L. (2008). Overcoming Impediments to Cell Phone Forensics. In Proceedings of the 41st Hawaii International Conference on Systems Sciences.

Justice, C., Wu, H., & Walton, E. (2009). *Mobile Forensics in Healthcare*. Paper presented at the Proceedings of the 2009 Eighth International Conference on Mobile Business.

Morrisey, Sean. (2010, July). New DFM recruit Sean Morrisey writes about the iPhone forensic tool Lantern. DFM Newsletter July 2010. Retrieved March 19, 2011 from http://www.digitalforensicsmagazine.com/newsletter/DFM-Newsletter07.html

NIST. (2007). SP800-101 Guidelines on Cell Phone Forensics (pp. 104). Retrieved from http://csrc.nist.gov/publications/nistpubs/800-101/SP800-101.pdf.

NIST. (2010). Test Results for Mobile Device Acquisition Tool: Zdziarski's Method (draft). October 2010. Retrieved from http://www.ntis.gov/search/product.aspx?ABBR=PB2011104749

Owen, P., Thomas, P., & McPhee, D. (2010). *An Analysis of the Digital Forensic Examination of Mobile Phones* Paper presented at the 2010 Fourth International Conference on Next Generation Mobile Applications, Services and Technologies.

Palmer, G. (2001). *A Road Map for Digital Forensic Research.* Paper presented at the First Digital Forensics Research Workshop (DFWRS). Retrieved from http://www.dfrws.org/2001/dfrws-rm-final.pdf

Pollitt, M. M. (2007). *An Ad Hoc Review of Digital Forensic Models*. Paper presented at the Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4155349

The Sarbanes-Oxley Act of 2002. (2010). Retrieved December 16, 2010, from http://www.gpo.gov/fdsys/pkg/PLAW-107publ204/content-detail.html

Schuessler, J. H., & Ibragimova, B. (2009). *Portable Privacy: Mobile Device Adoption*. Paper presented at the Annual Security Conference. Retrieved from *www.security-conference.org/SecurityConf_2009_Proc/Papers/4.doc*

Scoop, E. (2010). DrFirst™ Creates Stunning E-Prescribing Experience on iPhone Retrieved December 16, 2010, from http://www.emrconsultant.com/forum/topic/722-drfirst-creates-stunning-e-prescribing-experience-on-iphone/

Wright, J. & Adler, M. (2010). Session 209-Securing Application Data. Apple World Wide Developers Conference 2010. San Francisco, CA, USA. Retrieved December 29, 2010 from http://developer.apple.com/videos/wwdc/2010/

Yarow, J. (2010). Best Buy CEO: iPad Is Cannibalizing Laptop Sales By As Much As A Shocking 50%. Retrieved December 16, 2010, from http://www.businessinsider.com/best-buy-ceo-ipad-is-cannibalizing-laptop-sales-2010-9

Zdziarski, J. (2010). The Zdziarski Method. Retrieved December 26, 2010, from http://viaforensics.com/education/white-papers/iphone-forensics/zdziarski/

Zareen, A. & Baig, S. (2010). Mobile phone forensics: Challenges, analysis, and tool classification. In Proceedings of 5th International workshop on Systematic Approaches to Digital Forensic Engineering, 47-55. May 2010, Oakland, CA, USA.

A PRACTITIONERS GUIDE TO THE FORENSIC INVESTIGATION OF XBOX 360 GAMING CONSOLES

Dr. Ashley L Podhradsky
Drexel University

Dr. Rob D'Ovidio
Drexel University

Cindy Casey
Drexel University

ABSTRACT

Given the ubiquitous nature of computing, individuals now have nearly 24-7 access to the internet. People are not just going online through traditional means with a PC anymore, they are now frequently using nontraditional devices such as cell phones, smart phones, and gaming consoles. Given the increased use of gaming consoles for online access, there is also an increased use of gaming consoles to commit criminal activity. The digital forensic community has been tasked with creating new approaches for forensically analyzing gaming consoles. In this research paper the authors demonstrate different tools, both commercial and open source, available to forensically analyzing gaming consoles, specifically the Xbox 360. Used Xbox 360 gaming consoles were purchased online through popular auction sites for the purpose of this research.

Keywords: Digital Forensics, Identity Theft, Xbox 360 Gaming Console, Cyber Crime

1. INTRODUCTION

Technology has introduced new mediums for criminal and misuse activity. While the crimes and misuse are not new, the medium they are carried out on is. Therefore, the digital forensic community has to work to create new standards, tools, and approaches to investigating gaming consoles.

While many gaming consoles exist, Microsoft's Xbox 360 is the most popular among American consumers, selling over thirty-nine million consoles, six million more than their top competitor the PS3. (Bloomberg Businessweek, 2010). With this rise in popularity, the Xbox 360 has also become a popular medium for criminals. When Bill Gates first announced his plans for the Xbox 360 gaming system in January 2000, at the International Electronic Consumers Show in Las Vegas, some critics proclaimed that this new console was nothing more than a "...PC in a black box (Official Xbox Magazine staff , 2005)." These critics were not too far off the mark. The Xbox 360 is not only similar to a personal computer - it is actually *more* powerful than most average personal computers. The hardware and technical specifications found in today's Xbox 360 console includes a detachable 250GB hard drive, an IBM customized power –PC based CPU containing three symmetrical cores each capable of running 3.2 GHz, a 512 MB GDDR3 RAM (which reduces the heat dispersal burden and is capable of transferring 4 bits of data per pin in 2 clock cycles for increased throughput), and 700 MHz DDR (theoretically supplying a swift 1400 MB per second maximum bandwidth) memory (Berardini, 2005).

Given the advanced hardware, high storage capacities and online access, the Xbox 360 has become a favorite medium for cybercrimes.

2. CRIMINAL ACTIVITY ON GAMING CONSOLES

The latest gaming consoles by Microsoft, Sony, and Nintendo provide users with computing and Internet functionality that is similar to the functionality offered to users of traditional computing devices (i.e. desktop and laptop computers running Windows, Macintosh, and Linux operating systems). The Xbox 360, for example, allows users to access social networking services such as Facebook and Twitter, stream Internet radio through Last.fm, and watch movies via Netflix. The PS3 allows users to send instant messages and create chat rooms to banter with other users over the PlayStation Network. It also allows users to access web-based email (e.g. Hotmail, Gmail, Yahoo! Mail) and websites through its proprietary browser. The Nintendo Wii allows users to send email messages, including messages that contain picture attachments, to other Wii users and users of third-party email services. The Wii, PS3, and Xbox 360 all offer users the ability to store media files on a hard drive or in flash memory.

The functionality of the PS3, Wii, and Xbox 360 provide offenders with a powerful tool to use for committing and supporting criminal activity. The communication options available through these gaming consoles are particularly helpful to criminals. Text, voice, and video communication options within gaming environments and through consoles menus provide offenders with easy access to a population of suitable targets for victimization for crimes that produce economic and social harms.

Criminal activity that produces economic harm is expressed in terms of monetary damages (Criminal Intelligence Service Canada, 2007). These damages can be borne by individuals, communities, businesses, and governments and can be committed by a single person or an organized criminal group. Subscriber data (e.g. name, address, phone number, credit card number, gaming network ID, and gaming network password) connected to an account for an online video game console community can be exploited by criminals for direct financial gain or sold to third-parties for their misuse. Virtual currencies and virtual goods amassed by a game player can, at times, be converted into real-world currency through in-game transactions or third-party services (e.g. EBay, PlayersAuctions, and IGE) and are, thus, attractive targets for economic fraud.

Media reports document the involvement of gaming consoles in a variety of crimes aimed at illicit financial gain, including video game piracy (McHugh, 2011), cracking/hacking (Rivington, 2007; McMillan, 2011), identity theft (Lemos, 2007), credit card fraud (Evers, 2007), and phishing (Fried, 2005; Deleon, 2008; Constantin, 2010). For example, Harris (2009) reports on the theft and, subsequent, sale of more than 500,000 Xbox Live account credentials. He also notes that the credentials sold for approximately £5 per account.

Unlike crimes that produce economic harm, crimes that produce social harm are not expressed in terms of monetary damages (Criminal Intelligence Service Canada, 2007). Instead, criminal activity involving social harm is expressed in terms of the physical and psychological damages to the victim.

Children who use gaming consoles and respective online networks are particularly vulnerable to crimes producing social harm. When playing games with other people over the Internet, children often find themselves immersed in environments devoid of the traditional guardians (e.g. parents and teachers) who serve to protect them in the physical world. Media reports have linked gaming consoles to the victimization of children in cases of rape (Hill, 2009), child pornography (Bush, 2008; Weinstein, 2009; Peterson, 2010), online harassment/bullying (Snow, 2007; Fujji, 2010), and child sexual solicitation (Bullock, 2009; Cavalli, 2009; Potter, 2009). Hitt (2011), for example, details a case in which a 36-year old woman traveled from Florida to Maryland to meet a 13-year old boy she met in an Xbox Live chat room. During her visit, the woman engaged in sexual activity with the boy. She was subsequently charged with rape and child molestation. Chat transcripts discovered during the investigation also showed that the offender exchanged sexually explicit images and videos with her victim.

3. XBOX 360 GAMING CONSOLE

The file data format used on the Xbox 360 is FATX, which is an offshoot of the more familiar FAT32, found on older computers and storage devices (Paul K. Burkea P. C., 2006). In fact, the two possess virtually identical format and file data layouts. Unlike the FAT32 however, the FATX does not contain the backup boot or file system information sectors found in FAT32. Additionally, FATX does not support Unicode, which is often utilized by examiners when performing forensic analyses (World Lingo , 2010). The reasoning behind these variations in the file format is that the Xbox 360 was designed primarily for entertainment as opposed to productivity. Thus, redundancy and legacy are apparently forfeited in order to increase the system's speed.

Some of the identifying data which can potentially be retrieved from consoles include, but are not limited to, a user's name, address, telephone number, and credit card information. Credit cards are used to purchase games through the Live Arcade, pay for Xbox 360 Live membership, and buy merchandise such as gamer icons and console themes at Xbox 360's Live Marketplace. One popular movie subscription service, Netflix (Netflix, 2011), even permits its members to rent movies using credit cards directly though their Xbox 360 consoles. Other personal information includes profile data, chat transcripts, blog files and online history. In fact, the Xbox 360 is even capable of keeping a gamers' blog for the user by monitoring the account and automatically generating blog entries about their daily gaming activities.

In addition to gaming consoles becoming incidental to a crime, such as with identity theft, they are also increasingly becoming the actual *instrument* of the crime (i.e.: using the Xbox 360 to transfer and store child pornography).

Given the abundance of data that is retrievable on Xbox 360 consoles, there is an increasing demand to learn more about what tools and approaches are favorable in acquiring data on game consoles. While many tools exist, as with traditional computer forensics, not all tools are created equal.

For this research, two Xbox 360 gaming consoles were purchased randomly from an online auction site and a popular classified forum respectively. An additional Xbox 360 hard drive was retrieved after being discarded, bringing the total tested drives to three. The researchers acknowledge the sample size is small, however they feel it is appropriate due to the fact the major testing is on the software, not the drives.

4. THE INVESTIGATIVE PROCESS

Once removed from the consoles (if applicable), the drives were extracted using T10 and T4 Torx wrenches. Although some forensic examiners report problems accessing data due to locked drives, we did not encounter any difficulties. A variety of open source and commercial tools were utilized to examine the drives. Also, before each tool was used both pre and post Md5 and SHA-1 hashes were recorded for validation purposes utilizing EnCase**.** Direct checksums were also obtained using Linux to curtail dependency and maintain objectivity on the software being tested. The reasoning for utilizing such a wide array of tools was twofold. First, there is not a great deal of information available to date regarding the structure and forensic examination of gaming consoles. This is not because gaming consoles are new per se, but rather that they have evolved so rapidly over the past decade. Secondly, no one tool was capable of presenting the drives in their entirety. The software used to examine the Xbox 360 drives included the following:

- *XPlorer360*- Freeware tool that allows access to three Xbox partitions and memory cards. Xplorer360 allows access to both physical and logical areas of the drive
- *FTK 3.0*- Forensic Toolkit (FTK), produced by AccessData is a commercial suite of applications for forensic analysis of digital media, including Xbox consoles
- *FTK Imager*- Freeware tool from AccessData which allows users to forensically image and analyze drives
- *Modio*- Freeware modding tool that allows Xbox users open their system to allow for customized use of their console

- *wxPirs*- Freeware tool that allows extraction of access to PIRS (themes or gamertags), LIVE (content downloaded from Xbox Live), or CON (internal files specific to Xbox) container files on Xbox 360's
- *ProDiscover Basic*- Freeware tool based on the commercial ProDiscover- allows viewing of each sector to determine data storage locations
- *Digital Forensic Framework* (DFF)- Is an open source tool that aids in the collection and analysis of digital evidence
- *Hex Editor XV132* – Freeware hex editing tool that runs on memory and doesn't need to be installed on the host system, incorporate a built in hex to string and allows bookmarks
- *XFT 2.0- C*ommercial Xbox toolkit developed by Protowise Labs that allows for access to configuration, modification, and user files, included recovering deleted files
- *Data Rescue's DD (DrDD*)- Freeware tool that recovers deleted files off of corrupted storage devices or partitions, while not designed for gaming consoles, it was used to determine functionality
- EnCase Forensic v6 – Commercial forensic analysis tool by Guidance Software (Guidance Software , 2011)

In addition to the above software, several operating systems were also employed during our analysis. This was done to not only to eliminate the possibility that any of the software limitations encountered were the direct result of an incompatible OS, but also to gain a clearer understanding of the FATX file structure. The operating systems utilized for this study were:

- Windows XP
- WIN 7 (Ultimate)
- Red Hat Fedora 14
- Ubuntu 10.10

Determining which operating system to use created somewhat of a dichotomy at times. While the majority of the tools available only operate in a Windows environment, the Linux operating system appeared to be the most compatible with the actual gaming console itself. In fact, gamers seeking to download and play unsigned copies of Xbox 360 games, or elicit superior gaming and dashboard options, can modify their console using Linux. This is referred to as soft-modding or simply modding. Microsoft discourages these types of system changes, which if executed will void the system's warranty. (Microsoft, 2010)

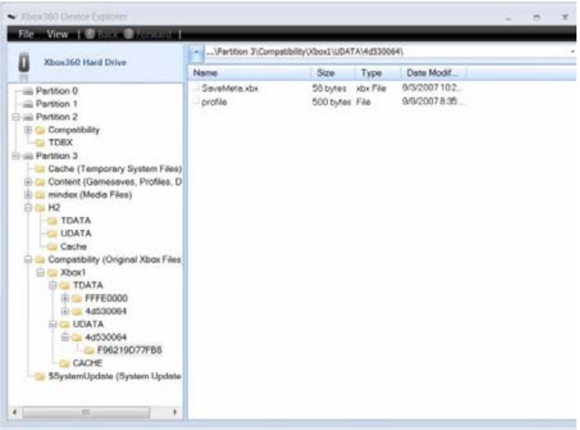
In a recent effort to discourage console modifications, Microsoft released an Xbox 360360 update in early August 2009. This was referred to as the "homebrew lockout" by the Free60 Project, an organization which both promotes and supports users running homebrew applications and Linux operating systems on their Xbox 360360 gaming consoles. The update overwrote the first stage boot loader (responsible for starting the system when it is turned on) thus causing any updates or modifications made by the user to render their system useless. (Free60 Project, 2009) This information can be of significant importance to digital examiners who are seeking to establish or understand the system's bootstrapping process and subsequent drive structure,particularly given how thorny this task can be.

Because the Xbox 360 does not contain the same type of BIOS found in a PC, it should not be expected to boot like the typical PC. In fact, as early as 2002, MIT researcher, Andrew Huang, noted in his detailed study of the Xbox 360's structure that the Xbox 360 contains a "secret boot block"

(Huang, 2001). Perhaps this was an attempt by Microsoft to deter tampering and possibly initially, although not very successfully, as a security mechanism. This information is pertinent because if the boot block is a decoy – then what else might be a red herring?

An example of this ambiguity was found upon examination of the hard drive's partitions. Partition 1, the second partition encountered when opening an Xbox 360 drive, appears to be empty – that is, when it can be found. There could be several reasons for this. It might be reserved for future use or simply just not accessible. Another option is that it could be a lure – a hard drive honey pot of sorts to deflect, and possibly detect, unauthorized access or changes.

Partition 1 was only viewable on two of the hard drives examined, including one sample containing a second or merged set of files. These integrated or legacy files were located on Partition 3, as seen in the capture below using the open source utility, Modio. (Image 1)

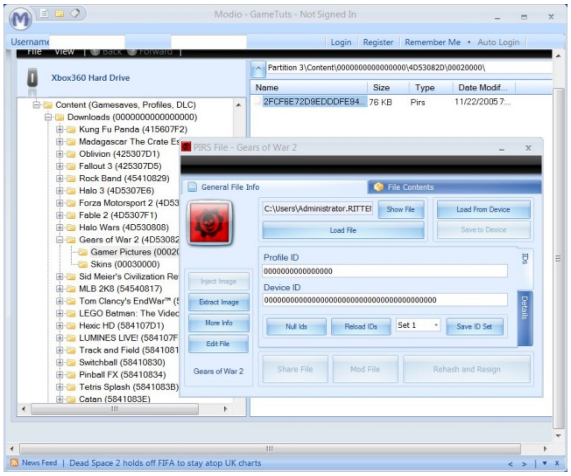


...\Partition 3\Compatibility\Xbox1\UDATA\4d530064\

Name	Size	Type	Date Modif...
SaveMeta.xbx	58 bytes	xbx File	8/3/2007 10:2...
profile	500 bytes	File	8/8/2007 8:35...

Image 1- Partitions as viewed in Modio

Modio is a modding utility that allows Xbox 360 users to manipulate their consoles. It is also handy for viewing image files on the fly without needing to export them first into another program. (Image 2) However, the option to extract files is also available. Although not yet tested by NIST, further evaluation of this utility might prove valuable to law enforcement agencies.



Partition 3\Content\0000000000000000\4D53082D\00020000\

Name	Size	Type	Date Modif...
2FCF6E72D9EDDDFE94...	76 KB	Pirs	11/22/2005 7:...

C:\Users\Administrator.RITTEI

Profile ID
0000000000000000

Device ID
0000000000000000000000000000000000000000

Set 1

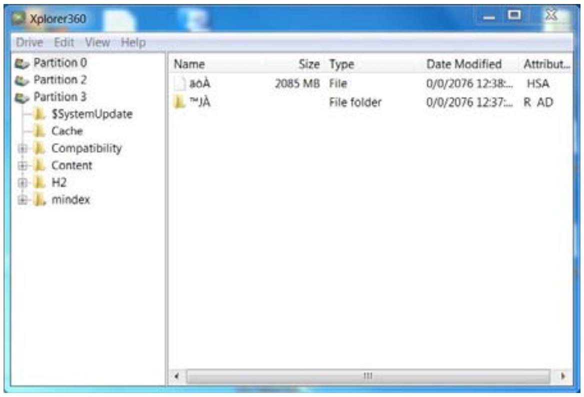
Image 2 – Image viewed in Modio

The hard drives were accessed using a USB 2.0 to SATA adaptor with a 50/60 Hz power supply cable. Writing access to USB adaptors was disabled via the registry in Windows and driver-level write blocking in Linux. Imaging with Access Data's Forensic Toolkit 3.0 (FTK) was a timely process which did not yield extremely productive results. The limited results obtained could be attributed to the FATX file structure of the Xbox 360. The extracted files were inspected by examining the raw data to determine if the drives were intact, deleted, or reformatted.

All three of the drives exhibited signs of being overwritten as evidenced by large sections of zeros in non-program specific files. It would be difficult at best however to declaratively state the drives were reformatted without further studies as each operating system has its own unique way of performing this process and while the Xbox 360 does share some similarities with a PC, it cannot truly be measured using the same criteria. (Computer Gyaan, 2010)

Xplorer360

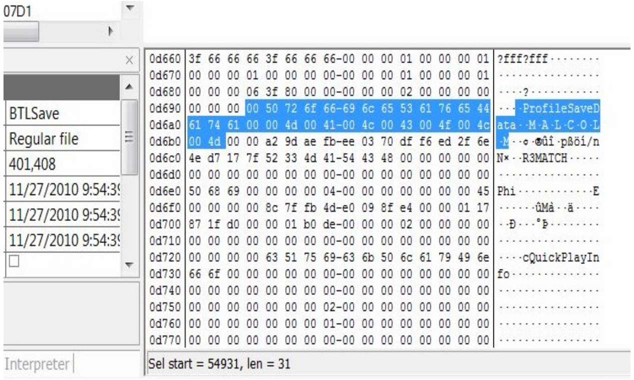
One of the more useful tools employed was a utility called Xplorer360. Xplorer360 is an open source program that enables gamers to open and view, edit, or export data from their Xbox 360 hard drives through their PC. The results were very swift with the hard drive opening in under a minute. Partitions and their subsequent subfolders are displayed in the left hand pane. More detailed information about a selected file or directory is displayed in the right pane. Although earlier studies of the Xbox 360 drive found that Partition 0 was an empty partition (Bolt, 2011), our analysis found two drives that did exhibit files on Partition 0. (Image 3) The empty partition was initially attributed to the extra file mentioned earlier on Partition 3, Xbox 3601 (Partition 3\Compatibility\Xbox 3601), which when observed using traditional forensic tools such as FTK 3.0, appeared to be on the only drive in our study that possessed an empty partition 0. However, after utilizing popular modding tools such as Modio and Explorer360, we were able to ascertain that the two drives containing data in partition 0 *included* the drive with the additional Xbox360 folder. The drive which *did not* contain viewable data in Partition 0 was the newer of the three drives as ascertained from sector 4 (07-02-09). This indicates that the empty Partition 0 may be the result of the August 2009 update which as mentioned earlier, reportedly overwrote the first stage boot loader.



Name	Size	Type	Date Modified	Attribut...
aöÀ	2085 MB	File	0/0/2076 12:38:...	HSA
™JÀ		File folder	0/0/2076 12:37:...	R AD

Image 3 -Partition 0, Viewed in Xplorer360 showing a JA folder and an aoA file

Ironically, although FTK 3.0 did not generate any remarkable user data independently, additional data was revealed later using FTK Imager. After the drive's contents were opened and dumped using Xplorer360, the extracted files were opened in FTK Imager for analysis. One test drive produced a file containing a user's name. This file, which contained profile saved data, was identified as Partition3\Content\0000000000000000\4D5707D4\00000001\BTLsave, last modified on 8/28/2007. (Image 4) Other personal data obtained from the same drive included a user's first name and a partial or abbreviated city name. This was later confirmed by comparing the name discovered with the name and location of the individual who originally owned the console.



07D1

BTLSave
Regular file
401,408
11/27/2010 9:54:3
11/27/2010 9:54:3
11/27/2010 9:54:3
□

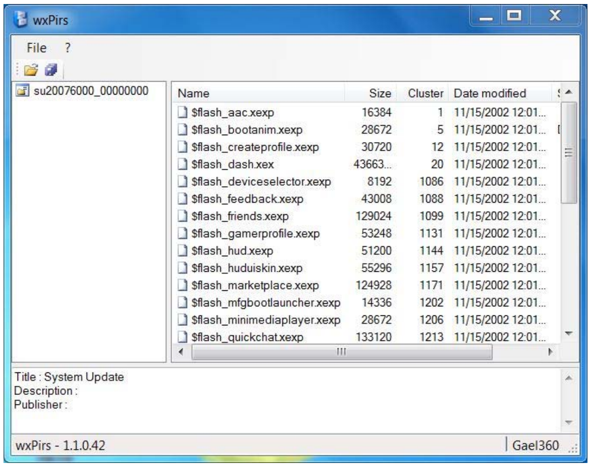
Interpreter

```
0d660 3f 66 66 66 3f 66 66 66-00 00 00 01 00 00 00 01 ?fff?fff········
0d670 00 00 00 01 00 00 00 00-00 00 00 01 00 00 00 01 ················
0d680 00 00 00 06 3f 80 00 00-00 00 00 02 00 00 00 00 ····?···········
0d690 00 00 00 00 50 72 6f 66-69 6c 65 53 61 76 65 44 ····ProfileSaveD
0d6a0 61 74 61 00 00 4d 00 41-00 4c 00 43 00 4f 00 4c ata··M·A·L·C·O·L
0d6b0 00 4d 00 00 a2 9d ae fb-ee 03 70 df f6 ed 2f 6e ·M··¢·®ûî·pßöí/n
0d6c0 4e d7 17 7f 52 33 4d 41-54 43 48 00 00 00 00 00 N×··R3MATCH·····
0d6d0 00 00 00 00 00 00 00 00-00 00 00 00 00 00 00 00 ················
0d6e0 50 68 69 00 00 00 00 04-00 00 00 00 00 00 00 45 Phi············E
0d6f0 00 00 00 00 8c 7f fb 4d-e0 09 8f e4 00 00 01 17 ······ûMà··ä····
0d700 87 1f d0 00 00 01 b0 de-00 00 00 02 00 00 00 00 ··Ð···°Þ········
0d710 00 00 00 00 00 00 00 00-00 00 00 00 00 00 00 00 ················
0d720 00 00 00 00 63 51 75 69-63 6b 50 6c 61 79 49 6e ····cQuickPlayIn
0d730 66 6f 00 00 00 00 00 00-00 00 00 00 00 00 00 00 fo··············
0d740 00 00 00 00 00 00 00 00-00 00 00 00 00 00 00 00 ················
0d750 00 00 00 00 00 00 00 02-00 00 00 00 00 00 00 00 ················
0d760 00 00 00 00 00 00 00 01-00 00 00 00 00 00 00 00 ················
0d770 00 00 00 00 00 00 00 00-00 00 00 00 00 00 00 00 ················
```

Sel start = 54931, len = 31

Image 4 – Profile saved data revealing a user's name as seen in FTK Imager

In partition 3, under system update files (Partition3\$SystemUpdate) was a 6.96 MB Pirs file named su20076000_00000000. Extracting this file and opening it with wxPirs revealed a list of xexp files (Image 5). WxPirs is another open source utility commonly used by gamers seeking to modify their gaming consoles. It enables users to open PIRS, CON, and LIVE files - commonly found on the Xbox 360360 drive.



su20076000_00000000

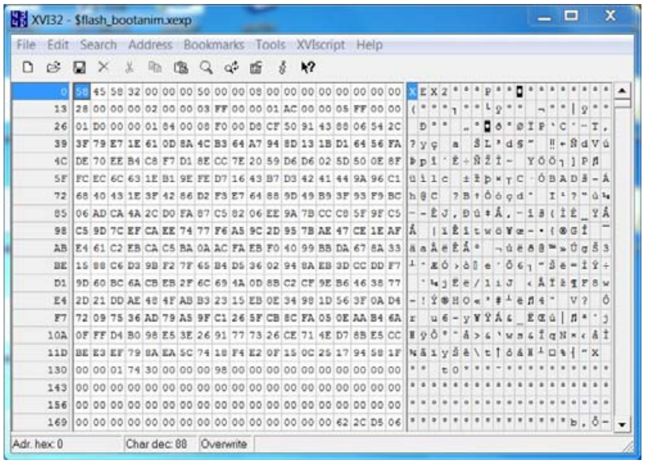
Name	Size	Cluster	Date modified
$flash_aac.xexp	16384	1	11/15/2002 12:01...
$flash_bootanim.xexp	28672	5	11/15/2002 12:01...
$flash_createprofile.xexp	30720	12	11/15/2002 12:01...
$flash_dash.xex	43663...	20	11/15/2002 12:01...
$flash_deviceselector.xexp	8192	1086	11/15/2002 12:01...
$flash_feedback.xexp	43008	1088	11/15/2002 12:01...
$flash_friends.xexp	129024	1099	11/15/2002 12:01...
$flash_gamerprofile.xexp	53248	1131	11/15/2002 12:01...
$flash_hud.xexp	51200	1144	11/15/2002 12:01...
$flash_huduiskin.xexp	55296	1157	11/15/2002 12:01...
$flash_marketplace.xexp	124928	1171	11/15/2002 12:01...
$flash_mfgbootlauncher.xexp	14336	1202	11/15/2002 12:01...
$flash_minimediaplayer.xexp	28672	1206	11/15/2002 12:01...
$flash_quickchat.xexp	133120	1213	11/15/2002 12:01...

Title : System Update
Description :
Publisher :

wxPirs - 1.1.0.42 | Gael360

Image 5 - Partition3\$SystemUpdate\ su20076000_00000000 extracted from Modio as viewed in wxPirs.

The xexp files were then extracted from wxPir and opened further with a Hex Editor (XV132). Once opened in the Hex Editor we could see that the files contained symbol table data - most likely used for linking programs to other programs. Xexp files are software development files that store information about a program and that program's functions. (Microsoft, 2005) This particular system update was found on all three of the hard drives. (Image 6)


XVI32 - $flash_bootanim.xexp
File Edit Search Address Bookmarks Tools XVIscript Help

```
   0  58 45 58 32 00 00 00 50 00 00 08 00 00 00 00 00 00 00 00
  13  28 00 00 00 02 00 00 03 FF 00 00 01 AC 00 00 05 FF 00 00
  26  01 D0 00 00 01 84 00 08 F0 00 D8 CF 50 91 43 88 06 54 2C
  39  3F 79 E7 1E 61 0D 8A 4C B3 64 A7 94 8D 13 1B D1 64 56 FA
  4C  DE 70 EE B4 C8 F7 D1 8E CC 7E 20 59 D6 D6 02 5D 50 0E 8F
  5F  FC EC 6C 63 1E B1 9E FE D7 16 43 B7 D3 42 41 44 9A 96 C1
  72  68 40 43 1E 3F 42 86 D2 F3 E7 64 88 9D 49 B9 3F 93 F9 BC
  85  06 AD CA 4A 2C D0 FA 87 C5 82 06 EE 9A 7B CC C8 5F 9F C5
  98  C5 9D 7C EF CA EE 74 77 F6 A5 9C 2D 95 7B AE 47 CE 1E AF
  AB  E4 61 C2 EB CA C5 BA 0A AC FA EB F0 40 99 BB DA 67 8A 33
  BE  15 88 C6 D3 9B F2 7F 65 B4 D5 36 02 94 8A EB 3D CC DD F7
  D1  9D 60 BC 6A CB EB 2F 6C 69 4A 0D 8B C2 CF 9E B6 46 38 77
  E4  2D 21 DD AE 48 4F AB B3 23 15 EB 0E 34 98 1D 56 3F 0A D4
  F7  72 09 75 36 AD 79 A5 9F C1 26 5F CB 8C FA 05 0E AA B4 6A
 10A  0F FF D4 B0 98 E5 3E 26 91 77 73 26 CE 71 4E D7 8B E5 CC
 11D  BE E3 EF 79 8A EA 5C 74 18 F4 E2 0F 15 0C 25 17 94 58 1F
 130  00 00 01 74 30 00 00 00 98 00 00 00 00 00 00 00 00 00 00
 143  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
 156  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
 169  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 62 2C D5 06
```

Adr. hex: 0 Char dec: 88 Overwrite

Image 6 - $flash_bootanim.xexp file extracted from wxPirs as viewed in XV132

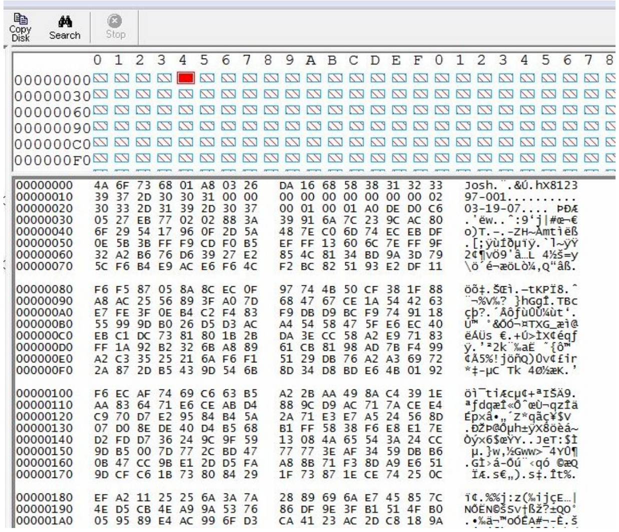
These system update files were identified as belonging to an update released by Microsoft in January 2007. (Billo, 2007) Apparently, similar to the August 2009 update discussed earlier, this was possibly another attempt to keep gamers from modifying their consoles. It is also interesting to note that the August 2009 update was not found in the system update folder on any of the drives examined.

A closer inspection of the sectors on each drive was performed using ProDiscover Basic and Digital Forensic Framework (DFF). ProDiscover Basic is the demo-freeware version of Technology Pathway's ProDiscover Forensics. It enables digital examiners to scrutinize a hard drive's clusters and files hidden in slack space. Digital Forensic Framework (DFF) is an open source cross-platform tool for examining digital media. It is a rather efficient utility which enables the user to find hidden data. While ProDiscover was not useful for drive acquisition, DFF was. Once the drives were extracted using DataRescue's DD (DrDD) however, ProDiscover was very instrumental in our research.

On two of the drives, including the one with the assimilated systems, the first piece of data observed was found on sector two - ©Axb (programming code belonging to Microsoft. In the other drive, the first sector containing data was sector four. All three drives had a rather interesting find in sector four, the name JOSH, followed by some digits and a date, as indicated in image 7 and table 1.

Drive	Name	Digits	Date
001	JOSH	97-001	03-19-07
002	JOSH	49-001	07-02-09
003	JOSH	78-001	08-07-08

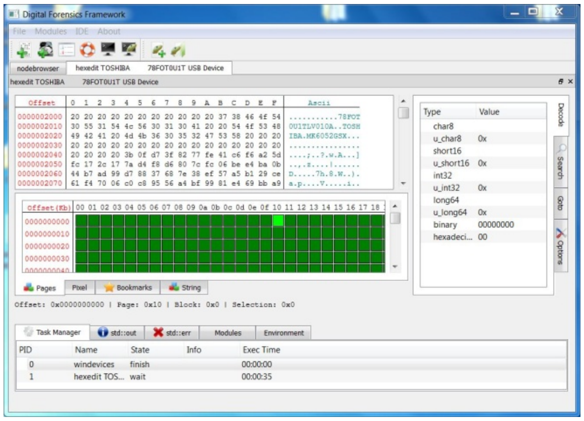
Table 1 – Sector 4 data found



```
          0 1 2 3 4 5 6 7 8 9 A B C D E F 0 1 2 3 4 5 6 7 8
00000000
00000030
00000060
00000090
000000C0
000000F0

00000000   4A 6F 73 68 01 A8 03 26   DA 16 68 58 38 31 32 33   Josh.¨.&Ú.hX8123
00000010   39 37 2D 30 30 31 00 00   00 00 00 00 00 00 00 02   97-001..........
00000020   30 33 2D 31 39 2D 30 37   00 01 00 01 A0 DE D0 C6   03-19-07.... ÞÐÆ
00000030   05 27 EB 77 02 02 88 3A   39 91 6A 7C 23 9C AC 80   .'ëw..ˆ:9'j|#œ¬€
00000040   6F 29 54 17 96 0F 2D 5A   48 7E C0 6D 74 EC EB DF   o)T.–.-ZH~Àmtìëß
00000050   0E 5B 3B FF F9 CD F0 B5   EF FF 13 60 6C 7E FF 9F   .[;ÿùÍðµïÿ.`l~ÿŸ
00000060   32 A2 B6 76 D6 39 27 E2   85 4C 81 34 BD 9A 3D 79   2¢¶vÖ9'â…L 4½š=y
00000070   5C F6 B4 E9 AC E6 F6 4C   F2 BC 82 51 93 E2 DF 11   \ö´é¬æöLò¼‚Q"âß.

00000080   F6 F5 87 05 8A 8C EC 0F   97 74 4B 50 CF 38 1F 88   öõ‡.ŠŒì.—tKPÏ8.ˆ
00000090   A8 AC 25 56 89 3F A0 7D   68 47 67 CE 1A 54 42 63   ¨¬%V‰? }hGgÎ.TBc
000000A0   E7 FE 3F 0E B4 C2 F4 83   F9 DB D9 BC F9 74 91 18   çþ?.´Âôƒ ùÛÙ¼ùt'.
000000B0   55 99 9D B0 26 D5 D3 AC   A4 54 58 47 5F E6 EC 40   U™ °&ÕÓ¬¤TXG_æì@
000000C0   EB C1 DC 73 81 80 1B 2B   DA 3E CC 58 A2 E9 71 83   ëÁÜs €.+Ú>ÌX¢éqƒ
000000D0   FF 1A 92 B2 32 6B A8 89   61 CB 81 98 AD 7B F4 99   ÿ.'²2k¨‰aË ˜{ô™
000000E0   A2 C3 35 25 21 6A F6 F1   51 29 DB 76 A2 A3 69 72   ¢Ã5%!jöñQ)Ûv¢£ir
000000F0   2A 87 2D B5 43 9D 54 6B   8D 34 D8 BD E6 4B 01 92   *‡-µC Tk 4Ø½æK.'

00000100   F6 EC AF 74 69 C6 63 B5   A2 2B AA 49 8A C4 39 1E   öì¯tiÆcµ¢+ªIŠÄ9.
00000110   AA 83 64 71 E6 CE AB D4   88 9C D9 AC 71 7A CE E4   ªƒdqæÎ«Ôˆœ Ù¬qzÎä
00000120   C9 70 D7 E2 95 84 B4 5A   2A 71 E3 E7 A5 24 56 8D   Épxâ•„´Z*qãç¥$V
00000130   07 D0 8E DE 40 D4 B5 68   B1 FF 58 38 F6 E8 E1 7E   .ÐŽÞ@Ôµh±ÿX8öèá~
00000140   D2 FD D7 36 24 9C 9F 59   13 08 4A 65 54 3A 24 CC   Òý×6$œŸY..JeT:$Ì
00000150   9D B5 00 7D 77 2C BD 47   77 77 3E AF 34 59 DB B6    µ.}w,½Gww>¯4YÛ¶
00000160   0B 47 CC 9B E1 2D D5 FA   A8 8B 71 F3 8D A9 E6 51   .GÌ›á-Õú¨‹qó ©æQ
00000170   9D CF C6 1B 73 80 84 29   1F 73 87 1E CE 74 25 0C    ÏÆ.s€„).s‡.Ît%.

00000180   EF A2 11 25 25 6A 3A 7A   28 89 69 6A E7 45 85 7C   ï¢.%%j:z(‰ijçE…|
00000190   4E D5 CB 4E A9 9A 53 76   86 DF 9E 3F B1 51 4F B0   NÕËN©šSv†ßž?±QO°
000001A0   05 95 89 E4 AC 99 6F D3   CA 41 23 AC 2D C8 18 9A   .•‰ä¬™oÓÊA#¬-È.š
```

Image 7 – Sector 4 in ProDiscover Basic

This could signify a number of things including a digital ID, some type of Microsoft numbering or cataloging scheme, or the developer's signature (i.e.; Joshua Gilpatrick, Microsoft Xbox 360 Program Manager). Later, we encountered files with a similar structure (i.e.:CON hx8123 97-001 03-19-07). Information regarding the hard drive itself was located in sector ten. (Image 8)



```
hexedit TOSHIBA    78FOT0U1T USB Device

Offset      0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F   Ascii
0000002000  20 20 20 20 20 20 20 20 20 20 20 37 38 46 4f 54  ...........78FOT
0000002010  30 55 31 54 4c 56 30 31 30 41 20 20 54 4f 53 48  0U1TLV010A..TOSH
0000002020  49 42 41 20 4d 4b 36 30 35 32 47 53 58 20 20 20  IBA.MK6052GSX...
0000002030  20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20  ................
0000002040  20 20 20 20 3b 0f d7 3f 82 77 fe 41 c6 f6 a2 5d  ....;..?.w.A...]
0000002050  fc 17 2c 17 7a d4 f8 d6 80 7c fc 06 be e4 ba 0b  ..,.z....|......
0000002060  44 b7 ad 99 d7 88 37 68 7e 38 ef 57 a5 b1 29 ce  D.....7h.8.W..).
0000002070  61 f4 70 06 c0 c8 95 56 a4 bf 99 81 e4 69 bb a9  a.p....V.....i..

Offset(Kb)  00 01 02 03 04 05 06 07 08 09 0a 0b 0c 0d 0e 0f 10 11 12 13 14 15 16 17 18
0000000000
0000000010
0000000020
0000000030
0000000040

Offset: 0x0000000000 | Page: 0x10 | Block: 0x0 | Selection: 0x0
```

Type	Value
char8	
u_char8	0x
short16	
u_short16	0x
int32	
u_int32	0x
long64	
u_long64	0x
binary	00000000
hexadeci...	00

PID	Name	State	Info	Exec Time
0	windevices	finish		00:00:00
1	hexedit TOS...	wait		00:00:35

Image 8 – Sector 10, Hard Drive Information as seen in DFF

Examining the Xbox 360 drive using EnCase can be extremely productive - depending on what you are looking for. Image 9 shows some of the data obtained on one of the drives imaged with EnCase. In this particular instance, we can see NAT (Network Address Translation) rules for a site called Bungle.net, where Halo players can have their stats tracked or purchase games and merchandise. (Bungie, 2011)

Microsoft defines three categories of Nat on their consoles- open, moderate, and closed. These attributes, or policies, control the amount of user access to Live services. The ports used are UDP (User Datagram Protocol) ports 3074, 5060, and 5061. (OAI Networks, 2011) Considering that UDP is a connectionless protocol, this could present a considerable vulnerability (ie: UDP 5060 and weak SIP or Brute Force Attack) of which users are not warned about. Thus, when gamers who are not familiar with NAT or VoIP weaknesses elect to change their settings in an effort to host games or communicate with other players, they are also unknowingly introducing more vulnerabilities into their system.

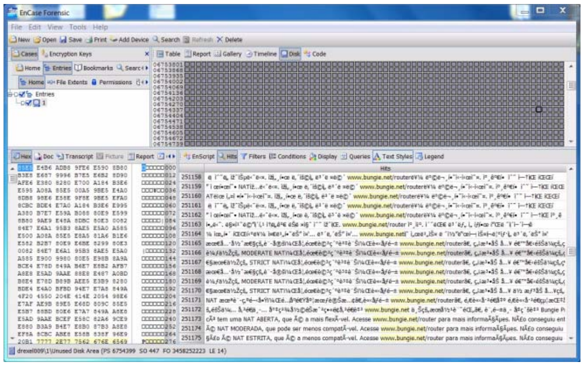
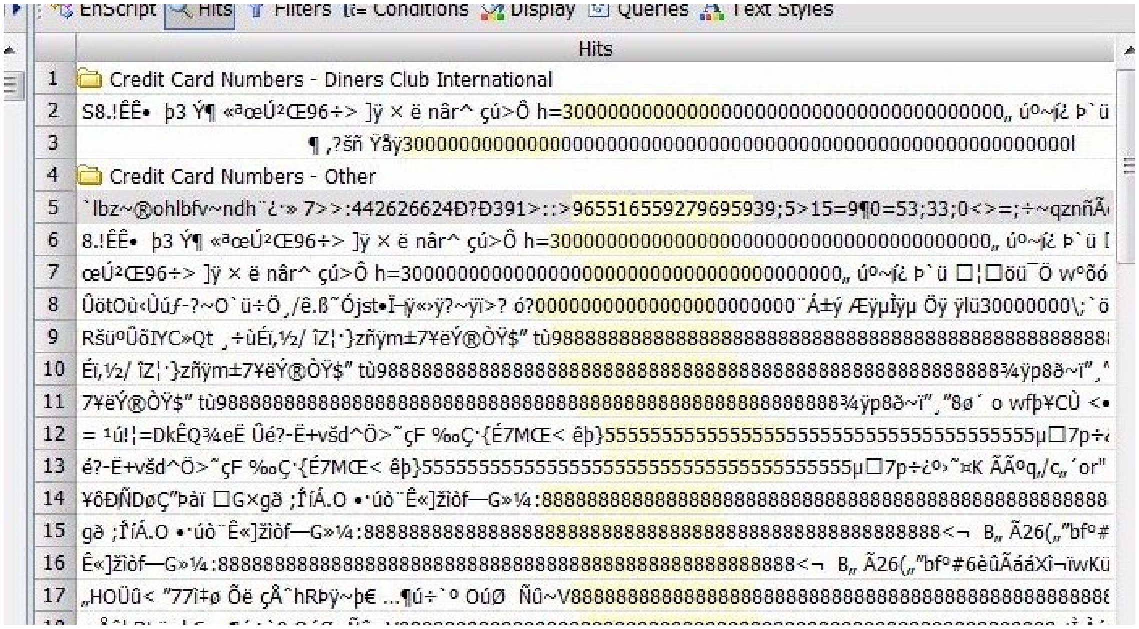

EnCase Forensic
File Edit View Tools Help
New Open Save Print Add Device Search Refresh Delete
Cases Encryption Keys
Table Report Gallery Timeline Disk Code
Home Entries Bookmarks Search
Hex Doc Transcript Picture Report
EnScript Hits Filters Conditions Display Queries Text Styles Legend
Hits

Image 9 –Microsoft's defined NAT as viewed in EnCase

Another benefit of utilizing EnCase is its ability to discover credit card information on a hard drive by looking for numbers encoded with ASCII digit characters that match valid credit card company identifiers. These numbers are then run against the Luhr formula (an algorithm used to validate credit cards, social security numbers, and other identification numbers). (University of Michigan, 2008) Performing a fast scan on one of the drives resulted in a possible credit card hit. (Image 10) Although this does not definitively prove there are any credit card numbers on the hard drive, it is highly probable given the results obtained. The Bank Identification Number in this hit identifies this as a Bank of America Discover Card. (BinBD, 2011)

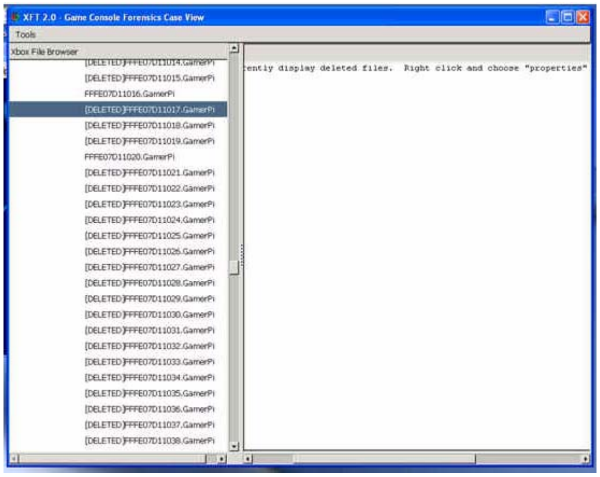


```
                    Hits
1   Credit Card Numbers - Diners Club International
2   S8.!ÊÊ•  þ3 Ý¶ «ªœÚ²Œ96÷> ]ÿ × ë nâr^ çú>Ô h=30000000000000000000000000000000000000„ úº~í¿ Þ`ü
3                           ¶ ,?šñ Ÿåÿ300000000000000000000000000000000000000000000000000000000l
4   Credit Card Numbers - Other
5   `lbz~®ohlbfv~ndh¨¿·» 7>>:442626624Ð?Ð391>::>965516559279695939;5>15=9¶0=53;33;0<>=;÷~qznñÃ
6   8.!ÊÊ•  þ3 Ý¶ «ªœÚ²Œ96÷> ]ÿ × ë nâr^ çú>Ô h=30000000000000000000000000000000000000„ úº~í¿ Þ`ü [
7   œÚ²Œ96÷> ]ÿ × ë nâr^ çú>Ô h=30000000000000000000000000000000000000„ úº~í¿ Þ`ü □¦□öü¯Ö wºõó
8   ÛötOù‹Ùúƒ-?~O`ü÷Ö ¸/ê.ß˜Ójst•Ī─ÿ«›ÿ?~ÿï>? ó?0000000000000000000000¨Á±ý ÆÿµĬÿµ Öÿ ÿlü30000000\;`ö
9   RšüºÛõIYC»Qt ¸÷ùÉí,½/ îZ¦·}zñÿm±7¥ëÝ®ÒŸ$" tù988888888888888888888888888888888888888888888888
10  Éí,½/ îZ¦·}zñÿm±7¥ëÝ®ÒŸ$" tù98888888888888888888888888888888888888888¾ÿp8ð~ï"¸"
11  7¥ëÝ®ÒŸ$" tù98888888888888888888888888888888888888888888888888888¾ÿp8ð~ï"¸"8ø´ o wfþ¥CÙ <•
12  = ¹ú!¦=DkÊQ¾eË Ûé?-Ë+všd^Ö>˜çF ‰Ç·{É7MŒ< êþ}555555555555555555555555555555555555555µ□7p÷¿
13  é?-Ë+všd^Ö>˜çF ‰Ç·{É7MŒ< êþ}555555555555555555555555555555555555555µ□7p÷¿º›˜¤K ÃÃºq,/c„´or"
14  ¥ôÐÑDøÇ"Þàï □G×gð ;ĬíÁ.O •·úò¨Ê«]žìòf—G»¼:888888888888888888888888888888888888888888888888
15  gð ;ĬíÁ.O •·úò¨Ê«]žìòf—G»¼:8888888888888888888888888888888888888888888888888<¬  B„ Ã26(„"bfº#
16  Ê«]žìòf—G»¼:8888888888888888888888888888888888888888888888888<¬  B„ Ã26(„"bfº#6èûÃáâXî¬ïwKü
17  „HOÜû< "77î‡ø Õë çÅˆhRÞÿ~þ€ …¶ú÷`º OúØ  Ñû~V8888888888888888888888888888888888888888888888
```

Image 10 – EnCase credit card hit

A new tool recently developed to address the need for forensic software capable of obtaining information from nontraditional devices is XFT 2.0 Game Console Forensic Toolkit, developed by David Collins, a computer scientist at Sam Hoston State University in Texas and distributed by Protowise Labs. (Protowise Labs, 2011) XFT 2.0 features both FATX and XTAF (derived from MS-DOS) file system mounting and preview, file hashing, recovery of deleted files, and file type identification. It is designed to run on Windows operating systems and features a user-friendly interface, although when tested on both Windows XP and WIN 7, the utility did not run as smoothly on WIN 7.

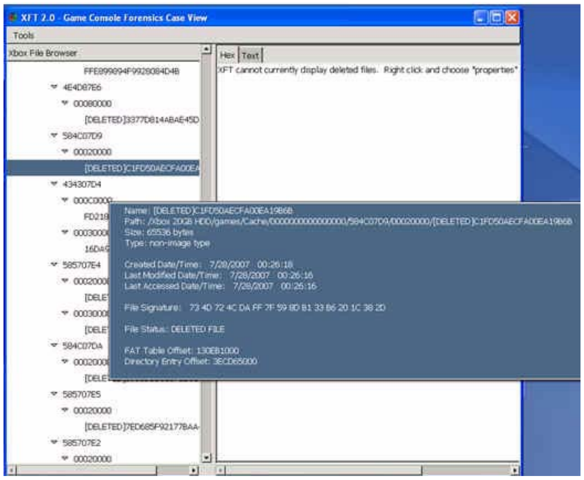
While we were able to see the names of deleted files, we were unable to actually view their contents. When attempting to view deleted files the message "XFT cannot currently display deleted files. Right click and choose "properties" for disk offset and starting cluster" was obtained, as seen ini mage 11. By right-clicking on a selected deleted file, the user is given the option to export, hash, or view the properties of that file (Image 12). This information can prove very useful for law enforcement agencies in cases involving child sexual exploitation where the hash values obtained can be compared against known values from the CVIP (Child Victim Identification Program) database (FBI, 2011). Although other forensic tools tested performed hashes, XFT was the only tool which showed the deleted files from the Xbox 360 drives.



```
[DELETED]FFFE07D11014.GamerPi
[DELETED]FFFE07D11015.GamerPi
FFFE07D11016.GamerPi
[DELETED]FFFE07D11017.GamerPi
[DELETED]FFFE07D11018.GamerPi
[DELETED]FFFE07D11019.GamerPi
FFFE07D11020.GamerPi
[DELETED]FFFE07D11021.GamerPi
[DELETED]FFFE07D11022.GamerPi
[DELETED]FFFE07D11023.GamerPi
[DELETED]FFFE07D11024.GamerPi
[DELETED]FFFE07D11025.GamerPi
[DELETED]FFFE07D11026.GamerPi
[DELETED]FFFE07D11027.GamerPi
[DELETED]FFFE07D11028.GamerPi
[DELETED]FFFE07D11029.GamerPi
[DELETED]FFFE07D11030.GamerPi
[DELETED]FFFE07D11031.GamerPi
[DELETED]FFFE07D11032.GamerPi
[DELETED]FFFE07D11033.GamerPi
[DELETED]FFFE07D11034.GamerPi
[DELETED]FFFE07D11035.GamerPi
[DELETED]FFFE07D11036.GamerPi
[DELETED]FFFE07D11037.GamerPi
[DELETED]FFFE07D11038.GamerPi
```

```
ently display deleted files.  Right click and choose "properties"
```

Image 11 – XFT message – "…cannot currently display dleted files."



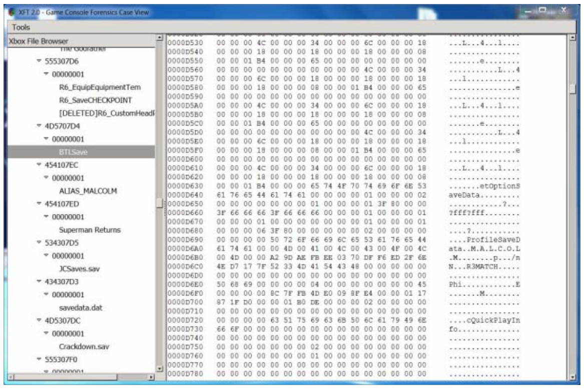
```
FFE899894F9928084D48
▾ 4E4D87E6
  ▾ 00080000
      [DELETED]3377D814ABAE45D
▾ 584C07D9
  ▾ 00020000
      [DELETED]C1FD50AECFA00EA
▾ 434307D4
  ▾ 000C0000
      FD218
  ▾ 0003000
      16DA
▾ 585707E4
  ▾ 0002000
      [DELE
  ▾ 0003000
      [DELE
▾ 584C07DA
  ▾ 0002000
      [DELE
▾ 585707E5
  ▾ 00020000
      [DELETED]7ED685F92177BAA-
▾ 585707E2
  ▾ 00020000
```

```
XFT cannot currently display deleted files.  Right click and choose "properties"
```

```
Name: [DELETED]C1FD50AECFA00EA1986B
Path: /Xbox 20GB HDD/games/Cache/0000000000000000/584C07D9/00020000/[DELETED]C1FD50AECFA00EA1986B
Size: 65536 bytes
Type: non-image type

Created Date/Time:   7/28/2007   00:26:18
Last Modified Date/Time:   7/28/2007   00:26:16
Last Accessed Date/Time:   7/28/2007   00:26:16

File Signature:   73 4D 72 4C DA FF 7F 59 8D B1 33 B6 20 1C 38 2D

File Status: DELETED FILE

FAT Table Offset: 130EB1000
Directory Entry Offset: 3ECD65000
```

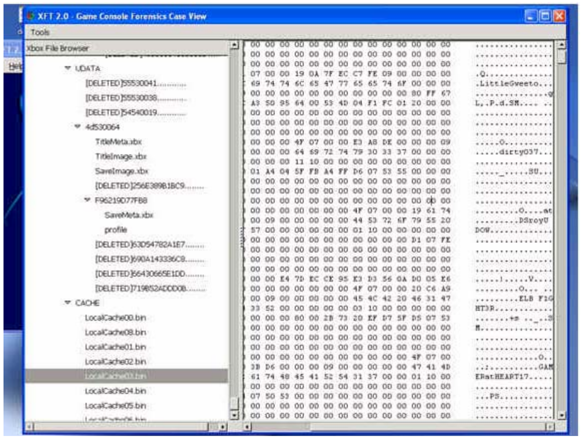
Image 12 – Viewing the properties of a deleted file in XFT

Other information discovered with XFT 2.0 included user names (Image 13) and the user's player list containing the gamer tags of other Xbox 360 players. (Image 14) This finding is extremely significant because it can not only aid law enforcement seeking to establish a connection between users, but it can also pose a risk to anyone who has been in contact with a user whose system has been compromised. Gamer tags can be searched through any number of gamer databases or social networking sites to gain additional information about a player.



```
0000D530   00 00 00 4C 00 00 00 34 00 00 00 6C 00 00 00 18   ...L...4...l....
0000D540   00 00 00 18 00 00 00 18 00 00 00 18 00 00 00 08   ................
0000D550   00 00 01 B4 00 00 00 65 00 00 00 00 00 00 00 00   .......e........
0000D560   00 00 00 00 00 00 00 00 00 00 00 4C 00 00 00 34   ...........L...4
0000D570   00 00 00 6C 00 00 00 18 00 00 00 18 00 00 00 18   ...l............
0000D580   00 00 00 18 00 00 00 08 00 00 01 B4 00 00 00 65   ...............e
0000D590   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D5A0   00 00 00 4C 00 00 00 34 00 00 00 6C 00 00 00 18   ...L...4...l....
0000D5B0   00 00 00 18 00 00 00 18 00 00 00 18 00 00 00 08   ................
0000D5C0   00 00 01 B4 00 00 00 65 00 00 00 00 00 00 00 00   .......e........
0000D5D0   00 00 00 00 00 00 00 00 00 00 00 4C 00 00 00 34   ...........L...4
0000D5E0   00 00 00 6C 00 00 00 18 00 00 00 18 00 00 00 18   ...l............
0000D5F0   00 00 00 18 00 00 00 08 00 00 01 B4 00 00 00 65   ...............e
0000D600   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D610   00 00 00 4C 00 00 00 34 00 00 00 6C 00 00 00 18   ...L...4...l....
0000D620   00 00 00 18 00 00 00 18 00 00 00 18 00 00 00 08   ................
0000D630   00 00 01 B4 00 00 00 65 74 4F 70 74 69 6F 6E 53   .......etOptionS
0000D640   61 76 65 44 61 74 61 00 00 00 00 01 00 00 00 02   aveData.........
0000D650   00 00 00 00 00 00 00 01 00 00 00 01 3F 80 00 00   ............?...
0000D660   3F 66 66 66 3F 66 66 66 00 00 00 01 00 00 00 01   ?fff?fff........
0000D670   00 00 00 01 00 00 00 00 00 00 00 01 00 00 00 01   ................
0000D680   00 00 00 06 3F 80 00 00 00 00 00 02 00 00 00 00   ....?...........
0000D690   00 00 00 00 50 72 6F 66 69 6C 65 53 61 76 65 44   ....ProfileSaveD
0000D6A0   61 74 61 00 00 4D 00 41 00 4C 00 43 00 4F 00 4C   ata..M.A.L.C.O.L
0000D6B0   00 4D 00 00 A2 9D AE FB EE 03 70 DF F6 ED 2F 6E   .M........p.../n
0000D6C0   4E D7 17 7F 52 33 4D 41 54 43 48 00 00 00 00 00   N...R3MATCH.....
0000D6D0   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D6E0   50 68 69 00 00 00 00 04 00 00 00 00 00 00 00 45   Phi............E
0000D6F0   00 00 00 00 8C 7F FB 4D E0 09 8F E4 00 00 01 17   .......M........
0000D700   87 1F D0 00 00 01 B0 DE 00 00 00 02 00 00 00 00   ................
0000D710   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D720   00 00 00 00 63 51 75 69 63 6B 50 6C 61 79 49 6E   ....cQuickPlayIn
0000D730   66 6F 00 00 00 00 00 00 00 00 00 00 00 00 00 00   fo..............
0000D740   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D750   00 00 00 00 00 00 00 02 00 00 00 00 00 00 00 00   ................
0000D760   00 00 00 00 00 00 00 01 00 00 00 00 00 00 00 00   ................
0000D770   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
0000D780   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

Image 13 – User name viewed in XFT


XFT 2.0 - Game Console Forensics Case View
Tools
Xbox File Browser
UDATA
[DELETED]55530041............
[DELETED]55530038............
[DELETED]54540019............
4d530064
TitleMeta.xbx
TitleImage.xbx
SaveImage.xbx
[DELETED]256E389B1BC9........
F96219D77FB8
SaveMeta.xbx
profile
[DELETED]63D54782A1E7........
[DELETED]690A143336C8........
[DELETED]66430665E1DD........
[DELETED]719852ADDD08........
CACHE
LocalCache00.bin
LocalCache08.bin
LocalCache01.bin
LocalCache02.bin
LocalCache03.bin
LocalCache04.bin
LocalCache05.bin
LocalCache06.bin

```
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
07 00 00 19 0A 7F EC C7 FE 09 00 00 00 00   .Q..............
69 74 74 6C 65 47 77 65 65 74 6F 00 00 00   .LittleGweeto...
00 00 00 00 00 00 00 00 00 00 00 80 FF 67   ...............g
A3 50 95 64 00 53 4D 04 F1 FC 01 20 00 00   L..P.d.SM.... ..
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 4F 07 00 00 E3 A8 DE 00 00 00 09   .....O..........
00 00 00 64 69 72 74 79 30 33 37 00 00 00   .....dirty037...
00 00 00 11 10 00 00 00 00 00 00 00 00 00   ................
01 A4 04 5F FB A4 FF D6 07 53 55 00 00 00   ....._.....SU...
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 4F 07 00 00 19 61 74   .........O....at
00 09 00 00 00 00 00 44 53 72 6F 79 55 20   .........DSroyU
57 00 00 00 00 00 00 01 10 00 00 00 00 00   DOW.............
00 00 00 00 00 00 00 00 00 00 00 D1 07 FE   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 E4 7D EC CE 95 E3 D3 56 0A D0 05 E6   .....}.....V....
00 00 00 00 00 00 00 4F 07 00 00 20 C6 A9   .........O... ..
00 09 00 00 00 00 00 45 4C 42 20 46 31 47   .........ELB F1G
33 52 00 00 00 00 00 03 10 00 00 00 00 00   HT3R............
00 00 00 80 00 2B 73 20 EF D7 5F D5 07 53   .......+s .._..S
00 00 00 00 00 00 00 00 00 00 00 00 00 00   M...............
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 4F 07 00   .............O..
3B D6 00 00 00 09 00 00 00 00 00 47 41 4D   ..;..........GAM
61 74 48 45 41 52 54 31 37 00 00 01 10 00   ERatHEART17.....
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
07 50 53 00 00 00 00 00 00 00 00 00 00 00   ...PS...........
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

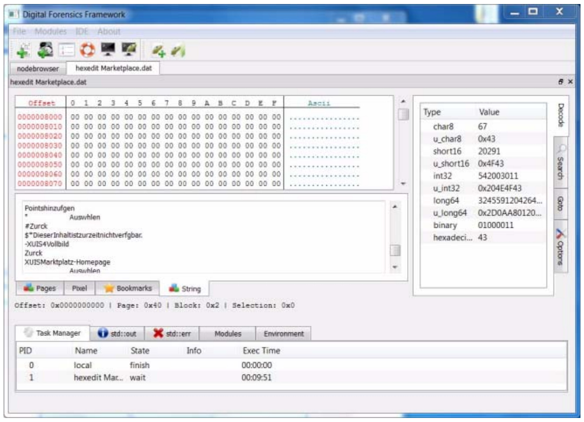
Image 14 – Cache showing a player's list in XFT

While XFT does not enable users to read larger files such as databases, it does enable the option to export the data. In one example, we exported the marketplace database for closer examination using notepad. After a quick look through the file, we came to the text "Purchase History Items", and decided to take a closer look in DFF. Once in DFF, strings of text in German, Italian, and French were discovered. (Table 2) (Image 15)

Item	Language	Information
per maggiori informazion**.**	Italian	for greater information
ore dopo aver selezionato	Italian	hours after to have selected
inhalt ist zur zeit nicht	German	contents are at present not
dejouer les	French	to thwart them

Table 2 – Example of foreign languages found in marketplace.dat file

Because Xbox 360 is an international platform, one might expect to see multiple languages in the marketplace data file. However, it presents forensic examiners with another challenge and should be kept in mind when examining the contents of the drive.



```
Offset      0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F   Ascii
0000008000  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008010  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008020  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008030  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008040  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008050  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008060  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000008070  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
```

```
Pointshinzufgen
*          Auswhlen
#Zurck
$*DieserInhaltistzurzeitnichtverfgbar.
-XUIS4Vollbild
Zurck
XUISMarktplatz-Homepage
           Auswhlen
```

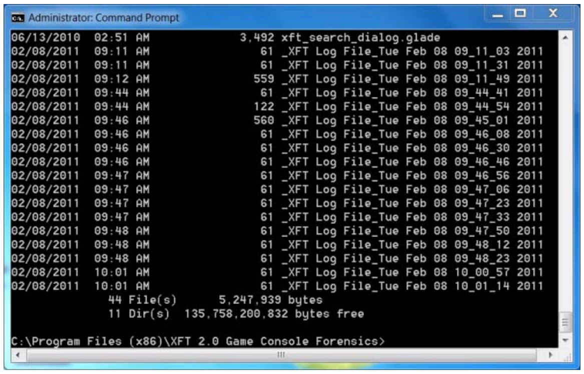
```
Offset: 0x0000000000 | Page: 0x40 | Block: 0x2 | Selection: 0x0
```

Type	Value
char8	67
u_char8	0x43
short16	20291
u_short16	0x4F43
int32	542003011
u_int32	0x204E4F43
long64	3245591204264...
u_long64	0x2D0AA80120...
binary	01000011
hexadeci...	43

PID	Name	State	Info	Exec Time
0	local	finish		00:00:00
1	hexedit Mar...	wait		00:09:51

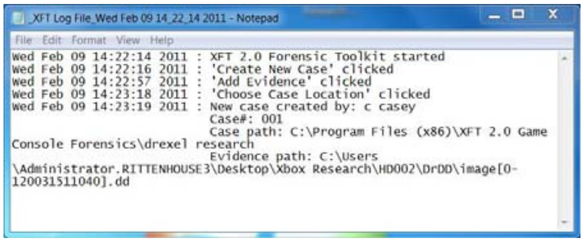
Image 15 – Marketplace database in DFF

Although XFT was designed specifically to examine Xbox 360 drives, we were unable to acquire the drive through the program without first extracting the data using DrDD. While the drives were tested both before and after acquisition, it is problematic at best to claim with any certainly that the extracted data was not altered during the extraction because we were transferring FATX data using tools, which even if tested and given a green light by NIST, were not designed to acquire or examine FATX files. This can create quite a quagmire when working within a legal framework. One feature of XFT which addresses this dilemma is its ability to keep an electronic "chain of custody" of the data being examined. Each time data is accessed through the program, it is logged in a file until the case is manually is closed. (Images 16 and 17)



```
06/13/2010  02:51 AM             3,492 xft_search_dialog.glade
02/08/2011  09:11 AM                61 _XFT Log File_Tue Feb 08 09_11_03 2011
02/08/2011  09:11 AM                61 _XFT Log File_Tue Feb 08 09_11_31 2011
02/08/2011  09:12 AM               559 _XFT Log File_Tue Feb 08 09_11_49 2011
02/08/2011  09:44 AM                61 _XFT Log File_Tue Feb 08 09_44_41 2011
02/08/2011  09:44 AM               122 _XFT Log File_Tue Feb 08 09_44_54 2011
02/08/2011  09:46 AM               560 _XFT Log File_Tue Feb 08 09_45_01 2011
02/08/2011  09:46 AM                61 _XFT Log File_Tue Feb 08 09_46_08 2011
02/08/2011  09:46 AM                61 _XFT Log File_Tue Feb 08 09_46_30 2011
02/08/2011  09:46 AM                61 _XFT Log File_Tue Feb 08 09_46_46 2011
02/08/2011  09:47 AM                61 _XFT Log File_Tue Feb 08 09_46_56 2011
02/08/2011  09:47 AM                61 _XFT Log File_Tue Feb 08 09_47_06 2011
02/08/2011  09:47 AM                61 _XFT Log File_Tue Feb 08 09_47_23 2011
02/08/2011  09:47 AM                61 _XFT Log File_Tue Feb 08 09_47_33 2011
02/08/2011  09:48 AM                61 _XFT Log File_Tue Feb 08 09_47_50 2011
02/08/2011  09:48 AM                61 _XFT Log File_Tue Feb 08 09_48_12 2011
02/08/2011  09:48 AM                61 _XFT Log File_Tue Feb 08 09_48_23 2011
02/08/2011  10:01 AM                61 _XFT Log File_Tue Feb 08 10_00_57 2011
02/08/2011  10:01 AM                61 _XFT Log File_Tue Feb 08 10_01_14 2011
              44 File(s)      5,247,939 bytes
              11 Dir(s)  135,758,200,832 bytes free

C:\Program Files (x86)\XFT 2.0 Game Console Forensics>
```

Image 16 – XFT access log as viewed in DOS


_XFT Log File_Wed Feb 09 14_22_14 2011 - Notepad
File Edit Format View Help

```
Wed Feb 09 14:22:14 2011 : XFT 2.0 Forensic Toolkit started
Wed Feb 09 14:22:16 2011 : 'Create New Case' clicked
Wed Feb 09 14:22:57 2011 : 'Add Evidence' clicked
Wed Feb 09 14:23:18 2011 : 'Choose Case Location' clicked
Wed Feb 09 14:23:19 2011 : New case created by: c casey
                           Case#: 001
                           Case path: C:\Program Files (x86)\XFT 2.0 Game
Console Forensics\drexel research
                           Evidence path: C:\Users
\Administrator.RITTENHOUSE3\Desktop\Xbox Research\HD002\DrDD\image[0-
120031511040].dd
```

Image 17 – Example of an XFT access log

Up until this point, we looked exclusively at Windows based tools. However, when examining an Xbox 360 drive, investigators can also obtain valuable information using Linux. Upon initial assessment, examiners can try to boot the console with Linux to determine if the system has been modified. Drives mounted to a computer running Linux (or machines booted with a Linux CD or bootable USB) can be searched using common Linux commands such as grep to look for files, or a defined string of text. The abundance of gamer sites and forums dedicated to Xbox 360 modding with Linux may also prove a valuable resource. If the budget is available, an analysis workstation can be built and dedicated to examining Xbox 360 drives. It is recommended however that the hardware of the machine being deployed for this workstation is compatible with the latest Linux kernel (2.6). (Paul K. Burkea P. C., 2006) During our research, we encountered repeated kernel errors while trying to examine the test drives in Linux.

5. CONCLUSION

Although many of the tools tested discovered the same or identical data, there was no single tool adept enough to perform the task independently. Furthermore, with the exception of XFT, evidence obtained from these tools may not necessarily be admissible in a court of law.

It is no longer feasible to examine devices such as gaming consoles, smart phones, and iPods using *"electronic ethnocentricity"*. Using computers to measure where data is, and how it should be structured or stored, will simply no longer suffice. As devices evolve, so must the examiner's methodologies. Technology has passed the age where we can use one or two tools, and by pushing a few buttons, have all of our evidence appear before our eyes and arranged automatically into neat little reports. This is not to suggest that program developers should not continue to create software to address these new needs, but rather that digital investigators may need to think outside the "box" when examining devices like gaming consoles.

By looking at a small sampling of drives using multiple tools and operating systems, we were slowly able to begin constructing a model of the Xbox 360 gaming console structure. While this was just a sampling of Microsoft's Xbox 360 architecture, it enabled us to find two user names, city, a user profile, a cache containing a player's list, and a credit card number. When we reference the seller address from eBay we are able to have a name, address, and credit card number; a complete identity. If the investigators had stuck exclusively to conventional techniques, or tools designed to acquire data from computer hard drives, they would have missed some of this data.

Given the increase of crimes using gaming consoles such as the Xbox 360, there needs to be more research conducted to help determine appropriate tools and approaches for forensically sound data identification and acquisition.

6. FUTURE WORK

Future work includes testing additional tools to determine the best acquisition method for gaming consoles, specifically the Xbox 360. Furthermore, the researchers aim to establish and verify date/time stamps on Xbox 360 data. For example, the researchers were able to recover "buddy lists," and if you are able to cross reference actions with the "buddy lists" and data/time stamps you would be able to build activity and communication timelines. This would be extremely helpful in criminal cases such as child exploitation, fraud or other criminal activities.

The researchers will also continue to work on developing best acquisition methods for emerging, non-traditional devices such as smart phones and other internet capable devices.

7. ABOUT THE AUTHORS

Dr. Podhradsky is an Assistant Professor of Computing and Security Technology at Drexel University. Dr. D'Ovidio is a Professor of Criminal Justice at Drexel University, and Cindy Casey is a student in the Computing and Security Technology program at Drexel University.

8. REFERENCES:

Becker, D. (2001, 1 6). *Microsoft got game: Xbox 360 unveiled*. Retrieved 11 17, 2010, from CNET News :
http://news.cnet.com/Microsoft-got-game-Xbox 360-unveiled/2100-1040_3-250632.html?tag=untagged

Berardini, C. (2005, 12 5). *The Xbox 360 360 System Specifications* . Retrieved 11 17, 2010, from Team Xbox 360: http://hardware.teamXbox 360.com/articles/Xbox 360/1144/The-Xbox 360-360-System-Specifications/p1

Billo, J. (2007, 1 23). *Xbox 360 360 dashboard update and efuses* . Retrieved 1 22, 2011, from Jake Billo's weblog: http://jakebillo.com/Xbox 360-360-dashboard-update-and-efuses

BinBD. (2011). *Bin Checker*. Retrieved 2 7, 2011, from Bin Database Search - Bin Checker: http://www.bindb.com/bin-database.html

Bloomberg Businessweek. (2010, 3 11). *Microsoft's Xbox 360 Sales Beat Wii, PS3 in February on "BioShock"*. Retrieved 11 17, 2010, from Bloomsberg.com:
http://www.businessweek.com/news/2010-03-11/microsoft-s-Xbox 360-sales-beat-wii-ps3-in-february-on-bioshock-.html

Bolt, S. (2011). *Xbox 360 360 Forensics; A Digital Foresnics Guide to Examining Artifacts.* Burlington : Syngress.

Bungie. (2011, 2 4). *Halo*. Retrieved 2 7, 2011, from Bungie.net: http://www.bungiestore.com/

Bullock, H. (2009, 2 6). *Accused Sexual Predator Edward Stout Met Victim Through Xbox*. Retrieved 2 18, 2011, from KSFN-TV: http://abclocal.go.com/kfsn/story?section=news/local&id=6643907

Bush, E. (2008, 10 24). *Virginia Man Arrested for Child Pornography over Xbox Live*. Retrieved 2, 18, 2011, from Planet Xbox 360:
http://www.planetxbox360.com/article_5559/Virginia_Man_Arrested_for_Child_Pornography_o

Carmody, T. (2010, 11 3). *How Motion Detection Works in Xbox 360 Kinect*. Retrieved 11 17, 2010, from Wired: http://www.wired.com/gadgetlab/2010/11/tonights-release-Xbox 360-kinect-how-does-it-work/all/1

Cavalli, E. (2009, 3 17). *Animal Crossing is Pedophile Haven*. Retrieved 2 18, 2011, from Wired: http://www.wired.com/gamelife/2009/03/missouri-police/

Computer Gyaan. (2010, 11 2). *Disk formatting and Data recovery*. Retrieved 1 14, 2011, from Computer Gyaan:
http://www.muamat.com/classifieds/174/events/2010-12-30/11663_Disk_formatting_and_Data_recovery.html

craigslist. (2010). *Offical Site*. Retrieved 12 23, 2010, from craigslist: http://www.craigslist.org

Constantin, L. (2010, 10 19). *Phishers Target Xbox Players Via Fake Gamertag Changer*. Retrieved 2 18, 2011, from Softpedia: http://news.softpedia.com/news/Phishers-Target-Xbox-Players-via-Fake-Gamertag-Changer-161812.shtml

Criminal Intelligence Service Canada (2007). *Integrated Threat Assessment Methodology*.

Ottowa, Ontario: Criminal Intelligence Service Canada.

Deleon, N. (2008, 8 5). *Be Careful, There's a Phishing Scam Going Around Xbox Live*. Retrieved 2 18, 2011, from CrunchGear: http://www.crunchgear.com/2008/08/05/be-careful-theres-a-phishing-scam-going-around-xbox-live/

Ebay. (2010). *Xbox 360 Console Listings*. Retrieved 11 30, 2010, from Ebay Online Auction : http://video-games.shop.ebay.com/Systems-/139971/i.html?_nkw=Xbox 360+console&_catref=1&_fln=1&_trksid=p3286.c0.m282

Evers, J. (2007, 3 20). *Microsoft Probes Possible Xbox Live Fraud*. Retrieved 2 16, 2011, from CNet News: http://news.cnet.com/2100-7349_3-6169060.html

FBI. (2011). *Child Victim Identification Program (CVIP)*. Retrieved 2 8, 2011, from Department of Justice/Federal Bureau of Investigation: http://foia.fbi.gov/cvip.htm

Federal Trade Commission . (2007). *2006 Identity Theft Survey Report.* McLean: Synovate.

Free60 Project. (2009, 8 11). *849x System Update*. Retrieved 1 4, 2011, from Free60 Project Achieves: http://free60.org/old/849x_System_Update.html

Fried, I. (2005, 5 25). *Microsoft Plugs Phishing Hole in Xbox Site*. Retrieved 2 18 2011, from CNet News: http://news.cnet.com/Microsoft-plugs-phishing-hole-in-Xbox-site/2100-1029_3-5720241.html

Fujji, M. (2010, 4 15). *Man Arrested for Intimidating Witness Over Xbox Live*. Retrieved 2 18, 2011, from College News: http://www.collegenews.com/index.php?/article/witness_intimidated_over_xbox_live_041520101235235/

Harris, M. (2009, 1 23). *Online Games Open Door to ID Theft*. Retrieved 2 17, 2011, from Tech Radar: http://www.techradar.com/news/internet/online-games-open-door-to-id-theft-610380

Hill, C. (2009, 1 29). *Teenage Boy Accused of Repeatedly Raping 12-Year Old He Met on Xbox Live*. Retrieved 2 18, 2011, from NY Daily News: http://www.nydailynews.com/money/2009/01/29/2009-01-29_teenage_boy_accused_of_repeatedly_raping.html

Hitt, B. (2011, 1 8). *Women Raped Boy, 13, She Met Playing Xbox Online*. Retrieved 2 19, 2011, from KTLA: http://www.ktla.com/news/landing/ktla-xbox-mom-rape,0,2120242.story

Huang, A. ". (2001, 5 26). *Keeping Secrets in Hardware: the Microsoft Xbox 360 Case Study*. Retrieved 1 7, 2011, from Massachusetts Institute of Technology - Artificial Intelligence Laboratory: http://web.mit.edu/bunnie/www/proj/anatak/AIM-2002-008.pdf

Javelin Strategy and Research. (2010, 10 2). *2010 Identity Fraud Survey Report: Identity Fraud Continues to Rise.* Retrieved 11 16, 2010, from Javelin Strategy and Research Library: https://www.javelinstrategy.com/research/brochures/Brochure-170

Lemos, R. (2007, 3 23). *Account Pretexters Plague Xbox Live*. Retrieved 2 18, 2011, from The Register: http://www.theregister.co.uk/2007/03/23/xbox_live_pretexting/

McHugh, M. (2011, 2 4). *Pirated Microsoft Software Funded Mexican Drug Cartel*. Retrived 2 18, 2011, from Digital Trends: http://www.digitaltrends.com/computing/pirated-microsoft-software-funded-mexican-drug-cartel/

McMillan, R. (2011 2 14). *Spanish Police Arrest Alleged Nintendo Hacker*. Retrieved 2 18, 2011, from PCWorld: http://www.pcworld.com/businesscenter/article/219598/spanish_police_arrest_alleged_nintendo_hacker.html

Microsoft . (2010, 10 20). *Article ID: 906502 - How to format an Xbox 360 360 Hard Drive or Memory Unit*. Retrieved 12 23, 2010, from Microsoft Support : http://support.microsoft.com/kb/906502

Microsoft. (2005). *.exp Files as Linker Input*. Retrieved 1 5, 2011, from MSDN Library, Link Input Files: http://msdn.microsoft.com/en-us/library/se8y7dcs(v=vs.80).aspx

Microsoft. (2010, 10 20). *How to format an Xbox 360 360 Hard Drive or Memory Unit*. Retrieved 11 30, 2010, from Microsoft Support Search Microsoft SupportSearch Microsoft.comSearch the web : http://support.microsoft.com/kb/906502

Microsoft. (2010, 10). *Xbox 360 LIVE and Games for Windows LIVE Terms of Use*. Retrieved 1 12, 2011, from Microsoft Xbox 360: http://www.Xbox 360.com/en-US/Legal/livetou

Netflix. (2011). *How does Netflix work?* Retrieved 2 7, 2011, from NetFlix: http://www.netflix.com/Default?mqso=80012928

OAI Networks. (2011). *Strict, Moderate, and Open NAT - Load balancing Xbox 360 Game Servers* . Retrieved 2 7, 2011, from Tech Tips: http://www.cainetworks.com/support/how-to-NAT-strict-open.html

Official Xbox 360 Magazine staff . (2005, 12 13). *The Complete Hisotry of Xbox 360*. Retrieved 11 17, 2010, from CVG Gaming : http://www.computerandvideogames.com/article.php?id=131066

Paul K. Burkea, P. C. (2006). *Xbox 360 Forensics.* Retrieved 11 18, 2010, from Journal of Digital Forensic Practice: http://dx.doi.org/10.1080/15567280701417991

Paul K. Burkea, P. C. (2006). *Xbox 360 Forensics.* Retrieved 2 9, 2011, from Journal of Digital Forensic Practice, Volume 1, Issue 4 December 2006 , pages 275 - 282 : http://www.informaworld.com/smpp/section?content=a779635437&fulltext=713240928

Peterson, D. (2010, 8 30). *Former Walt Disnet World Employee Arrested on Xbox Child Porn Charges*. Retrieved 2 17, 2011, from Examiner: http://www.examiner.com/disney-travel-in-national/former-walt-disney-world-employee-arrested-on-xbox-child-porn-charges

Potter, N. (2009, 3 13). *PlayStation Sex Crime: Criminal Used Video Game to Get Girl's Naked Pictures*. Retrieved 2 18, 2011, from ABC News: http://abcnews.go.com/print?id=7009977

Protowise Labs. (2011). *XFT 2.0 Game Console Forensics is Released*. Retrieved 2 8, 2011, from XFT 2.0 Game Console: http://protowise.com/?tag=xft-Xbox 360-forensics

Rivington, J. (2007, 8 20). *Wii and PS3 Vulnerable to Hacks and Phishing*. Retrieved 2 18 2011, from Tech Radar: http://www.techradar.com/news/gaming/consoles/wii-and-ps3-vulnerable-to-hacks-and-phishing-161313

Snow, B. (2009, 12 13). *Gamer Arrested for Shooting Threats on Xbox Live*. Retrieved 2 18, 2011, from GamePro: http://www.gamepro.com/article/news/152848/gamer-arrested-for-shooting-threats-on-xbox-live/

Team Xbox 360. (2005, 9 8). *Xbox 360 Live Facts 'n Stats.* Retrieved 11 30, 2010, from Team Xbox 360: http://news.teamXbox 360.com/Xbox 360/9194/Xbox 360-Live-Facts-n-Stats/

Technology Pathways. (2011). *ProDiscover Demo Download*. Retrieved 2 7, 2011, from Technology Pathways: http://www.techpathways.com/Demo.htm

The President's Identity Theft Task Force. (2007). *Combating Identity Theft: A Strategic Plan.* Washington: U.S. Government.

University of Michigan. (2008, 6 20). *Tools for Discovering Credit Card and Social Security.* Retrieved 2 7, 2011, from Information Technology Security Services : http://www.safecomputing.umich.edu/tools/download/ccn-ssn_discovery_tools.pdf

Weinstein, N. (2009, 3 14). *Man Charged with Alleged Child Porn Via PS3*. Retrieved 2 18, 2011, from CNet News: http://news.cnet.com/8301-10797_3-10196553-235.html

Wells, E. C. (2008, 1). *Sustaining Gen Y's Interests*. Retrieved 11 30, 2010, from Today's Garden Center: http://www.todaysgardencenter.com/trends/sustainability/?storyid=340

World Lingo . (2010). *Comparison of file systems* . Retrieved 11 18, 2010, from World Lingo : http://www.worldlingo.com/ma/enwiki/en/Comparison_of_file_systems

9. BIBLIOGRAPHY

DataRescue. (2010). Retrieved 2011, from DrDD- DataRescue's DD freeware: http://www.datarescue.com/photorescue/v3/drdd.htm

Digital Forensics Framework. (2009). Retrieved 2010, from Digital Forensics: http://www.digital-forensic.org/digital-forensics-framework/community/

EnCase. (2011). Retrieved 2011, from Guidance Software: http://www.guidancesoftware.com/

Hex Editor XVI32. (2009). Retrieved 2001, from Freeware Hex Editor XVI32: http://www.chmaas.handshake.de/delphi/freeware/xvi32/xvi32.htm

Modio. (2010). Retrieved 2011, from Game-Tuts:
http://www.game-tuts.com/community/index.php?pageid=modio

Prodiscover. (2010). Retrieved 2011, from Technology Pathways:
http://www.techpathways.com/DesktopDefault.aspx?tabindex=8&tabid=14

wxPirs . (2010). Retrieved 2011, from Xbox-Scene :
http://www.xbox-scene.com/xbox360-tools/wxPirs.php

Xbox-Scene. (2009). Retrieved 2011, from Xbox 360 PC Tools:
http://www.xbox-scene.com/xbox360-tools/xplorer360.php

XFT 2.0 Game Console Toolkit Released. (2009). Retrieved 2011, from Video Game Device Forensics: http://consoleforensics.com/xft-2-0-game-console-toolkit-released/

BACKTRACK IN THE OUTBACK - A PRELIMINARY REPORT ON A CYBER SECURITY EVALUATION OF ORGANISATIONS IN WESTERN AUSTRALIA

Craig Valli, Andrew Woodward and Peter Hannay
secau – Security Research Centre
Edith Cowan University
Perth Western Australia

ABSTRACT

The authors were involved in extensive vulnerability assessment and penetration testing of over 15 large organisations across various industry sectors in the Perth CBD. The actual live testing involved a team of five people for approximately a four week period, and was black box testing. The scanning consisted of running network and web vulnerability tools, and in a few cases, exploiting vulnerability to establish validity of the tools. The tools were run in aggressive mode with no attempt made to deceive or avoid detection by IDS/IPS or firewalls. The aim of the testing was to determine firstly whether these organisations were able to detect such hostile scanning, and secondly to gauge their response. This paper does not extensively analyse the resultant empirical data from the tests this will be the subject of several other papers.

Of the 15 agencies investigated, only two were able to detect the activity, and only one of these escalated this to authorities. Many had intrusion detection or prevention systems, but these did not appear to detect the scanning which was conducted. Others did not have any form of detection, only logging without active monitoring and some had no persistent logging of anything. Of those who did detect, the lack of a formal incident response and escalation plan hampered their ability to respond and escalate appropriately. Many of these organisations had recently, or very recently undergone penetration testing by external audit or IT companies, and yet there were still numerous vulnerabilities, or their system did not detect the scan. The conclusion is that organisations need to be very specific about what their needs are when engaging external agents to conduct network security testing, as current penetration testing is giving them a false sense of security

1. INTRODUCTION

This paper examines issues uncovered as a result of vulnerability assessment of information systems across 15 large organisations across various industry sectors in Perth, Western Australia. This assessment involved a team of 5 staff for a period of 3 months. There was a variety of assessments performed including documentation and policy review as well as live enumeration and penetration testing of systems over an extended period of time. The NIST SP800-115 document defines penetration testing as attempting to break in to a system (Scarfone *et al*, 2008). The reality is that most audit firms who conduct penetration testing are really only performing vulnerability scanning. This is not to be confused with vulnerability assessment, a process which examines overall security posture of an organisation, and examines network configuration, policy, procedure, compliance, governance and change management, which are all often causes of any vulnerability found by scanning. An appropriate analogy would be that vulnerability identifies the symptoms of security issues, whereas vulnerability assessment finds the cause of the disease. As such, treating the symptoms found by a vulnerability scanner is analogous to taking a pain killer to treat a sore throat, whereas a vulnerability assessment would determine that antibiotics are needed to treat the cause of the infection.

This paper will focus on issues surrounding the penetration testing of the systems and issues uncovered in this process.. The organisations were told that their systems were going to be tested and that they should use normal escalation procedures should they detect an attack or compromise of a system. The organisations were not told the nature of the testing nor its duration, magnitude or frequency, they were told simply when testing would start. At the conclusion of the testing period organisations were given an exit interview to give feedback but also to check how well if at all detection of attacks had occurred and what if any action had been taken. This paper outlines some of the macro issues and errors that are still being perpetrated by organisations.

2. THE RULES OF ENGAGEMENT

The main idea or thrust behind the penetration testing was to enumerate and attack information systems used by the organisation for service delivery. This focus encompassed not only conventional email and web systems but also VPNs, video conferencing systems and a variety of bespoke systems that had external IP. The other primary directive was that there was to be no specialised attacks or advanced enumeration techniques used in the conduct of the testing. This meant attacks had to resemble those that could be mounted by novice users who downloaded freely available tools and used online information sources to educate themselves and perpertrate any malfeasance. An example of this was the web testing tool nikto (Sullo & Lodge, 2011) that was used in default modes no IDS/IPS evasion techniques were utilised. Nmap (Fyodor, 2002) similarly was used with the nmapfe frontend and selections of options were taken from these default menus to perform port scanning, service identification and operating system.

The attack intensity also escalated in magnitude as the testing progressed for example initial enumeration was done doing scans that probed every 15 seconds to highly aggressive all ports all service scans that emanated 50-100Mbytes of traffic, across entire B Class address spaces in 5-20 minutes. Likewise, password brute force attempts initially at low connection rates ~ 1 attempt every 5 secs to literally the complete set of dictionaries on the Openwall CD exhausted as quickly as the tool or the connection could carry them. The latter with even basic bandwidth monitoring would have detected.

The attack platforms were that of a home user ADSL account supplemented by cheap cloud based virtual servers for instance no server used cost more than $70 for a years subscription and had a bandwidth limit of 1TB of traffic a month. It should be noted that the servers were on fast high speed links and were capable of delivering sustained attacks of large volume.

There were 3 people conducting probing of the 15 targets from a total pool of 12 real IP addresses. The attacks were consistently from these IPs across the 15 targets, however timing was such that any co-ordination of the attacking IPs would have been coincidental beyond each attackers set of IPs. As to escalation the relevant authorities were aware of the testing period and the attacking IPs.

3. TOOLS USED

The tools used were freely available and well known attack tools they were primarily sourced from BackTrack 4 CDs on local laptops. The virtual servers all used Ubuntu 10.10 default installs as the base system that was supplemented with commonly used binaries for security testing and penetration such as nmap, nikto and others.

For enumeration principally nmap was used for service and system enumerations, this was supplemented by the use of httprint and nikto or other specific tools as needed when enumerating or fingerprinting services. As previously mentioned this tools utilised nothing other than default options available in menus, to reflect the reality of a relative computer novice.

For system wide attack again default tools were used perpertrate attacks against identified services or operating system platforms, these included nessus and metaploit. The approach with attack was an increase in magnitude initial attack profile was attack against an enumerated service for instance running a SQL injector against an identified SQL server again using default or noisy methods. This type of attack would not be specific for instance if the scan reported a SQL server on a Microsoft platform an SQL tool that attacked other platforms was utilised as well meaning that any even poorly configured IDS should have detected a series of attacks.

Having attempted lower magnitude attacks these then were systematically escalated to full noise indiscriminate brute force attacks. An example is metasploit autopwn was used against a host with impunity and basic limitation was bandwidth of connection to carry the attack, no evasion, no tweaks.

The final stage of attack was that of social engineering using USB memory sticks as the vector that was simply dropped or left within the business building perimeter. The USB vector was not designed to autoboot and activate at insertion of the drive. The USB had 3 files on it namely a readme.txt, a modified binary called encryptor.exe and a false file called crypted.vol that contained random characters. For the USB to call home via a DNS request the human actor had to run the encryptor.exe and attempt a password.

4. RESULTS AND DISCUSSION

The extensive data from the testing are still being analysed however the following statements put the extent of the exposures uncovered in perspective for the purposes of the discussion.

- All organisations were readily and easily enumerated with only 2 organisations being aware of the probing.
- All organisations had significant exposures uncovered in the network scanning and testing.
- All except one of the organisations detected the intense scans and attacks of the system.
- All except one of the organisations did any tangible, credible and trackable escalation of incidents.
- Some of the organisations logging and record keeping is that poor that no evidence could be located post testing.
- Only one organisation was not compromised by USB stick attack. Two external IT providers to the organisations were also compromised.
- All USBs were effective within less than 48 hours of being dropped at the organisations.

4.1 Escalation and responses or lack thereof

All organisations showed no or extremely poor escalation of incident to authorities. Only two of the 15 organisations escalated the attacks to authorities for further investigation. This is alarming in that 13 organisations failed to detect and effectively respond to sustained attacks on their systems.

To their credit, two organisations provided some response but again there are concerns in their level of response. One organisation undertook what can be best described as multiple agency contact, basically contacting anyone who would listen. This mass alerting was conducted against the organisations policy which had a person in a designated position would make the call and then only to one agency. This demonstrates poor organisational awareness of policy, which may indicate a lack of training or familiarity with escalation.

Another organisation had succumbed to crying wolf or demonstrated Hawthornian effects in response, such that they were contacting agencies and reporting attacks from the attack team when in fact they had been idle on that organisation for 10 days. Basically, other IPs that were actually attacking from a home based DSL account within the Western Australian ISP IP address spaces were being attributed to the attack teams efforts and escalated to responder agencies.

An intentional ruse was effective in that while attacking with the home based DSL accounts simultaneously high intensity attacks and probes were being perpetrated from the large bandwidth virtual server accounts with no reporting of these apparent by any of the responding organisations.

Feedback from responder agencies has been that escalation and reporting was inadequate and presented no real opportunity for defending systems. In particular one responding agency has resolved to undertake an education and advice program to inform operators how best to report a cyber attack in order to get resolution of the attack.

4.2 Technology tokenism

Some of the organisations had expensive dedicated security appliances that were deployed and inadequately managed, which indicates these organisations are suffering from technology tokenism. It could be that staff were not trained in the use of these appliances, which indicates strongly that organisations should look at the total investment cost which includes ongoing training and support for staff. Also, many of the network borne threats are complex, multi-partite and asymmetric. The modern security appliance is a highly complex system and needs constant adjustment to get optimal performance from it.

4.3 Security is so inconvenient

IT Staff from several organisations reported a lack of acceptance or recognition of risk in IT systems by upper management. One organisation relayed that they had in fact tried to secure USB ports by disabling them through policy management afforded by Windows XP. Soon after deployment was enacted they were summarily told to undo this by top line management as it was not convenient and USB posed no real threat or risk to the organisation. There were also similar vignettes communicated where executives was not aware of or did not want to acknowledge the clear and present danger that not deploying or enabling security measures brought to the organisation.

4.4 Post incident forensics

Another stage of this activity is the investigation of post incident ability to respond to evidentiary requirements and also provide data for analysis of incidents. The analysis of any log files or intrusion data is not yet complete however there have been uncovered significant issues already in this phase.

Several of the organisations have not been able to provide any tangible log file data. There are several reasons, the most alarming is that preservation of log files is not occurring beyond a short time window of a week to a few days dependant upon logging activity i.e the log files are live and simply utilise fifo. There is no daily archiving or storage of log data in these organisations, which under WA state law is a breach of the State Records Act, let alone the fact its basic security practice. Any argument that storage space or performance of appliance is a significant issue is a very tired IT industry meme. Hard disk storage is incredibly cheap and devices are sufficiently powerful that any logging is now in the realms of 1-2 per cent of CPU and if an IPS or device is that marginal bigger issues are afoot.

Several of the organisations do not know how to extract data from their IDS/IPS, firewall systems when this information was requested they have supplied HTML documents taken from their system management consoles. This is clearly an inadequate response.

4.5 Penetration Testing vs. Vulnerability Assessment

Of concern is that nearly all organisations examined in this research had recently paid external companies to conduct penetration tests against their infrastructure. The evidence presented as a result of examining these 15 organisations is that penetration testing seems to have almost zero value, whilst having a very high cost, both in monetary and security terms.

The profile of companies employed to conduct such penetration testing are commonly audit organisations for which their major business is financial audit. However, the growth in the use of the internet for e-commerce and other core business functions has seen these organisations branch out into IT security auditing, or so called ethical hacking. As such, when an organisation requests an external audit of their organisations, an evaluation of the health of general computer controls as they relate to financial system access is also conducted. Increasingly, organisations are also being sold ethical hacking or penetration testing in relation to their internet facing infrastructure. In addition to being part of a financial audit, organisations are also using these same audit firms to conduct ad-hoc assessment of their network infrastructure as part of change management or configuration changes.

There are anecdotal reports from some organisations that the companies conducting these tests commonly ask them to add them to a firewall or IPS white list or to turn off certain security features so that they can conduct the test. Such an approach may allow for testing of an individual component with a companies defence infrastructure, but it certainly does not test or evaluate the security of an organisation as a whole. Standard practice for such organisations is to use recent graduates armed with a tool (commonly Nessus and Nmap) to run scans against the target organisation who requested the test. Whilst the people using the tools may have been adequately trained and instructed, they are far from network security experts, or even ethical hackers, as they sometimes refer to themselves.

In defence of the audit organisations, they are likely only providing the service which they are asked to perform. That is, organisation A asks for, and receives a penetration test of their firewall. Is this a useful test of organisation A's security? No, but it is what they requested. Having said that, there appears to be an ethical issue in relation to charging large amounts of money for a test which is largely worthless, regardless of whether the organisation specifically requested it or not.

Of far greater value to an organisation is a vulnerability assessment which assesses the overall security posture of an organisation, including such aspects as policy, procedure, physical security, change management and governance. For example, a penetration test may find an open port on a firewall that should have been closed. In that instance, the recommendation is to close the port. A vulnerability assessment, through an examination of the firewall rule sets, would also pick up that a port was open. However, the recommendation would then be to look at change management and policy and procedure in relation to network security as to *why* the port was open, and to prevent such an issue occurring in the future.

5. CONCLUSION

This engaged research has resulted in uncovering significant issues that need addressing in organisations with respect to preparedness to attack, response, escalation and investigation of external attack of cyber systems.

All of these organisations have an IT department and in some cases have personnel responsible for security which mitigates the resourcing defence that many organisations put forward, some also outsource their daily IT security to specialist firms. Many have also paid large amounts of money to external audit agencies to conduct penetration tests against their infrastructure. However in defence of the IT staff in these IT departments often security of systems is compromised by poor management decisions as result of poor understanding of IT based risk. This is all too a common theme in investigations of this sort and something both sides of the IT management divide need to work on.

There was largely systemic failure to detect and respond to the attacks. Only 2 out of the 15 organisations provided any semblance of coherent response to the attacks, the other 13 can only be categorised risk wise as extreme. The work has uncovered that there is significant fundamental work that needs to be undertaken in these organisations before any semblance of an IT security posture or awareness could be proclaimed.

6. REFERENCES

Fyodor. (1998). "Remote OS detection via TCP/IP stack fingerprinting." Retrieved 10 May, 2002, from http://www.insecure.org/nmap/nmap-fingerprinting-article.txt.

Scarfone, K., Souppaya, M., Cody, A. & Orebaugh, A. (2008). Technical Guide to Information Security Testing and Assessment. National Institute of Standards and Technology. Gaithersburg, Maryland

Sullo, C. and D. Lodge (2011). Nikto2.

DIGITAL FORENSICS INVESTIGATION IN A COLLEGIATE ENVIRONMENT

Robert E. Johnston, CISSP
92 Carriage House
Enfield, CT 06082-6042
Telephone: 860-776-2055 Cell: 860-539-9206
Fax: 860-741-6418 (by arrangement)
E-mail: bjohnston@e-computer-security.com

Connecticut Community Colleges
System Office
Connecticut Community Colleges
61 Woodland Street
Hartford, Connecticut 06105
Telephone: 860-244-7763 Fax: 860-244-7886
E-mail: rjohnston@commnet.edu

ABSTRACT

Creating, building, managing a cost effective digital forensics lab including a team of qualified examiners can be a challenge for colleges [1] with multiple campuses in multiple towns, counties and states. Leaving such examination responsibilities to each of the campuses results in not only disparity in the results but more than likely excessive duplication of efforts as well as the potential for compromise of evidence. Centralizing the forensic efforts results in a team that is not subject to the political pressures of a campus and virtually eliminates the possibility of examiner favoritism. Learn what it takes to create a cost effective centralized digital forensics lab. It sounds simple but is truly quite complex when you consider the chain-of-custody issue as well as the management support needed during initial implementation. There will be resistance at some of the campuses while others will welcome the removal of a burden. We will also examine why such a lab is necessary and what can be learned about compliance to existing policy as well as the possibility of identifying the need for additional policy/standards.

Keywords: digital forensics investigation malware criminal chain-of-custody centralized lab

1. THE CHALLENGE

Implementing centralized digital forensics investigation within a widespread enterprise can be difficult. There are numerous fiefdoms involved, many of which hold self-serving interests which are contrary to such a project. No matter how much sense it may make it is not uncommon to meet massive resistance. The initial acceptance of the concept will often be the greatest challenge its advocate will ever face!

Knowing your organization including the individuals involved in blocking or supporting such a project is usually necessary in the collegiate environment. After all, commonly each campus is quite independent from central management whose role is primarily that of obtaining funding and setting budgets with some over site relative to more sensitive issues which certainly vary from college to college and private versus the public sector. Some intelligence gathering is often essential to achieve success.

Do not jump in whole hog without knowing the terrain. Plan….plan….plan….

2. JUSTIFYING THE PROJECT

Your greatest challenge is the justification. The balance of effort will be a piece of cake in comparison but quite tedious at times. Care must be taken to ensure clarity and understandability. Often the justification will be read by those who are unfamiliar with information technology and especially the whole concept of digital forensics. To many digital forensics is just another term and very possibly is simply an adaptation of the term forensics which has become so popular today in the field of law enforcement. Many cannot relate that forensics in science and information technology is an analysis technique which ensures that should illegal activity, whatever that might be, be found that the evidence is preserved in a manner consistent with that which is acceptable to law enforcement and the courts.

More than likely very few if any digital forensic investigations regarding malware infestations will uncover criminal activity. However, it is entirely possible and you must be prepared. How embarrassing would it be to the college if a staff or faculty member was detected with sums of child pornography on the system including trading/sale of same but could not be prosecuted due to inadmissible evidence? Take it a step further and envision that identifying that staff member to law enforcement results in a determination that the staff member is also a child molester yet could not be prosecuted due to the principle of “fruit of the poisonous tree” [2]. Surely, this something everyone at every level wishes to avoid.

Preservation of evidence is not the only justification. The remainder relates to traditional management concepts/needs.

2.1 Control

In order to ensure that the evidence (the malware infected device; specifically, the hard drive in the case of an infected computer) is preserved in a manner satisfactory to law enforcement, etcetera it is imperative that effective control is maintained throughout the process commonly referred to as the “chain of custody” [3].

This requires the creation of detailed records of the handling and storage of a physical drive from the time it is taken into possession by the information technology staff through and until the drive is successfully and properly forensically imaged. In an ideal world the physical drive would be preserved until it is established that there is not a criminal concern. However, in reality this is not practical in most environments. The number of duplicate drives at each site likely would be excessive.

While efficient, this centralized process does tie up each drive for several days even with the creation of a forensic image archive. There is the transportation in both directions as well as the time in the forensic lab. Unless generously configured there will be times when a drive sits for two to three days at the forensic lab until it has been successfully archived. As a result the average amount of time a drive to be investigated is out of service is likely five business days. Also, the drive will remain unusable for another day or so until it is wiped and reimaged. Thus, it is impractical to leave the impacted user without a computer. Therefore each campus will need to keep a sum of drives on hand and ready to go when infections occur.

In addition there is the issue of control while in the forensics lab. As will be seen later, there are additional control benefits in the decentralized lab.

2.2 Savings

The question to be answered is whether the work which needs to be done is being accomplished and, if so is it complying with all of the issues relative to evidence preservation? If no, an analysis is required to determine why and identify the savings that can be realized through centralization. If yes, then the issue is a comparison of costs between the current processes versus that of centralization.

More than likely if you are considering a centralized digital forensics lab either the current process is not effective or does not exist but there should be a perceived need. Need takes many forms, constant reinfection being one of them along with compromise of PII [4] or PCI DSS [5] information as well as other information under development which should not be disclosed until ready; in other words competitive information which also exists in the collegiate world, especially within the private sector.

2.2.1 Constant Reinfection

The primary cause of constant reinfection is the failure of a specific campus that does not follow proper procedures when an infection occurs. With a centralized digital forensic lab such failures become readily apparent and corrective actions can be initiated.

2.2.2 PII Compromise

When PII is compromised the rules/regulations/laws vary from jurisdiction in addition to the ethical obligations. Hopefully your college already has a published policy regarding compromised PII. When compromised PII is detected by a centralized digital forensic lab you are assured that the resulting actions meet current requirements. The embarrassment that might occur should non-compliance be discovered and reported by the media could result in incalculable damage.

2.2.3 PCI DSS Compromise

Failure to comply with the Payment Card Industry Data Security Standard [5] likely will result in unfavorable media coverage as well as the real potential for the loss of rights to process payment cards on campus in a convenient manner.

2.2.4 Compromise of Competitive Information

Development of new majors/minors and other strategies including the development of new for fee services are commonly business confidential until they are made public.

2.3 Staffing

The initial staffing size is difficult to calculate but certainly should be far less than when such examinations are conducted at each campus. Staffing size is also dependent upon the working model. Experience indicates that the decentralized lab requires a smaller staff in addition to offering other advantages and efficiencies. In a typical environment, when a thorough examination is conducted of each case to include production of a written report that can be read and understood at the campus by non-information technology professionals, the average time per case is two hours.

2.4 Influence, Bias, etcetera

In a centralized model the examiners are sheltered from all forms of overt influence and bias as well as friendships. It is quite common in a collegiate environment to “protect their own”, especially when the unknown or misunderstood is encountered. All too often senior staff and faculty become concerned that the case may impact their career, especially when they are unaware of peers encountering similar problems. Dealing with the matter centrally and properly managing the entire process can and should eliminate this concern.

3. CREATING, BUILDING AND MANAGING A COST EFFECTIVE DIGITAL FORENSICS LAB

This is not a seat-of-the-pants project. Careful planning will result in successful implementation with little or no disruption to existing operations. Key to this process is choosing the appropriate model and while there are perceived advantages to both, the decentralized model offers greater flexibility and opportunity.

3.1 Models

3.1.1 Traditional Lab

The traditional lab is totally centralized and frequently is completely isolated from all other information technology activities. This represents a great deal of cost which can be minimized in the decentralized model. Typically, in the traditional lab the forensic examiners are solely responsible for all activities from creating the archives to mounting the drives to be examined on their dedicated forensic examination work station. Commonly, the examiner works from the console of the work station.

There are variations, many of which will be described in the decentralized lab. However, most of the initial costs of the traditional lab cannot be avoided.

3.1.2 Decentralized Lab

The concept of a decentralized lab is foreign to many yet much of its structure is similar to the traditional lab and many of its features can be implemented in the traditional lab.

Basic Concept: Compartmentalize the many responsibilities of the digital forensics lab thus ensuring a higher level of confidence and trust in its integrity while allowing some of the activities to be performed “remotely”.

Forensic computers: Locate in a truly secure data center, preferably not located on the campus of any of the colleges. Day to day support of the forensic computers is performed by operations staff to the extent necessary to mount and dismount cases being examined.

Examination/Archive copies: Examination and archive copy functions are commonly created by the same operations staff which supports the forensic computers. Thus, once a drive to be examined arrives on site only operations staff trusted to support forensics ever handles the original physical drive as well as all copies.

Forensic Examiners: Examiners access their assigned forensic computer remotely even when they are physically based on site. Thus, examiners can be located anywhere they are able to connect securely into the forensic network. Thus, should there be qualified forensic examiners on one or more of the campuses they can be reassigned to the new forensic team. In addition, in today’s world of digital mobility a valued team member can be retained should it be necessary for that team member to not live in the region.

Forensic Network: The forensic network must be carefully architected to be isolated from the balance of the college network and access to that rigidly managed and monitored as well as restricted to forensic staff only! Logical maintenance of the forensic computers is the responsibility of the assigned examiner. Physical maintenance is the responsibility of forensic trusted operations staff.

3.2 Building the Lab

The cost of building such a lab can often be minimized if the college’s network architecture already has a centralized data center providing common services to all of the campuses. For those without this option must consider whether to co-locate on an existing campus or completely off-site. Costs can be minimized with co-location providing that the forensic staff work environment is isolated from the general campus environment. Failure to do so compromises many of the benefits of a centralized digital forensics lab.

3.2.1 Hardware

Hardware must be robust but not necessarily state-of-the-art. Forensic tools have not been that quick to jump to the latest hardware architecture and likely will not abandon support for earlier platforms which support XP. There may be some concerns regarding XP relative to Internet access yet since such actual access should only be performed in a virtual mode that is not likely to be a near term issue.

Clearly, XP platforms being replaced with Win7 etcetera can be utilized in the lab. Components of the platform will require replacement for best performance as well as maximizing memory and external ports. Also, some hardware write blocks, at least one per forensic computer will be needed. As hard drives keep growing in size it may be appropriate to examine the case drive directly and based on the result determine whether an archive image copy is needed.

3.2.2 Software

There are numerous software tools available. While there are other examination tools, serious consideration should be given to choosing Encase [6]. It does require some training/experience to be effective with Encase, but in the end it is the tool which is trusted in law enforcement circles should they become involved. Beyond Encase, a trusted VM tool is needed as well as several other tools which should be considered:

- Automated registry decoder; e.g., Registry Ripper [7]
- View the Registry in native mode; e.g., Registry Viewer [8]
- Tool to locate and identify PII/PCI DSS data; e.g., Identity Finder [9]
- Tool to evaluate links; e.g., Link Examiner [10]
- Linux-like environment for Windows making it possible to port software running on POSIX systems (such as Linux, BSD, and Unix systems) to Windows; e.g., Cygwin [11]
- Possibly a network meeting tool; e.g., TeamViewer [12]
- VM tool; there are many to choose from.
- Sandbox Tool; e.g., Sandboxie [13]
- Key Recovery; e.g., Recover Keys [14]

3.2.3 Staffing

All members of the forensic team must be chosen for their skills, experience and trustworthiness. Fortunately it is very likely that you will be able to identify within your current professional staff. If not, perhaps within faculty. In today's job market it is possible you can locate key staff locally at reasonable cost. Choose your staff carefully as their duties require not only competency and loyalty but also trustworthiness.

4. LEARNING OPPORTUNITIES/ADVANTAGES

Unlike having forensic examiners at each campus, operating a centralized digital forensics lab will provide benefits difficult to achieve without one.

4.1 Image Maintenance

Hopefully there are image [15] standards in place. Due to the challenge of distribution and installation from a centralized facility most colleges provide imaging standards on each campus. Ideally, there are standards set centrally which describe image content, frequency of refresh, etcetera.

Forensic examinations can readily identify where those standards are not being maintained and thus corrective action can be initiated.

4.2 Consistent Practices

One of the challenges of managing multiple locations is that of consistency. In addition, it is not that uncommon to come across practices at one location which are an improvement over that which is practiced at other sites. Whether the result is that of bringing all sites into alignment or learning what is better than a current practice, it is nothing but distinct value.

Another concern is that of inconsistency; for example, a situation develops that results in management/clients/media noting a problem that could have been avoided had it followed a practice at site D, and why aren't sites B, C and E also following site D's model?

4.3 Building Trust/Confidence with the Campuses IT [16] Security Staff

While it is not uncommon for the IT Security teams at each campus to be initially wary when a centralized digital forensic facility is created, when done carefully it will result in a trust relationship which otherwise might not have been built. Over time it is more than likely that a query will be received from an IT Security staff member or manager regarding a specific incident. When clear concise explanations are offered while avoiding the implication of blame, trust develops; especially when it is possible to point out how such situations can be avoided in the future.

5. SUMMARY

Selling the concept of a centralized digital forensics facility/lab in the collegiate environment can be challenging. The basic premise of campus independence/autonomy will always be an issue. However, the fact of the matter is that there is much to be gained and learned through centralizing digital forensics as well as a potential significant cost savings.

No two colleges or campuses are identical. For campuses, location in terms of distance from the college is a large consideration/influence. None-the-less, a serious examination of the potential benefits of a centralized digital forensics lab should be performed.

6. AUTHOR'S BIOGRAPHY

Robert E. Johnston, CISSP, is an experienced information security professional, Bob has performed security services from coast-to-coast, and overseas, regarding all aspects of information security, including contingency planning, for large and small businesses. Having been the senior information security officer for major financial institutions he brings the vision and experience of a senior corporation executive and the broad knowledge developed while servicing his consulting clients. Bob maintains technical competence in critical areas including Networks/Internet/Intranet components (e.g., HTML, JAVA, Active-X, TCP/IP, Firewalls, E-mail), Information Security (e.g., PGP, RACF, CA-Top Secret, CA-ACF2, CICS, Cryptography), LAN/WAN (e.g., Windows 95, 98, NT, W2K, XP, Vista, Win7) as well as all hardware platforms, security concepts, standards and policies. He has served as an expert witness in Federal/State criminal/civil cases and GAO Administrative hearings and, conducted several successful computer forensic investigations within the financial services sector as well as within the undergraduate collegiate sector. A recognized expert, he has made more than 100 presentations worldwide and has written more than 100 articles published in numerous periodicals and journals. As a highly skilled information security professional, the International Information Systems Security Certification Consortium (ISC)^2 awarded him the designation of Certified Information Systems Security Professional (CISSP) in 1995.

7. REFERENCES

[1] college – For the purposes of this paper a college is defined as the parent college of a group of colleges at varying locations. Often the colleges not on the primary campus are run quite independently; almost as if they are not truly affiliated with the parent college.

[2] fruit of the poisonous tree – https://secure.wikimedia.org/wikipedia/en/wiki/Fruit_of_the_poisonous_tree

[3] chain of custody – https://secure.wikimedia.org/wikipedia/en/wiki/Chain_of_custody

[4] PII – https://secure.wikimedia.org/wikipedia/en/wiki/Personally_identifiable_information

[5] PCI DSS a.k.a. Payment Card Industry Data Security Standard – https://secure.wikimedia.org/wikipedia/en/wiki/PCI_DSS

[6] Encase – https://secure.wikimedia.org/wikipedia/en/wiki/EnCase and http://www.guidancesoftware.com/

[7] Registry Ripper a.k.a. RegRipper – http://regripper.net/?page_id=120

[8] Registry Viewer – http://www.softpedia.com/get/Tweak/Registry-Tweak/Registry-Viewer.shtml, http://accessdata.com/media/en_us/print/techdocs/Registry%20Viewer.pdf and http://accessdata.com/downloads/current_releases/rv/AccessData%20Registry%20Viewer.exe

[9] Identity Finder – http://www.identityfinder.com/

[10] Link Examiner – http://www.simplecarver.com/free/ and http://www.analogx.com/contents/download/network/lnkexam/Freeware.htm

[11] Cygwin – http://www.cygwin.com/

[12] Team Viewer – http://www.teamviewer.com/en/index.aspx

[13] Sandbox Tool – https://secure.wikimedia.org/wikipedia/en/wiki/Sandbox_%28computer_security%29 and http://www.sandboxie.com/

[14] Recover Keys – http://recover-keys.com/

[15] image – operating system image; http://publib.boulder.ibm.com/infocenter/tivihelp/v13r1/index.jsp?topic=/com.ibm.tivoli.tpm.img.doc/bootsrv/csfi_images.html

[16] IT – information technology; https://secure.wikimedia.org/wikipedia/en/wiki/Information_technology

APPENDIX – SAMPLE FORENSICS EXAMINATION OPERATION

AUTHOR

Robert E. Johnston, CISSP, November 1, 2010, eMail: bjohnston@e-computer-security.com

OVERVIEW

This paper was prepared for a professional discussion group that wanted a basic explanation of a forensics lab. Since the group consisted of virtually all private sector business security professionals you will find that it avoids reference to the collegiate environment and I tried my best to make it usable in the private sector. Common abbreviations are not explained and abbreviations created for convenience in the paper are “explained” the first time they occur. In addition, you will find for your convenience a complete list of abbreviations at the end of this document.

INTRODUCTION

Forensics Labs can take many forms. The reason for preparing this model is that it was requested by someone who wanted “model procedures” to which I responded that there is not truly a model that is uniform to all situations. I believe that the following dissertation will make that abundantly clear yet possibly assist him in his assignment/endeavor.

This is based on an existing “successful” lab supporting an enterprise consisting of 12 remote locations and a central office, all within a single state. Some of the practices contained herein clearly will not work due to physical distances elsewhere. Understand that the distance from the central office to any remote site does not exceed 60 miles with the majority within 30 miles. On the other hand, why does the lab exist?

After all, there are commercial labs committed to the recovery of information; criminal labs intended to identify illegal activity as well as many others including the enterprise which focuses upon network compromise including PII, PCI and HIPAA issues. The lab to be illustrated is concerned with network compromise, PII and to a limited extent PCI matters. At the same time such labs cannot ignore the possibility of the discovery of illegal activity whether fraud, extortion, child pornography or other criminal activity. While an enterprise might consider such possibility to be infinitesimally small, the possibility should not be ignored!

Once one starts examining a hard drive, it is amazing what might be discovered. It truly ranges from criminal activity to massive waste of resources and time as well as proper usage of enterprise resources. While, for the most part the discovery of such activity not in the best interest of the enterprise must be concluded on an individual basis, that option does not exist for some activity that must be reported to law enforcement as the result of legislation. Thus, it is incumbent upon every forensic activity to ensure that the “chain of custody” is maintained lest damage to the image of the enterprise and/or violation of law occur when such activity is revealed but cannot be prosecuted and possibly the offender cannot be reprimanded under corporate guidelines.

CHAIN OF CUSTODY

When an event occurs at a Remote Office (RO) a notice is sent to the Central Office (CO) advising of the issue and requesting advice as to the necessity of a forensic examination. A prompt reply is provided confirming the need or offering technical advice when one is not needed.

Systems to be forensically examined have their drive(s) removed by authorized IT personnel at the RO and a record kept of that individual as well as the reason for submission and details regarding the drive’s identity on the RO Control Sheet (CS). The drive(s) is/are transported to the CO by one of several authorized individuals and their identity is recorded along with the date and time on their CS. The CO CS will record all drive handling from imaging through to return. The forensic examiners (FEs) never touch the original drive, imaged drive or the archived version.

THE FORENSIC LAB

Forensic labs are designed in many forms while, hopefully, meeting the objectives of management and excellent business practices. All sorts of issues must be taken into consideration including available resources (space, staff, objectives and etcetera). Many labs are the actual work space of the examiners. Others adopt a more flexible environment by placing the lab in the data center environment where the operations staff supports the forensic computers and the examiners connect to them remotely, never having physical contact with the hardware.

Having operations perform all of the drive handling issues ensures knowing where the responsibility lies as well as having the FEs totally focused upon their responsibilities of case examination and reporting.

There clearly are advantages to both, but when all is said and done, procedurally many find the latter arrangement to be the most advantageous. Once the drive is imaged and archived the image drive is mounted on a forensic computer and the responsible FE notified; all of which is documented on the CO CS.

THE EXAMINATION

FE activity commences with the creation of a virtual drive for a malware scan for all cases. However, before the malware scan may be started specific “history files” must be created so that the result of the malware scan and further activity can be properly documented. Using a naming standard created for the forensic examinations all of the preliminary work of creating the temporary storage directory (TSD) and propagating much of the content including copies of all quarantined items, an extract of each of the major components of the registry and a boiler plate copy of the FE’s report (FER) to be populated as the FE continues through to completion is created by a custom program created for this activity. The FE then initiates the rescan directing the result be stored in a sub-directory of the TSD.

Then, a copy of the RO notice is created in the TSD and a summary of its content entered into the FER. Dependent upon the reason for examination the FE proceeds to review the many resources captured in the TSD while the malware scan continues to completion. Once completed any malware infestations detected will be documented in the FER. When necessary, commonly for every examination, the drive will be opened with a forensic examination tool and the details of the content of the drive will be examined for the specifics needed to document and close the case.

When completed the FE will submit the FER to management for final disposition. From the FE’s perspective the case is closed and all documentation is noted as closed including entries in the CS of the RO and CO. Drive final disposition is also documented in the CS and the content of the TSD is transferred to the permanent history file.

Other summary reports are created on a monthly basis from the FERs for use by management in understanding just what is being examined and understanding the issues which might warrant further action to preclude repetition.

ABBREVIATIONS

CO – Central Office

CS – Control Sheet

FE – forensic examiners

FER – forensic examiner’s report

RO – Remote Office

TSD – temporary storage directory

Subscription Information

The Proceedings of the Conference on Digital Forensics, Security and Law is a publication of the Association of Digital Forensics, Security and Law (ADFSL). The proceedings are published on a non-profit basis.

The proceedings are published in both print and electronic form under the following ISSN's:

ISSN: 1931-7379 (print)

ISSN: 1931-7387 (online)

Subscription rates for the proceedings are as follows:

Institutional	-	Print & Online:	$120	(1 issue)
Institutional	-	Online:	$95	(1 issue)
Individual	-	Print:	$25	(1 issue)
Individual	-	Online:	$25	(1 issue)

Subscription requests may be made to the ADFSL.

The offices of the Association of Digital Forensics, Security and Law (ADFSL) are at the following address:

Association of Digital Forensics, Security and Law
1642 Horsepen Hills Road
Maidens, Virginia 23102
Tel: 804-402-9239
Fax: 804-680-3038
E-mail: office@adfsl.org
Website: http://www.adfsl.org

Contents