

Integrated Digital Forensic Process Model

by

Michael Donovan Köhn

Submitted in fulfilment of the requirements for the degree
Magister Scientia (Computer Science)
in the Faculty of Engineering, Built Environment and Information Technology
University of Pretoria, Pretoria

November 2012

Publication data:

Michael Donovan Köhn. Integrated Digital Forensic Process Model. Master's dissertation, University of Pretoria, Department of Computer Science, Pretoria, South Africa, November 2012.

Electronic, hyperlinked versions of this thesis are available online, as Adobe PDF files, at:

<http://icsa.cs.up.ac.za/>

<http://upetd.up.ac.za/UPeTD.htm>

Integrated Digital Forensic Process Model

by

Michael Donovan Köhn

E-mail: mkohn@cs.up.ac.za

Abstract

The Information and Communications Technology (ICT) environment constitutes an integral part of our daily lives. Individual computer users and large corporate companies are increasingly dependent on services provided by ICT. These services range from basic communication to managing large databases with corporate client information. Within these ICT environments something is bound to go wrong for a number of reasons, which include an intentional attack on information services provided by an organisation. These organisations have in turn become interested in tracing the root cause of such an incident with the intent of successfully prosecuting a suspected malicious user.

Digital forensics has developed significantly towards prosecuting such criminals. The volumes of information and rapid technological developments have contributed to making simple investigations rather cumbersome. In the digital forensics community a number of digital forensic process models have been proposed encapsulating a complete methodology for an investigation. Software developers have also greatly contributed toward the development of digital forensics tools. These developments have resulted in divergent views on digital forensic investigations.

This dissertation presents the IDFPM - Integrated Digital Forensic Process Model. The model is presented after examining digital forensic process models within the current academic and law enforcement literature. An adapted sequential logic notation is used to represent the forensic models. The terminology used in the various models is examined and standardised to suit the IDFPM. Finally, a prototype supports a limited selection of the IDFPM processes, which will aid a digital forensic investigator.

Keywords: Digital Forensics, Digital Forensic Process Models, Digital Evidence, Digital Forensic Framework.

Supervisor : Prof. J. H. P. Eloff

Department : Department of Computer Science

Degree : Master of Science

There are a great many things about architecture that are hidden from the untrained eye.

– Frank Gehry

Acknowledgements

- Thank you God for giving me the strength, wisdom, endurance and the patience to do this, despite the constant challenges.
- I would like to thank my mother and father, Gary and Magsie, for their constant support, love and encouragement. Thank you for always listening and giving much needed advice. I would also like to thank my brothers Jonathan and Alexander, and sister Vanessa for their enthusiasm. None of this would have been possible without your support.
- Thank you Prof Jan Eloff who has insisted that I complete this milestone in my life. Thank you for pushing me through the difficult times with your knowledge and experience. Thank you Prof Mariki Eloff for your soft and kind words that inspired me to see this through. You have both become great friends in this journey.
- Thank you Prof Kourie, Prof Olivier, Prof Venter and Prof Carsten for listening when I asked questions, guiding me toward possible answers. Your help is highly appreciated.
- My friends and colleagues in the ICSA lab over the past number of years. Thank you Emmanuel Adigun, Kamil Reddy, Maciej Rossudowski, Marco Slaviero, Neil Croft, Johann Fourie, Francois Mouton, Pedro de Souza and Waldo Delpont. You have all taught me important lessons and I am thankful for your patience.
- My friends and colleagues at the University need a special thanks. Will van Heerden, Marius Riekert, Jan Kroeze, Pierre Rautenbach, Nelis Franken, Koos de Beer, Hossana Twinomurizi, Katherine Malan and Ronald Klazar. The journey would have been less interesting without you.
- Thank you Eugene Wessels, Francois Malan and Elsamaria Eloff, who have patiently nudged me along the difficult path. I am always grateful for your ability to make things sound fine.

- Thank you Isabel Claassen who made the words come to life.
- Lastly, I would like to thank all my students. Special thanks to Frederick van Staden who willingly *refactored* my *horrible* code and added an attractive GUI.

Contents

List of Figures	8
List of Tables	9
Chapter 1 Introduction	10
1.1 Introduction	10
1.2 Motivation for the Study	13
1.3 Problem Statement and Objective	14
1.4 Research Methodology	14
1.5 Terminology Used	15
1.5.1 Framework	15
1.5.2 Digital Forensics	16
1.5.3 Investigation	16
1.5.4 Process	16
1.6 Dissertation outline	16
Chapter 2 Digital Forensics and Investigations	19
2.1 Introduction	19
2.2 Forensic Science	20
2.3 Digital Forensics as a Discipline	21
2.4 Digital Forensics	21
2.5 Digital Forensics Defined	23
2.6 Digital Forensic Investigations	24
2.6.1 Types of Digital Forensic Investigations	24

2.6.2	Goal of a Digital Forensic Investigation	26
2.7	Conclusion	28
Chapter 3	Digital Evidence	29
3.1	Introduction	29
3.2	Digital Evidence	30
3.2.1	Where Digital Evidence is found	30
3.2.2	Defining Digital Evidence	31
3.2.3	Characteristics of Digital Evidence	33
	Locard’s Exchange Principle	33
	Digital Stream of Bits	34
3.3	Digital Evidence and Metadata	34
3.4	Legal Principles of Digital Evidence	36
3.4.1	Circumstantial and Hearsay Nature of Digital Evidence	36
3.4.2	Authorisation to Conduct a Digital Forensic Investigation	37
3.4.3	Authenticity of Digital Evidence	38
3.4.4	Scientific Method	38
	Digital Evidence Principles as Formulated in Frye and Daubert	39
3.4.5	Best Evidence Rule	40
3.5	Best Practice and Standard Operating Procedures during Investigations	41
3.5.1	Best Practice	42
3.5.2	Standard Operating Procedures – SOPs	43
3.6	Conclusion	44
Chapter 4	Digital Forensic Process Models – DFPMs	45
4.1	Introduction	45
4.2	Motivation for Using Sequential Logic	46
4.3	Process Representation in Sequential Logic	47
4.4	Lee	48
4.4.1	Process	48
4.4.2	Terminology	48
4.4.3	Comments	49

4.5	Kruse and Heiser	50
4.5.1	Process	50
4.5.2	Terminology	50
4.5.3	Comments	50
4.6	NIJ Electronic Crime Scene Investigation	51
4.6.1	Processes	51
4.6.2	Terminology	51
4.6.3	Comments	53
4.7	NIJ Forensic Examination of Digital Evidence	54
4.7.1	Process	55
4.7.2	Terminology	55
4.7.3	Comments	56
4.8	Casey	56
4.8.1	Process	56
4.8.2	Terminology	57
4.8.3	Comments	57
4.9	Digital Forensics Research Workshop Group - DFRWS	59
4.9.1	Process	59
4.9.2	Terminology	59
4.9.3	Comments	59
4.10	Reith	60
4.10.1	Process	60
4.10.2	Terminology	60
4.10.3	Comments	61
4.11	Carrier and Spafford	62
4.11.1	Process	62
4.11.2	Terminology	63
4.11.3	Comments	63
4.12	Baryamureeba	64
4.12.1	Process	64
4.12.2	Terminology	65

4.12.3	Comments	65
4.13	Ciardhuáin	66
4.13.1	Process	67
4.13.2	Terminology	67
4.13.3	Comments	69
4.14	Cohen	70
4.14.1	Process	70
4.14.2	Terminology	70
4.14.3	Comments	71
4.15	Some Common Ground	71
4.16	Conclusion	75
Chapter 5	Integrated DFPM	76
5.1	Introduction	76
5.2	Construction and Interpretation of DFPMs	77
5.3	Reducing the Number of Processes in the DFPMs	78
5.4	Integrated DFPM - IDFPM	83
5.5	Conclusion	87
Chapter 6	The complete IDFPM	88
6.1	Introduction	88
6.2	Documentation	89
6.3	Preparation	90
6.3.1	Policy/Procedure	91
6.3.2	Infrastructure Readiness	91
6.3.3	Operational Readiness	91
6.4	Incident	92
6.4.1	Detect	92
6.4.2	Assess	93
6.4.3	Confirm	93
6.4.4	Notify	93

6.4.5	Authorise	93
6.4.6	Deploy	93
6.5	Incident Response	94
6.5.1	Approach Strategy	95
6.5.2	Search	95
6.5.3	Seize	95
6.5.4	Recover	96
6.5.5	Preserve	96
6.5.6	Transport	96
6.5.7	Store	96
6.6	Digital Forensic Investigation	96
6.6.1	Collect	98
6.6.2	Authenticate	98
6.6.3	Examine	98
6.6.4	Harvest	98
6.6.5	Reduce	99
6.6.6	Identify	99
6.6.7	Classify	100
6.6.8	Organise	100
6.6.9	Compare	100
6.6.10	Hypothesis	100
6.6.11	Analyse	100
6.6.12	Attribute	101
6.6.13	Evaluate	101
6.6.14	Interpret	101
6.6.15	Reconstruct	101
6.6.16	Communicate	101
6.6.17	Review	102
6.7	Presentation	102
6.7.1	Present a Report	103
6.7.2	Decide	103

6.7.3	Disseminate	103
6.8	Additional IDFPM Requirements	103
6.8.1	Role Players	104
6.8.2	Location	104
6.9	The IDFPM Support Structure	104
6.9.1	Law	105
6.9.2	Principles	105
6.9.3	Experience	106
6.10	Conclusion	106
Chapter 7 The IDFPM and Digital Forensics Tools		108
7.1	Introduction	108
7.2	Tools	109
7.3	The Integrated Digital Forensics Process Model Prototype - IDFPMP	113
7.3.1	IDFPM Processes Identified in the IDFPMP	113
7.3.2	Functional Requirements of the IDFPMP	114
7.3.3	Use Case of the IDFPMP	115
7.3.4	Technical Platform Specification of the IDFPMP	116
7.3.5	Interface Design of the IDFPMP	116
7.3.6	Process Description of the IDFPMP	117
7.3.7	External Output Structures of the IDFPMP	120
	XML Metadata Index	121
	IDFPMP MySQL Database	122
7.4	Conclusion	123
Chapter 8 Conclusion		124
8.1	Introduction	124
8.2	Revisiting the Problem Statement and Research Objectives	124
8.3	Main Contribution	125
8.4	Future Research	126
Bibliography		127

Chapter A Publications and Contributions	136
A.1 Framework for a Digital Forensic Investigation	136
A.2 UML Modelling of Digital Forensic Process Models (DFPMs)	145
A.3 Isolating a cloud instance for a digital forensic investigation	159
A.4 Integrated Digital Forensic Process Model	167

List of Figures

5.1	The IDFPMP illustrated as a process flow diagram	86
6.1	The <i>prepare</i> process flow diagram in the IDFPMP followed by the <i>detect</i> sub-process, not illustrated here	90
6.2	The <i>incident</i> process flow diagram in the IDFPMP followed by the <i>approach strategy</i> sub-process, not illustrated here	92
6.3	The <i>incident response</i> process flow diagram in the IDFPMP followed by the <i>collect</i> sub-process, not illustrated here	94
6.4	The <i>investigation</i> process flow diagram in the IDFPMP followed by the <i>present a report</i> sub-process, not illustrated here	97
6.5	The <i>presentation</i> process flow diagram in the IDFPMP	102
6.6	The framework within which the IDFPMP will operate	105
7.1	IDFPMP processes as found in the IDFPMP	114
7.2	IDFPMP use case	115
7.3	IDFPMP interface	116
7.4	IDFPMP main frame tabs	117
7.5	IDFPMP results interface	117
7.6	IDFPMP evidence loader	118
7.7	IDFPMP file structure loaded	119
7.8	IDFPMP processing complete	119
7.9	IDFPMP XML file structure loaded	120
7.10	XML metadata index structure	122
7.11	IDFPMP database output	123

List of Tables

4.1	DFPMs	73
5.1	DFPMs with reduced process listing	82
7.1	Tools supporting the DFPM processes	111
7.2	The left panel of the IDFPMP icon process description	118

Chapter 1

Introduction

O villain, villain, smiling, damned villain!

Hamlet - Shakespeare

1.1 Introduction

Users in the Information and Communications Technology (ICT) environment cover a wide spectrum from individual computer users to large corporations. The ICT working environments are experiencing increased computer use for other than work-related reasons. User activities may include but are not limited to browsing the Internet for private purposes and using online search engines for work-related information. Large organisations on the other hand conduct online business and retain a visible profile to attract new investors.

Advances in social networking, mobile technology, various cloud computing and storage solutions have increased the information flow options available to organisations. This has weakened the control of the organisation over information. The increasing activity in ICT-focused environments has also led to an increase in computer and network-related misuse. These range from employees, using simple password cracking tools to gain access to managerial account information, to fraud and theft of company resources.

Increased computer and network misuse, among many others, has led to an increase in computer-related investigations. A typical investigation will have some hypothesis or

observable phenomenon that is verified by some proof [104]. Auditing logs have been useful to draw some correlation between user activity and verification of a suspicious activity during these investigations. Non-repudiation is included if these activities can be attributed to a specific user without the user being able to deny such an action. However, in computer environments, audit logs cannot be used to prove user actions beyond dispute [13].

Although these developments have led to auditing being key in answering many of the questions related to user activity [30, 88, 37], they have not been sufficient in finding the root cause of suspicious activities.

Digital forensics has made some rapid developments over the past few years. One of the main reasons for this is the advancement in tools and systems that allow ordinary computer users to be more proficient in performing difficult audit tasks [104]. Current literature and Internet searches point to numerous sources that give a trivial and easy tutorial on how to perform simple tasks aimed at gaining access to computers [66]. The Internet has also enabled the ordinary computer user to access all types of information such as pornography, illegal software, copied music and confidential documents. This has demanded great strides in computer security mechanisms in an effort to counteract such activities and a growing need for forensic tools to gather accurate digital evidence for prosecuting harmful criminals who commit all types of computer-related incidents. Garfinkel [41] and Beebe [8] submit a lack in digital forensic standardisation and process, which is resulting in limited prosecution.

Advances in and the ease of use of forensic tools have created a false sense of security based on the credibility of the evidence produced by these forensic tools. A misconception has been created, namely that the man in the street can conduct a computer forensic investigation. The forensic tools used have various features that facilitate digital forensic investigations [22, 14]. In a court of law the process followed in gathering the digital evidence as well as the digital evidence itself is important. Unfortunately the court proceedings focus on scrutinising the validity of the process followed in evidence handling before considering the evidential value.

In South Africa there are a number of cases where digital evidence has been highlighted as pivotal to the conclusions drawn. One such case is the Motata case where a

digital cellular telephone recording was ruled admissible evidence after the original had been deleted [96]. In this case the court investigated the admissibility of certain audio media recorded on a cellular telephone. The original audio recording was deleted from the cellular telephone once the audio recording was copied to a computer.

In the Vermooten case, the only copy of an original testament was on the deceased's portable computer [95]. The question was whether the document had been modified since the death of the deceased. Here the stored computer document was admitted as an original and authentic document. In both the above cases the authenticity of stored digital files was questioned. In the Vermooten case the digital document was accepted as the original, but in the Motata case the copied audio recording was accepted as sufficient proof of the existence of an original.

Numerous procedures have been proposed for the collection of digital forensic evidence. Committees such as the Digital Forensic Research Workshop Group (DFRWS) [32] and the American Society of Digital Forensics and eDiscovery (ASDFED) [4] have proposed processes to be followed in the collection of digital evidence. From the above it follows that there is no standard forensic process in place to be followed by digital forensic investigators. It would be a serious mistake for a forensic investigator to ignore the procedure of evidence collection in cases where the evidence aids in proving the case and leaves no doubt in the minds of those having to decide on the matter. Where evidence is presented without proof of thorough procedure, the defense may question the forensic procedure followed to collect the digital evidence. The famous American court case of Simpson is an example where the forensic process was scrutinised by the defense [108]. In this case the crime scene evidence was collected, but a robust evidence collection process was not followed. Because of this the evidence was invalidated by the defense.

Tools such as Encase have been accepted as a reliable solution in computer crime investigations [49]. Both the process followed when using Encase and the resulting digital evidence are accepted as reliable. Other tools have also been used successfully, such as FTK [2] and SleuthKit [15]. Some are commercially available and others are open source. Many of these tools have been validated and accepted as reliable by the American judiciary [49]. However, the evidence collection process and the digital evidence presentation are vital in a successful prosecution.

1.2 Motivation for the Study

The field of digital forensics is relatively new [7, 84]. Various methodologies have been proposed, yet there is no accepted standard procedure within the digital forensics community. Several forensic techniques are used with success where others have failed to prove events beyond doubt. Reasons for this failure include inadequate resources, lack of sufficient training and shortage of funding. Experts are scarce and expensive [3], while the lack of a professional association governing the actions of experts has been criticised [3, 84]. Within the research community there are also inconsistencies in terminology used to describe various processes [26, 25].

The current literature reveals a number of proposed frameworks, models and procedures that have been put in place in an attempt to formally describe an effective digital forensic investigation process. In digital forensic investigations, various processes focus on different actions performed, such as the data extraction. Others tend to be more concerned with the analysis of the data extracted from the digital media.

There is an acute need for approaching digital forensic investigations in an integrated and formal manner that allows a flexible approach towards the changing environments of digital forensics, yet maintains a strict procedural line to satisfy the courts in their scrutiny of the evidence. Not only is the evidence itself important in a digital forensic investigation, but the process followed in acquiring the evidence before final presentation is crucial. Numerous court decisions have given guidelines to determine the standard required from the evidence collected and process followed in its collection.

There are, however, very few guidelines for establishing a formal digital forensic standard. Each investigation is dealt with separately and every investigator has a specific procedure that he/she will follow in a specific type of investigation. The valuable experience gained from previous investigations is seldom stored by forensic investigators. Evidence types found on computers are discarded or destroyed after each investigation. Process exceptions in unique cases are documented and placed in storage, never to be looked at again. Most importantly, court findings are not recorded for easy access by digital forensic investigators.

1.3 Problem Statement and Objective

Digital forensics is a relatively new field in the greater ambit of forensics. For a number of years numerous processes have been proposed in an attempt to capture a process that will allow a complete investigation. Many software developers have seized the opportunity to develop a tool to manifest some of the proposed processes. Indeed some have even developed their own digital forensic methodology, such as Encase [22]. Most of these tools provide partial solutions based on their understanding of the essential process components. Some of them focus on different phases or subcomponents of an investigation. However, no integrated approach towards digital forensic investigations has been established yet. The shortcomings of the investigation process are generally exploited in court where both the digital evidence and the digital forensic process are scrutinised by the court and opposing stakeholders.

The primary objective of this dissertation is to investigate existing digital forensic models reported on in the known published literature to determine whether these can be integrated into a single digital forensic process model. Evidence produced by implementing this process model in an investigation will ensure that it can withstand legal scrutiny in a court of law. The digital evidence presented is thus the result of a rigorous digital evidence collection process. The dissertation will however be limited to the development of an integrated digital forensic process model and will not include a validation study.

A secondary objective is to investigate whether any part of the process model can be automated, without compromising the validity of the process model. The purpose here is to alleviate the time-consuming investigative processes on behalf of the investigators. Such processes are generally aimed at reducing the data to be analysed during the investigation.

1.4 Research Methodology

As was mentioned in the previous section, the dissertation focuses on two main issues. The first involves the lack of a standardised digital forensic process in the investigative community. The second involves the tremendous volumes of data to be analysed during a digital forensic investigation.

To address these problems, the following methodology is listed as a road map.

- Examine current literature on digital forensic process models. Review and identify essential components of current digital forensic process models so as to construct an integrated digital forensic process model.
- Construct a model in selecting the essential components identified in the literature study. Integrate these components into a single integrated digital forensic process model.
- Construct a consolidated framework within which the digital forensic process models are executed. Consider various perspectives to accommodate all possible role players within an investigation.
- Examine the integrated digital forensic process model to identify which processes can possibly be automated to ease the time-consuming task of data reduction in future investigations. Use set theory as a basis for finding unique user data.

The formal approach outlined above includes an analysis of the current state of investigative process models. Once an integrated process model has been designed, a prototype should be developed to validate the proposed framework.

1.5 Terminology Used

In this dissertation the following terms have to be defined in advance because they are sometimes used differently in the literature.

1.5.1 Framework

The term framework is used extensively in this dissertation. In the literature, a number of other terms are often used, for instance architecture. Framework is defined as a structure for supporting, specifically a *skeletal support used as the basis for something being constructed or a structure supporting something* [101].

1.5.2 Digital Forensics

Digital forensics is a specific, predefined and accepted process applied to digitally stored data or digital media that use scientific proven and derived methods, based on a solid legal foundation, to produce after-the-fact digital evidence [77]. It aims at deriving the set of events or actions indicating a possible root cause, where reconstruction can be used to validate the scientifically derived conclusions. This definition is discussed in Chapter 2.

1.5.3 Investigation

The online dictionary gives the following definition for an investigation: *The act or process of investigating*. A second definition is a *detailed inquiry or systematic examination* [40]. An investigation is primarily defined as a careful search or examination in order to discover facts. In a digital forensic investigation the facts that are discovered form part of the evidence presented in court. Chapter 2 discusses an investigation in detail.

1.5.4 Process

Throughout the literature on digital forensics, the terms process and model also appear frequently. Model is an example of a pattern someone might follow [101]. In contrast, process has a number of meanings which are all considered important, such as given by the free online dictionary is *a series of actions or changes, or a series of natural developments which result in an overall change*. Another definition of a process includes computer actions. Computers perform operations on data in order to obtain the required information [40]. For the purposes of this research, most of the meanings have been considered and regardless of the underlying operations performed, the process must enable effective data extraction to aid in further investigation.

1.6 Dissertation outline

This dissertation is structured as follows:

Chapter 1: Introduction

The current chapter includes an introduction, motivation for the study, problem statement and research objectives. The research methodology is presented. Terminology used in the dissertation is listed.

Chapter 2: Digital Forensics Investigations

Chapter 2 describes digital forensics within the broader context of sciences. Digital forensics is defined for the purposes of this dissertation by listing the essential elements of the definitions found in the current literature. The goal of a digital forensic investigation is also given.

Chapter 3: Digital Evidence

Chapter 3 deals with aspects of digital evidence exclusively. These aspects include what digital evidence is, where it is found and how an investigator should deal with such evidence. Best practice and standard operating procedures are listed. Rules relating to scientific testimony are discussed.

Chapter 4: Digital Forensic Process Models – DFPMs

Chapter 4 introduces Digital Forensic Process Models. Each DFPM selected from the current literature is briefly discussed. A process description in sequential logic notation is given and the terminology described. The DFPMs are briefly commented on by listing areas for commendation and shortcomings.

Chapter 5: Integrated DFPM

Chapter 5 integrates the various DFPMs discussed in Chapter 4. The construction of Table 4.1 is described, including how it should be interpreted. The existing process descriptions are reduced by looking at the process terminology of the DFPMs discussed in Chapter 4. The Integrated DFPM is constructed by grouping similar processes together.

Chapter 6: The complete IDFPM Framework

Chapter 7 discusses the IDFPM terminology as used in the complete framework. The IDFPM rests on three essential pillars which are the law, digital evidence principles and investigator experience.

Chapter 7: The IDFPM and Digital Forensics Tools

Chapter 7 briefly describes a number of tools available to an investigator during an investigation. A prototype, specifically developed for the research project, addresses a limited number of the IDFPM processes. The processes of the prototype that supports the IDFPM are listed and discussed.

Chapter 8: Conclusion and Future Work

Chapter 8 concludes the dissertation.

Chapter 2

Digital Forensics and Investigations

How poor are they that have not patience!

Othello – Shakespeare

2.1 Introduction

This chapter aims to place the field of digital forensics as a science within the broader spectrum of disciplines. A few definitions of digital forensics are given. These definitions are critically discussed in addition to comments found in the literature. The main elements are identified from which digital forensics is defined for the purposes of this dissertation.

Digital forensic investigation as used in this dissertation is discussed by investigating definitions found in the literature. Forensic science is briefly introduced, followed by some reflection on its definition and historic origin in Section 2.2. Section 2.3 places forensic science as a discipline among other sciences, whereafter the term digital forensics as defined in the literature is discussed in Section 2.4. Section 2.5 provides a definition of digital forensics for the purposes of this dissertation. Sections 2.6.1 and 2.6.2 define digital forensic investigations and state the goal of such an investigation.

2.2 Forensic Science

Forensics is a well-established science where various contributions have been made to achieve results in the criminal justice systems [23]. The origins of the word *forensic* can be traced back to Latin. *Forensis* is defined as *before the forum* [76]. Forensic science is also often referred to as forensics. The application of forensic sciences is applicable in criminal and civil actions. Therefore forensics is effectively a synonym for *legal* or *related to courts* [109].

In Roman times, cases were presented in a forum to a group of individuals who would hear the matter. The accuser and the accused gave speeches based on their version of events. The best-presented argument would determine the fate of the accused. The modern use of the word forensic is used with this background in mind, therefore involves a *form of legal evidence* and a *category of public presentation* [109].

In its early development, forensic science lacked standard practice, which in turn allowed many criminals to escape liability. Public hearings were conducted with forced confessions and manipulated witness testimony. During the 5th century, Germanic law stipulated that a medical expert should determine cause of death [71]. A 1247 Chinese book contains a procedure to be followed in the event of a suspicious death [71]. In 1248 the first written account of forensics is found in a case where the application of medicine and entomology were used to solve a case [71]. In an article titled *A History of Digital Forensics*, Pollitt gives a more complete overview of digital forensics [81].

Logic and *procedure* were increasingly applied to forensic science to determine the root cause in incident investigations [109]. During early development the focus was on the *process* followed and conclusions were formulated from deductions based on rational *logic*.

Forensic sciences are increasingly being applied to other well-established sciences such as accounting, economics and management. Forensic and audit functions are often both found in large organisations, such as audit companies, bank services and tax and revenue services. Examples in South Africa are seen in KPMG, ABSA and SARS respectively.

Forensics finds application within a broad spectrum of sciences of which digital forensics is one such branch within the sciences; therefore the aim of Section 2.4 is to define the term *digital forensics*.

2.3 Digital Forensics as a Discipline

To appropriately position digital forensics as a field within the broader context of the sciences, and to understand the nature of digital forensics, the discussion below explores where digital forensics should be placed.

The Oxford Dictionary defines the word forensic as *relating to or denoting the application of scientific methods to the investigation of crime and of or relating to courts of law* [101]. This definition has two distinct components that require further explanation. Firstly, scientific techniques and methods must be applied in an investigation. Secondly, the scientific methods and techniques used have to be legally relevant when the evidence is examined by a court of law. By implication, several legal rules of evidence, best practice and principles have to be followed to ensure the evidence is admissible (to be examined in Chapter 3). Evidence presented with the proper procedure is used to prove or disprove that the suspected set of events has occurred [61]. Proper procedure is crucial in any investigation where a matter will be dealt with in court.

Forensics is the art or study of argumentation and formal debate [33]. Ciardhuáin emphasises the application of scientific knowledge to legal problems [74, 50, 24]. Digital forensics is therefore dependent on science and law [22, 103].

2.4 Digital Forensics

Digital forensics is often defined from a selective perspective of the person involved in an investigation [54]. This section lists and discusses a couple of important definitions generally accepted in the literature. Common elements are extracted from the definitions to formulate an inclusive definition for digital forensics.

Carrier states that computer forensics, digital forensics and media analysis are often used to describe the relatively new field of digital forensics [13]; for this reason the terms are used interchangeably in this dissertation. Computer forensics [70, 27, 107, 119, 60] and digital forensics [56, 62, 77, 97, 26] are both used in the literature to describe the sub-branch of forensic sciences. Digital forensics seems to be the generally accepted term.

The definition given by Palmer in 2001 is one of the first references to *digital* forensics,

and it broadens the scope of possible media to be included in an investigation:

Digital Forensics 1: Digital forensics is the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from digital sources for the purposes of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorised actions shown to be disruptive to planned operations [77].

This definition formulated by Palmer is generally accepted to be an all-inclusive definition. The procedure is included by the set of listed processes: *preservation, collection, validation, identification, analysis, interpretation* and *presentation*. The scientific method used to execute the process should be scientifically proven or validated. Pollitt states that digital forensics is not a single process but a group of tasks and processes in an investigation [27]. Robbins does not prescribe the process as methodically as Palmer [107], but nevertheless includes it in his definition. Formulating a process list in the definition is, however, rather limiting and should be avoided.

Reconstruction is listed as an element to aid in finding a root cause or simulating the anticipated events leading to a state. Reconstruction is based on empirical research which confirms scientific method. Within digital forensics, mention is often made of the fact that a different investigator using different tools should reach the same conclusion [106]. The actions, being unauthorised or shown to be disruptive to planned operations, must be known or identifiable. This is the focus of the work done by Tan on forensic readiness [100, 90].

In 2005 Willassen redefined digital forensics as follows:

Digital Forensics 2: Digital forensics is the practice of scientifically derived and proven technical methods and tools towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of after-the-fact digital information derived from digital sources for the purpose of facilitating or furthering the reconstruction of the events as forensic evidence [117].

At first glance, this seems to be a copy of Palmer, but there are a few important differences. First, Willassen replaces *use* with *practice*. Secondly, not only the method is important, but also tools used to produce evidence. Thirdly, the criminal element is removed from the definition, which broadens the scope of application to include digital forensics as a means for finding originating causes in other types of investigations. Lastly, the root cause is not included. Certain investigations will not have a specific root cause. Willassen states that the investigation will be after the fact, or post incident [117].

A striking similarity with Palmer is that the *process* can be used to aid in providing evidence, make a possible reconstruction of events; and help in preventing future similar incidence [80].

From the definitions listed earlier and the discussion above the common elements identified in a digital forensics definition are:

- Scientific derived and proven methods and tools;
- Applied to some digital media or digital data;
- After the fact or *post facto*;
- In a specific predefined process or accepted process;
- Considering legal principles;
- Extract digital evidence;
- Indicating a set of events or actions being the root cause; and
- Used to aid in reconstruction, decision and future prevention on a previously taken action or *a-priori* [64].

The term digital forensics as used for the purposes of this dissertation is defined in Section 2.5.

2.5 Digital Forensics Defined

Considering the definitions discussed in Section 2.4, digital forensics is subsequently defined in this dissertation as follows:

Digital Forensics is a specific, predefined and accepted process applied to digitally stored data or digital media using scientific proven and derived methods, based on a solid legal foundation, to extract after-the-fact digital evidence with the goal of deriving the set of events or actions indicating a possible root cause, where reconstruction of possible events can be used to validate the scientifically derived conclusions.

The legal foundation is not the focus of this dissertation, but is seen as the departure point in various methodologies [74].

2.6 Digital Forensic Investigations

Digital investigations, digital forensic investigations, forensic examination and forensic investigations have been used to describe an *investigation* where a digital device forms part of the incident. For the purposes of this dissertation the term digital forensic investigation (DFI) is used. The terms will however be used interchangeably in this section to accurately reflect various author opinions. The successful outcome of a DFI is the presentation of digital evidence (to be discussed in Chapter 3). A DFI is conducted by an appropriately qualified investigator [3].

Section 2.6.1 describes the various references to types of investigations, whereafter the definition used in this dissertation is given. Section 2.6.2 discusses the successful outcome of a digital forensic investigation.

2.6.1 Types of Digital Forensic Investigations

A *forensic examination* is a scientific process of analysing media, file systems, devices or content for information of probative value [27]. Probative value is the weight that the evidence will carry in determining the outcome. Forensic examinations can be extended to include examinations of digital images where the authenticity of digitally stored images are determined. Techniques include photogrammetry but do not include photos taken to document the crime scene [73]. Digital forensic examination is part of the greater digital forensic process.

A digital device is the focus of a digital investigation where an incident has occurred [13]. Carrier defines a digital investigation as a *process* that formulates and tests a *hypothesis* to produce a reasonable explanation about digital *events* or the *state* of digital data [20, 14]. The hypothesis and digital device determine which investigation techniques are used.

The formulation of the hypothesis considers the state data is in and the action taken on data to bring about the change. Data is generally in one of two states, *original* and *modified*. The *action* taken during an incident will bring about a change in state. The action determines the type of incident [13].

According to Irons, digital investigation is a generic term used to describe any activity that involves computer or digital technology to produce forensically robust evidence [55]. Forensic computing is used as a technique to examine and analyse data [67, 78, 52, 85].

A *digital forensic investigation* is the process to determine and relate extracted information and digital evidence to establish factual information for judicial review [54]. Ioeng and Cohen emphasise the need to establish factual information as the outcome of an investigation [54, 26].

Carrier proposes the use of the term digital forensic investigation rather than digital forensics [14], and reasons that forensic science addresses a limited hypothesis. DNA can be used in a digital forensic investigation to determine the relationship between the suspect and the device at the physical crime scene [22]. The DNA forensic process does not encompass an entire investigation, but only a limited question such as *does the fingerprint found at the crime scene match that of the suspect?*. In this instance, physical science forensics aids digital forensic science to determine the answer to a question in the greater digital forensic investigation, namely - *was the suspect at the crime scene?*

Physical forensics is used to individualise or determine the unique source of an object in the context of the incident [25]. *Digital investigation* or a DFI is the process of identifying potential evidence *and* identifying the unique source [14].

A *forensic investigation* of digital evidence is employed as a post-event response to an incident [90]. Live forensics or Network forensics is conducted in real time, and would therefore not technically be called a forensic investigation [63]. Traditional or *dead* forensics involves the recovery of evidence from computer systems that have been powered

down [57]. Cloud forensics is a recent development in digital forensic investigations [31, 65]. Mobile forensics has also been researched to include investigations using cellular technology [83, 28, 68].

A *DFI* is therefore a special type of a investigation where the scientific procedures and techniques used will allow the results – digital evidence – to be admissible in a court of law. The result of a DFI must have a legal basis. Evidence cannot be directly viewed and some tool will be used to examine the state of the data. One of the tools to observe the state of digital data is indirect data observation. This is similar to being told about something instead of seeing it for yourself, formally known as *hearsay* in rules of evidence. The weight you attribute to the evidential value is based on the extent to which the tool is trusted [118, 78]. The confidence in DFIs is based on the level of trust of the hardware and software used to collect and analyse the data [16].

For the purposes of this dissertation, a digital forensic investigation is defined as follows:

Digital forensic investigation or *DFI* is a special type of investigation where the scientific procedures and techniques used will allow the results – digital evidence – to be admissible in a court of law.

2.6.2 Goal of a Digital Forensic Investigation

Any investigation conducted has some outcome, or very specific purpose. An investigation is generally initiated with the idea to establish facts about an event that has taken place. According to Kruse, forensics is conducted to determine the root cause of an event or incident [60]. The primary goal in establishing the root cause is to ensure that the investigation is conducted in a manner to withstand legal scrutiny, when the matter is serious enough to warrant it. However, any investigation should be conducted methodically to ensure that the conduct of the investigator is of such a nature that the validity of the evidence produced will not be questioned. It should be noted that various types of DFIs exist. These include live forensics, proactive forensics and network forensics [9, 89, 68]. Beebe and Clark suggest a second-tier phase to the DFI, which specifically anticipates steps followed in specific incident investigations [9]. Types of incidents include drug activity, financial crimes and child pornography.

Rogers [89] propose that a digital evidence triage to aid the investigator. The evidence triage consists of the user usage profile, Internet usage and chronological timeline activity. The specific user evidence is found in home directories, the registry and file properties. Depending on the type of investigation conducted the evidence triage will guide an investigator to possible evidence, if other traces have possibly been removed.

The fundamental point of departure for any investigation is to answer basic questions about evidence. In addition to knowing what happened, there is a need to know who is responsible [60]. Every investigator should ask six key questions during an investigation: what, why, how, who, where and when [54, 9]. The *what* is determined by the data attributes or metadata; *why* refers to the motivation; *how* is the procedure followed to initiate the incident or isolate the necessary evidence; *who* are the people involved; *where* refers to the location and *when* refers to time. The following paragraphs give some clarification on some of the questions, however not all will be discussed.

Finding the person *who* performed an alleged action is vital when trying to lay blame for an action on a computer. As soon as the person has been identified, a much faster avenue is opened to finding the rest of the pieces of evidence. An interview can be conducted to ask pressing questions about motive [22]. In a civil claim the claimant and respondent or other legal persona might be important. An aspect to be considered here is that the investigator must never be suspected of fabricating the evidence and the person it can be attributed to as being the author.

What one finds as evidence might have no relation to the crime and could probably not even make sense. Evidence can be found in states, objects or events. Various models have been proposed to mathematically express what the investigator will find as determined by the hypothesis [17, 19, 18, 25, 44, 45, 86, 85].

These models tend to complicate simple cases, in which one often only has to prove a very limited hypothesis. It may be a picture of a married couple, taken by a camera, where the *connection* of the camera to the computer is in question. The answer here may be as simple as that the picture was found on the computer in the file format of a *.jpeg* file, therefore the camera is somehow connected to the computer the picture was found on. *What* evidence is found should not be confused with *how* the evidence was found [54].

This might sound very easy from a physical crime scene scenario point of view. However, in a DFI the *where* is one of the most contentious issues when dealing with digital evidence, especially in view of the strict requirements of the law of evidence. Depending on where the evidence is found, a number of factors are considered. Firstly, the evidence might be in a country where there is no mention of the crime being committed – often referred to as the principle of legality [10]. Secondly, there might be no trace of the evidence. Locards principle assures that it must be somewhere [23], Cohen specifically includes examination and traces as a process in an investigation [25]. Thirdly, there are cases where no blame can be attributed to anyone but an innocent computer programmer. Lastly there may be some unknown exchange between two autonomous systems transacting on a scheduled batch run. The legal counsel is also interested in the geographic location of the incident [54].

Patel addresses the issue quite extensively in a number of published articles [46, 78]. Time and synchronisation are often disputed, as is seen in the limited number of cases dealing with time [95].

Chapter 4 deals extensively with the question of *how* evidence was found in an investigation. The precision of the investigation and the process followed determine the accuracy of incident reconstruction.

The purpose of this section is to highlight the purpose of a DFI, in other words to find evidence to support a legal conclusion.

2.7 Conclusion

This chapter described the nature and origin of forensic sciences. Digital forensics was defined after a discussion of some prominent definitions found in the current literature. Various types of digital forensic investigations were discussed.

The goal of a digital forensic investigation was stated as legally acceptable evidence to be presented at an appropriate forum. The who, what, where, when, why and how that an investigator must ask continually during an investigation were briefly discussed.

Chapter 3 deals exclusively with digital evidence.

Chapter 3

Digital Evidence

*Is this a dagger which I see before me,
The handle toward my hand? Come, let me clutch thee;
I have thee not, and yet I see thee still.
Art thou not, fatal vision, sensible
To feeling as to sight? or art thou but
A dagger of the mind, a false creation,
Proceeding from the heat-oppressed brain?
I see thee yet, in form as palpable
As this which now I draw.*

Macbeth, Act II, scene i – Shakespeare

Out, damned spot! out, I say!

Macbeth, Act V, scene i – Shakespeare

3.1 Introduction

Chapter 2 introduced terminology related to digital forensic investigations and ultimately the goal of such an investigation; digital forensics was defined for the purposes of this dissertation. A number of aspects of this definition still need further discussion. The current chapter will address some of the aspects that were only briefly mentioned as part of the definition of digital forensics.

The definition of digital forensics as defined in Section 2.5 is repeated here for the sake of convenience:

Digital Forensics is a specific, predefined and accepted process applied to digitally stored data or digital media using scientific proven and derived methods, based on a solid legal foundation, to extract after-the-fact digital evidence with the goal of deriving the set of events or actions indicating a possible root cause, where reconstruction of possible events can be used to validate the scientifically derived conclusions.

Chapter 4 discusses the specific, predefined and accepted processes in detail.

The current chapter will examine the aspects of the definition as indicated. Section 3.2 describes where digital evidence is found, defines digital evidence and lists general characteristics of digital evidence. Metadata also aids an investigator to successfully produce relevant digital evidence; hence this concept is briefly discussed in Section 3.3. Digital evidence is usually used to find the root cause of an incident and depending on the seriousness of the incident, the evidence may need to be presented in a court. Section 3.4 discusses some legal principles that investigators must regard during an investigation. Section 3.5 briefly investigates best practice and standard operating procedures.

3.2 Digital Evidence

A digital forensic investigation aims to produce *digital evidence*. This digital evidence can be used towards constructing a possible set of events or actions resulting in such digital evidence. The current section will discuss where digital evidence is found, define it and give some insight into digital evidence characteristics.

3.2.1 Where Digital Evidence is found

According to the definition of digital forensics, scientific methods and tools are applied to digital media or digital data. Digital evidence is therefore found on digital or electronic devices [27]. Digital evidence is also resident on digital media. Digital media typically include DVDs, CDs, flash memory technology and digital tape drives, but many more

technologies can be listed here. A complete listing is given in the First Responders' Guide [50].

3.2.2 Defining Digital Evidence

Various definitions for digital evidence is found in the literature [25, 50, 44, 11, 117, 44, 27, 74, 14, 21]. Three definitions of the selected list are discussed further in this section.

Digital Evidence 1: Digital evidence is data or information stored on or transmitted by a digital device [50, 44, 11, 117, 44].

Definition 1 above lists data and information as an element of digital evidence. For the purposes of this dissertation the definitions of data and information are taken from the South African *Electronic Communications and Transactions Act* (ECTA) [87]. ECTA defines data as the electronic representation of information in any form [87]. Information is specifically defined as any data generated, sent, received or stored by electronic means [87]. Digital evidence therefore is data or information.

Digital evidence is found on [117, 50, 74] and produced by [44] digital devices. Various devices are listed in the literature and include a wide variety of devices that can manipulate information [117, 50].

Digital devices have the capability to produce information. Although definition 1 limits the processes to storage and transmission, ECTA also includes generation and reception of data. Cohen states that a sequence of *events, processes or actions* initiated on a digital device can change the state in which data is stored, transmitted, stored or received [25].

Definition 1 indicates that digital evidence is data or information found on a digital device, and that the device is capable of manipulating the data or information. Definition 1 therefore defines digital evidence in its simplest form.

Digital Evidence 2: Digital evidence is data that supports or refutes a hypothesis formulated during a digital forensic investigation [14].

Digital evidence includes an hypothesis as a construct of the definition. Carrier proposes that the investigator formulates a hypothesis at the initial stages of the investi-

gation to give the investigation an objective. Evidence found in support of the hypothesis should be admitted in court. This concept will be expanded on in definition 3.

Digital Evidence 3: Digital evidence is information of probative value stored or transmitted in binary form that may be relied upon in court [27, 74].

Definition 2 includes *probative value* and *reliance in court* as elements of the definition. In the FORZA model proposed by Ioeng, digital evidence presented in court must be reliable and relevant to the legal question [54].

Reliability of digital evidence is maintained by the chain of custody and by adequately addressing data integrity. According to Ioeng, relevance affects the weight of digital evidence once successfully admitted to court [54]. Ioeng further introduces the concept of reconnaissance where a variety of different methods, practices and tools are developed to produce reliable digital evidence. Reconnaissance as a concept of the scientific method is discussed further in Section 3.4.4.

Weight and admissibility should be seen within the context of relevance. Digital evidence is relevant when it specifically deals with the legal question that needs to be answered. Stated slightly differently, digital evidence is used to determine the outcome of the legal question once the relevance is determined [25].

Once evidence is relevant in finding a possible explanation for the root cause, admissibility is determined by arguments of the litigants based on various legal rules and principles [92]. An investigator must ensure that the digital evidence is reliable by following an acceptable procedure that is accurately documented. If the investigator follows an acceptable procedure and makes the necessary investigation notes, there is no need for concern about the legal principles and rules that the litigants will argue. The *weight* that the digital evidence bears is determined by the court [92] and has probative value which the court will use to determine the outcome [25]. Weight and probative value are similar concepts.

The investigator has a direct influence on the relevance of the evidence, whereas the other aspects are not within his control. The investigator can however indirectly influence weight and admissibility by following a well-documented procedure during the investigation.

All the above factors – reconnaissance, relevance, admissibility and weight – are used to determine the finding of the court.

The definition for digital evidence used for the purposes of this dissertation is formulated as follows:

Digital evidence is data or information of probative value stored on or transmitted by a digital device that supports or refutes a hypothesis formulated during a digital forensic investigation relied upon in court to determine the outcome of a legal question.

The following section introduces some characteristics of digital evidence.

3.2.3 Characteristics of Digital Evidence

Il est impossible au malfaiteur d'agir avec l'intensité que suppose l'action criminelle sans laisser des traces de son passage.

It is impossible for the criminal to act, especially considering the intensity of a crime, without leaving traces of his presence.

Locard [23]

This section gives some indication of the nature and characteristics of digital evidence. These characteristics can aid and challenge investigators during an investigation.

Locard's Exchange Principle

Edmond Locard postulated the principle that when two items make contact, there will be an exchange [23]. The Locard principle is often quoted in the forensic sciences and is applicable in digital forensics investigations [53].

When an incident occurs, a perpetrator will leave trace evidence and remove trace evidence at the scene. This exchange is known as the Locard exchange principle. Several techniques have been developed in conventional forensic sciences to successfully prosecute criminals. Techniques used include blood analysis, DNA matching and fingerprint testing. These techniques are used to confirm the presence of a suspected person at

a physical scene. Culley applies this principle when he states that where there is an interaction with a computer system, traces will be left [29].

Digital Stream of Bits

Cohen refers to digital evidence as a *bag of bits*, which in turn can be organised in sequences to represent information. The information in sequential bits will seldom make sense and tools are required to represent these structures logically in a human readable form [25].

The context in which digital evidence is found also aids the investigator during the investigation. Metadata is used to describe data more accurately and is helpful in determining the context of digital evidence. The next section introduces and discusses metadata characteristics that are useful in an digital forensic investigation.

3.3 Digital Evidence and Metadata

Metadata describes digital data stored on digital devices [114]. Metadata increases the information about the content and even the context in which data is found on a digital device. The bits of information are typically structured as files and folders. These files and folders can be viewed by various tools available to an investigator.

Metadata aids an investigator during an investigation to make a number of inferences about the data found on a device. The metadata can give an indication of

- when the file was created, accessed and modified, often referred to as the Modification, Access and Creation or MAC times associated with the file;
- where the file is stored, by the file path;
- what the function of the file is by examining the extension and header data in the file;
- the owner and creator of a file;
- size of the file, which can be an indication that it is not consistent with a known file; and

- hash signature by which the data can uniquely be identified.

The above list of metadata attributes is used in this dissertation to determine which data is unique on a device by comparing it with a list of known files. The unique user data can then be attributed to a specific user or possible suspect. The unique history of the file over time can also be tracked, provided that enough metadata is collected during an investigation.

File metadata includes creation, modification and access times. Within the digital forensics community this specific metadata cluster is known as MAC times [38]. The creation time of a file gives some indication of the time the file was created on the system. The access time indicates when the file was last viewed and the content of the file remains unchanged. The modification time indicates that the content of the file has changed at a specific time.

A file might be an evidence exhibit in a case where time is critical. Since an investigator may not always find the required *time* information as part of the file content, this is where the file metadata may be helpful. Once the metadata has been examined, an investigator could conclude that the file was created on a specific date or not.

Time-based analysis as proposed by Patel [46] makes extensive use of file metadata to reconstruct a possible sequence of events.

Possible problems an investigator might face is that the metadata is not a true reflection of sequential events found on a device. Firstly, the metadata may be manipulated making it difficult to draw accurate conclusions on a specified file's time attributes. Secondly, within a networked environment, time synchronisation on the connected devices might not be the same and may affect the accuracy of metadata transmitted over a network.

An operating system tool such as the *list* command in Linux can be used to find metadata. The Linux manual pages specify that using the *ls -l* command, i.e. list with the long format option, displays the following metadata attributes for each file: file permissions, number of links, owner name, group name, number of bytes in the file, abbreviated month, day of month file was last modified, hour file last modified, minute file last modified, and the path name [5].

The original source of information can be determined if enough metadata of the file

is known. Pollitt [27] draws attention to the meaning of a *copy*, *duplicate* and *original* of digital evidence. These terms are briefly described here:

- *Original* digital evidence is defined as those physical and logical data objects associated with the items at the time of seizure [27].
- A *copy* of data is an accurate reproduction of information contained in the data objects independent of the original physical item. A copy stored in RAM will be a copy of an original found elsewhere [27]. Confusingly, this *original* data is stored on the same digital device *or* another digital device where it was previously created or transmitted from.
- *Duplicate* digital evidence is an accurate digital reproduction of all the logical data objects contained on the original physical item. The reproduction is independent of the physical item or device [27].

3.4 Legal Principles of Digital Evidence

In the definition given in Section 2.5, legal principles should be considered. This section briefly lists some aspects that an investigator should consider when conducting a digital forensic investigation.

Culley states that courts, tribunals and various other committees increasingly require digital evidence. The reason provided is the fact that many aspects of our lives have become dependent on computers [29]. Therefore, it is important for the investigator to be aware of the legal requirements that are generally accepted within the legal fraternity. Ignoring legal procedures as these listed here will generally result in evidence being inadmissible [3, 69, 53].

3.4.1 Circumstantial and Hearsay Nature of Digital Evidence

Digital evidence is circumstantial and hearsay in nature [25, 21, 53]. Within the legal context, these characteristics must be considered by the investigator. Direct evidence

establishes a fact, where circumstantial evidence may suggest a fact [21]. Circumstance is also often referred to as context in the literature.

The hearsay nature of digital evidence can be explained by the following illustration. An eyewitness can usually confirm a fact. The investigator discovers the traces of the actual event, but does not see the actual event. The investigator can only establish the evidence found and a possible explanation. An exact reproduction of actions to produce the digital evidence may not be possible. The author or creator of the digital evidence may not be present in court. In such cases the evidence is hearsay in nature and the original is not re-produceable [92].

The context in which the digital evidence is found will also enable the investigator to make certain inferences [25]. Digital evidence is seen as an abstract of some event if the context is removed [21, 53]. Cohen formulates the following two rules about digital evidence where the context cannot be determined [25]:

1. An investigator cannot place a person at a specific location in time and place to prove that something happened in the physical world; and
2. digital evidence traces show that a certain sequence of events are consistent with certain traces resulting from a specific sequence of events.

Cohen's second rule poses a challenge to the investigator. The challenge is that a number of different sequences of events can result in a specific state on a device. A simulated sequence of events might not be the same as the actual sequence of events that caused the state being investigated [85].

Authorisation and authenticity of digital evidence are essential in a successful digital forensic investigation.

3.4.2 Authorisation to Conduct a Digital Forensic Investigation

Authorisation is required from an appropriate authority before an investigation can be initiated [56]. The appropriate authority can be obtained from company management when an internal disciplinary investigation is conducted. In extreme cases a court will have to issue a search warrant. In some instances the permission of the owner of a device

may be sufficient [22]. Depending on the seriousness of the investigation, there should be a legal basis for the collection of digital evidence before such evidence can be collected.

3.4.3 Authenticity of Digital Evidence

Digital evidence that is extracted from a digital source must be authentic. Data that is extracted from an original source is authentic if it can be verified as authentic. Digital evidence is verified by hashing the data content [12]. The hash value will ensure the authenticity of digital evidence required during later stages of the investigation and presentation at trial [50].

The legal requirements on authenticity that must be satisfied are the following [53]:

- The content of the data must be unchanged.
- The data is a duplicate from an determinable original source.
- The metadata presented should be accurate.

The legal requirements can easily be achieved by using technological features provided by digital forensics tools. MD5 or SHA-1 are generally used as technology to authenticate data within these tools. Pollitt lists the following requirements to verify the authenticity of the original image against a processed image [27]:

1. The duplicate image must be a true and accurate representation of the original.
The person who captured or is present during the imaging process must testify that the original has been imaged.
2. The imaging process of the original must be documented in a log.
3. The imaging process must provide sufficient detail for an equally competent investigator to reproduce the process and reach the same conclusion.

3.4.4 Scientific Method

Another requirement of the digital forensics definition given in Section 2.5 involves the scientific method used. The scientific methods used to produce digital evidence must be trusted. There should be some proof that the method or tool used is reliable and works.

Digital evidence is usually introduced by expert testimony [25]. Expert testimony is introduced except where non-experts can bring clarity to non-scientific issues. Issues based on scientific, technical or specialised knowledge can only be addressed by experts. This general rule is contained in the Federal Rules of Evidence as used in the United States of America [25].

The next section briefly introduces the rules surrounding expert testimony, where scientific method is applied to produce such digital evidence testimony. The rules were formulated in the cases of Frye [22] and Daubert [25].

Digital Evidence Principles as Formulated in Frye and Daubert

Various tools and technologies have been developed by investigators to successfully present digital evidence in court. The tools and technology are developed through scientific methods based on the requirements of individual cases. These methods are refined and subsequently used in other investigations. Each investigation poses new challenges to investigators, who then have to develop new tools to enable successful production of digital evidence. When developing new tools or presenting evidence in court an expert should be aware of the following requirements that guide the courts in reaching a finding.

These requirements specifically deal with instances where scientific testimony is presented by experts using some form of scientific method [21, 22, 25].

In *Frye v United States* [22] the court had to decide on the admissibility of a polygraph test as evidence. The court concluded that testimony given by an expert must have a scientific basis that is sufficiently established and accepted.

The rule formulated in *Daubert v Merrell Dow Pharmaceuticals (92-102)*, 509 U.S. 597 (1993) [25] is accepted as a fundamental principle in digital evidence presentation by an expert in court [12, 25, 22]. Where a *scientific method or technique* is used to produce digital evidence by an expert in the United States, the following requirements have to be met when applying this *unproven technique*:

- The theory, method or technique can be and has been tested.
- The theory, method or technique has been subject to peer review and publication.
- Error or potential error is known.

- The theory is generally accepted in the scientific community.
- Testimony of the expert is based on some special skill.

3.4.5 Best Evidence Rule

The original evidence is preferred when evidence is presented in court. The best evidence rule is formulated to be the original writing [25].

In South Africa, digital evidence is classified as documentary evidence [92, 53]. Documentary evidence requires that the evidence produced must be relevant, admissible and authentic [53, 92]. A document is defined as any written thing capable of being evidence, including a printed document [92].

Where documentary evidence is tendered to prove the truth of the contents, it may amount to hearsay. So the document remains circumstantial and hearsay until supported by oral evidence in support [53].

The best evidence rule provides that if a party wants to rely on evidence in a document, the following three requirements have to be satisfied [92]:

- The contents of the document may be proved only by the original, which is the best evidence.
- The authenticity of the document must be proved to satisfy the court about its originality.
- The document must duly be certified as original.

ECTA specifically provides that the rules of evidence must not be applied to deny the admissibility of a data message because it is not in original form, if it is the *best evidence* the producing party can obtain [87].

Section 3.4 has dealt with the basic minimum requirements against which the courts will evaluate digital evidence to be admitted. Digital forensic investigators have to be conscious of the legal principles listed above to ensure that evidence is readily admitted when presented in court. In cases where an investigator acted without authorisation, the admissibility of the evidence will be challenged by the opposing party. In cases

where unsound scientific methods are used to extract evidence, the admissibility will be challenged. Where the authenticity of the evidence cannot be proved the evidence will not readily be admitted without challenge from the opposition. Evidence may be challenged on various other grounds if the above requirements are not met.

3.5 Best Practice and Standard Operating Procedures during Investigations

Digital forensic investigators are often faced with the question of having to explain *how* the digital evidence was collected. The evidence presented in court is not challenged as often as how the evidence was found. Over time, investigators have formulated standard operating procedures and best practice guidelines when dealing with digital evidence.

The Cambridge Online Dictionary defines best practice as *a working method, or set of working methods, which is officially accepted as being the best to use in a particular business or industry, usually described formally and in detail* [75]. Standard operating procedures (SOPs) are a set of written instructions documenting a repetitive activity followed by an organisation [72, 22]. In the case of a digital forensic investigation, SOPs must be performed every time a device is investigated [22]. A best practice will develop into a standard operating procedure over time.

Ruiben states that every piece of digital evidence should be challenged to ensure that an investigator followed a rigorous process [91]. The method or process is constantly emphasised as critical during a digital forensic investigation. The best practice principles outlined are seen as a solid base from which specific standard operating procedures should develop.

Best practice and standard operating procedures form part of the digital investigation process that should continually be developed and tested with current and emerging techniques and technologies. They also form part of the investigation in that the process followed must continually be recorded during the investigation process, especially where a deviation from standard is performed.

3.5.1 Best Practice

Within the established literature, two documents are often referenced when dealing with best practice in the field of digital forensics. The first document is the *Good Practice Guide for Computer – Based Electronic Evidence* or ACPO guide developed by the Association of Chief Police Officers in the United Kingdom. This is often referenced as the source of best practice principles in the literature [50, 27, 22].

The second document is the *European Convention on Cyber Crime* or ECCC guide [27]. The best practice principles listed in the ECCC guide are similar to those in the ACPO guide [50]. The four best practice principles listed in the ACPO guide are the following [1]

1. Data contained on a device that will be relied upon in court is not to be changed by any action taken during an investigation.
2. In exceptional circumstances it may be necessary to access the original data on a device, in which case a competent person with the necessary expertise must do so and will be liable to explain the relevance and implications of their actions.
3. A clear audit trail, log, chain of custody or other produceable record of all processes applied to digital evidence should be created and preserved separate from the investigation process. The processes should be reproducible by an independent third party to give the same result.
4. The investigating officer in charge of the investigation is responsible for ensuring that the law and these principles are adhered to [29].

The first two best practice principles above deal with data integrity, and mention specifically that the data must not be changed. Where the data is changed, there should be a good reason for the change or necessary exposure to risk. A person dealing with digital evidence should be appropriately qualified [27]. Pollitt states that the use of the word evidence suggests that the investigator is a person who is recognised by the courts [27]. This indicates that the investigator not only has to be qualified and competent to conduct an investigation [21], but must have knowledge of the nature of digital evidence.

The third best practice principle requires an investigator to make investigative notes. These notes should reflect how the investigator conducted the investigation. There are various ways in which an investigation log can be created, and some of these are included in tools such as Encase. The process followed to produce digital evidence must be fully documented, preserved and available for review. These documents include the chain of custody and investigation notes compiled by the investigator [27].

The important aspect of best practice principle three is the requirement of reproducibility. Another investigator following the same steps as the original investigator should reach the same conclusion. The chain of custody is enforced by placing the responsibility for the evidence and any actions taken on such evidence on the person in possession of such evidence [27]. Casey proposes that best practice should be developed to become standard operating procedure to enable the second investigator to reach the same conclusion as the first investigator [22].

The last best practice principle is broad in scope. The requirement is to be aware of the law and enforce these best practice principles. A limited set of legal principles that investigators should be aware of were discussed in Section 3.4.

The principles served as a departure point in many documents that have subsequently developed. Among these documents are the *Report on Digital Evidence* compiled by Pollitt.

The best practice guidelines above indicate how an investigator should deal with digital evidence at all times during an investigation. The following section describes the standard operating procedures that have developed over time when dealing with digital evidence.

3.5.2 Standard Operating Procedures – SOPs

Procedures are the step-by-step guide followed by the investigator during an investigation [22]. The Standard Operating Procedures (SOPs) that an investigator follows should be documented within the organisation conducting investigations on a regular basis. The prescribed document gives an outsider a view of what the SOP is when dealing with digital evidence in a specific instance.

SOP documents must be compiled for organisations working in the field of digital

forensics or conducting investigations on a regular basis. The SOP documents must be compiled and revised for procedures regularly conducted by investigators [27]. The documents must also be revised regularly to ensure legal compliance and technological developments.

Data collected during an investigation in a scientific manner must use acceptable methods, such as the rule formulated in the case of Daubert.

The National Institute of Justice in the United States suggests that a SOP compiled by an investigator should contain a title, purpose, equipment, materials, standards and controls [27].

Procedures including a step-by-step description of how the investigation is conducted must be fully documented in a SOP document. Appropriate measures must be put in place to prevent the degradation of original digital evidence collected.

3.6 Conclusion

This chapter dealt with various aspects of digital evidence. An investigator dealing with digital evidence must be aware of what digital evidence is, where it is found and how it should be dealt with.

Digital forensics is a young science and remains relatively new to both scientists and lawyers. Developments in the legal and scientific communities are important to lawyers and investigators alike. The communities should be aware of these developments and must readily accommodate the changes. Many unsolved technical, legal and even ethical issues will continually challenge both these communities [78]. These challenges will only be addressed in time as the issues become known through appropriate communication.

Best practice and SOP are mere guidelines for investigators. These documents however give insight into how an investigator conducted an investigation. Chapter 4 examines a limited set of proposed digital forensic process models.

Chapter 4

Digital Forensic Process Models – DFPMs

Though this be madness, yet there is method in't.

Hamlet – Shakespeare

4.1 Introduction

In Chapter 2 applicable terminology in digital forensics was introduced and discussed. The focus of Chapter 3 is on digital evidence. This chapter will introduce and deal with various process models known in the current literature.

Digital forensic process models (DFPMs) feature extensively in law enforcement and academic discussion. Terminology used in the processes will be listed and briefly defined where new terms are introduced.

The ACPO [74] process model is used in digital forensic investigations. This DFPM includes the recovery phase, collection phase, examination phase, analysis phase and report or statement phase. The ACPO DFPM is based on the earlier work of existing process models and the processes in the DFPM are only briefly described. The ACPO DFPM is not represented in a formal notation.

A number of model representations were considered to represent each model uniformly. These include UML Activity, Use Case Diagrams [59] and Finite State Ma-

chines [20, 25]. For the purposes of this study, the ordering of the events is considered to be important. Process modelling was also considered, but a simple mathematic notation is used as formulated by Moore and Mealy [116]. The motivation for sequential logic is given in Section 4.2. Each DFPM is shown in a notation similar to sequential logic. This is done to identify similarities and differences within the sequence of tasks required for conducting a digital forensic investigation.

Each DFPM will therefore be introduced, the process described in sequential logic notation and the terminology defined. Differences and similarities of DFPM will be listed, compared and commented on as the DFPMs are introduced.

Each DFPM uses specific terminology, and the original terminology is described, after which the terminology will be standardised to avoid any confusion. As an illustration some DFPMs are referred to as *Frameworks*, others as *Models*. In this chapter the terms *DFPM* or *Process* will be used, while *Framework* constitutes the subject matter of Chapter 6.

The dissertation will discuss the following selected DFPMs: Lee [61], Kruse and Heiser [60], NIJ Electronic Crime Scene Investigation [50], NIJ Forensic Examination of Digital Evidence [74], Casey [21], DFRWS [77], Reith [84], Carrier and Spafford [18], Baryamureeba [7], Ciardhuáin [24] and Cohen [25]. These eleven DFPMs are well known within the research community and widely used in academia and digital forensic investigations.

4.2 Motivation for Using Sequential Logic

Although various alternative representations were mentioned in Section 4.1, sequential logic is used in this chapter. In terms of sequential logic, the evaluation of the current state is determined by the previous state *and* the current state. This can be explained by looking at the order of two processes in a digital forensic investigation. Data extraction from the digital media must be completed before evidence can be discovered during examination. In short, data must be extracted before examination can start. Sequential logic requires the ordering of processes to be considered as this example illustrates.

4.3 Process Representation in Sequential Logic

The notation below is deduced from the description of sequential logic [116]. The values have been directly replaced with the process steps. The sequential logic used in this chapter is known as the Mealy machine [116]. The Mealy machine is a sequential logic circuit where the output is dependent on the input and the current internal state. For the circuit to render a *true* value, all the conditions in a specific order must also be *true*. Therefore the circuit representing all the conditions will only be valid if all the conditions are satisfied.

The sequential logic circuit is given as follows:

$$\langle x \rangle = \langle x_1, x_2, x_3, \dots, x_n \rangle \text{ where } x_i \in \{0, 1\}$$

x_i is the set of conditions part of the circuit, evaluating either to *true* or *false*. The following additional conditions also hold:

$$z = \varphi(\langle x \rangle) \text{ so that } z = 1$$

For any set of conditions used to evaluate to *true*, $z = 1$, all the conditions of the circuit have to evaluate to *true*. φ can be replaced by any circuit criteria or set of conditions.

If all the conditions within a circuit evaluate to *true*, the ordering has occurred in the listed order of the conditions contained in the circuit. Therefore the following also holds:

$$(x_1 \wedge x_2 \wedge \dots \wedge x_n) = 1 \text{ and } (x_i = 1) \prec (x_j = 1) \text{ for all } (i < j)$$

The above notation is used in the remainder of this chapter to enable an analysis of the various DFPMs by using a standardised notation. The various circuits or DFPMs are represented by φ . φ is replaced by the process name in the remainder of the chapter. x_i or each condition within the circuit is a single process within the DFPM. x_i is replaced by the current process name. The adaption is illustrated here as:

$$DFPM = \{start \Rightarrow next \Rightarrow then \dots end\}$$

\prec is replaced by \Rightarrow ; and $()$ is replaced by $\{ \}$. In certain instances sub-processes are also indicated, where the notation is adapted accordingly. \parallel indicates that a process occurs at the same time as another process. \Leftrightarrow is used in a single DFPM, which indicates that the previous process can be repeated after the current process has been executed.

4.4 Lee

The forensic scientist Henry Lee formulated a Scientific Crime Scene Investigation model [61]. This model is formulated to accommodate investigations that use forensic science. It does not find specific application within the digital forensics field, but is a good departure point nonetheless.

4.4.1 Process

The process can be represented as follows:

$$\text{Lee} = \{ \textit{Recognise} \Rightarrow \textit{Identify} \Rightarrow \textit{Individualise} \Rightarrow \textit{Reconstruct} \}$$

where

$$\textit{Recognise} = \{ \textit{Document} \Rightarrow \textit{Collect and Preserve} \} \tag{4.1}$$

$$\textit{Identify} = \{ \textit{Classify} \Rightarrow \textit{Compare} \}$$

$$\textit{Individualise} = \{ \textit{Evaluate} \Rightarrow \textit{Interpret} \}$$

$$\textit{Reconstruct} = \{ \textit{Reconstruct} \Rightarrow \textit{Report and Present} \}$$

4.4.2 Terminology

The terminology as described in the work of Lee is listed below.

- *Recognition* is where items or patterns are seen to be potential evidence. The investigator must know what to look for and where to find it. This process has two sub-activities, namely *documentation* and *collection and preservation*. Documenting evidence is an important aspect of an investigation, where any action by any person is clearly documented. Collection is where the evidence is collected from

the crime scene, thereafter *bagged and tagged*. Digital evidence must be preserved once safely contained.

- *Identification* of the various types of evidence is done next. Evidence is classified and compared, usually into categories such as physical, biological, chemical and other standard types [24].
- *Individualisation* is where evidence is linked to a particular individual or event. The evidence is then evaluated and interpreted.
- *Reconstruction* is where evidence objects and events are linked to account for a possible event sequence. During reconstruction, possible event sequences are reported and presented [61].

4.4.3 Comments

Lee deals specifically with the crime scene investigation and not with the entire investigation process [24]. He does not include a detailed approach applicable to electronic crime scenes [24], but advocates that the investigation must be systematic and methodical. The model is primarily aimed at physical evidence, but can be adapted to include digital evidence at a digital crime scene investigation. The physical evidence is usually processed for trace evidence such as blood and DNA forensic analysis, where trace evidence on digital media is the data itself.

In Equation 4.1 there would be little change in the process formulation. The procedure dealing with digital evidence will have to be included as a sub-procedure. *Recognise* will have to include a digital evidence type such as CD-ROMs, DVDs, laptops and other possible incident-related objects. Once these evidence exhibits have been identified as a possible medium enabling an incident, they will have to be *classified* accordingly. A suitable digital forensic examiner will also have to be notified to report after the digital forensic examination is complete.

4.5 Kruse and Heiser

The model proposed by Kruse and Heiser in 2001 consists of three main processes within a framework [60]. This DFPM is used as a basic starting point from which the processes can be expanded, depending on the circumstances of the investigation. Every action taken during the investigation must be thoroughly documented. Documentation is seen to be a continuous process during the entire investigation.

4.5.1 Process

$$\text{Kruse and Heiser} = \{Acquire \Rightarrow Authenticate \Rightarrow Analyse\} \quad (4.2)$$

4.5.2 Terminology

Kruse and Heiser refer to a Computer Forensic Investigation Framework [60]. The framework consists of the following listed and defined processes:

- *Acquire* the evidence without altering or damaging the original.
- *Authenticate* the acquired evidence against the original image.
- *Analyse* the evidence image without modification [60].

4.5.3 Comments

The Kruse and Heiser process is a high-level abstraction of a basic DFPM and closely resembles the best practice enunciated by AOCF [1]. During acquisition, easily explainable steps taken must be documented. Kruse and Heiser include storage and transport under evidence acquisition. Evidence is authenticated by an integrity calculation, such as a hash function and a timestamp. Common hash functions used are MD5 and SHA1 [60].

The basic Kruse and Heiser process structure is recognisable in many of the DFPMs, such as in the models of NIJ [50], Reith [84] and Carrier [7]. The reason for this is that they contain the initial steps taken in many digital forensic investigations. However, the *analysis* is lacking much needed detail in many respects. A final report on the investigation is also lacking. The DFPM can be slightly modified to remedy these defects.

$$\text{Modified Kruse and Heise} = \{Acquire \Rightarrow Authenticate \Rightarrow Examine \Rightarrow Analyze \Rightarrow Report\} \quad (4.3)$$

4.6 NIJ Electronic Crime Scene Investigation

The United States National Institute of Justice (NIJ) created a working group to prevent cyber crime. The working group is known as the Technical Working Group for Electronic Crime Scene Investigation (TWGECSI). In 1999 an initial planning draft was compiled by the National Institute of Science and Technology (NIST). The final draft was approved by the working group members in November 2000, and published in 2001 [50].

The NIJ document lists the four basic steps to a forensic investigation listed in Equation 4.5. The primary objective of the NIJ document is to effectively capture the essential features of any general electronic crime scene and specifically to aid first responders when encountering a digital crime scene to ensure evidence admissibility in court.

4.6.1 Processes

The document lists two separate processes that should be completed during an Investigation. The first responder will complete the following steps at the crime scene:

$$\text{NIJ First Responder} = \{Recognise \Rightarrow Document \Rightarrow Collect \Rightarrow Package\} \quad (4.4)$$

The main DFPM specifies the following steps to be included:

$$\text{NIJ Investigator} = \{Collect \Rightarrow Examine \Rightarrow Analyse \Rightarrow Report\} \quad (4.5)$$

4.6.2 Terminology

The procedure terminology for the first responder is listed first:

- The first responder must secure the crime scene first and foremost [50]. This must be done to avoid any further or later contamination of potential digital evidence.

The safety of all personnel must be ensured. Any person or thing that can harm the crime scene should be removed. Once the crime scene is secured potential evidence must be *recognised* and identified. Information known at the time the first responder is notified will aid the responder in knowing what to seize as specified in the warrant [22]. A plan of action is formulated to ensure that the integrity of the evidence is maintained. A brief preliminary interview is conducted with possible witnesses and/or suspects.

- The crime scene must be *documented*. This will include notes, diagrams, photographs, video recording detailing the complete crime scene without touching anything surrounding the crime scene. Writing impressions on note pads, paper shredders and other destructive devices may be in the proximity of the crime scene. Trace, biological and latent prints near the crime scene must also be documented. Any process that will potentially damage the digital extraction of evidence must be avoided, such as chemicals. If the computer is on, leave it on, if it is off, leave it off. An appropriate digital forensic examiner is to be consulted in such a case. This is a difficult situation in which to make a correct decision.
- The evidence is *collected and preserved*. Collection and preservation done by the first responder is of a physical nature, for instance, the computer is collected at the scene and preserved. Preservation is where the evidence is isolated to ensure that it is not contaminated [22].
- The physical evidence is then *packaged* and transported to a suitable location where the investigation will be completed. The location is usually a digital forensics laboratory.

At this point, the terminology below would seem repetitive, but this procedure will take place in a digital forensics laboratory. This is the second procedure described in the NIJ document. The tasks are specifically to be conducted after the physical evidence has arrived at the location where the investigation was completed. Each of the four listed steps has a number of sub-procedures to be completed. These steps are typically completed by the digital forensic examiner and not by the first responder.

- *Collection* includes the searching for and the recognition, collection and documenting of digital evidence. The evidence is extracted from the physical medium in the form of data. The collection may be in real time. If this occurs, then it may be classified as live forensics. Digital evidence that may be lost, such as data stored on volatile memory modules [43], must also be included.
- *Examination* is the phase during which the evidence is made visible. The origin and significance of the evidence is determined. The objective of the examination is to aid in document discovery and the compilation of a complete evidence repository. Any hidden or obfuscated data is made visible before reconstruction can begin. This phase is to be completed by a forensic practitioner with the appropriate qualifications [50]. What the qualifications should be is not mentioned in the NIJ document. Examination is seen as a technical review done by the forensic examiner.
- During *analysis* the product of the examination phase is scrutinised. The evidence significance and probative value are determined. These are evaluated to aid in evidence presentation. The agency or investigative team as a whole is to aid in the analysis phase, thus the whole investigative team must be present.
- The *reporting* phase includes a detailed outline of the examination process and a complete listing of the data collected. Notes, including all documentation, must be preserved for discovery and subsequent testimony.

4.6.3 Comments

The NIJ Electronic Crime Scene Investigation proposed a DFPM for electronic crime scene investigations. The guide lists several processes. Firstly, the first responder must be trained to recognise, document, collect and package possible evidence at a crime scene. Secondly, the digital forensic examiner or investigation process is listed. The first responder and investigator will generally not be the same person, therefore the DFPM authors are conscious of the practical execution of an investigation.

There seems to be differing opinions on the meanings of *examination* and *analysis* in the literature. According to the Oxford Dictionary, *examination* includes the notion

of a *detailed inspection* [101]. The same Dictionary defines *analysis*, among others, as a *detailed examination of the elements or the structure of something* [101]. The definitions are similar. Therefore, *examination and analysis* is defined as a *detailed inspection of the structure*. Analysis subdivides the bigger problem into smaller components. The NIJ document states that the purpose of the examination is to make the evidence visible and to explain its origin and significance [50]. The purpose of the analysis is to look at the product of the examination for its probative value and significance to the case [50]. It is similar to the adage *divide and conquer*.

The guide is not intended to detail analysis techniques. Significantly, the guide states that local or domestic legislation and regional rules pertaining to evidence must be studied and included.

In Chapter 2 the first responder and digital forensic investigator or examiner are listed as important people in a DFPM.

4.7 NIJ Forensic Examination of Digital Evidence

In April 2004 the NIJ compiled a special report on Forensic Examination of Digital Evidence as a guide specific for law enforcers. The recommendations in the report should not be seen as policy or legal mandates. The report is also not the only correct course of action but a guide compiled by a committee with experience in the field of digital forensic examinations. The motivation for compiling the report lies in the wave of electronic crime dealt with by law enforcement agencies, financial institutions and firms with a digital forensics infrastructure capability. The report should therefore be seen as a general guide in the development of policies and procedures [74].

The guide emphasises the need to be ready and have forensic procedures and protocols in place in the event of being confronted with electronic media containing incriminating evidence.

The guide addresses investigating high-technology crimes, creating a digital forensic suite and presenting digital evidence in a court room. It was also written as generic as possible to ensure that past, present and rapidly developing technological developments are included. It should however be read with current technological restraints and gen-

erally accepted practices in mind. The focus is also on common situations encountered during a digital forensic examination and investigators are reminded to follow the general principles and procedures.

4.7.1 Process

The process suggested in the forensic examination guide is more focused than that in the first responders guide.

$$\text{NIJ Examination} = \{ \textit{Policy and Procedure Development} \Rightarrow \textit{Assessment} \Rightarrow \textit{Acquisition} \Rightarrow \textit{Examination} \Rightarrow \textit{Document and Report} \} \quad (4.6)$$

4.7.2 Terminology

The terminology in the digital forensic examination guide is also more specific than in the first responders guide. For the sake of brevity, examiner is taken to mean digital forensic examiner.

- *Policy and Procedure Development* is required for the establishment and operation of a digital forensics unit which should include a mission statement, personnel requirements and administrative considerations, such as training, software licensing and resource commitment.
- *Assessment* is where the examiner assesses the scope of the case thoroughly and determines an appropriate course of action.
- *Acquisition* - digital evidence is fragile by nature and can be altered and damaged or even destroyed by improper handling or examination. Therefore examination of evidence should be conducted on a copy of the original evidence, and the original evidence should be acquired in a manner that protects and preserves the integrity of the evidence.
- *Examination* is to extract and analyse digital evidence, where extraction is the process of recovering data from the media. Analysis was already defined in Section 4.6.2.

- *Documenting and reporting* include the actions and observations documented throughout the DFPM. Depending on the type of investigation, the report will be written to list all facets of the investigation. The written report concludes with an objective opinion on the findings.

4.7.3 Comments

This DFPM includes an emphasis on preparation, which has been identified as forensic readiness by Tan [100]. Policies and procedures are encouraged to form part of the preparation process [74]. These should be developed with legal compliance in mind. Appropriate resources need to be identified before an investigation is initiated, therefore the acquisition cannot commence without some proper approach strategy. Considering the strict requirements of sequential logic, the investigation will fall short of meeting the stringent requirements of forensic presentation. For a DFPM to be successfully applied, each process must be completed as thoroughly as possible. The NIJ Electronic Crime Scene Investigation document also includes case examples and template documents to guide an investigator or Cyber Incident Response Team (CIRT) towards a successful investigation [74].

4.8 Casey

In 2000 Casey proposed a DFPM for processing and examining digital evidence. This DFPM can be applied to various investigations, including standalone computer systems and networked environments [24]. The Casey DFPM has been revisited in 2004 to include a number of other processes [21]. The 2000 Casey DFPM and subsequent improvements are all listed below.

4.8.1 Process

The 2000 Casey DFPM involves the following listed processes:

$$\text{Casey} = \{ \textit{Recognition} \Rightarrow \textit{Preservation} \Rightarrow \textit{Classification} \Rightarrow \textit{Reconstruction} \}$$

where

$$\textit{Preservation} = \{ \textit{Collect} \Rightarrow \textit{Document} \}$$

$$\textit{Classification} = \{ \textit{Compare} \Rightarrow \textit{Individualise} \}$$

(4.7)

4.8.2 Terminology

- *Recognition* will be where the investigator recognises a similar pattern that might have presented itself in the past. This is a form of investigator experience based on previous investigations and could include a database of previously solved investigations.
- *Preservation* consists of two sub-processes, *collect* and *document*, followed by *classification*.
- During *classification*, evidence objects are *compared* and *individualised*.
- Individualisation is where evidence is attributed to an origin or creator [25].
- During *reconstruction*, a possible event sequence of the crime is traced by reconstructing the possible event sequence that most accurately reflects the events as they could have happened during the actual crime or incident. Reconstruction is generally required to prove *how* a certain result is achieved for various purposes.

4.8.3 Comments

The DFPM by Casey is similar to the one originally proposed by Lee. The focus of Casey's model is on processing and examining the digital evidence [24], however analysis will differ substantially from physical analysis of non-digital evidence types. The first and last processes, which are *recognition* and *reporting* are identical in both these DFPMs. The 2004 model is significantly extended to include processes identified in the

development of various other DFPMs in the literature. The all-inclusive updated 2004 DFPM is given as follows:

$$\begin{aligned}
 \text{Casey 2004} = \{ & \textit{Incident Recognition} \Rightarrow \textit{Assessment} \Rightarrow \textit{Identification} \\
 & \textit{and Seizure} \Rightarrow \textit{Preservation} \Rightarrow \textit{Recovery} \\
 & \Rightarrow \textit{Harvesting} \Rightarrow \textit{Reduction} \Rightarrow \textit{Classification} \\
 & \Rightarrow \textit{Analysis} \Rightarrow \textit{Reporting} \} \tag{4.8}
 \end{aligned}$$

where

$$\begin{aligned}
 \textit{Preservation} &= \{ \textit{Collect} \Rightarrow \textit{Document} \} \\
 \textit{Classification} &= \{ \textit{Organise} \Rightarrow \textit{Compare} \Rightarrow \textit{Individualise} \}
 \end{aligned}$$

The influence of the Digital Forensics Research Workshop Group discussed in Section 4.9 is clearly seen here. The analysis of Casey offered by Ciardhuáin is somewhat too critical because Casey does not adequately address many of the critical changes [24]. The terminology used by Casey in this model seems to have a different meaning, but the following illustration can be seen as a refinement of previous models:

$$\textit{Examine} = \{ \textit{Recovery} \Rightarrow \textit{Harvesting} \Rightarrow \textit{Reduction} \Rightarrow \textit{Classification} \} \tag{4.9}$$

For the purposes of this dissertation, Equation 4.9 is of considerable importance. *Reduction* is the process where data is significantly reduced to attribute data to specific users. For the purposes of this improvement, *preservation* and *classification* are still deemed inclusive in the broader DFPM as in the 2000 DFPM. Casey also differentiates between temporal, relational and functional analysis in the 2004 model [21]. This dissertation will however not include an analysis on this level of granularity.

4.9 Digital Forensics Research Workshop Group - DFRWS

The Digital Forensics Research Working Group (DFRWS) developed a framework with the following listed processes: identification; preservation; collection; examination; analysis; presentation, and decision [77]. The process elements are the same as the definition used in this dissertation for Digital Forensics as seen in Chapter 2.

4.9.1 Process

$$\begin{aligned} \text{DFRWS} = \{ & \textit{Identify} \Rightarrow \textit{Preserve} \Rightarrow \textit{Collect} \Rightarrow \textit{Examine} \\ & \Rightarrow \textit{Analyse} \Rightarrow \textit{Present} \Rightarrow \textit{Decide} \} \end{aligned} \quad (4.10)$$

4.9.2 Terminology

Most of the terminology listed in the process has been defined above. The meanings of the terms or processes are the same as those already listed. A comprehensive exposition can be found in the original DFRWS paper [77].

4.9.3 Comments

In this DFPM, processes are termed tasks and individual tasks consist of elements. This DFPM is seen as an important foundation for much work within the digital forensics community. Two crucial stages of the investigation, namely *presentation* and a point of *decision* are included as processes [24]. The decision is seen as important because an investigator can decide whether the investigation has produced relevant evidence to prove an reported incident. Preservation in this DFPM addresses case management technologies, imaging technologies and time synchronisation. The chain of custody is also specifically mentioned and accommodated within the DFPM. This DFRWS DFPM can be applied to many investigations, including standalone digital systems and networked systems [93].

Many of the subsequent DFPMs, as the one seen here, follow a linear sequential model. This indicates that subprocesses are gradually being avoided. The influence of

this will be seen in the architecture proposed in Chapter 6.

4.10 Reith

The Abstract Digital Forensic Model (ADFM) was proposed by Reith, Carr and Grunch in 2002 [84]. Reith specifically explores the possibility of using traditional forensic techniques within the digital environment. Previous DFPM shortcomings identified by the authors are listed as:

- a standardised framework for digital forensic tool development;
- keeping the framework abstracted from technology to accommodate future technologies;
- a methodology used to enable members of the judiciary to relate technology to non-technical observers; and
- accommodating non-digital evidence within the abstraction [84].

4.10.1 Process

In Section 4.9 reference was made to the linear approach adopted by many of the DFPMs. The process list indicates this approach where the processes are *identification*, *preparation*, *approach strategy*, *preservation*, *collection*, *examination*, *analysis*, *presentation*, and *returning evidence* [84].

$$\begin{aligned} \text{Reith} = \{ & \textit{Identify} \Rightarrow \textit{Prepare} \Rightarrow \textit{Approach Strategy} \\ & \Rightarrow \textit{Preserve} \Rightarrow \textit{Collect} \Rightarrow \textit{Examine} \Rightarrow \textit{Analyse} \\ & \Rightarrow \textit{Present} \Rightarrow \textit{Return} \} \end{aligned} \quad (4.11)$$

4.10.2 Terminology

- *Identification* is where the type of incident is recognised from possible known indicators. These indicators are typically known to an experienced investigator.

- *Preparation* includes obtaining the correct tools, required techniques, search warrants, monitoring authorisation and management support.
- *Approach strategy* is dynamically formulated based on the impact on the bystanders and technology used. The aim is to minimise any impact on the victim and maximise the collection of the unattained evidence. This process draws inspiration from forensic readiness as defined by Tan [100, 82].
- *Preservation* is the process of securing, isolating and preserving the physical and digital state of the evidence.
- *Collection* is where the physical scene is documented and the digital evidence is duplicated using accepted standard procedures. Specific information is not given as to what standard procedure is preferred. Digital data is seldom extracted at the crime scene. The preferred method is to isolate the physical evidence and collect the digital evidence in a forensic laboratory.
- *Examination* is the detailed systematic search of evidence relating to the suspected incident. The focus is to locate and identify potential evidence. A detailed document is constructed for analysis.
- *Analysis* is where the significance of the examined evidence is explored. Fragments of data can be reconstructed, from which conclusions may be drawn. Reith indicated at this point that examination and analysis can be repeated to support the incident theory or hypothesis. Analysis is not as technically demanding as examination, hence more people can aid in proving the hypothesis.
- *Presentation* includes the summary of the investigation and the findings.
- Evidence must be *returned* to the proper owner after the investigation has been completed.

4.10.3 Comments

The shortcomings identified by Reith, Carr and Grunch have been addressed here and the listed advantages of the DFPM include the list in Section 4.10. This DFPM is broad

enough to include a physical device such as a printer or other devices collected from an incident scene and hence accommodates future technological developments.

The model is specifically abstracted from technology dependent procedures [24]. The model can be used in the development of more detailed methods required in specific investigation types. Baryamureeba states that the chain of custody is not mentioned in this model [7].

This framework can be used to categorise various incident types [93]. Selamat states that the approach strategy is a duplication of the preparation phase [93]. Responding to the incident will require some form of preparation, thus giving an indication of the approach strategy to use [7]. Notification of an incident also gives prior knowledge for which an investigator can prepare.

4.11 Carrier and Spafford

Carrier and Spafford propose a DFPM with five groupings and seventeen phases in total. The DFPM is named the Integrated Digital Investigation Process (IDIP). The groups are listed as readiness, deployment, physical and digital investigation with a review grouping [18]. This model is in contrast to the linear development approaches mentioned in previous sections.

4.11.1 Process

$$\begin{aligned}
 \text{Carrier and Spafford} = \{ & \textit{Readiness} \Rightarrow \textit{Deployment} \Rightarrow \textit{Physical} \\
 & \textit{Investigation} \parallel \textit{Digital Investigation} \\
 & \Rightarrow \textit{Review} \}
 \end{aligned} \tag{4.12}$$

where the digital and physical investigations occur simultaneously, and

$$\begin{aligned}
 \textit{Readiness} &= \{\textit{Operational Readiness} \Rightarrow \textit{Infrastructure Readiness}\} \\
 \textit{Deployment} &= \{\textit{Detection and Notification} \\
 &\quad \Rightarrow \textit{Confirmation and Authorisation}\} \\
 \textit{Physical Investigation} &= \{\textit{Preservation} \Rightarrow \textit{Survey} \Rightarrow \textit{Documentation} \Rightarrow \textit{Search} \\
 &\quad \textit{and Collection} \Rightarrow \textit{Reconstruction} \Rightarrow \textit{Presentation}\} \\
 \textit{Digital Investigation} &= \{\textit{Preservation} \Rightarrow \textit{Survey} \Rightarrow \textit{Documentation} \Rightarrow \textit{Search} \\
 &\quad \textit{and Collection} \Rightarrow \textit{Reconstruction} \Rightarrow \textit{Presentation}\}
 \end{aligned}
 \tag{4.13}$$

4.11.2 Terminology

The terminology used within the DFPM is similar to the definitions given in many of the sections before. During the *review* phase the whole investigation is reviewed and areas of improvement are identified. It is interesting to note that this DFPM includes a physical and digital investigation to be conducted concurrently. For a full discussion the original paper by Carrier and Spafford can be referenced [18]. Baryamureeba [7] also gives a detailed overview of the terminology used in the paper as will be set out in Section 4.12.

4.11.3 Comments

The DFPM includes processes in the investigation to accommodate issues such as data protection, acquisition, imaging, extraction, interrogation, ingestion and normalisation, analysis and reporting [7]. High-level processes are included for both the physical and logical or digital crime scenes. Baryamureeba questions the practicality of the model. The following might illustrate this problem. The primary crime scene is where the crime is initiated. The target of location or the victim, which is the secondary crime scene, is not included as part of the investigation. However, including the physical crime scene is a notable contribution.

The digital crime scene is defined as a virtual environment within the context of

hardware and software [18]. The crime or incident exists in this virtual environment. This DFPM also differs from previously defined DFPMs in that it consists of groups of phases to be completed. Reconstruction and review of an investigation are important. This confirms the principle that a different investigator should reach a similar conclusion if the same steps are followed methodically. Baryamureeba is also critical of the various sub-categories that make implementation difficult – another indication of a linear standardised approach sought by many of the authors.

The clear differentiation of a physical and digital crime scene seems trivial, but this distinction is critical for practical purposes of an investigation.

4.12 Baryamureeba

The Enhanced Integrated Digital Investigation Process (EIDIP) DFPM makes a clear distinction between the physical and digital crime scene investigation processes [7]. This DFPM is an extension of the DFPM in Section 4.11.

4.12.1 Process

$$\text{Baryamureeba} = \{ \textit{Readiness} \Leftrightarrow \textit{Deployment} \Leftrightarrow \textit{Traceback} \\ \Leftrightarrow \textit{Dynamite} \Leftrightarrow \textit{Review} \}$$

where

$$\begin{aligned} \textit{Readiness} &= \{ \textit{Operational Readiness} \Rightarrow \textit{Infrastructure Readiness} \} \\ \textit{Deployment} &= \{ \textit{Detection and Notification} \Rightarrow \textit{Physical Crime Scene} \\ &\quad \textit{Investigation} \Rightarrow \textit{Digital Crime Scene Investigation} \\ &\quad \Rightarrow \textit{Confirmation} \Rightarrow \textit{Submission} \} \\ \textit{Traceback} &= \{ \textit{Digital Crime Scene Investigation} \Rightarrow \textit{Authorisation} \} \\ \textit{Dynamite} &= \{ \textit{Physical Crime Scene Investigation} \Rightarrow \textit{Digital} \\ &\quad \textit{Crime Scene Investigation} \Rightarrow \textit{Reconstruction} \\ &\quad \Rightarrow \textit{Communication} \} \end{aligned} \quad (4.14)$$

4.12.2 Terminology

- *Readiness* is listed to include the training of personnel and having sufficient infrastructure to deal with the investigation.
- *Deployment* involves providing mechanisms to detect incidents and confirm such incidents. There are five sub-processes. The first sub-process is to detect the incident and notify the appropriate authority. Second, the physical crime scene is examined and potential evidence is identified. Third, the potential evidence is subjected to a digital examination. Fourth, confirmation of the incident is given to obtain legal approval for a search warrant. Evidence is then presented to the appropriate forum.
- In the *Traceback* phase, the physical crime scene is tracked down to lead to the identification of devices used in the execution of the act. Firstly, the primary crime scene is obtained from evidence collected during deployment. This typically includes finding the host computer within a networked environment. Secondly, authorisation is obtained to permit further investigation of the acquired evidence.
- The *Dynamite* phase investigates the primary crime scene. This is aimed at collecting and analysing evidence items found at the primary scene to find the incident perpetrators. The phase involves four sub-processes. The physical evidence found at the crime scene is examined. Secondly, the digital crime scene is examined. Thirdly, possible events are reconstructed to formulate a possible hypothesis. Fourthly, the final interpretations are communicated in a presentation to the appropriate forum.
- The *Review* phase is performed last, where the whole investigation is reviewed and areas of improvement are identified.

4.12.3 Comments

This DFPM is based on the work of Carrier and Spafford. In the original paper this DFPM is represented as a waterfall type model. This is indicated in Equation [4.14](#)

with the bi-directional arrow. The listed processes given for the physical and digital investigation processes differ and do not occur simultaneously. The process list includes readiness, deployment, traceback, dynamite and review phases. Each of the listed processes includes a number of sub-processes to be completed during the investigation. Table 4.1 later on illustrates that the terms digital and physical investigation have different meanings within each of the sub-processes. This only increases confusion on the already broad spectrum of terminology listed so far.

The main objective of the Baryamureeba DFPM is to separate the physical investigation from the digital investigation [7]. However, this results in a complicated adaption of the Carrier and Spafford DFPM. In Baryamureeba's model a new phase is introduced where the primary crime scene is identified in the traceback phase. The primary crime scene is where the incident originated. Reconstruction is done only once in this DFPM, when all the necessary evidence has been collected.

The digital crime scene is processed in a virtual environment created by hardware and software [7]. The phases listed are preservation, survey, search and collection and documentation. The preservation phase includes the duplication of digital media. During the survey the investigator identifies and separates potential useful data from the imaged set. Hidden, deleted, manipulated or damaged data files are recovered during the search and collect phase. Documentation involves the extensive documenting of the evidence found, which in turn is useful in the presentation phase.

4.13 Ciardhuáin

The DFPM proposed by Ciardhuáin is probably the most all-inclusive and comprehensive to date. The steps or phases are also called activities. The following activities listed in this DFPM are *awareness, authorisation, planning, notification, search for and identification of evidence, collection, transportation, storage, examination, hypothesis, presentation, proof/defence, and dissemination* [24]. Although the steps are discussed in depth by the authors of the paper, only the listed differences will be repeated here. This DFPM is also a linear representation of a DFPM.

4.13.1 Process

$$\begin{aligned}
 \text{Ciardhuáin} = \{ & \textit{Awareness} \Rightarrow \textit{Authorise} \Rightarrow \textit{Plan} \Rightarrow \textit{Notify} \Rightarrow \textit{Search}/ \\
 & \textit{Identify} \Rightarrow \textit{Collect} \Rightarrow \textit{Transport} \Rightarrow \textit{Store} \Rightarrow \textit{Examine} \\
 & \Rightarrow \textit{Hypothesise} \Rightarrow \textit{Present} \Rightarrow \textit{Prove/Defend} \Rightarrow \textit{Disseminate} \}
 \end{aligned} \tag{4.15}$$

4.13.2 Terminology

Processes follow the waterfall model, in other words one process is followed by another in sequence. Certain sequences can be repeated if needed. The sequence of examine, hypothesise, present and prove/defend will often be repeated as the evidence pool grows during the investigation [24].

- *Awareness* is defined as the phase during which the investigators are made aware that a crime has taken place; the crime is reported to some authority. An intrusion detection system can also trigger such awareness. Ciardhuáin specifically includes this in the DFPM because the method of awareness could influence the investigation [24]. In some instances the investigation will have to be conducted with or without the knowledge of the instigator. In some cases the co-operation of the investigator can be expected, in cases such as internal investigations where the parties would like to find the route cause of the incident. Awareness can be internal or external to an organisation.
- *Authorisation* is where the type of investigation has been identified and now the appropriate authorisation may be required to proceed. Authorisation is also acquired internally or externally.
- *Planning* is influenced by information within and outside the organisation that will impact on the investigation. Outside factors include legal and other requirements that are not determined by the investigators, while internal factors include policies of the organisation, prior investigative knowledge and procedures. The scope can also be backtracked if the full requirements of the investigation are not included in the planned scope. Externally imposed policies, regulations and legislation, external information, information distribution and organisational policies can influence the planning phase.

- During *notification* the stakeholders or investigated subject are informed that an investigation is taking place. In cases where the investigated subject must not know that an investigation is taking place, this step is omitted. Other interested parties can also be informed during this step that there is an investigation in progress.
- The *search and identification* of evidence is the stage during which the location of the potential evidence is identified. In large investigations this may include finding information routes of information flows over ISPs. Authorisation will probably have to be revisited in cases of multiple jurisdictions.
- *Collection* occurs when the investigator takes physical possession of the evidence to be preserved and analysed. Ciardhuáin includes hard disk imaging and seizing of entire computers. This process is the focus of much of the literature on digital forensics [24]. Mistakes and incorrect procedure during this process will render evidence in later stages useless and inadmissible in court proceedings. Where questionable procedure is followed or cannot appropriately be explained during a court hearing, the digital evidence could be inadmissible. Many legal practitioners will focus on the collection procedure that was followed to find a questionable procedure that will invalidate the incriminating evidence.
- After collection, the evidence is *transported* to a suitable location for forensic examination. It is important that the integrity of the evidence is not affected during transfer – whether physically or digitally. Digital evidence is stored in a safe location before examination. Such integrity must also be ensured at the storage location.
- *Examination* is the core process of the digital investigation. A large number of techniques have to be used to access, find and extract evidence from the collected media. Large volumes of data can be the subject of an investigation and automated techniques may be required to aid the investigator. Ciardhuáin specifically mentions that during examination some automated techniques are required to aid the investigator [24].
- The *hypothesis* formulated by the investigator is based on the examination of the

evidence. The hypothesis is the investigator's construction of events or a possible sequence of events leading to the origin of the reported violation. The document compiled during the investigation must reflect the findings of the digital forensic examiner. Backtracking during examination is expected as the examiner gains insight into the investigation. There can be internal and external challenges to the formulated hypothesis. An example may be legal relevance of evidence found during an investigation.

- *Presentation* is where the hypothesis is presented to people other than the investigators, such as a jury or management. A decision will then be made on the basis of the presented findings.
- The *proof/defence* is where an investigator challenges his original hypothesis. The investigator will have to defend the findings, or prove that the events occurred as explained in the presentation.
- *Dissemination* of the lessons learnt is the final activity, if required. Policies and procedures influencing future investigations have to be integrated with current policies and procedures.

4.13.3 Comments

According to Ciardhuáin, the reason for proposing this model is the fact that the focus of other DFPMs is on processing digital evidence, whereas this model incorporates the whole investigation process. A lack of standardised terminology also seems to be a problem that is identified and that needs a solution [24]. The process only gives guidance on *what* must be done and not *how*. *How* would typically include specifics such as tools and technology that must be used. The training needed before an investigator is qualified to do an investigation, as well as best practices, common experiences and the development of standards are also identified as important future research topics.

Ciardhuáin also specifically notices that information flows are not addressed in any of the previous DFPMs proposed. The main problem is where and how the chain of custody is compiled. Different legal systems, best practices and languages are some difficulties that investigators could encounter.

Awareness, transport, storage, reconstruction and dissemination are considered irrelevant according to a survey conducted by Ciardhuáin [24]. The rest of the proposed DFPM activities were considered most relevant.

Perumal [79] states that awareness is a three-step process that includes the complaint, the investigation and the prosecution. In this DFPM only the complaint step is included.

Future work includes policy development on criminal investigations, auditors, civil litigation, system administrator investigations and judicial inquiries [24].

4.14 Cohen

The DFPM proposed by Cohen consists of seven listed processes or phases, which strongly correlate with those of the NIJ Digital Forensic Examination. The DFPM subsists within a digital forensic framework proposed by Cohen, where the focus of this DFPM is digital forensic examination [25].

4.14.1 Process

$$\text{Cohen} = \{ \textit{Identification} \Rightarrow \textit{Collection} \Rightarrow \textit{Transportation} \Rightarrow \textit{Storage} \\ \Rightarrow \textit{Examination and Traces} \Rightarrow \textit{Presentation} \Rightarrow \textit{Destruction} \}$$

where

$$\textit{Examination} = \{ \textit{Analysis} \Rightarrow \textit{Interpretation} \Rightarrow \textit{Attribution} \Rightarrow \textit{Reconstruction} \} \quad (4.16)$$

4.14.2 Terminology

- *Analysis* is where evidence is understood and characterised relative to the legal issue at hand.
- *Interpretation* takes the results of analysis to producing meaningful statements. The statements give meaning to the situation, legal and technical terms.
- *Attribution* involves drawing conclusions about causes and effects. The links that exist between them are identified and documented. A particular cause will give rise

to an effect, conversely a particular effect may or may not be caused by a certain cause.

- *Reconstruction* is the process by which a set of mechanisms that are similar to those identified has caused the effect of the digital evidence produced. Reconstruction is therefore a process where the investigator lists certain assumptions and limitations to most accurately present how evidence came to exist.

4.14.3 Comments

The focus of the Cohen DFPM is on the examination of digital evidence. It is interesting to compare the term examination as suggested by Casey in Section 4.8 and by Cohen here. This clearly indicates the need for some standardisation of terminology. The issue is constantly mentioned by various authors but never sufficiently addressed.

The set of activities included in *examination* given by Casey and Cohen respectively can be compared as follows:

Casey: *Examination* = {*Recovery, Harvesting, Reduction, Classification*};

where the set given by

Cohen: *Examination* = {*Analysis, Interpretation, Attribution, Reconstruction*}.

Clearly not a single sub-process within the two identified sets even bear the same meaning. A possible explanation is that the interpretation of the term *examine* and *analyse* have been swapped by the authors. This problem is fully addressed in Chapter 5.

Section 4.15 represents all the DFPMs that have been discussed here and summarises them in a single table so as to identify differences and similarities.

4.15 Some Common Ground

In this chapter a number of DFPMs have been discussed. The list includes Lee [61], Kruse and Heiser [60], NIJ Electronic Crime Scene Investigation [50], NIJ Forensic Examination of Digital Evidence [74], Casey [21], DFRWS [77], Reith [84], Carrier and Spafford [18], Baryamureeba [7], Ciardhuáin [24] and Cohen [25].

Each of the listed DFPMs has a description of its processes in sequential logic. In

Table 4.1 each of the processes has been assigned a numeric value to correspond with the order in which they appear in a specific DFPM. The DFPMs have a number of process steps that are taken during the investigation process, as indicated by the numeric value. The process terminology or descriptive actions has been standardised to be represented as actions – hence the terms are all given in the form of verbs. The standardised actions are seen in the rows of Table 4.1. The step in the DFPM is now represented by a numeric value in the corresponding column position. The following paragraph provides an example where the process descriptions are now represented as numeric values.

Lee lists four main steps or processes, which include *recognise*, *identify*, *individualise* and *reconstruct*. Each one of the steps is numbered as steps 1, 2, 3 and 4. The sub-processes are indicated by step 2.1 and step 2.2, namely *classify* and *compare* respectively. Where a DFPM has a process with sub-processes, the main process is listed, followed by the sub-processes. An example of this can be seen in Cohen where *examination* is indicated as step 5 in Table 4.1. The sub-processes *analysis*, *interpretation*, *attribution* and *reconstruction* are respectively represented as steps 5.1, 5.2, 5.3 and 5.4.

Carrier and Spafford list *digital investigation* and *physical investigation* to occur simultaneously, therefore both sub-processes have been assigned the same numeric value of step 3 within the sequential logic representation. The same can be seen in the Ciardhuáin DFPM where *search* and *identify* are represented by the same numeric value as step 5.

Baryamureeba has different sub-process numbers assigned to *digital investigation* and *physical investigation*. Here the processes do not occur simultaneously as in the Carrier and Spafford DFPM, but at different times during the investigation. The problem is that *digital investigation* and *physical investigation* in Baryamureeba's model have a different interpretation with every repetition.

The NIJ first responder process is indicated in the same column as the NIJ investigator as seen in Table 4.1.

Table 4.1: DFPMs

	Lee	Kruse and Heiser	NIJ First Responder	NIJ Investigator	NIJ Examination	Casey 2004	DFRWS	Reith	Carrier and Spafford	Baryamureeba	Ciardhuáin	Cohen
Recognise	1					1						
Document	1.1		2		5	4.2			3.3			
Collect	1.2		3	1		4.1	3	5	3.5		6	2
Preserve	1.2					4	2	4	3.1			
Identify	2					3	1	1			5	1
Classify	2.1					8						
Compare	2.2					8.2						
Individualise	3					8.3						
Evaluate	3.1											
Interpret	3.2											5.2
Reconstruct	4					4			3.6	4.3		5.4
Report	4.1			4	6	10						
Present	4.1						6	8	3.7		11	6
Acquire		1			3							
Authenticate		2										
Analyse		3		3		9	5	7				5.1
Examine			1	2	4		4	6			9	5
Package			4									
Policy/Procedure					1							
Assess					2	2						
Seizure						3						
Recover						5						
Harvest						6						
Reduce						7						
Organise						8.1						
Decide							7					
Prepare								2				
Approach Strategy								3				
Return								9				
Readiness									1	1		
Operational									1.1	1.1		

Continued on next page

Table 4.1 – continued from previous page

	Lee	Kruse and Heiser	NIJ First Responder	NIJ Investigator	NIJ Examination	Casey 2004	DFRWS	Reith	Carrier and Spafford	Baryamureeba	Ciardhuáin	Cohen
Infrastructure									1.2	1.2		
Deployment									2	2		
Physical Investigation									3	2.3 4.1		
Digital Investigation									3	2.4 3.1 4.2		
Review									4	5		
Detect									2.1			
Notify									2.1	2.5	4	
Confirm									2.2	8		
Authorise									2.2	3.2	2	
Survey									3.2			
Search									3.4		5	
Traceback										3		
Dynamite										4		
Submit										2.6		
Communicate										4.4		
Become aware											1	
Plan											3	
Prove/Defend											12	
Disseminate											13	
Transport											7	3
Store											8	4
Hypothesise											10	
Trace												5
Destroy												7
Attribute												5.3

Table 4.1 is a summary of the DFPM processes described in this chapter. The eleven DFPMs are listed in the columns, whereas the rows list the ordered processes required to complete a successful digital forensic investigation. Terms described in each DFPM

differ in many instances. There are however instances where the listed terms are actually synonyms for terms listed in other DFPMs. *Collect* and *Acquire* are not duplicated in any of the listed DFPMs. They do however have similar meaning as confirmed by a Thesaurus [102].

4.16 Conclusion

In this chapter, various DFPMs used extensively in the academic environment were discussed. Each DFPM was discussed briefly, a process listing was given in sequential logic notation and the terminology was described. Thereafter each DFPM was briefly commented on by listing commendations and shortcomings.

The aim of the chapter was to identify some standard method for conducting a digital forensic investigation. This was done by identifying the essential processes that must be completed for digital evidence to be admitted as evidence in a court proceeding. In the discussion, various problems were identified, such as that a standardised terminology is lacking to describe processes used within the digital forensic community.

Chapter 5 aims to reduce the process listing to identify the essential process components necessary in a DFPM.

Chapter 5

Integrated DFPM

*When we mean to build, We first survey the plot, then draw the model;
And when we see the figure of the house, Then must we rate the cost of the
erection.*

– *William Shakespeare*

5.1 Introduction

In Chapter 4 a selection of existing DFPMs was discussed.

The focus of the current chapter is to investigate how the various DFPMs can be integrated into a single DFPM based on the DFPMs discussed in Chapter 4. The terminology as discussed in Chapter 4 will be examined to find similar meanings whereby the number of required processes can be effectively reduced in the Integrated DFPM.

The IDFPM integration will address the primary objective as formulated in the problem statement, namely to determine whether the various DFPMs investigated can be integrated successfully. The eleven DFPMs are integrated by completing the following steps:

1. Understanding how Table 4.1 is constructed and interpreted.
2. Reducing the number of rows, which implies a reduction in the number of processes.
3. Construct an Integrated DFPM.

The following paragraphs will address the steps enumerated above.

5.2 Construction and Interpretation of DFPMs

The sequential logic notation introduced in the previous chapter is used to determine which processes in the various DFPMs are duplicates that can be removed. Table 4.1 compiled in the previous chapter is a summarised representation of all the DFPMs discussed; this table is to be used to remove similar processes based on the descriptions provided by the various DFPM processes.

The DFPM proposed by Lee is seen in the first column and used for illustration purposes here. The process listing order is indicated by the numerical index as seen in the rows of Table 4.1:

1. Recognise;
2. Identify;
3. Individualise;
4. Reconstruct.

Sub-processes are indicated by using a dot notation, where step *1.1* indicates the first sub-process of the first process. The first process will be to *recognise*, where the first sub-process is to *document*.

The complete DFPM listing as proposed by Lee is given grammatically as:

1. Recognise, including documenting while collecting and preserving;
2. Identify, while classifying and comparing;
3. Individualise, while evaluating and interpreting; and
4. Reconstruct, including reporting and presenting.

The DFPM by Lee is represented clearly and precisely by using the sequential logic notation proposed previously as:

$$\text{Lee} = \{ \textit{Recognise} \Rightarrow \textit{Identify} \Rightarrow \textit{Individualise} \Rightarrow \textit{Reconstruct} \}$$

where

$$\textit{Recognise} = \{ \textit{Document} \Rightarrow \textit{Collect and Preserve} \}$$

$$\textit{Identify} = \{ \textit{Classify} \Rightarrow \textit{Compare} \}$$

$$\textit{Individualise} = \{ \textit{Evaluate} \Rightarrow \textit{Interpret} \}$$

$$\textit{Reconstruct} = \{ \textit{Reconstruct} \Rightarrow \textit{Report and Present} \}$$

Each of the processes has been constructed with a sequential logic notation, which is a clear indication of the process ordering chosen by the various authors. In the construction of the notation the essential descriptions are identified and included in the DFPM. Table 4.1 lists all the DFPMs and process descriptions as discussed in Chapter 4. The table is an alternative representation of the DFPMs and is used as a summary of the DFPMs discussed. The table indicates that the various DFPMs do not all have the same terminology to describe the processes. The aim of Chapter 5 is therefore to reduce the number of process descriptions so as to standardise the terminology used in a digital forensic investigation.

5.3 Reducing the Number of Processes in the DFPMs

The number of processes in all the DFPMs is determined by the number of rows. Process descriptions are identified and compared with the other listed DFPM descriptions to effectively reduce the number of process descriptions that have similar meaning – which will result in less rows. For the purposes of the reduction process, no distinction is made between processes and sub-processes.

The processes are reduced by applying the following steps:

1. The row or process description with the highest recurrence is identified.

2. If there is a DFPM that does not have the highest occurring process description, then the DFPM descriptions are studied to find similar descriptions.
3. If there is a description in the DFPM that has a similar description, it is removed from the list of descriptions and assigned to the highest occurring process description.
4. If there is not a similar process description, then the description is retained as a process in the DFPM listing.
5. These steps are repeated until all the process descriptions have been examined.

The *collect* process appears in nine (9) of the listed DFPMs and *acquire* appears in two (2). *Collect* and *acquire* are synonyms [102], and therefore have similar meaning. The *collect* and *acquire* processes descriptions listed in Chapter 4 have the same objective.

Collection is defined in section 4.6.2, quoted here for convenience.

Collection includes the searching for and the recognition, collection and documenting of digital evidence. The evidence is extracted from the physical medium in the form of data. The collection may be in real time. If this occurs, then it may be classified as live forensics. Digital evidence that may be lost, such as data stored on volatile memory modules, must also be included.

Kruse and Heiser [60] in section 4.5.2 is included here to show the correlation with the NIJ Forensic Examination in section 4.7, where the investigator must *acquire the evidence without altering or damaging the original*.

Acquisition is defined in section 4.7.2, quoted here for convenience.

Acquisition is given to be digital evidence that is fragile by nature and can be altered and damaged or even destroyed by improper handling or examination. Therefore examination of evidence should be conducted on a copy of the original evidence, and the original evidence should be acquired in a manner that protects and preserves the integrity of the evidence.

The term/word/verb *acquire* is defined the same in both cases. The definition for *collection*, as discussed above, gives practical effect to *acquisition*. Both the *collect* and *acquire* processes is where digital evidence is extracted from a digital medium on which the investigation will be conducted.

The two process descriptions are mutually exclusive processes. Set theory is used to illustrate this as follows:

If

α is a set of DFPMs that have *collect* as process description; and

β is the set of DFPMs that have *acquire* as process description

then $\alpha \setminus \beta \therefore collect \equiv acquire$.

The new process description is assigned the name of *collect*, because it appears in the majority of the DFPMs. The DFPM process listing has been reduced by one process description. To use a dictionary would detract from the original definitions provided by the authors of the various DFPMs. In such a case the scientific approach or basis would have been voided by not including terminology as it is used and understood within the digital forensic scientific community.

The same method is applied to all the process descriptions. The following is a list of process descriptions that are found to be the same:

- *prepare* \equiv *ready* \equiv *plan*
- *detect* \equiv *recognise* \equiv *become aware*
- *disseminate* \equiv *destroy* \equiv *return*
- *preserve* \equiv *package*
- *classify* \equiv *organise*
- *report* \equiv *present*
- *assess* \equiv *survey*
- *attribute* \equiv *individualise*

The first word in each line is the new process description. The nineteen process descriptions listed above have thus been reduced to eight descriptions.

Report and *present* are similar in that a report will have to be compiled before being presented. The process is renamed to *present a report*.

In the DFPM proposed by Carrier and Spafford [18] the physical and digital investigations are indicated as separate processes. For the purposes of reducing the number of processes these processes are removed, but the sub-processes are retained. The *readiness* process by the same authors is also removed, and the sub-processes are renamed to *operational* and *infrastructure readiness* respectively. The result of these reductions provides for a linear process description, which is presumably not what the authors had in mind. The physical and digital investigations are therefore executed in parallel; by the same proposition the operational and infrastructure readiness processes also run in parallel.

The Baryamureeba [7] DFPM proposes *dynamite* and *traceback* processes. These processes can safely be removed because the sub-process listing described the process listing. Baryamureeba's [7] DFPM contains a *submit* process which is removed because *presenting a report* is taken to have the same meaning.

The Cohen [25] DFPM *traces* process is removed and is included with *examine*.

The Ciardhuáin [24] DFPM includes a *proof* and *defence* process. These are functions performed by parties to a court hearing and are therefore excluded. The *proof* is seen as the digital evidence presented in court as included in the report.

The processes and sub-process number indices have not been changed for the purpose of reducing the number of rows in the table. Some of the process numbers have been removed during the reduction process. The ordering of the process listings have however been changed to represent some logical flow of an investigation. Therefore no specific meaning is attached to the process listing as such, but most of the processes still follow some degree of the original ordering.

Table 5.1: DFPMs with reduced process listing

	Lee	Kruse and Heiser	NIJ First Responder	NIJ Investigator	NIJ Examination	Casey 2004	DFRWS	Reith	Carrier and Spafford	Baryamureeba	Ciardhuáin	Cohen
Policy/Procedure					1							
Operational Readiness									1.1	1.1		
Infrastructure Readiness									1.2	1.2		
Detect	1					1			2.1		1	
Assess					2	2			3.2			
Notify									2.1	2.5	4	
Confirm									2.2	8		
Authorise									2.2	3.2	2	
Deploy									2	2		
Document	1.1		2		5	4.2			3.3			
Approach Strategy								3				
Search									3.4		5	
Seize						3						
Recover						5						
Transport											7	3
Preserve	1.2		4			4	2	4	3.1			
Store											8	4
Collect	1.2	1	3	1	3	4.1	3	5	3.5		6	2
Authenticate		2										
Examine			1	2	4		4	6			9	5
Harvest						6						
Reduce						7						
Identify	2					3	1	1			5	1
Classify	2.1					8						
Organise						8.1						
Compare	2.2					8.2						
Hypothesise											10	
Analyse		3		3		9	5	7				5.1
Attribute	3					8.3						5.3
Evaluate	3.1											
Interpret	3.2											5.2

Continued on next page

Table 5.1 – continued from previous page

	Lee	Kruse and Heiser	NIJ First Responder	NIJ Investigator	NIJ Examination	Casey 2004	DFRWS	Reith	Carrier and Spafford	Baryamureeba	Ciardhuáin	Cohen
Reconstruct	4					4			3.6	4.3		5.4
Communicate										4.4		
Review									4	5		
Present a report	4.1		4	6	10	6	8	3.7	2.6	11	6	
Decide						7						
Disseminate							9				13	7

5.4 Integrated DFPM - IDFPM

The previous section reduced the number of process descriptions in all the examined DFPMs. However, all the DFPMs examined are still represented in Table 5.1. The purpose of this section is to integrate the eleven DFPMs into a single integrated DFPM that encapsulates the essential processes identified previously.

The new DFPM is called the Integrated DFPM or IDFPM and consists of the following processes: preparation, incident, incident response, physical investigation, digital forensic investigation, presentation. The processes are performed by appropriately qualified personnel who are conscious of the ACPO principles.

The *documentation* process is included in the IDFPM as a continuous process. The documenting process includes the investigation documents and chain of custody recorded as accurately as possible throughout the entire investigation. The diagrammatic representation on the IDFPM is illustrated in Figure 5.1. In developing the policies and procedures in an organisation, it is essential to seek legal advice so as to ensure robust documentation that will be able to withstand legal scrutiny.

Infrastructure and Operational Readiness are processes that occur in parallel. These processes will overlap extensively when setting up a digital forensic organisation. The documentation should continuously be developed after each investigation to ensure

that it is in line with decisions reached and new developments in technology.

When the incident is detected, the situation should be assessed and confirmed before notification is sent to authorise an investigation. At the incident scene it is not always easy to determine what one may find. It is therefore important to have an approach strategy in place before searching the premises. During incident response one may not always find physical evidence to seize; this is why recovery, seizure and preservation occur in parallel before evidence is transported and stored.

The digital forensic investigation's *collect* sub-process may occur directly after potential evidence is preserved. This will be common in network or live forensic investigations. The sub-process listing from *hypothesise* up to *review* should be repeated during the *digital forensic investigation* process to continually test the hypothesis formulated.

The decision reached during *presentation* should be recorded in *preparation*; this will aid investigators in future investigations when faced with similar incidents.

The IDFPM listing is given in sequential logic notation as follows:

$$\begin{aligned}
 \text{IDFPM} = \{ \{ & \textit{Preparation} \Rightarrow \textit{Incident} \Rightarrow \textit{Incident Response} \\
 & \Rightarrow \textit{Physical Investigation} \parallel \textit{Digital Forensic} \\
 & \textit{Investigation} \Rightarrow \textit{Presentation} \} \parallel \textit{Document} \}
 \end{aligned}
 \tag{5.1}$$

where

$$\begin{aligned}
 \textit{Preparation} &= \{\textit{Policy/Procedure} \Rightarrow \textit{Operational Readiness} \\
 &\quad || \textit{Infrastructure Readiness}\} \\
 \textit{Incident} &= \{\textit{Detect} \Rightarrow \textit{Assess} || \textit{Confirm} \\
 &\quad \Rightarrow \textit{Notify} \Rightarrow \textit{Authorise} \Rightarrow \textit{Deploy}\} \\
 \textit{IncidentResponse} &= \{\textit{Approach Strategy} \Rightarrow \textit{Search} \\
 &\quad \Rightarrow \textit{Recover} || \{\textit{Seize} \Rightarrow \textit{Preserve}\} \\
 &\quad \Rightarrow \textit{Transport} \Rightarrow \textit{Store}\} \\
 &\quad \wedge \{\textit{Preserve} \Rightarrow \textit{Digital Forensic Investigation}\} \\
 \textit{Digital Forensic Investigation} &= \{\textit{Collect} \Rightarrow \textit{Authenticate} \Rightarrow \textit{Examine} \Rightarrow \textit{Harvest} \\
 &\quad \Rightarrow \textit{Reduce} \Rightarrow \textit{Identify} \Rightarrow \textit{Classify} \Rightarrow \textit{Organise} \\
 &\quad \Rightarrow \textit{Compare} \Rightarrow \textit{Hypothesise} \Rightarrow \textit{Analyse} \Rightarrow \textit{Attribute} \\
 &\quad \Rightarrow \textit{Evaluate} \Rightarrow \textit{Interpret} \Rightarrow \textit{Reconstruct} \\
 &\quad \Rightarrow \textit{Communicate} \Rightarrow \textit{Review}\} \\
 &\quad \wedge \{\textit{Reconstruct} \Rightarrow \textit{Hypothesise}\} \\
 \textit{Presentation} &= \{\textit{Report/Present} \Rightarrow \textit{Decide} \\
 &\quad \Rightarrow \textit{Disseminate}\}
 \end{aligned}$$

The IDFPM is alternatively shown in Figure 5.1 as a process flow diagram. Both these representations will be used in Chapter 6 when the terminology is discussed.

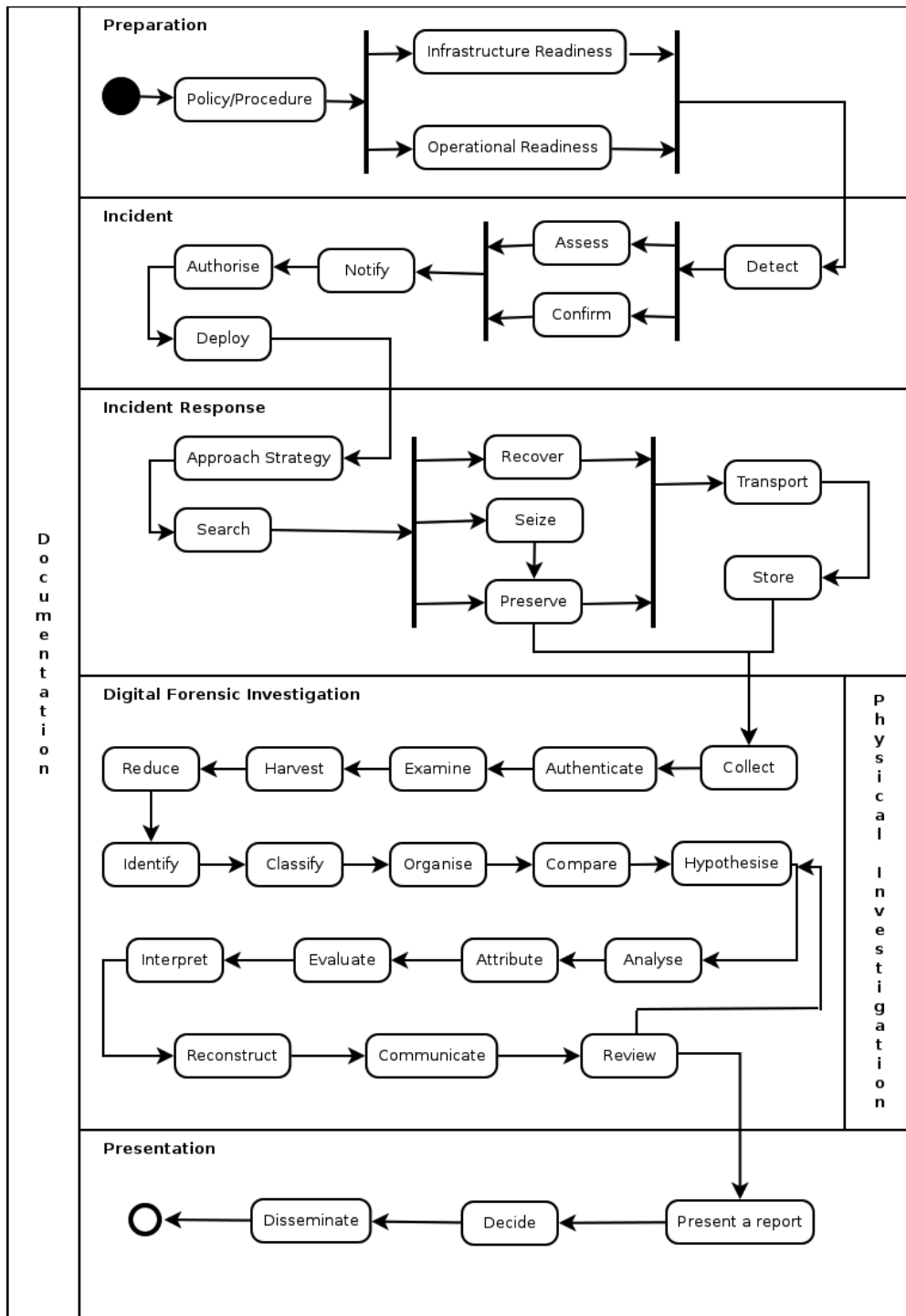


Figure 5.1: The IDFPM illustrated as a process flow diagram

5.5 Conclusion

This chapter explained how Table 5.1 was constructed and should be interpreted. The number of processes were reduced by identifying similar process descriptions as discussed in Chapter 4. The IDFPM was constructed by grouping similar processes together. The IDFPM was illustrated as sequential logic notation in Equation 5.1 and alternatively as a process model in Figure 5.1. Chapter 6 defines and discusses the process and sub-process descriptions of the IDFPM.

Chapter 6

The complete IDFPM

The difference between something good and something great is attention to detail.

– Charles R. Swindoll

6.1 Introduction

Chapter 5 introduced a sequential logic notation and process diagram for the IDFPM. The aim of the current chapter is to discuss the process and sub-process descriptions of the IDFPM for the purposes of the proposed model in detail. The process diagram and sequential logic notation are given for each one of the IDFPM processes, whereafter the descriptions are provided. The descriptions are a combination of the various DFPM definitions discussed in Chapter 4.

Section 6.8 briefly discusses the additional requirements of an investigation, which include the role players during an investigation and location. The IDFPM is supported by three essential pillars that include the law, digital forensic principles and investigator experience. The process descriptions listed guide the investigator as to how the digital forensic investigation is conducted.

The following paragraphs will address documentation, discuss the IDFPM in detail and thereafter discuss the additional requirements.

6.2 Documentation

Documentation is critical in any digital forensic investigation and is a continuous process in the IDFPM. Any deviation from the prescribed documentation should thoroughly be documented to ensure the chain of evidence is maintained; this will also ensure compliance with the best practice principles [1]. The primary purpose of the documentation is to serve as an investigation log to the investigator who will ultimately testify, in many instances long after the incident occurred, what procedure and investigative techniques were used to admit the final digital evidence.

Documentation starts during preparation when the organisation must compile a policy and procedure document on its approach to digital forensic investigations. After the detection of the first incident, the scene must be fully documented to enable easy physical reconstruction by the investigator during the digital forensic investigation. First responders arriving at the incident are not always equipped to effectively deal with documentation as required by the digital forensic investigation. It is therefore important that they are adequately trained to have a minimum required level of training before deployed to the incident [36, 105]. It is equally important that the first responders record the physical scene as accurately as possible to aid both the physical and digital forensic investigators as the investigation progresses.

The method of documenting is not as important as ensuring that every sub-process is accurately described in documentation for later reference. The documentation should as a minimum include investigator notes to enable another investigator to reach the same conclusion. The investigator must ensure that the chain of custody and chain of evidence are fully and accurately documented.

The documentation will form the basis of the digital evidence ultimately submitted to court. The digital evidence produced is presented in a report document with the findings of the investigator. The report presented includes the methods and techniques used during the investigation, the documented digital evidence presented, the chain of evidence, chain of custody and the expert opinion of the investigator.

6.3 Preparation

Preparation is the single most critical process in the IDFPM. This is where the organisation enables itself to deal effectively with various types of incidents. Tan encapsulates this process by stating that forensic readiness has two main objectives, firstly to maximise the collection of credible digital evidence from an incident environment, and secondly to minimise the cost of a forensic incident response [100].

Before any digital forensics investigation can be initiated, the organisation to conduct such an investigation, will have to be prepared for the specific type of investigation. Digital forensic readiness addresses the notion of how evidence effectiveness can be maximised at a minimal operational cost [100]. The focus of forensic readiness is on various types of investigations where an operational and infrastructure readiness is already established. Operational and infrastructure readiness is included as a component in the preparation process. As an example, a mobile forensic investigation will follow the same IDFPM used in forensic readiness, but a mobile forensic kit would have to be purchased on an infrastructure and operational level.

The sequential logic notation of preparation is given as:

$$\begin{aligned}
 \textit{Preparation} = \{ & \textit{Policy/Procedure} \Rightarrow \textit{Operational Readiness} \\
 & || \textit{Infrastructure Readiness} \}
 \end{aligned}
 \tag{6.1}$$

Figure 6.1 shows the *prepare* process as a process flow diagram.

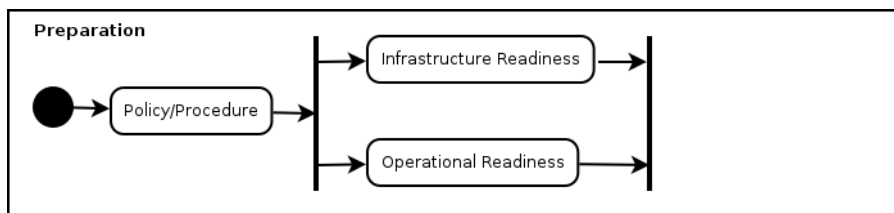


Figure 6.1: The *prepare* process flow diagram in the IDFPM followed by the *detect* sub-process, not illustrated here

The definition of *preparation* used for the purpose of this dissertation is the ability to maximise the effective production of digital evidence from a digital forensic investigation

while minimising the organisational cost of the investigation [100]. The two sub-processes are now discussed separately.

6.3.1 Policy/Procedure

Drafting of the initial policy and procedure documents need to be in place before an organisation can successfully initiate digital forensic investigations. The documents should contain a clear description of how the organisation will deal with digital evidence for the purposes of an investigation. The objective of the documents is to ensure that a minimum standard of integrity is maintained during investigations when dealing with digital evidence. The organisation must be aware of international standards that might be applicable in instances where various disciplines are interested in the findings of the investigation.

6.3.2 Infrastructure Readiness

Infrastructure readiness is rather difficult to define, due to the constant developments in technology. The primary objective with this process is to enable the organisation to effectively deal with different types of incidents to be investigated. This will mean that the organisation has to obtain the correct tools and technology in time to effectively conduct a digital forensic investigation [84, 24].

6.3.3 Operational Readiness

Operational readiness is determined by factors internal and external to the organisation. External factors include, but are not limited to, the legal system, territory legislation, rules of evidence and type of investigations conducted. Internal factors include the training of appropriately qualified personnel [7, 3]. The investigators have to be fully aware of both types of influences within the organisation and possible limitations. Any defects will certainly be exploited during presentation of the digital evidence findings.

6.4 Incident

An incident may be any action performed to compromise the confidentiality, availability and integrity of an information system. Digital forensics specifically deals with data found on digital media. The incident scope will have to be determined by the type of investigation conducted.

The sequential location of the incident is given as :

$$\begin{aligned}
 \textit{Incident} = \{ & \textit{Detect} \Rightarrow \textit{Assess} \parallel \textit{Confirm} \\
 & \Rightarrow \textit{Notify} \Rightarrow \textit{Authorise} \Rightarrow \textit{Deploy} \}
 \end{aligned}
 \tag{6.2}$$

Figure 6.2 shows the sub-processes followed during the *incident* process.

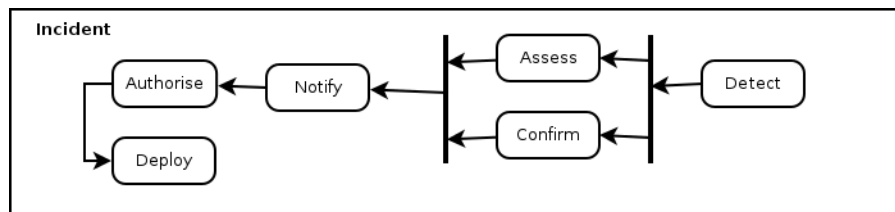


Figure 6.2: The *incident* process flow diagram in the IDFPM followed by the *approach strategy* sub-process, not illustrated here

6.4.1 Detect

An incident is detected by an automated incident detection system, or a similar set of event sequences is recognised by an investigator, based on possible previous experience. A comparative database should be developed to include possible unknown types of incidents, how they were detected, what the findings were and possible guidelines in expediting the investigation.

An incident can also be brought to the attention of the appropriate authorities by some external reporting method.

6.4.2 Assess

The detected anomaly is assessed by an investigator of an automated expert system to determine the appropriate course that the investigation should take.

6.4.3 Confirm

The incident detected should be confirmed by some other source before action is taken towards an incident response.

6.4.4 Notify

Once confirmation of an incident is verified, the investigators should be notified to initiate an incident response.

6.4.5 Authorise

Before any incident can be investigated, the suitable authority must be informed of the investigation. The authority should grant permission for the investigation to be initiated. This will include instruction from an attorney, a police warrant or other effective authorisation. The level of authorisation required is determined by the type of incident to be investigated.

An internal organisation investigation will also require authorisation and some form of informed consent from employees. Incidents are often detected covertly and dealt with overtly within an organisation. In these instances it is imperative that the organisation's policies and procedures are studied to determine any possible investigative limitation.

6.4.6 Deploy

All the sub-procedures listed above build up to the effective deployment of resources to respond to the incident detected. Defects in the sub-processes will be exploited during the following processes of the IDFPM.

6.5 Incident Response

Once the incident sub-processes have been completed, the incident response is initiated. The incident response process listing is given as:

$$\begin{aligned}
 \textit{Incident Response} &= \{ \textit{Approach Strategy} \Rightarrow \textit{Search} \\
 &\Rightarrow \textit{Recover} \parallel \{ \textit{Seize} \Rightarrow \textit{Preserve} \} \\
 &\Rightarrow \textit{Transport} \Rightarrow \textit{Store} \} \\
 &\wedge \{ \textit{Preserve} \Rightarrow \textit{Digital Forensic Investigation} \}
 \end{aligned}
 \tag{6.3}$$

Figure 6.3 shows the sub-process listing during the *incident response*.

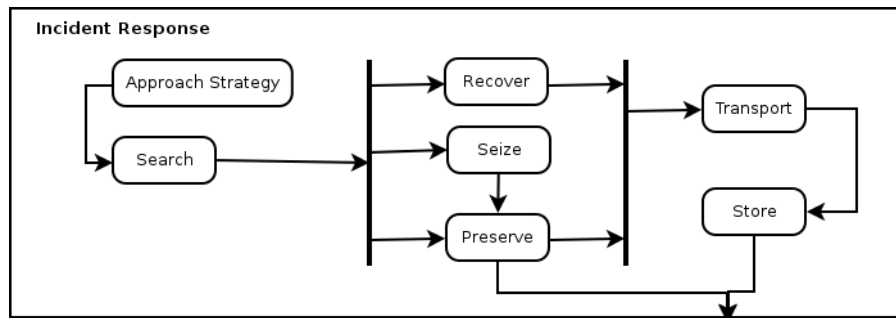


Figure 6.3: The *incident response* process flow diagram in the IDFPM followed by the *collect* sub-process, not illustrated here

The first responders typically arrive at the incident scene. Every investigation is different and it is impossible to determine what the first responders at the scene will encounter. Depending on the type of investigation, witnesses need to be safeguarded, suspects need to be detained as soon as possible after arrival and potential evidence must be secured. The first responder is the first custodian to maintain the chain of evidence and custody of potential digital evidence. The first responder must be able to accurately describe the scene in the initial drafting of documentation; these include photographs, video and sketches [16].

6.5.1 Approach Strategy

The approach strategy is determined by the type of investigation, the known facts [16] and the geographic location of the incident scene. An approach strategy is formulated after a brief interview has been conducted with witnesses and/or suspects. The objective of the approach strategy is to initialise a robust chain of evidence and chain of custody, while minimising possible damage to potential digital evidence.

6.5.2 Search

The location of physical evidence is determined by the approach strategy within the physical world. The digital evidence is located during the search sub-process within the cyber world. Digital evidence can potentially be found at various locations, central or distributed, within the cyber world depending on the incident.

Searching is limited to determining the exact location of digital evidence ultimately used in the investigation. Detection and notification determine the primary scene to where the first responders are deployed. The primary scene is usually only an entry point into a broader information system. The encompassing information system is a possible host within which the digital evidence must be located.

Potential digital evidence must be searched for at the incident scene. The incident location can be an initial point, end point or intermediate point in the incident. A perpetrator will usually be at an initial point while a victim is found at an end point. Intermediate points include, but are not limited to, a server used to mask the real attacker, ISPs and routers.

6.5.3 Seize

Seizure of digital evidence is dependent on a number of circumstances. A user computer can be packaged easily for transport and storage. A first responder will often encounter an incident within a larger information system. In these instances, the data located after a search should be duplicated immediately. This is in effect an extraction of an exact copy of digital evidence from the incident scene. The physical media is not seized in such instances, but the data is preserved for the purposes of the digital investigation.

6.5.4 Recover

Recovery occurs when the original system is restored to a functioning original state with additional security features to prevent similar future incidents [16]. This sub-process is not applicable to all types of investigations.

6.5.5 Preserve

Preservation is the securing, isolation and preserving of the digital and physical state of evidence [21]. The seized physical evidence is packaged and then transported to be stored at a suitable location, or alternatively the digital evidence is extracted during collection at the outset of the digital forensic investigation at a digital forensics laboratory. If the digital evidence is not capable of being transported, it must be preserved at the incident scene.

6.5.6 Transport

Evidence is transported to a location for secure storage of the original evidence and collection of digital evidence for the digital forensic investigation. The integrity, chain of evidence and chain of custody must be accountable during all stages of transportation.

6.5.7 Store

The physical and digital evidence must be stored in a secure pre-determined location. A standard should be implemented to ensure that the storage location is practical and sufficiently secured for the purposes of storing digital evidence. A selected number of factors should be considered, such as protection against water damage [22], possible malicious activity and theft.

6.6 Digital Forensic Investigation

The heart of the IDFPM is the digital forensic investigation. The processes listed will determine the success of the investigator's findings, which will ultimately be presented

in court. The sub-process listing is given in sequential notation as follows:

$$\begin{aligned}
 \textit{Digital Forensic Investigation} = & \{ \textit{Collect} \Rightarrow \textit{Authenticate} \Rightarrow \textit{Examine} \Rightarrow \textit{Harvest} \\
 & \Rightarrow \textit{Reduce} \Rightarrow \textit{Identify} \Rightarrow \textit{Classify} \Rightarrow \textit{Organise} \\
 & \Rightarrow \textit{Compare} \Rightarrow \textit{Hypothesise} \Rightarrow \textit{Analyse} \Rightarrow \textit{Attribute} \\
 & \Rightarrow \textit{Evaluate} \Rightarrow \textit{Interpret} \Rightarrow \textit{Reconstruct} \\
 & \Rightarrow \textit{Communicate} \Rightarrow \textit{Review} \} \\
 & \wedge \{ \textit{Reconstruct} \Rightarrow \textit{Hypothesise} \}
 \end{aligned}
 \tag{6.4}$$

Figure 6.4 shows the *investigation* sub-process listing in detail.

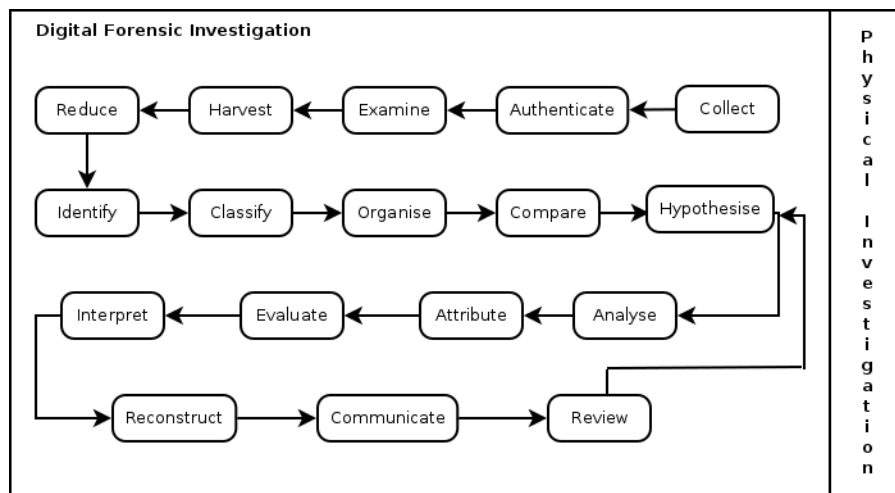


Figure 6.4: The *investigation* process flow diagram in the IDFPM followed by the *present a report* sub-process, not illustrated here

The physical investigation process occurs in parallel with the digital investigation if the crime is not isolated to the digital space. The focus of the physical investigation is to analyse DNA, fingerprints and other possible physical evidence obtained from the incident scene. The focus of this dissertation is limited to the digital forensic investigation sub-processes.

6.6.1 Collect

Collection of digital evidence is where the investigator takes physical possession of the original media. Two bit-by-bit images of the original data is produced by methodically following accepted best practice procedures, ensuring that the original data is not modified. One copy will be the investigation working copy and the other will be preserved in storage to maintain an exact copy of the original evidence.

The result of the collect sub-process is a copy of the original digital evidence, usually on another similar storage media. The digital evidence is a physical copy of the data set which has no logical data structure. Encase will produce a unique file type that will not be readable on most operating systems. The Linux *dd* command will produce a raw data file which will not be recognised by most operating systems in its current form.

6.6.2 Authenticate

The collected data attains legal validity by verifying the extracted data as genuine. A hash value of the original data and copied data is calculated. The hash value of both data sets must be exactly the same. Data is authenticated by using a unique one-way hash signature, usually MD5 or SHA-1.

6.6.3 Examine

Examination is generally known to be the process where the investigator makes digital evidence visible or extracts the data into a human readable form. Obfuscated data, which can be deleted or hidden data, is processed using sound digital forensic methods to conduct an effective investigation. With the use of digital forensics tools such as Encase [47] and Scalpel [34] the sub-process has largely been automated.

6.6.4 Harvest

Once all the data has been rendered visible by examination, the data is harvested by giving a logical structure to the entire dataset. The file and folder structure is indexed to give structure to the data collected from the original media. It may well happen that

the file allocation tables or disk indexing is deleted in some investigations.

The examination process will ensure that files, such as partially deleted files, are recognised collection from the original evidence medium. The partially discovered files and folders are then harvested. The harvesting process will produce a logical structure, the raw data is represented as information. The partially deleted files processed during examination will be visible to the extent that they were discovered or made visible during examination.

Collect, authenticate, examine and harvest will follow in processing sequence. On task level the following will happen. Collection extracts the raw data from the original digital media as bits and bytes. A tool such as Encase [49] or the Linux *dd* command is used to collect the data. Cohen states that you have a bag of bits after the data is collected [25]. The raw data is then authenticated and verified, ensuring the copied raw data is a representation of the original data. MD5 or SHA-1 is often used in combination to produce an authentic signature of the original data. The raw data is processed to identify possible metadata traces during examination. Examination is executed by a number of tools which have the ability to scan for file header and footer data. Harvesting will produce a logical structured data set, where the extracted raw data is now structured information. The harvested information can be mounted and read by the original file system, such as NTFS [57].

6.6.5 Reduce

The data analysed in a digital forensic investigation can be quite large. The data is reduced by identifying known data elements. This is done by using metadata and unique identifiers, such as MD5, to eliminate known system files and various other application data. The data remaining will be modified data or data that can uniquely be attributed to the users of a specific computer system.

6.6.6 Identify

Identification occurs when the investigators use the known digital evidence data to identify a possible incident to be investigated.

6.6.7 Classify

During classification, digital evidence is grouped by data with similar identifying patterns. Depending on the type of investigation, identified the data should be classified accordingly.

6.6.8 Organise

The digital evidence is organised in a manner so as to expedite the digital forensic investigation by focusing on the identified incident type and data classified. The digital evidence is restructured to suitably conduct the identified investigation.

6.6.9 Compare

If similar incidents have occurred in the past and are known to the investigator, the known classifications should be used to compare the current digital forensics data with similar past incidents.

6.6.10 Hypothesis

Up to this point in the investigation the investigator has only dealt with what is possibly known from the digital evidence. The investigator will have to formulate a hypothesis based on assumptions inferred from the digital evidence by the previous sub-processes. The crux of the hypothesis is to determine a possible root cause of the incident.

6.6.11 Analyse

During analysis, the organised data is thoroughly investigated and tested against the hypothesis formulated. During this sub-process, the legal validity of possible digital evidence is questioned by considering factors such as relevance, admissibility and weight. The hypothesis is tested by identifying best possible evidence.

6.6.12 Attribute

When the digital evidence is attributed to a specific user, the digital evidence is aligned to be linked with a particular individual or event that lies at the root cause of the incident.

6.6.13 Evaluate

The findings of the investigator are evaluated to determine whether the hypothesis formulated holds true.

6.6.14 Interpret

When the findings have been evaluated and the hypothesis holds true, the digital evidence is interpreted to produce meaningful statements in the legal context of a technical subject.

6.6.15 Reconstruct

A sequence of events inferred from the digital evidence known to the investigator is used to reconstruct a possible event sequence that reflects the incident result as accurately as possible. Reconstruction is not a finding based on the original digital evidence, nor is it established as factual. It is generally used to explain *how* the incident might have occurred.

6.6.16 Communicate

The digital evidence and investigator findings are communicated to the relevant interested parties. In most instances this will be the authority that authorised the incident response and subsequent digital investigation.

6.6.17 Review

The investigation results are reviewed and tested against the original hypothesis. Areas of improvement are identified to refine possible findings for the purposes of presentation and reporting. The organisation will also determine how to proceed with the incident.

Review is a sub-process through which the investigation is refined. This sub-process can either proceed to the presentation of a report during the presentation process. Alternatively, review is followed by the hypothesis sub-process to form a cycle that is repeated until the incident can be explained by producing a valid hypothesis with sound relevant admissible digital forensic evidence to support the findings.

6.7 Presentation

The final process in the IDFPM involves presentation of the final report. Figure 6.5 shows the final three sub-processes which are now discussed. The process is also shown as a sequential logic notation as follows:

$$\begin{aligned}
 \textit{Presentation} &= \{ \textit{Report/Present} \Rightarrow \textit{Decide} \\
 &\Rightarrow \textit{Disseminate} \}
 \end{aligned}
 \tag{6.5}$$

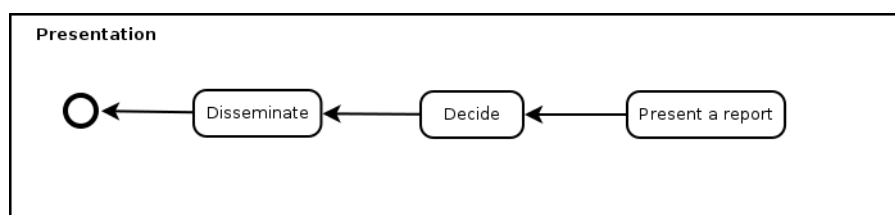


Figure 6.5: The *presentation* process flow diagram in the IDFPM

Presentation occurs when the hypothesis is presented to people other than the investigators, such as a jury or management. A decision will then be made based on the findings.

6.7.1 Present a Report

The presentation of a report involves the compilation of a detailed report detailing the entire investigation process, the chain of evidence, the chain of custody and ultimately the investigator findings that are formulated in an opinion to be presented in court. All other relevant documentation that was compiled during the investigation and that might be relevant in reaching a decision is included in the final presentation report. The legal processes of litigation, if applicable, will become the focus of the processes that follow.

6.7.2 Decide

Based on the presentation report, a decision is made on the person to whom the incident can be attributed. The decision must be recorded in some database for future reference.

6.7.3 Disseminate

Dissemination is the final activity of the IDFPM. In this sub-process of the investigation, the outcome of the investigation is used to review the existing policies and procedures of the organisation. The original digital evidence is also returned to the rightful owner.

6.8 Additional IDFPM Requirements

The additional requirements listed here function as a support structure for the IDFPM, namely role players and location.

Ioeng proposed a framework in 2006 that was derived from the Zachman framework [121] development of Enterprise Security Architecture. The framework was adapted to applications of forensics and hence referred to as FORensics ZACHman framework. The framework links the role players and their different processes together [54]. It is probably for this reason that the framework should be seen as a taxonomy for digital forensic investigation processes to function [98].

6.8.1 Role Players

Role players include a digital forensic examiner [25], case leader, system owner, legal advisor, digital forensic specialist, forensic investigator, forensic analyst and legal prosecutor [54].

A first responder at the incident scene may encounter a victim, witness or suspect [1]. The first responder may be an investigator or a team member of the evidence recovery organisation. An external consulting witness can be contacted if additional expertise is required [1]; the consultant will seldom be present at the incident scene, but may be contacted specifically for the digital investigation.

Armstrong indicates that digital forensics is a multi-disciplinary field where a wide range of subject specialisation needs to be understood [6]. Therefore an unlimited number of role players can be identified in potential incident investigations and the proposed IDFPF Framework is not limited to specific role players.

6.8.2 Location

The locations that are important in a digital forensic investigation can be listed as the initial incident scene, the digital forensic laboratory and a venue where the digital evidence will ultimately be presented [50]. Other unknown locations may have to be considered, depending on the type of incident, such as cases where distributed or networked incidents are at issue.

The following paragraph provides essential foundation pillars on which the IDFPF should operate. The IDFPF should take cognisance of the legal requirements of a country. The digital forensic principles as encapsulated in the ACPO [1]. Investigator experience is also critical in the success of an investigation.

6.9 The IDFPF Support Structure

There are three pillars that support the IDFPF, namely law, principles and experience. These pillars are required for a digital forensic investigation to be conducted in a manner that will enable the report presented to withstand legal scrutiny.

The framework is illustrated in Figure 6.6.

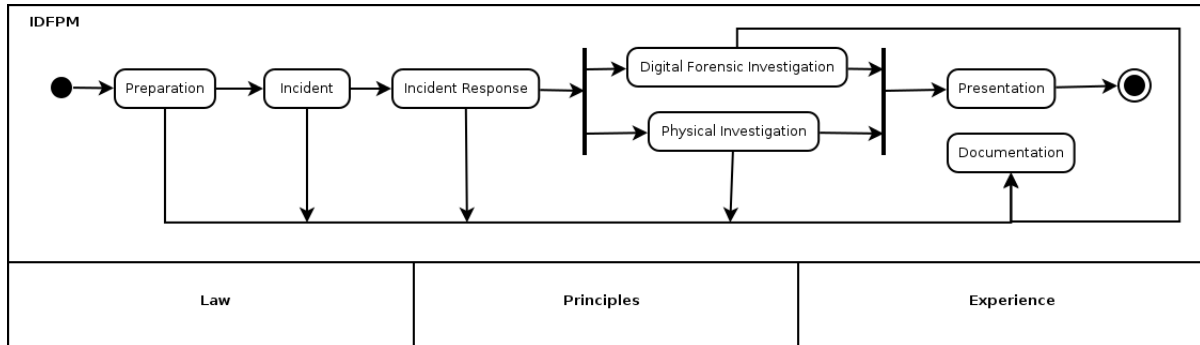


Figure 6.6: The framework within which the IDFPM will operate

Each one of the three pillars will be discussed in more detail in the next few paragraphs.

6.9.1 Law

The primary focus and challenge of digital forensics is not only technical, and the legal framework within which it operates should also be considered [51, 58].

The first and probably the most important pillar on which the IDFPM rests is law. The legal rules and standards required in the IDFPM are dependent on the country in which the investigation is conducted. It is therefore important to ensure that legal advice is sought in developing the organisation’s policies and procedures. The legal framework of each country will determine specifically how the IDFPM will operate.

The process of collecting digital evidence is assumed to be a legal process and the appropriate rules of evidence in different regions should be observed [27]. The rules of evidence are determined by the legal framework that applies in each country [52]. Pollitt states that the appropriate authority is necessary for digital evidence to be admissible in court, which can only be enforced by the appropriate authority qualified by law [27].

6.9.2 Principles

The IDFPM rests on specific principles, albeit legal or within the field of digital forensics. The APCO [1] principles have become well-known in the digital forensic community.

These principles should form an integral part of every investigation conducted by a digital forensic investigator.

The Best Practice and Standard Operating Procedure [27] documents are also essential and should be formulated prior to an investigation. These documents should be adapted to accommodate new types of incidents and developing technology. Circumstances may exist; where no best practice regarding a specific incident exists; in these cases the investigator should focus on documenting everything as extensively as possible. The reason for this is twofold: firstly to withstand legal scrutiny and secondly to aid in formulating a future best practice.

6.9.3 Experience

Experience refers to the experience of previous investigators and investigations. The findings of investigations, as well as the personal experience from other investigators, can be recorded to form a reference database that will aid other investigators.

Experience is often referred to in the literature, but a solution on how to address the difficult challenges is difficult to find [18, 35]. Ruiben often refers to an *experienced* investigator [91], but there is no existing standard against which this can successfully be measured [3]. Of the three pillars on which the IDFPM should rest, this one poses the biggest problem due to a lack of communication between disciplines and development within the digital forensics discipline.

The role players are closely linked to the experience required in digital forensics investigations. The experience of various specialists is included as the third and final pillar of the IDFPM.

6.10 Conclusion

The current chapter described the terminology as used in the IDFPM. The IDFPM is not able to function in isolation and is strengthened by additional requirements and a support structure. Within the digital forensic community much needs to be done to strengthen the three proposed pillars of the IDFPM, namely law, digital forensic principles and investigator experience.

Chapter 7 briefly introduces a number of tools that will enable an investigator to conduct a digital forensic investigation. These tools are examined to determine which processes in the IDFPM are supported by the listed tools. The chapter concludes with a prototype supporting some of the IDFPM processes.

Chapter 7

The IDFPM and Digital Forensics Tools

As technology advances, it reverses the characteristics of every situation again and again. The age of automation is going to be the age of 'do it yourself'.

– Marshall McLuhan

7.1 Introduction

Chapter 6 provided a detailed terminology listing of the processes included in the IDFPM. The additional IDFPM requirements an investigator should use as a departure point were also listed and briefly discussed as the law, digital forensics principles and investigator experience.

The IDFPM provides a framework that will aid an investigator in procedural aspects of a digital forensic investigation. The initial research question of this dissertation identified the examination and integration of various process models as an all-encompassing process resulting in the IDFPM. Tools and technology offer an effective solution to many of the procedural approaches in such an investigation. The aim of this chapter is, firstly, to provide an overview of digital forensic tools that were available for a digital forensic investigation at the time of writing this dissertation. Secondly, it aims to discuss the

IDFPM prototype that was specifically developed for this research project by addressing a limited number of IDFPM processes.

7.2 Tools

A number of digital forensic tool suites are available for investigators to conduct an investigation. Encase [49] and FTK [2] are well-known and readily accepted digital forensic suites. Various other digital forensic suites and tools are emerging to comply with the increasing demand that technology places on investigators. The challenges during investigations include the increasing size of digital evidence acquired for examination and rapidly developing technology such as cloud solutions and mobile devices.

A selected number of other useful digital forensic tools are briefly introduced and discussed. The tools listed are selected on the basis of solving a number of critical IDFPM processes.

- EnCase Forensic is a computer forensics product by Guidance Software which is used to acquire, analyse and report on digital evidence in various cases [112]. Encase is seen as the leading digital forensics software suite [49]. Encase has established a well-known reputation in digital forensic investigations and is readily accepted in court proceedings where digital evidence is presented [48].
- FTK Forensic Toolkit is a computer forensics software suite developed by AccessData [113]. FTK includes a wide variety of tools to aid the investigator in a forensic investigation. FTK Imager is a standalone application that is used to image hard drive disks. This results in a raw data image that calculates MD5 hash values and confirms data integrity [2]. Forensic Toolkit is used inter alia to recover deleted information, analyse email data, do keyword searches and perform password cracking [2, 113].
- SANS Investigative Forensics Toolkit, or SIFT, is a virtual machine computer forensics suite with various tools included to conduct a forensic examination [99] and a digital forensic investigation. SIFT is able to securely deal with raw image

files, including multiple file systems and evidence formats. File Systems include, but are not limited to, NTFS, HFS, UFS and ext3.

- Computer Online Forensic Evidence Extractor (COFEE), a product of Microsoft, has been specifically developed to aid investigators in extracting evidence from a Windows-based computer [110]. COFEE consists of a suite of 150 tools with a GUI used to collect data from a computer being investigated. COFEE is used for automated data extraction after configuration and generates reports from the data collected.
- PTK Forensics (PTK) is a non-free, commercial GUI for the digital forensics tool known as *The Sleuth Kit* [115] that runs on Linux, Apache, MySQL and PHP (LAMP) architecture. PTK is used for digital evidence acquisition and indexing during an investigation. The indexes are stored in an SQL database to be used for referencing. All data is verified during acquisition by hash signature calculation.
- DFF (Digital Forensics Framework) is a free and Open Source computer forensics software developed by (Re)discover Digital Investigation [39]. DFF can collect, preserve and examine digital evidence without data modification. DFF also includes tools that analyse digital evidence, recover data, construct search indexes and perform volatile memory forensic analysis.
- The Coroner's Toolkit (TCT) is a suite of computer security programs by Dan Farmer and Wietse Venema, used to assist in digital forensic analysis [111]. TCT is mainly a Linux-based tool used for data recovery and analysis.
- The Sleuth Kit (TSK) is a collection of Unix- and Windows-based tools used in forensic analysis of computer systems [15], maintained and written by Brian Carrier. TSK is used for data extraction and analysis of Windows-, Linux- and Unix-based computer images in an investigation. An interesting tool included in the suite is *mactime* which creates a timeline of all files based on their MAC times. TSK has been popularised by a custom front-end application named Autopsy.
- Bulk extractor is a forensic application used to scan disk images and extract useful information [94]. The results are stored in a separate file that is used to inspect

the image contents easily. Bulk extractor is platform independent and is capable of producing quick and effective results, with pre-programmed useful classification lists that include bank account details, email lists, telephone numbers and various histograms. BEViewer is a Java-based GUI that simplifies the complicated commands used to extract information from the evidence image [42].

Table 7.1 shows the processes of the IDFPM that are supported by the tools listed. The tools that have been identified to aid the investigator with the IDFPM provide limited support for all the processes required in a digital forensic investigation.

Table 7.1: Tools supporting the DFPM processes

	Encase	FTK Forensic Toolkit	SIFT	COFEE	PTK Forensics	DFE	The Coroner's Toolkit	The Sleuth Kit	Bulk Extractor
Policy/Procedure									
Operational Readiness									
Infrastructure Readiness									
Detect									
Assess									
Notify									
Confirm									
Authorise									
Deploy									
Document	*	*							
Approach Strategy									
Search									
Seize									
Recover	*	*	*	*	*	*	*	*	*
Transport									
Preserve									
Store									
Collect	*	*							
Authenticate	*	*	*	*	*	*	*	*	

Continued on next page

Table 7.1 – continued from previous page

	Encase	FTK Forensic Toolkit	SIFT	COFEE	PTK Forensics	DFE	The Coroner's Toolkit	The Sleuth Kit	Bulk Extractor
Examine	*	*	*	*	*	*	*	*	*
Harvest	*	*	*	*	*	*	*	*	*
Reduce	*	*							*
Identify	*	*	*	*	*	*	*	*	*
Classify	*	*							*
Organise									
Compare									
Hypothesise		*							
Analyse	*	*							
Attribute									
Evaluate									
Interpret									
Reconstruct	*	*							
Communicate									
Review									
Present a report	*	*							
Decide									
Disseminate									

Table 7.1 illustrates the deficiency in existing tools that support the complete IDFPM. All the tools examined have support for *examining, harvesting and identification* of digital evidence. *Reduction* and *classification* are supported by Encase, FTK and Bulk Extractor. The *compare* and *attribute* processes of the IDFPM are not currently supported by the tools examined. The processes that enjoy the attention of tool developers are limited to the *digital forensic investigation* process. The aim of developing a prototype is to address some of the deficiencies in the current list of tools examined. The prototype developed will specifically provide support for the *compare* and *attribute* processes of the IDFPM.

7.3 The Integrated Digital Forensics Process Model Prototype - IDFPMP

This section gives a brief overview of a prototype developed specifically for the purposes of this study. The IDFPMP is a digital forensic tool developed to address a select few processes in the IDFPM. The supported processes include *collect*, *authenticate*, *examine*, *harvest*, *reduce*, *classify*, *compare* and *attribute*, as described in the IDFPM.

The IDFPMP is a metadata index tool used to extract unique user data from an evidence file structure. The user evidence is attributed to the user by using file metadata. The IDFPMP conveniently deals with large volumes of digital evidence by eliminating known data files. The unique user data will direct the course of the investigation where the investigation is dependent on user-generated data. Duplicate files are removed and their history can conveniently be tracked within the file structure over time.

The following section indicates which of the IDFPMP processes are supported within the IDFPM.

7.3.1 IDFPM Processes Identified in the IDFPMP

The IDFPMP is a prototype built to include a limited solution to the broader processes in the IDFPM. The primary objective of the IDFPMP is to reduce the number of files an investigator should examine. The affected processes identified in the broader IDFPM framework are shown in Figure 7.1. The processes that the IDFPMP supports are *collect*, *authenticate*, *examine*, *harvest*, *reduce*, *classify*, *compare* and *attribute*. The IDFPMP scope is limited to the *digital forensic investigation* process of the IDFPM. The processes have been selected because the *compare* and *attribute* processes are dependent on the remaining processes listed.

The prototype has potential to be expanded to include more IDFPM processes. Future developments include an advanced search function, a graphic reconstruction of the history of a file and an increase in the number of classifications to more effectively reduce the amount of evidence an investigator must examine. A time-stamp mechanism can be included to monitor changes in a file structure over time within an organisation. This time-stamp history tracking capability will ensure digital forensic readiness within the

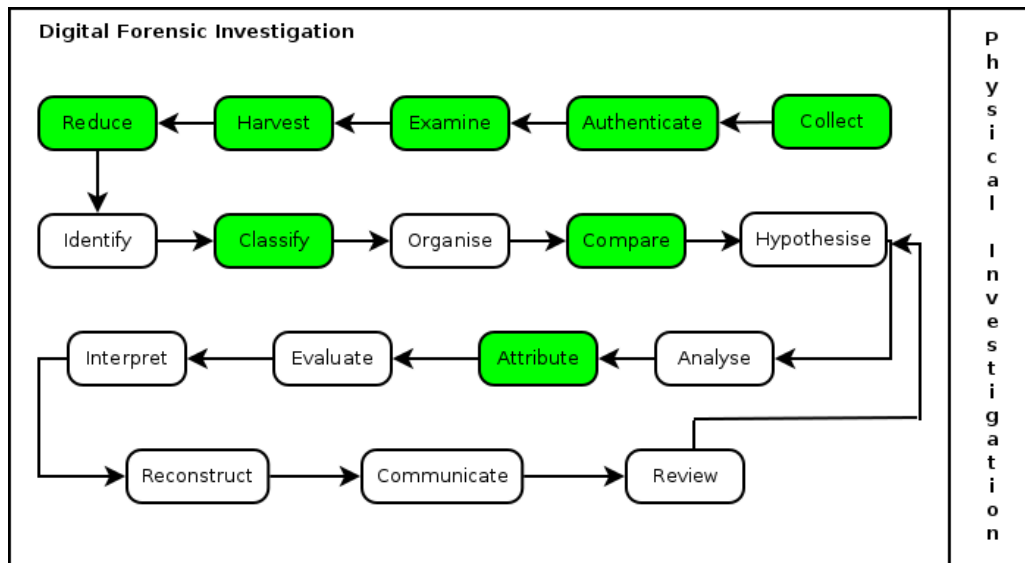


Figure 7.1: IDFPMP processes as found in the IDFPMP

organisation. An investigator will know when and by whom file content was modified, and at which particular point in time.

The following aspects of the IDFPMP are discussed in the sections below:

- The functional requirements of the IDFPMP
- Use case of the IDFPMP
- The technical platform specification of the IDFPMP
- The interface design of the IDFPMP
- The process description, which provides a brief description of the processes supported in the IDFPMP and how they are dealt with in the IDFPMP
- The output structures of the IDFPMP, which detail the structure and output provided by the prototype and aid the investigator in further analysis

7.3.2 Functional Requirements of the IDFPMP

The purpose of the IDFPMP is to extract a metadata index from a file structure during a digital forensic investigation. The metadata index is used to aid an investigator in

reducing digital evidence that needs to be examined in an investigation. The unique user data is extracted from the file structure by identifying known data. The file structure root folder is used as the IDFPMP input. The file structure is traversed using a breadth-first search, while extracting the metadata of the files and folders. The IDFPMP output is an XML structure, which is a metadata abstraction of the file structure.

The IDFPMP database is populated with the XML metadata extracted from a file structure. The IDFPMP database is used in subsequent investigations to determine which data is unique to a specific user. The unique user data is extracted by comparing the database content with the current XML metadata index.

7.3.3 Use Case of the IDFPMP

The IDFPMP processes affected by the IDFPMP are *reduction*, *comparison* and *attribution*. The use case for the IDFPMP is illustrated in Figure 7.2.

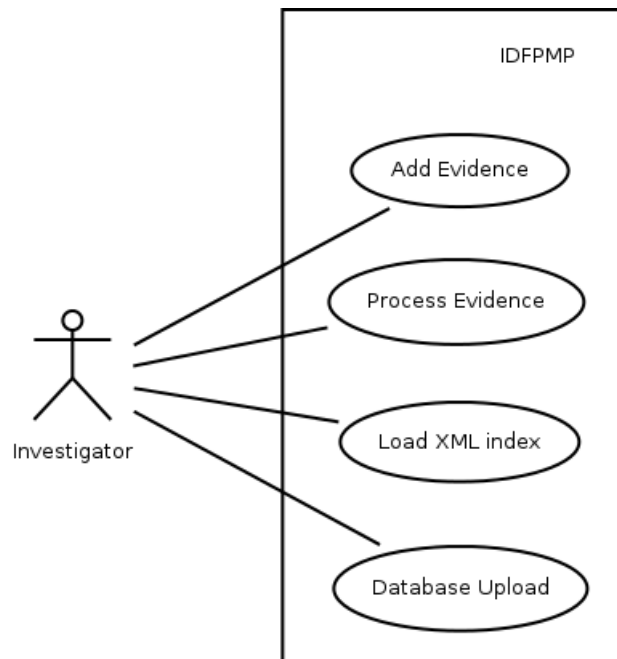


Figure 7.2: IDFPMP use case

7.3.4 Technical Platform Specification of the IDFPMP

The IDFPMP is developed in Java, which will enable the investigator to extract evidence from a number of platforms with ease. The platforms on which the IDFPMP has been tested include Windows, Linux and OS X. Java 1.7 is used for the primary reason that it includes a number of libraries that the IDFPMP uses extensively.

The IDFPMP database used for development is MySQL, as it can effectively deal with the potential large volumes of data that will accumulate over time.

The IDFPMP was tested on an Apple Mac with an Intel Core i7 2GHz processor with 8GB DDR3 RAM.

7.3.5 Interface Design of the IDFPMP

Figure 7.3 shows the use interface of the IDFPMP. The main window has five icons on the far left panel. These five icons are associated with the processes in the IDFPMP. They allow the investigator to add evidence, process evidence, load a previously compiled XML index file, upload the results to the IDFPMP database and close the application. These processes will be discussed further in Section 7.3.6.

The main viewing area has two functional tabs that the user can select to view the IDFPMP output. The tabs will now be discussed separately.

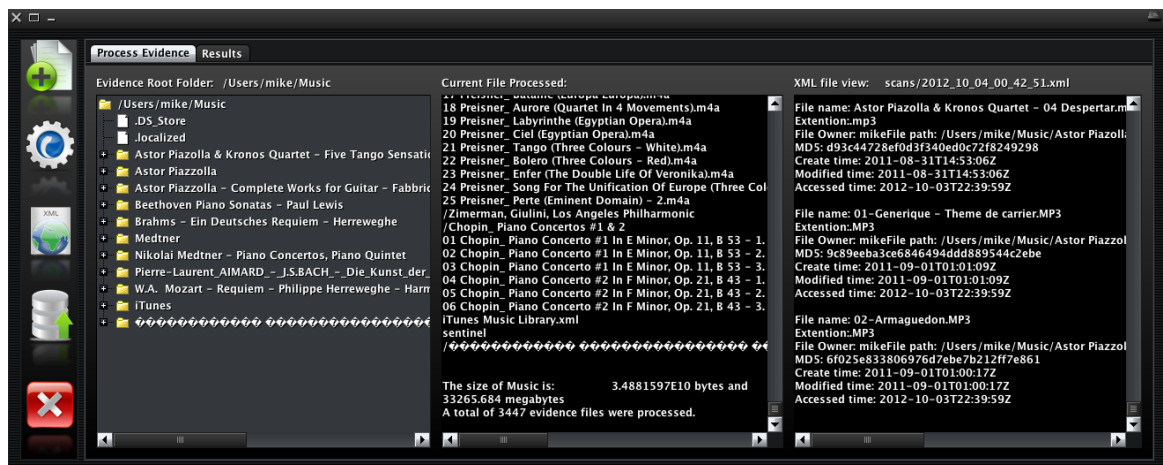


Figure 7.3: IDFPMP interface

The main window panel has two functional tabs, Process Evidence and Results, as seen in Figure 7.4.

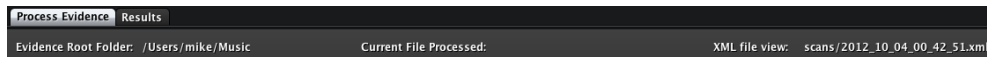


Figure 7.4: IDFPMP main frame tabs

The Process Evidence tab enables the investigator to see the chosen process output as shown in Figure 7.3. The three viewing areas of the Process Evidence tab show the various stages of the evidence during execution. The left viewing area shows the file structure currently loaded for processing. The middle viewing area lists the file currently being processed. The viewing area on the right is the XML metadata index compiled by the IDFPMP.

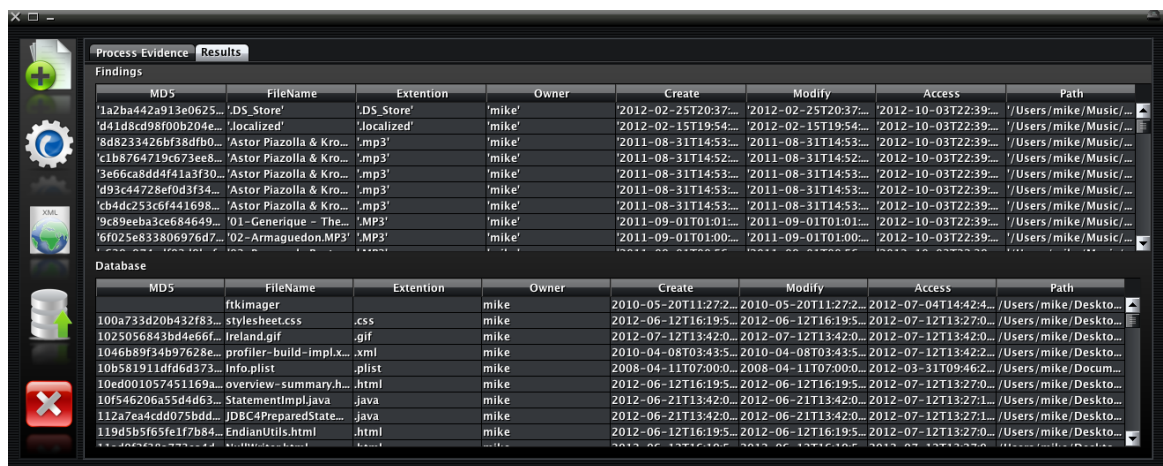


Figure 7.5: IDFPMP results interface

The Results tab, as seen in Figure 7.5, provides the investigator with a view of the output after the evidence has been processed. The two viewing areas allow the user to examine the results of the IDFPMP. The top viewing area shows the current XML file as before database integration. The bottom viewing area shows the existing IDFPMP database.

7.3.6 Process Description of the IDFPMP

The process descriptions are supported by the icons on the left panel as seen in Table 7.2.






	Add Evidence
	Process Evidence
	Load XML
	Upload to Database
	Close Application

Table 7.2: The left panel of the IDFPMP icon process description

The processes are described as follows:

- The Add Evidence icon allows the investigator to add a file structure to be processed. Figure 7.6 shows the interface when the icon is clicked.

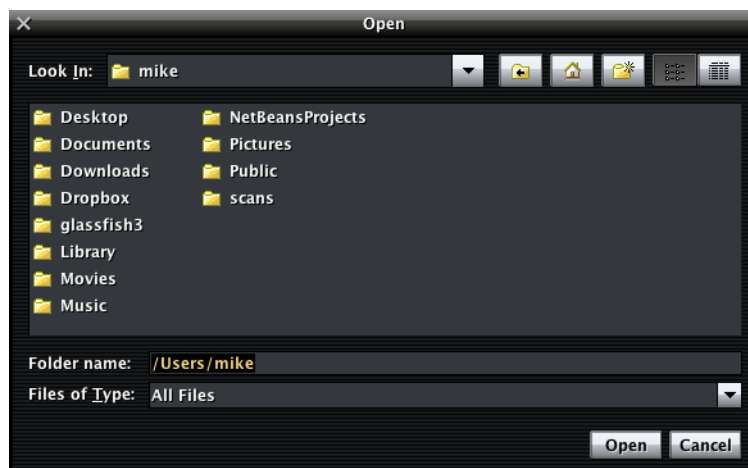


Figure 7.6: IDFPMP evidence loader

Once the evidence is added, the investigator will see the file structure selected as shown in Figure 7.7.

- The Process Evidence icon will initiate the process of examining the selected file

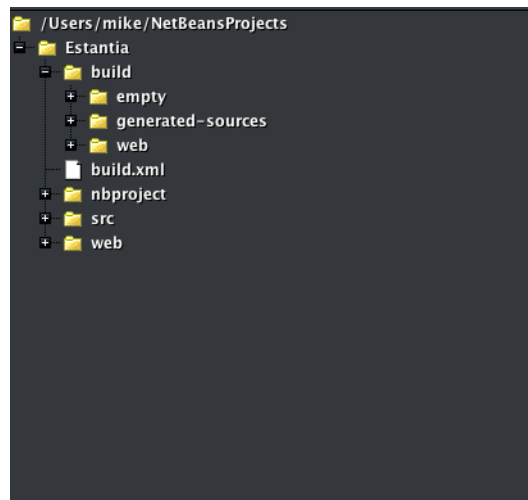
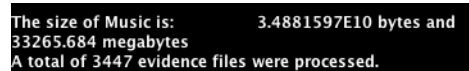


Figure 7.7: IDFPMP file structure loaded

structure and compiling the XML metadata index. The file structure loaded is traversed, breadth first, while extracting the file and folder metadata. The metadata extracted is limited to the file name, file extension, file owner, file path, MD5 hash and the MAC times associated with the file. All files are processed, including hidden and locked files. Folder metadata is included in the XML index, but the MD5 hash and file extension are excluded in the file index structure.

Once the processing is complete, a message is displayed. The user is prompted with the total number of files examined, processing time and the size of the file structure processed, as shown in Figure 7.8. In this case 3447 files with a total size of 33.4 GB were examined in 10 minutes. The resulting XML metadata index is only 2 MB.



```
The size of Music is: 3.4881597E10 bytes and
33265.684 megabytes
A total of 3447 evidence files were processed.
```

Figure 7.8: IDFPMP processing complete

Figure 7.9 shows the XML as seen by the investigator in the IDFPMP interface. The XML file constructed will not be included in the database at this point. The purpose of separating the XML and database processes is to allow the investigator

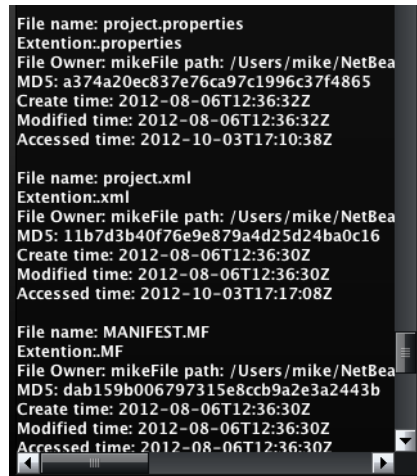


Figure 7.9: IDFPMP XML file structure loaded

to determine which evidence must be loaded to the database and whether the data is significant enough to upload to the database when convenient.

- The Load XML icon allows the investigator to load a XML metadata index. This is especially helpful if the investigator does not want to include everything in the database, or is in a location where access to a database is restricted. The XML metadata structure is loaded and displayed when the Results tab is selected in the Findings viewing area.
- The Database Upload icon allows the user to upload the XML metadata to the IDFPMP database. The XML file metadata content is classified by comparing the metadata with the data in the IDFPMP database. The data is classified as known files, difference files and history files. These classifications are discussed below.
- Close Application icon enables the investigator to terminate the application at any point.

7.3.7 External Output Structures of the IDFPMP

The IDFPMP will output two separate data structures: the XML metadata index and the IDFPMP MySQL database. These data structures are discussed separately next.

XML Metadata Index

The XML file contains the standard header as specified in the XML specification [120]. The XML file is named by combining the date and time the evidence was loaded. A sample name is *2012_10_04_00_42_51*, which indicates that the XML file was processed on 4 October 2012 at 00:42am. The XML file is saved in the scan folder with the data and time the evidence is loaded. The XML file can be uploaded to the database at any stage after the XML metadata index has been constructed.

The IDFPMP database and the XML index contain the same information. The IDFPMP uses the metadata index to classify the files as difference files, known files and history files. The three XML classification files that are compiled before the metadata is uploaded to the IDFPMP database are structured as follows:

- Difference files are files that are unique to the file structure processed by the IDFPMP. Difference files are not in the IDFPMP database. The primary test used to determine unique user files is to compare known MD5 hash signatures on file content. Any hash not found in the database will be classified as a difference file.
- Known files are the files found in the IDFPMP database. They are recorded in a separate XML file which the investigator can use at a later stage. The files listed in here include known system files that will generally not contribute to the investigation. If the contents of a known file are changed, it will be classified as a difference file. Other changes to metadata such as file extensions will not affect this classification, because the test is purely based on file content.
- History files are the files in the database that have a known hash, but some of the other metadata recorded is different to that of the recorded information in the database. Various history files can be selected from the database to determine ownership over time. These files are used to determine the original creator of a file over time. The test is however not absolute and other tests will have to be implemented.

The XML metadata index file is structured as illustrated in Figure 7.10.


```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ForensicComputerStructure>
  <FileStructure>
    <File>
      <fileName>MANIFEST.MF</fileName>
      <extension>.MF</extension>
      <owner>mike</owner>
      <filePath>/Users/mike</filePath>
      <md5>dab159b006797315e8ccb9a2e3a2443b</md5>
      <create>2012-08-06T12:36:41Z</create>
      <modify>2012-08-06T12:36:41Z</modify>
      <access>2012-08-13T09:27:21Z</access>
    </File>
  </FileStructure>
</ForensicComputerStructure>
```

Figure 7.10: XML metadata index structure

The metadata recorded in the XML file is limited to the file name, file extension, file owner, file path, MD5 hash and MAC times, as shown in in Figure 7.10.

The XML structure as seen above groups the metadata as forensic computer structure, file structure and file tags. The Forensic Computer Structure tag is a single evidence node loaded, which is the root folder of the evidence to be processed. The File Structure tag groups all the files for the purposes of an investigation instance together. The File tag will be all the individual items processed by the IDFPMP.

IDFPMP MySQL Database

The IDFPMP evidence index uploaded to the database is determined by the investigator. Not all evidence processed will be uploaded automatically because the database will become redundant in information retention.

The database has eight columns, which correspond with the eight attributes collected in the XML file. The primary key of the findings table is the MD5 generated by the file content. The remaining columns are filename, extension, file owner, create, modify, access and filepath. Folders do not have an MD5 or extension in the database, but all other data is recorded for folders. Figure 7.11 illustrates the structure as discussed here.

md5	filename	extention	fileowner	create	modify	access	filepath
10f546206a...	StatementImpl.java	.java	mike	2012-06-21T...	2012-06-21T...	2012-07-12T...	/Users/mike/Desktop/mi...
11049fd87f...	2-03 BWV 235 (3) Grati...	.m4a	mike	2010-03-07T...	2010-03-07T...	2012-03-31T...	/Users/mike/Music/iTune...
111cbbec24...	8-12 Sonata No. 26 in...	.m4a	mike	2011-04-10T...	2011-04-10T...	2012-03-31T...	/Users/mike/Music/iTune...
112a7ea4cd...	JDBC4PreparedStateme...	.java	mike	2012-06-21T...	2012-06-21T...	2012-07-12T...	/Users/mike/Desktop/mi...
114271e5ce...	03. [Earl Wild, piano] Ill...	.flac	mike	2011-09-02T...	2011-09-02T...	2012-10-03T...	/Users/mike/Music/Medt...
1146aa99c4...	2-09 Bach_Orchestral...	.m4a	mike	2011-01-28T...	2011-01-28T...	2012-03-31T...	/Users/mike/Music/iTune...
115f528700...	13 Brahms_Variations...	.m4a	mike	2011-01-30T...	2011-01-30T...	2012-03-31T...	/Users/mike/Music/iTune...
1168c4420f...	13. Ballada Para Mi Mu...	.mp3	mike	2011-09-01T...	2011-09-01T...	2012-10-03T...	/Users/mike/Music/Astor...
116c3b3427...	13 Etude X.m4a	.m4a	mike	2011-01-30T...	2011-01-30T...	2012-03-31T...	/Users/mike/Music/iTune...
119d5b5f65...	EndianUtils.html	.html	mike	2012-06-12T...	2012-06-12T...	2012-07-12T...	/Users/mike/Desktop/mi...
11ad0f2f38a...	NullWriter.html	.html	mike	2012-06-12T...	2012-06-12T...	2012-07-12T...	/Users/mike/Desktop/mi...
11b17f8d57...	2-26 VI-8. Rezitativ.m4a	.m4a	mike	2010-03-07T...	2010-03-07T...	2012-03-31T...	/Users/mike/Music/iTune...
11bd4424de...	08 - Contrapunctus VII...	.flac	mike	2011-09-11T...	2011-09-11T...	2012-10-03T...	/Users/mike/Music/Pierre...
11c2873c79...	WatchableWriter.java	.java	mike	2012-06-21T...	2012-06-21T...	2012-07-12T...	/Users/mike/Desktop/mi...
11c2b65512...	08 Chopin_Nocturne #...	.m4a	mike	2011-01-29T...	2011-01-29T...	2012-03-31T...	/Users/mike/Music/iTune...

Figure 7.11: IDFPMP database output

7.4 Conclusion

The IDFPMP serves only as a proof of concept and is not a contribution in itself. The focus of the dissertation was on the digital forensic process models. The two processes identified were seen as processes which are currently not included in the tools used in digital forensics. Before implementing the IDFPMP, various DFPMPs were examined, compared and then integrated into a consolidated final IDFPMP framework or model. Current tools addressing digital forensic solutions were identified and compared against a formal digital forensic process model. The shortcomings in the current tools were identified by comparing the tools with the various processes in the IDFPMP.

The aim of the prototype was not to implement the entire IDFPMP to enable a complete investigation, but to draw attention to *how* shortcomings in the current tools can be addressed by formal analysis of DFPMPs. The tools in industry have a limited scope and this prototype has addressed only the limitation identified by an analysis on a selected number of known tools which are readily used by DFIs.

This chapter briefly introduced a number of tools that would have been available to the investigator at the time of writing this dissertation. The innovative IDFPMP was discussed with the functional requirements being the focus of the discussion. The IDFPMP processes that are supported by the IDFPMP were also illustrated and discussed briefly.

The aim of the IDFPMP is not to be a complete solution to the IDFPMP framework as proposed in this dissertation. Both the Digital Forensic Tools and the IDFPMP primarily focus on the Digital Forensic Investigation process of the IDFPMP. This leaves scope for further development on a number of the IDFPMP processes.

The following chapter will conclude this dissertation.

Chapter 8

Conclusion

It's more fun to arrive at a conclusion than to justify it.

– Malcolm Forbes

8.1 Introduction

The current dissertation focusses on the Digital Forensic Process Models (DFPMs) used in a Digital Forensic Investigation. The study was conducted by examining a limited number of the DFPMs encountered in the current literature. The study was motivated by the lack of consistent terminology used in the research community and the non-standardised approach towards digital forensic investigations. Chapter 1 proposed a problem statement to the current dissertation. In closing, the problem statement and research objectives are revisited to determine to which extent the objectives have been met. The current chapter will conclude with a discussion of the current study's main contribution to the research and suggestions on further research as identified in this dissertation.

8.2 Revisiting the Problem Statement and Research Objectives

The primary objective of this dissertation was to investigate existing digital forensic process models within the known published literature to determine whether these can be integrated into a single digital forensic process model. The resulting IDFPM is a framework produced after examining a number of the existing digital forensic process models.

Evidence produced by implementing this process model in an investigation will ensure that the evidence can withstand legal scrutiny in a court of law. The complete IDFPM rests on three pillars:

law, digital forensic principles and previous investigator experience. These pillars will ensure that the digital evidence presented in court is – as a result of a rigorous digital evidence collection process – based on solid legal principles.

A further objective was to investigate whether any part of the process model could be automated and still produce a valid process model. The IDFPMP was developed and is able to automate a selected number of the IDFPM processes. The identified processes aid an investigator to effectively reduce the large volumes of digital evidence by producing an XML index of the collected evidence. The XML index is constructed by metadata that will be used to compare existing known data with data that is unique to a specific device or user.

8.3 Main Contribution

The main contribution of this dissertation is the development of an Integrated Digital Forensic Process Model or IDFPM. The IDFPM has made the following contributions:

- The development of the IDFPM framework that aids investigators to conduct a digital forensic investigation. The IDFPM is a result of an analysis of various digital forensic models in the current literature that have been successfully combined into an Integrated Digital Forensic Process Model or IDFPM.
- UML process flow diagrams and sequential logic were used to represent the processes in the selected process models. Sequential logic was used to eliminate the redundant processes, while the process flow diagrams were used to depict the IDFPM. Sequential logic processes are dependent on previous processes, which are critical elements in a digital forensics investigation. The final IDFPM was presented in both notations to illustrate previous process dependencies of the IDFPM.
- Terminology used in a number of digital forensic process models were redefined from existing process models. The definitions were by no means standardised, but an attempt was made towards establishing a standardised list of terms to be used by digital forensic investigators and the IDFPM.
- The integrated digital forensic process model was examined to identify the processes that could possibly be automated to alleviate the time-consuming task of data reduction in future investigations. The IDFPMP that was specifically developed for this dissertation supported the *collect*, *authenticate*, *examine*, *harvest*, *reduce*, *classify*, *compare* and *attribute* processes in the IDFPM.

8.4 Future Research

Although the proposed IDFPMP achieved the set of objectives to the extent described in the section above, it still suffers some limitations. These limitations provide opportunities to extend and support the work described in this dissertation by means of a number of future research projects:

- The IDFPMP is an integrated digital forensic process model. The model successfully integrates a number of process models, yet needs to be verified by an investigator in a number of investigations. The IDFPMP limitations have to be identified and corrected.
- The standardised terminology should be tested with case studies to determine their suitability in description.
- A limited number of the IDFPMP processes have been implemented in the IDFPMP. The prototype can be extended by including a number of automated processes to further alleviate the burden of investigators.

DFPMPs will continue to form an integral part of digital forensic investigations. The IDFPMP developed in this dissertation should be seen as a limited contribution to the body of knowledge available in industry and academia.

Bibliography

- [1] 7safe. Good Practice Guide for Computer-Based Electronic Evidence, 2008.
- [2] Access Data. Forensic Toolkit. <http://accessdata.com/products/computer-forensics/ftk>, September 2011.
- [3] C.L. Allinson. *Legislative and Security Requirements of Audit Material for Evidentiary Purpose*. PhD, Queensland University of Technology, <http://www.qut.edu.au/>, September 2004.
- [4] The American Society of Digital Forensics and eDiscovery. The American Society of Digital Forensics and eDiscovery. <http://www.asdfed.com/>, September 2011.
- [5] Apple Corporation. *Mac OS X Developer Library*. Apple, <https://developer.apple.com/library/mac>, version 10.6.6 edition, October 2012.
- [6] C. Armstrong and N. Jayaratna. Teaching Computer Forensics: Uniting Practice with Intellect. In *Proceedings of the 8th colloquium for Information Systems Security Education*, West Point, New York, June 2004.
- [7] V. Baryamureeba and F. Tushabe. The Enhanced Digital Forensic Investigation Process Model. In *Proceedings of the 4th Annual Digital Forensic Research Workshop, Baltimore, MD*. Citeseer, 2004.
- [8] N.L. Beebe. Digital Forensics Research: The Good, the Bad, and the Unaddressed. In *Fifth annual IFIP WG 11.9 international conference on digital forensics*, 2009.
- [9] N.L. Beebe and J.G. Clark. A hierarchical, Objective-Based Framework for the Digital Investigation Process. *Digital Investigation*, 2005.
- [10] J. Burchell. *Principles of Criminal Law*. Juta, Third edition, 2005.
- [11] D. Byers and N. Shahmehri. A systematic evaluation of disk imaging in Encase 6.8 and LinEn 6.1. *Digital Investigation*, 1(10), 2009.

- [12] B.D. Carrier. Open Source Digital Forensics Tools - The Legal Argument. *@stake Research Report*, 2002.
- [13] B.D. Carrier. *File System Forensic Analysis*. Addison Wesley, first edition, 2005.
- [14] B.D. Carrier. File System Forensic Analysis. <http://www.digital-evidence.org/>, September 2011.
- [15] B.D. Carrier. Sleuth Kit. <http://www.sleuthkit.org/>, September 2011.
- [16] B.D. Carrier and E.H. Spafford. Getting Physical with the Digital Investigation Process. Technical Report 29, CERIAS, 2003.
- [17] B.D. Carrier and E.H. Spafford. Defining Event Reconstruction of Digital Crime Scenes. *Journal of Forensic Sciences*, 2004.
- [18] B.D. Carrier and E.H. Spafford. An Event-Based Digital Forensic Investigation Framework. *Digital Forensic Research Workshop (DFRWS)*, 2004.
- [19] B.D. Carrier and E.H. Spafford. Automated Digital Evidence Target Definition Using Outliers Analysis and Existing Evidence. *Digital Forensic Research Workshop (DFRWS)*, 2005.
- [20] B.D. Carrier and E.H. Spafford. Categories of Digital Investigation Analysis Techniques Based on the Computer History Model. *Digital Investigation*, 2006.
- [21] E. Casey. *Digital Evidence and Computer Crime*. Elsevier Academic Press, 2004.
- [22] E. Casey, editor. *Handbook of Computer Crime Investigation: Forensic Tools and Technology*. Elsevier Academic Press, first edition, 2007.
- [23] W.J. Chisum and B.E. Turvey. Evidence Dynamics: Locard's Exchange Principle & Crime Reconstruction. *Journal of Behavioral Profiling*, 1(1), January 2000.
- [24] S.O. Ciardhuáin. An Extended Model of Cybercrime Investigations. *International Journal of Digital Evidence*, 3, 2004.
- [25] F. Cohen. *Digital Forensic Evidence Examination*. Fred Cohen & Associates, Second edition, 2009.
- [26] F. Cohen. Toward a Science of Digital Forensic Evidence Examination. In Kam-Pui Chow and Sujeet Sheno, editors, *Advances in Digital Forensics VI*, volume 337 of *IFIP Advances in Information and Communication Technology*, pages 17–35. Springer Boston, 2010.
- [27] Computer Analysis Response Team. *Report on Digital Evidence*, October 2001.

- [28] N.J. Croft and M.S. Olivier. Sequence Release of Privacy Accurate Call Data Record Information in a GSM Forensic Investigation. In *Proceedings of the ISSA 2006 from Insight to Foresight Conference*, 2006.
- [29] A. Culley. Computer Forensics: Past, Present and Future. *Digital Forensics*, 8:32–36, 2003.
- [30] G. Dahli. Forensic Accounting And Auditing: Compared And Contrasted To Traditional Accounting And Auditing. *American Journal of Business Education*, 1(2):115–126, 2008.
- [31] W. Delpont and M.S. Olivier. Isolating Instances in Cloud Forensics. In G.L. Peterson and S. Sheno, editors, *IFIP International Conference Digital Forensics*, pages 187–200. Springer, 2012.
- [32] DFRWS. Digital Forensics Research Workshop. online, September 2011.
- [33] Online Dictionary. Online Dictionary, 10 2011.
- [34] Digital Forensics Solutions. Scalpel. <http://www.digitalforensicssolutions.com/Scalpel/>, October 2012.
- [35] J.H.P. Eloff and A. Granova. Identity Theft. *Computer Fraud & Security*, 2004(11):7–11, October 2004.
- [36] R.F. Erbacher. Computer Forensics: Training and Education. In *Proceedings of the 5th Security Conference, Las Vegas, NV*, April 2006.
- [37] B.K.L. Fei, J.H.P. Eloff, M.S. Olivier, and H.A. Venter. Analysis of Web Proxy Logs. In *Advances in Digital Forensics II*, volume 222, pages 247–258. Springer, 2006.
- [38] Forensics Wiki. Mac Times. http://www.forensicswiki.org/wiki/MAC_times, March 2012.
- [39] Digital Forensic Framework. Rediscover Digital Investigation. <http://www.digital-forensic.org/>, September 2012.
- [40] The Free Online Dictionary. The Free Online Dictionary, September 2011.
- [41] S.L. Garfinkel. Digital Forensics Research: The next 10 years. *Digital Investigation*, 7:64–73, August 2010.
- [42] S.L. Garfinkel. Be Viewer. https://github.com/simsong/bulk_extractor/wiki/BEViewer, October 2012.
- [43] S.L. Garfinkel and A. Shelat. Remembrance of Data Passed: A Study of Disk Sanitization Practices. *IEEE Security & Privacy*, pages 17–27, January 2003.

- [44] P. Gladyshev. *Formalising Event Reconstruction in Digital Investigations*. PhD thesis, University College Dublin, August 2004.
- [45] P. Gladyshev. Finite State Machine Analysis of a Blackmail Investigation. *International Journal of Digital Evidence*, 4(1), 2005.
- [46] P. Gladyshev and A. Patel. Formalising Event Time Bounding in Digital Investigations. *International Journal of Digital Evidence*, 4(2), 2005.
- [47] Guidance Software. Encase. <http://www.guidancesoftware.com/forensic.htm>, April 2004.
- [48] Guidance Software. Encase search technology validated. <http://investors.guidancesoftware.com/releasedetail.cfm>, March 2009.
- [49] Guidance Software. Guidance software. <http://www.guidancesoftware.com/forensic.htm>, September 2011.
- [50] C.S. Guides. *Electronic Crime Scene Investigation: A Guide for First Responders*. United States Department of Justice, 2001.
- [51] M. Hannan. To Revisit: What is Forensic Computing? In *Australian Computer, Network & Information Forensics*, pages 103–111. School of Computer and Information Science, Edith Cowan University, Western Australia, 2004.
- [52] M. Hannan, S. Frings, V. Broucek, and P. Turner. Forensic Computing Theory & Practice: Towards developing a methodology for a standardised approach to Computer misuse. In *1st Australian Computer, Network and Information Forensics Conference*, Perth, November 2003.
- [53] J. Hershensohn. I.T. Forensics: The Collection and Presentation of Digital Evidence. In H.S. Venter, editor, *ISSA*, 2005.
- [54] R.S.C. Ieong. Forza - digital forensic investigation framework that incorporate(s) legal issues. *Digital Investigation*, 3(supplement):29–36, September 2006.
- [55] A.D. Irons, P. Stephenson, and R.I. Ferguson. Digital Investigation as a distinct discipline: A pedagogic perspective. *Digital Investigation*, 6(1):82–90, 2009.
- [56] K.J. Jones, R. Bejtlich, and C.W. Rose. *Real Digital Forensics: Computer Security And Incident Response*. Addison Wesley, first edition, 2006.
- [57] R. Koen and M.S. Olivier. An Open-Source Forensics Platform. In *Advances in Digital Forensics IV*, pages 325–334, 2008.

- [58] M.D. Köhn, J.H.P. Eloff, and M.S. Olivier. Framework for a Digital Forensic Investigation. In H.S. Venter, editor, *Information Security South Africa (ISSA)*, 2005.
- [59] M.D. Köhn, J.H.P. Eloff, and M.S. Olivier. UML modelling of Digital Forensic Process Models. In H.S. Venter, editor, *Information Security South Africa (ISSA)*, 2008.
- [60] W.H. Kruse and J. Heiser. *Computer Forensics: Incident Response Essentials*. Addison Wesley, first edition, March 2002.
- [61] H.C. Lee, T.M. Palmer, and M.T. Miller. *Henry Lee's Crime Scene Handbook*. Academic Press, First edition, 2001.
- [62] R. Leigland and A.W. Krings. A Formalization of Digital Forensics. *International Journal of Digital Evidence*, 3(2), 2004.
- [63] M. Lessing and B. von Solms. Live Forensic Acquisition as Alternative to Traditional Forensic Processes. Citeseer.
- [64] J. Lowry, R. Valdez, and B. Wood. Adversary Modeling to Develop Forensic Observables. In *Digital Forensic Research Workshop*, 2004.
- [65] N. Manyathi. 2012 Lex Informatica Cyber Law Conference. *De Rebus - South African Attorneys Journal*, pages 17–19, November 2012.
- [66] S. McClure, J. Scambray, and G. Kurtz. *Hacking Exposed: Network Security Secrets and Solutions*. McGraw-Hill, sixth edition, 2009.
- [67] R. McKemmish. What is Forensic Computing? *Trends and issues in crime and criminal justice*, 118:1–6, 1999.
- [68] A. Mylonas, V. Meletiadis, B. Tsoumas, L. Mitrou, and D. Gritzalis. Smartphone forensics: A proactive investigation scheme for evidence acquisition. In Gritzalis D, editor, *Proceedings of the 27th IFIP International Information Security and Privacy Conference*, pages 249–260. Springer, 2012.
- [69] A. Nieman. *Search and Seizure, Production and Preservation of Eelectronic Evidence*. PhD thesis, North-West University, Potchefstroom, 2006.
- [70] M.G. Noblett, M.M. Pollitt, and L.A. Presley. Recovering and Examining Computer Forensic Evidence. *Forensic Science Communications*, 2(4), October 2000.
- [71] North Carolina Association for Biomedical Research. Bio Science. http://www.aboutbioscience.org/wib_forensic_science.html, June 2011.

- [72] U.S. Office of Environmental Information. *Guidance Software for Preparing Standard Operating Procedures (SOPs)*. United States Environmental Protection Agency, April 2007.
- [73] U.S. Department of Justice. *Handbook of Forensic Services*. Federal Bureau of Investigation, Laboratory Division, 2003.
- [74] National Institute of Standards and Technology (NIST). *Forensic Examination of Digital Evidence: A Guide for Law Enforcement*. United States Department of Justice National Institute of Justice, <https://www.ncjrs.gov/pdffiles1/nij/199408.pdf>, April 2008.
- [75] Cambridge Online Dictionaries. Cambridge Online Dictionaries. <http://dictionary.cambridge.org/dictionary/british/best-practice>, March 2012.
- [76] Oxford University Press. The Oxford Dictionary. <http://oxforddictionaries.com/>, October 2012.
- [77] G. Palmer. A Road Map for Digital Forensics Research. Technical report, Digital Forensics Research Workshop Group, August 2001.
- [78] A. Patel and S.O. Ciardhuain. The Impact of Forensic Computing on Telecommunications. *IEEE Communications Magazine*, pages 64–67, November 2000.
- [79] S. Perumal. Digital Forensic Model Based on Malaysian Investigation Process. *International Journal of Computer Science and Network Security*, 9(8):38–44, August 2009.
- [80] C.P. Pfleeger and S.L. Pfleeger. *Security in Computing*. Prentice Hall, Fourth edition, 2007.
- [81] M.M. Pollitt. A History of Digital Forensics. In K. Chow and S. Sheno, editors, *Advances in Digital Forensics VI*, 337, pages 3–15. Springer Berlin Heidelberg, 2010.
- [82] R. Rawlingson. A Ten Step Process for Forensic Readiness. *International Journal of Digital Evidence*, 2, 2004.
- [83] I. Ray. Remote Upload of Digital Evidence over Mobile ad-hoc Networks. In *Advances in Digital Forensics II*, volume 222. Springer, 2006.
- [84] M. Reith, C. Carr, and G. Gunsch. An Examination of Digital Forensic Models. *International Journal of Digital Evidence*, 1(3):1–12, 2002.
- [85] S. Rekhis and N. Boudriga. A Temporal Logic-Based Model for Forensic Investigation in Networked System Security. In *Proceedings of the 2005 ACM symposium on Applied computing*, SAC '05, pages 287–291, New York, NY, USA, 2005. ACM.

- [86] S. Rekhis, J. Krichene, and N. Boudriga. DigForNet: Digital Forensic in Networking. In Sushil Jajodia, Pierangela Samarati, and Stelvio Cimato, editors, *Proceedings of The IFIP International Information Security Conference*, volume 278 of *IFIP International Federation for Information Processing*, pages 637–651. Springer Boston, 2008.
- [87] Republic of South Africa. Electronic Communications Act. Government Gazette, February 2010.
- [88] G.G. Richard, V. Roussev, and L. Marziale. Forensic discovery auditing of digital evidence containers. *Digital Investigation*, 4:88–97, 2007.
- [89] M.R. Rogers and K. Seigfrid. The future of computer forensics: a needs analysis survey. *Computers and Security*, 23(1):12–16, February 2004.
- [90] R. Rowlingson. A Ten Step Process for Forensic Readiness. *International Journal of Digital Evidence*, 2(3), 2004.
- [91] G. Ruibin and M. Gaertner. Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework. *International Journal of Digital Evidence*, 4(1), 2005.
- [92] P.J. Schwikkard and S.E. van der Merwe. *Principles of Evidence*. Juta, Second edition, 2002.
- [93] S.R. Selamat, R. Yusof, and S. Sahib. Mapping process of digital forensic investigation framework. *International Journal of Computer Science and Network Security*, 8(10), October 2008.
- [94] AFFLIB Open Source Computer Forensic Software. Bulk extractor. http://afflib.org/software/bulk_extractor, September 2012.
- [95] South Gauteng High Court. Vermooten v The Master of the Court. Court Proceedings, 2009.
- [96] South Gauteng High Court. Motata v State. <http://www.saflii.org/za/cases/ZAGPJHC/2010/134.pdf>, November 2010.
- [97] T. Stallard and K. Levitt. Automated Analysis for Digital Forensic Science: Semantic Integrity Checking. In *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*, pages 160–167. IEEE, Computer Society, 2003.
- [98] P. Stephenson. Modeling of Post-Incident Root Cause Analysis. *International Journal of Digital Evidence*, 2(2), 2003.
- [99] System Administration Networking and Security Institute (SANS). Computer Forensics and Incident Response. <http://computer-forensics.sans.org/community/downloads>, September 2012.

- [100] J. Tan. Forensic Readiness. *Cambridge, MA:@ Stake*, July 2001.
- [101] The Oxford Dictionary. The Oxford dictionary. <http://www.askoxford.com>, April 2011.
- [102] thesaurus.com. Thesaurus. <http://thesaurus.com/>, October 2011.
- [103] J. I. Thornton. Uses and Abuses of Forensics Science. *American Bar Association Journal*, 69:288–292, March 1983.
- [104] E. Tittel, editor. *Scene of the Cybercrime: Computer Forensics Handbook*. Syngress Shinder Books, First edition, August 2002.
- [105] J.P. Venter. Process Flows for Cyber Forensics Training Operations. *CSIR Researchspace*, 2006.
- [106] S. von Solms, C. Louwrens, C. Reekie, and T. Grobler. A Control Framework for Digital Forensics. In M.S. Olivier and S. Shenoi, editors, *Advances in Digital Forensics II*, volume 222 of *IFIP Advances in Information and Communication Technology*, pages 343–355. Springer Boston, 2006.
- [107] C. Walker. Computer Forensics: Bringing the Evidence to Court. *Online: http://www.infosecwriters.com/text_resources/pdf/Computer_Forensics_to_Court.pdf as on*, 12, 2007.
- [108] J. Walraven. The Simpson Trial Transcripts. <http://simpson.walraven.org/>, February 1995.
- [109] Wikipedia. Forensic Science. http://en.wikipedia.org/wiki/Forensic_science, October 2011.
- [110] Wikipedia. Computer Online Forensic Evidence Extractor. <http://en.wikipedia.org/wiki/COFEE>, September 2012.
- [111] Wikipedia. The Coroner’s Toolkit. http://en.wikipedia.org/wiki/The_Coroner%27s_Toolkit, September 2012.
- [112] Wikipedia. Encase. <http://en.wikipedia.org/wiki/Encase>, August 2012.
- [113] Wikipedia. Forensic Toolkit. http://en.wikipedia.org/wiki/Forensic_Toolkit, August 2012.
- [114] Wikipedia. Metadata. <http://en.wikipedia.org/wiki/Metadata>, February 2012.
- [115] Wikipedia. PTK Forensics. http://en.wikipedia.org/wiki/PTK_Forensics, September 2012.
- [116] Wikipedia. Sequential Logic. http://en.wikipedia.org/wiki/Sequential_logic, October 2012.
- [117] S.Y. Willasen and S.F. Mjøl̄snes. Digital Forensic Research. *Retrieved December*, 30(2007):92–097, 2005.

- [118] M. Wojcik, H.S. Venter, J.H.P. Eloff, and M.S. Olivier. Applying Machine Trust Models to Forensic Investigations. In Martin Olivier and Sujeet Sheno, editors, *Advances in Digital Forensics II*, volume 222 of *IFIP Advances in Information and Communication Technology*, pages 55–65. Springer Boston, 2006.
- [119] H.B. Wolfe. Computer Forensics. *Computers & Security*, 22(1):26–28, 2003.
- [120] World Wide Web Consortium. Extensible markup language (xml) 1.0 (fifth edition). <http://www.w3.org/TR/REC-xml/>, October 2012.
- [121] J.A. Zachman. *The Zachman Framework for Enterprise Architecture: Primer for Enterprise Engineering and Manufacturing*. Zachman International, 2003.

Appendix A

Publications and Contributions

A.1 Framework for a Digital Forensic Investigation

```
author    = {M. Kohn and M.S. Olivier and J.H.P. Eloff},
title     = {Framework for a Digital Forensic Investigation},
booktitle = {ISSA},
year      = {2006},
pages     = {1-7},
ee        = {http://icsa.cs.up.ac.za/issa/2006/Proceedings/Full/101_Paper.pdf},
crossref  = {DBLP:conf/issa/2006},
```

Framework for a Digital Forensic Investigation

Michael Kohn¹, JHP Eloff² and MS Olivier³

[1mkohn@cs.up.ac.za](mailto:mkohn@cs.up.ac.za), [2eloff@cs.up.ac.za](mailto:eloff@cs.up.ac.za), [3molivier@cs.up.ac.za](mailto:molivier@cs.up.ac.za)

**Information and Computer Security Architectures Research Group (ICSA)
Department of Computer Science
University of Pretoria**

Abstract - Computer Forensics is essential for the successful prosecution of computer criminals. For a forensic investigation to be performed successfully there are a number of important steps that have to be considered and taken. The aim of this paper is to define a clear, step-by-step framework for the collection of evidence suitable for presentation in a court of law. Existing forensic models will be surveyed and then adapted to create a specific application framework for single computer, entry point forensics.

1. Introduction

Over the past few years, computer forensics has risen to the fore as an increasingly important method of identifying and prosecuting computer criminals. Prior to the development of sound computer forensics procedures and techniques, many cases of computer crime were left unsolved. There are many reasons why an investigation might not lead to a successful prosecution, but the predominant one is a lack of preparation. The organization investigating the suspicious behaviour often lacks the tools and skills required to successfully gather evidence. Individuals attempting to investigate such suspicious activity may also lack the financial resources or tools to conduct such an investigation adequately and ensure that the evidence is undisputable in all circumstances. Moreover, there are instances when all of the above have been adequately put in place by an organization, but, due to a lack of training and correct procedure, the evidence collected can easily be disputed.

As a result, computer forensics seeks to introduce cohesion and consistency to the wide field of extracting and examining evidence obtained from a computer at a crime scene. In particular, the extraction of evidence from a computer is performed in such a way that the original incriminating evidence is not compromised. This is also useful when presenting a case without the support of legal expertise, as is often the situation since many organizations and individuals do not have in-house or personal legal representation.

This paper will propose a three phase framework that can be followed systematically to produce forensically sound evidence. The framework is an adaptation or combination of several existing forensic models.

The paper is structured as follows: the subsequent section will clarify important terminology used in the field of forensics; the third section will briefly discuss some

generally accepted frameworks; section four will introduce the proposed forensic framework, and closing remarks will be made in section five.

2. Background

According to the Oxford dictionary, the word forensic is defined as “relating to or denoting the application of scientific methods to the investigation of crime” and “of or relating to courts of law” [8]. At first, this appears to be quite a broad definition, but what is important in the first definition is that scientific methods are used in the investigation and the second definition emphasizes the fact that forensic activity usually relates to courts of law. Nonetheless, not all cases investigated end up in court. Examples are internal investigations and disciplinary hearings [7]. In conclusion, what would seem to be important is that, when a forensic investigation is launched, it is conducted in a scientific way and with a legal base as support.

Some authors make a clear distinction between computer and digital forensics [5]. Yet, for the purposes of this paper, no real distinction is made. Computer forensics can be defined as “analytical and investigative techniques used for the preservation, identification, extraction, documentation, analysis and interpretation of computer media (digital data) which is stored or encoded for evidentiary and/or root cause analysis” [7].

There are, however, methods which can help circumvent the, often tedious, task of ascertaining which factors are applicable to a particular forensic investigation. All organizations should have standards, policies and procedures in place that can assist in such an investigation. Standards that are important here are ISO17799 [10] and COBIT [11]. These standards do not cover a forensic investigation, but could be used to aid it.

As well as internal standards and policies, there are several legislative measures that support organizations attempting to prosecute computer crimes. In South Africa, there are a number of important Acts that can be referenced. These include the Electronic Communications and Transactions (ECT) [12] and the Promotion of Access to Information Act (PAIA) [13]. These, however, do not provide any clear guidelines as to how a forensic investigation should be conducted to ensure legal appropriateness.

Consequently, an important way for most organizations to protect themselves against computer crime is to institute internal policies and procedures which specify exactly what constitutes harmful action against or within an organization. These, however, are beyond the scope of this paper since there are a wide variety of possible solutions that can and have effectively been used.

Thus far it has been determined that implementing certain Standards, like ISO17799, can be a useful initial step by an organization towards effectively protecting its information and assets. Moreover, that specific policies and procedures should also be implemented within an organization to help protect the internal integrity of information and assets.

Once these basics are in place, the next step is to apply a sound forensic framework, which will consistently gather evidence suitable for presentation in a court of law, to ensure that criminal behaviour can be successfully prosecuted.

The Oxford dictionary defines a framework as “a supporting or underlying structure” [9]. A computer forensic framework can be defined as a structure to support a successful forensic investigation. This implies that the conclusion reached by one computer forensic expert should be the same as any other person who has conducted the same investigation [7].

A framework is also dependent on a number of structures. In the case of computer forensics, or forensics in general, legislation has to be considered to be of prominent importance. A forensic investigation has to be conducted in a scientific manner and must comply with all legal requirements, as set out in the second definition of forensics above. Evidence will have to be collected in this manner irrespective of the purpose i.e. internal investigation, disciplinary hearing or court case.

3. Frameworks

There is an old saying that prevention is better than cure. When applied to forensic frameworks this would seem to imply that preparation is the key to conducting a successful forensic investigation. Although preparation is important, it is impossible to be prepared for all types of behaviour. A sound base of previous knowledge and experience will always help, but a suggestion or documented case is not a complete resolution to solving a problem.

The number of forensic models that have been proposed reveals the complexity of the computer forensic process. Most focus on either the investigation itself or emphasize a particular stage of the investigation.

Kruse and Heiser refer to a computer forensic investigation methodology with three basic components. They are: acquiring the evidence; authenticating the evidence, and analyzing the data [1]. These components focus on maintaining the integrity of the evidence during the investigation.

The United States of America’s Department of Justice proposed a process model for forensics. This model is abstracted from technology. This model has four phases: collection; examination; analysis, and reporting. [5] There is a correlation between the ‘acquiring the evidence’ stage identified by Kruse and Heiser and the ‘collection’ stage proposed here. ‘Analyzing the data’ and ‘analysis’ are the same in both frameworks. Kruse has, however, neglected to include a vital component: reporting. This is included by the Department of Justice framework.

The Scientific Crime Scene Investigation Model proposed by Lee consists of four steps. They are: recognition; identification; individualisation, and reconstruction [1]. These steps only refer to a part of the forensic investigation process. These steps all clearly fall

within the ‘investigation’ stage of the process; there is neither a ‘preparation’ nor ‘presentation’ stage either side.

Casey proposes a framework similar to Lee. This framework focuses on processing and examining digital evidence. The steps included are: recognition; preservation; classification, and reconstruction [3]. In both Lee and Casey’s models, the first and last steps are identical. Casey also places the focus of the forensic process on the investigation itself.

The Digital Forensics Research Working Group (DFRW) developed a framework with the following steps: identification; preservation; collection; examination; analysis; presentation, and decision [4]. This framework puts in place an important foundation for future work and includes two crucial stages of the investigation. Components of an investigation stage as well as presentation stage are present.

Reith proposed a framework that includes a number of components that are not mentioned in the above frameworks. The full listed components are: identification; preparation; approach; strategy; preservation; collection; examination; analysis; presentation, and returning evidence [5]. This comprehensive process offers a number of advantages, as listed by the authors. For example, a number of the components can be included in other stages of an investigation, as will be shown later.

The model proposed by Ciardhuáin is probably the most complete to date. The steps or phases are also called ‘activities’. The model includes the following activities: awareness; authorization; planning; notification; search for and identify evidence; collection; transportation; storage; examination; hypothesis; presentation; proof/defense, and dissemination [6]. The steps are discussed in depth by the authors of the paper.

From the proposed frameworks mentioned above, the following can be seen quite clearly:

- Each of the proposed models builds on the experience of the previous;
- Some of the models have similar approaches;
- Some of the models focus on different areas of the investigation.

Perhaps the best way to balance the process is to ensure the focus remains on achieving the overriding goal: to produce concrete evidence suitable for presentation in a court of law.

4. Proposed Framework

The previous section outlined several important forensic frameworks. In this section a new framework will be proposed. The aim is to merge the existing frameworks already mentioned to compile a reasonably complete framework. The framework proposed in this paper has three stages. They are: preparation; investigation, and presentation. The previously proposed frameworks’ phases are grouped into these three stages. These stages also comply with the definition of forensics in general. If a forensic investigation

conducted these three stages as a minimum, there would be little doubt that a proper forensics process had been followed.

The aim of this paper is not to propose a complete framework exhibiting a number of finite steps. The grouping of defined steps into three, broad stages ensures a more adaptable framework. The preparation, investigation and presentation stages are illustrated in the following diagram.

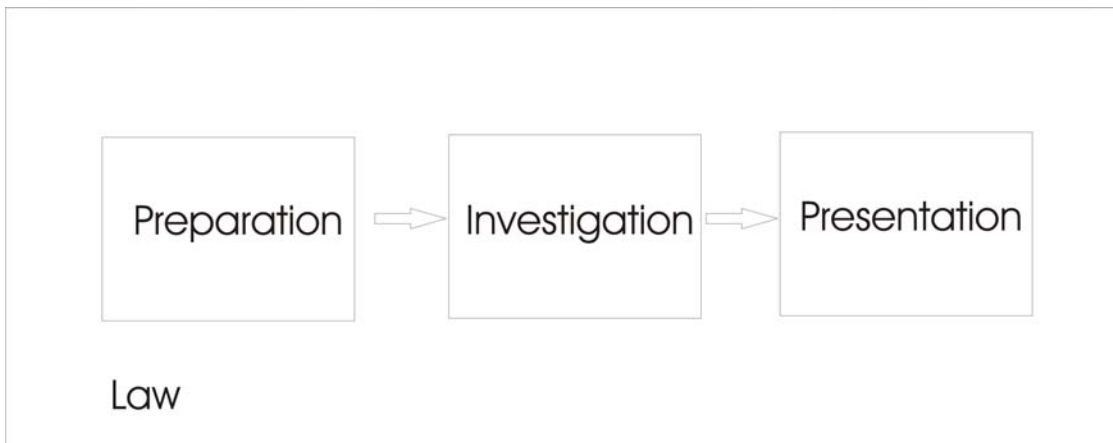


Figure 1: Investigation Stages

Figure 1 illustrates the order in which these stages should be conducted. It is also suggested that this framework should form part of a cycle within the investigation process.

All the phases mentioned by previous frameworks can be incorporated into this framework. This framework also sets a legal base as foundation. The reason for this is so that a clear understanding of what the legal requirements are is established right at the start of the investigation and informs each subsequent step or phase. By focusing on this end goal and deciding what legal norms are to be used, the most applicable framework and integral steps will become clear.

The Preparation stage of the investigation should include the following:

- Standards used in the organization;
- Policies and procedures in place to assist in the investigation;
- Training;
- Legal advice;
- Notification to the correct authorities;
- Documentation of previous incidents;
- Planning, also known as an ‘approach strategy’.

The Investigation stage should include the following:

- Searching for and identifying evidence on a computer;
- Collection of the evidence from the computer (original is duplicated);
- Transportation of the evidence to a secure environment;

- Storage of evidence collected at the scene;
- Examination of the evidence using the proper tools (finding incriminating evidence);
- Analysis (looks at the product of the examination to determine the significance and value of the evidence found).

The final stage of any forensic investigation should include a Presentation stage. This stage is important because it satisfies the key requirement specified by the definition of the word ‘forensic’. This stage will include the following vital steps:

- Presenting the analysis, and
- Proving the analysis.

The steps in the final, Presentation stage of the investigation should prove the hypothesis reached during the investigation. The evidence presented should also hold up in court if the proposed framework and all previous steps were followed correctly.

The proposed framework draws on the experience of other authors [1] [2] [3] [4] [5] [6] and this research has highlighted two important points. Firstly, that knowledge of the relevant legal base prior to setting up the framework is vital since this will have a bearing on the whole investigative process. Secondly, that the process should include three stages — preparation, investigation and presentation — to meet the basic requirements of the definition of the word ‘forensic’.

5. Conclusion

The aim of this paper is to establish a clear guideline of what steps should be followed in a forensic process. These steps, in turn, should enable us to clearly define a framework that can be used in a forensic investigation. A study of previously proposed frameworks revealed that a number of steps or phases overlapped one another and that the difference was mainly one of terminology.

No new steps were added in the framework proposed in this paper. Instead, similar tasks were grouped into the stages required by a forensic investigation. The stages required are preparation, investigation and presentation. This framework can easily be expanded to include any number of additional phases required in the future.

It is, however, important to note that there are several levels of abstraction in the process. Nonetheless, two requirements were identified as needed at every level: the legal requirements of a specific system and documentation of all the steps taken.

6. Bibliography

[1] Baryamureeba, V. and Tushabe, F.: **The Enhanced Digital Investigation Process Model** Digital Forensics Research Workshop. 2004.

- [2] Carrier, B. and Spafford, EH.: **Getting Physical with the Investigation Process** International Journal of Digital Evidence. Fall 2003, Volume 2, Issue 2, 2003.
- [3] Casey, E.: **Digital Evidence and Computer Crime**, 2nd Edition, Elsevier Academic Press, 2004.
- [4] National Institute of Justice. **Results from Tools and Technologie Working Group**, Governors Summit on Cybercrime and Cyberterrorism, Princeton NJ, 2002.
- [5] Reith, M., Carr, C. and Gunsch, G.: **An Examination of Digital Forensic Models**, International Journal of Digital Evidence. Fall 2002, Volume 1, Issue 3, 2002.
- [6] Ciardhuáin, SO.: **An Extended Model of Cybercrime Investigations**, International Journal of Digital Evidence. Summer 2004, Volume 3, Issue1, 2004.
- [7] Van Solms, SH. and Lourens, CP.: **A Control Framework for Digital Forensics**, IFIP 11.9, 2006.
- [8] http://www.askoxford.com/concise_oed/forensic?view=uk : forensic accessed on 7 June 2006.
- [9] http://www.askoxford.com/concise_oed/framework?view=uk : framework accessed on 7 June 2006.
- [10] **Information Technology – Security techniques – Codes of Practice for information security management**. International Organisation for Standardization and the International Electrotechnical Commission. ISO/IEC 17799. 2005.
- [11] Information Security, Audit and Control Association (ISACA). July 2000. COBIT 3rd Edition Control Objectives.
<http://isaca.org>.
- [12] Electronic Communications and Transactions (ECT) Act 25 of 2002. South Africa.
- [13] Promotion of Access to Information Act (PAIA), Act 2 of 2000. South Africa.

M Kohn, JHP Eloff and MS Olivier, "Framework for a Digital Forensic Investigation," in HS Venter, JHP Eloff, L Labuschagne and MM Eloff (eds), *Proceedings of the ISSA 2006 from Insight to Foresight Conference*, Sandton, South Africa, July 2006 (Published electronically)

©The authors

Source: <http://mo.co.za>

A.2 UML Modelling of Digital Forensic Process Models (DFPMs)

```
author    = {M. Kohn and J.H.P. Eloff and M.S. Olivier},  
title     = {UML Modelling of Digital Forensic Process Models (DFPMs)},  
booktitle = {ISSA},  
year      = {2008},  
pages     = {1-13},  
ee        = {http://icsa.cs.up.ac.za/issa/2008/Proceedings/Full/25.pdf},  
crossref  = {DBLP:conf/issa/2008},
```


UML MODELLING OF DIGITAL FORENSIC PROCESS MODELS (DFPMs)

Michael Köhn¹, J.H.P. Eloff², MS Olivier³

^{1,2,3}Information and Computer Security Architectures (ICSA)
Research Group
Department of Computer Science
University of Pretoria
South Africa

¹mkohn@cs.up.ac.za, ²eloff@cs.up.ac.za, ³molivier@cs.up.ac.za

ABSTRACT

A number of forensic processes have been used successfully in the field of Digital Forensics. The aim of this paper is to model some of these processes by using the Unified Modeling Language (UML) - specifically the behavioural Use Cases and Activity diagrams. This modelling gives a clear indication of the limitations of these processes. A UML-based comparison is made of two prominent DFPMs that are currently available in the literature. This is followed by a newly proposed DFPM as developed by the authors.

KEY WORDS

Digital Forensics, Digital Forensic Process Model, Process Modelling, Unified Modelling Language, UML

UML MODELLING OF DIGITAL FORENSIC PROCESS MODELS (DFPMs)

1 INTRODUCTION

The authors of this paper argue that a Digital Forensic Process Models (DFPM) in particular and the field of digital forensic investigations in general can benefit from the introduction of a formal modelling approach. In this paper we propose that UML [1] would be a suitable paradigm for modelling forensic processes. Most of the modelling representations for forensic investigations found in the current literature are made in a rather informal and intuitive way [?, 2]. Thus it is argued that because of the value of a forensic investigation and the formal field of forensic investigation can benefit from introducing a formal modelling approach. Some of these formal modelling approaches include Z-specification, relational algebra and UML modelling. UML modelling is the vehicle chosen for this paper because it provides a structured and behavioural approach that is needed for a forensic investigation. UML is an accepted formal specification for the modelling of processes. This paper will focus on modelling two existing DFPMs, that of Kruse [3] and that of the United States Department of Justice (USDOJ) [4]. The UML that will be used will be limited to Use Case and Activity Diagrams.

Digital forensics has experienced a number of rapid advances to date. This can be seen in the tools that have been developed for forensic investigations such as Encase¹ and Forensic Tool Kit (FTK)². These tools try to encompass the whole digital forensic process into one tool. Encase, which has done this with great success has been accepted in the United States and other countries as a reliable forensic investigation tool [5]. A number of the tools that do not form part of the greater investigation are nevertheless of some use and do assist. Knoppix³ is one such tool that offers limited forensic capability. In the event of encountering a computer that is turned off, it could aid the investigator in possibly finding material without tampering with the integrity of the data. From this it is clear that a digital forensic investigation is made up of multiple facets, which include technology, procedure and legal components. Thus it seems that there is a need for an integrated DFPM.

¹Encase online: <http://www.guidancesoftware.com/>

²Access Data online: <http://www.accessdata.com/>

³Knoppix online: <http://www.knoppix.org/>

A number of DFPMs that have been developed since 2000 aim to assist the investigator in reaching a conclusion upon completion of the investigation. DFPMs used in investigations with success include — but are not limited to — those proposed by Kruse [3], the United States Department of Justice (USDOJ) [4], Casey [6], Reith [7] and Ciardhuin [2].

According to the Oxford online dictionary, the term forensic is defined as “relating to or denoting the application of scientific methods to the investigation of crime” and “of or relating to courts of law”⁴. From this definition it is clear that the ultimate goal of a digital forensic investigation is to present some form of evidence in a court of law using the correct legal procedures with scientific backing.

Closer examination of DFPMs reveals no apparent problem, but a number of questions do arise. Who are the actors that will interact with the system or defined process? Are the role players clearly defined? Do some of these models have short comings? Is it possible to combine some features of existing DFPMs in order to construct an ideal DFPM? To answer these questions, a formal way of comparison is needed to explore some of these problems.

The remainder of the current paper is structured as follows. Section 2 presents some background to the paper and refers to related work performed with regard to forensic processes. In section 3 the Kruse and USDOJ DFPM is modelled in UML using Activity and Use Case Diagrams. Some comments are also made on these two DFPMs. Section 4 contains the result of a brief comparison between the Kruse and USDOJ DFPMs. Section 5 introduces a new integrated model called InteDFPM, which combines the Kruse and USDOJ DFPMs. The paper is concluded in Section 6.

2 BACKGROUND AND RELATED WORK

Digital forensics has been accepted as the process of “analytical and investigative techniques used for the preservation, identification, extraction, documentation, analysis and interpretation of computer media (digital data) which is stored or encoded for evidentiary and or/or root cause analysis” [8]. Most of the proposed DFPMs use some elements of the above definition as point of departure for the development of such a process, such as [3, 4, 6, 9, 7, 2]. These DFPMs are listed in Figure 1. The names of the DFPMs are given in the left margin, while the processes included in each of these models are

⁴The Oxford Dictionary: <http://www.askoxford.com>

listed along the top.

	Acquire	Authenticate	Analyze	Collection	Examination	Reporting	Recognition	Identification	Individualisation	Reconstruction	Preservation	Classification	Presentation	Decision	Preparation	Approach Strategy	Returning Evidence	Awareness	Authorization	Planning	Notification	Transpoptation	Storage	Hypothesis	Proof/defence	Dissemination
Kruse	*	*	*																							
USDOJ			*	*	*	*																				
Casey							*			*	*	*														
DFRWS			*	*	*			*			*		*	*												
Reith			*	*	*			*			*		*		*	*	*									
Ciardhuain				*	*							*						*	*	*	*	*	*	*	*	*

Figure 1: Current DFPMs

The investigation phase of the process constitutes the main focus of most DFPMs. In [4, 9, 7] examination, analysis and collection are included, as this is where most of the activities taking place as part of the investigation are conducted. This focus on investigation is dangerous for a number of reasons. Forensics generally should have a goal of presenting evidence in some form and providing some factual basis to substantiate the investigation’s finding.

In the analysis of some of the DFPMs as seen in Figure 1 one can clearly see the additions that have been made over time. These DFPMs have become increasingly complex. The terminology used in the models is a factor that contributes to creating this unnecessary complexity. Many terms are quite similar to those used in other DFPMs to describe a similar concept. For example, ‘Acquire’ used in the Kruse DFPM and ‘Collect’ used in the USDOJ DFPM would probably amount to the same process — the activities may overlap in many respects.

On examining Figure 1, the reader may agree that there is indeed a need to refine these DFPMs in order to create an integrated model that encapsulates components derived from the given/selected few DFPMs.

3 UML MODELLING

For the purposes of this paper we will be modelling the Kruse and USDOJ DFPMs. The two different types of behavioural UML models that are used

will be the Activity and Use Case Diagrams. Only a high-level system depiction will be presented in all diagrams.

3.1 Kruse

The Kruse model of computer forensics consists of three main processes or phases. The first is acquire the evidence while ensuring that the integrity of the data is maintained. Secondly, authenticate the acquired data, while checking the integrity of the extracted data against the original data. Authentication in digital forensics is usually done by comparing data of the original MD5 hash with the copied MD5 hash [10]. Thirdly, analyse the data without tainting the integrity of the data. This process involves the most intense part of the investigation into the Kruse model.

It is also worth mentioning that the Kruse DFPM is designed specifically for computer-related crimes [3].

3.1.1 UML Activity Diagram

The Kruse DFPM Activity diagram is represented in Figure 2.

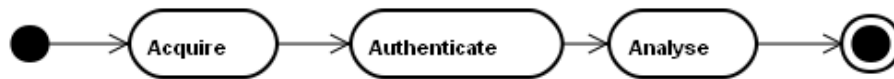


Figure 2: Kruse Activity diagram

The three processes follow one after the other: Acquire, Authenticate and then Analyse. These processes commence with a starting state and end with a finishing state.

3.1.2 UML Use Case Diagram

The Kruse DFPM Use Case is represented in Figure 3. This figure also depicts the different role players.

The three main role players that interact with the system are the Investigator, the Prosecution and the Defence. The Investigator can be specialised to a First Responder, which can be any one of the following: Emergency Response Team or System Administrator. The Prosecution and the Defence will be role players in a criminal matter only. The system consists of three

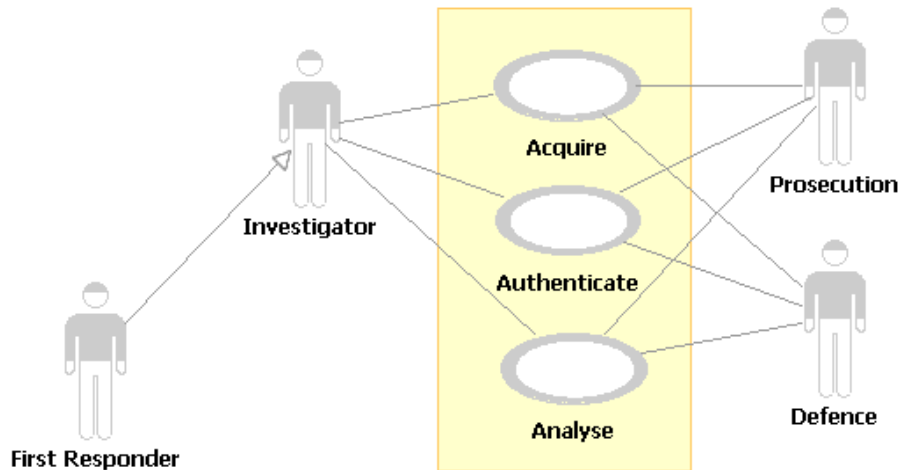


Figure 3: Kruse Use Case Diagram

Use Cases: Acquire, Authenticate and Analyse. The system boundary is depicted by the large rectangle containing the three use cases.

3.1.3 Comments on Kruse DFPM

It should be noted that this is truly an oversimplification of the Kruse DFPM. Each of the use cases in Figure 3 and the activity diagram in Figure 2 will be expanded to include subprocesses.

The activity diagram is clear and it is obvious to see that an investigation starts, runs its course and stops. The main concern is that no real evidence document or report is generated during the investigation. The Kruse DFPM however states in its specifications that documentation and chain-of-custody reports should be maintained during each of the processes.

The use case clearly indicates that the investigator will interact with each one of the processes. Kruse states that in many instances the investigator will not be the same person. The 'Acquire' activity is always encountered by the First Responder and the other two use cases can be performed in a laboratory environment. The court is mentioned throughout the specification, but there is no clear interaction with the system.

3.2 United States Department of Justice (USDOJ)

The USDOJ [4] model accounts for four phases namely collection, examination, analysis and report. The collection phase involves searching for the evidence, recognising that the evidence would be applicable to the specific case, collecting the evidence, while documenting every step taken in the process. The main aim of the second phase, examination, is to reveal any hidden or obscure data. The origin of the original data and its significance are important in providing a visual output that will be used in the analysis process. The third phase involves analysis and the visual product of the examination process is the input to this analysis. Here a case will be built and evidence will be constructed to prove the particular crime. Baryamureeba [?] states that the analysis phase will also determine the probative value, which would actually be the function of the courts. The outcome of this phase would be to produce evidence that would serve to prove the elements of a specific crime. Every step is also documented throughout. The final phase in the USDOJ model is the report phase. During this phase a complete report will be compiled to document the process followed from the beginning of the investigation. The product will be the final evidence report presented in court. Contained in this document is the chain-of-custody report, complete investigation documentation and presentable evidence.

One of the design principles in the USDOJ DFPM is to abstract the process from any specific technology [4].

3.2.1 UML Activity Diagram

The Activity Diagram of the USDOJ DFPM is given in Figure 4.

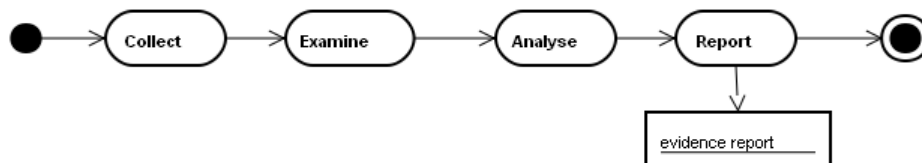


Figure 4: USDOJ Activity diagram

The process commences with a starting state. The data is collected from the digital device, after which it is examined and then analysed. During the

report phase, an evidence report is created as an object output. After the completion of the evidence report, the process stops.

3.2.2 UML Use Case Diagram

The Use Case diagram of the USDOJ DFPM can be seen in Figure 5.

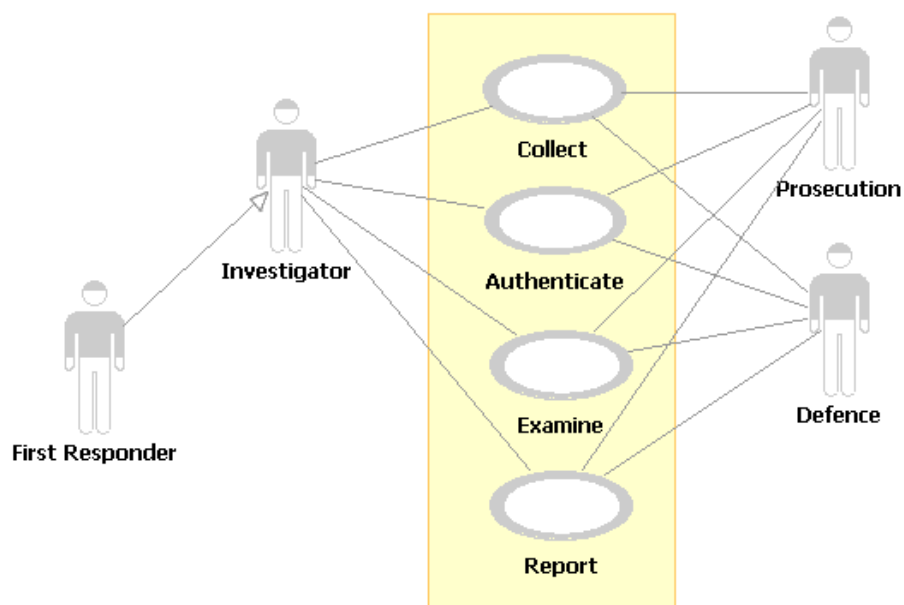


Figure 5: USDOJ Use Case Diagram

In Figure 5 there are three actors: the Investigator, the Prosecution and the Defence. The First Responder is a specialisation of Investigator. An Investigator can be any one of the following: police officer, manager or a forensic investigator. The DFPM is specifically set up for First Responders. There are four use cases in the system, namely, Collection, Examination, Analysis and Reporting.

3.2.3 Comments on the USDOJ DFPM

In the USDOJ Activity diagram, the processes are executed one after the other. There is one apparent difference, which involves the fact that during the Reporting process an evidence report is generated as an output. This

will ultimately be used in a matter before the court. The evidence report will contain all the evidence collected during the investigation, including the chain-of-custody document and presentable evidence. It should be noted that the current paper will not consider what a court considers to be presentable evidence.

The Use Case diagram in Figure 5 does not show the court as a role player that interacts with the system. In the USDOJ specification the court is often mentioned, but no emphasis is placed on the fact that the court ultimately will evaluate the presented evidence report in its finding. There is also no clear indication as to how and when the court must evaluate the document. Nevertheless, an important contribution by the USDOJ DFPM is the fact that an evidence report document is in fact produced.

4 COMPARISON BETWEEN THE TWO DFPMs

Similarities between the Kruse and USDOJ DFPMs are apparent: Firstly, although the models use different terminology ('Acquire' and 'Collect') to describe the first phase, the processes are actually the same. For our purposes we will refer to both as 'Collect' in the remainder of the paper. Secondly, both models have an 'Analysis' phase, resulting in an Analyse process.

There are however also a number of significant differences that cannot be ignored. These include the fact that Kruse's DFPM explicitly validates the integrity of the data in an authentication process, while the USDOJ DFPM includes an examination process. The latter might not always be needed, as data is often hidden and obscured from an investigator. This process will also compromise the integrity of the data. Finally, the DFPM of the USDOJ includes the compilation of a report process, while the Kruse DFPM does not.

5 InteDFPM: INTEGRATED DFPM

The Kruse and USDOJ DFPMs have been modelled using UML Activity and Use Case diagrams. In this section we propose to integrate and expand the two DFPMs into a combined DFPM containing the best elements of both DFPMs. This combined model is called the InteDFPM.

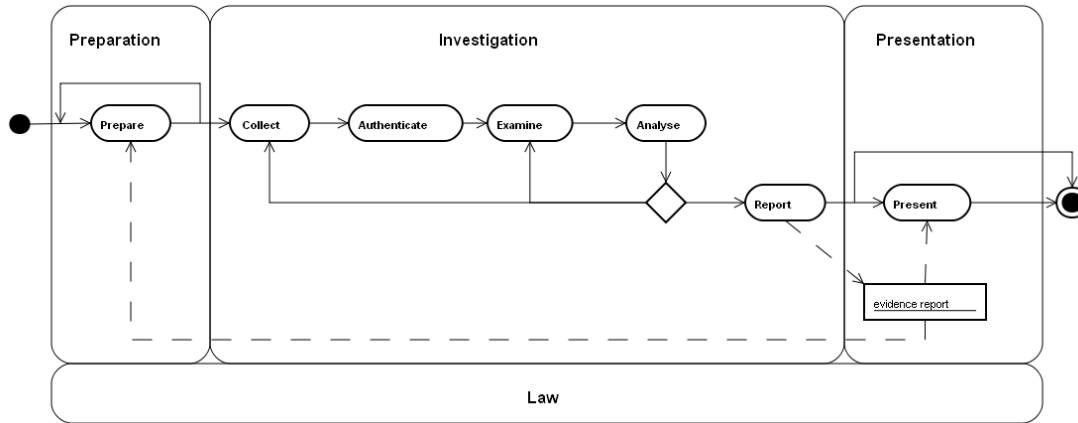


Figure 6: *InteDFPM Activity Diagram*

5.1 UML Activity Diagram

Figure 6 shows the InteDFPM superimposed on a framework proposed by Köhn et al [11]. This framework has three phases: Preparation, Investigation and Presentation. Note that the sub-processes are not included. The law is the foundation for this framework as illustrated by the row along the bottom of Figure 6. The implication is really to ensure that everything is based on sound legal principles so as to withstand legal scrutiny in court.

Two processes have been added to the Activity diagram to integrate with the Köhn framework. These are ‘Prepare’ in the Preparation phase and ‘Present’ in the Presentation phase.

The whole process is triggered by a criminal action (not indicated in Figure 6), which constitutes the starting point. Prepare is the first step and will not be elaborated on here. The rest of the processes follow logically — from Prepare to Collect, Authenticate, Examine and then Analyse. Authentication, is included between the Collection and Examination steps to ensure the data integrity of the data before the Examination is started. Examination can modify the contents of the data such as in the case of hidden files, compressed files and other forms of data obfuscation. The data has to be authenticated before any of this happens. If this is not done, there might be a dispute in court concerning the validity of the material.

A decision point follows the Analysis process. The primary investigator will consider whether to examine more data or to collect more data from the

original source. Once this decision point has reached depletion an evidence report is compiled as part of the Report process. This process will include the compilation of presentable evidence, chain-of-custody reports and complete documentation compiled during the investigation phase. The evidence document is the output of the Investigation phase.

Eventually the evidence report will be an input to the Presentation process. This is where the court will also have the opportunity to evaluate the evidence. It should be noted that the present process can be excluded in the event of not finding sufficient evidence or other relevant factors.

The court finding will be an input to further investigations. This will help investigators to prepare for unforeseen factors that were previously unknown.

5.2 UML Use Case Diagram

Figure 7 illustrated the Integrated Use Case Diagram for the combined Kruse and USDOJ DFPMs.

Figure 7 corresponds to a large extent with the separate Kruse and USDOJ Use Case diagrams. Collect, Authentate, Examine, Analyse and Report are the required use cases.

The system will interact with the following role players: the Investigator can be specialised to be either a First Responder or Other. A specialised Investigator can be any type of Investigator specified by a number of DFPMs. The Investigator will interact with almost all the use cases. It must be noted that it is not always the same person investigating the data. Thus the Investigator does not remain the same person throughout the course of the investigation.

The Prosecution and Defense will be interested in the steps taken in each of the use cases. The Court will examine the evidence report generated by the Report use case. The Court's interaction will change when there is a dispute about the steps taken during investigation. In such a case the Court will evaluate all the use cases. Ultimately, the Court will be interested only in the findings presented in the evidence report, and it will reach a finding based on the presented evidence. The Court will also determine the admissibility and weight of each of the pieces of evidence included in the evidence report.

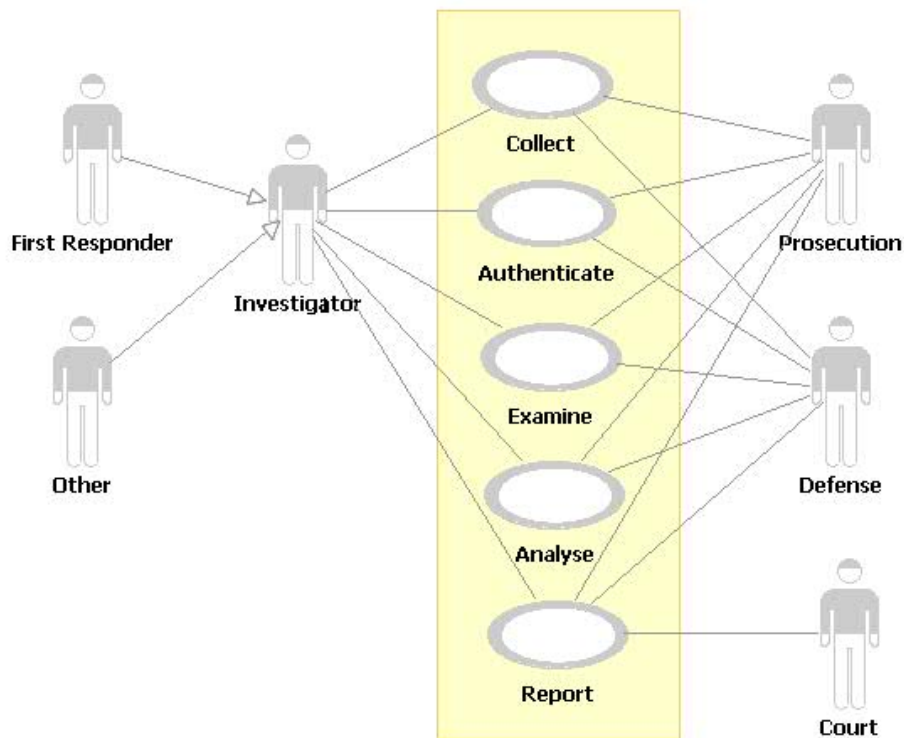


Figure 7: InteDFPM Use Case diagram

6 CONCLUSION

The aim of this paper was to model two DFPMs from the current literature. Activity and Use Case diagrams from the behavioural UML specification were used for this purpose. An Integrated DFPM (InteDFPM) was proposed by combining the Kruse and USDOJ DFPMs, after which the InteDFPM was superimposed on a framework proposed by Köhn [11]. The InteDFPM Use Case Diagram was also presented.

By modelling the DFPMs using UML, it becomes clear that there are a number of shortcomings in the design of the DFPMs. Who are the role players that interact with the system? Neither the Kruse DFPM nor the USDOJ DFPM makes any definitive statement on who the role players should be, except that there must be an Investigator. Furthermore, both DFPMs use different terminology. These problems have been addressed in the paper.

A very important action that is missing both in the above architecture and in the DFPM is the criminal act itself. Future work should explore the possibility of including the criminal act and subsequently including it into the InteDFPM. Other DFPMs should also be investigated for possible incorporation into the InteDFPM.

References

- [1] G. Booch, J. Rumbaugh, and I. Jacobson, *The Unified Modeling Language User Guide*. Addison Wesley, 1999.
- [2] S. O. Ciardhuain, “An extended model of cybercrime investigations,” *International Journal of Digital Evidence*, vol. 3, 2004.
- [3] W. Kruse and J. Heiser, *Computer Forensics: Incident Response Essentials*. Addison Wesley, 2002.
- [4] Technical Working Group for Electronic Crime Scene Investigation, *Electronic Crime Scene Investigation: A Guide for First Responders*, United States Department of Justice, 2001.
- [5] T. Wilsdon and J. Slay, “Towards a validation framework for forensic computing tools,” in *First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE’05)*, 2005, pp. 48–55.
- [6] E. Casey, *Digital Evidence and Computer Crime*. Elsevier Academic Press, 2004.
- [7] M. Reith, C. Carr, and G. Grunsch, “An examination of digital forensic models,” *International Journal of Digital Evidence*, vol. 1, 2002.
- [8] S. van Solms, C. Louwrens, C. Reekie, and T. Grobler, “A control framework for digital forensics,” in *IFIP 11.9*, 2006.
- [9] Digital Forensics Research Workshop, *A Road Map for Digital Forensics Research*, 2001. [Online]. Available: <http://www.dfrws.org/2001/dfrws-rm-final.pdf>
- [10] A. Culley, “Computer forensics: past, present and future,” *Digital Forensics*, vol. 8, pp. 32–36, 2003.
- [11] M. D. Köhn, J. H. P. Eloff, and M. S. Olivier, “Framework for a digital forensic investigation,” in *Information Security South Africa (ISSA)*, H. S. Venter, Ed., 2005.

A.3 Isolating a cloud instance for a digital forensic investigation

```
author    = {W. Delpport and M. Kohn and M.S. Olivier},  
title     = {Isolating a cloud instance for a digital forensic investigation},  
booktitle = {ISSA},  
year      = {2011},  
ee        = {http://icsa.cs.up.ac.za/issa/2011/Proceedings/Research/Delpport_Olivier_Kohn.pdf},  
crossref  = {DBLP:conf/issa/2011},
```

Isolating a Cloud Instance for a Digital Forensic Investigation.

Waldo Delpont

Information and Computer Security
Architectures Research Group
Department of Computer Science
University of Pretoria
South Africa
Email: wdelpont@cs.up.ac.za

Martin S. Olivier

Information and Computer Security
Architectures Research Group
Department of Computer Science
University of Pretoria
South Africa
Email: molivier@cs.up.ac.za

Michael Köhn

Information and Computer Security
Architectures Research Group
Department of Computer Science
University of Pretoria
South Africa
Email: mkohn@cs.up.ac.za

Abstract—Cloud Computing is gaining acceptance and increasing in popularity. Organizations often rely on Cloud resources to effectively replace their in house computer systems. In a Cloud environment an instance is typically accepted to be a virtual system resource established within that Cloud. Multiple instances can be contained a single node. The Cloud itself consists of multiple nodes. The Cloud structure has no predefined or fixed boundaries.

Digital Forensics (DFs) can be considered the science of finding a root cause of a particular incident. Isolating the incident environment is generally accepted within the Forensic Community to be an integral part of a Forensic process. We consider this isolation is also needed in a Digital Forensic Investigations (DFIs). The isolation prevents any further contamination or tampering of possible evidence.

In order to isolate the incident the Cloud instance is isolated. The node instance is effectively placed in a controlled environment to enable a controlled DF investigation to be conducted. This paper will introduce possible techniques to isolate these Cloud instances to facilitate an investigation. The techniques include, but are not limited to Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, Man in the Middle (MITM) and Let's Hope for the Best (LHFTB). A discussion of each of these techniques will be given. This discussion will include a description of each techniques, the advantages and disadvantages of using the techniques and the visibility of the techniques.

Index Terms—Cloud Computing, Digital Forensic.

I. INTRODUCTION

As long as people are involved there will be crime. The need for Digital Forensics exists because some of these crimes or other incident is taking place on computer system [1]. The introduction of clouds complicated the digital forensic process. There is a belief that a digital investigation a cloud can be difficult to do. The need to have formal and proven methods to conduct a digital investigation on a cloud became apparent.

As computer related technologies continuous to expand a logical expansion in online technologies was cloud computing [2]. Cloud computing enables service providers to provide virtual systems to their clients. It enables the service providers to maintain a large number of independent services in a single cloud infrastructure.

In a cloud an instance must be isolated when it becomes apparent that an incident happened on that particular instance.

This isolated helps preserve the integrity of the evidence collected from the instance. One of the problems to preserve the integrity of an instance is an attribute of clouds [1]. In a cloud the data from instance may share the storage of multiple instances and may not be in a constant place in the cloud. To preserve the integrity of the evidence the location on the cloud must be known and must be protected from tampering and contamination. Another complexity is that other instances on the same node may belong to other users. Users expect at least availability and privacy of their instance provided by the service provider [3]. The Digital Forensic process must be done in a manner that will not result in the privacy of other instances being lost and the availability of the instance must be effected in the smallest manner possible.

This paper intends to introduce new techniques to isolate instances on a cloud. These techniques are Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, MITM and LHFTB In Section 2 a brief description of Digital forensics is provided. It also provides information about Computer, Network and Cloud Forensics. Section 3 provides a description of Cloud Computing. In Section 4 we discuss why there is a need to isolate the instance on a cloud. Section 5 introduces the techniques to isolate an instances and compares them to each other. A summary and the conclusion are given in Section 6.

II. DIGITAL FORENSICS

To define digital forensics one first needs to define forensics. Forensics is a formal and proven approach to the gathering of evidence and processing of a crime scene. Sometimes used in the court of law [4]. From this definition it can be said that digital forensics must be based on sound scientific methods and techniques. It can be added that digital forensic can aid in the court of law. The digital forensic process helps in answering the who, what, when, where and how of an investigation [1].

In a digital forensics process a live or dead analysis can be followed [1]. The normal computer forensic process uses dead analysis, in a dead analysis the system is turned off as soon as the examination team acquires it and images are made of

the storage mediums, the analysis is then conducted on the images [5]. The other approach is a live analysis where the computer is kept on and evidence gathered from the computer in the environment that is on the system. There are advantages and disadvantages of both. The main disadvantage of a dead analysis is the fact that some information may be lost because it is in a buffer or the RAM. The problem with a live analysis is that the evidence can be destroyed or modified without the intent to do so.

In order to obtain admissible evidence a well defined forensic process needs to be followed. Cohen [6] proposes a model for the digital forensic examination that consists of seven phases. The phases are Identification, Collection, Transportation, Storage, Examination and traces, Presentation and Destruction.

Identification : In the identification phase possible evidence is identified as evidence.

Collection : Once the evidence is identified it is collected. The integrity of the evidence must be preserved while the evidence is gathered.

Transportation : The collected evidence must be transported. The evidence is collected at a crime scene and the rest of the digital forensic examination will happen at a different location. The evidence is moved to an examination lab where there is the necessary equipment to do a digital forensic examination. The normal manner to ensure that the evidence integrity is kept is to copy the evidence and keep the original in a safe place and move the copy.

Storage : The digital forensic process may be a lengthy process, while the examination is on-going or even done the evidence needs to be stored in a manner so that the evidence will not degrade and become inadmissible.

Examination and traces : The Examination and traces phase consist of four sub categories, they are Analysis, Interpretation, Attribution and Reconstruction [6]. The Examination phase will try to explain route of evidence, from creation to state it is in now. The last step is to try and create the same output from the original evidence.

Presentation : The next step is to present the findings. The presentation can take various forms. A report containing the outline of the examination process and the evidence that was found can be created. In some cases the examiners should be able to testify in the court of law. The report and the testimony content will be summaries of the previous phases. If a presentation contains faults or inaccuracies it will have a negative effect on the evidence that was gathered even making the evidence inadmissible.

Destruction : The last step in the digital forensic examination is the demolition of the gathered evidence. The period can range from immediate destruction to seventy years after the case. The time period is influenced by various factors including data sensitivity and case severity.

Documentation is a continuous process and needs to happen in all phases of the digital examination. One of the main aids to help preserve the integrity of the evidence is documentation. The documentation should at least include the

name of the evidence, the place the evidence is gathered. The documentation should also include the processes followed in identifying, retrieving, storing and transporting the evidence. The documentation should also mention the chain of custody when the examination was in progress. There have been several cases where the outcome of the case was influenced by the documentation.

A. Computer Forensics

Computer forensics is related to the forensics of computer components and their content [7]. The field of computer forensics attempts to narrow the search for evidence to the computer itself, the content on the computer and devices attached to the computer.

B. Network Forensics

Network forensics was introduced to help solve attacks on networked systems. The evidence of a Network Forensic investigation is collected from the data sent over the physical network consisting of a network containing at least two computers [7]. One method of gathering possible evidence is by capturing and analysing network traffic. Other sources of network forensic evidence are logs from servers, users browsers settings and router information. Network Forensics can be done live. The problem with live network forensics is that significant hardware resources on a network consisting of more nodes than a typical home network [8].

C. Cloud Forensics

Cloud forensics is Digital Forensics applied on Cloud Computing [9]. Cloud Forensics is a subset of Computer Forensics as a cloud runs on a network and consists of network equipment. Cloud Forensics also entails Computer Forensics as a cloud consists of nodes that are computers. A cloud also consists of instances which are a special case of a computer instance. This means that Cloud Forensics ties Computer and Network Forensics together. This does not mean it is Digital forensics. Cloud Forensics is also a sub category of Digital Forensics.

III. CLOUD COMPUTING

Cloud computing is a relative old term but has been adopted quickly the last couple of years [2]. Cloud Computing builds on different forms of distributed computing. It ties the distributed computing together with virtualization. Cloud Computing enables a service provider to provide a flexible, cost effective and on-demand infrastructure to its clients instead of the clients running their own infrastructure. There is no standardized definition for cloud computing [10]. For the purpose of this paper Cloud Computing will be defined as a distributed computing architecture providing flexible, cost effective and on-demand infrastructure to users over some form of network by using virtualization to create virtual resources on the abstracted hardware.

The users of cloud infrastructure are provided a virtual computer with which can be interacted usually through the

Internet [1]. This virtual computer can also be known as an instance. Normally an instance can be accessed from anywhere in the world depending on the security setup. The instance can be a small instance used by a single user to store backups of files or it can be a server running the website and database of a company. A client only pays the service provider for services rendered. If the requirements of the client change it is an easy process to change the scope of the instance to accommodate the new requirements of the client. If a new instance is required the task of stating and setting up an instance is trivial. On most Cloud systems an instances can be launched from an image that contains most of the needed software. This images were created with a specific task that it needs to perform. An image might be created that serves as a basis for a web server and another image for home computers.

The service provider is responsible for maintaining the Confidentiality Integrity and Availability (CIA) of the instances on a hardware level. The user is responsible for protecting the CIA on a higher level e.g. the content of files [11].

The value that can be added from Cloud computing is significant primarily to small and medium sized businesses [12]. It enables businesses to have access to servers without the initial start-up cost and they have no maintenance cost on hardware level. As the businesses grows their infrastructure can easily be changed to adapt to the growth.

Cloud Computing is growing and is estimated to become a billion dollar industry this year [9]. The reason for this is that some of the largest IT related companies has implemented or is implementing cloud computing. Some of the large companies are Google, Microsoft, IBM and Amazon [11], [1]. These company state they will provide CIA to their customers by using various techniques.

IV. THE NEED TO ISOLATE A CRIME SCENE

In a “real word” forensic process the crime scene is isolated [13]. The isolation helps protect the possible evidence from contamination and loss of continuity. If any contamination happens or the continuity is lost all the evidence gathered from the investigation admissibility might get lost. To help protect the admissibility of the evidence a crime scene is dived into separate parts to aid in the isolation. These parts can only be entered by authorized personnel using authorized manner. A path is sent out where the personnel can walk in and around the crime scene. A log is kept of where personnel are and what they are doing.

Multiple instances can reside on a single cloud node. A user of an instances expect that there is confidentiality in place to protect the data on that instance [11]. When a Digital Forensics Investigation is done on a cloud there must be methods in place to prove that the privacy and confidentiality of the users has been protected. We prove to users that their instances CIA was protected by using tested method that are accepted and known to protect clients CIA. To have proven methods to follow in an DFI the methods must be based on reliable technique to collect and preserve evidence.

In the cloud environment we want to protect the instance that we are going to investigate from tampering and contamination. In order to provide admissible evidence the evidence needs to be protected. Gathering evidence is one of the aims of a DFI. If the evidence is suspected to be invalid by any means it will not be able so serve as admissible evidence. In order to add the evidence admissibility the evidence needs to be protected from contamination and tampering.

Is a normal DFI it is accepted that assets may be seized. As stated above in a cloud environment there can be multiple instances running on a single cloud node. This makes it improbable that assets my be seized [2].

We feel it is necessary to isolate an instance on a cloud node. The controlled environment will aid in protecting the instance from contamination and tampering. This controlled environment where an instance is isolated is going to be used for the DFI.

V. ISOLATION OF A CRIME SCENE IN A CLOUD

As stated a cloud node can contain multiple instances and the nodes needs to be cleared when doing an DFI. The methods for clearing include moving the suspicious instance to another node or moving the uninvolved instances too other nodes. The CIA of the other instances is protected when moving the suspicious instance. This can result in the loss of possible evidence. When we move an instance data may get lost or the instance might realize it is being moved and tamper with evidence. To protect the evidence the other instances are moved from the node. Care must be taken when moving the instances in order to protect their CIA.

When isolating a cloud instance the investigator must consider a we live or dead analysis is applicable. The techniques that are suited for each type of analysis may differ. When doing a live forensics analysis we want to stop the instance from tampering with evidence. If a dead analysis is chosen the other instances must be protected from the consciousness of the power outage. It must be decided what looses and risks are acceptable before staring with an DFI in a cloud.

The techniques that are proposed are Instance Relocation, Server Farming, Failover, Address Relocation, Sandboxing, Man in the Middle (MITM) and Let’s Hope for the Best (LHFTB).

A. Instance Relocation

Instance relocation means that an instance is moved inside the cloud. This is done by moving the instance from one node to another. This can be done manually or automatically. When it is done manually the administrators of the cloud will usually move the instance by some means. Automatic relocation is done by the cloud operating system. When the instance is moved it can be done in three possible ways. The existing instance can be ended and a new one created. Another option is where a new instance is created and the old instances is destroyed once the new instance is created. The other option is where the instance is logically moved. This entails that the

data is moved from one node to another without the instance being destroyed.

To move an instance we divide an instance in three units that must be moved. These units include data on secondary storage, the content of the virtual memory e.g. swap memory and the running processes.

1) *Manual Instance Relocation*: When an instance is moved manually it is up to an administrator or investigator to move the instance. The possible methods to manually relocate an instance is a subset of the methods giving above. Either the existing instance can be ended and a new one created or a new instance is created and the old instance is destroyed once the new instance is created. When an existing instance is ended all of the units must be protected or saved. There is a verity of methods available. The storage can be copied to an image file using tools including `dd_rescue` [14]. The content of the virtual memory can also be written to files also using `dd_rescue`. Once all the files are created the original instance is removed and a new instance created. The new instance will receive all the content of the old instance. The new instance can be created with the same network address as the old instance but on a different node. One problem is the process. It is hard to store process in a manner that can restore the instances later to the new instance. The other method involves creating a new instance and moving all of the units to the new instance and then removing the old instance. Once the new instance is created the storage content can be moved to the new instance. The running process can be moved using methods designed to move processes between computers [15]. Some of the proposed methods by Milošević have been testes and proven as valid process moving methods. The virtual memory is harder to move and care must taken to move it. It is difficult to move because while the instance is being used the virtual memory is in a constant state of change. When the new instance has all the units of the original instance the original instance is removed. The new instance must then be set to have the same network addresses as the old instance to receive the network traffic.

2) *Automatic Instance Relocation*: The cloud operating system will move the instance in the Automatic Instance Relocation technique. The methods used to move is implemented by the creators of the cloud operating system. The creators must insure the method can be proven and is reliable. The means it uses may be the same as described above or be other methods. The reasons for an instance to be moved by the system includes, but is not limited to, the administrator or investigator asking the system to move an instance and load balancing. The administrator or investigator asks the system to move an instance for the purpose of an DFI other possible reasons might include conflict of interest between instances on a single node. The load balancing functionality might be implemented in cloud operating systems. When the systems notices that instances on a node are extremely resource dependent and other nodes have lost of resources available it might try to balance the load of the nodes. This functionality can be used by an investigator. The investigator

forces an instances to be resource intensive then the system will move it away from the node. The node is cleared by the system itself.

These instance relocation techniques enables a node to be cleared for an DFI. The way in which they are moved can be controlled and monitored. The service provider can prove to its customers that it is protecting their CIA. The cloud operating system manufactures can implement reliable methods to do a successful DFI on their cloud system.

If the instances are moved in a manner that violates their CIA the service provider may be influenced negatively. The customers may experience downtime of their instance or loss of data. They can then leave the service provider or charge for down time. If the cloud operating system moves the instances it might be hard for the investigator to prove they are using reliable methods. This adds reliance from the manufactures to be involved in an DFI.

These techniques can be hard to implement. As discovered by experimentation the storage media can be easily be copied but it is a non trivial task to send it to the new instance and keep that instance running. The hard drives where copied as a whole and the process of overwriting system files can result in the new instance failing. The process can be moved if the operating system of the instance supports the functionality. This can be an effective method to clear a cloud node if there is no build in functionality if implemented correctly.

B. Server Farming

A server farm is a multi-node system [16]. In web server farms the web-site is split over two or more nodes. The user interacting with the website only sees the functionality of a single server. In the server multiple nodes are used to deal with the website. The server farm uses some form of routing to route request between nodes from users. The server farms use distribution technologies to enable this service. This distribution aids in the Quality of service of the website. There is no single point of failure. When a node fails the router will stop sending request to that node.

In a cloud multiple instances can be created that is logically the same instance but over multiple nodes. Multiple instances work together and appears as one instances. The load for the logical instance is spread over the actual instances. When a single node fails the remaining instances will continue to function. This enables examiners to terminate instances on the same node and to isolate the suspicious instance on a node. Small server farms of the uninvolved instances needs to be created at the start of the investigation. They can be created by adding just one instance to the farm. This means there will be two instances in the farm. Once the server farm is working the original instance can be removed.

To enable Server farming on clouds it needs to be implemented by the cloud operating systems creators. The cloud infrastructure must provide for the rerouting of network traffic. The cloud infrastructure must also allow multiple instances to exist over multiple nodes that can interact. The process of creating a server farm for the sole purpose of an DFI might

put unnecessary load on the cloud. If the cloud provides the functionality to provide availability to its clients it can just be used to aid an DFI.

Although Server farming can be resource expensive it can aid the service provider manage their clients CIA. The instances can be removed from the node without a loss of availability.

This technique relies on cooperation from the cloud operating system creators. If the implementation is wrong the DFI can result in the loss of CIA of other users on the cloud.

C. Failover

In a failover environment there is at least one server replicating an other server [17], [18]. The replicating server is commonly known as the backup server. If the primary server fails the backup server can immediately take over. This means that all the data and processes of the primary server is replicated on the backup server. Failover was introduced to provide high availability for websites. In 1999 E-Bay lost an estimated 5 million dollars when there servers failed [19]. If there where failover technology implemented this problem could have been prevented. Failover can be provided in several ways. Possible methods are Client-based failover, DNS-based failover and IP-address take over [17]. In Client-based failover the client knows of both the primary server and the backup server. If the primary server is unresponsive the client communicates with the backup server. When using DNS-based failover the DNS server redirects traffic to the backup server when the primary server fails. In IP-address take over the backup server takes over the IP-address of the primary server when the backup server notices the primary server has failed.

To implement failover an adaptation of the IP-address take over will be used. The original instance is replicated creating a backup instance. Once the original instance is killed the backup instance will take over the IP of the original instance. The method in which the instance is replicated is open to the DFI team. To replicate the same units as for Instance Relocation needs to be moved. The units include data on secondary storage, the content of the virtual memory e.g. swap memory and the running processes.

The failover technique will result in virtually no availability loss of the instance. The failover can be implemented by the DFI team. There is almost reliance on the cloud operating system manufacturers. This technique also does not use a lot of resources of the cloud.

There will be a loss in availability and some data may get lost. This loss can cause loss of CIA. If the loss is acceptable the method may be used.

D. Address relocation

Address relocation can be seen as when network traffic is relocated to other computer. The network traffic is directed by either the router or DNS server to other computer because of some reason. A network packet is sent so a specific IP address. The computer which has the IP address might be unavailable

and the packet is sent to other computer without the sender being aware of the change. The rerouting mechanism also makes it appears as if the packets that are returned to sender are sent from original computer. The Address relocation can be seen as a special case of the DNS-based failover method. A backup server is maintained in some or other form. When it is detected that the main computer has failed the traffic is routed to the backup server.

Create a replica instances of the uninvolved nodes. Once they are created use the clouds internal network DNS server or other method to redirect all traffic to the new instance. If the clouds DNS server cant be changed use an extra instance. This instance will serve as a middle ground to the instance and the internal DNS. The instance is another level of DNS. The instance can be used to interact with multiple instances but is controlled by the administrator of the system and not the system. The top level DNS can be configured when an DFI is in progress to redirect all the traffic to a replica created. The primary instance can then be removed.

The switch overtime from primary instance to replica instance can be insignificant if the replication is correctly implemented.

This method relays on replication working correctly. If the replication is incorrect the Address relocation is inefficient. The replication will help keep the instances CIA. This method also adds the complexity of two DNS server running on the cloud. The Service Providers might argue this technique is a waste of cloud resources.

E. Sandboxing

In program security a sandbox is a controlled environment where a program can execute [20], [21], [22]. A program cannot escape the sandbox and cannot effect other programs outside the sandbox. It is used to stop malicious programs from harming other programs on the same computer by limiting the interactions between the programs. A sandbox is created by software controlling the interaction of the program with other programs.

In terms of a cloud we will isolate an instance by placing it in a sandbox. The sandbox will prevent it from interacting with other instances. The other instances will then be protected from harm. To enable this functionality two approaches can be followed. The cloud operating system can launch a sandbox application. The other option is where the investigator launches an application on the instance. This application will monitoring all communication channels. It creates a virtual box around the instance. The instance can do what it wants inside the box but will not be able to do anything outside the box. This application will run on the network of the instance. Networking is the communication method an instance has with the rest of the cloud. The sandbox application will monitor network traffic and block were needed.

The sandbox techniques aids the service provider in protecting the CIA of the other instances. The other instances are protected while the DFI is being done and the instance that is being investigated is boxed in and continues as normal.

Information can be lost while the instance is sandboxed. The instance might realize that it is placed in a sandbox and try to tamper with possible evidence. It might be difficult to block the network traffic in a manner that can be proven to be accepted in the field of DFI.

These techniques help the service provider in the CIA of other instances but evidence loss can occur. The instance can be sandboxed while the other instances are moved from the node. Once they are removed an DFI can be performed. This DFI can be a live or dead investigation. Once all other instances are off it can be decided which method is preferred. The sandbox may add a live forensics as the instance is kept in a controlled environment.

F. Man in the Middle

The term MITM can be used in network security to describe an Man in the Middle Attack (MITMA) [20]. An MITMA is a combination of potential threats in computer security. These threats include interception, interruption, modification and fabrication. Interception is where an other entity gains access to an assist. Usually the interception is unknown to the sender and receiver. The assist is delivered to the receiver and a copy to the entity. Interruption is where an assist is lost. The assist may be blocked, deleted or any other form of destruction of the assist. Modification is where the assist gets modified in some way. The receiver receives a changed version of the assist. Fabrication is where a new assist is created. The sender sends the original assist and the receiver receives the assist created by the entity. An MITMA is where the entity places itself between the sender and receiver. It receives all the assists from the sender and sends assists to the receiver. The assists are vulnerable to interception, interruption, modification and fabrication.

To allow an MITM to be used in clouds to assist in an DFI an entity will be created that exists between the cloud instance and the hardware of the cloud. This entity can be part of or use the virtualization software of the cloud. The data going from the instance to the hardware and from hardware to the instance can be analysed. The hardware includes but is not limited to the network, CPU, RAM and hard drive. This enables a forensic process to be done on all the data being used in an instance. The forensic process will be a live forensic investigation.

The entity can be kept inactive when there is no suspicion of wrong doing on an instance. This minimizes cost of being ready for an DFI in term of computation cost. When there is a suspicion of an instance the MITM entity can be activated. Once activated the MITM entity will analyse all actions of the instance and the data going from and to the instance. It will stop the instance from deleting data on storage media and RAM. The MITM entity will allow an investigator to access the resources of the instances without the instance being aware of the analysis. The investigator can also observe the actions the instance is or trying to perform. To enable the MITM to exist between the instance and hardware it must be added by the creators of the cloud software or by a using

company. To aid in the evidence admissibility the MITM must be implemented using proven methods.

An advantage of this method is that the instance does not know it is being analysed. It can prevent the instance from destroying evidence also from doing the suspicious activity. Other advantages include that the instance can function as expected and other instances will not be affected by the DFI. The techniques also aid in the protection of other instances. The instances that are being investigated can logically be blocked from communicating with other instances.

A potential problem is implementing it. There is a reliance on the cloud operating system manufacturers. The cloud operating system manufacturers might not feel the need to add this functionality. To enable a company to add the functionality the software must be reverse engineered. Once the software is reverse engineered the MITM must be added. Both of these approaches has problems. The cloud operating systems creators might not make the functionality available to only its own employees or might create the functionality sub standard. The admissibility of evidence might be lost because of bad implementations. The problem with reverse engineering is the reverse engineering. Most software packages have a term of use. This term usually permits the reverse engineering of the software. This opens a change that the company using an MITM they added might be sued. There is also the problem that proving the implemented as correct can be challenging because it was not implemented in a normal manner.

We believe that these techniques has the potential to be a valid techniques to do an DFI in clouds if the cloud manufacturers agrees to implement a reliable and proven MITM functionality in their software. The MITM might also be used to with other techniques. The other techniques clear the node of instances and a controlled live forensic process can be followed on the instance.

G. Let's Hope for the Best

The usual procedure is followed for doing an DFI [5] in the LHFTB technique. The node is turned off and taken to a controlled environment. Images of the hard drives of the node are made. These images are then analysed. A potential difficulty is that a node can contain multiple instances. The hard drives of the node can contain multiple virtual hard drives. The investigator must know how the cloud operating systems stores information. Information from other instances may not be used. This violates the CIA of the other users. It can be difficult to piece together the original virtual hard drive and credible evidence may get lost.

A possible advantage for LHFTB is that a suspicious instance has no warnings. This means that the instance possible will not interfere with possible evidence.

A potential problem is that on a single node can contain multiple instances. These instances can be lost. This violates the agreement between the service provider and the client. Uninvolved client lost their availability. Another problem is that running information is lost. The information in RAM and the network is lost and cannot be used.

We propose that this technique is not used on its own and that the other technique must be combined. First an MITM must be started on the instance that needs to be investigated. The RAM and other information can be acquired from the MITM. Other instances must then be moved from the node. The MITM also aids in the instance moving process it protects the instances being moved and keeps the investigated instance in a controlled environment. Then the power must be removed and images made. This creates a controlled and monitored DFI.

VI. CONCLUSION

Cloud computing is a rapid growing technology [11]. A DFI might be hard to do in a cloud because of various reasons [1]. On a cloud one node can contain multiple instances. The possible evidence can share a drive with several other instances data. The evidences needs to be protected. In the “real world” a crime scene is isolated to protect the evidence. If a digital crime scene on the cloud is isolated it can aid the evidences admissibility.

This paper introduced possible techniques to isolate an instance on a cloud. The techniques introduced where Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, MITM and LHFTB. A brief discussion of each of these techniques where given.

It can be seen from the discussion that no one technique proposed a perfect solution. The techniques may be combined to provide a feasible method to isolate a cloud instance. The differences between some of the techniques are small and may be seen as the same. The differences of the techniques allows them to be used in different environments.

We want to implement the techniques in the future to test them in an experimental environment.

REFERENCES

- [1] S. Biggs and S. Vidalis, “Cloud computing: The impact on digital forensic investigations,” in *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*, November 2009, pp. 1 – 6.
- [2] M. Vouk, “Cloud computing - issues, research and implementations,” in *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on*, June 2008, pp. 31 – 40.
- [3] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J. Molina, “Controlling data in the cloud: outsourcing computation without outsourcing control,” in *Proceedings of the 2009 ACM workshop on Cloud computing security*, ser. CCSW '09. New York, NY, USA: ACM, 2009, pp. 85 – 90. [Online]. Available: <http://doi.acm.org/10.1145/1655008.1655020>
- [4] “Definition of Forensic,” April 2011, oxford Dictionary. [Online]. Available: <http://www.oxforddictionaries.com/definition/forensic?view=uk>
- [5] M. A. Caloyannides, N. Memon, and W. Venema, “Digital forensics,” *Security Privacy, IEEE*, vol. 7, no. 2, pp. 16 – 17, March 2009.
- [6] F. Cohen, *Digital Forensic Evidence Examination*, 2nd ed. Livermore, CA: Fed Cohen & Associates, February 2010.
- [7] B. Fei, “Data visualisation in digital forensics,” Master’s thesis, University of Pretoria, 2007.
- [8] V. Corey, C. Peterman, S. Shearin, M. Greenberg, and J. Van Bokkelen, “Network forensics analysis,” *Internet Computing, IEEE*, vol. 6, no. 6, pp. 60 – 66, November 2002.
- [9] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, “Cloud forensics: an overview,” *IFIP International Conference on Digital Forensics*, vol. 7, 2011.
- [10] I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud computing and grid computing 360-degree compared,” in *Grid Computing Environments Workshop, 2008. GCE '08*, nov. 2008, pp. 1 –10.
- [11] R. Lu, X. Lin, X. Liang, and X. S. Shen, “Secure provenance: the essential of bread and butter of data forensics in cloud computing,” in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '10. New York, NY, USA: ACM, 2010, pp. 282–292. [Online]. Available: <http://doi.acm.org/10.1145/1755688.1755723>
- [12] G. Reese, *Cloud Application Architectures : Building Applications and Infrastructure in the Cloud*, 1st ed., A. Oram, Ed. O’Reilly Media, 2009.
- [13] P. White, *Crime Scene to Court: The Essentials of Forensic Science*, 3rd ed., P. White, Ed. Royal Society of Chemistry, 2010.
- [14] K. Garloff, “dd_rescue,” Computer Program, version 1.23. [Online]. Available: <http://www.garloff.de/kurt/linux/ddrescue/>
- [15] D. S. Milojević, F. Douglas, Y. Paindaveine, R. Wheeler, and S. Zhou, “Process migration,” *ACM Comput. Surv.*, vol. 32, pp. 241–299, September 2000. [Online]. Available: <http://doi.acm.org/10.1145/367701.367728>
- [16] E. Casalicchio and S. Tucci, “Static and dynamic scheduling algorithms for scalable web server farm,” in *Parallel and Distributed Processing, 2001. Proceedings. Ninth Euromicro Workshop on*, 2001, pp. 369 –376.
- [17] K. Singh and H. Schulzrinne, “Failover, load sharing and server architecture in sip telephony,” *Computer Communications*, vol. 30, no. 5, pp. 927 – 942, 2007, advances in Computer Communications Networks. [Online]. Available: <http://www.sciencedirect.com/science/article/B6TYP-4KYY4GT-1/2/4faf31d97db80455a5a5eb986648fcb6>
- [18] I. Kuzminykh, “Failover and load sharing in sip -based ip telephony,” in *Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2008 Proceedings of International Conference on*, February 2008, pp. 420 – 422.
- [19] R. Zhang, T. Abdelzaher, and J. Stankovic, “Efficient tcp connection failover in web server clusters,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, march 2004, pp. 1219 – 1228 vol.2.
- [20] C. P. Pfleeger and S. L. Pfleeger, *Security in Computing*, 4th ed. Prentice Hall, 2006.
- [21] C. Greamo and A. Ghosh, “Sandboxing and virtualization: Modern tools for combating malware,” *Security Privacy, IEEE*, vol. 9, no. 2, pp. 79 –82, April 2011.
- [22] M. Smith, T. Friese, M. Engel, and B. Freisleben, “Countering security threats in service-oriented on-demand grid computing using sandboxing and trusted computing techniques,” *Journal of Parallel and Distributed Computing*, vol. 66, no. 9, pp. 1189 – 1204, 2006, security in grid and distributed systems. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WKJ-4K66F0M-1/2/5ec06dfeedd5b7fa56fb84ba4b6fef39>

A.4 Integrated Digital Forensic Process Model

Computers & Security

Currently being reviewed

Integrated Digital Forensic Process Model

M.D. Kohn¹, (corresponding author)

M.M. Eloff², J.H.P. Eloff^{1,3},

¹Department of Computer Science, University of Pretoria,
Lynnwood Road, 0001, Pretoria, South Africa
Tel: +27 83 457 7112

²Institute for Corporate Citizenship, Unisa,
PO Box 392, Pretoria, 0003, South Africa
Tel: +27 12 433 4604

³SAP New Business and Technology (Mobile Empowerment) Africa, Suite 173, Private
Bag X 25, Lynnwood Ridge 0040, South Africa
Tel: +27 12 999 9100

¹mkohn@cs.up.ac.za

²eloffmm@unisa.ac.za

^{1,3}jan.eloff@sap.com

Abstract

Digital forensics is an established research and application field. Various process models exist describing the steps and processes to follow during digital forensic investigations. During such investigations, it is not only the digital evidence itself that needs to prevail in a court of law; the process followed and terminology used should also be rigorous and generally accepted within the digital forensic community. Different investigators have been refining their own investigative methods, resulting in a variety of digital forensic process models. This paper proposes a standardized Digital Forensic Process Model to aid investigators in following a uniform approach in digital forensic investigations.

Keywords: Digital Forensics, Computer Forensics, Digital Forensic Investigations, Process Models, Digital Forensic Process

1. Introduction

Digital forensics, also known as computer forensics, first presented itself in the 1970s (Pollitt, 2010). During the first investigation, financial fraud proved to be the root cause on the suspect computer. Over the past years digital forensics has become increasingly important in cases where electronic devices are used in the perpetration of a crime. Garfinkel (2010) provides a recent historic overview of digital forensic developments. The initial focus of digital forensic investigations was on crimes committed by using computers, but the field has expanded to include different devices where digitally stored information can be manipulated and used for various other criminal related activities.

Digital forensic investigations are common practice in law enforcement and commerce. Rapidly developing technology has resulted in various methods used

by investigators to establish the root cause of an incident. This has in turn resulted in a number of digital forensic investigation approaches being proposed, developed and refined. Garfinkel (2010) and Beebe (2009) submit a lack in digital forensic standardization and process, which is resulting in limited prosecution.

The aim of this paper is to investigate some of the most prominent process models used in digital forensic investigations and, after a comparative analysis of these process models, to propose an Integrated Digital Forensic Process Model or IDFPM. The IDFPM consists some of the prominent processes as extracted from the process models examined. The proposed IDFPM is a contribution towards a standardized DFPM regarding processes and terminology.

The remainder of the paper is structured as follows: the background section discusses some of the definitions of digital forensics in order to derive at working definitions for both digital forensics and digital forensic investigations. Section 3 discusses a number of existing process models within the current literature, while Section 4 introduces the Integrated Digital Forensic Process Model. The paper is concluded in Section 5.

2. Background

The main purpose of the background section is to define digital forensics, a digital forensic investigation, as well as the goal of such an investigation as used in this paper.

2.1 Digital Forensics

Digital forensics is often defined from the limited perspective of the person involved in an investigation (Ioeng, 2006). This section lists and discusses some of the definitions generally accepted within the literature. Common elements are extracted from the definitions to formulate an inclusive definition of digital forensics as proposed in this paper.

Computer forensics, digital forensics and media analysis, all of which are terms used to describe the relatively new field of digital forensics (Carrier, 2005), are used in the literature to describe this sub-branch of forensic sciences (Noblitt et al., 2000). The term *digital forensics* was chosen for the use in this paper.

Palmer (2001) defined *digital forensics* as “the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from digital sources for the purposes of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorized actions shown to be disruptive to planned operations”.

This definition is generally accepted to be an all-inclusive definition. The steps followed in this process are *preservation, collection, validation, identification,*

analysis, interpretation and presentation. Pollitt (2001) states that digital forensics is not a single process but a group of tasks, steps or sub-processes followed during an investigation. It is for this reason that a digital forensic process must be flexible to accommodate various technologies. A static process will limit future developments in the digital forensics field. Robbins (2012) does not prescribe the process as methodically as Palmer, but nevertheless includes a process in the definition. Formulating a fixed process list in a definition of digital forensics should however be avoided.

Reconstruction is listed as an element to aid in finding a root cause or simulating the events leading to an investigation. In digital forensics one investigator should reach the same conclusion as another, even when using different tools (von Solms et al., 2006). Unauthorized actions or actions shown to be disruptive to planned operations, must be known or identifiable, prior and during a digital forensic investigation. This knowledge will aid process development and certainty in producing the evidence.

Willassen and Mjøl̄snes (2005) redefined *digital forensics* as “the practice of scientifically derived and proven technical methods and tools towards the (process steps similar to Palmer) after-the-fact digital information derived from digital sources for the purpose of facilitating or furthering the reconstruction of the events as forensic evidence”.

The main difference between these two definitions is that Willassen and Mjøl̄snes (2005) removed the criminal element from the definition, which broadens the scope of application to include digital forensics in various types of investigations, such as commercial investigations.

Considering these two definitions, as well as other definitions in the literature, digital forensics is for the purposes of this paper, defined as a specific, predefined and accepted process applied to data stored digitally or digital media using scientific proven and derived methods, based on a solid legal foundation, to extract after-the-fact digital evidence with the goal of deriving the set of events or actions indicating a possible root cause, where reconstruction of possible events can be used to validate the scientifically derived conclusions.

2.2 Digital Forensic Investigation

A *digital forensic investigation* or DFI is the process to determine and relate extracted information and digital evidence to establish factual information for judicial review (Ioeng, 2006). Ioeng (2006) and Cohen (2010) emphasize the need to establish factual information as the outcome of such investigation.

Carrier (2005) proposes the use of the term DFI rather than digital forensics, and reasons that forensic science addresses a limited hypothesis. In a digital forensic investigation, DNA can be used to determine the relation between the suspect and the device at the physical crime scene (Casey, 2007). In this instance,

forensic science aids digital forensic science to determine a solution in the greater DFI, namely to prove whether the suspect was at the crime scene. Non-Digital Evidence can aid an investigator in a DFI to directly establish the presence of the suspect at the scene of the incident. A DFI is therefore the process of identifying potential evidence, digital or non-digital, *and* identifying the unique source of the potential evidence (Carrier, 2005). Cohen (2009) includes attribution as a process in his DFPM where the resulting digital evidence is linked to a specific person.

A *forensic investigation* of digital evidence is employed as a post-event response to an incident (Rowlingson, 2004). A *DFI* is therefore a special type of investigation where the scientific procedures and techniques used will allow the results, in other words digital evidence, to be admissible in a court of law. Since digital evidence is contained on some electronic media it cannot be viewed with the naked eye, some tool will be used to examine the state of this digital data. Tools used to observe the state of digital data constitute an indirect data observation. Tools used in DFIs include, but are not limited to, Encase (Guidance Software, 2011) and FTK (Access Data, 2011). The weight attributed to the evidential value is based on the extent the tool is trusted (Wojcik et al., 2006; Patel and Ciardhuáin, 2000). The confidence in DFIs is based on the level of trust in the hardware and software used to collect and analyze the data (Carrier and Spafford, 2003). Trust in the ability and experience of the investigator also contributes to the level of confidence in a digital forensic investigation when expert testimony is presented in court.

DFI, for the purpose of this paper, is therefore defined, as a special type of investigation where the scientific procedures and techniques used will allow the results – digital evidence – to be admissible in a court of law or a disciplinary forum in a corporate organization.

Any investigation conducted has a very specific purpose or outcome namely admissible digital forensic evidence that will aid judicial review. An investigation is generally initiated with the aim to establish some facts about an event that has taken place. The primary goal in establishing a possible root cause is to ensure that the investigation is conducted in a manner that will withstand legal scrutiny when the matter is serious enough to warrant it. However, any investigation should be conducted methodically to ensure that the conduct of the investigator is of such a nature that the validity of the evidence produced cannot be questioned. It should be noted that various types of DFIs exist. These include live forensics, proactive forensics and network forensics (Beebe and Clark, 2004; Rogers et al., 2006; Mylonas et al., 2012). Beebe and Clark (2004) suggest a second-tier phase to the DFI, which specifically anticipates steps followed in specific incident investigations. Types of incidents include drug activity, financial crimes and child pornography.

Rogers et al. (2006) propose a digital evidence triage to aid the investigator. The evidence triage consists of the user usage profile, Internet usage and

chronological timeline activity. The specific user evidence is found in home directories, the registry and file properties. Depending on the type of investigation conducted the evidence triage will guide an investigator to possible evidence, if other traces have possibly been removed.

The fundamental point of departure for any investigation is to answer basic questions about the evidence. In addition to knowing what happened, there is a need to know who is responsible (Kruse and Heiser, 2002). Zachman (2003) developed a framework adapted by Ioeng (2006) and Beebe and Clark (2004) proposing that every investigator should ask the following six key questions during an investigation (Ioeng, 2006; Beebe and Clark, 2004): what, why, how, who, where and when. *What* is determined by the data attributes or metadata, *why* refers to the motivation, *how* is the procedure followed to initiate the incident or isolate the necessary evidence, *who* are the people involved, *where* refers to the location and *when* refers to time.

The next section deals extensively with the question of *how* evidence was found in an investigation. The *how* question is addressed by the steps of a process, but the steps also have to be defined. These definitions are an attempt to formulate standardized terminology, which should be used in the digital forensics community. The precision of steps followed in the investigation and the process adhered to determine the accuracy of incident reconstruction. Various authors have described these steps in a Digital Forensic Process Model or DFPM.

3 Selected Existing Digital Forensic Process Models (DFPMs)

The primary objective of a DFPM is to aid the investigator to explain *how* specific digital evidence is found on a device. A number of DFPMs exist in the current literature. These process models have not yet been standardized formally. The initial process descriptions include broad phased descriptions such as preparing for an electronic investigation and validating electronic evidence (Wolfe, 2003). Phase descriptions include reconstruction and hypothesis testing (Carrier and Spafford, 2004b). Detailed process descriptions are found in DFPMs proposed by Ciardhuáin (2004), Carrier and Spafford (2003). The listed authors use different representations to illustrate the DFPMs, including descriptions, process models and state diagrams.

To uniformly represent the DFPMs as discussed in this paper, a number of representations were considered. These include, but are not limited to, UML Activity, Use Case Diagrams (Köhn et al, 2008) and Finite State Machines (Carrier and Spafford, 2006; Cohen, 2009). For the purposes of this paper the ordering of the events or processes are considered critical.

Sequential logic is a simple mathematic notation used in the remainder of the paper as formulated by Moore and Mealy (Wikipedia, 2011). Sequential logic is proposed to represent the DFPMs because the circuit outcome is dependent on

the input and the current internal state. For the circuit to evaluate true, all the conditions of the previous states must be true. The circuit will fail if the current state is not positively completed. This will allow an investigator to revisit previous steps in the process, but he/she will not be able to continue if a step is not complete or fails.

The sequential logic notation is however adapted to uniformly represent each of the DFPMs, where the list values have been directly replaced with the process steps. The adapted sequential notation is illustrated here as:

$$\text{DFPM} = \{start \Rightarrow next \Rightarrow then \dots end\}$$

In certain instances where sub-processes are indicated, these will be illustrated when introducing the DFPMs. Parallel processes are indicated by ||. \Leftrightarrow is used where a previous process can be repeated after executing the current process. Each DFPM is shown using the adapted sequential logic notation. This is done to identify similarities and differences within the sequence of tasks when conducting a digital forensic investigation. Terminology used in the DFPMs is listed and briefly defined where new terms are introduced. Some brief comments conclude the discussion on each DFPM.

The paper is limited to a discussion of the following DFPMs: Lee et al., (2001), Casey (2004), Carrier and Spafford (2004a), Baryamureeba and Tushabe (2004), Ciardhuáin (2004) and Cohen (2009). Numerous other DFPMs were examined in the initial research but the discussion in this paper has been limited to the listed DFPMs as they are the DFPMs that have the most detailed sub-processes listing and are contained in those not listed here. All the process descriptions included in the DFPMs are discussed.

3.1 Lee

Henry Lee formulated a Scientific Crime Scene Investigation model (Lee et al., 2001) to accommodate investigations that use forensic science. Lee's model does not find specific applications within the digital forensics field, but is a good departure point nonetheless. This DFPM is represented clearly and precisely by using the sequential logic notation proposed previously as:

$$\text{Lee} = \{Recognize \Rightarrow Identify \Rightarrow Individualize \Rightarrow Reconstruct\}$$

where

$$\text{Recognize} = \{Document \Rightarrow Collect and Preserve\}$$

$$\text{Identify} = \{Classify \Rightarrow Compare\}$$

$$\text{Individualize} = \{Evaluate \Rightarrow Interpret\}$$

Reconstruct = {*Reconstruct* \Rightarrow *Report and Present*}

The terminology used in the Lee DFPM is described as follows:

Recognition is where items or patterns are seen to be potential evidence. The investigator must know what to look for and where to find it. This sub-process has two activities, namely *documentation* and *collection and preservation*. Documenting evidence is an important aspect of an investigation, where any action by any person is clearly documented. Beebe and Clark (2004) suggest that every phase of an investigation must be thoroughly documented throughout the entire investigation process. Collection is where the evidence is collected from the crime scene, before being *bagged and tagged*. Digital evidence must obviously be preserved once safely contained. Documentation should be done while collecting and preserving the evidence. Collection and preservation is seen as a single process step in this DFPM.

Identification of the various types of evidence is done next. Evidence is classified and compared, usually into categories such as physical, biological, chemical and other standard types (Ciardhuáin, 2004).

Individualization is where evidence is linked to a particular individual or event. The evidence is then evaluated and interpreted. *Reconstruction* is where evidence objects and events are linked so as to account for a possible sequence of events. During reconstruction, possible event sequences are reported and presented (Lee et al., 2001). Similarly report and present is considered to be a single process step in this DFPM.

Ciardhuáin (2004) criticizes Lee's model because it deals specifically with the crime scene investigation and not with the entire investigative process. However, Ciardhuáin (2004) does not include a detailed approach that may be applicable to electronic crime scenes, but advocates that the investigation must be systematic and methodical. The model focuses primarily on physical evidence, but can be adapted to include digital evidence during a digital crime scene investigation. The physical evidence is usually processed for trace evidences such as blood and DNA forensic analysis, where trace evidence on digital media is the data itself.

3.2 Casey

In 2000 Casey proposed a DFPM for processing and examining digital evidence. This DFPM can be applied to various investigations, including standalone computer systems and networked environments (Ciardhuáin, 2004). In 2004 Casey revised his DFPM to include a number of additional processes (Casey, 2004). The 2000 Casey DFPM is listed as:

Casey 2000 = {*Recognition* \Rightarrow *Preservation* \Rightarrow *Classification* \Rightarrow *Reconstruction*}

where

Preservation = {*Collect* ⇒ *Document*}

Classification = {*Compare* ⇒ *Individualize*}

The DFPM by Casey (2000) in 2000 is similar to the one proposed by Lee. Casey's model is aimed at processing and examining digital evidence (Baryamureeba and Tushabe, 2004), however his analysis will differ substantially from the physical analysis of non-digital evidence types. The first and last processes, namely recognition and reporting, are identical in both these DFPMs. The 2004 Casey model is significantly extended to include various processes identified in the development of various other DFPMs reported on in the literature. The all-inclusive updated 2004 DFPM is given as follows:

Casey 2004 = {*Incident Recognition* ⇒ *Assessment* ⇒ *Identification and Seizure* ⇒ *Preservation* ⇒ *Recovery* ⇒ *Harvesting* ⇒ *Reduction* ⇒ *Classification* ⇒ *Analysis* ⇒ *Reporting*}

where

Preservation = {*Collect* ⇒ *Document*}

Classification = {*Organize* ⇒ *Compare* ⇒ *Individualize*}

The terminology used by Casey is described as follows:

Recognition will be where the investigator recognizes a similar pattern that might have presented itself in the past. This is a form of investigator experience based on previous investigations and could include a database of previously solved investigations.

Preservation consists of two sub-processes, *collect* and *document* (Casey, 2004). During preservation the digital evidence should firstly ensure quality and continued availability; and secondly maintain the integrity of the evidence during the entire investigation process.

During *classification*, evidence objects are *compared* and *individualized*. *Individualization* is where evidence is attributed to an origin or creator (Cohen, 2009).

During *reconstruction*, the sequence of the crime is traced by reconstructing the possible event sequence that most accurately reflects the events as they could have happened during the actual crime or incident. Reconstruction is generally required to prove *how* a certain result is achieved for various purposes.

3.3 Carrier and Spafford

Carrier and Spafford propose a DFPM with five sub-processes and seventeen activities in total. The DFPM is named the Integrated Digital Investigation Process (IDIP). The sub-processes are listed as readiness, deployment, physical investigation, digital investigation and review (Carrier and Spafford, 2004a). This model is in contrast to the linear development approaches mentioned in previous sections.

The Carrier and Spafford DFPM is given as:

Carrier and Spafford = {*Readiness* ⇒ *Deployment* ⇒ *Physical Investigation* // *Digital Investigation* ⇒ *Review*}

where the digital and physical investigations occur simultaneously, and

Readiness = {*Operational Readiness* ⇒ *Infrastructure Readiness*}

Deployment = {*Detection and Notification* ⇒ *Confirmation and Authorization*}

Physical Investigation = {*Preservation* ⇒ *Survey* ⇒ *Documentation* ⇒ *Search and Collection* ⇒ *Reconstruction* ⇒ *Presentation*}

Digital Investigation = {*Preservation* ⇒ *Survey* ⇒ *Documentation* ⇒ *Search and Collection* ⇒ *Reconstruction* ⇒ *Presentation*}

The terminology used in this DFPM is similar to the definitions given for the previous models. During the *review* phase the whole investigation is reviewed and areas of improvement are identified. It is interesting to note that this DFPM includes a physical and digital investigation to be conducted concurrently. For the full discussion see Carrier and Spafford (2004a).

The DFPM includes sub-processes during investigation to accommodate issues such as data protection, acquisition, imaging, extraction, interrogation, ingestion and normalization, analysis and reporting (Baryamureeba and Tushabe, 2004). High-level processes are included for both the physical and logical or digital crime scenes. Baryamureeba and Tushabe, (2004) questions the practicality of the model. The following illustrates this problem. The primary crime scene is where the crime is initiated. The target of location of the victim, which is the secondary crime scene, is not included as part of the investigation. Carrier and Spafford's inclusion of the physical crime scene is however a notable contribution.

Making a clear differentiation between a physical and digital crime scene seems

trivial, but this distinction is critical for the practical execution of an investigation.

3.4 Baryamureeba

The Enhanced Integrated Digital Investigation Process (EIDIP) DFPM also makes a clear distinction between the physical and digital crime scene investigation processes (Baryamureeba and Tushabe, 2004). This DFPM is an extension of the DFPM proposed by Carrier and Spafford. The Baryamureeba DFPM is given as:

Baryamureeba = {*Readiness* \Leftrightarrow *Deployment* \Leftrightarrow *Traceback* \Leftrightarrow *Dynamite* \Leftrightarrow *Review*}

where

Readiness = {*Operational Readiness* \Rightarrow *Infrastructure Readiness*}

Deployment = {*Detection and Notification* \Rightarrow *Physical Crime Scene Investigation* \Rightarrow *Digital Crime Scene Investigation* \Rightarrow *Confirmation* \Rightarrow *Submission*}

Traceback = {*Digital Crime Scene Investigation* \Rightarrow *Authorization*}

Dynamite = {*Physical Crime Scene Investigation* \Rightarrow *Digital Crime Scene Investigation* \Rightarrow *Reconstruction* \Rightarrow *Communication*}

The terminology used by Baryamureeba and Tushabe (2004) is defined as follows:

Readiness includes the training of personnel and the provision of sufficient resources and infrastructure to deal with the investigation.

Deployment provides mechanisms to detect and confirm incidents, and it consists of five sub-processes. The first is to detect the incident and notify the appropriate authority. Secondly, the physical crime scene is examined to identify potential evidence. Thirdly, the potential digital evidence is subjected to a digital examination of potential evidence. Fourthly, confirmation of the incident is given to obtain legal approval for a search warrant. Lastly the evidence is presented to the appropriate forum.

In the *Traceback* phase, the physical crime scene is tracked down to identify devices used in the execution of the crime. Firstly the primary crime scene is reconstructed from evidence collected during deployment. This typically includes finding the host computer within a networked environment. Secondly, authorization is obtained to permit further investigation of the acquired evidence.

The *Dynamite* phase investigates the primary crime scene. It is aimed at

collecting and analyzing evidence items found at the primary scene so as to find the incident perpetrators. The phase involves four sub-processes. The physical evidence found at the crime scene is examined. Secondly, the digital crime scene is examined. Thirdly, possible events are reconstructed to formulate a possible hypothesis. Fourthly, the final interpretations are communicated in a presentation to the appropriate forum.

Review is performed last, where the investigation is reviewed and areas of improvement are identified.

The Baryamureeba DFPM builds on the work of Carrier and Spafford. Carrier and Spafford propose a waterfall type model in their original paper, which does allow splash back to previous phases. Baryamureeba adapts their process flow enabling the investigator to backtrack to previous phases, which is indicated here with a bi-directional arrow between phases. The listed sub-processes given for the physical and digital investigation processes differ and do not occur simultaneously. The sub-process listing is readiness, deployment, traceback, dynamite and review. Each of the listed sub-processes includes a number of activities to be completed during the investigation.

The main objective of the Baryamureeba DFPM is to separate the physical investigation from the digital investigation (Baryamureeba and Tushabe, 2004). However, this results in a complicated adaption of the Carrier and Spafford DFPM. In Baryamureeba's DFPM a new phase is introduced where the primary crime scene is identified in the traceback phase. The primary crime scene is the place where the incident originated. Reconstruction is done only once in this DFPM, when all the necessary evidence has been collected.

The digital crime scene is processed in a virtual environment created by hardware and software (Baryamureeba and Tushabe, 2004). The phases listed are preservation, survey, search and collection, and documentation. The preservation phase includes the duplication of digital media. During the survey the investigator identifies and separates potential useful data from the imaged set. Hidden, deleted, manipulated or damaged data files are recovered during the search and collect phase. Documentation involves the extensive documenting of all the evidence found, which in turn is useful in the presentation phase.

3.5 Ciardhuáin

The DFPM proposed by Ciardhuáin (2004) is probably the most all-inclusive and comprehensive to date. The steps or phases are also called activities. The following activities are listed in this DFPM: awareness; authorization; planning; notification; search for and identification of evidence; collection; transportation; storage; examination; hypothesis; presentation; proof/ defense, and dissemination. The steps are discussed in depth by Ciardhuáin (2004). Only the terms that have not previously been listed are introduced here. Ciardhuáin's DFPM is a linear representation and is represented as follows:

Ciardhuáin = {*Become aware* ⇒ *Authorize* ⇒ *Plan* ⇒ *Notify* ⇒ *Search/Identify*
⇒ *Collect* ⇒ *Transport* ⇒ *Store* ⇒ *Examine* ⇒ *Hypothesize* ⇒ *Present*
⇒ *Prove/Defend* ⇒ *Disseminate*}

Processes follow the waterfall model, in other words processes follow one another in sequence. Certain sequences can be repeated if needed. The sequence of examine, hypothesis, presentation and prove/ defend will often be repeated as the evidence pool grows during the investigation (Ciardhuáin, 2004).

Awareness is defined as the phase during which the investigators are made aware that a crime has taken place, i.e. the crime is reported to some authority. An intrusion detection system can also trigger such awareness. Ciardhuáin (2004) specifically includes this in the DFPM because the method of becoming aware could influence the investigation. The investigation will have to be conducted regardless if the investigator has prior knowledge of the type of incident, or not. The co-operation of various parties can be expected, e.g. in cases such as internal investigations where the parties would like to find the root cause of the incident. Awareness can be internal or external to an organization.

Authorization is where the type of investigation has been identified and now the appropriate authorization may be required to proceed. Authorization is also acquired internally or externally.

Planning is influenced by information within and outside the organization that will impact on the investigation. Outside factors include legal and other requirements that are not determined by the investigators, while internal factors include policies of the organization, prior investigative knowledge and procedures. The scope can also be backtracked if the full requirements of the investigation are not included in the planned scope. Externally imposed policies, regulations and legislation, external information, information distribution and organizational policies can influence the planning phase.

During *notification* the stakeholders or investigated subject is informed that an investigation is taking place. In cases where the subject investigated must not know that an investigation is taking place, this step is omitted. Other interested parties can also be informed that there is an investigation in progress during this step.

The *search and identification* of evidence is where the location of the potential evidence is identified. In large investigations this may include finding routes of information flows over ISPs. Authorization will probably have to be revisited in cases of multiple jurisdictions.

Collection occurs when the investigator takes physical possession of the evidence to be preserved and analyzed. Ciardhuáin (2004) includes hard disk imaging and seizing of entire computers in this step. The primary focus of the current literature on digital forensics is on the collection of digital evidence.

Mistakes and incorrect procedures during this process will render evidence in later stages useless, and therefore inadmissible in court. Where questionable procedure is followed or cannot appropriately be explained during a court hearing, the digital evidence could be ruled inadmissible. Many legal practitioners will focus on the collection procedure followed to find a questionable procedure in an attempt to invalidate the incriminating evidence.

After collection, the evidence is *transported* to a suitable location for forensic examination. It is important that the integrity of the evidence is not affected physically or digitally during transfer. Digital evidence is stored in a safe location before examination. The integrity of the evidence must also be ensured at the storage location.

Examination is the core process of the digital investigation. A large number of techniques have to be used to access, find and extract evidence from the collected media. When large volumes of data constitute the subject of an investigation, automated techniques may be required to aid the investigator. Ciardhuáin (2004) specifically mentions that during examination some automated techniques are required to aid the investigator.

The *hypothesis* formulated by the investigator is based on his/her examination of the examined digital evidence. The hypothesis is the investigator's proposed construction of events or a possible sequence of events leading to the reported violation. The document compiled during the investigation must reflect the findings of the digital forensic examiner. Backtracking during examination is expected as the examiner gains insight into the investigation. The formulated hypothesis may present the investigator with internal and external challenges. An external challenge could for example be the legal relevance of evidence found during an investigation. An internal challenge could be that there is no digital evidence to support the formulated hypothesis.

Presentation is where the hypothesis is presented to people other than the investigators, such as a jury or management. A decision will then be made on the basis of the presented findings.

The *proof / defense* is where the digital forensic examiner questions or substantiates his/her original investigation hypothesis. The investigator will have to defend the findings, or prove that the events occurred as explained in the presentation.

Dissemination of the lessons learnt is the final activity, if required. Policies and procedures influencing future investigations have to be integrated with current policies and procedures.

According to Ciardhuáin (2004), the reason for proposing this model is the fact that, while other DFPMs focus on processing digital evidence, this model incorporates the whole investigation process. A lack of standardized terminology

also seems to be an identified problem that needs a solution. The process only gives guidance on *what* must be done and not *how*. This DFPM does not include specifics such as tools and technology to be used, or the training needed before an investigator is qualified to do an investigation. Best practices, common experiences and the development of standards are identified as important future research topics.

Ciardhuáin (2004), Carrier and Spafford (2004a) also remark that information flows are not addressed in any of the previous DFPMs discussed. The main problem in this regard is where and how the chain of custody is compiled. Different legal systems, best practices and languages are some difficulties that investigators could encounter.

Awareness, transport, storage and dissemination are considered irrelevant according to a survey conducted by Ciardhuáin (2004). The remainder of the proposed DFPM activities was considered relevant. In contrast to the view expressed by Ciardhuáin that awareness is irrelevant, Perumal (2009) states that awareness should be extended to be a three-step process. The sub-processes proposed by Perumal include the complaint, investigation and prosecution. The Ciardhuáin DFPM only includes the complaint step.

Ciardhuáin's (2004) later work includes policy development on criminal investigations, auditors, civil litigation, system administrator investigations and judicial inquiries.

3.6 Cohen

The DFPM proposed by Cohen (2009) consists of seven listed processes or phases. The focus of this DFPM is the digital forensic examination. The Cohen DFPM is given as:

Cohen = {*Identification* ⇒ *Collection* ⇒ *Transportation* ⇒ *Storage* ⇒ *Examination and Traces* ⇒ *Presentation* ⇒ *Destruction*}

where

Examination = {*Analysis* ⇒ *Interpretation* ⇒ *Attribution* ⇒ *Reconstruction*}

The terminology as used by Cohen is described as follows:

Analysis is where evidence is understood and characterized relative to the legal issue at hand. Beebe and Clark (2004) propose an iterative sub-process listing as survey, extract and examine during analysis. The sub-processes analysis includes the physical media, media management, file system, application and network hierarchy.

Interpretation takes the results of analysis to produce meaningful statements.

The statements give meaning to the legal and technical situation. *Attribution* involves drawing conclusions about causes and effects. The links that exist are identified and documented. A particular cause will give rise to an effect; conversely, a particular effect may or may not be caused by a certain action or incident.

Reconstruction is the process by which a set of mechanisms that are similar to those identified has caused the effect of the digital evidence produced. Reconstruction is therefore a process where the investigator lists certain assumptions and limitations to most accurately present how evidence came to exist.

The focus of the Cohen DFPM is on the examination of digital evidence. It is interesting to compare the examination sub-process listing given by Casey with that of Cohen. This clearly indicates the need for some standardization of terminology. The issue is constantly mentioned by various authors but never sufficiently addressed.

A comparison of the set of activities included under *examination* by Casey (2004) and Cohen (2009) respectively reveals the following two sets:

Casey *Examination* = {*Recovery, Harvesting, Reduction, Classification*}

and

Cohen *Examination* = {*Analysis, Interpretation, Attribution, Reconstruction*}.

Clearly not a single sub-process within the two identified sets bears the same meaning. A possible explanation for this discrepancy is that the interpretations of the term *examine* and *analyse* has been swapped by the authors.

4. Construction of an Integrated DFPM (IDFPM)

This section introduces a new DFPM by integrating the six DFPMs discussed in the previous paragraphs. The three factors that we have considered include the terminology used, the process ordering and parallel processes.

The terminologies used to describe the processes in the DFPMs often differ, but there are similarities. The DFPM process descriptions are studied to find similar meaning in the terminology so as to effectively reduce the number of required processes. Eliminating processes from the DFPMs that have similar objectives also reduces duplicate processes.

Carrier and Spafford list *digital investigation* and *physical investigation* to occur simultaneously. In the Baryamureeba DFPM, *detection* and *notification* also occur simultaneously.

The Integrated Digital Forensic Process Model or IDFPM consists of the following

processes: *preparation, incident, incident response, physical investigation, digital forensic investigation and presentation.*

The IDFPM process listing is given as:

$$\text{DFPM} = \{\{ \text{Preparation} \Rightarrow \text{Incident} \Rightarrow \text{Incident Response} \Rightarrow \text{Physical Investigation} \parallel \text{Digital Forensic Investigation} \Rightarrow \text{Presentation} \} \parallel \text{Documentation} \}$$

where

$$\text{Preparation} = \{ \text{Policy/Procedure} \Rightarrow \text{Operational Readiness} \parallel \text{Infrastructure Readiness} \}$$

$$\text{Incident} = \{ \text{Detect} \Rightarrow \text{Assess} \parallel \text{Confirm} \Rightarrow \text{Notify} \Rightarrow \text{Authorize} \Rightarrow \text{Deploy} \}$$

$$\text{Incident Response} = \{ \text{Approach Strategy} \Rightarrow \text{Search} \Rightarrow \{ \text{Recover} \parallel \{ \text{Seize} \Rightarrow \text{Preserve} \} \parallel \text{Preserve} \} \Rightarrow \{ \text{Transport} \Rightarrow \text{Store} \Rightarrow \text{Collect} \} \}$$

$$\text{DFI} = \{ \text{Collect} \Rightarrow \text{Authenticate} \Rightarrow \text{Examine} \Rightarrow \text{Harvest} \Rightarrow \text{Reduce} \Rightarrow \text{Identify} \Rightarrow \text{Classify} \Rightarrow \text{Organize} \Rightarrow \text{Compare} \Rightarrow \text{Hypothesize} \Rightarrow \text{Analyze} \Rightarrow \text{Attribute} \Rightarrow \text{Evaluate} \Rightarrow \text{Interpret} \Rightarrow \text{Reconstruct} \Rightarrow \text{Communicate} \Rightarrow \text{Review} \} \wedge \{ \text{Reconstruct} \Rightarrow \text{Hypothesize} \}$$

$$\text{Presentation} = \{ \text{Report/Present} \Rightarrow \text{Decide} \Rightarrow \text{Dissemination} \}$$

The Documentation process is included in the IDFPM as a continuous process and includes the investigation documents and chain of custody recorded as accurately as possible throughout the entire investigation. The diagrammatic representation on the IDFPM is illustrated in Figure 1. When developing the policies and procedures in an organization it is essential to ensure that legal advice is sought to ensure any documentation will be able to withstand legal scrutiny.

Infrastructure and operational readiness are processes that occur in parallel. These processes will overlap extensively when setting up a digital forensic organization. The documentation should continuously be updated after each investigation to ensure that it is in line with decisions reached and new developments in technology.

When an incident is detected, the situation should be assessed and confirmed before notification of the incident is sent to authorize an investigation. At the incident scene it is not always easy to determine what one may find and it is therefore important to have an approach strategy in place before searching the premises. During incident response, one may not always find physical evidence to seize; this is why recovery and preservation may occur in parallel before the digital evidence is transported and stored. The incident response phase is considered to be where the incident is located. The digital forensic investigation

to follow is largely lab based.

The digital forensic investigation *collect* sub-process may occur directly after potential evidence has been preserved. This will be common in network or live forensic investigations. The sub-process listing from *hypothesize* up to *review* should be repeated during the digital forensic investigation process to continually test the hypothesis formulated.

The decision reached during *presentation* should be recorded in preparation to aid investigators in future investigations when faced with similar incidents.

Finally, the IDFPM should not be seen as a static process model. The IDFPM can and must develop and integrate current methods, tools and technologies as they develop. The terminology used must also be expanded to accommodate any of the developments.

5. Conclusion

The paper briefly discussed a number of important definitions that are integral to a digital forensic investigation. Definitions for digital forensic and digital forensic investigations were proposed. Various Digital Forensic Process Models or DFPMs are identified in the current literature. The DFPMs identified all have differing approaches. A selected number of DFPMs were introduced and discussed by listing an adapted process description using sequential logic notation, with the terminology used in each model explained. The DFPMs were compared with each other, and the essential processes required in an integrated digital forensic process model, were identified and abstracted.

In one of the previous sections various problems were identified in the existing DFPMs, such as differing terms that actually refer to the same processes or steps, or the conflicting terminology reflecting different interpretations of a process step. Therefore, the IDFPM is not just a merging of existing DFPMs, but an integration of the discussed DFPMs and a purification of the terminology used, resulting in an all-encompassing standardised IDFPM.

The main contribution of the paper was the introduction of this Integrated DFPM or IDFPM, which is a process model consisting of the selected DFPMs discussed. The IDFPM is a standardized DFPM with proposed standardized terminology.

References

Access Data. FTK. 2011. <http://accessdata.com/products/computer-forensics/ftk> (accessed September 2011).

Baryamureeba V, Tushabe F. 2004. The Enhanced Digital Forensic Investigation Process Model. Digital Forensics Research Workshop (DFRWS). Baltimore:

Citeseer, 2004.

Beebe NL, Clark JG. 2004. A hierarchical, objectives-based framework for the digital investigations process. Digital Forensics Research Workshop (DFRWS). Baltimore: St. Paul: West Publishing Co., 2004. 147-167.

Beebe NL. 2009. Digital forensics research: the good, the bad, and the unaddressed. In: Fifth annual IFIP WG 11.9 international conference on digital forensics; January 2009.

Carrier BD, Spafford EH. 2003. Getting Physical with the Digital Forensic Process. International Journal of Digital Evidence volume 2, issue 2 (2003).

Carrier BD, Spafford EH. 2004a. An event-based digital forensic investigation framework. Digital Forensics Research Workshop (DFRWS). 2004.

Carrier BD, Spafford EH. 2004b. Defining Event Reconstruction of Digital Crime Scenes. Journal of Forensic Sciences (Wiley Online Library), 2004.

Carrier BD, Spafford EH. 2006. Categories of Digital Forensics Investigation Analysis Techniques based on the Computer History Model. Digital Investigation (Elsevier), 2006.

Carrier BD. 2005. File System Forensic Analysis. Addison Wesley, 2005.

Casey E. 2004. Digital Evidence and Computer Crime. Elsevier Academic Press, 2004.

Casey E. 2007. Handbook of Computer Crime Investigation: Forensic Tools and Technology. 1st ed. Elsevier Academic Press, 2007.

Ciardhuáin SO. 2004. An Extended Model of Cybercrime Investigations. International Journal of Digital Evidence 3 (2004).

Cohen, F. 2009. Digital Forensic Evidence Examination. 2nd ed. Fred Cohen & Associates, 2009.

Cohen, F. 2010. Towards a Science of Digital Forensic Evidence Examination. Edited by C Kam-Pui and S Shenoi. IFIP Advances in Information and Communication Technology. Boston: Springer, 2010. 17-35.

Garfinkel SL. 2010. Digital forensics research: The next 10 years, The Proceedings of the Tenth Annual DFRWS Conference, Digital Investigation, Volume 7, Supplement, August 2010, Pages S64–S73, <http://dx.doi.org/10.1016/j.diin.2010.05.009>,

Guidance Software. 2011. Forensic. 2011. <http://www.guidancesoftware.com/forensic.htm> (accessed September 2011).

- Ioeng RSC. 2006. Forza - Digital Forensic Investigation Framework Incorporate Legal Issues. Digital Investigation, September 2006: 29-36.
- Köhn MD, Eloff JHP, Olivier MS. 2008 UML modelling of Digital Forensic Process Models. Information Security South Africa (ISSA). 2008.
- Kruse WH, Heiser J. 2002 Computer Forensics: Incident Response Essentials. 1st ed. Addison Wesley, 2002.
- Lee HC, Palmer TM, Miller MT. 2001 Henry Lee's Crime Scene Handbook. 1st ed. Academic Press, 2001.
- Mylonas A, Meletiadiis V, Tsoumas B, Mitrou L, Gritzalis D. 2012. Smartphone forensics: A proactive investigation scheme for evidence acquisition. Gritzalis D, et al., editors, in Proc. of the 27th IFIP International Information Security and Privacy Conference. Springer; AICT-376; 2012. p. 249-260.
- Noblitt MG, Pollitt MM. 2000. Recovering and Examining Computer Forensic Evidence. Forensic Science Communications Volume 2, No. 4 (October 2000).
- Palmer GA. 2001. Roadmap for Digital Forensic Research. Digital Forensics Research Workshop (DFRWS), 2001.
- Patel A, Ciardhuáin SO. 2000. The Impact of Forensic Computing on Telecommunications. IEEE Communications Magazine, November 2000: 64-67.
- Perumal S. 2009. Digital Forensic Model based on Malaysian Investigation Process. International Journal of Computer Science and Network Security Volume 9, No. 8 (August 2009): 38-44.
- Pollitt MM. 2001. Report on Digital Forensics. 13th INTERPOL Forensic Science Symposium. Computer Analysis Response Team, 2001.
- Pollitt MM. 2010. A History of Digital Forensics. IFIP International Conference Digital Forensics. 2010. 3-15.
- Robbins J. 2012. An Explanation of Computer Forensics. Retrieved February 10, 2012, from <http://www.pivx.com/forensics>
- Rogers MK, Goldman J, Mislán R, Wedge T, Debrotá S. 2006. Computer Forensics Field Triage Process Model. Conference on Digital Forensics, Security and Law. 2006.
- Rowlingson RA 2004 Ten Step Process for Forensic Readiness. International Journal of Digital Evidence volume 2, no. 3 (2004).
- von Solms SH, Louwrens C, Reekie C, Grobler TA. 2006. Control Framework for Digital Forensics. Edited by S Sujeet and MS Olivier. IFIP Advances in Digital

Forensics and Communication Technology. Boston: Springer, 2006. 343-355.

Wikipedia. 2012. Sequential Logic. [http://en.wikipedia.org/wiki/Sequential logic](http://en.wikipedia.org/wiki/Sequential_logic) (accessed October 2012).

Willassen SY, Mjølunes, SF. 2005. Digital Forensics Research, Telektronikk Volume 101 No. 1 – 2005 pp. 92-97. Retrieved from <http://www.telenor.com/telektronikk/> (accessed on 10 October 2008)

Wojcik M, Venter HS, Eloff JHP, Olivier MS. 2006. Applying machine trust models to Forensic Investigations. Edited by Olivier M. and Sheno, S. Advances in Digital Forensics II (Springer) volume 222 (2006): 55-65. Wolfe HB. Computer Forensics. Computers & Security (Elsevier) volume 22, no. 1 (2003): 26-28.

Wolfe, HB. 2003. Computer Forensics. Computers & Security (Elsevier) 22, no. 1 (2003): 26-28.

Zachman JA. 2003. The Zachman Framework for Enterprise Architecture: Primer for Enterprise Engineering and Manufacturing, Zachman International 2003

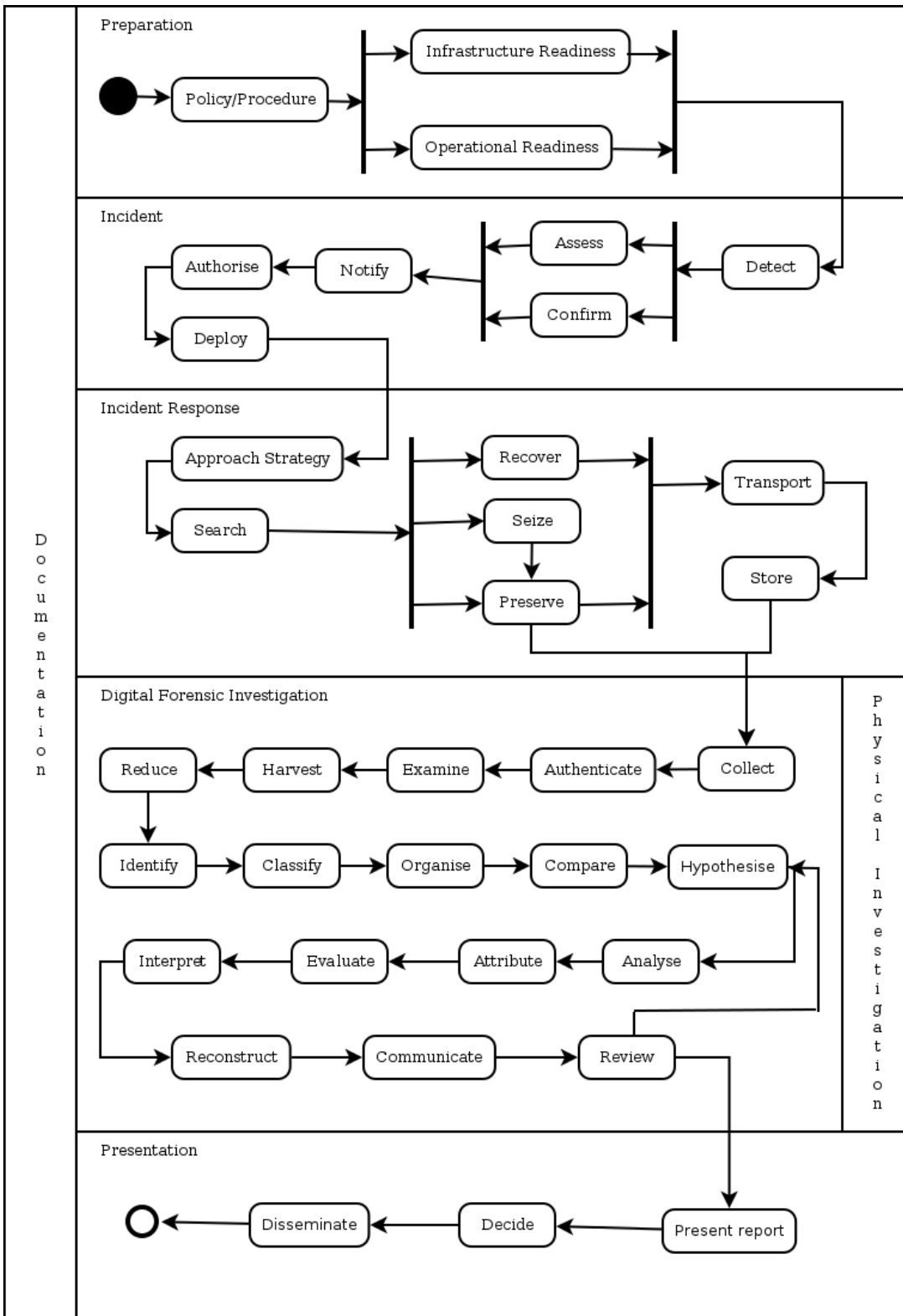


Fig. 1. The IDFPM illustrated as a process flow diagram

Author Biographies

Michael Kohn is currently a candidate advocate in the High Court in South Africa. His main contribution is towards developing the elective presentation of Digital Evidence by way of rigorous internationally accepted Digital Forensics Procedures. He has given testimony in numerous courts.

Michael is interested in Trust, Network Forensics, Privacy Encryption, Cyber Forensics, Cyber Crime, Cyber Law and procedural and legal aspects of Digital Evidence.

He has a number of conference publications and is involved with the Standardisation of the Digital Forensics Process at the South African Bureau of Standards (SABS).

Mariki M Eloff received a PhD computer science degree in 2000 from the then Rand Afrikaans University, South Africa, now known as the University of Johannesburg.

She is a full professor and chief researcher at the Institute of Corporate Citizenship at Unisa.

Prof Eloff is deputy chair of the Unisa Employment Disability Forum, a council member of The Independent Living Centre for Disabled Persons in South Africa as well as a member of the National Council for Persons with Physical Disabilities in South Africa. In 2010 she received the Unisa Women in Research award for Research Leadership.

Jan Eloff holds a PhD in computer science from the University of Johannesburg, South Africa, previously known as the Rand Afrikaans University.

He is currently the Research Director of SAP Research Pretoria specialising in Mobile Empowerment. He is also appointed as an Extraordinary Professor in Computer Science at the University of Pretoria.