Conference on
Digital Forensics, Security and Law
Oklahoma City, Oklahoma
April 23-25, 2008

Conference Organizer
Glenn S. Dardick
Longwood University
Virginia, USA

Chair
David P. Biros
Oklahoma State University
Oklahoma, USA

Copyright © 2008 ADFSL, the Association of Digital Forensics, Security and Law. Permission to make digital or printed copies of all or any part of this journal is granted without fee for personal or classroom use only and provided that such copies are not made or distributed for profit or commercial use. All copies must be accompanied by this copyright notice and a full citation. Permission from the ADFSL is required to make digital or printed copies of all or any part of this journal for-profit or commercial use. Permission requests should be sent to Dr. Glenn S. Dardick, Association of Digital Forensics, Security and Law, 1642 Horsepen Hills Road, Maidens, Virginia 23102 or emailed to office@adfsl.org.

ISSN 1931-7379
Contents

Schedule ........................................................................................................................................4
Digital Forensic Certification versus Forensic Science Certification.................................7
Nena Lim
Network Forensic Investigation of Internal Misuse/Crime in Saudi Arabia:
A Hacking Case ..........................................................................................................................15
Abdulrazaq Al-Murjan and Konstantinos Xynos
The Virtual Digital Forensics Lab: Expanding Law Enforcement Capabilities ..................33
Sean A. Ensz and Mark R. McCoy
How Virtualized Environments Affect Computer Forensics ...........................................39
Diane Barrett
SiMPLE - Rethinking the Monolithic Approach to Digital Forensic Software .................51
Craig Valli
Data Mining Techniques in Fraud Detection ........................................................................57
Rekha Bhowmik
The Cyber-Workplace – Identifying Liability Issues in the Information Age
and Managing E-Risk ..............................................................................................................73
Nigel Wilson
Designing Laboratories for Small Scale Digital Device Forensics ..................................81
Richard P. Mislan and Tim Wedge
Schedule

WEDNESDAY, April 23, 2008

09:00am Registration (all day)

AM Workshop

- 10:00am - Computer Forensics Examiners as Private Investigators: The Role of Academia in the Debate
  Gary C. Kessler, Champlain College, USA

12:00noon Lunch

PM Workshop

- 01:30pm - Digital Forensics: The Role of the "Real First Responders"
  David P. Biros, Oklahoma State University, USA

THURSDAY, April 24, 2008

09:00am Registration (all day)

10:00am Start of Conference – Welcoming Remarks

AM Sessions

- 10:30am - Digital Forensic Certification Versus Forensic Science Certification
  Nena Lim, Örebro University, Sweden

- 11:10am - Multi-Agency Cybercrime Data Repository
  David P. Biros, Oklahoma State University, USA

12:00noon Lunch

PM Sessions

- 01:30pm - Small System Digital Device Labs
  Richard Mislan, Purdue University, USA

- 02:10pm - The Virtual Digital Forensics Lab - Expanding Law Enforcement Capabilities
  Mark McCoy, University of Central Oklahoma, USA
  03:00pm - Break

- 03:15pm - How Virtualized Environments Affect Computer Forensics
  Diane Barrett, University of Advancing Technology, USA

- 03:55pm - Electronic Discovery Update: After Zubulake and the New Federal Rules
  Milt Luoma, Metropolitan State University, USA
FRIDAY, April 25, 2008

08:00am Breakfast

AM Sessions

• 08:30am - SiMPLEx - Rethinking the Monolithic Approach to Digital Forensic Software
  Craig Valli, Edith Cowan University, Australia

• 09:05am - Data Mining Techniques for Fraud Detection
  Rekha Bhowmik, Sam Houston State University, USA

• 09:40am - Using N-gram Analysis for Forensic Author Identification and Text Relatedness
  Carole Chaski, ALIAS Technology LLC and Institute for Linguistic Evidence, Inc, USA

  10:15pm - Break

• 10:45am - The Cyber-Workplace – Liability Issues and Risk Management
  Nigel Wilson, Bar Chambers, South Australia, Australia

• 11:25am - The Evolving Role of the Digital Forensics Expert: Protective Orders
  Glenn S. Dardick, Longwood University, USA

12:00noon End of Conference – Closing Remarks
Digital Forensic Certification versus Forensic Science Certification

Nena Lim
Örebro University
Sweden
nena.lim@oru.se

ABSTRACT
Companies often rely on certifications to select appropriate individuals in disciplines such as accounting and engineering. The general public also tends to have confidence in a professional who has some kinds of certification because certification implies a standard of excellence and that the individual has expert knowledge in a specific discipline. An interesting question to the digital forensic community is: How is a digital forensic certification compared to a forensic science certification? The objective of this paper is to compare the requirements of a digital forensic certification to those of a forensic science certification. Results of the comparison shed light on the maturity level of the digital forensic discipline and reveal what can be improved to enhance the confidence and trust of the general public on the digital forensic profession.

Keywords: certification, recertification, digital forensics, computer forensics, forensic science

1. INTRODUCTION
The general public has confidence in professionals with certification because certification is a proof that individuals meet a minimum standard and are capable of doing their jobs properly. Such qualification is particularly important to digital forensic professionals because they often need to present as expert witnesses in courts where both their work, such as the methodology and tools used, and their qualifications are under close scrutiny (Nelson et al., 2005). It is noteworthy that collecting and analyzing digital evidence appropriately represent only part of a forensic investigation process. The expertise and qualification of the digital forensic investigators often could have a significant impact on the reliability of the findings in the eyes of judges and juries.

Despite the importance of certification to digital forensic professionals, no prior study has examined this issue. The objective of this paper is to compare the requirements of one of the digital forensic certifications to those of a closely related discipline -- forensic science. Results of the comparison shed light on the maturity level of the digital forensic discipline and reveal what can be improved to enhance the confidence and trust of the general public on the digital forensic profession.

2. DIGITAL FORENSIC CERTIFICATION
Digital forensic is the “process of identifying, preserving, analyzing, and presenting digital evidence in a manner that is legally acceptable” (McKemmish, 1999). It emerged as a discipline in 1980s (Mohay et al., 2003). Unlike well-established disciplines such as accounting and engineering, digital forensics discipline has many certifications available. The variety of certification could be quite confusing even to the digital forensic professionals themselves. Some of the certifications include Certified Information Forensics Investigator (CIFI), Certified Computer Examiner (CFE), GIAC Certified Forensics Analyst (GCFA), just to name a few. With so many digital forensic certifications available, choosing a representative one is not an easy task. In this paper, we decided to use the number of certificate holders as an objective indicator of the representativeness of a certification. Unfortunately, most computer forensic certification granting bodies do not indicate their numbers of certificate holders on their web-sites. Based on the limited available information, we chose GIAC Certified Forensics Analyst (GCFA) because it has more than 1100 certificate holders and GIAC has certified
over 20,000 professionals.

GCFA is one of the certifications granted by Global Information Assurance Certification (GIAC). GIAC is a professional body established in 1999. It offers a suite of more than 20 certifications which cover expertise in computer security and digital forensics (Frisk, 2008). All GIAC certifications are structured across four levels (level 3 to level 6). As the coverage of each GIAC certification is rather specific, information systems or security professionals will have to obtain different certifications from GIAC to show they have knowledge in different areas. GCFA is a level 5 certification and it provides assurance that a certified individual has the knowledge and skills necessary to undertake forensic analysis and incident investigation.

3. FORENSIC SCIENCE CERTIFICATION

Similar to digital forensic, forensic science is the application of science to identify, preserve, analyze, and present evidence in a legally acceptable manner. The main difference between the two disciplines is that forensic science emphasizes physical evidence instead of digital evidence. The forensic science certification is chosen for comparison purposes because forensic science is closely related to digital forensics. It will be interesting to see how closely the certification requirements of these two relatively new disciplines match up to each other. Among the various forensic science certifications, the certification provided by the Board of Forensic Document Examiners (BFDE) is included in the comparison in this study because it is one of the six accredited boards under the Forensic Specialties Accreditation Board, Inc. (FSCB). (None of the FSCB accredited boards is related to digital forensic.) As the FSCB is sponsored by the American Academy of Forensic Sciences (AAFS), the National Forensic Science Technology Center (NFSTC), and the National Institute of Justice, certifications granted by the accredited boards under FSCB are likely to gain more confidence from the general public. We believe the BFDE was established around 2005 but the exact date is unclear.

4. CERTIFICATION COMPARISON

Table 1 summarizes the information of the GCFA and BC-BFDE certifications. According to their websites, in early 2008 the number of certificate holders of GCFA and BC-BFDE are 1134 and 15 respectively. Apart from the relatively longer history of GIAC, we believe the big discrepancy in number of certificate holders is also due to greater demand of digital forensic professionals and more stringent certification requirements of the BFDE.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Digital Forensic</th>
<th>Forensic Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certification Granting Association</td>
<td>Global Information Assurance Certification (GIAC)</td>
<td>Board of Forensic Document Examiners (BFDE)</td>
</tr>
<tr>
<td>Certification title</td>
<td>GIAC Certified Forensics Analyst (GCFA)</td>
<td>Board Certified by the Board of Forensic Document Examiner (BC-BFDE)</td>
</tr>
<tr>
<td>Number of Certificate holders</td>
<td>1134</td>
<td>15</td>
</tr>
<tr>
<td>Levels of certification</td>
<td>Two (Silver and Gold)</td>
<td>One</td>
</tr>
<tr>
<td>Application fee</td>
<td>None</td>
<td>$100</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>Certification fee</td>
<td>Silver: $899 ($499 with SANS training, $325 for recertification) Gold: $299</td>
<td>$500</td>
</tr>
<tr>
<td>Annual fee</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
| Certification requirement | Silver level:  
  - Pass an online examination  
  
  Gold level:  
  - Silver certification  
  - Complete a written technical report |  
  - Baccalaureate degree (can be substituted by professional experience)  
  - Basic training in forensic document examination including attendance as a spectator at a trial where a forensic document examiner presented expert witness testimony  
  - Good moral character  
  - Sign an agreement to abide by Code of Ethics and Code of Professional Responsibility  
  - Two letters of recommendation  
  - Satisfactorily pass a background check  
  - Pass all examinations  
  - Currently employed in the profession |
| Examination structure | Silver level:  
  - A proctored open book online examination consisting of 150 multiple choice questions  
  - 4 hours is allocated for the examination  
  - Passing score is 70%  
  - Must pass within 4 months upon account activation  
  
  Gold level:  
  - A satisfactory 20-page technical report  
  - Must complete within 6 months |  
  - A proctored written exam consists of around 250 multiple choice questions covering 9 sections of knowledge  
  - A proctored performance examination  
  - A total of 8 hours allocated for all examination  
  - Study guide provided  
  - No fixed passing score but applicants must pass all 9 sections |
| Valid period      | 4 years       | 5 years    |
4.1 Certification Requirements

GCFA offers two levels of certification (Silver and Gold) whereas BFDE offers only one level of certification. To obtain a GCFA Silver certification, applicants do not need to have any prior experience in digital forensic. All they have to do is pass an online examination. Certificate applicants have four months to complete the examination requirement upon their application submission for the certification. Certificate holders of GCFA Silver certification can advance to the Gold level if they complete a technical report under an advisor within 6 months.

The requirements set for a BC-BFDE certification are more comprehensive than those of GCFA. Apart from passing all examinations, certificate applicants generally need to have a baccalaureate degree although such educational requirement sometimes can be substituted by professional experience. BC-BFDE applicants need to show that they had basic training in forensic document examination prior to their application. This includes an attendance as a spectator at a trial where a forensic document examiner presented expert witness testimony. Moreover, applicants are expected to have good moral character and are required to sign an agreement to abide themselves by the Code of Ethics and Code of Professional Responsibility. All applicants need to pass a background check. They should also currently be employed in the profession and submitted two letters of recommendation.

4.2 Examination Structure

Candidates of GIAC certification generally are expected to take training courses provided by the SANS (SysAdmin, Audit, Network, Security) Institute before they take certification examinations. For the GCFA certification, candidates are expected to take a course titled “System Forensics, Investigation & Response.” Certificate applicants who obtain the GCFA certification without attending the SANS training are said to have passed the challenge certification. The certification fee for challenge applicants is $899 but it is reduced to $499 for applicants who have completed the SANS training. Topics covered in the SANS course include the following:

- Forensic definitions
- Incident response and volatile evidence gathering
- Core forensic methodology
- File system essentials and forensics
Network forensics
Timeline analysis
Forensic toolkits
Media analysis using the Sleuthkit
Hash comparisons
Autopsy forensic browser
Windows forensic
NTFS/FAT examination
Application foot printing and analysis
Legal consideration

The GCFA Silver level examination is a four-hour proctored online examination. The examination is open book and consists of 150 multiple choice questions. The passing score is 70 percent. The prerequisite for obtaining a GCFA gold certification is a GCFA silver certification. GCFA Gold level applicants are required to complete a 20-page technical report covering in digital forensic under the supervision of an adviser. Technical reports are evaluated using four criteria: technical accuracy, clear explanation of advanced concepts, extension of ideas beyond courseware, and organization of report.

The BFDE examination includes a proctored written examination and a proctored performance examination. The written examination lasts for four hours and includes around 250 multiple choice questions from the following nine sections. Certificate applicants need to pass all sections to obtain an overall pass but there is not a fixed passing score.

Foundation skills
Gathering evidence
Analyzing handwriting
Analyzing falsified documents
Analyzing features of paper and media
Analyzing impact and non-impact images
Use of laboratory instruments
Evaluating evidence and presenting case findings
Demonstrating knowledge of legal procedures

In addition to the written examination, certificate applicants of BFDE also need to undertake a practical examination which requires candidates to conduct examination in a simulated case. During the practical examination, applicants are required to examine case documents, render an opinion, and provide an argument to support the opinion. The practical examination also lasts for four hours. Unlike GCFA, no course is provided specifically for the BFDE applicants. However, all BFDE applicants receive a study guide.

4.3 Recertification Requirement

To maintain the confidence of the general public and employers in their work, certified professionals in many disciplines these days are expected to keep abreast of the latest development and keep their skills and knowledge current. As a result, certification granting bodies often require its certificate holders to undergo some kinds of recertification process.
The validity period of the two certifications is similar. A GCFA certification is valid for four years whereas a BC-BFDE certification is valid for five years. Nevertheless, GCFA and BFDE adopt different approaches toward recertification as shown in Table 1. A certificate holder of GCFA needs to retake the complete certification examination every four years. Similar to its certification requirements, the BFCD has a comprehensive list of requirements for recertification. Certificate holders of BC-BFDE seek for recertification need to be currently working in the document examination profession. Moreover, they are required to take 60 credit hours of continuing forensic education and undertake proficiency test twice within five years. (It is, however, unclear how the proficiency examination is related to the normal certification examinations.) Recertification applicants need to show they have contributed to the profession through publication, presentation, or participation in research. More importantly, the BFDE has additional recertification requirements which involve monitoring the behavior of its certificate holders. First, BFDE certificate holders need to submit a transcript of two complete case files for a peer review audit. Second, if they render an opposite opinion in trial against another forensic expert in more than five occasions within a five-year period, the Board will appoint a review committee to review the certificate holders’ work.

5. CONCLUSION

The objective of the above certification comparison is to provide a glimpse of the current certification situation in digital forensic as well as a closely related discipline. For both digital forensic professionals and the other stakeholders in digital forensic such as employers, the above comparison results provide them with some ideas how the digital forensic discipline compared against other disciplines. Of course, with the variety of digital forensic certifications currently available, the comparison results are not meant to be definite. For example, while GCFA applicants do not need to have any relevant digital forensic experience, some digital forensic certifications such as Cyber Security Forensic Analyst (CSFA) and EnCase Certified Examiner (EnCE) have such requirements (Lim, 2008).

Overall, the comparison results show that certification requirements for a forensic science certification such as BC-BFDE are more stringent than those of a digital forensic certification both in terms of practical experience and practical examination. Moreover, the numerous specific digital forensic certifications could be quite confusing for other stakeholders of the discipline. For example, while GCFA covers incident response, GIAC offers a specific certification in that area titled GIAC Certified Incident Handler (GCIH). In fact, such an overlap can be confusing even for the security and digital forensics professional who do not hold those certifications. We believe that it might not be a bad idea for the digital forensic community to consider following an overall certification approach. That is, it might want to consider offering an overall certification but allowing its certificate holders to specialize in a specific area. In this way, the digital forensic certification might have a clearer identity. The general public might have more confidence in the discipline if more rigorous requirements on practical experience are set.

6. REFERENCES


Network Forensic Investigation of Internal Misuse/Crime in Saudi Arabia: A Hacking Case

Abdulrazaq Al-Murjan
Information Security Research Group
Faculty of Advanced Technology
University of Glamorgan
Pontpridd, Wales, UK
Aalmurja@glam.ac.uk

Konstantinos Xynos
Information Security Research Group
Faculty of Advanced Technology
University of Glamorgan
Pontpridd, Wales, UK
kxynos@glam.ac.uk

ABSTRACT

There are ad-hoc guidelines and a limited policy on computer incident response that does not include computer forensic preparation procedures (e.g. logging incidents). In addition, these guidelines do not consider the requirement of Islamic law for admissible evidence at an organisational level in Saudi Arabia. Network forensic investigation might breach the Saudi law if they follow ad-hoc or international digital forensic standards such as Association of Chief Police Officers (ACPO) guidelines. This might put the organisation in a costly situation when a malicious employee sues an Islamic court. This is because the law of Saudi Arabia is complying with Islamic (Al Sharia) law. Network forensic investigators should comprehend Islamic legal requirements for admissible evidence such as privacy of a suspect, integrity and availability of evidence. These legal requirements should be translated into information technology to conduct the processes of digital forensic. These processes include searching for, collecting, preserving and presenting electronic evidence in an Islamic court. Although insider abuse/crime have not been usually reported to the law enforcement in Saudi Arabia, a hacking case is provided and examined in order to highlight shortcomings for producing e-evidence at an organisational level in Saudi Arabia. Furthermore, this case shows that there is a conflict between the technical (ad-hoc) process of collecting e-evidence which has been followed at an organisational level by network forensic investigators and the main principle of forensic procedure in Saudi Arabia. It also illustrates that there is no technical investigative standard for digital evidence. Moreover, this research addresses these issues by proposing a technical investigative standard for digital evidence. As a result of this standard, network forensic investigation is able to produce e-evidence with respect to the principles of forensic procedure in Saudi Arabia.

Keywords: Internal threats, malicious insider, network forensic investigation, hacking, formal controls for digital forensics, technical controls for digital forensics, informal controls for digital forensics, forensic procedure in Saudi Arabia

1. INTRODUCTION

Internal threats to information technology may have a significant impact on organisations (Dillon 1999; Melara and Sarriegui, 2003; Magklaras and Furnell 2004) because they usually lead to disclosure of information, modification, denial of service (DoS), illegal use, identity theft or repudiation (Gollmann 2006). These threats refer to crime/abuse when the fundamentals of an
organisation's security policy, confidentiality, integrity and availability are breached and violated by insiders. According to the British Department of Trade and Industry (DTI) in association with PriceWaterhouseCoopers who published the 'Information Security Breaches Survey 2006', malicious security incidents in organisations stemming from insiders are almost double that of those originating from unauthorised outsiders (PriceWaterHouseCoopers 2006).

Numerous attempts have been made to correctly define what a malicious insider is. According to Schneier (2000), 'a malicious insider' is defined as an employee who is an expert involved in the design of the system he is now committing an attack against. However, there are a number of internal threats that only need non-technical skills to commit a crime such as stealing the password of another user. Randazzo et al. (2005) found that 87% of internal incidents in the banking sector used simple user commands and were non-technical. Therefore, according to Schneier's definition, an employee who steals a password by looking over a colleague’s shoulder at a system that she/he is unauthorised to use, and then uses this to create a threat, is not identified as an insider (Rowlingson 2005). However, Rowlingson (2005:295) defines ‘an insider’ as “Someone who has skills, knowledge, resources or access, considered privileged to, or under the control of, an organisation”. This definition is broader as it also covers skills, knowledge, resources and access.

This paper accepts this definition because there are four situations that are exploited by malicious insiders to abuse or commit a crime towards the organisation’s resources: these insiders are trusted, privileged to perform specific tasks, have physical access to a target system, and have knowledge of where the valuable resources are.

When an insider commits a crime or abuses a system, internal network forensic investigation should be ready to deal with internal abuse/crime such as to prove that a crime has been committed and identify the insider. It is often the case that an electronic crime has been committed and the guilty party needs to be identified. For this reason network forensic readiness is required. When internal abuse or crimes occur, digital forensic readiness is usually required in order to reveal electronic evidence (e-evidence) from networks or computers that are destined for use in court. Furthermore, Endicott-Popovsky and Frincke (2006) define network forensic readiness as “maximizing the ability of an environment to collect credible digital evidence while minimizing the cost of an incident response”. However, Kent and Ghavalas (2005) confirm that a large number of organisations do not have any processes or procedures in place for handling security computer events which results in mishandling admissible evidence. This is because the digital forensic practice is usually ad-hoc and lacking in widely accepted theoretical models or principles (Taylor et al. 2007).

In fact, the task of implementing a network forensic investigation is a complicated issue because it is required to comprehend the legal system. Endicott-Popovsky et al. (2007:5) believe that “Implementing network forensic readiness in an organization will require accepting an expanded role for systems and network administrators, as well as an understanding of how legal requirements for admissible evidence can be translated into information system requirements i.e., what network data to collect and where; what tools and procedures to use and how; who should be trained and in what topics; etc. Adapting a tool alone or a technique will not be sufficient.” It is also not an easy task because of the nature of the crime (i.e. electronic) and the evidence (i.e. digital). Electronic evidence can be exploited to hide unlawful activities. The following list provides a few of the activities:

- Anonymity
- The rapid development and variety of methods to commit a crime
- The difficulty of confirming that a crime has been committed
- Hard to find evidence
- Ability to conceal evidence
Ease in destroying evidence (Casey 2004a; Al-Alma 2004; Technical Working Group for Electronic Crime Scene Investigation 2001)

In addition, the purpose of this research is to propose a framework of network forensic investigation at an organisational level, in order to improve the process of revealing e-evidence that is destined for use in an Islamic court. Saudi Arabia is chosen because the source of constitution of Saudi Arabia is Al Sharia law which comes from the Qur’an and Al Sunnah1. Moreover, there is no defined way to analyse electronic crimes. A case study of an insider hacking at an organisational level, in Saudi Arabia, is analysed and examined in order to reveal the issues facing digital forensic investigators handling e-evidence. In this context, this research will address the following issues:

- How does network forensic investigation work at an organisational level in Saudi Arabia?
- What are the major problems facing network forensic investigations at an organisational level in Saudi Arabia?
- What is the proposed solution to improve the process of network forensic investigation at an organisational level in Saudi Arabia?

2. COMPUTER CRIME IN SAUDI ARABIA

800 million dollars (£400 million) have been invested by the Saudi government in the first phase of the implementation of the Country's e-Government program (Ruiz 2006). The Saudi government is keen to develop a high quality of government services in order to promote an attractive environment for foreign investments (Ruiz 2006).

According to the annual report of Communication and Information Technology Commission in Saudi Arabia, the number of internet users in Saudi Arabia has increased from 1,000,000 in 2001 to 4,700,000 in 2006. It also indicates that broadband subscribers have increased from 14,000 in 2001 to 220,000 in 2006 (Communication and Information Technology Commission 2006:14). However, the government of Saudi Arabia is highly concerned about computer crimes. According to Naif Arab Academy For Security Sciences (NAASS), financial exploitation is of major concern. The cost of computer crimes in the Middle East was approximately $600,000,000 (£306,618,704) in 2000 (Al-Qasem and Al-Zhrani 2004). In the same year, the cost to Saudi companies and the public sector reached $150,000,000 (£76,655,735) (25% of the total cost of computer crime in the Middle East) (Al-Qasem and Al-Zhrani 2004).

There are a number of forms of behaviour which may lead to computer abuse/crime based on religion and public interest in Saudi Arabia. These are impersonation, espionage, information disclosure, modification, information theft, defamation adultery (false accusation of adultery), materials contradicting religion (pornography, gambling and dangerous religious ideas), forming illegal relationships by chatting via network communications, and threatening national security such as terrorist websites (Al-Sanad 2004; Mansour 2002; Al-Qasem and Al-Zhrani 2004; Al-Alma 2004). In Saudi Arabia, a survey of common cyber crimes (Alminshawi 2003) was conducted and found that the most common internet crimes were hacking, financial related offences, sexual abuse, and immoral behaviour such as obscene email, opponent websites and piracy. Importantly, the study also confirms that hacking is the highest cyber crime in the Saudi society.

In 2007, the law has been amended in Saudi Arabia to combat the growing threat of cyber-crime (IT in Saudi Arabia 2007). This law deals with offences, such as hacking or the use of internet resource to spread terrorism. This law establishes that website defacing is a crime worthy of punishment, when data theft can carry a fine of more than $130,000 or even a maximum one year prison sentence. The Saudi law has also applied this punishment to those found guilty of defamation using electronic means.

---

1 The Qur’an: According to Muslims, it is the speech of God.
Al Sunnah: The saying and teaching of the Prophet Mohammed.
or those who unlawfully break into private computer networks. Spreading malware could result in paying a fine of $800,000 and spending up to four years in jail, also those found guilty of distributing what is considered to be obscene martial face the same punishment. Users who create websites with pornographic content or content that defames humanity, or promotes drug use will be punished with fines of up to $1,300,000 and five years jail (IT in Saudi Arabia 2007). Table 1 shows these e-crimes with its punishments.

The Communication and Information Technology Commission in Saudi Arabia has also started the establishment of a national centre for the Saudi Arabian Computer Emergency Response Team (CERT-SA) in 2007. The team is planning to play a critical role in awareness, detection, prevention, coordination and response to information security incidents at a national level (Communication and Information Technology Commission 2006). Although this is an improvement for Saudi Arabia towards the protection of information technology, the centre is required to prove that a crime has been committed and provide the link between suspects and victims. Therefore, translating the principle of the traditional forensic investigation process in Saudi Arabia into information technology is the most important goal to be achieved.

Is the cyber crime law enough to protect information technology? Unfortunately the answer is no. E-crime presents one of the major challenges to Saudi Law enforcement and corporate security professionals because there is a lack of digital forensic readiness (Al-Qasem and Al-Zhrani 2004; Mansour 2002). Corporate security professionals in Saudi Arabia find that it is difficult to detect and prove when an e-crime has taken place and finding the link between suspects and victims. Al-Anazi (2003) confirms that most Saudi companies could not identify malicious insiders, because there is a lack of digital forensic investigation guidelines. He also confirms that detection and identification of the source of attack in the Saudi private sector was limited. Detection of the incident without identifying the source of attack was also low (Al-Anazi 2003). Section 3 will discuss the issues facing digital forensic investigation to identify the source of attack in order to improve the process of identifying a malicious insider.

Therefore, the major problem facing corporate security professionals in Saudi Arabia is a lack of network forensic readiness such as searching for, collecting and analysing e-evidence especially with respect to procedural Islamic criminal law (these procedures will be discussed in section 5).

<table>
<thead>
<tr>
<th>Section</th>
<th>Types of Information Crime</th>
<th>Sanctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 3</td>
<td>Sniffing and hijacking data and communications without permission; Blackmail; Unauthorized access to web sites in order to modify, destroy and delete data.</td>
<td>Prison for maximum of one year and a fine of maximum Saudi Riyals (SR) 500,000 (£73,965), or one of these sanctions.</td>
</tr>
<tr>
<td>Section 4</td>
<td>Computer fraud; Unauthorized access to bank account details in order to obtain data, money or services.</td>
<td>Prison for maximum of three years and a fine of maximum SR 2,000,000 (£295,238), or one of these sanctions.</td>
</tr>
<tr>
<td>Section 5</td>
<td>Unauthorized access to private data in order to sabotage or circulate it. Making the network out of services (logic bomb) Denial of Services</td>
<td>Prison for maximum of four years and a fine of maximum SR 3,000,000 (£443,786), or one of these sanctions.</td>
</tr>
<tr>
<td>Section 6</td>
<td>Materials against public interests or personal’s privacy by creating, saving or sending these materials such slanders and pornography. Creating illegitimate websites in order to sell or advertise illegitimate activities (which contradict with Islam religion) such as gambling, drags and alcohols.</td>
<td>Prison for maximum of five years and a fine of maximum SR 3,000,000 (£443,786), or one of these sanctions.</td>
</tr>
<tr>
<td>Section 7</td>
<td>Creating a website for terrorist group in order to assist them to communicate, share their data or collect money and support their ideas. Unauthorized physical or logical access to national security information.</td>
<td>Prison for maximum of ten years and a fine of maximum SR 5,000,000 (£739,644), or one of these sanctions.</td>
</tr>
</tbody>
</table>

Table 1. Saudi cyber-law
To address this issue, the old policy of security in organisations should be changed to accommodate the cyber crime law. Previously, organisations could only protect against cyber crime but now it is possible to bring the criminals to justice through Islamic law. The authors concluded this section by emphasising that Saudi Arabia needs forensic investigations to deal with e-crime with respect to Al-Sharia law (Al-Sanad 2004; Alminshawi 2003; Al-Qasem and Al-Zhrani 2006). Figure 1 illustrates how a malicious insider commits a crime to an organisation’s resources or using these recourses to commit other crimes. This figure also demonstrates why digital forensic mishandling occurs.

The following section will illustrate a case study where network and system administrators wrongly mishandled computer evidence.

![Figure 1 Main Issues of Handling E-Evidence](image)

### 3. A CASE STUDY OF AN INSIDER IN SAUDI ARABIA

One of the biggest companies in Saudi Arabia has a limited policy on computer incident response that does not include computer forensic preparation procedures (e.g. logging incidents). The following case demonstrates these issues that took place, before the Saudi law of cyber crime was issued.

A malicious insider logged into the domain using the domain administrator account and illegally added his account and another account to the Internet groups (manager’s groups). When the information security department detected this security incident event they disabled both accounts and the malicious insider’s PC. He then used the domain administrator again and enabled both accounts and modified the configurations on his PC so that no one could access his PC. Figure 2 demonstrates the electronic evidence that a malicious insider’s account was moved to open the Internet policy (will be discussed in the next section).
3.1 Analyzing the Case Study

According to the law of cyber crime in Saudi Arabia, hacking is considered as illegal entry to a computerized system. Illegal entry is defined as “an entry with a deliberate manner to the PC, a website, or information system, or network computers without being authorized” (IT in Saudi Arabia 2007).

Two types of policies are implemented for Internet access in this organization. The first type is confined access to only one hour for normal users; whereas the second type is open (unlimited) access to the Internet for managers and above. The insider gained unauthorized access in order to modify the policy by moving his account and another account from confined access to unlimited access.

Table 2 illustrates that there is a lack of investigative information about the malicious attack. Network forensic readiness is required to illustrate how, when and by whom the security incident occurred, what methodology is used to attack the system, and all of this with respect to Saudi procedural criminal law. Figure 4 demonstrates a fictional attack methodology that might have been used by the insider against the system in Saudi Arabia. Therefore, network forensic readiness will play a critical role in handling computer crime evidence. This is achieved by understanding the legal requirements for admissible evidence and how that translates into information system requirements:

- What happened and how;
- What network data to collect and from where;
- What tools and procedures to use and how;
An important question has to be answered: why are corporate digital forensics investigators in Saudi Arabia not able to handle e-evidence? Research conducted by Dhillon (1999) and Melara and Sarriegui, (2003) on internal security controls provide a partial solution. Although, they examine and enhance internal security controls: formal, technical and informal, when one of the three kinds of control measures is not correctly implemented, a malicious insider poses a major risk to an organisation’s resources. Therefore, the authors believe that their proposals can also be improved to examine, enhance and implement network forensic readiness at an organisational level. The following section will examine why corporate digital forensic investigators are not able to handle e-evidence. The definition of formal, technical and informal controls will be amended in order to adapt them for network forensic investigation at an organisational level.

### Table 2. Information about crime and investigation

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Hacker case in Saudi Arabia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Facts related to case</strong></td>
<td></td>
</tr>
<tr>
<td>Types of attack</td>
<td>Hacking (gaining access)</td>
</tr>
<tr>
<td>Reason of attack</td>
<td>To access the Internet</td>
</tr>
</tbody>
</table>

| **Facts related to network forensic investigation** |
| Investigator time | Not identified |
| Consequence | No prosecution |
| Evidence | Insufficient |
| Investigator | System and network administrators |
| Network forensic readiness | Ad hoc |

**Figure 4.** Attack methodology of the malicious insider against the system

#### 3.1.1 Lack of formal controls for digital forensics

Formal controls for network forensic investigation are responsible for establishing proper policy and procedure to ensure that an organisation is able to collect and gather digital evidence, before an e-
crime takes place. Comprehension of the legal requirements for admissible evidence is required in order to translate it into information system requirements.

The majority of digital forensic investigations in Muslim countries have not understood the Islamic legal requirements for admissible evidence because their governments only recently have issued a law on combating cyber crimes. For example Saudi Arabia issued the law in 2007 and the United Arab Emirates (UAE) issued their cyber crimes law in 2006 (Gulf news report 2007). Formal controls in organisations are still not complete.

Furthermore, they do not have any processes or procedures in place for handling events that result in the requirement to produce admissible evidence (Kent and Ghavalas 2005). This is because the digital forensic practice is usually ad-hoc and lacking in widely accepted theoretical models or principles (Taylor et al. 2007). By creating guidelines on how to collect and analyse evidence and seizing hardware, it is possible to present the evidence to court. As previously mentioned, Al-Anazi (2003) confirms that in the Saudi private sector the detection and identification process of a malicious insider is limited and just detection is also very low.

Therefore, in order to identify why detection and identification of a malicious insider is limited in the majority of organisations in Saudi Arabia, the incident response policy for the case study was examined:

1. The incident policy is still under research by the information security group in the organisation and is not fully implemented yet;
2. There are no clear guidelines on:
   - How to report e-crime;
   - What data should be collected and how;
   - Where to find it;
   - What procedures to follow;
   - What tools to use to collect the evidence;
   - Who is responsible to collect the evidence;
   - How to protect the integrity of the evidence;
   - How to preserve the evidence;
   - How to protect the privacy of the suspects.
3. The policy does not consider the requirement of Saudi law (the requirement of the law will be discussed in the section 4).

Therefore, the security incident policy of this organisation is insufficient and ineffective in handling e-evidence because the translation of the Islamic legal system into information system is completely absent from this policy. The absence and misunderstanding of the legal system in the policy will lead to mishandling e-evidence. Section 4 will address this main issue by examining the principles of Al Sharia law in order to enhance formal controls.

On the other hand, the United States of America (USA) has addressed these issues and a short brief of this procedure is given. In the USA, the procedure of collecting electronic evidence from both computer networks and computer systems has been well documented in order to assist public and private forensic investigations in handling admissible electronic evidence based on the legal system of the United States (Technical Working Group for Electronic Crime Scene Investigation 2001). However, these procedures can not be implemented in Saudi Arabia, since there are differences the law. In most Western cultures publication or ownership of pornographic materials is not a crime, whereas under Al Sharia law they are a crime (Alama 2004).

The next section will look at why technical tools do not support digital forensic investigation to produce e-evidence.
3.1.2 Deficiency of technical controls for digital forensics

After the shortcomings of formal controls facing digital forensic investigation to deal with e-evidence were illustrated, the question will be answered: What is the impact of shortcomings of formal controls? What tools are missing to support digital forensic investigation? Before finding out the answer, this research will discuss in general the issues preventing digital forensic investigation from handling e-evidence.

Technical controls for network forensic investigation are tools and techniques responsible for detecting unacceptable behaviour, gathering and storing digital evidence at a technical level, such as Intrusion Detection System (IDS). There are a number of issues facing the investigation in order to collect evidence from internal threats. One of the main issues facing forensic investigations is to produce admissible evidence. From a digital forensic perspective, there is a lack of tools. The main reason for this is that tools are designed with information security in mind, rather than evidence processing (Casey 2004b). Another issue is that although network-based log files provide clues, they present a problem to investigators (Mandia et al. 2003). The logs are stored in many different formats, such as syslog and SNMP traps, and in different places each time which can cause confusion. Due to these facts, special software is required every time to access and read the logs. Therefore, these difficulties prevent digital forensic investigation from producing admissible evidence to prove an e-crime has been committed and identify a malicious insider.

Crime scene of the case study

Some information about the operating system of the organisation’s network will be given in order to comprehend the shortfalls and mishandling of the case study. The organisation uses a client/server Operating System (OS). A client is an organisation user’s computer which requests data or service from an organisation’s server; whereas a server provides services or data to a client such as email, and directory services.

Active Directory (AD) (crime scene) is a critical database of users, computers and network resources and makes the network resources accessible to users and applications (Microsoft 2003). The groups are stored in the AD and are monitored by Microsoft Operations Manager (MOM). MOM can monitor, manage, and secure a wide range of resources including computers, applications, Web server farms, and corporate servers (Microsoft Corporation 2005). When the insider breached the policy for accessing the Internet, MOM detected it and sent an email to the AD administrator.

From the point of digital forensic, there are a number of shortcomings in this case study and they are:

1. Audit trails were not reviewed on both sides. The aim of the audit trails is to provide a summary of information of events that can be tracked to aid in identifying events that have happened and which compromised the information systems in an organisation (Blyth and Kovacich 2006). The trails usually record a user through their identification name or number. The identification is unique and ties the user to the activities on the system (Blyth and Kovacich 2006). The audit trails will then assist network forensic investigators to identify the link between a malicious insider and the crime.

The organisation uses a Client/server OS, therefore it is important for digital forensics to review the audit trails that are located in the security event logs on both the client and server:

The server:

- Audit logs: the investigators did not successfully track the logon of the administrator domain account into the AD, which would provide the link between a malicious insider and the crime;
- Auditing policy change: the investigators missed out on a lot of evident clues that the policy was changed;
Proxy logs: the investigators did not analyse the proxy log that proved that a malicious insider was browsing the Internet for more than the allowed period of time (i.e., one hour).

The client (a malicious insider):

- Audit logs: failure and successful events were also not examined to prove the link between a malicious insider and the crime.

2. This organisation is dependent only on one mechanism (i.e., MOM) to protect the AD;
3. A malicious insider’s computer was not examined to find out the methods that were used to gain the administrator access because the investigators did not discover what tools were used by the insider.
4. The conflicting time stamps between the security incident report and the Active Directory’s log. The log of the AD was only reviewed and a snapshot of the log indicates that there is a compromise of integrity. The security incident report shows the crime had taken place at 11:00 AM; whereas the AD log recorded the last logon was 11:10 AM. Figure 5 illustrates that last logon to the System was 11:10. As a result of this conflict, e-evidence is not acceptable in the court.
5. The authors interviewed corporate security professionals in the organisation and inspected the Intrusion Prevention System (IPS) and found that the IPS and IDS had not been updated for a year. Therefore, when a malicious insider used a brute force attack, the IDS could not detect it.
6. A block writer was not used to protect the integrity of the electronic evidence.

![Figure 5. Evidence of the time stamp of last logon](image)

3.1.3 Lack of informal controls for digital forensics

Informal controls are responsible for training computer forensic personnel, educating and increasing the awareness of end-users, which this is out scope of the research. In fact, there is a lack of digital
forensic experts that deal with internal threats and there are no training programmes to develop their technical skills (Alserhani 2004; Al-Sannd 2004; Al-Alama 2004; Al-Qassem and Al-Zhrani 2006). It is also vital to train end-users on how to deal with e-crime and how to report e-crime; for example by not switching a computer off when s/he has detected a crime. According to our case, it appears that the system and network administrators have no skills in handle e-evidence.

Kent and Ghavalas’s (2005) propose that the addressed proactive approach naturally results in improved handling of incidents. Their paper is aimed at enhancing and improving the skills of forensic experts. Figure 6 demonstrates the malicious mechanism that was used, the shortcomings of the network forensic investigation and the reason for mishandling e-evidence in our case study.

After determining the main issues that prevent digital forensic investigation from correctly handling e-evidence, it is important for digital forensic experts to understand the Islamic legal system when handling evidence in Saudi Arabia.

![Diagram of A malicious Insider's Mechanism]

**Figure 6.** Major issues of mishandling evidence at the organisational level in Saudi Arabia

### 4. FORENSIC PROCEDURES BASED ON ISLAMIC LAW

This section will discuss the traditional forensic procedure in Saudi Arabia in order to apply it on to information systems. It will also assist the investigators in understanding the legal requirements of Islamic law in order to handle and produce e-evidence at the organisational level in Saudi Arabia.

The forensic process and procedure has been created by the country’s ruler for collecting evidence, investigating and securing a verdict in order to punish the guilty based on Al Sharia law (Dafiri 2003). When a crime has taken place in a Muslim society, evidence is required to support the allegations because the Qur’an states “bring forth your proofs, if you are truthful.” (The noble Qur’an 27:64). The aim of forensic investigations in Islam is to know the truth and achieve justice between people. This procedure includes the following steps:

#### 4.1 Initial preparation: Inference (A stage towards proving the committing of a crime)

After a crime has been reported, the first step is to ensure the crime has been committed because the Messenger of Allah, (peace upon him) reported that “…but evidence on the plaintiff and denied the
right to” (Binothaimeen 2004).

Inference is defined as providing evidence based on the sources of Al Sharia law. It is also defined as providing private evidence because it is not mentioned in the Qur'an and the Sunna (Dafiri 2003). For instance, there is no process and procedure on how to collect evidence of an advanced crime such as the hacking of a system and impersonation within computers.

Nowadays, inference is called a process of collected evidence in forensic procedures. It is aimed at making sure that a crime has been committed because it is written in the Qur'an that “If an evil-doer comes to you with a report, looks carefully into it, lest you harm a people in ignorance…” (The Noble Qur'an 49:6).

The goal of this phase is achieved by collecting a large amount of information about a specific crime to decide whether to take forward the investigation or to close the case because of insufficient evidence.

In the Islamic law, it is important to provide the evidence when the victim reports the crime. Methods of proving crime in Al Sharia law is a controversial issue because there are two points of view (AlKarmi 2005; Al-Zohaili 1994). The first point of view believes that these methods are limited to only specific methods such as witnesses, confession and oath. These views are based on the Qur'an and the Sunna (Al-Zohaili 1994). The second point of view believes that these methods are unlimited to include any method that leads to the truth such as witness, confession, strong evidence, bearing testimony (Al Qarinah) and a scientific method.

The definition of evidence from Al-Sharia is a sign which could assist in finding an answer to a mystery (Al-Zohaili 1994). There are a number of rules for collecting evidence in Al-Sharia as follows (Al-Zohaili 1994):

- It is based on scientific techniques: evidence should not be guessed or predicted but the evidence should be extracted from a scientific process;
- It provides a link between a crime and its victim or a crime and its perpetrator. If there is a strong link between them this evidence is called strong otherwise it is weak evidence. Strong evidence is acceptable in Al-Sharia as a main method of proof in Al Sharia law; whereas weak evidence is unacceptable because it is based on prediction.

Therefore, the initial preparation is important when collecting and accumulating information about the crime that will give the investigator a clear picture; such as hearing witnesses and interviewing the victim (Dafiri 2003).

These days’ computer abuses/crimes require special methods to prove and identify the offender. Searching, collecting and analysing information is required to extract the evidence in order to achieve justice.

To apply the initial preparation phase to computer abuse/crime, the proposed procedure below should be followed to ensure evidence that a crime has been committed:

- Interview the victim;
- Hear the witnesses;
- Determine the method to be used to detect and identify the incident such as IDS system, log file analysis or system administrator;
- Collect information about the operating system, file structure, types of attacks, when it happened, how, and where and the consequences of attacks (The International Association of Computer Investigative Specialists 2007);
- Determine the condition of a targeted machine (i.e., switched off or on)(ACPO 2005);
- Determine the role of the computers in a crime (e.g., a computer is a target, tool or store of evidence) (Casey 2004a);
- Record the procedure of initial preparation phase.
As previously mentioned, this phase will assist the investigator in understanding the role that the computers have in the crime. This information will be useful for the next phase because the investigator can determine what digital evidence he/she has to look for and what kind of tools. For instance, if there is a crime that involves a certain tool, the forensic investigator will look for software tools such as a Trojan Horse. Therefore, partial imaging of the drive or selective files will only be required. Initial preparation might also solve the dispute of the privacy of people because it is likely to determine the evidence before the investigation has been conducted.

4.2 Forensic investigation phase

Investigation is defined in Islam as proving the crime with admissible evidence (Dafiri 2003). Forensic investigation is a set of legal procedures that have been followed by investigators before a trial in order to discover the truth and identify the offender by inspecting and analysing the evidence of crime (Dafiri 2003). There are a number of forensic investigation processes such as:

4.2.1 Searching process

This is a process of investigation that has been conducted to search for specific evidence of a crime within private property (Dafiri 2003). The searching process is conducted only after the suspect is identified, the crime has been committed and sufficient evidence is collected. It is a technical process of finding a link between a crime and a suspect in order to prove or refute a crime.

According to article 45 of the forensic process in Saudi Arabia, an investigator has permission only to search for information and things related to a particular crime (Dafiri 2003). Therefore, the ultimate goal of this phase is to search for information that relates to the specific crime and to protect the privacy of people. In fact, the process of searching breaches Al-Sharia law because it violates the privacy of people, the Qur’an states “spy not” (The noble Qur’an 49:12). However, if it is necessary, this process is allowed but it is very restricted (Al-Sannd 2004). It must be stopped immediately when evidence has been found (Dafiri, 2003).

To apply a searching process to computer abuse/crime, these steps have to be followed:

- It is vital to determine the condition of a computer whether it is switched off or on because some information might not be recovered if the computer is switched off such as volatile data (e.g., ARP cache and routing table). ACPO guidelines are useful because they refer to the condition of a computer and suggest what to do in both situations;
- The computer system should be examined physically. Documentation should include a physical description and detailed notation of any irregularities (The International Association of Computer Investigative Specialists 2007).
- A picture of the screen should be taken when it is switched on;
- To protect the integrity of the evidence, whole imaging of the offender’s computer is required in the investigation of computer crime. However, it breaches the suspect’s privacy, based on Islamic law. Therefore, partial file copying may solve this problem;
- It is not allowed for the forensic collector of evidence to examine the evidence because it violates the privacy of people;
- Investigation should be started where the data of evidentiary value is most likely to be found such as (The International Association of Computer Investigative Specialists 2007):
  - A full directory listing should be made to include filenames, time stamp and so forth;
  - Files created by users should be examined using file viewers such as e-mail and database;
  - Operating system files should be examined such as registry, temporary, cache and history files;
  - Only an authorised person is to analyse the evidence;
  - It is useful to examine unused and unallocated space volume for previously deleted data, deleted folders and slack space data;
- All findings should be recorded;
All the processes should be recorded.

4.2.2 Seizing process

The main aim of the searching process is to seize the evidence for a specific crime. Seizing is retaining evidence with legal authority in order to protect the integrity and availability of evidence (Dafiri 2003). Seizing is not only to prove an allegation but is also to disprove an allegation.

The articles 55, 56, 57 and 59 in the forensic procedures in Saudi Arabian law could be applied to seize computer crime evidence (Dafiri 2003):

- It is allowed to seize the computer without accessing the data and the date and time of seizing the computer should be recorded;
- It is allowed to keep the seized contents of a computer, CDs, flash memory, floppy disks and papers away from the suspects and saved in a safe box (The International Association of Computer Investigative Specialists 2007);
- It is allowed to seize any printouts and images related to this crime;
- It is allowed for the owner to have the seized papers and documents returned;
- The investigator is responsible for protecting the confidentiality of the seized items.

4.2.3 Inspection Phase

The aim of this phase is to give the whole picture of the crime by proving the link between a crime scene and a suspect or between a suspect and a victim. This phase is important because it will prove or refute the link between a suspect, a crime and the seized items in order to provide an answer.

To apply this process to computer abuse/crime, these processes should be followed:

- The original data should not be examined in order to protect the integrity of evidence;
- The name of the suspect should be kept from the examiner to avoid lying and cronyism;
- There should be a logical order of the inspection process;
- When the contents of a seized computer are received by an examiner, the examiner should check and sign for them;
- Investigators may choose to implement a forensically sound operating system. The use of physical write-blocking hardware or software may be used in operating system environments that are not forensically sound;
- There should be one examination of the copy data;
- Examine the log files which is located in IDS, router, firewall, DHCP, and so on;
- The evidence should be extracted;
- All the processes should be recorded.

4.2.4 Expert witness

An expert witness is someone who has a high level of skill and knowledge. He has scientific knowledge of advanced methods for proving the process of linking a suspect and a crime scene. It is allowed in Islam to have an expert witness (Dafiri 2003), it is written in the Qur’an that “So ask of those who know the Scripture if you know not” (The noble Qur’an 16:43). There are a number of conditions when using an expert witness such as (Dafiri 2003):

- An expert is required to provide the process of linking a suspect and a crime scene;
- An expert is required to provide a report at the required time;
- An expert can have recourse to another expert;
- It is allowed for both a victim and a suspect to have recourse to a private consultant to examine the processes of the expert;
- Both a victim and a suspect can reject the report of the expert.
4.3 Reporting

The final step, digital forensic investigation is required to answer what type of crime, type of evidence, how to collect, what happened, when and who. Table 3 demonstrates the main principles of forensic procedure in Al Sharia law.

<table>
<thead>
<tr>
<th>Principles of forensic procedure in Saudi Arabia</th>
<th>Translated into information system requirements</th>
</tr>
</thead>
</table>
| Chain of evidence                              | • Make sure a crime has taken place by reviewing the methods of detecting attacks such as IDS, IPS, users and security professionals.  
• What types of e-crime;  
• Starting to build a case from the log file of the detecting mechanism until reporting by making a link between each log based on the time stamp. |
| Privacy of suspect                              | • Make sure there is a link between a suspect and a computer crime by reviewing the network’s and system’s logs;  
• Looking for only specific information;  
• stop immediately when evidence has been found;  
• Make an image of selective files by using en-case. |
| Integrity of evidence                           | • No action is taken on the original copy;  
• Make an image of selective files;  
• Use block writing. |
| Availability of evidence                        | • Keep the evidence in safe place such as CD, flash memory and printout. |

Table 3: main principles of forensic procedure in Al Sharia Law

5. CONCLUSION

The hacking case showed that organisations in Saudi Arabia are in desperate need for an investigative standard. The authors found that network forensic investigation at the organisational level uses ad-hoc guidelines to collect e-and analyse evidence. There is also a limited policy on computer incident response that does not include computer forensic preparation procedures (e.g. logging incidents). Moreover, these guidelines do not include the translation of the principle of forensic investigation process in Saudi Arabia into information systems. As a result, mishandling of e-evidence usually occurs and prevents investigators from collecting valuable evidence, based on Al Sharia law. To address these shortcomings, the chain of evidence, the privacy of the suspect, the integrity and availability of the evidence are considered in the process of network forensic investigation.
In the future the existing forensic procedure, in Saudi Arabia, will be enhanced with applicable digital forensic guidelines. This will then lead to an investigative standard for digital evidence. The investigative standard will include many existing computer forensic procedures in order to enhance formal and technical controls of network forensic investigation at the organisational level in Saudi Arabia. Figure 7 demonstrates the main contents of a technical investigative standard for digital evidence applicable for Al Sharia law.

REFERENCES


The Noble Qur'an in the English Language. Madinah, Saudi Arabia.


Blyth, A. and G. L. Kovacich (2006), Information Assurance, Springer, USA.,


Endicott-Popovsky and D. Frincke (2007), ‘Embedding Forensics Capabilities into Networks’,


Mandia, K., C. Prosise, et al. (2003), Incident Response & Computer Forensics, Corel Ventura, U.S.A.


MicrosoftCorporation (2003), 2279B Planning, Implementing, and Maintaining a Microsoft Windows Server 2003 Active Directory Infrastructure, Microsoft,USA.


The Virtual Digital Forensics Lab: Expanding Law Enforcement Capabilities

Sean A. Ensz
University of Oklahoma
200 Felgar Street, Norman, Oklahoma 73019
405.325.3954 Office
405.325.1633 Fax
ensz@ou.edu

Mark R. McCoy
University of Central Oklahoma
Forensic Science Institute
100 N. University, Edmond, Oklahoma 73034
405.974.5617 Office
405.974.3871 Fax
mmccoy@ucok.edu

ABSTRACT
Law enforcement is attempting to respond to the growing and complex need to examine all manner of digital evidence using stand-alone forensic workstations and limited storage solutions. Digital forensic investigators often find their cases stalled by cumbersome and inflexible technology limiting their effectiveness. The Virtual Digital Forensics Lab (VDFL) is a new concept that applies existing enterprise host, storage, and network virtualization technologies to current forensic investigative methods. This paper details the concept of the VDFL, the technology solutions it employs, and the flexibility it provides for digital forensic investigators.

Keywords: Virtual Digital Forensics, digital forensic investigations, law enforcement, virtual lab, Digital Forensics

1. INTRODUCTION
Law enforcement investigators have attempted to respond to the growing and complex need to investigate all matter of computer related incidents by using stand-alone forensic workstations and limiting storage solutions. Forensic investigators often find that their cases are held up by cumbersome and inflexible technology that limits their effectiveness. The need to store and examine large quantities of data and the need to provide easy access to examination results to investigators in remote locations has changed to face of the digital forensics laboratory. This paper details the concept of the Virtual Digital Forensics Laboratory (VDFL), the technology solutions it employs, and the flexibility it provides for digital forensic investigators.

2. VIRTUALIZATION
A Virtual Computer Forensics Lab (VCFL) is a new concept that applies existing enterprise virtualization technology to current forensic investigative methods. Virtualization technology was introduced in the 1960s to allow the full use of mainframe hardware, but more recently virtualized network, storage and workstation technologies have matured to the point where they can be used to effectively overcome computer forensics lab constraints. Today virtualization is helping many
Information Technology (IT) organizations solve problems with scalability, security, and management. Virtualization can help computer forensic labs do the same.

A computer forensics lab must be able to keep pace with the technology it analyzes, and it must allow investigators secure remote access to forensic tools. Virtualized hosts and virtualized storage, along with strong network encryption, allow organizations the flexibility for multiple investigators to collaborate using the same evidence, while using as many virtual forensic workstations as needed, with a storage system that can scale to hundreds of terabytes.

Virtualization technology is the abstract layer that resides between what is presented and the physical hardware. There are three core virtualized technologies needed to create a virtual lab environment. They are virtual private networks, virtual machines, and virtualized storage. A fourth (non-virtualized) component, using two-factor identity management technologies, is also needed to create a secure and confidential lab environment. This technology can be applied to existing computer forensics labs to create a complete virtualized layer that still meets rigid ASCLD (American Society of Crime Laboratory Directors) requirements (ASCLD, 2008).

3. VIRTUAL PRIVATE NETWORK

VPN connections are accomplished through the use of network firewalls that create an encrypted tunnel between the user and the network being accessed. VPN technology uses strong encryption mechanisms that make it almost impossible to snoop the network traffic. Recently, a new type of VPN called an SSL VPN became available that allows traffic to essentially be tunneled through a web proxy.

VPN technology can be used to provide secure remote access by computer forensic investigators to work cases from a remote office. (See Figure 1). The firewalls can be configured in such a way that prevents remote workstations from accessing any hosts on the internet other than the virtual workstations. Once the original evidence has been duplicated and placed on virtualized storage, it can be accessed from a fully functioning virtual workstation loaded with any forensic software needed by the investigator. The remote investigator should see no difference between working a case within the lab or remotely.

Figure 1.

4. VIRTUALIZED STORAGE

Enterprise level storage virtualization that facilitates access from multiple workstations and servers requires the use of Storage Area Network (SAN) architecture. SAN is a storage system that uses
remote storage arrays in such a way that a disk appears to be locally attached. The physical disks use Redundant Array of Independent Disks (RAID) technology to protect against disk failure and are scalable to allow additional storage to be added when needed. Many large regional and state computer forensic labs currently use SAN technology, but they fail to fully take advantage of the benefits virtualized storage provides. Virtualized storage is considered the process of abstracting logical storage from the physical disks. Traditional drive mapping requires groups of physical disks (called a LUN) to be assigned to a host. The virtualization system, however, presents the logical storage space for data storage and the controller handles the process of mapping it to the physical disks. This technology helps by providing efficient utilization, shared storage without restrictions and increased performance.

Efficient utilization is accomplished by managing disks as a single pool and presenting disk resources to any server. Utilizing all disk resources all the time provides for maximum efficiency. Shared storage without restrictions is provided by removing the limitations of physical drives by aggregating them into logical, virtual volumes. Write virtual volumes to the entire disk pool or any subset of the disk pool; provision without restrictions. Increased performance is gained by scaling performance linearly across all the available drives; making each disk drive's performance characteristics accessible to each server. Any volume can simultaneously utilize all of the disk drives in the shared pool to access data (Compellent, 2008).

Virtualization technology can also include presenting a set volume size that is larger than the sectors allocated on the physical disk drives (called thin provisioning). With traditional storage, disk space must be pre-allocated and is often not fully utilized by the host. This can create a large amount of storage space that goes unused. Consider a situation in which five forensic investigators each request 500 gigabytes of SAN space to work their case, but they only end up using 300 gigabytes of actual storage space. In this scenario, 1 terabyte of space, using a traditional SAN, would go unused. With thin provisioning the 1 terabyte would still be available to work other cases.

5. VIRTUAL WORKSTATIONS

Virtualization is an abstraction layer that decouples the physical hardware from the operating system. Using virtual machine software, it is possible create a virtualized forensic workstation environment that completely emulates all aspects of a real physical workstation. This technology allows multiple virtual workstations to run on a single server that can act independently of each other with granular permissions. Access to storage and network resources can also be tightly controlled and monitored to only allow virtual machines to access the resources assigned to the investigator. Since all hardware at the virtual layer is standard, it is possible to create uniform forensic workstation builds that can be duplicated for each new case. Forensic investigators will be able to use a pristine build that has not been contaminated by previous cases.

Virtualization provides for partitioning, isolation, and encapsulation. Using partitioning multiple applications and operating systems can be supported within a single physical system. Computing resources are treated as a uniform pool to be allocated to virtual machines in a controlled manner. Virtual machines are completely isolated from the host machine and other virtual machines. If a virtual machine crashes, all others are unaffected. Data does not leak across virtual machines and applications can only communicate over configured network connections. With encapsulation, complete virtual machine environment is saved as a single file; easy to back up, move and copy. Standardized virtualized hardware is presented to the application to guarantee compatibility (VMWare, 2008).

It is possible to start and assign multiple virtual machines to investigators that allow them to work multiple cases at one time. Investigators often contend with downtime because a workstation is tied up making images, or conducting text searches or data carving. With a robust virtual machine infrastructure it is possible to assign multiple virtual workstations, with dual processor support and ample memory, to an investigator with varying operating systems if needed.
6. PROPOSED DESIGN

The transition from the traditional forensic lab to a virtualized forensic lab is as simple as adding a virtualization technology layer over existing infrastructure and processes. The lab design will meet, and in some cases exceed, stringent standards for data security, evidence handling, and investigation techniques. Strong network encryption and two-factor authentication schemes will be used to ensure confidentiality and integrity of all lab equipment. Evidence handling and investigative processes will meet standards set for by the ASCLD Laboratory Accreditation Board Manual (ASCLD, 2008).

6.1 Network

The network design will use hardware based site-to-site VPN tunnels between remote investigator PC and the physical computer forensics lab. The use of hardware VPN appliances ensures that only equipment properly configured by lab personnel will be able to gain entry into the secure network. The VPN tunnels will use strong 128-bit AES (Advanced Encryption Standard currently adopted by NIST and FIPS) encryption algorithms to prevent data leakage and network sniffing over internet links. Remote VPN appliances will be configured in such a way that investigator workstations will be restricted to only access the hosts on the VPN. Internet access will be completely restricted from the workstations. The design, using VPN technology, will mimic a standalone physical network in which all workstations are connected to the same physical switch. See Figure 2 for a logical diagram of the virtual digital forensics lab design.

6.2 Storage

The storage will incorporate an enterprise SAN solution that will provide redundancy, fault tolerance, and scalability. The solution will also use cutting edge controllers that will provide virtualization technology, providing flexibility and ease of use. New virtual volumes will be created for each new investigation that can be resized and duplicated as needed. Virtual storage technology allows snapshots to be taken of existing virtual volumes if multiple investigators want to work with the same evidence. Virtualization technology is also on the horizon that will allow a virtual volume to be mirrored to a tape drive that will allow duplicate forensic images to be simultaneously written to a tape backup while a working forensic copy is written to the virtual volume. This is done entirely in the background without any additional steps needed by the forensic technician acquiring the original evidence. Destruction write technology will also be used to wipe physical sectors on the SAN hard drives before new data is written.

6.3 Workstations

The virtual forensic workstations will be running an enterprise version of VMware on server hardware to provide flexibility and performance. The remote investigator will use a low cost PC to access, through the VPN tunnel, the powerful virtual forensic workstation running locally in VMware. The Microsoft remote desktop client will be used to gain access to the virtual forensic workstation. Once the remote desktop connection is made the investigator will be required to enter his or her username followed by a PIN and six digit number presented on a key fob. The virtual forensic workstation will only have access to the virtual volume containing a working copy of the evidence that was assigned to the investigator.

Case workflow will occur in the following stages:

- Initial intake and tagging of original evidence
- Virtual Volume is created on the SAN to meet case storage requirements
- Original evidence is imaged to a virtual volume on the SAN
- New virtual workstation is started and assigned to an investigator
- Virtual volume on SAN is mapped to the virtual workstation
- Virtual workstation is granted network access to investigator remote PC
• Investigator initiates a remote desktop connection and authenticates
• Investigator begins forensic analysis

Figure 2.

REFERENCES
AUTHORS BIOGRAPHIES

Sean A. Ensz, CISSP, GSEC, EnCE, RHCE, is an Information Technology Security Analyst at the University of Oklahoma.

Mark R. McCoy is currently an Assistant Professor of Criminal Justice at the University of Central Oklahoma. In June 2008, he will join the faculty of the UCO Forensic Science Institute as the Digital Evidence and Cyber Security Program Administrator. He recently retired after over 20 years of service with the Oklahoma State Bureau of Investigation. He is a member of the International Association of Computer Investigative Specialists and is a Certified Forensic Computer Examiner. He was the first supervisor of the OSBI Computer Crime Unit and has conducted hundreds of forensic examinations involving digital evidence. He has a Masters Degree in Forensic Science and a Doctorate in Occupational and Adult Education.
How Virtualized Environments Affect Computer Forensics

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

ABSTRACT
Virtualized environments can make forensics investigation more difficult. Technological advances in virtualization tools essentially make removable media a PC that can be carried around in a pocket or around a neck. Running operating systems and applications this way leaves very little trace on the host system. This paper will explore all the newest methods for virtualized environments and the implications they have on the world of forensics. It will begin by describing and differentiating between software and hardware virtualization. It will then move on to explain the various methods used for server and desktop virtualization. Next, it will describe the fundamentals of a traditional forensic investigation and explain how virtualization affects this process. Finally, it will describe the common methods to find virtualization artifacts and identify virtual activities that affect the examination process.

Keywords: Hardware-assisted, Hypervisor, Para-virtualization, Virtual Machine, virtualization, VMware, Moka5, MojoPac, Portable Virtual Privacy Machine, VirtualBox,

1. INTRODUCTION
According to a research published by Gartner in February of this year, there are nearly 100 providers of products adapted for the server virtualization management marketplace [1]. Fewer than 5 million PCs were "virtualized" in 2006; by 2011, that figure will rise to between 480 million and 846 million [2].

Virtualization of both clients and servers has several attractive benefits that are fueling the use of virtual machine environments (VMEs). With more emphasis being placed on going green and power becoming more expensive, virtualization offers cost benefits by decreasing the number of physical machines required within an environment. A virtualized environment offers reduced support by making testing and maintenance easier. On the client side, the ability to run multiple operating environments allows a machine to support applications and services for an operating environment other than the primary environment.

The deployment of virtualization software is nearly a given for servers using 64-bit processors and many have built-in virtualization capabilities. However, as the use of VMEs increases, computer attackers are increasingly interested in detecting the presence of VMEs, both locally and across the network. There are some specific uses of VME technology that are driving the underground toward deploying techniques for virtual machine detection as well as an increase of virtual environments and applications that can be run from a USB device. The interest in the use and detection of VMEs is not limited to those that want to spread malware writers or conceal activities. When malicious code is released that makes use of its own VME, it will become essential for anti-malware researchers to find ways to detect the VME. Additionally, computer forensics professionals will be required to detect and examine such environments.

2. HOW VIRTUALIZATION WORKS
Is it quite complex to virtualize to an operating system. There are several published papers
explaining how virtualization works. For example, Keith Adams and Ole Agesen from VMware have written “A Comparison of Software and Hardware Techniques for x86 Virtualization”. Building on this concept is “Attacks on More Virtual Machine Emulators” by Peter Ferrie. In order for virtualization to happen, a hypervisor is used. The hypervisor controls how access to a computer's processors and memory is shared. A hypervisor or virtual machine monitor (VMM) is a virtualization platform that provides more than one operating systems to run on a host computer at the same time. This section will take a brief look at the underlying technologies of virtualization.

2.1 Hardware

A Type 1 native or bare-metal hypervisor is software that runs directly on a hardware platform. The guest operating system runs at the second level above the hardware. These hardware-bound virtual machine emulators rely on the real, underlying CPU to execute non-sensitive instructions at native speed [3]. However, since they execute instructions on a real CPU, there are some changes to the environment, in order to share the hardware resources between the guest operating system and the host operating system. In hardware virtualization, a guest operating system is run under control of a host system, where the guest has been ported to a virtual architecture which is almost like the hardware it is actually running on. This technique allows full guest systems to be run in a relatively efficient manner [4]. The guest OS is not aware it is being virtualized and requires no modification. Full virtualization is the only option that requires no hardware assist or operating system assist to virtualize sensitive and privileged instructions. The hypervisor translates all operating system instructions on the fly and caches the results for future use, while user level instructions run unmodified at native speed [5].

2.2 Paravirtualization and Hardware Assist

Paravirtualization involves modifying the OS kernel to replace nonvirtualizable instructions with hypercalls that communicate directly with the virtualization layer hypervisor. The hypervisor also provides hypercall interfaces for other critical kernel operations such as memory management and interrupt handling [5].

The Virtual Machine Interface (or VMI) was developed by VMware as a mechanism for providing transparent paravirtualization. The VMI interface works by isolating any operations which may require hypervisor intervention into a special set of function calls. The implementation of those functions loads a "hypervisor ROM" [6]. This design also allows the same binary kernel image to run under a variety of hypervisors, or, with the right ROM, in native mode on the bare hardware.

Hardware vendors are rapidly embracing virtualization and developing new features to simplify virtualization techniques. In hardware assist virtualization, the guest state is stored in Virtual Machine Control Structures (VT-x) or Virtual Machine Control Blocks. Second generation hardware assist technologies are in development that will have a greater impact on virtualization performance while reducing memory overhead [5]. Both AMD and Intel have announced future development roadmaps for this technology.

Table 1 is a comparison of the hardware virtualization types as listed by VMware in “Understanding Full Virtualization, Paravirtualization, and Hardware Assist”.

Table 1
<table>
<thead>
<tr>
<th>Technique</th>
<th>Full Virtualization with Binary Translation</th>
<th>Hardware Assisted Virtualization</th>
<th>OS Assisted Virtualization / Paravirtualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary Translation and Direct Execution</td>
<td>Exit to Root Mode on Privileged Instructions</td>
<td>Hypercalls</td>
<td></td>
</tr>
<tr>
<td>Guest Modification / Compatibility</td>
<td>Unmodified Guest OS Excellent compatibility</td>
<td>Unmodified Guest OS Excellent compatibility</td>
<td>Guest OS codified to issue Hypercalls so it can’t run on Native Hardware or other Hypervisors Poor compatibility; Not available on Windows OSes</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Fair</td>
<td>Better in certain cases</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Fair</td>
<td>Better in certain cases</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Fair</td>
<td>Better in certain cases</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Fair</td>
<td>Better in certain cases</td>
</tr>
<tr>
<td>Used By</td>
<td>VMware, Microsoft, Parallels</td>
<td>VMware, Microsoft, Parallels, Xen</td>
<td>VMware, Xen</td>
</tr>
<tr>
<td>Guest OS Hypervisor Independent?</td>
<td>Yes</td>
<td>Yes</td>
<td>XenLinux runs only on Xen Hypervisor VMI-Linux is Hypervisor agnostic</td>
</tr>
</tbody>
</table>

Table 1: Summary comparison of x86 processor virtualization techniques[5]

2.3 Hosted hypervisors

A Type 2 or hosted hypervisor is software that runs within an operating system environment and the guest operating system runs at the third level above the hardware. The hypervisor runs as an application or shell on another already running operating system. Operating systems running on the
hypervisor are then called guest or virtual operating systems. A virtual machine monitor (VMM) provides a layer of software between the operating system(s) and hardware of a machine to create the illusion of one or more virtual machines (VMs) on a single physical platform. A virtual machine entirely encapsulates the state of the guest operating system running inside it [7]. Virtual machines are composed entirely of software and contain no hardware components whatsoever. As a result, virtual machines offer a number of distinct advantages over physical hardware.

Thus, the host can boot to completion, and launch any number of applications as usual, with one them being the virtual machine emulator. That emulator then sets up some CPU-specific control structures and uses the VMLAUNCH (Intel) or VMRUN (AMD) instruction to place the operating system into a virtualized state. At that point, there are effectively two copies of the operating system in existence, but one (the host) is suspended while the other (the guest) runs freely in the new state. Whenever an interesting event (an intercept, interrupt, or exception) occurs, the host operating system (the virtual machine emulator) regains control, handles the event, and then resumes execution of the guest operating system [3].

2.4 Embedded Hypervisors

Dell is exploring the concept of embedding a lightweight hypervisor in the firmware of some future Dell servers. Embedding a hypervisor into a system could offer a number of benefits including:

- Reducing software installation to a simple file copy
- Reducing software updates to a simple file copy
- Windows, Linux, BSD Unix and just about anything else could run simultaneously and at the same time
- Obsolete devices could appear to be present even though they’ve not be manufactured or supported for years
- New technology, such as hardware supporting advanced graphics, high-speed networking and storage, could be included in the system and appear to software as well-known, much loved, older devices that are compatible with generations of software [8].

The Qtopia Phone Edition demo has been designed to run in a Linux virtual machine using VMware Player and Palm OS released Garnet VM software for the Nokia series of Internet Tablet devices in November of 2007. Garnet VM is a virtual machine software application for running Garnet OS-based applications in a Linux environment that essentially acts as an emulator allowing Palm OS applications to run on a Nokia N770, N800 and N810 Internet Tablet [9]. By 2012, more than 50% of new smartphones shipped will include hardware virtualization support [10].

3. VIRTUAL TECHNOLOGY IN BUSINESS

As mentioned that the beginning of this paper, there are nearly 100 providers of products adapted for the server virtualization management marketplace [1]. IDC Research predicts that spending on virtualization will reach almost $15 billion worldwide by 2009. This section discusses some of the major players and more popular products used for virtualization.

3.1 VMware and VMware Preconfigured Appliances

Perhaps the best know virtualization vendor is VMware. VMware invented virtualization for the x86 platform in the 1990s to address underutilization and other issues, overcoming many challenges in the process [11]. VMware offers a wide variety of services and products. Of particular interest are the preconfigured appliances that can be readily downloaded and implemented using VMware Player. VMware Player makes it simple to quickly evaluate one of the many virtual appliances available through the VMware Virtual Appliance Marketplace. A virtual appliance is a pre-built, pre-configured and ready-to-use enterprise software application on a virtual machine. With VMware Player, anyone
can quickly and easily experience the benefits of preconfigured products without any installation or configuration hassles [11].

3.2 Microsoft Virtual Products

Microsoft provides a full suite of technologies to enable an integrated, end-to-end virtualized infrastructure. Using familiar interfaces and common management consoles, a virtualized environment based on Microsoft technologies simplifies infrastructure management and delivers powerful capabilities [12]. Microsoft’s emphasis on virtualization technologies is rooted in creating what is called a Dynamic IT environment. Microsoft’s solution includes servers, desktops, and applications virtual machine management and virtualization acceleration. Recently Microsoft has released beta versions of Hyper-V and Application Virtualization. Hyper-V provides software infrastructure and basic management tools in Windows Server 2008 that can be used to create and manage a virtualized server computing environment. Hyper-V requires an x64-based processor, hardware-assisted virtualization, and hardware data execution protection. Microsoft Application Virtualization 4.5 Beta includes new capabilities designed to aid in the support of large-scale virtualization implementations across more sites and enable multiple delivery options [12]. Microsoft has recently begun offering pre-configured Virtual Hard Disks (VHDs) that can be downloaded evaluated similar the virtual appliance market.

3.3 XenSource

Another major virtualization competitor is XenSource. XenSource distributes its hypervisor as free, open-source software but sells related products. Simon Crosby, the firm's technology chief, likens this to giving away an engine in order to sell a car around it. He believes this approach will help to spread virtualisation more quickly and prevent VMware from establishing a Microsoft-like dominance of the market. XenSource now has the necessary backing, having recently been acquired by Citrix, another software firm, for $500m. Some people think that if Microsoft fails to catch up with VMware on its own, it will buy Citrix [13]. The Citrix Delivery Center offers comprehensive end-to-end virtualization solutions, with server, application and desktop virtualization – purpose built to enable IT to deliver applications to users anywhere [14].

3.4 Parallels Preconfigured Appliances

Parallels is similar in VMware in that it provides virtualization solutions along with preconfigured appliances except for the Macintosh platform. Parallels provides server, desktop, automation and management solutions. Parallels announced the availability of a Template Catalog for Parallels Virtuozzo Containers. The company offers a library of more than 350 software downloads that can be used to easily create and manage operating systems and applications running in virtual environments [15]. The Parallels Virtuozzo Templates are comparable to VMware's Virtual Appliance Marketplace. Parallels also offers service providers access to more than 100 applications certified according to the Application Packaging Standard (APS) for software-as-service (SaaS) solutions.

3.5 Virtualization Boxes

Besides the above named companies, there is a wide variety of other companies that offer virtual solutions. Since the market is growing at a very fast pace, included in this section are only the solutions the author found might be relevant to the computer forensics realm.

InnoTek’s VirtualBox is a mature virtualization tool that runs on Windows and Linux and supports Windows (including Vista), Linux, OS/2 Warp, OpenBSD, and FreeBSD as guest operating systems. Like many companies have done, InnoTek has split its product into two editions: An open source version and a full version with additional features aimed at enterprise customers[16]. Sun Microsystems, Inc. announced that it entered into a stock purchase agreement to acquire InnoTek in mid February of this year.
Pano Logic offers a complete desktop virtualization solution. The Pano device is a zero client. It has no CPU, no memory, no operating system, no drivers, no software and no moving parts. The Pano device connects keyboard, mouse, display, audio and USB peripherals over an existing IP network to an instance of Windows XP or Vista running on a virtualized server [17]. Pano Logic has moved the PC and all its software off the desktop and into the data center. A management device sits between the Pano device and the virtualization server.

The InBoxer Anti-Risk Appliance combines powerful email archiving, electronic discovery, and real-time content monitoring in a single appliance. The InBoxer virtual appliance separates the software from the hardware it is running on and can adopt storage virtualization [18].

4. VIRTUAL TECHNOLOGY FOR INDIVIDUAL USE

The use of virtualization is growing in the individual use market as well as the corporate environment. This section explores the technology being used with personal computer that do not alter the current environment, but use a USB device to run the virtual environment, thereby leaving the original system intact.

4.1 MojoPac

MojoPac is developed by RingCube. It now has a MojoPac enterprise solution which includes mojostation, mojodrive and mojonet. MojoPac Usher is an application that can be installed on host computers to enable MojoPac to run with a limited mode host login. MojoPac’s virtualization technology encapsulates a complete Windows desktop environment, including applications, files and settings isolating it from the underlying host PC. This virtualized environment can be loaded onto a host computer, a portable USB storage device or network attached storage and run on any Windows host computer [19]. Ring Cube announced the end of February that it has surpassed 100,000 registered users of its MojoPac software platform. This rapid adoption further demonstrates RingCube’s innovative leadership in the Desktop Virtualization product category to enable users to gain greater mobility, productivity, and access to their desktops, applications, and data. “The market for portable virtual workspaces is just beginning to develop and the response to MojoPac has been overwhelming,” said Pete Foley, CEO of RingCube Technologies. “To capture over 100,000 registered users in just over 1 year of the first available download for registered users validates the significant value and benefit customers have gained using MojoPac’s revolutionary approach of virtualizing desktop environments and the strength of our product offering”[20]

4.2 Moka5

Moka5 LivePCs contain everything needed to run a virtual computer: an operating system and a set of applications. LivePCs can be run from a USB flash drive, USB hard disk, iPod, or a desktop computer [21]. LivePCs can be created that are similar to VMware and Parallels concept. A LivePC can be downloaded from a repository of public LivePCs created in a LivePC Library. LivePCs are run on Moka5 Engine. This technology will stream and prefetch LivePCs so they can be shared and automatically updates the LivePCs as the maintainers make changes. There is a BareMetal Edition Beta of the Moka5 Engine that is installed on a separate disk partition and the desktop computer boots directly into Moka5 Engine.

4.3 Portable Virtual Privacy Machine

The Portable Privacy Machine by MetroPipe contains a complete virtual Linux machine with privacy-enabled Open Source Internet applications. Carry your Internet applications, email, bookmarks, history, web cookies, and download files in your pocket. The Portable Privacy Machine is based on Damn Small Linux (DSL) and QEMU releases[22]. QEMU is a generic, open source processor emulator.
4.4 Preconfigured virtual appliances

VMware hosts about 725 virtual appliances that can easily downloaded and installed. Earlier it was mentioned that Parallels offers more than 350 virtual appliances. Available ready to go, are over 1,000 virtual appliances that anyone can use.

5. ADOPTION OF A STANDARD FOR PACKAGING VIRTUAL MACHINES

On November 27, 2007, the Distributed Management Task Force, Inc. created an open standard for system virtualization management "With the ever-increasing adoption of virtualization, DMTF aims to simplify and provide ease-of-use for the virtual environment by creating an industry standard for system virtualization management," said Winston Bumpus, DMTF president. "Our role also extends to ensure the success of this standard, so we are thrilled to host the first-ever SVPC plugfest to test early implementations for interoperability."[23] This standard recognizes supported virtualization management capabilities for discovering virtual computer systems, managing the lifecycle of virtual computer systems, controlling virtual resources and monitoring virtual systems.

6. HOW THESE TECHNOLOGIES AFFECT FORENSIC INVESTIGATIONS

Traditionally virtual machines have been used to create contained environments for malware isolation or to examine suspect machines. VMware can be used to mount a dd image. Applications like LiveView create a VMware virtual machine out of a raw (dd-style) disk image or physical disk. This allows the forensic examiner to boot the image or disk and gain an interactive, user-level perspective of the environment without modifying the underlying image or disk. Because all changes made to the disk are written to a separate file, the examiner can instantly revert all of his or her changes back to the original pristine state of the disk[24]. Virtual Forensic Computing (VFC) utilizes VMware's VMPlayer and the forensic disk mount tool Mount Image Pro, to re-create a subject machine in a matter of seconds. However, now instead of using virtual environments to examine machines, virtual environments themselves need to be examined.

Virtualization technology allows mobile employees to leave hardware behind and take only software with them. Entire environments can now be carried on micro devices such as a USB drive or iPod. Organizations are exploring the possibilities of downloading a virtual machine from a browser on a borrowed machine. All these technological changes present new challenges to the traditional methods of performing computer forensics.

6.1 MojoPac

MojoPac technology was described earlier. Listed below are issues that this technology presents:

- All documents and personal items can be copied to the drive, before launching.
- Once started, access to the local hard drive is eliminated.
- Access to CD and removable drives is still possible.
- May need administrative rights on the host machine in order to run.
- Currently will only run on Windows XP
- MojoPac has it’s own separate registry and shell
- Programs of the same name may be running on both at the same time
- MojoPac implements paging between memory and the hard drive to take place on the host PC instead of on the portable drive
- Process is RingThreeMainWin32
- Browsing and multimedia history stays inside MojoPac
- According to website, after a user Exits and Ejects MojoPac, there is no trace left behind on the Host PC
6.2 Moka5
Moka5 technology was described earlier. Listed below are issues that this technology presents:

- Installs VMware Player
- Asks whether you want to leave it installed for easier load next time
- Moka5 Engine will stream and prefetch LivePCs
- Any changes made during a session are captured in separate file systems on a ramdisk
- Creates folders in the my documents folder for Live PC

6.3 Portable Virtual Privacy Machine
Portable Virtual Privacy Machine technology was described earlier. Listed below are issues that this technology presents:

- Very small LINUX distro designed to boot from a USB drive
- No installation needed
- Just plug the drive into any Windows or Linux computer, and click on the Virtual Privacy Machine icon

6.4 Preconfigured virtual appliances
Preconfigured virtual appliances were described earlier. Listed below are issues that this technology presents:

- No installation needed, runs via VMware Player
- Virtual Applications can be combined. Example: BackTrack2 with Metasploit 3

7. WHAT TO LOOK FOR
In many of the aforementioned technologies, virtual devices are exclusive to the virtual machine and are files on the host. For example, VMware creates virtual adapters as well as files with extensions: .vmx, .vmdk, .vmsn., and vmss. “What Files Make Up a Virtual Machine?” posted on VMware’s website is an explanation of all the files extensions that are associated with VMware along with the purpose of the file. Since some forensic software lists these extensions as unknown file types, a forensic examiner should become familiar with these files. The same goes for Microsoft’s virtual products where the vhd format is utilized.

The host’s critical resources such as memory, processor time, video, and sound are shared with the virtual machines. In applications such as MojoPac, the host resources must be utilized for better performance. Log files are created by most software; virtual machines are no exception, look for these. Since many of these technologies use a USB drive for access, there will be remnants in the registry. The March 2007 edition of Digital Investigation has an article titled “Tackling the U3 trend with computer forensics”. Here Andy Spruill and Chris Pavan explore the artifacts left behind by U3 devices. This article is of particular interest in the investigation of virtual devices run from USB devices. When investigating virtual machines, some general items to examine include:

- MRU cache
- Link files
- Prefetch files
- Page file
- Unique identifiers associated with the program
• Artifacts in processes, file system, and/or registry
• Artifacts in memory
• VME-specific virtual hardware, processor instructions and capabilities

In the corporate environment, Application-layer security, such as application proxies can capture some evidence that can help track actions. Application-layer firewall logging can capture more than the IP address and port number. Application-layer firewalls are capable of intercepting packets traveling to or from an application such as a browser. This provides a more thorough examination of network traffic and can capture evidence from applications such as Moka5 and Portable Virtual Privacy Machine. Corporations also have the option of not allowing removable media. This can eliminate the issues that arise from using many of the technologies mentioned here.

The home environment becomes a bit more difficult. If the user is computer savvy, finding tracks may be almost impossible. Devices are becoming smaller with larger capacity and can easily be hidden. Home environments need to be examined very closely for all CDs and removable devices.

8. CURRENT CHALLENGES

A virtual machine located inside forensic software cannot currently be examined by the software. Most software reports the virtual machine files as unknown file types. Although the virtual machine can be exported or loaded into another virtual machine, when that suspect virtual machine is loaded the information inside the original virtual machine changes.

In his presentation on the Effectiveness of Hash Sets, Douglas White of the National Institute of Standards & Technology (NIST) compares physical and virtual OS installations. There is a difference in the number of files in each type of installation. His research shows the differences in physical vs. virtual machines appear to be due to devices:

• Virtual machines use abstract/generic device interfaces
• Physical machines require vendor specific drivers

This being said, any investigation now must first determine if the device being examined in real or virtual. Determining if the device is real can be done in several ways. In November of 2004, Joanna Rutkowski published the Red Pill or how to detect VMM using (almost) one CPU instruction [25]. The Red Pill focuses on detecting virtual machine usage without looking for file system artifacts based on relocation of sensitive data structures. Scoopy Doo and Jerry are tools that detect a VMware fingerprint. When Scoopy Doo is run, it simply states: This is/is not a virtual machine. These tools can be found at: http://www.trapkit.de/research/vmm/index.html. On this website, Tobias Klein also poses the question “is it possible to break out of a VM (to reach the Host OS or to manipulate other VMs)? This is quite an interesting question as the implications can be great since virtualization is based on isolated environments. For those more adventurous, Snoopy Pro is available. This tool analyzes virtual traffic between the device and driver.

When examining virtualized environments, it is important to reflect on what is being captured. Tools available to examine virtual environments are limited. The Volatility Framework 1.1.1 is a collection of tools, for the extraction of digital artifacts from volatile memory (RAM) images. The framework is intended to introduce people to the techniques and complexities associated with extracting digital artifacts from volatile memory images and provide a platform for further research into this area [26]. In May 2007 Network General added virtual server forensics. The company added modules that let IT personnel peer into the workings of VMware’s ESX and Microsoft Virtual Server. However, when Henderson and Dvorak, who members of the Network World Lab Alliance, tested the virtual-machine-monitoring capabilities, they found it takes a lot of preparation and configuration work to yield useful data [27].
9. WHAT THE FUTURE HOLDS?

Virtualization appears to have a definite hold on the market and companies are competing fiercely to develop and implement products for this environment. At the end of May 2007, May Google silently acquired application virtualization startup GreenBorder. GreenBorder uses application virtualization for security containment of desktop software like browsers, email clients, and rich-media players[28]. Along with all these changes and technologies, challenges will come.

Our court system already has a difficult time with cyber crime. Earlier this year a federal grand jury issued a subpoena to MySpace.com in a case where a teenage girl committed suicide. Federal prosecutors are considering charging Lori with defrauding MySpace for creating a false account. In another recent cyber crime case, the judge ruled that there was no crime because it was a faceless crime. The judicial system is not equipped to keep up with the changing face of crime. Since virtual machines and environments are being used, can crimes committed be construed as virtual crimes? Susan W. Brenner explored the question: Is There Such a Thing as "Virtual Crime"? Her research show that the actual entry into the computer or computer system presumably occurs in the “virtual world,” as would the steps an offender intends to take in order to commit an offense. This fact is not enough to prevent the liability for the offenders conduct. There is still a legally cognizable harm, such as the offender’s entering an area to which she does not have lawful access and thereby violating the owner of that area’s right to exclude those to whom she has not granted access. As to this fact, it is conceptually irrelevant whether the location that is unlawfully accessed exists in the physical world or in the virtual world; the harm to the owner of that area is logically indistinguishable [29]. However, the possibility of a challenge based on virtual environments still exists.

What happens when we have kiosks that a user downloads a virtual environment into a browser, commits a crime, and then deletes the virtual machine? This can happen anywhere. Virtual social networks continue to grow. How will crime be investigated in Second Life? Not long ago there were people stealing WOW gold and selling it on eBay.

In virtualization, there is the ability to roll back or delete a bad or defective machine. With the Federal Rules of Civil Procedure governing data retention, will virtual machines need to be included in an organization’s data retention policy? As investigators find ways to examine virtual machines, ill the processes be questioned as to the original evidence file? Borrowing the last line from “Attacks on More Virtual Machine Emulators” by Peter Ferrie: “One thing is clear – the future looks complicated”.

REFERENCES


http://blogs.zdnet.com/virtualization/?p=205

Posted By: Ryan Kairer on Tuesday, November 13, 2007.


http://www.citrix.com/English/ps2/products/product.asp?contentID=683711


[16] Running, J. VirtualBox: InnoTek's virtualization goes open source


[18] InBoxer Virtual Appliance http://www.inboxer.com/


http://www.businesswire.com/portal/site/home/?epi_menuItemID=8529ea2ad8631dcd3bb97904c6908a0&epi_menuID=887566059a3aed6efaa9e27a808a0c&epi_baseMenuID=384979e8cc48c441ef01305c6908a0c&ndmViewId=news_view&newsLang=en&newsId=20080226005603


http://www.metropipe.net/ppm.php?SID=4e38766ea6a3fab4792ced91b2bdbe48

http://www.dmtf.org/newsroom/pr/view?item_key=70d5d3ba78d39488626f838397a3d1e9812e5d40


[25] Rutkowski, J. Red Pill... or how to detect VMM using (almost) one CPU instruction
http://invisiblethings.org/papers/redpill.html


SiMPLE - Rethinking the Monolithic Approach to Digital Forensic Software

Dr Craig Valli
Edith Cowan University
Perth, Western Australia
Phone +61-8-93706013 Email c.valli@ecu.edu.au

ABSTRACT

This paper outlines a collaborative project nearing completion between the sec.au Security Research Group at Edith Cowan University and Western Australian Police Computer Crime Squad. The primary goal of this project is to create a software tool for use by non-technical law enforcement officers during the initial investigation and assessment of an electronic crime scene.

This tool will be designed as an initial response tool, to quickly and easily find, view and export any relevant files stored on a computer, establishing if further expert investigation of that computer is warranted. When fully developed, the tool will allow investigators unprecedented real time, on site access to electronic evidence whilst maintaining complete forensic soundness.

Keywords forensic, triage, images, project, police, case study

1. INTRODUCTION

The primary goal of this project is to create a software tool for use by non-technical law enforcement officers during the initial forensic investigation and assessment of an electronic crime scene. This project is being conducted by the sec.au Security Research Group at Edith Cowan University in conjunction with the Western Australian Police Computer Crime Squad.

This tool will be designed as an initial response tool, to quickly and easily find, view and export any relevant files stored on a computer, establishing if further expert investigation of that computer is warranted. When fully developed, the tool will allow investigators unprecedented real time, on site access to electronic evidence whilst maintaining complete forensic soundness. This tool is customer focused and should provide an efficient, effective and reliable response to electronic crime. It is anticipated that this will significantly benefit evidence discovery by

- Improved service delivery
- Facilitate at scene interview of suspects
- Enable at scene comparisons of images to environment (Children at Risk)
- Significantly reduce property seizure
- Reduce demand for specialist forensic examination by enabling immediate and local resolution of issues

It is anticipated that the software tool will, in combination with the development of training and protocols for its use, allow law enforcement investigators to collect electronic evidence that can be introduced into the scrutiny of a judicial process. This project has incorporated extensive testing to an international standard which will be used to prove the developed tool is fit for purpose.

The Computer Crime Squad (CCS) is responsible for providing the Western Australia Police (WAPS) with expert services in the investigation, identification, preservation, interpretation and articulation of electronic evidence. The ubiquitous nature of computers, the Internet, mobile telephones, computer networks and other electronic devices in society today has resulted in new and complex challenges for
law enforcement. Australians are avid consumers of technology having one of the highest adoption rates in the world. For law enforcement this translates into more offenders using technology to commit crimes, and electronic devices being a potential source of evidence for all crimes.

To add further complexity to policing within Western Australia the state is geographically sparse with a land area of 1 million square miles. Some crime scenes are located over 3000 kilometres from the main office of the computer crime squad in Perth. This means that personnel and equipment must be transported at considerable expense to these remote locations for on-site triage and for the presentation of evidence in court.

2. ONE SIZE DOES NOT FIT ALL

One of the major issues facing police is the ability to respond to electronically initiated crime. Much of the existing response capability is based around the use of solutions that have a significant investment in training of staff to a level where they become proficient enough to analyse and present electronic evidence in a court of law. The software and hardware solutions whilst in of themselves are very powerful forensic analysis tools in the hands of suitably qualified experts are in some regards a self deprecating loop. The deprecating loop is as a result of increasingly high levels of specialisation and training required for the expert use of these increasingly complex “forensic” tools even for simple tasks. So it logically and legally follows that use of these expert tools by other than suitably acquitted staff may rightfully produce polemic evidence or outcomes. Therefore, instead of having a simple purpose built tool or process that can be used by individuals with minimal training to affect a simple profiling exercise, organisations are increasingly compelled to use complex and expensive software driven by well-trained, niche specialists to undertake even simple preliminary investigations. A further effect is that of a positive feedback loop for the investigation of incidents, which overtime, regardless of magnitude, size and seriousness, will require increasingly higher levels of expertise and equipment.

This feedback loop is the equivalent of using a large laser level guided 30 tonne bulldozer to dig a 20cm x 20cm x 20 cm deep hole in free flowing sand. Although technically feasible, the particulars of a task can be accomplished far more easily, efficiently and effectively with a purpose built trenching shovel. This is the argument underpinning fundamental principle of design used in this project and that is that less is more.

Currently, there are a multitude of security and forensics related bootable CDs available for download on the Internet or distributed restrictively by police forces that are potentially suitable for use by experienced and expert users for forensic analysis. There are however, several serious forensic problems and quality issues with the use of these CDs as “baseline” systems on which to build systems. Firstly, the author challenges any user of either KNOPPIX STD (Cumming, 2008) Helix (e-fense, 2008) or any similar monolithic bootable CDs to describe the purpose and function of all security related binaries on these systems. It is possible to by using these CDs to produce polemic outcomes for a prosecuting party. For instance, Helix contains known malcode/malware and penetration testing tools none of which should have a place on a system that is being used for initial examination of a suspect’s computer or device, yet some investigators persist in using Helix for initial examinations and acquisition.

The issues of binary abundances aside there are basic and fundamental forensic operational issues with these CD systems. A stark example is a default installation of KNOPPIX STD by default mounts any swap spaces on the hard disks that are present in the computer. This is not an ideal outcome when the only evidence that may be of any use in a prosecution now resides in a contaminated swap file as result of the use of the CD by a forensic novice.

Due to the large number of binaries in existence on these CDs the subsequent menus are extensive, complex and require significant understanding. Each binary not required for a particular purpose relating to the forensic process being undertaken potentially also introduces another compounding
variable into the forensic integrity equation. Furthermore, this large monolithic approach is similar to the bulldozer example above. Take a simple task such as cleaning media and then imaging and validating the subsequent image. Why do you need to use something as complex as Helix, Knoppix or even monolithic commercial tools to do this. It would be smarter, less contentious and more prudent to use a small purpose built and tested CD containing the operating system and only the required specialist binaries. In this case a validated bootable CD with dd or other suitable imaging program and requisite utilities to produce hashes of the acquired images and requisite system logs would be sufficient.

Furthermore, a problem with the multi-purpose security or forensics CD is that many of them will start various additional services some of which may compromise the forensic integrity of the computer under investigation. In addition, this mode of operation causes often unnecessary use of computer resources in terms of processor and memory, which can be better utilised in performing the required forensic task at hand.

One of the advantages of the bootable LINUX environments is that they can be fully customised to suit a specific need or requirement. For experienced users this is a relatively simple task to accomplish and normally requires a CD burner and a PC with sufficient RAM and hard disk space for decompression and recompression of the resultant CD image. This well established process allows for the production of highly customised, highly specific niche CDs to be produced for use in variety of ways. For this reason the system under development has utilised a Linux bootable CD environment for its development.

Fitness to task
As mentioned previously one of the major problems is the availability of solutions that will allow a non digital forensic specialist or private investigator the ability to preview evidence at scene in a forensically sound manner. While EnCase (Software 2008), Forensic Tool Kit (AccessData 2008), Helix or KNOPPIX STD etc are all competent tools they are not suited for use by the mainstream police or private investigators for the same could be contentious at best in a court of law.

One of the major issues faced by most computer crime teams globally is the possession and distribution of offensive or illegal images via the Internet. These images are typically sexual in nature and normally relate to paedophilia, bestiality, rape and other criminal acts. One of the advantages for the offender in remote Australia is that the accessibility of the Internet has made it easier for them to acquire images of an illicit or illegal nature. The potential offender’s machines are often located in remote locations and normally are in regions where specialist crime facilities are scant and reduced in size. This has a serious implication for policing as seizure or on-site triage of such equipment requires specialist skills or expert status not normally found in these rural and remote locations. Excluding remote areas, there is strong evidence to suggest that the possession of illegal or illicit images is increasingly becoming the volume crime of the digital crime area with successful campaigns netting a wide range of offenders in geographical locations and societal strata (Scoop, 2008; McAuliffe, 2001).

This current status quo scenario has precipitated the need for a validated tool for use by mainstream police to enable them to undertake sound digital forensic examination of topical images found on a computer. From this scenario several design considerations are mandatory for the tool/CD and should:

- Be able to be used with a minimum of computer knowledge or training to extract images
- Be forensically valid and minimise polemy in its modus operandii and construction
- Allow for easy extraction of evidence and duplication to forensically clean media.
- Allow for subsequent presentation of evidence from the media.
- Use validated and tested tools in its production.
3. THE SIMPLE APPROACH

The approach by the development team has been the production of a single purpose, specialised tool to use for profiling of a suspect's computer that has a simple to use interface. Unlike other similar CD projects, the startup of the CD and subsequently the program is treated as a first part of an atomic and complete process that concludes with successful output of the suspect images to DVD in a readily readable format. All means of input other than that which is required has been trapped or disabled, reducing the ability for someone to subvert or pervert the process.

The SiMPLE approach is to search only for topically found files i.e. no file carving or overtly forensic process is used to locate candidate files for examination. This does limit the search capabilities of the tool but the law enforcement advisers believe this to be the least contentious approach. The law enforcement experience has been also that offender’s motivation in viewing this type of material is one of quick access and viewing for gratification which is counter to the use of erasure and encryption processes.

The project has had staged development and testing throughout and has been so far broken into four distinct but coupled phases. Each phase has been subject to extensive and rigorous testing. These stages are the production of base operating system, production of image indexing, production of image viewer and output of results.

**Base Operating System**
The underlying kernel is a Gentoo kernel that is monolithic containing all possible drivers relevant to the preview and output of suspect material. The team determined that the ad-hoc loading of kernel modules would increase the complexity of verification and potentially lead to possible complications with evidence and the expert ability of the user. The kernel although monolithic with respect to drivers needed for display and output of images does have a reduction or elimination of unnecessary drivers. The kernel itself is stripped for example of all network drivers and the SiMPLE CD is incapable of initiating a network session. This reduction removes the contention that the investigating machine was compromised by a network borne Trojan, virus or RPC malcode during preview of the images. It also disallows the loading of network file shares and other network based services that may introduce issues into the preview.

Furthermore, all unnecessary binaries from a default install have been removed. The design tenet is if it does not have a purpose in the indexing, preview and output it does not have a place on the CD. This tenet further reduces dispute about the CD eliminating the possibilities for misuse or misapplication unlike the typical monolithic security/forensics CD such as Helix or KNOPPIX STD that contain hundreds of unnecessary binaries. The system is also designed to operate on a Pentium 3 standard PC with 256Mbytes of RAM.

**Image indexing**
This process currently spawns as a series of scripts that search the hard drive for topically stored images or movies. The resultant topical files found are indexed, hashed and their location stored in a database structure. This database structure is then used to retrieve details about the located files for use by other parts of the program or process for example final output of the results to DVD media.

**Image viewer**
The image viewer is a simple to use application that allows an officer or investigator to preview the located files visually. It creates an initial thumbnail view of all the located files, the user is then able to select a larger view of the file (in the case of a graphic) or allow the viewing of a video file by clicking on it with the mouse. Then, if the investigator wishes they can select the image for later output to the DVD by selection via simple check box.


Output of results
The final process is the output of results to sterile media, in this case a blank DVD. USB memory sticks were considered initially but issues with management and handling with respect to continuity (chain of evidence) were thought to be unnecessarily complex. A USB-based DVD burner is connected to the suspect computer allowing for the export of material of interest. In addition to exportation of the selected images and files, the program also outputs a full forensic log of all activity undertaken during the process. In addition, to the program activity logs, all the relevant system based logs are also written to the DVD. All of the data stored on the DVD is in readily accessible HTML format, which allows viewing of the suspect’s material on a wide range of platforms. Contained in the output is the file hash, file location (full path), file size and other relevant information.

Other relevant design features
The system also has some operational exception handling incorporated. The system will halt for particular instances for example if encrypted volumes or non Windows based partitions are located on the machine. These exceptions could possibly indicate obfuscation attempts by the person of interest and require higher grade of examination/scrutiny for resolution.

4. CONCLUSION
The SiMPLE project is nearing its first official beta and represents almost 4000 hours of effort from initial idea, to proof of concept and now release beta. The tool has already generated significant interest with the Australian Law Enforcement community with its simple, single task orientation.

Already the outcomes of this project have sparked the production and development of another application to extract registry keys and ownership identifiers from laptops proffered for sale on the second hand market or in the possession of a person of interest. The Laptop Inspection and Recovery System (LIARS) will be using the same base kernel and development approach of SiMPLE.

Operationally for police a tool like this, has significant potential to impact workloads and clear-up rates of digital enabled crime. This impact is that is allows non-forensically trained police officers to undertake profiling of a suspects computer in-situ and immediately.

Finally, SiMPLE is about equipping police and investigators after a minimal amount of training with easy to use tools and techniques to recover rudimentary digital evidence. We can no longer rely on a monolithic and wholly expert approach to digital evidence collection if we are to combat crime. The systems have to be easy to use and have much of the expert knowledge embedded in the tool and its modus operandii. In the same way the most existing police are trained and able to take fingerprint evidence from the scene, SiMPLE is a tool aimed at enabling the digital equivalent.

REFERENCES
McAuliffe, Wendy (2001). “Commercial child porn ring bust leads to 100 arrests”, http://news.zdnet.co.uk/internet/0,1000000097,2092866,00.htm
http://www.scoop.co.nz/stories/PO0803/S00050.htm

Data Mining Techniques in Fraud Detection

Rekha Bhowmik
Department of Computer Science
Sam Houston State University
Huntsville, Texas
rbb002@shsu.edu

ABSTRACT
The paper presents application of data mining techniques to fraud analysis. We present some classification and prediction data mining techniques which we consider important to handle fraud detection. There exist a number of data mining algorithms and we present statistics-based algorithm, decision tree-based algorithm and rule-based algorithm. We present Bayesian classification model to detect fraud in automobile insurance. Naïve Bayesian visualization is selected to analyze and interpret the classifier predictions. We illustrate how ROC curves can be deployed for model assessment in order to provide a more intuitive analysis of the models.

Keywords: Data Mining, Decision Tree, Bayesian Network, ROC Curve, Confusion Matrix

1. INTRODUCTION
Data mining refers to extracting or “mining” knowledge from large amount of data. There are a number of data mining techniques like clustering, neural networks, regression, multiple predictive models. Here, we discuss only few techniques of data mining which would be considered important to handle fraud detection. They are i) Bayesian network, for classifying risk group, and ii) Decision tree, for creating descriptive model of each risk group.

Data Mining is associated with (a) supervised learning based on training data of known fraud and legitimate cases and (b) unsupervised learning with data that are not labeled to be fraud or legitimate. Bedford’s law can be interpreted as an example of unsupervised learning (Bolton et al. 2002). The direct application of these methods to forensic accounting is limited due to almost complete nonexistence of large sets of fraud training data (Bolton et al. 2002; Jensen, 1997).

Insurance fraud, credit card fraud, telecommunications fraud, and check forgery are some of the main types of fraud. Insurance fraud is common in automobile, travel. The Uniform Suspected Insurance Fraud Reporting Form, adopted by the NAIC Antifraud Task Force 2003, replaced the prior Task Force form. This form standardizes insurance fraud data for the insurance industry and makes it easier to report and track. Fraud detection involves three types of offenders (Baldock, 1997): i) Criminal offenders, ii) organized criminal offenders who are responsible for major fraud, and iii) offenders who commit fraud (called soft fraud) when suffering from financial hardship. Soft fraud is the hardest to lessen because the cost for each suspected incident is usually higher than the cost of the fraud (National White Collar Crime Center, 2003). Types i) and ii) offenders, called hard fraud, avoid anti-fraud measures (Sparrow, 2002).

We present data mining techniques which are most appropriate for fraud analysis. We present automobile insurance example. Three data mining techniques used for fraud analysis are: i) Bayesian network, ii) Decision tree, and iii) backpropagation. Bayesian network is the technique used for classification task. Classification, given a set of predefined categorical classes, determines which of these classes a specific data belongs to. Decision trees are used to create descriptive models. Descriptive models are created to describe the characteristics of fault.

The remainder of this paper is organized as follows. In Section 2, we present the existing fraud detection systems and techniques. Section 3 the three classification algorithms and application. Section
2. EXISTING FRAUD DETECTION SYSTEMS

A fuzzy logic system (Altrock et al. 1995) incorporated the actual fraud evaluation policy using optimum threshold values. The result showed the chances of fraud and the reasons why an insurance claim is fraudulent. This system predicted slightly better results than the auditors. Another logic system (Cox et al. 1995) used two approaches to imitate the reasoning of fraud experts, i) the discovery model, uses an unsupervised neural network to find the relationships in data and to find clusters, then patterns within the clusters are identified, and ii) the fuzzy anomaly detection model, which used Wang-Mendel algorithm to find how health care providers committed fraud against insurance companies. The EFD system (Major et al. 1995) integrated the expert knowledge with statistical information to identify providers whose behavior did not fit the rule.

The hot spots methodology (Williams et al. 1997) performed a three step process: i) k-means clustering algorithm for cluster detection is used because the other clustering algorithms tend to be computationally expensive where the datasets are very large, ii) C4.5 algorithm, the resulting decision tree can be converted to a rule set and pruned, and iii) visualization tools for rule evaluation, building statistical summaries of the entities associated with each rule. (Williams, 1999) extended the hot spots methodology to use genetic algorithms to generate and explore the rules.

The credit fraud model (Groth et al. 1998) suggested a classification technique with fraud/legal attribute, and a clustering followed by a classification technique with no fraud/legal attribute. Kohonen's Self-Organizing Feature Map (Brockett et al. 1998) was used to categorize automobile injury claims depending on the size of fraud suspicion. The validity of the Feature Map was then evaluated using a back propagation algorithm and feed forward neural networks. Result showed that the method was more reliable and consistent compared to the fraud assessment.

Classification techniques have proved to be very effective in fraud detection (He et al. 1998; Chen et al. 1999) and therefore, can be applied to categorize crime data. The distributed data mining model (Chen et al. 1999) uses a realistic cost model to evaluate C4.5, CART, and naïve Bayesian classification models. The method was applied to credit card transactions. The neural data mining approach (Brause et al. 1999) uses rule-based association rules to mine symbolic data and Radial Basis Function neural network to mine analog data. The approach discusses the importance of use of non-numeric data in fraud detection. It was found that the results of association rules increased the predictive accuracy.

SAS Enterprise Miner Software (SAS e-intelligence, 2000) depends on association rules, cluster detection and classification techniques to detect fraudulent claims. The Bayesian Belief Network (BBN) and Artificial Neural Network (ANN) study used the STAGE algorithm for BBN in fraud detection and backpropagation for ANN (Maes et al. 2002). STAGE repeatedly alternates between two stages of search: running the original search method on objective function, and running hill-climbing to optimize the value function. The result shows that BBNs were much faster to train, but were slower when applied to new instances. FraudFocus Software (Magnify, 2002) automatically scores all claims. The scores are sorted in descending order of fraud potential and generate descriptive rules for fraudulent claims. FairIsaac(Weatherford et al. 2002) recommended backpropagation neural networks for fraudulent credit card use. The ASPECT group (Weatherford et al. 2002) focused on neural networks to train current user profiles and user profiles histories. A caller’s current profile and the profile history are compared to find probable fraud. (Cahill et al. 2002) build on the adaptive fraud detection framework (Fawcett et al. 1997) by applying an event-driven approach of assigning fraud scores to detect fraud. The (Cahill et al. 2002) framework can also detect types of fraud using rules. This framework has been used in both wireless and wired fraud detection systems. (Ormerod el al. 2003) used dynamic BBNs called Mass Detection tool to detect fraudulent claims, which then used a
rule generator called Suspicion Building Tool.

The different types of fraud detection are: internal, insurance, credit card, and telecommunications fraud detection. Internal fraud detection consists in determining fraudulent financial reporting by management (Lin et al. 2003; Bell et al. 2000), and abnormal retail transactions by employees (Kim et al. 2003). There are four types of insurance fraud detection: home insurance (Bentley, 2000; Von Altrock, 1997), crop insurance (Little et al. 2002), automobile insurance fraud detection (Phua et al. 2004; Brockett et al. 2002; Stefano et al. 2001; Belhadji et al. 2000), and health insurance (Yamanishi et al. 2004; Riedinger et al. 2002). A single meta-classifier (Phua et al. 2004) is used to select the best base classifiers, and then combined with these base classifiers’ predictions to improve cost savings (stacking-bagging). Automobile insurance fraud detection data set was used to demonstrate the stacking-bagging problem. Credit card fraud detection refers to screening credit applications (Wheeler et al. 2000), and/or logged credit card transactions (Foster et al. 2004; Fan, 2004; Chen et al. 2004; Chiu et al. 2004; Kim et al. 2002; Maes et al. 2002; Syeda et al. 2002). Telecommunications subscription data (Cortes et al. 2003; Cahill et al. 2002; Rosset et al. 1999; Moreau et al. 1997), and/or wired and wireless phone calls (Kim et al. 2003; Burge et al. 2001) are monitored. Credit transactional fraud detection has been presented by (Foster et al. 2004) and bad debts prediction (Ezawa et al. 1996). Employee/retail (Kim et al. 2003), national crop insurance (Little et al. 2002), and credit application (Wheeler et al. 2000). Literature focus on video-on-demand websites (Barse et al. 2003) and IP-based telecommunication services (McGibney et al. 2003). Online sellers (Bhargava et al. 2003) and online buyers (Sherman, 2002) can be monitored by automated systems. Fraud detection in government organisations such as tax (Bonchi et al. 1999) and customs (Shao et al. 2002) has also been reported.

We discuss below supervised data mining technique to detect crime using Bayesian Belief Networks, Decision trees, and Artificial Neural Networks.

2.1 Bayesian Belief Networks

Bayesian Belief Networks provide a graphic model of causal relationships on which class membership probabilities (Han et al. 2000) are predicted, so that a given instance is legal or fraud (Prodromidis, 1999). Naïve Bayesian classification assumes that the attributes of an instance are independent, given the target attribute (Feelders et al. 2003). The aim is to assign a new instance to the class that has the highest posterior probability. The algorithm is very effective and can give better predictive accuracy when compared to C4.5 decision trees and backpropagation (Domingos et al. 1996; Elkan et al. 2001). However, when the attributes are redundant, the predictive accuracy is reduced (Witten et al. 1999).

2.2 Decision Trees

Decision trees are machine learning techniques that express independent attributes and a dependent attribute in a tree-shaped structure that represents a set of decisions (Witten et al. 1999). Classification rules, extracted from decision trees, are IF-THEN expressions in which the preconditions are logically ANDed and all the tests have to succeed if each rule is to be generated. The related applications include the analysis of instances from drug smuggling, governmental financial transactions (Mena et al. 2003), and customs declaration fraud (Shao et al. 2002) to more serious crimes such as drug related homicides, serial sex crimes (SPSS, 2003), and homeland security (James et al. 2002; Mena et al. 2003). Data mining methods have solved security and criminal detection problems. [Mena, 2003] reviewed the subject (intelligent agents, link analysis, text mining, decision trees, self-organizing maps, machine learning, and neural networks) for security managers, law enforcement investigators, counter-intelligence agents, fraud specialists, and information security analysts. C4.5 (Quinlan et al. 1993) is used to divide data into segments based and to generate descriptive classification rules that can be used to classify a new instance. C4.5 can help to make predictions and to extract crime patterns. It generates rules from trees (Witten et al., 1999) and handles numeric attributes, missing values, pruning, and estimating error rates. C4.5 performs slightly better than CART and ID3 (Prodromidis,
1999) in terms of predictive accuracy. The learning and classification steps are generally fast (Han et al. 2000). However, performance decrease can occur when C4.5 is applied to large datasets. C5.0 shows marginal improvements to decision tree induction.

### 2.3 Artificial Neural Networks

Artificial Neural Networks represent complex mathematical equations with summations, exponentials, and parameters to copy neurons (Berry et al. 2000). They have been applied to classify crime instances such as burglary, sexual offences, and known criminals’ facial characteristics (Mena et al. 2003b). Backpropagation neural networks can process a large number of instances with tolerance to noisy data and has the ability to classify patterns on which they have not been trained (Han et al. 2000). They are appropriate where the results of the model are more important (Berry et al. 2000). However, backpropagation require long training hours, extensive testing, retaining parameters like the number of hidden neurons, learning rate (Bigus, 1996).

### 3. APPLICATION

The steps in crime detection are: i) classifiers, ii) integrate multiple classifiers, iii) ANN approach to clustering, and iv) visualization techniques to describe the patterns.

#### 3.1 Bayesian Network

Bayesian Network is a Directed Acyclic Graph, where each node represents a random variable and is associated with the conditional probability of the node given its parents. This model shows each variable in a given domain as a node in the graph and dependencies between these variables as arcs connecting the respective nodes. That is, all the edges in the graphical model are directed and there are no cycles.

For the purpose of fraud detection, we construct two Bayesian networks to describe the behavior of auto insurance. First, a Bayesian network is constructed to model behavior under the assumption that the driver is fraudulent (F) and another model under the assumption the driver is a legitimate user (NF), see Figure 3. The ‘fraud net’ is set up by using expert knowledge. The ‘user net’ is set up by using data from non fraudulent drivers. During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks (the observed user behavior x derived from his toll tickets) and propagating it through the network, we can get the probability of the measurement x under two above mentioned hypotheses. This means, we obtain judgments to what degree an observed user behavior meets typical fraudulent or non-fraudulent behavior. These quantities we call p(x|NF) and p(x|F). By postulating the probability of fraud P(F) and P(NF) = 1-P(F) in general and by applying Bayes’ rule, we get the probability of fraud, given the measurement x,

\[
P(F|x) = \frac{P(F)p(x|F)}{P(x)}
\]

where, the denominator p(x) can be calculated as

\[
P(x) = P(F)p(x|F) + P(NF)p(x|NF)
\]

The chain rule of probabilities is:

Suppose there are two classes \( C_1, C_2 \) for fraud and legal respectively. Given an instance \( X = (X_1, X_2, ..., X_n) \) and each row is represented by an attribute vector \( \mathbf{A} = (A_1, A_2, ..., A_n) \)

The classification is to derive the maximum \( P(C_i|X) \) which can be derived from Bayes’ theorem as given in the following steps:

i) \( P(\text{fraud}|X) = \frac{[P(\text{fraud} | X) P(\text{fraud})]}{P(X)} \)

\[
P(\text{legal}|X) = \frac{[P(\text{legal} | X) P(\text{legal})]}{P(X)}
\]
As $P(X)$ is constant for all classes, only $[P(\text{fraud} | X) P(\text{fraud})]$ and $[P(\text{legal} | X) P(\text{legal})]$ need to be maximized.

ii) The class prior probabilities may be estimated by:

$$P(\text{fraud}) = \frac{s_i}{s}$$

Here, $s$ is the total number of training examples and $s_i$ is the number of training examples of class fraud.

iii) A simplified assumption of no dependence relation between attributes is made. Thus,

$$P(X | \text{fraud}) = \prod_{k=1}^{n} P(x_k | \text{fraud})$$

and

$$P(X | \text{legal}) = \prod_{k=1}^{n} P(x_k | \text{legal})$$

The probabilities $P(x_1 | \text{fraud})$, $P(x_2 | \text{fraud})$ can be estimated from the training samples:

$$P(x_k | \text{fraud}) = \frac{s_{ik}}{s_i}$$

Here, $s_i$ is the number of training examples for class fraud and $s_{ik}$ is the number of training examples of class with value $x_k$ for $A_k$.

### 3.1.1 Application

We present Bayesian learning algorithm to predict occurrence of fraud. Using the “Output” classification results for Table 1, there are 17 tuples classified as legal, and 3 as fraud. To facilitate classification, we divide the age of driver attribute into ranges:

<table>
<thead>
<tr>
<th>Instance</th>
<th>Name</th>
<th>Gender</th>
<th>Age_driver</th>
<th>fault</th>
<th>Driver_rating</th>
<th>Vehicle_age</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>David Okere</td>
<td>M</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>legal</td>
</tr>
<tr>
<td>2</td>
<td>Beau Jackson</td>
<td>M</td>
<td>32</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>fraud</td>
</tr>
<tr>
<td>3</td>
<td>Jeremy Dejean</td>
<td>M</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>legal</td>
</tr>
<tr>
<td>4</td>
<td>Robert Howard</td>
<td>M</td>
<td>35</td>
<td>1</td>
<td>0.33</td>
<td>1</td>
<td>legal</td>
</tr>
<tr>
<td>5</td>
<td>Crystal Smith</td>
<td>F</td>
<td>22</td>
<td>1</td>
<td>0.66</td>
<td>8</td>
<td>legal</td>
</tr>
<tr>
<td>6</td>
<td>Chibuike Penson</td>
<td>M</td>
<td>36</td>
<td>0</td>
<td>0.66</td>
<td>6</td>
<td>legal</td>
</tr>
<tr>
<td>7</td>
<td>Collin Pyle</td>
<td>M</td>
<td>42</td>
<td>1</td>
<td>0.33</td>
<td>3</td>
<td>legal</td>
</tr>
<tr>
<td>8</td>
<td>Eric Penson</td>
<td>M</td>
<td>39</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>fraud</td>
</tr>
<tr>
<td>9</td>
<td>Kristina Green</td>
<td>F</td>
<td>29</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>legal</td>
</tr>
<tr>
<td>10</td>
<td>Jerry Smith</td>
<td>M</td>
<td>33</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>legal</td>
</tr>
<tr>
<td>11</td>
<td>Maggie Frazier</td>
<td>F</td>
<td>42</td>
<td>1</td>
<td>0.66</td>
<td>3</td>
<td>legal</td>
</tr>
</tbody>
</table>
Table 2 shows the counts and subsequent probabilities associated with the attributes. With these simulated training data, we estimate the prior probabilities:

The classifier has to predict the class of instance to be fraud or legal.

\[ P(\text{fraud}) = \frac{s_i}{s} = \frac{3}{20} = 0.15 \]
\[ P(\text{legal}) = \frac{s_i}{s} = \frac{17}{20} = 0.85 \]

Table 2: Probabilities associated with attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
<th>Count</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>M</td>
<td>13</td>
<td>13/17, 3/3</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>4</td>
<td>4/17, 0/3</td>
</tr>
<tr>
<td>age_driver</td>
<td>(20, 25)</td>
<td>3</td>
<td>3/18, 0</td>
</tr>
<tr>
<td></td>
<td>(25, 30)</td>
<td>4</td>
<td>4/18, 0</td>
</tr>
<tr>
<td></td>
<td>(30, 35)</td>
<td>3</td>
<td>3/18, 1/2</td>
</tr>
<tr>
<td></td>
<td>(35, 40)</td>
<td>3</td>
<td>3/18, 1/2</td>
</tr>
<tr>
<td></td>
<td>(40, 45)</td>
<td>3</td>
<td>3/18, 0</td>
</tr>
<tr>
<td></td>
<td>(45, 50)</td>
<td>2</td>
<td>2/18, 0</td>
</tr>
<tr>
<td>fault</td>
<td>0</td>
<td>5</td>
<td>5/17, 0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>12</td>
<td>12/17, 3/17</td>
</tr>
<tr>
<td>driver_rating</td>
<td>0</td>
<td>6</td>
<td>6/17, 1/3</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>5</td>
<td>5/17, 0</td>
</tr>
<tr>
<td></td>
<td>0.66</td>
<td>3</td>
<td>3/17, 0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>3/17, 2/3</td>
</tr>
</tbody>
</table>

We use these values to classify a new tuple. Suppose, we wish to classify \( X = (\text{Crystal Smith}, \text{F}, 31) \). By using these values and the associated probabilities of gender and driver age, we obtain the following estimates:

\[ P(X | \text{legal}) = \frac{4}{17} \times \frac{3}{18} = 0.039 \]
\[ P(X | \text{fraud}) = \frac{3}{3} \times \frac{1}{2} = 0.500 \]
Thus, likelihood of being legal = 0.039 * 0.9 = 0.0351

Likelihood of being fraud = 0.500 * 0.1 = 0.050

We estimate $P(X)$ by summing up these individuals likelihood values since $X$ will be either legal of fraud:

$$P(X) = 0.0351 + 0.050 = 0.0851$$

Finally, we obtain the actual probabilities of each event:

$$P(\text{legal} \mid X) = \frac{(0.039 \times 0.9)}{0.0851} = 0.412$$
$$P(\text{fraud} \mid X) = \frac{(0.500 \times 0.1)}{0.0851} = 0.588$$

Therefore, based on these probabilities, we classify the new tuple as fraud because it has the highest probability.

Since attributes are treated as independent, the addition of redundant ones reduces its predictive power. To relax this conditional independence is to add derived attributes which are created from combinations of existing attributes.

Missing data cause problems during classification process. Naïve Bayes classifier can handle missing values in training datasets. To demonstrate this, seven missing values appear in dataset. The naïve Bayes approach is easy to use and only one scan of the training data is required. The approach can handle missing values by simply omitting that probability when calculating the likelihoods of membership in each class. Although the approach is straightforward, it does not always yield satisfactory results. The attributes usually are not independent. We could use subset of the attributes by ignoring any that are dependent on others. The technique does not handle continuous data. Dividing the continuous values into ranges could be used to solve this problem, but the division of the continuous values is a tedious task, and how this is done can impact the results.

### 3.2 DECISION TREE-BASED ALGORITHM

A decision tree (DT) is a tree associated with a data base that has each internal node labeled with an attribute, each arc labeled with a predicate that can be applied to the attribute, and each leaf node labeled with a class. Solving the classification problem is a two-step process: i) decision tree induction- construct a DT, and ii) apply the DT to determine its class. Rules can be generated that are easy to interpret. They scale well for large databases because the tree size is independent of the database size.

DT algorithms do not easily handle continuous data. The attribute domains must be divided into categories. Handling missing is difficult. Since the DT is constructed from the training data, overfitting may occur. This can be overcome via tree pruning.

#### 3.2.1 C4.5 Algorithm

The basic algorithm for decision tree is as follows:

1. Suppose there are two classes for fraud and legal. The tree starts as a single node $N$ representing the training samples.
ii) If the samples are of the same class *fraud*, then the node becomes a leaf and is labeled as *fraud*.

iii) Otherwise, the algorithm uses an entropy-based measure as a heuristic for selecting the attribute that will best separate the samples into individual classes.

The entropy, or expected information needed to classify a given sample is:

\[
I(fraud, legal)= - \left( \frac{\text{NumberFraudSamples}}{\text{NumberSamples}} \right) \log_2 \left( \frac{\text{NumberFraudSamples}}{\text{NumberSamples}} \right) - \left( \frac{\text{NumberLegalSamples}}{\text{NumberSamples}} \right) \log_2 \left( \frac{\text{NumberLegalSamples}}{\text{NumberOfSamples}} \right)
\]

iv) Expected information or entropy required to classify into subsets by test attribute E is:

\[
E(A) = \sum \left[ \left( \frac{\text{NumberTestAttributeFraudValues}}{\text{NumberSamples}} \right) + \left( \frac{\text{NumberTestAttributeLegalValues}}{\text{NumberSamples}} \right) \right] I(\text{TestAttributeFraudValues, TestAttributeLegalValues})
\]

v) Expected reduction in entropy is:

\[
\text{Gain}(A) = I - E(A)
\]

The algorithm computes the information gain of each attribute. The attribute with highest information gain is the one selected for test attribute.

vi) A branch is created for each known value of the test attribute. The algorithm uses the same process iteratively to form a decision tree at each partition. Once an attribute has occurred at a node, it need not be considered in any of the node’s descendents.

The iterative partitioning stops only when one of the conditions is true: a) all examples for a given node belong to the same class, or b) there are no remaining attributes on which samples may be further partitioned. If this is the case, a leaf is created with the class in majority among samples, c) there are no samples for the branch test-attribute. In this case, a leaf is created with the majority class in samples

### 3.3 Rule Based Algorithm

One way to perform classification is to generate if-then rules. There are algorithms that generate rules from trees as well as algorithms that generate rules without first creating DTs.

#### 3.3.1 Generating Rules from a Decision Tree

The following rules are generated for the Decision Tree (DT).

If driver age =25, then class = legal
If (driver_age =40) \land (vehicle_age =7), then class = legal
If (driver_age =32) \land (driver_rating =1), then class = fraud
If (driver_age \leq 40) \land (driver_rating =1) \land (vehicle_age =2), then class = fraud
If (driver_age > 40) \land (driver_age \leq 50) \land (driver_rating =0.33), then class = legal
4. MODEL PERFORMANCE

4.1 Confusion Matrix

There are two ways to examine the performance of classifiers: i) confusion matrix, and ii) to use a ROC graph. Given a class, C_j, and a tuple, t_i, that tuple may or may not be assigned to that class while its actual membership may or may not be in that class. With two classes, there are four possible outcomes with the classification as: i) true positives (hits), ii) false positives (false alarms), iii) true negatives (correct rejections), and iv) false negatives (ALGORITHM positive and true negative represent correct actions. False positive occurs if the actual outcome is legal but incorrectly predicted as fraud. False negative occurs when the actual outcome is fraud but incorrectly predicted as legal. A confusion matrix (Kohavi and Provost, 1998), Table 3a, contains information about actual and predicted classifications. Performance is evaluated using the data in the matrix. Table 3b shows confusion matrix built on simulated data. It shows the classification model being applied to the test data that consists of 7000 instances roughly split evenly between two classes. The model commits some errors and has an accuracy of 78%. We also applied the model to the same data but to the negative class with respect to class skew in the data. The quality of a model highly depends on the choice of the test data. We also that that ROC curves are not so dependent on the choice of test data, at least with class skew.

Table 3a: Confusion Matrix

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>legal</td>
<td>fraud</td>
</tr>
<tr>
<td>legal</td>
<td>TP</td>
</tr>
<tr>
<td>fraud</td>
<td>FN</td>
</tr>
</tbody>
</table>

Table 3b: Confusion matrix of a model applied to test dataset

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Accuracy: 0.78</th>
<th>recall: 0.86</th>
<th>precision: 0.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>legal</td>
<td>3100</td>
<td>1125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fraud</td>
<td>395</td>
<td>2380</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A number of model performance metrics (Table 3c) can be derived from the confusion matrix.

Table 3c: Performance metrics

<table>
<thead>
<tr>
<th>Model performance metrics</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (AC)</td>
<td>$AC = \frac{a + d}{a + b + c + d}$</td>
</tr>
<tr>
<td>Recall or true positive rate (TP)</td>
<td>$TP = \frac{d}{c + d}$</td>
</tr>
<tr>
<td>False positive rate (FP)</td>
<td>$FP = \frac{b}{a + b}$</td>
</tr>
<tr>
<td>True negative rate (TN)</td>
<td>$TN = \frac{a}{a + b}$</td>
</tr>
<tr>
<td>False negative rate (FN)</td>
<td>$FN = \frac{c}{c + d}$</td>
</tr>
</tbody>
</table>
The accuracy determined above (Table 3b) may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases (Kubat et al., 1998). Suppose there are 1500 cases, 1460 of which are negative cases and 40 of which are positive cases. If the system classifies them all as negative, the accuracy would be 97.3%, even though the classifier missed all positive cases. Other performance measures are geometric mean (g-mean) (Kubat et al., 1998), and F-Measure (Lewis and Gale, 1994). For calculating F-measure, $\beta$ has a value from 0 to $\infty$ and is used to control the weight assigned to $TP$ and $P$. Any classifier evaluated using g-mean or F-measure will have a value of 0, if all positive cases are classified incorrectly.

To easily view and understand the output, visualization of the results is helpful.

Naïve Bayesian visualization provides an interactive view of the prediction results. The attributes can be sorted by the predictor and evidence items can be sorted by the number of items in its storage bin. Attribute column graphs help to find the significant attributes in neural networks. Decision tree visualization builds trees by splitting attributes from C4.5 classifiers.

Cumulative gains and lift charts are visual aids for measuring model performance. Lift is a measure of a predictive model calculated as the ratio between the results obtained with or without the predictive model. For instance, if 105 of all samples are actually fraud and a naïve Bayesian classifier could correctly predict 20 fraud samples per 100 samples, then that corresponds to a lift of 4.

<table>
<thead>
<tr>
<th>Table 4: Costs of Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>fraud</strong></td>
</tr>
<tr>
<td>True Positive(Hit) cost = number of hits * average cost per investigation</td>
</tr>
<tr>
<td>False Negative(miss) cost = number of misses * average cost per claim</td>
</tr>
</tbody>
</table>

Table 4 shows that True Positives (hits) and False Positives (false alarms) require cost per investigation. False alarms cost are the most expensive because both investigation and claim costs are required. False Negatives (misses) and True Negatives (correct rejection) are the cost of claim.

### 4.2 Relative Operating Characteristic Curve

Another way to examine the performance of classifiers is to use a Relative Operating Characteristic (ROC) curve, (Swets, 1988). A ROC graph is a curve that depicts the performance and performance tradeoff of a classification model (Fawcett, 2004, Flach, 2004) with the False Positives along X-axis and the True Positives along the Y axis. The point (0, 1) is the perfect classifier: it classifies all positive cases and negative cases correctly. It is (0, 1) because the false positive FP is 0, and the TP rate is 1. The point (0, 0) represents a classifier that predicts all cases to be negative, while the point (1, 1) corresponds to a classifier that predicts every case to be positive. Point (1, 0) is the classifier that is incorrect for all classifications. An ROC curve or point is independent of class distribution or error
costs (Provost et al., 1998). It sums all information contained in the confusion matrix, since FN is the complement of TP and TN is the complement of FP (Swets, 1988). It provides a visual tool for examining the exchange between a classifier to correctly identify positive cases and the number of negative cases incorrectly classified.

We introduce new performance metrics to construct ROC curves (in confusion matrix terms), the TP Rate (TPR) and the FP Rate (FPR):

\[
TPR(\text{recall}) = \frac{TP}{TP+FN} \\
FPR = \frac{FP}{TN + FP}
\]

The classifier is mapped to the same point in the ROC graph regardless of whether the original test set with sampled down negative class is used illustrating that ROC graphs are not sensitive to class skew.

One way of comparing ROC points is by using an equation that equates accuracy with the Euclidian distance from the perfect classifier, the point (0, 1). We include a weight factor that allows defining relative misclassification costs. We define \( AC_d \) as a distance based performance measure:

\[
AC_d = 1 - \sqrt{W \cdot (1 - TP)^2 + (1 - W) \cdot FP^2}, \text{ where } W \text{ ranges from 0 to 1, that is used to assign relative importance to false positives and false negatives. } AC_d \text{ ranges from 0 for the perfect classifier to } \sqrt{2} \text{ for a classifier that classifies all cases incorrectly. } AC_d \text{ differs from } g\text{-mean}_1, g\text{-mean}_2 \text{ and } F\text{-measure in that it is equal to 0 only if all cases are classified correctly. In other words, a classifier evaluated using } AC_d \text{ gets some credit for correct classification of negative cases, regardless of its accuracy in correctly identifying positive cases.}
\]

4. CONCLUSIONS

We studied the existing fraud detection systems. To predict and present fraud we used Naïve Bayesian classifier. We looked at model performance metrics derived from the confusion matrix. We illustrated how ROC curves can be deployed for model assessment. Performance metrics such as accuracy, recall, and precision are derived from the confusion matrix. ROC analysis provides a highly visual account of a model’s performance. It is strong with respect to class skew, making it a reliable performance metric in many important fraud detection application areas.
REFERENCES


Williams, G. J. and Huang, Z.(1997). ‘Mining the Knowledge: Mine the Hot Spots Methodology for Mining Large Real World Databases’, 10th Australian Joint Conference on Artificial Intelligence, Published in Lecture Notes in Artificial Intelligence, Springer-Verlag, December, Perth, Western Australia.
The Cyber-Workplace – Identifying Liability Issues in the Information Age and Managing E-Risk

Nigel Wilson
Barrister
Bar Chambers
34 Carrington Street
Adelaide South Australia 5000
Australia
nigel.wilson@barchambers.com.au

ABSTRACT

The information age provides numerous opportunities for modern society but also presents significant challenges in identifying liability issues and in managing risk. Technological change has occurred rapidly and is continuing at the same time as other major trends and changes are taking place in society and, in particular, in the workplace. The prospect of global liability and the complexity of jurisdictional differences present a considerable hurdle to the uniform regulation of liability issues. General legislation and legal principles have been readily applied to the cyber-world and to modern business practices and the workplace. Where necessary, legislatures have introduced specific legislation to regulate unfair or inappropriate business and workplace practices which involve the use of technology in an unsuitable manner. Consistent with international objectives, a central element of the regulation of e-commerce and the cyber-space economy will be the protection of individual human rights, particularly the right to privacy. Human rights concepts raise special challenges in relation to the regulation of the modern, technology-intensive economy and workplace. Appropriate data protection and document retention policies will be a critical component of effective risk management and compliance programs in the information age.

1. THE CYBER-WORKPLACE

A vast amount of work has been done and time and resources have been spent by International agencies and national governments in examining the likely impact of Information and Communication Technologies (ICTs) on the modern workplace and the human condition.[1]

A key feature of the stated international position regarding the regulation of cyber-space and ICTs is the necessity for the protection of fundamental human rights, particularly the right to privacy.

1.1 The International Position

Historically, periods of major international conflict have been marked by ICT development and growth either at the time or shortly thereafter. In the post-World War II it was resolved by Article 12 of the Universal Declaration of Human Rights (1948) that:

“No-one shall be subjected to arbitrary interference with his/her privacy, family, home or correspondence, nor to attacks upon his/her honour or reputation. Everyone has the right to protection of the law against such interference or attacks.”

Article 17 of the International Covenant on Civil and Political Rights (1966) is to similar effect.

In 1997 then United States President Bill Clinton stated a framework for global electronic commerce in which five core principles were identified:

- the private sector should lead;
- governments should avoid undue restrictions on electronic commerce;
• where government involvement is needed, its aim should be to support and enforce a predictable, minimalistic, consistent and simple legal environment for commerce;
• governments should recognize the unique qualities of the internet;
• electronic commerce over the internet should be facilitated on a global basis.[2]

More recently, the United Nations General Assembly resolved in the Millennium Declaration in 2000:

“To ensure that the benefits of new technology, in conformity with the recommendations contained in the Economic and Social Council (ECOSOC) 2000 Ministerial Declaration, are available to all.”

The ECOSOC 2000 Ministerial Declaration stated that ICTs:

• are central to the emerging global knowledge-based economy;
• can accelerate growth;
• can promote sustainable development;
• assist in eradicating poverty in developing countries and countries in transition.

However, the following key issues and concerns were identified:

• the “new economy” creates opportunities for economic growth and social development;
• the majority of the world population still lives in poverty and remains untouched by the information and communication technology (ICT) revolution;
• there was a potential for economic development by developing countries to close the “digital divide” and in so doing ICTs should be utilized to foster “digital opportunity”.

Subsequently, the United Nations General Assembly resolved that legal systems should:

• protect the confidentiality, integrity and the availability of data and computer systems from unauthorized impairment;
• ensure that criminal abuse is penalized.[3]

1.2 Regulating the Cyber-Workplace at the National Level

In both common law and civil law jurisdictions around the world existing general laws have been readily applied to the modern economy and the cyber-space environment. When necessary, specific legislation has been enacted to prohibit or regulate practices which are not in the public interest. Some of this legislation reflects general human rights principles and, in effect, prohibits or regulates behaviour or conduct which constitutes an inappropriate or unwelcome interference with an individual’s privacy, family, home or correspondence.

2. THE IMPACT OF TECHNOLOGY ON MODERN SOCIETY IN PERSPECTIVE

The pace of technological change and its increasingly ready acceptance can be demonstrated by analyzing, specifically, the short history but immediate impact of telephones, computers and the
internet on modern society. An appreciation of the impact of the ICT revolution to date can assist in planning for and coping with future developments. The identifiable trends to date include: constant innovation, speed of growth and change, increasing (often exponential) scale of activity and production, high risk but possible high return, global impact and the benefit of predictable, uniform regulatory frameworks.

2.1 Telephone

The telephone was invented in 1876 by Alexander Graham Bell. By 1880 the Bell Company had leased only 100,000 instruments.

By contrast, in 2007 Apple’s new “Iphone” is estimated to have sold between 500,000 to 700,000 units in the first weekend of its sales. Each phone retailed for approximately US$499 to US$599. Accordingly, approximately US$250 million in sales are estimated to have occurred in one weekend alone.

Trends in the relative cost of telephone usage also demonstrate the vast economies of scale in the international telecommunications system. The cost of a telephone call from New York to London was approximately a dollar in 1950, six cents in 1990 and is essentially “free” today using the internet.

2.2 Computers

The first rotor machines were the subject of the Enigma patent in 1918.

During World War II electro-mechanical “bombers” were developed together with the top secret Colossus computer. The Electronic Numerical Integrator and Computer was developed between 1943 and 1946.

By 1965 Intel founder, Mr Graham Moore, described what became known subsequently as “Moore’s law”: that the number of transistors on a computer chip doubles every two years. As a result, a musical birthday card bought today has more computing power than the fastest main frame computers of the 1970s.

2.3 The Internet

The internet was invented in 1969 and used predominantly for email and file transfers. The HTTP (Hypertext Transfer Protocol) and HTML (Hypertext Markup Language) protocols were developed in 1989. Business to consumer (B2C) and business to business (B2B) data exchange, communication and commerce has spawned as a result.

In March 2000 the “dot.com bubble” burst. However, the rate of internet usage is burgeoning. In 2006 58% of Australian households had an internet connection.[4] In 2007 it is reported that nearly a billion people use digital technology in their daily lives. Further, despite “the notorious dotcom collapses, estimates show that worldwide online trade exceeded US $2000 billion in 2002 with predicted increases in excess of US $12,800 billion by 2006: the European Union alone is expected to experience on-line trade rising from €77billion in 2001 to €2.2trillion by 2006”.[5]

3. OTHER MAJOR TRENDS IN THE MODERN WORKPLACE

The technology changes occurring in the cyber-workplace and society are also occurring at the same time as a number of other significant changes.

Major studies have identified the following trends:

- a shifting workforce composition including older workforce and an ageing population together with an increasingly female participation in the workforce;
- an increasingly skilled workforce with emphasis on “knowledge” based industries;
organisational changes in which firms are becoming more specialized and are increasingly vertically disintegrated;
• the nature of the employment environment has changed from the traditional employer-employee relationship towards an increasing use of independent contractors, temporary workforce and, in some industries, “e-lancing”;
• work locations now include temporary locations and “remote” workplaces;
• workplace education and training now includes ICT-based training.[6]

These trends must also be borne in mind in seeking to understand the way in which modern business is conducted and to regulate its activities.

4. LIABILITY ISSUES IN THE MODERN WORKPLACE

Until recently, domestic regulation of modern business activities has not emphasized individual human rights. The common law has been reluctant to protect an individual’s right to privacy.[7] However, an increasing number of jurisdictions are adopting international principles of human rights into domestic law.

The liability and regulatory issues for the information age include:

• global liability issues;
• jurisdiction – based issues;
• risk issues;
• data and document retention issues;
• human rights issues.

4.1 Global Liability Issues

Globalisation of commerce and trade gives rise to a potential liability in every jurisdiction in which a website is viewed or an email is published.[8] Provided the jurisdictional basis exists, current consumer protection legislation has the capacity to apply extra-territorially, for example, to misleading advertising on the internet[9] and to the operation of websites outside a country’s jurisdiction engaging in inappropriate business practices.[10] Courts have recognized the need for international co-operation in meeting the needs of consumers in the internet world[11] as well as the need to regulate companies located within a jurisdiction but operating outside that jurisdiction.[12]

4.2 Jurisdiction – based issues

Modern commerce is being conducted on an international basis. The cyber-space environment raises issues regarding the location of the “worker” and the data or transactions in which they are involved. Each of these may differ from the location of the employer. Further, “home” offices and ICTs may contain important information which may be owned by the employer or others. The “home” office may not satisfy specific occupational health and safety regulations which may apply in the employer’s office or traditional workplace. Further, insurance policies which may apply to e-risk events arising from economic activity are usually jurisdiction-specific and contain United States exclusions. The internet is often described as “borderless”.

4.3 Risk Issues

Risk issues for the modern economy include viruses damaging own systems and being forwarded to third parties. Third parties (hackers etc) have the capacity to damage systems through unauthorized access and sabotage. Data protection of confidential information will be paramount. The detection of fraud and other criminal practices will be a key consideration.[13] The protection of intellectual property is the subject of considerable international regulation and comity but the relative ease with
which technological innovation can be reproduced or reverse-engineered and the relatively short operational life of new technologies mean that enforcement is often not effective or timely.[14]

If an “e-risk” event occurs within an organization the possible financial consequences include trading losses, business interruption, personnel downtime, data retrieval costs, reputation loss and restoration or remedial costs. The organization the subject of such an event may itself be responsible to other parties (eg customers or clients for privacy intrusions or suppliers to whom duties of care or contractual obligations are owed).

### 4.4 Data and Document Retention Issues

The “paperless office” has become an expression which has not been reflected in reality. Innovation in rights management of data and documentation and the ability of software to control the recipient of a document and how long it is accessible[15] gives rise to issues regarding data and document retention. In subsequent litigation, the failure to establish suitable policy and system control procedures, including control of access to relevant databases, programs, logging of changes, backup practices and audit procedures, can give rise to documents being rendered inadmissible.[16]

### 4.5 Human rights issues

Common law protection of a right to privacy has been inconsistent.

General legislation protecting privacy has the capacity to regulate breaches of privacy principles.[17] The central concept in the protection of privacy is the notion of personal information which is information or an opinion which identifies an individual or allows their identity to be readily worked out from the information. In the event of failure to comply with the principles then, for example in Australia, the Privacy Commissioner has the power to investigate a complaint or investigate on the Commissioner’s own initiative an act or practice which may be a breach of privacy (even if no complaint is made) and seek an order (injunction) from the court to stop conduct that does or would breach the privacy principles. For example, inadvertent disclosure of customer email addresses has been sanctioned.[18]

In addition, some jurisdictions have recently enacted human rights legislation which is reflective of the international charter of human rights in which the right of a person not to have his or her privacy, family, home or correspondence unlawfully or arbitrarily interfered with is protected.[19] For example, specific legislation has been enacted to prohibit:

- “spam” making it illegal to send or cause to be sent “unsolicited commercial electronic messages”; [20]
- unsolicited telemarketing calls making it illegal to make unsolicited telemarketing calls to numbers listed on the register.[21]

The adoption of broad human rights principles raises complications for regulation in the information age:

- some jurisdictions have ratified international human rights conventions but have not legislated for their application domestically;[22]
- the expense and delay involved in the enforcement of human rights principles;
- the perception that human rights principles involve public law concepts (eg. judicial review) rather than private law rights and remedies including rights to compensation;
- the interpretation and enforcement of human rights principles has been far from predictable, simple and consistent.
5. RISK MANAGEMENT

As with all risk management, the key elements for risk management of liability issues in the cyber-workplace will include:

- appropriate training and supervision;
- assessment of the threat, system characteristics and the physical and cyber environments in which those systems operate in a documented and comprehensive manner;[23]
- effective protocols and compliance. Specifically, in relation to modern technology these include:
  - closed networks;
  - intranets;
  - firewalls;
  - anti-virus protection;
  - digital signatures; and
  - encryption security
- maintenance procedures and systems, including for managing and dealing with security breaches.

The inter-relationship in modern society between critical infrastructures (electric power, gas supply, water supply and waste treatment, rail transport and ICTs) has been described as “mutually and circularly dependent”. The International Risk Governance Council has concluded that “…our societies are most vulnerable to disruptions of electric power supply and disruptions to, or degradation of, ICT services”. It was their judgment that “a significant problem for owners, managers and regulators is that the public and many officials in government have limited knowledge of the vulnerabilities of these systems and of the risk factors that have increased during the past several decades.”[24] The challenge for individuals, businesses and governments will be to identify relevant risks and to put in place appropriate risk management strategies or policy frameworks.

6. CONCLUSION

Cyber-space has the very real prospect of leading to a digital divide between nations and the people within them rather than fostering digital opportunity. The identification and regulation of liability issues will be a key component of the equitable allocation of ICTs worldwide. A fundamental factor in the successful achievement of such a worthwhile goal will be an awareness of the relevance, and consistent application, of human rights principles to an area which has historically been marked by a “survival of the fittest” and a “first to market” mentality.

What cannot be overlooked is that human rights “should be seen as informing almost everything lawyers and courts do”.[25] This includes the regulation of the modern business and work-place environment both now and in the future.

One individual whose corporation has so revolutionized the modern economy and has been a driving and dominant force in the information age has said:

“During the last decade, digital technology has changed the world in profound and exciting ways. Today we communicate instantly with people we care about without worrying about the traditional limitations of time and location. At work, we collaborate with colleagues in distant cities ... But these changes are just the beginning.”[24]
If the current stage of ICT development is in its infancy then the challenge to society and the legal environment of regulation, liability allocation and risk management will be to strike a balance between innovation and competition and the protection of fundamental human rights in the modern economy and the cyber work-place.

NOTES

This paper is a revised and updated version of two papers presented by the author in December 2007 and January 2008.

The first paper was presented in December 2007 at the Second International Conference on Legal, Security and Privacy Issues in Information Technology (LSPI) which was held in Beijing, China. The LSPI paper was entitled “The Workplace of the Future – Liability Issues and Risk Management” and was published in the conference proceedings book “Cyberlaw, Security and Privacy” (Edited by S. Kierkegaard), 2007. The LSPI paper is to be published in 2008 in the International Journal of Liability and Scientific Enquiry.

The second paper was presented in January 2008 at the E-Forensics 2008 Conference (The First International Conference on Forensic Applications and Techniques in Telecommunications, Information and Multimedia) at the University of Adelaide in Adelaide, South Australia.

REFERENCES

[7] A common law right to privacy was left open by the High Court of Australia in Australian Broadcasting Corporation v Lenah Game Meats Pty Ltd (2001) 208 CLR 199. Subsequent cases in the State Supreme Courts of Australia have in one instance upheld a right to privacy (Grosse v Purvis [2003] QDC 151) and in another case in a different State of Australia not found a right to privacy (Giller v Procopets [2004] VSC 113).
[13] For example, an employer was found vicariously liable for the fraud of its employee who accessed and transferred monies from the bank account of a person for whom she was responsible for caring: Ffrench v Sestilli [2007] SASC 241.


[21] See the *Do Not Call Register Act 2006* (Commonwealth of Australia).


ABSTRACT

The ubiquity of small scale digital devices (SSDD), the public’s ever increasing societal dependence on SSDD, and the continual presence of SSDD at all types of crime scenes, including non-technical and violent crimes, demand a formalized curriculum for the education and training of future cyber forensic examiners. This paper presents the various SSDD forensics labs currently in use and under development for future use at the Purdue University Cyber Forensics Laboratory. The primary objective of each module is to provide specific real-world cases for the learning, comprehension, and understanding of hands-on investigative techniques and methodologies. The purpose of this paper is to outline those elements that will make effective Small Scale Digital Device Forensics labs.

Keywords: Forensics, Cyber Forensics, Digital Evidence, Education, Training, Small Scale Digital Devices, Personal Digital Technologies

1.0 INTRODUCTION

As Small Scale Digital Devices (SSDD) have become so pervasive in the daily social fabric of our lives, they have also presented a wealth of information for our forensic examiners. As personal storage units of all that is important to our social lives, these devices tell of who we know, who we communicate with, and what we find worth capturing.

As treasure troves, these SSDD are not easy “nuts to crack.” Over the past ten years, the world of computer forensics has had the luxury of a limited handful of file system formats, being ensconced in forensic standards and systems developed to acquire and analyze these evidentiary hard drives. As SSDD forensics is still in its infancy, forensic examiners are still debating the standards and methods for acquiring these idiosyncratic devices (Ayers et al., 2007; SWGDE, 2007; Ayers et al., 2006; Gratzner, Naccache, Znaty, 2006; InterPol, 2006; McCarthy, 2005; ACPO/NHTCU, n.d.; IOCE, 2000).

2.0 SMALL SCALE DIGITAL DEVICES

By definition a Small Scale Digital Device is “a small form factor device which utilizes permanent or temporary memory in conjunction with embedded chips to perform a variety of tasks” (Harrill and Mislans, 2007). Categorically, SSDD have been further defined by the various types of devices:

- Personal Digital Assistants
Cellular Telephones
Audio / Video Devices
Gaming Devices
Other Devices

As the name suggests, SSDD are characterized by their physical size which appears to be diminishing over time. As these devices grow smaller, and more personal, they have extended functionalities storing massive amounts of information. Furthermore, many of these SSDD provide the means of synchronizing information to another computer, whether through wired or wireless networks such as WiFi or Bluetooth.

2.1 Personal Digital Assistants

In 1992, Jeff Hawkins spawned the beginning of the Personal Digital Assistant rage that ran well through the last part of the century. Soon after, names like Newton, Nino, Palm, iPaq, and Zaurus became handheld household names. Providing Personal Information Management (PIM) applications like calendars, contacts, task lists, and memo pads, these devices became the device du jour for the business world.

2.2 Mobile Telephones

In 1973, when Martin Cooper placed the first call on his “Brick Phone”, a revolution started that has forever changed the social fabric of our daily lives. Mobile phones of today are so ubiquitous (Jansen, Ayers, 2006), that rather than just placing calls, these devices provide technologies for Short Message Service (SMS) messaging, Multi-Media Messaging Service (MMS) messaging, Instant Messaging (IM), electronic mail, Web browsing, multimedia capturing and playback, electronic document previewing, basic Personal Information Management (PIM) applications (e.g., contacts, calendar, etc.) and financial transactions (Willassen, 2003).

2.3 Audio/Video Devices

In 2001, Apple Computer started a revolution changing the face of the music industry overnight. With the birth of the iPod, a whole world of music lovers started “ripping” their CD collections to store on these portable music devices. Not long after the birth of the iPod, and many years after the beta vs. VHS wars, came the world of portable digital video players. Digital video players have become so prevalent thanks to products like Apple's Video iPod that it is not uncommon to carry entire seasons of syndicated television in one’s pocket or purse. Unfortunately, today’s criminals have also found that these large portable hard drives can store much more information than just music (Marsico and Rogers, 2005).

Finally, an overarching characteristic of SSDD is the variety of the different manufacturers, service providers, operating systems, technologies, form factors, and data and power cables. The combinatorial explosions that culminate from these various categories and characteristics of SSDD lend to a wide variety of physical and virtual possibilities (Gratzner, Naccache, Znaty, 2006).

3.0 FORENSICS OF SMALL SCALE DIGITAL DEVICES

In today’s wired society, many crime scenes are littered with some type of digital evidence (Robinson and Smith, 2001). More times than not, crimes scene evidence is found on a thumb drive, a cell phone, or another type of SSDD. Unfortunately, first responders don’t always realize the potential time-criticality and sensitivity of such digital evidence. With the ever-increasing storage capacitance and the ever-decreasing physical size of these devices, it is imperative to prepare our future forensic examiners and investigators with the tools and techniques for analyzing these evidentiary devices.

SSDD Forensics is the science of recovering digital evidence from SSDD under forensically sound
conditions using accepted methods (Harrill and Mislan, 2007). SSDD, especially those with advanced capabilities, are a relatively recent phenomenon, not usually covered in classical computer forensics (Jansen and Ayers, 2005). Thus, it has become a major initiative at the Purdue University Cyber Forensics Laboratory to create a SSDD Forensics course with representative labs.

As digital technology evolves, the existing capabilities of these devices continue to improve rapidly. When SSDD are involved in a crime or other incident, forensic examiners require tools that allow the proper retrieval and speedy examination of information present on the device (Jansen and Ayers, 2006). Unfortunately, these exist as a “Swiss Army knife” collection of tools from various manufacturers, open source groups, or underground, black hat, or hacker sources. As an endorsement of the “Swiss Army knife” approach, the Scientific Working Group on Digital Evidence (SWGDE, 2007) suggests the following procedure when examining Mobile Phones:

1. Use proven and validated hardware/software solutions. If the phone includes a SIM card, examine the card with and without the handset.
2. Use open source, free, or manufacturer-specific tools.
3. Use wireless transfer methods such as Bluetooth or Infrared.
4. Use the suspect device to display data while photographing or videotaping the screen.
5. Transcribe information viewed on the device to include call logs, phone books, text messages, etc.
6. Use the suspect device to E-Mail or forward the data to an examination device. In the event this method is used, the examiner must document why this method was used and the steps taken. The examiner must also ensure that the data received is an accurate depiction of what was on the suspect device.

As with cell phones, the multiple attempts of data acquisition work well with other types of SSDD. These multiple attempts are directly related to the multiple tools that the students are exposed to: DirSnoop, Access Data FTK, GSM .XRY, Cellebrite UME36, Secure View, Device Seizure, SIMCon, SIMIS, iDEN Companion Pro, iDEN Phonebook Manager, iDEN Media Downloader, Blackberry Desktop Manager, Amber Blackberry Converter, Oxygen Phone Manager, MobilEdit!, TULP2G, Nokia PC Suite, Sony Ericson PC Suite, various Flasher Boxes, several of our own developed tools, and various other data synchronization tools.

The main reason so many tools are necessary for SSDD forensics is that combinatorial explosion mentioned earlier. A good example of this is the Motorola RAZR cell phone. To date there are over twelve versions of this single phone, ranging from the RAZR V3 to the V9, each with its own technological modifications. Beyond the minor form factor variations, the manufacturer, Motorola, has recently changed the data/power connection from Mini USB to the newer Micro USB standard (Open Mobile Terminal Platform, 2007). To add to this, each network service provider adjusts the operating system or application software to their specific needs. Beyond this, a user may implement changes specific to their personal preferences. With so many variables, it makes this field of forensics ever-so dynamic, with a challenge at every turn (McCarthy, 2005; Robinson and Smith, 2001).

3.1 Lab Modules

To prepare our students for our reality and their future, we have designed real-world forensic labs based on cases taken from the headlines or various examiners experiences. The details are well developed and the actual devices, tools, and techniques are used throughout the forensic acquisition, preservation, analysis, and presentation of each case.

Currently, we are working with five labs: a murder, a kidnapping, a meth lab, corporate IP theft, and military intelligence gathering. Each lab is unique including at least ten different individuals, background case information, details about the suspects and victims, and at least five small scale
digital devices. Each of these devices includes evidence that can either help or hurt the investigation. In some instances, the evidence may point the students towards the Internet, leveraging information (either planted or previously existed) to further their investigations.

The goal for each of these labs is to take the students from cradle to grave of an actual investigation, learning how to process each type of SSDD evidence using the actual forensic tools and techniques that may or may not support these devices. This is not only about learning the proper techniques to handle, acquire, analyze, and present the evidence, but also how to determine if the newly found information is useful or not. The added elements of realism and surprise provide for an ever-engaging and thought-provoking environment.

In each lab, students are presented with multiple SSDD, all interacting with each other at various levels. One unique case is the Meth Lab Explosion, which revolves around the remnants of a Meth Lab, several badly burned victims, and several suspects who may or may not have been involved in the activities which led up to the explosion. Given this scenario, the devices found on the scene include the melted thumb drive, charred cell phone, and its SIM card. Other evidence collected from the suspects includes the Blackberry and two iDEN phones.

The SIM Card was found in the charred cell phone, which was still in the hand of a young college girl who died in the horrific explosion. Also found at the scene was a melted thumb drive in the pocket of another young college boy. With the SIM Card alone, the students are possibly able to determine who she knew, who she had been communicating with (SMS, MMS), and who she might have been called last. The thumb drive may have files related to the activities of the event or may be totally unrelated. The Blackberry and the iDEN phones are related to the accident, but are from suspects who escaped or were not present at the explosion. Ideally, these devices will also provide further information for the investigation or lead to dead ends.

**4.0 EDUCATIONAL CONSIDERATIONS & EXPECTATIONS: CONTROLS AND VALIDATION**

**4.1 Validation in the Absence of Hashing**

The nature of many of these devices often precludes the use of hashing as a validation method. There remains a need to establish reliable methods for verifying what data was placed on the device pre-seizure, and that (preferably) no data was placed on the device post-seizure or, if data was placed on the device (due to practical constraints) the report includes comprehensive documentation explaining why, and placing a verifiable boundary on what data could and could not have been altered. Specific methods will vary from device to device, but instilling the necessity into the minds of future examiners is an important element in the development of lab exercises.

**4.2 Use of Known Values**

“How do you know if you’ve found all the evidence?” “How do you know if everything you’ve found is evidence?” These are questions that have plagued analysts for years, particularly in the field of digital recovery. In real life, it is usually impossible to answer these questions with certainty. In a training environment, it is imperative that students receive timely feedback to determine that the procedure they used was successful or not successful. There is a strong risk that ineffective or, worse, partially effective classroom techniques will carry over into on the job practices down the road. The specifics vary widely, but generally take one of two forms:

(1) The instructor inserts known values (which are unknown to the students) into piece of media and the student has to apply learned techniques, and ideally, some creative problem solving to find these values. Generally, this is a device that is being treated and handled as simulated evidence. This enables the instructor to reliably measure the degree of success of the students' evidence handling and processing methodologies as well as their interpretation of results.
(2) The student takes predetermined steps using noted values and then uses learned techniques and tools to observe the actual results. Because this involves actual changes to the device, this form is limited to devices that are NOT being treated as simulated evidence in the lab. This has the advantage of allowing the student to gain a greater understanding of the "cause and effect" relationship between a variety of user actions, and the digital artifacts that are created as a result of those actions.

4.3 Investigative Roles and the Big Picture

The process of retrieving useful information from a device and placing it in evidentiary context involves multiple roles (Ayers, Jansen, Cilleros, and Daniellou, 2006). Over the course of a semester, students are rotated through as many of these roles as possible as each simulated case is processed. This not only gives the student the experience of that role, but in dealing with the interactions between the roles as well. Full implementation of all roles in every exercise will not always be practical or necessary, and the emphasis in most cases will be on the role of the examiner. The typical “flow” of the ideal process will follow the steps outlined in Figure 1 below:

Acquisition/Seizure
- Can be a physical or paper exercise
- Students must take steps to preserve volatile memory and ensure RF isolation
- Multiple devices may be seized for a single case

Transport
(In most labs exercises, this will be combined with the acquisition phase)
- Mostly a paper exercise
- Students must preserve chain of custody, as well as data integrity (RF isolation, volatile memory)

Receiving
- Student in this role must verify previous students’ work
- Optionally, instructor can also insert “mistakes” in required, prefabricated documentation (i.e. search warrant) that the student must find and correct.

Assignment
(In some scenarios, this step will be omitted for smaller, and single item "cases")
- Student in this role, must "assign" specific devices to students filling the "examiner/analyst" role.

Processing / Analysis
- Using approved tools and methods, the student must extract information from the device(s)
- The student must recognize information of relevant evidentiary value

Reporting
- A student must take the role of lead investigator
- Results of the analysis of multiple devices must be correlated into a single, coherent report.

4.4 Establishing Goals and Objectives

On the surface, the requirement is deceptively obvious. As discussed earlier, the evidence to be found consists of known values, evidence handling procedures are standardized, and must be adhered to. Furthermore, the end result is supposed to be the concrete identification of a single individual to be held culpable for the commission of a defined illegal act. It is tempting, therefore, to see the process as some kind of flowchart with correct and incorrect steps to follow, or as a series of yes/no questions, with corresponding right and wrong answers that can be objectively graded the same way every time. This in turn, can easily draw the instructor to treat the process as a demonstration of the “application” category of the cognitive domain (Anderson, et al, 2001) where the student need only follow the correct steps every time to achieve success. Giving into that temptation would leave the student woefully unprepared for “the real world” and ignores several factors critical to the successful handling, reporting and presentation of digital evidence.

It is important to remember that while certain steps and procedures must always be observed regardless of the circumstances, there remain many variables that render a pure “flow chart process”
approach to investigation generally ineffective. The type of crime or suspected activity varies from case to case. Technical considerations also vary widely from case to case, the number and type of evidence items, the services that the suspect(s) or other related parties may or may not have used, their relationship to alleged activities all vary widely. Active attempts by investigative subjects to remove potential evidence will affect not only the results of the process, but the actual steps of the process.

The students’ work product is, in fact, a synthesis, and can often require a great deal of creativity. Students must not only use a tool correctly, but must first evaluate a variety of factors in selecting the correct tool to use. Throughout the process, the students must repeatedly form hypotheses, and must then devise and perform tests for each (or correctly recognize when a hypothesis cannot be tested with available evidence or resources).

Even though the simulated evidence may have been prepared, and 100% of the evidence artifacts a known factor to the instructor, it is not always reasonable to use the number of specific items found and recorded as a standard to measure success. It is not generally feasible for every potential digital artifact to be manually evaluated, and more than one path may lead to a correct, verifiable conclusion. We cannot hold the student to a standard that would be unachievable in the field.

It remains that there must be a “right answer”, and it must be correct and objective. It must be clear when the student has succeeded, and just as clear when the student has not. There are objective elements to a successful examination. Success of the examination can be determined by looking at these elements in the context of the specific case scenario. These elements include:

- Was the case solved? Was a correct conclusion reached? To be considered successful the evidence and its analysis should clearly show that an identified individual is the guilty party to such a degree that no reasonable person would disagree with the conclusion.
  
  ***Note that this does not necessarily require that every digital artifact on a device be the subject of human analysis.

- Were forensically sound procedures followed throughout the process? Was any evidence altered or lost as a result of the evidence handling or processing?

- Were the hypotheses formed reasonable? Were the tests devised to test those hypotheses effective and correctly executed and interpreted?

- Did the student reach and state conclusions that are not supported by the evidence found (regardless of the factual correctness of the conclusion)? Was there any misinterpretation of the evidence?

- Did the student look in reasonable appropriate places for both incriminating and exculpatory evidence?

5.0 CONCLUSIONS AND FUTURE PLANS

To build these labs, we have built up quite an extensive inventory of SSDD. On campus we worked with our telecommunication office to offload their inventory of old or broken phones and PDAs. In the community we worked our local Women’s Shelter and continually help them sort through their incoming inventory of donated phones. Globally, we are working with several recycling companies to increase our library of cell phones. To date, we have built up a collection of over three hundred unique SSDD and are currently in the development of a forensic acquisition tool testing database and a SSDD forensics knowledge base, providing a needed resource for federal, state, and local law enforcement forensic examiners.

In addition to gathering more equipment, we are also looking forward to additional forensic labs. The
current plans are to incorporate other devices such as portable gaming devices, handheld GPS systems, VOIP and satellite phones, and carputers. Additionally, other features we might add include practical exercises for accessing obstructed (password protected) devices. The one thing we can count on in this field is that the memory capacitance will increase, and the device will get smaller.

AUTHOR BIOGRAPHIES

Richard Mislan is an Associate Professor of Computer and Information Technology at Purdue University, West Lafayette, Indiana. His research interests include Small Scale Digital Device (SSDD) forensics, unusual sources of digital evidence, and the application of artificial intelligence techniques for improving efficiency in cyber forensics. He can be reached at rmislan@purdue.edu

Tim Wedge has the dual role of Computer Crime Specialist at the National White Collar Center, based in Glen Allen, Virginia and Research Scientist at Purdue University, West Lafayette, Indiana. His research interests include improved tool validation methodologies, and the development of tools for improving accuracy and precision in cyber forensics. He can be reached at twedge@nw3c.org

REFERENCES


Subscription Information

The Proceedings of the Conference on Digital Forensics, Security and Law is a publication of the Association of Digital Forensics, Security and Law (ADFSL). The proceedings are published on a non-profit basis.

The proceedings are published in both print and electronic form under the following ISSN's:

    ISSN: 1931-7379 (print)
    ISSN: 1931-7387 (online)

Subscription rates for the proceedings are as follows:

    Institutional - Print & Online: $120 (1 issue)
    Institutional - Online: $95 (1 issue)
    Individual - Print: $25 (1 issue)
    Individual - Online: $25 (1 issue)

Subscription requests may be made to the ADFSL.

The offices of the Association of Digital Forensics, Security and Law (ADFSL) are at the following address:

Association of Digital Forensics, Security and Law
1642 Horsepen Hills Road
Maidens, Virginia 23102
Tel: 804-402-9239
Fax: 804-680-3038
E-mail: editor@jdfsl.org
Website: http://www.adfsl.org
Contents

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

Digital Forensic Certification versus Forensic Science Certification ........................................7
Nena Lim

Network Forensic Investigation of Internal Misuse/Crime in Saudi Arabia: A Hacking Case ..............................................................................................................................15
Abdulrazaq Al-Murjan and Konstantinos Xynos

The Virtual Digital Forensics Lab: Expanding Law Enforcement Capabilities ..................33
Sean A. Ensz and Mark R. McCoy

How Virtualized Environments Affect Computer Forensics ................................................39
Diane Barrett

SiMPLE - Rethinking the Monolithic Approach to Digital Forensic Software ..................51
Craig Valli

Data Mining Techniques in Fraud Detection .................................................................................57
Rekha Bhowmik

The Cyber-Workplace – Identifying Liability Issues in the Information Age and Managing E-Risk ..............................................................................................................73
Nigel Wilson

Designing Laboratories for Small Scale Digital Device Forensics ..................................81
Richard P. Mislan and Tim Wedge