UML MODELLING OF DIGITAL FORENSIC PROCESS MODELS (DFPMs)

Michael Köhn[1], J.H.P. Eloff[2], MS Olivier[3]

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

[1]mkohn@cs.up.ac.za, [2]eloff@cs.up.ac.za, [3]molivier@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 Encase [1] 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. Knoppix [3] 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.

[1] Encase online: http://www.guidancesoftware.com/
[2] Access Data online: http://www.accessdata.com/
[3] Knoppix 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

[4]The 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	Transpoprtation	Storage	Hypothosis	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.

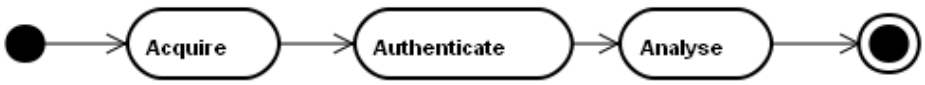

Figure 2: Kruse Activity diagram

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 Defense. 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

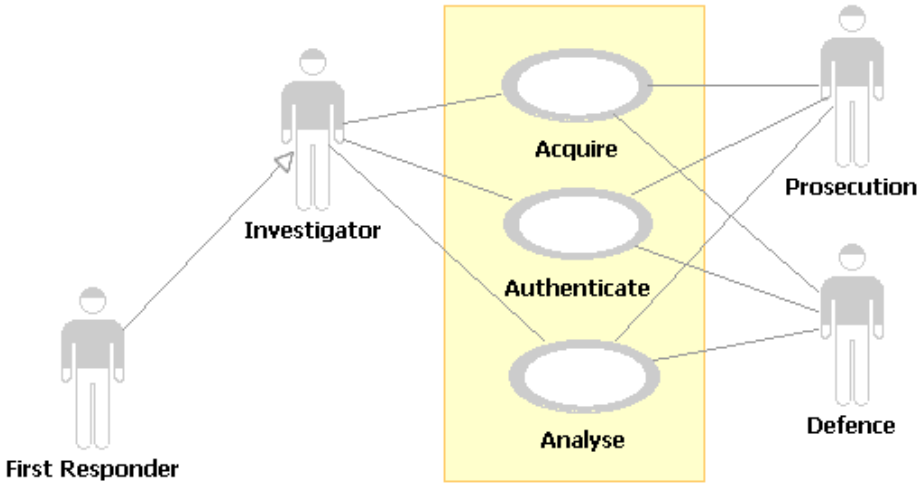

Figure 3: Kruse Use Case Diagram

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.

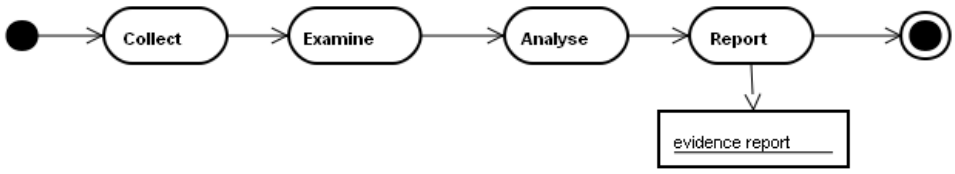

Figure 4: USDOJ Activity diagram

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.

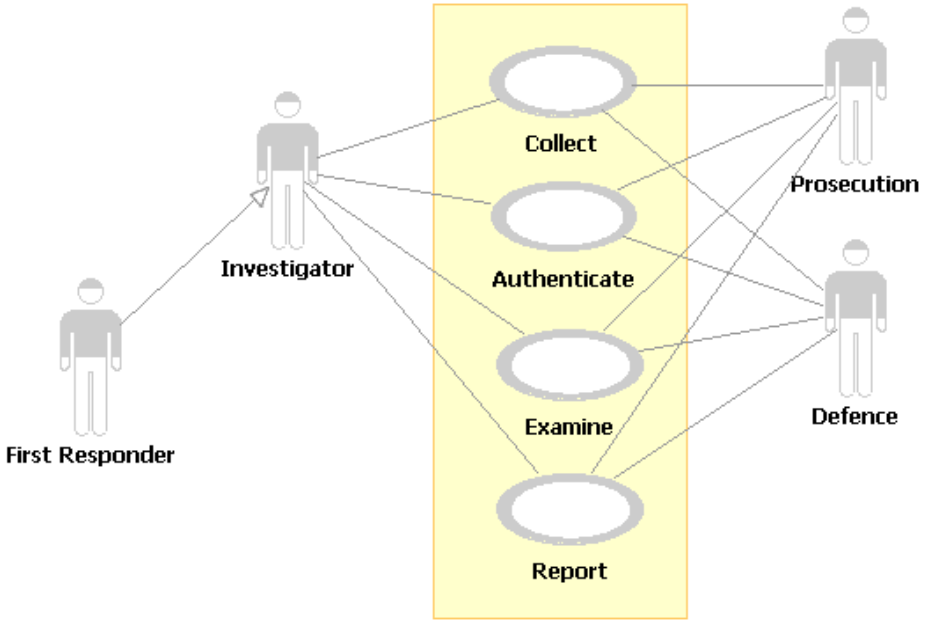

Figure 5: USDOJ Use Case Diagram

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.

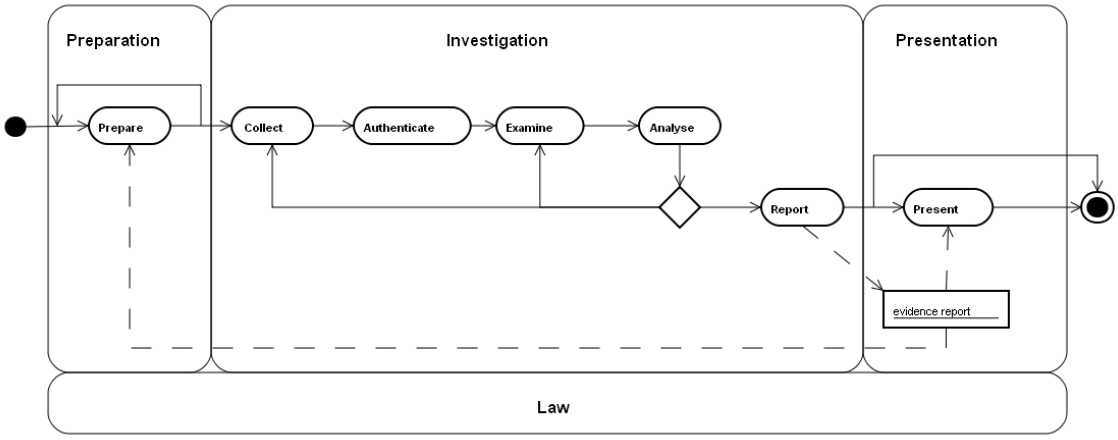

Figure 6: InteDFPM Activity Diagram

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 scruiteny 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, Examinate 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, Authente, 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 determin the admissibility and weight of each of the pieces of evidence included in the evidence report.

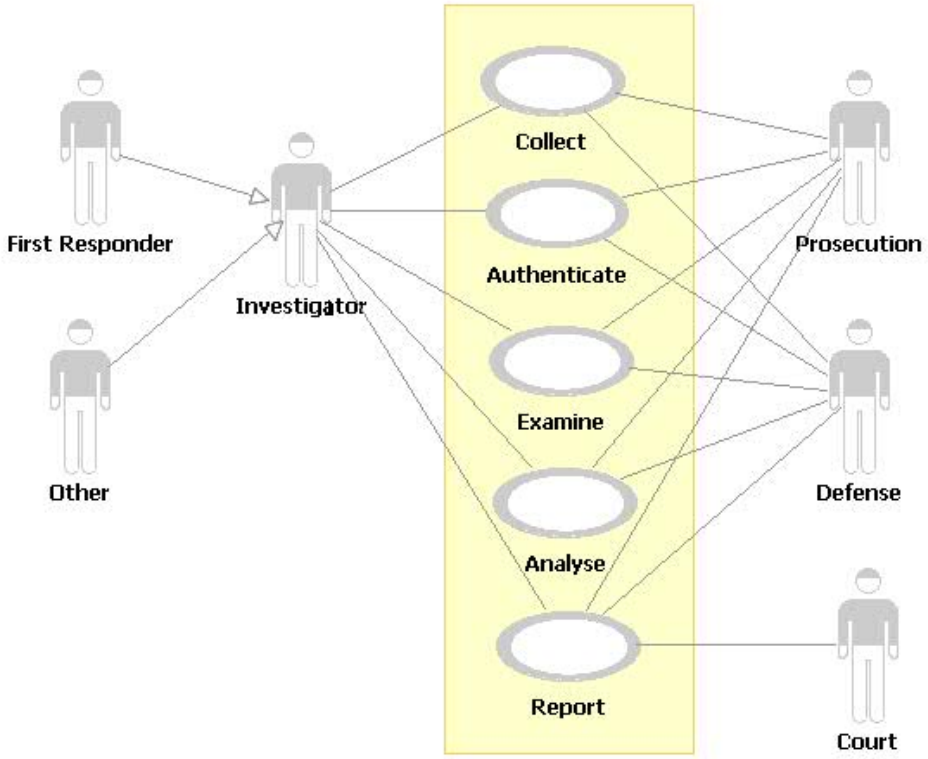

Figure 7: InteDFPM Use Case diagram

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 Responce 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 Acedemic 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.