State of the art of Digital Forensic Techniques

Enos K. Mabuto[1], H. S Venter[2]
Department of Computer Science
University of Pretoria, Pretoria, 0002, South Africa
Tel: +27 12 420 3654
Email: nasbutos@yahoo.co.uk[1], hsventer@cs.up.ac.za[2]

***Abstract* - Digital forensic standards have been developed that apply to the collection and preservation of digital evidence and they differ depending on the scene, nature and type of evidence being handled. For the successful prosecution and admissibility in court, certain accepted procedures have to be properly followed. Digital forensic examiners use different methods and tools to accomplish the same job during digital investigations and with the changing world of digital technology these tools and methods are variable to change. This paper analyzes and discusses the state of the art of digital forensic techniques as they are currently being used in the industry. Using the digital forensic process as a recognized scientific procedure, the paper looks at techniques applied in the acquisition and analysis phase.**

Keywords: digital forensics; digital forensic techniques

I. INTRODUCTION

Once an attack has occurred and a system or network has been compromised, it is essential to be able to sift through the evidence of what has happened. From a technical information technology perspective, this means knowing how to find, recognize, and locate the visible evidence of the crime. From a law enforcement perspective, this means knowing how to handle the evidence to make sure it is admissible in court [1]. A good investigator also needs to know the technicalities of where and how evidence can be located. Without the ability to reconstruct accurately what has been done, crucial evidence may be subject to question. More importantly, the qualifications of the digital forensic examiner can become an issue if the processing of the computer evidence was done haphazardly [2]. The digital forensic examiner has to use certain digital forensic techniques in order to conduct a digital forensic examination.

The digital forensic examiner has to follow the digital forensic process in order for evidence to be admissible in a court of law. The digital forensic process consists of a number of phases. It is widely accepted that the phase include acquisition, examination, analysis and reporting [7].

This paper however focuses only on digital forensic techniques as applied in the two most critical phases of the digital forensic process, which are the acquisition and analysis phases. The authors found that, from the research conducted, these two phases by far require the most need for action by a digital forensic examiner and, hence the reason for our focus on these two phases.

The remainder of the paper is structured as follows. Firstly some background on digital forensics in general and a description of the two most critical phases of the digital forensic process are given. The main contribution section then follows where fourteen digital forensic techniques are presented as found in literature and industry today. For each of the techniques a description is provided followed by examples of the tools that are used for each technique. One or more ways of conducting the procedure for each technique is iterated. A summary table summarizes the discussed digital forensic techniques in a table for the sake of convenience. The paper is then concluded in the last section.

II. BACKGROUND

In the following section the authors give a background of digital forensics including the definition, a digital forensic discussion and its goals. The digital forensic process is explained including a more detailed discussion of the acquisition and analysis phases.

A. Digital forensics

Digital forensics is the use of scientific methods for the identification, preservation, extraction and documentation of digital evidence derived from digital sources to enable successful prosecution [20]. Digital forensics must be based on the science of Information Communication Technology

(ICT) and within the requirements and interpretation of law [4].

According to the Oxford Dictionary, a technique is defined as a particular way of doing something, especially one you have to learn special skills [23]. In the field of digital forensics, a digital forensic technique consists of a procedure to be followed and often a certain tool to be used. A tool in this context can be a hardware or software tool.

In general the goal of digital forensic analysis is to identify digital evidence for an investigation. In the mid 1990's the International Organization of Computer Evidence (IOCE) [21] was established as a major development towards standardization in digital forensics. The main goal for IOCE was to "ensure harmonization of methods and practices among nations and guarantee the ability to use digital evidence collected by one state in the courts of another state" [5]. Another organization called the Scientific Working Group on Digital Evidence (SWGDE) [22] was formed in 1998 to "promulgate accepted forensic guidelines and definitions for the handling of digital evidence".

B. The digital forensic process

For the proper handling of evidence and in order to minimize errors in investigations it is necessary to have a structured way of handling investigations. This structured way is known as the digital forensic process.

Moreover, for the credibility of evidence digital forensic experts are usually asked to explain the process they used in collecting evidence in a court of law. This means that the digital forensic examiner should always know the digital forensic process and the appropriate toolsets used in a digital forensic investigation [6]. The digital forensic process can be categorized into four phases namely acquisition, examination, analysis and reporting. The process is common in different fields including mobile and network forensics. The process is used in investigations and has gained recognition in science [7] [8].The digital forensic process can also be defined as a number of steps from the original incident alert through to reporting of findings [9]. Recall that the authors only focus on the acquisition and analysis phases of the digital forensic process. Hence, the acquisition and analysis phases are discussed in more detail in the following subsections.

1) Acquisition

The acquisition phase describes how data will be acquired from different types of digital information sources. Data has to be acquired in a manner that maintains its integrity and authenticity [10].

The acquired data has to undergo forensic duplication or sector level duplication. A write blocker should be used in creating duplicates. The write blocker ensures that nothing is written to the original hard drive. Software imaging tools can also be used [11]. With imaging either a physical image (bit-by-bit image) can be created of the entire physical device or a logical image can be created which comprises of active directories and files available to the operating system [10].As a way of verifying the integrity of acquired data hashing is used. A digital hash conducts a mathematical algorithm of a device or file and provides a fingerprint that authenticates that the data has not been tampered or altered, and this fingerprint is maintained within the case file.

2) Analysis

The analysis phase describes how the data is processed. A hash analysis search can be conducted using hashing tools such as SHA-1, MD5 or using CRC applications. These tools conduct a mathematical analysis of a data storage device like a hard drive. By comparing hash values investigators can exclude large numbers of files that have no value to the case and hash comparing can be done between fingerprint and hash values of the data being examined. Enterprise forensic software like FTK and Encase can be used to compare hash values [10].Analysis is mainly about locating digital media and assembling them before interpreting the contents.

III. State of the art of Digital Forensic Techniques

In this section the authors identify fourteen digital forensic techniques as found in the literature. Each technique is discussed by describing a method which is followed by how to accomplish the task and the tool(s) used. A discussion on how the tool is used when needed is also provided. A summary table is shown which summarizes all digital forensic techniques discussed in this section.

A. Recovering deleted files

In cases were an attacker wants to cover his tracks, he can delete his files so that evidence cannot be found that links the crime to him. For this technique the authors illustrate two methods that can be used to recover deleted files.

Method one involves taking a forensic duplicate and makes it act like a real hard drive device. When the forensic duplicate is associated with a device the examiner can run tools just as if the original suspect drive were connected to the forensic workstation [3]. After running tools on the device the examiner can see what the partition table looks like, then mount the partition to recover deleted files using a forensic

tool like the Coroners tool kit by Dan Farmer and Wietse Venema, or using TASK, which was later named the Sleuth kit by Brian Carrier.

Method two involves the use of commercial forensic software. Deleted files are recovered automatically using Encase from Guidance software and Forensic Tool Kit from Access Data that recover files without user intervention. Evidence is first loaded into such a tool; the examiner then creates a case to add the evidence, and then adds a raw image to the case. The graphical user interface displays the recovered deleted files.

B. Production of time stamps and other meta data

Given a situation that access time, modification time or tags related to a file are required, there will be a need to acquire this data for appropriate court proceedings. For this technique the authors illustrate two methods that can be used to produce time stamps and other Meta data. This technique is for acquiring the metadata from all files that exist in the evidence i.e. file names and file sizes which are used for file name searches, timeline analysis and reporting.

Method one involves collecting the meta data from a logical file system and getting an output file with most of the information needed. The examiner's export the output file into their favorite spreadsheet or database program and perform queries, reconstruct the time line and then perform other search command to find missing hashes of the files used. Using a Perl script, the script connects the formats that are used by other file systems, such as Mac, files into a human readable format and prints out all the information to the console in a format that can be compared to other databases [3].

Method two involves using a commercial tool like Encase or FTK. Examiners can acquire metadata by computing, for example, hash values. After the hash values are computed they are computed, one can view the files from one of the evidence sets and then export it to a required format.

C. Removing known files

Sometimes it's necessary to exclude from an investigation all files that are commonly known like operating system files, for the purpose of saving time and concentrating only on other files more likely to contain evidence. For this technique the authors illustrate two methods that can be used to remove known files.

The first method requires a technique where examiners populate a directory with the data they want to add to a hash set. The data could be a few trusted programs or a complete trusted operating system .When they have the trusted data files, they can compare it to a list of hashes generated from evidence. Another command will produce all the files that do not match the trusted data set.

The second method involves a commercial tool like Encase. Encase has the capacity to build hash sets natively or import an existing set. When a hash set is provided to Encase, it populates the columns on the hash set and hash category. After comparing the known hashes to the hashes of the file or the unknown files, the known files are filtered filtering out. Encase can build a hash set from trusted files given to it. [3]

D. File signatures verifications

A file signature is data used to identify or verify the content of a file; this can be bytes within a file used to identify the format of the file. This technique is for determining the file content integrity generally against transmission errors or malicious attacks, such as file extension modification. The authors illustrate two methods that can be used to determine file signatures.

Method one involves using Sleuth Kit and Perl scripts to execute a command that compares the contents of the file to a so called magic file. The magic file contains information about the header and footer of several well-known types of files. A fragment is obtained which matches formats e.g. Microsoft Word documents. All such fragments are identified and imported into a spreadsheet or database. The examiner can then search for any files not matching their apparent signatures [3].

Method two uses the Encase tool. Encase is preloaded with a number of file signatures that one can view and it determines the file signatures by the search expression which is a series of hexadecimal values in the header and footer of a file that identifies the file type. The tool will simply report "bad signature".

E. String searching and file fragments

String searching is a technique employed by examiners when they know what they are looking for and they want to quickly acquire only the necessary files. One extracts all the strings from the binary data of the evidence file, then uses a string command to extract files with strings from the forensic image [3]. The examiner then uses the search command to search the strings for all keywords in the extracted text file and outputs them to a console.

F .Web browsing activity reconstruction

Web browsing activity reconstruction is a technique for trying to recover web browsing history of an attacker. This includes websites visited and any data sent over the internet. Encase, FTK and any browser history all include functionality

to examine a user's web browsing activities [3]. Encase utilizes a script referred to as E-script to parse the web browsing information found in the evidence and present it to the investigator. E-Script takes care of the logic of parsing potentially unknown file formats and presents it in an easy way to browse web page history

Internet explorer utilizes three techniques: web browsing history, cookies and temporary internet files.

web browsing history provides URLs of web sites visited recently.

Cookies are accepted while browsing the web which often contains valuable information like usernames and passwords.

Temporary Internet files are often referred to as the cache, which contain all temporary downloaded files from the web browsing.

G. Email activity Reconstruction

Email activity reconstruction is a technique used for the reconstruction of email data sent or received by an attacker. FTK and Paraben's Network Email Examiner are used. Parabens Network Email Examiner is used to convert the email repositories into a format that FTK can analyze, such as Microsoft Outlook PST files [3]. To reconstruct non-web based email in FTK, the examiners load the forensic image so that FTK identifies e-mail documents, then FTK will present all the email it detected. One can perform keyword searchs, on all the emails to identify relevant emails to investigate. After one has viewed all the emails examiners can create an email report with FTK.

H. Microsoft Windows registry reconstruction

Registry reconstruction is the technique for examining different types of log files, such as Windows event logs (such as system, application, and security logs) and application logs (such IIS, FTP logs)

The Microsoft Windows registry contains information such as installed programs, most recent documents, and most recent websites. The registry files are found in the windows configuration directory. Another method involves using FTK. Examiners can view the registry files by using file-registry view. FTK can automatically locate every registry file in the evidence. [3]

I. Analyzing unknown files

The objective of forensic analysis on files of unknown origin is to determine everything about a particular file, whether are executable or not executable. Because an executable file is not expected to contain word documents, an attacker can intend to hide information into another format. One method to analyze unknown files is by using a compiler such as Visual C++ Toolkit. With Visual C++ Toolkit files can be analyzed to reveal originality of file formats [3].

J. Software assisted

Software assisted refers to an amalgamation of almost all the techniques discussed in this section. The authors make mention of the more well known software often used in all digital forensic investigations. Therefore the reader will notice that some of these tools have already been mentioned during many of the techniques discussed. Numerous software packages is available for digital forensic investigations including but not limited to, the ones discussed in the following subsections.

1) AccessData

AccessData was Established in 1987 and offers products, services and training to digital forensic experts, government agencies, corporations, and legal firms [13]. They have a broad spectrum of standalone ad enterprise class solutions for legal review, E-discovery and compliance auditing. Available software packages include Ultimate Toolkit, Forensic Toolkit, Distributed Network Attack (DNA), Password Recovery Toolkit (PRTK), mobile phone examiner, Silent runner mobile, AD triangle and FTK Pro.

2). Guidance Software

Guidance software is recognized globally as one of the world leaders in digital forensic investigations. Services provided include litigation support, incident response and training for cooperate, law and government professionals [14]. Products include the Encase suite of packages i.e Encase Forensics, Encase Portable, Encase Enterprise, Encase E-discovery and Encase Cyber Security.

3). Parabens Forensic Tools

Paraben offers enterprise forensics, handheld forensics, hard drive forensics and network forensics software [15]. They also offer training programs and certification. In addition to these Paraben also offers open source software like p2p shuttle, p2p explorer and link 2.

4). Hot pepper technology

Hot pepper technology is a leader in custom applications and offers specialized digital forensic products for email extractions like Email Detective (EMD) [16].

5). Stepanet Datalifter

Since 1996 Stepanet Datalifter has been offering digital forensic investigating tools. Products include Datalifter Forensic Solutions and Datalifter Training Programs [17].

6). Digital intelligence software

Digital Intelligence Software offers software tools for identifying slack space, deleted files, imaging, partition un-hiding and write blocking. A particular product is Drive Spy [18].

K. Alternate data streams

Alternate Data Streams (ADS) are features introduced in Windows NTFS file system. The features are attached to a file and are generally not visible to the user. [19]. A file consists of different data streams, one stream hold the security information (access rights) another hold the real data expected to be in a file. There can be another stream with link information and other alternate data streams holding data the same way the standard file system does. ADS are totally hidden from the standard file system. That means the user can hide a lot of data in alternate data streams and nobody will even notice it from the standard file explorer.
NTFS's file stream can be detected with several command lines tool such as List Alternate Data Streams (LADS), Forensic Toolkit, Crucial ADS and Encase [12]. Prodiscover DFT can also examine alternate data streams and allows examination without altering the original data [10].

L. Live forensics

Sometimes imaging of an entire hard drive becomes complicated. For example, a hard drive with a size of tens of terabytes, can take several days to make a forensic copy. Making a forensic copy of a disk requires taking the disk offline. In such a case a company can suffer huge financial loss due while offline [10]. Live forensics is a technique when one leaves the system up and running That means taking a snapshot of the entire system, memory and disks whilst it's still running. However live forensics can affect other files like file timestamps, registry entries, swap keys and memory often checked by an hash [10]. Given authority one can install monitoring programs like Windows Forensic Toolchest (WFT) prior to incidents. Another option involves running forensic software from a USB to take snapshots in a bid to avoid changing much of the status of the system.

M. Self Organizing Maps

Self Organizing Maps (S.O.M) employ It is a neural network model and a data mining technique that can be used in digital forensics. SOMs are used to help investigators to get a visual snapshot of a hard drive enabling one to make better decisions on were to focus a digital forensic examination on a large disc. In so doing the examiner can conduct the forensics analysis process more efficiently and effectively.

The SOM is used to review interesting patterns. It allows mapping of high dimensional data onto a two dimensional map. The SOMFA tool has the ability to group data with similar characteristics and produce an ordered map which can then be visualized. Visualization techniques are applied to the two dimensional map and then displayed in the form of a hexagonal grid. It enables an interactive analysis with forensic data [14].

N. Recovering hidden files

Recovering hidden files can start from as simple as turning on the hidden property of folders to show hidden files in a Windows environment. Hidden files also include uncovering a covert channel such as steganography, identifying slack space and performing cryptanalysis on encrypted files.

Slack space is a type of unused space not being utilized by the actual data from the cluster size and is used to hide data because it does not show up in a directory or file system [10]. Steganography is a way of message concealment which can be applied to pictures, audio or videos. Programs like Steg detect, Steg break and visible noise can be used to detect steganography. Cryptanalysis can be used to perform encrypted data back into its readable form without having the encryption key. Cryptanalysis is often not successful depending on the strength of the encryption algorithm used [19]. Frequency analysis can be used for performing cryptanalysis on is another option for decrypting files. Explained on a higher level frequency analysis involves examining the encrypted text for repeated character strings and using the distance between the repeated string to calculate the key length [1].Password cracking can be used to uncover password protected files. The number of password cracking tools available on the market is legion. Table 1 summarizes the above discussion. For each technique the methods and tools used are shown.

TABLE I. Summary

	Digital Forensic Technique	Methods	Tools
A	Recovering deleted files	Forensic duplication	Coroners tool kit Sleuth Kit Encase Forensic Tool Kit
B	Production of time stamps and other Meta data	Performing queries and running find command	Perlscript Encase
C	Removing known files	Comparing md5 hashes	Encase FTK
D	File signatures verifications	Comparing header and footer of files for matching	Sleuth Kit Perl script Encase
E	String searching and file fragments	Running search command	Encase FTK
F	Web activity reconstruction	Recovering web browsing history, accepted cookies and temporary internet files	Encase FTK IE History
G	Email activity reconstruction	Converting email repositories to readable texts	FTK Parabens Network Email Examiner
H	Registry activity reconstruction	Examining windows system log files and application log files	FTK Reg edit
I	Analyzing unknown files	Using a compiler to reveal file originality	Visual C++ toolkit
J	Software assisted	Running software packages	FTK Encase Digital intelligence software Hot pepper technology Parabens Forensic Tools Stepanet Datalifter
K	Alternate Data Streams	Analyzing alternate data streams on NTFS	LADS Crucial ADS Prediscover DFT FTK Encase
L	Live forensics	Analyzing volatile processes	Windows Forensic Toolchest (WFT)
M	Self Organizing Maps	High definition data mapping	SOMFA
N	Recovering hidden files	Uncovering covert channels Decryption Cryptanalysis	Steg detect Steg Break Frequency analysis Password cracking

IV. Conclusion

The authors presented the techniques used in digital forensics and showed different methods of applying the technique, as well as tools used. The paper aimed to assist digital forensic investigators to have a better context of the state of the art of digital forensic techniques used in industry today. It is hoped this will pave the way for more research in enhancing or developing more advanced digital forensic techniques.

The research benefits both academic and industry researchers in knowing the current state of the art of digital forensic techniques in order to enable them to compare and contrast the different techniques in their endeavors of conducting digital forensic investigations.
The authors, themselves, plan to engage in future research which will involve identification of possible new digital forensic techniques and also to enhance some of the existing ones.

REFERENCES

[1] M.Cross, Scene of the Cyber Crime, 2rd ed.Syngress. 2008.p 500
[2]A.J.Marcella and R.S.Greenfield.Cyber forensics,London.Auerbach Publications. 2000
[3] K.J.Jones, R. Bejtlich and C.W. Rose. Real Digital Forensics, NewYork. Addison -Wesley. 2009
[4] J .Vacca, Digital forensics – Computer Crime Scene Investigation, Charles River Media. 2002
[5] E.Casy. Digital Evidence and Computer Crime-Forensic Science, Computers and the Internet, 2rd ed, Academic Press. 2004
[6] M.Solomon, G.Barett, D.Broom, N, Computer_Forensics. London.Sybex. 2005
[7] U.S National Institute of justice, "Electronic Crime Scene Investigation Guide: A guide for First Responders". 2001
[8] E.Casey, Handbook of Digital Forensics an Investigation, Academic Press. 2009. p 567
[9] E.Casey, Digital Evidence and Computer Crime, 2 rd ed, Elsevier. 2004
[10]A.Reys and J.Wiles. Cyber crime and digital forensics.Amorette Pedersen.Massachussetts. 2007.pp 243-263
[11] M. Van Horenbeeck, Technology Crime Investigation 24
[12]K.Mandia, C.Prosise, and M.Pepe, Incident_Response_and_Computer_Forensics, 2nd ed.NewYork.McGraw-Hill. 2003.p 72
[13] http://accessdata.com Accessed 04/05/2011
[14]http://www.guidancesoftware.com Accessed 04/05/2011
[15] http://www.paraben.com Accessed 04/05/2011
[16] http://www.hotpepperinc.com Accessed 04/05/2011
[17] http://www.datalifter.com Accessed 04/05/2011
[18] http://www.digitalintelligence.com Accessed 04/05/2011
[19] A.J.Marcella, and D.Menendez..Cyber forensics and a manual, 2rd ed,London.Auerbach Publications. 2008
[20] Kruse II and Heiser, 2002.p 93
[21] www.ioce.org Accessed 04/05/2011
[22] www.swgde.org Accessed 04/05/2011
[23] A.S. Hornby. Oxford Advanced Learners Dictionary. 7th ed. Oxford University Press.2006