A DIGITAL FORENSIC MODEL FOR COMPUTER NETWORKS

by

Tendai Sanyamahwe

(200604700)

Dissertation

submitted in fulfilment of the requirements of the degree

Masters of Commerce

in

Information Systems

in the

Faculty of Management and Commerce

of the

University of Fort Hare

Supervisor: Prof S. Flowerday
Co-Supervisor: Mr P. Tarwireyi

East London Campus

December 2011

DECLARATION

I, Mr Tendai Sanyamahwe, hereby declare that:

- The work in this dissertation is my own work.
- All sources used or referred to have been documented and recognised.
- This dissertation has not previously been submitted in full or partial fulfilment of the requirements for an equivalent or higher qualification at any other recognised educational institution.

Mr Tendai Sanyamahwe

_____

Signature

ABSTRACT

The Internet has become important since information is now stored in digital form and is transported both within and between organisations in large amounts through computer networks. Nevertheless, there are those individuals or groups of people who utilise the Internet to harm other businesses because they can remain relatively anonymous. To prosecute such criminals, forensic practitioners have to follow a well-defined procedure to convict responsible cyber-criminals in a court of law. Log files provide significant digital evidence in computer networks when tracing cyber-criminals. Network log mining is an evolution of typical digital forensics utilising evidence from network devices such as firewalls, switches and routers.

Network log mining is a process supported by presiding South African laws such as the Computer Evidence Act, 57 of 1983; the Electronic Communications and Transactions (ECT) Act, 25 of 2002; and the Electronic Communications Act, 36 of 2005. Nevertheless, international laws and regulations supporting network log mining include the Sarbanes-Oxley Act; the Foreign Corrupt Practices Act (FCPA) and the Bribery Act of the USA.

A digital forensic model for computer networks focusing on network log mining has been developed based on the literature reviewed and critical thought. The development of the model followed the Design Science methodology. However, this research project argues that there are some important aspects which are not fully addressed by South African presiding legislation supporting digital forensic investigations. With that in mind, this research project proposes some Forensic Investigation Precautions. These precautions were developed as part of the proposed model. The Diffusion of Innovations (DOI) Theory is the framework underpinning the development of the model and how it can be assimilated into the community.

The model was sent to IT experts for validation and this provided the qualitative element and the primary data of this research project. From these experts, this study found out that the proposed model is very unique, very comprehensive and has added new knowledge into the field of Information Technology. Also, a paper was written out of this research project.

Key words: Digital Forensics, Network Log Mining, Forensic Models

ACKNOWLEDGEMENTS

Firstly, I would like to thank Almighty God for his loving kindness and faithfulness. He has been my source of strength throughout. Ebenezer, thus far he has taken me.

Secondly, I would like to thank the following:

- ✓ Professor S. Flowerday, my supervisor for the supervision and help throughout this research project. He has been so supportive and hard working. May God reward him for his hard work.
- ✓ Mr Paul Tarwireyi, my co-supervisor for all the guidance and assistance.
- ✓ Department of Information Systems at UFH for providing a conducive environment and seminars with other IT students from other Universities.
- ✓ My family: my mother Florence Sanyamahwe, my brother Roy and my nieces Irene, Privilege and Tabonga for inspiring me. Thinking of them has helped me persevere. They are such a blessing to me.
- ✓ My fiancée Gay Makombore for her support and prayers which made it possible for me to persevere through the test of time. Her support is greatly appreciated.
- ✓ My fellow students who encouraged and motivated me.

The financial assistance of the Govan Mbeki Research and Development Centre (GMRDC) towards this research project is acknowledged. Opinions expressed in this dissertation and the conclusions arrived at, are those of the author, and are not necessarily to be attributed to the GMRDC.

TABLE OF CONTENTS

CHAPTER FOUR: ANALYSIS OF EXISTING DIGITAL FORENSIC MODELS FOR COMPUTER NETWORKS

CHAPTER FIVE: RESEARCH DESIGN AND METHODOLOGY

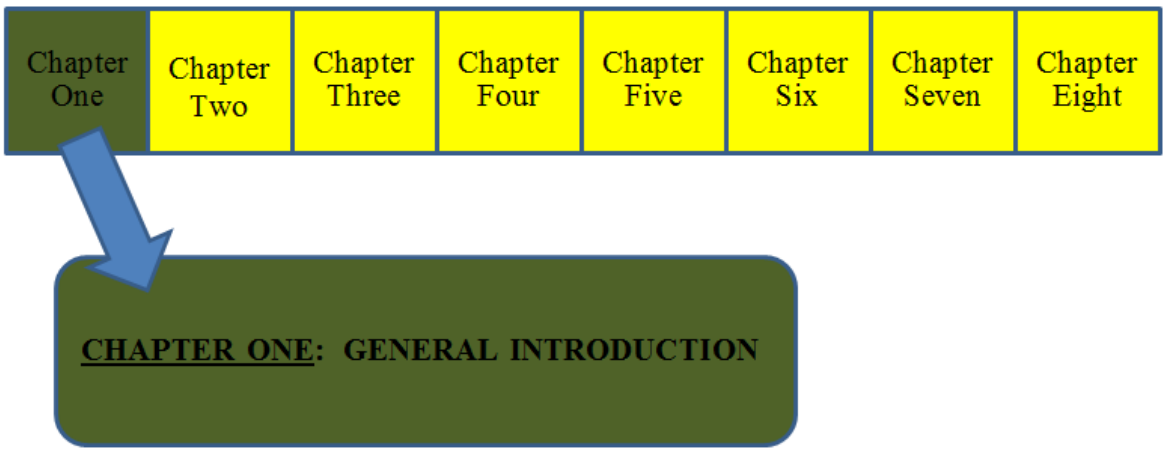
CHAPTER SIX: A LOG FILE DIGITAL FORENSIC MODEL

CHAPTER SEVEN: EXPERT REVIEWS

CHAPTER EIGHT: CONCLUSION

LIST OF FIGURES

LIST OF TABLES



1.1 Background

The Internet has become vital to most organisations since most of the information is now stored and transported in digital form both within and between organisations in large amounts through assimilated computer networks (KPMG, 2009). Various sectors, such as banking, shopping and communication have realised numerous benefits from the Internet radically changing the way many people conduct their businesses and their lives (Casey, 2008). Unfortunately, the volumes of data communicated over the Internet opens the door to abuse by criminals and terrorists who are able to remain relatively anonymous (Lim, 2008).

Organisations relying on the Internet for daily business processes face significant challenges to ensure that their networks operate safely and that their systems continue to provide critical services even in the face of attack (Stander, Dunnet & Rizzo, 2002). Moreover, it has become apparent that technology and the law do not seem to be in cooperation; instead, they grapple to find common ground (Jones, 2007a). Thus, cyber-criminals practise such criminal activities knowing that they can evade the law because of their anonymity. The lack of alignment between the law and technology causes some difficulties in computer network investigations, because there is no agreed standard model utilised in performing digital forensic investigations worldwide (Leigland & Krings, 2004; Perumal, 2009). Some other difficulties in digital forensic investigations are as a result of legislation differing from country to country. Forensic practitioners may fail to convict criminals, even if they hold potential evidence, if the information they have has not been acquired, stored and processed according to the law of that country (Perumal, 2009).

To address this issue, digital forensic practitioners have to follow a number of steps as part of the process of recovering digital footprints from computer networks as evidence of intrusion into an organisation's computer networks. Kohn, Eloff and Oliver (2006) state that the number of forensic models have added to the complexity of the field. In order to arrest and prosecute criminals involved in digital crime, practitioners must employ consistent and well-defined forensic procedures (Kohn e*t al*, 2006). Therefore, a need for standardised procedures in different scenarios/fields of digital forensics, such as database forensics, e-mail forensics and network forensics, are required. Leigland and Krings (2004) are of the view that this lack of standardisation makes the investigation processes difficult. Ieong (2006) states that a few key procedures from various authors are considered to be 'standard' procedures in digital forensic investigations. However, Ieong (2006) also notes that there are a number of discrepencies.

Digital forensics is a relatively new science. Derived as a synonym for computer forensics, its definition has expanded to include the forensics of all digital technology (Reith, Carr & Gunsch, 2002). Digital forensics has been defined as:

> *"the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations"* (Palmer, 2001, p16).

Network forensics is an evolution of typical digital forensics where the evidence is gathered from network devices such as firewalls, switches, routers and network management systems (log servers) by Information Technology security and network professionals gathering digital evidence for court cases (BBB Accredited Business, 2010). The process of network log mining is the fruition of network forensics in which forensic practitioners follow the digital footprints of the cyber-criminals until they reach the used machine and finds the particulars of the individual(s) who used the machine at the time the incident took place. The digital footprints of cybercriminals are usually recorded in the log files of a configured computer network device (Casey, 2008).

Munk, Kapusta and Svec (2010) are of the view that log files provide significant digital evidence in the computer networks when tracing cyber-criminals. This is because the log files of configured network device record all transactions which take place in an organisation's computer networks (Schuster, 2007). According to Forte (2004a), after tracking to the used machine, the forensic practitioners correlate the data found on both the log files of the used computer and the log files found in the victim organisation's computer networks. This has to be done following the legal procedures of criminal investigations (Das & Turkoglu, 2009). Therefore, network log mining is a process of discovering, extracting knowledge, modeling and analysing events recorded in the log files of computer networks in an effort to convict cyber-criminals in a court of law (Casey, 2008; Munk *et al*, 2010; Das & Turkoglu, 2009).

The lack of defined processes and procedures in this field results in incomplete evidence collection and inaccuracies in evidence interpretation (Casey, 2004b; Perumal, 2009). In addition to this complexity, the legal foundation, which is evolving slowly, restricts digital forensics (Ryan & Shpantzer, 2005; Hailey, 2010). Technology is improving on a daily basis while in most cases laws remain static, or take a significantly long time to update, and as a result, digital forensics is confined to potentially outdated laws. As a result, new concepts of forensics and advances in technology, which are not included in the judiciary, may face refutation in a court of law. Therefore, there is a need, firstly to standardise the process and secondly to comply with the law when retrieving log files from computer networks in order for the evidence to be legally admissible in a court of law (Perumal, 2009).

Law enforcement agencies have been busy trying to keep up with the criminal element that persistently abuses technology. In other words, law enforcement agencies have been trying to expand the terms of law in order to include cyber-crime in the judiciary (Reith *et al*, 2002). In this research project the term judiciary refers to the system of courts of justice in a country. In order for forensic practitioners to perform their job, there are a number of key steps that need to be considered and introduced (Kohn *et al,* 2006). These steps are encompassed in digital forensic models. Considering all that have been discussed already, the problem statement reflects the real challenge confronting network forensics.

1.2 Problem Statement

With increased connectivity due to the Internet, there are those who would utilise the Internet to exploit and cause harm to both individuals and organisations (Boeckeler, 2004; Perumal, 2009). In order for digital forensic practitioners to determine whether the network of an organisation has been compromised, various tools and techniques need to be utilised to determine the unauthorised intrusion into an organisation's computer networks.

Forensic practitioners are required to provide credible evidence in order to prosecute a suspect in a court of law (Perumal, 2009). Additionally, Chaula, Yngstrom and Kowalski (2005) say the ultimate responsibility for the credible recovery of log file entries from a compromised computer network lies with the forensic practitioner. The most serious problem is that if cyber-criminals have breached an organisation's computer network, the forensic practitioners need to determine the origin of the intrusion (Beebe & Clark 2005; Saboohi, 2008). This would be determined from log files acquired from computer networks using a process that has been standardised in the network forensic field (Beebe & Clark, 2005; Munk *et al*, 2010). The problem is that there is no standardised and widely accepted model for network forensic practitioners to follow (Reith *et al*, 2002; Lai, Gu, Jin, Wang & Li, 2011). This complicates the regulatory compliance issues thus, making it difficult to provide legally admissible evidence in a court of law.

1.3 Objective of the Study

The main objective of this research project is to produce a standardised model that will aid digital forensic practitioners in the retrieval of log files from computer networks (i.e. network log mining). Subsequently, it would assist with the correlation of log entries in a legally sound manner to ensure that the evidence is admissible in a court of law.

1.4 Research Questions

1.4.1 Main Research Question

What can be done to improve and provide a standardised process of recovering log file entries from an organisation's computer networks, in order to provide legally admissible evidence in a court of law?

1.4.2 Sub-questions

(i) What are the challenges faced in computer network investigations with respect to the recovery of log file entries from computer networks and the presentation of legally admissible evidence?

(i) What commonalities exist in models that have been used for log mining in computer network forensic investigations?

(iii) What procedures can be followed to ensure comprehensive digital evidence is collected correctly when conducting forensic log mining in computer networks?

1.5 Significance of Study

This study is important as it addresses the need to provide credible evidence that will be used in a court of law to convict suspected criminals. This needs to be done by recovering digital evidence from log files of computer network devices, and following a legal procedure through to the end of the case. Currently, there is no best practice or standard in the field of digital evidence investigation (Beebe & Clark, 2005; Carrier & Spafford, 2006; KPMG, 2009). A computer crime culprit may escape without punishment or an innocent person may suffer negative consequences simply on account of a forensic investigation that is inadequately and improperly conducted (Baryamureeba & Tushabe, 2004). Businesses lose millions of dollars when they cannot prosecute these cyber-criminals effectively (KPMG, 2009).

According to Rondganger (2008), network log mining is needed in cyber cases like the one involving the hacker Alistair Peterson, a Gauteng computer scientist, who headed an elaborate online bank-hacking syndicate. When he was caught in February 2008, he had already gathered R17 million by defrauding businesses, trust funds and corporate accounts. This issue was not solved by means of forensic practices; forensic investigators failed to convict him in a court of law, but later Alistair Peterson confessed when he developed a strong anti-virus (Rondganger, 2008). Peterson was not convicted because there was insufficient digital evidence recovered from computer networks to prove his unscrupulous behaviour. Perhaps, if there had been enough evidence from the log files from computer networks revealing Peterson's digital footprints, he might have been convicted.

1.6 Theoretical Framework

Hofstee (2006) states that, the purpose of reviewing literature is to provide the researcher with a better understanding of the research problem and a theory base of the work to be done and explain how the proposed research will fit into what has already been done. This study will review a variety of literature notably books, journals, conference papers and best practices.

This study will refer to the Diffusion of Innovations Theory (DOI). Diffusion of Innovations seeks to explain how innovations are adopted in a population. An innovation is an idea, behaviour, or object that is perceived as new by its audience (Robinson, 2009). Clarke (1999) says DOI purports to describe the patterns of adoption, explain the mechanism, and assist in predicting whether and how a new innovation will be successful. DOI theory sees innovations as being communicated through certain channels over time and within a particular social system (Rogers, 1995). In other words, DOI expresses how new ideas are received by a given population. Hence, it explains how a new model, proposed in this project, might be received by an audience.

Individuals are seen as possessing different degrees of willingness to adopt innovations and thus it is generally observed that the portion of the population, who adopt an innovation, is approximately normally distributed over time. Breaking this normal distribution into segments leads to the segregation of individuals into the following five categories of individual innovativeness (from earliest to latest adopters): innovators, early adopters, early majority, late majority, laggards (Rogers, 1995). The adoption curve is converted to a cumulative percent curve and a characteristic S-curve is generated that represents the rate of adoption of the innovation within a population. Thus, with the call for standardisation in the field of digital forensics, practitioners, lawyers, law enforcement agents as well as the judges will fall into the various adapter categories (Clarke, 1999). Creating a new process by which network forensic investigations should be conducted adds to a high degree of uncertainty inherent in all network forensic investigations. The research sub-questions are investigated in order to find out how the research main question can be solved considering the DOI theory.

(i) What are the challenges faced in computer network investigations with respect to the recovery of log file entries from computer networks and the presentation of legally admissible evidence?

Network log mining is a subset of digital forensics; therefore, most of the challenges faced during the investigations are comparatively the same as those associated in other fields of digital forensics. However, in this research project the challenges are analysed with respect to network log mining. The challenges involve the shortage of skilled labour familiar with network log mining. Numerous incompetent and unqualified practitioners have failed to convict criminals in a court of law, leading to organisations failing to trust forensic investigations (Taylor, Haggerty & Gresty, 2009). In addition, many organisations underestimate the admissibility and reliability of forensic investigations due to failure of conviction of cyber-criminals in a court of law and the inadequacy of legislation (Kent & Ghavalas, 2005). Nevertheless, various challenges will be analysed in this study in order to expose the depth of the problem in the field of network log mining.

(ii) What commonalities exist in models that have been used for log mining in computer network forensic investigations?

Computer crimes globally, are on the rise and unfortunately less than two percent of the reported cases result in conviction (Baryamureeba & Tushabe, 2004; KPMG, 2009). The process (model and approach) one adopts in conducting a digital forensics investigation is immensely crucial to the outcome of such an investigation. The model followed is a sequential process and is as follows: protect; discover; recover; reveal; access; analyse; print and provide consultation (Arthur & Venter, 2005). Overlooking one step or interchanging any of the steps may lead to incomplete or inconclusive results hence, wrong interpretations and conclusions.

A number of models that can be used in the process of seizing digital evidence from computer networks have been proposed worldwide (Beebe & Clark, 2005; Ieong, 2006; Perumal, 2009). As technology advances on a daily basis, there is a need to update the proposed forensic models so as to combat new tactics used by criminals. Some of the proposed models are presented in detail in Table 1.1.

Table 1.1: Digital Forensic Models (Own Compilation).

Model	Features of the model	Phases in the model
Kruse & Heisure (2004)	The model commonly known as the 3As.	acquiring, authenticating, analysis
Leigland & Krings (2004)	Formalises the forensic process by providing a tool to manage the procedures followed when dealing with compromised computer systems	identification, preservation, collection, examination, analysis, presentation, decision
Beebe & Clark (2005)	The model emphasises the preparation of the investigation before embarking on the technical practices of seizing digital evidence from digital sources	preparation, survey, incident response, extract, data collection, analysis, examination, presentation
Ieong (2006)	Introduction of FORZA model, which involves strengths from previously suggested models.	recognition, identification, individualism, reconstruction
Kent *et al.* (2006)	The model focuses on the important aspects of the forensic investigation	collection, examination, analysis, report
Freiling & Schwittay (2007)	A forensic model based on the analysis of the digital evidence	pre-analysis, analysis, post-analysis

Whilst these models have been proposed and are used in network log mining, there is no best practice or standardisation of the procedures followed; thus, many of the models are guidelines developed ad hoc for performing investigations (Leigland & Krings, 2004). This highlights the importance of the standardisation of procedures and techniques used in network log mining. On the other hand, Hershensohn (2005) says:

> "*Although the South African courts have not had much opportunity to litigate around this topic, we are fortunate in that we can look to the United States of America, and the United Kingdom, for rulings, which in many cases will apply to the scenarios at hand".*

Hershensohn (2005) reveals that digital forensic investigations are not conducted to full capacity in South Africa or even in Africa as a whole; hence, we need to borrow ideas from countries which have experts in the field. Also, KPMG (2009) reveals that African judiciaries refer to western countries about forensic issues. Various models will be explored, and the common elements identified and interrogated in order to provide a foundation for the development of a standardised model for log mining in computer networks; that is, a forensic model facilitating a detailed specification towards the extraction of log entries from computer network devices.

(iii) What procedures can be followed to ensure comprehensive digital evidence is collected correctly when conducting forensic log mining in computer networks?

A standardised model must be used when conducting network investigations. This is the model to be established in this study. When seizing digital evidence from computer networks, forensic law must be strictly observed in each and every step of the procedure. Also, considering that digital forensic examinations are crucial investigations (KPMG, 2009), it is important to consider confidentiality, integrity and availability of the evidence, because failure of one of these factors will result in the case rejected in a court of law.

The research design and methodology will describe how the researcher will generate a model for use in gathering digital evidence from computer networks to convict cyber-criminals in a court of law. The DOI theory elucidates how this model can be appreciated in a court a law.

1.7 Research Design and Methodology

1.7.1 Research Design

Collis and Hussey (2003) argue that positivistic and phenomenological paradigms are two extremes of the paradigm and only a few research projects would operate at these extremes. Figure 1.1 below illustrates the differences between positivistic and phenomenological approach.

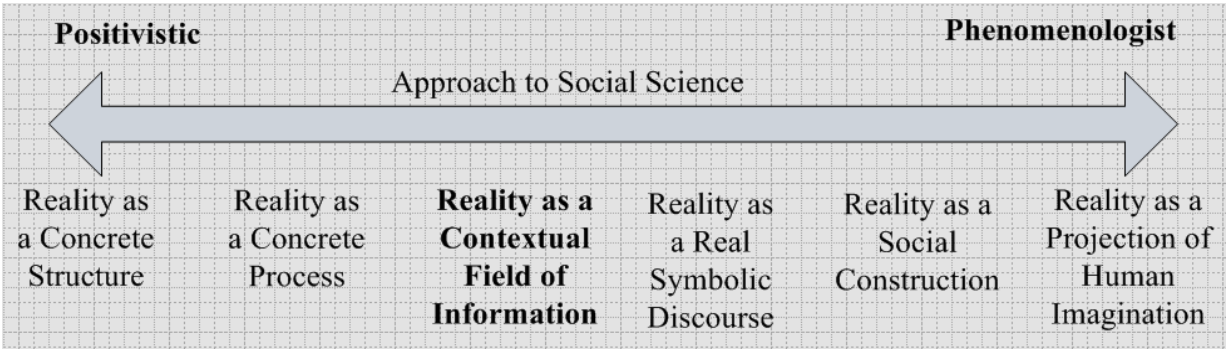


Fig1.1: Continuum of Core Ontological Assumptions: (Collis & Hussey, 2003)

When referring to the continuum of core ontological assumptions, this study falls into the "*Reality as a Contextual Field of Information"* category. In this category the social world is viewed as human beings continually processing information, learning and adapting to their environment (Van der Poll, 2004).

1.7.2 Research Methodology

The probability of a successful research project is greatly enhanced when the beginning is correctly defined as a precise statement of goals and justification (Congdon & Dunham, 1999).

Design Science (DS) research methodology, along with its research methods (data collection and data analysis) was followed. Design Science is a comprehensive problem solving process characterised by the detailed evaluation of a project, with the end goal being the creation of a model or artefact (Hevner & March, 2003). Design Science seeks to change the larger system of which the specific problem is a part, through the introduction of innovative models (Gabel, 2007). Design Science is technologically orientated and is essentially a problem solving process that leads to the development of an effective IT product, which is one of four types: constructs; methods; models or instantiations (Oates, 2006). The Design Science stage will produce a model to be used in computer network criminal investigations.

Design Science is used to assess what is being produced from each paradigm against each other in the context of business needs (Gabel, 2007). The model strived to be innovative, by solving an unsolved problem or solving a known problem in a more effective or efficient way. It should be noted however, that this study does not aim to solve the entire spectrum of problems in digital forensics, but rather produce a model, or artefact, that will add its part to the global collection of knowledge on the subject (Hevner & March, 2003).

After the model was developed, a validation stage took place, where the views and opinions of IT Experts were obtained. Thus, after the model was developed, it was sent to the Information Technology Security Experts (ITSEs) and they provided their opinions of the model. The opinions of the experts were used to obtain the most reliable consensus of opinion of a group of experts. Therefore, the experts clarified whether or not the model developed fulfilled the established requirements; that are functionality, reliability, completeness, usability and consistency (Oates, 2006). Accordingly, the feedback from the experts was used to refine the model developed in the Design Science stage.

1.7.2.1 Data Collection

Primary Data Collection

The primary data collection phase was conducted through the validation of the model by IT Experts. The created artefact (model), in a form of a paper (journal) document, was sent to the experts in four batches. After each batch the model was modified according to the feedback. In total, fourteen experts responded. The document allowed the ITSEs to express their views and opinions about the proposed forensic model. The experts in each batch were given four weeks to give their views about the new model. Most of the experts interviewed were local as well as international professors who are authors of most referenced papers concerning digital forensics, log file mining and IT auditing. Also, some of the experts are individuals working in forensic and/or IT auditing departments of some accounting firms in South Africa. From the twenty experts where the model was sent, fourteen experts responded and their responses were used to enhance the model. This provided the qualitative data element of the study.

1.7.2.2 Data Analysis

The qualitative data collected was analysed using mind maps. Mind mapping is a technique that allows a researcher to make use of available ideas to produce a new idea (Mandi, 2010). In other words, it helps one convert thoughts (ideas) generated, while creating linear thought that is needed the most (Bosley Group, 2010). The data was analysed as illustrated in the flow chart in Figure 1.2.

The flow chart has three iterations. At first, the existing forensic models from the literature were studied, their weakness and strength noted. Secondly, South African regulations, legislation and best practices in respect to digital forensics were studied from the literature. These include the Computer Evidence Act, 57 of 1983, the Electronic Communications and Transactions (ECT) Act, 25 of 2002 and the Electronic Communications (EC) Act, 36 of 2005 (Adams & Adams, 2000; South African Government Gazette, 2002; Universal Services and Access Fund Regulation (USAFR), 2008). The results from both studies formed the basis of the creation of a new forensic model, which was then sent to forensic experts. Lastly, expert reviews were analysed to refine the model. This marked the end of the process.

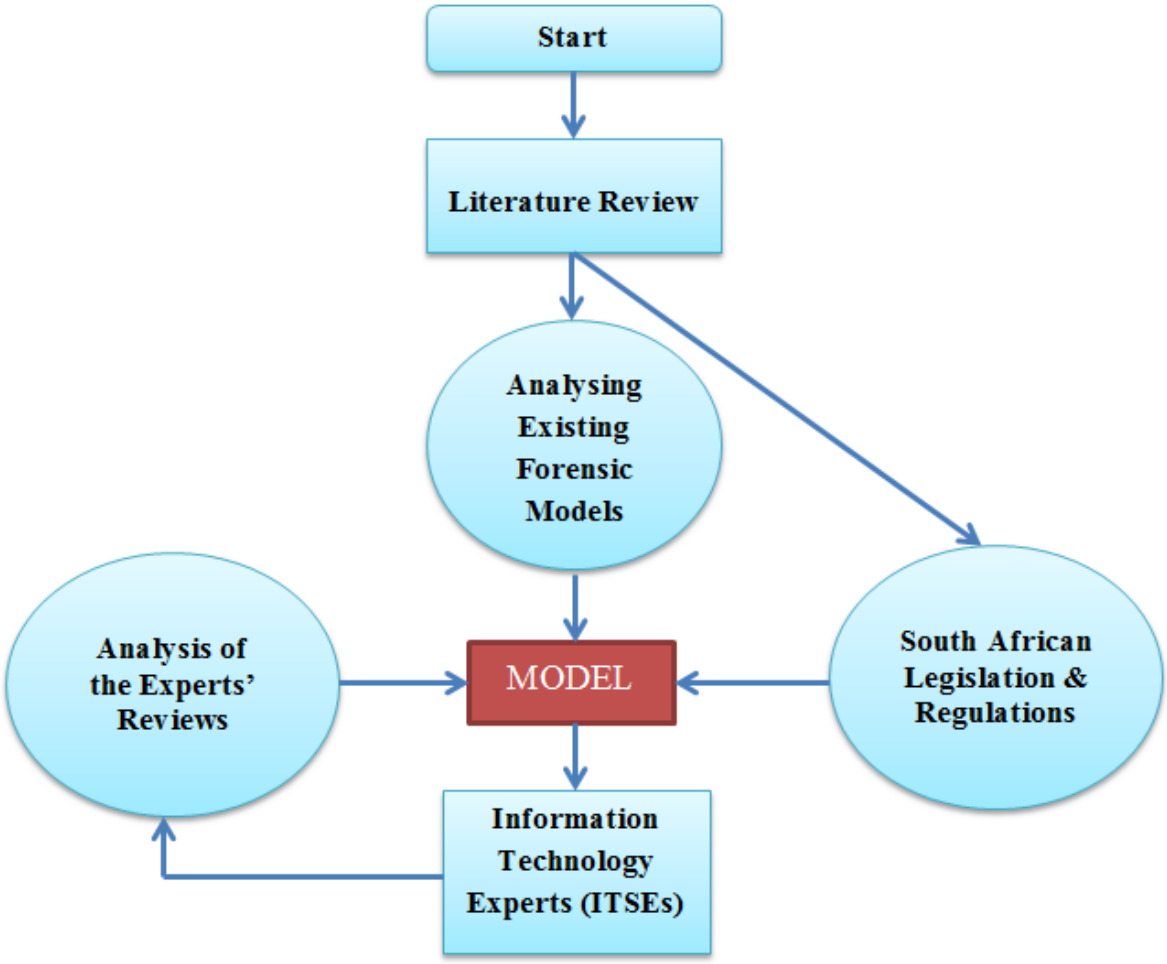


Fig 1.2: Data Analysis Structure (Own Compilation).

1.7.3 Highlights of Main Findings

From the expert reviews, this research project concluded that the proposed model was well-presented and could be easily understood, provided high technical quality, is unique, very comprehensive and has added new knowledge into the field of Information Technology and has common originality. Also, as part of achievement, a paper was written out of this research project.

1.8 Delimitation of the Study

Digital forensics is divided into three categories namely litigation support, digital media analysis and network investigation (Cardwell, Clinton, Cohen, Collins, Cornell, Cross *et al*, 2007). This study focused on the computer network investigations, with log files as the centre of attention. Thus, it was restricted to the detection of intrusion and seizure of digital evidence from log files of computer network devices. Existing literature formed the

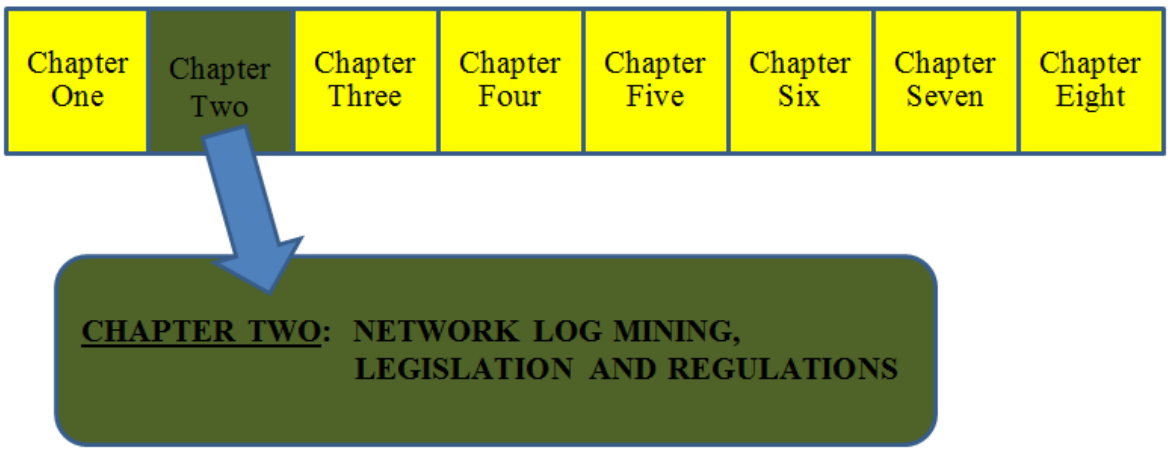
theoretical base from which the model for detection and seizure of evidence from log files was produced. The study followed the Electronic Communication Act, 36 of 2005; Computer Evidence Act, 57 of 1983; and the Electronic Communications and Transactions Act, 25 of 2002 of South African legislation, but also drew from other countries' laws and frameworks.

1.9 Ethical Considerations

The cases discussed are in the public domain and have been approved and judged by a court of law. Also during the study, the information was collected adhering to a strict confidentiality code in order to protect the privacy of organisations and individuals.

1.10 Research Project Outline

Chapter 1 presents the introduction, problem statement, research objectives and a briefing of the research methodology. Chapter 2 will provide a discussion on network log mining, laws and regulations. Chapter 3 presents the characteristics of cyber-crime and the challenges faced by forensic practitioners during network log mining. Chapter 4 reviews and analyses existing digital forensic models. Chapter 5 will discuss the research methodology, describing the methods used for data collection and data analysis. Chapter 6 will describe the proposed forensic model (artefact). Chapter 7 will outline the research analysis and findings and Chapter 8 will provide a summative conclusion of the results to determine if the research objective was addressed, and suggest areas for further research.



2.1 Introduction

In order to convict cyber-criminals who intrude into computer networks, digital forensic practitioners have to collect digital evidence from log files of network devices in a legally admissible manner (Forte, 2009). This chapter will first discuss the importance of network log mining security incidents. Subsequently, it will deliberate the issue of solidarity between legislation and digital forensics because the process of network log mining has to be conducted within the ambit of laws and regulations. Successively, the chapter will explain how digital evidence (proof from the log files) is viewed and esteemed in a court of law; and lastly, will provide some computer network case studies where log files played an important role.

2.2 Network Log Mining in Security Incidents

A typical organisation has many different devices such as firewalls, routers and servers which when configured, can all create log files of a user and system activity (Casey, 2008). Today many organisations allow users who are not employees, to access their computer networks through the use of marketing or commerce websites, e-mail and file transfers. Usually cyber-criminals take advantage of those opportunities to intrude into these organisations' private and confidential systems. Munk *et al*. (2010) are of the opinion that it is up to the network and the security departments of an organisation to ensure that access to certain information is restricted to only those services that the business wishes to make public; however, cyber-criminals usually use advanced technology to encroach into the targeted organisation's computer networks. Log files, for this reason, are used for tracing and debugging information; for that Forte (2009) says log

files are an important way for organisations to determine whether the tracing and debugging of information have been successful.

When people obtain access to an organisation's computer networks through the Internet, the log files show a record of all the activities each individual performed in the form of *Who* did *What* and *When* (Casey, 2008). Log files record each transaction executed by computer network devices at any time (Shebaro, Perez-Gonzalez & Crandall, 2010). Therefore, log files are an important source of digital forensic investigations because they usually connect a certain event to a point in time (Forte, 2004b). This is because log files contain information on transactions, connections and usage activities; so, the practice of information security through log file management provides solutions for the need for a balance between control and privacy. Thus, log files present the problem of unique identification of an intruder into an organisation's networks as the source of anonymous data about the user (Das & Turkoglu, 2009). The goal of using log files is to capture, model and analyse how individuals interact with the organisation's computer networks; hence, it is the method that provides significant digital evidence about how cyber-criminals navigate in the computer networks (Forte, 2004). Therefore, the log files are an excellent source to check the network anomalies for insider threats, e-discovery for data leakages and malicious activities for IT asset misuse as illustrated in Fig 2.1.

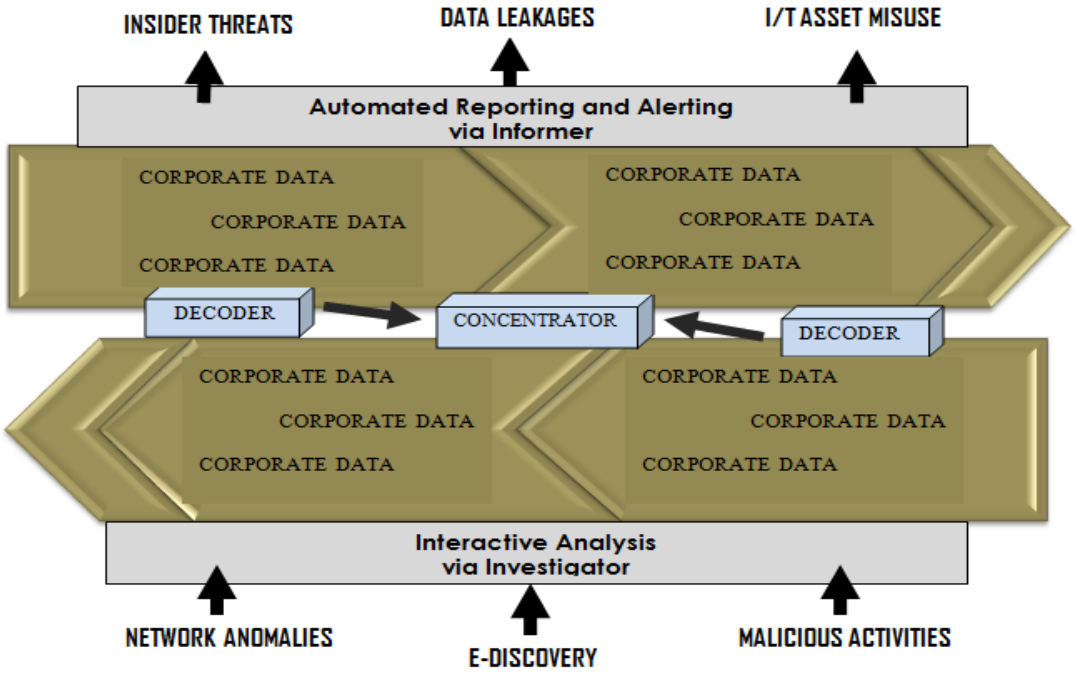


Fig 2.1: A description of Log Files use in an organisation's computer networks: (Forte, 2009)

Fig 2.1 shows a diagram illustrating how log files are used in the world of finance, especially to ensure compliance with acts such as Sarbanes-Oxley and the Gramm-Leach-Bliley (GLB) Act (Forte, 2009). From the perspective of most financial institutions like Deloitte and Touché, Ernst and Young, KPMG as well as banks, network forensics entails forensic investigations applied to information technology, predominantly the log files of computer networks (Schuster, 2007). Casey (2008) points out that network log mining is an extremely delicate discipline that requires extreme and thorough knowledge of the technical aspects of information systems in general. Financial institutions regard log files on computer network devices as important tools to use when tracing cyber-criminals through computer auditing as it is a way of convicting cyber-criminals and supported by the ISO27002 (KPMG, 2009). Moreover, Das and Turkoglu (2009) are of the view that log files are used in computer auditing because they address fraud and security extensively. Fig 2.2 illustrates how log files are used as a tool in computer auditing as a way of auditing events taking place in the computer networks of an organisation.

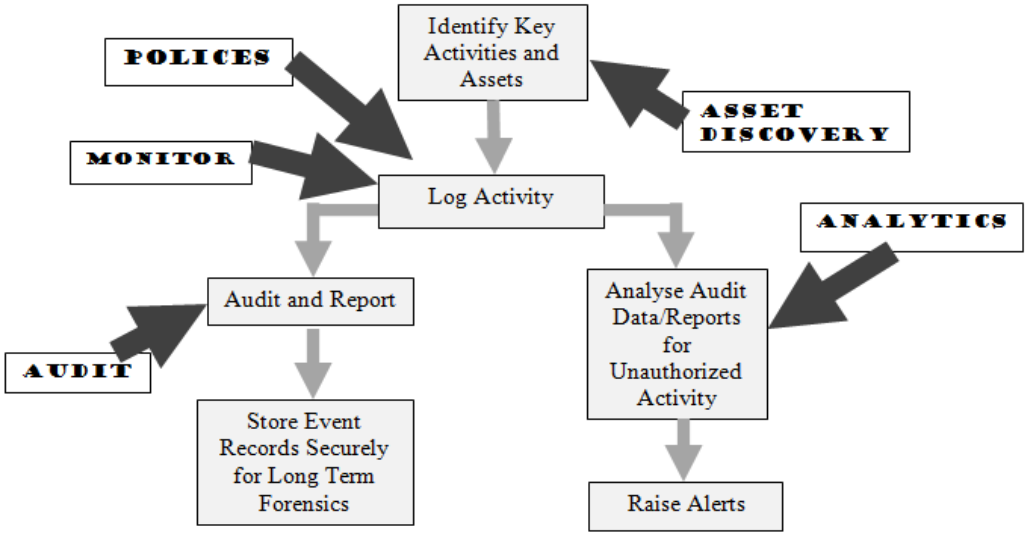


Fig 2.2: The Activity Auditing Lifecycle: (Forte, 2009)

From Fig 2.2 the inference is that when scrutinizing log files for business security purposes, log files are important in policy making, monitoring, asset discovery, auditing and analytics. Forte (2009) says policy making requires adequate security policies that account for key factors such as the laws in force in the nation where the company has operations, security management standards, risk factors and aspects relating to business. Therefore, log files are eloquent tools used to verify whether these important factors are being instigated. Correspondingly, asset discovery helps in identification of key assets and

activities which should be monitored through the use of log files (Forte, 2009). Thus, asset discovery is important in the sense that it allows precise mapping of the critical IT objects and thus, planning for how the assets will be monitored.

Based on the results of asset discovery, the next step is to monitor the assets. Organizational aspects comprise the assignment of roles and responsibilities; on the technical level, a logging process is defined to monitor the systems handling transactions which should have been identified during asset discovery (Tug, Sakiroglu & Arslan 2006; Forte, 2009). Lastly, log files are very important in auditing and analytics (KPMG, 2009). In most financial organisations, audits and analytics from log files are a dual process with maximum importance placed on the process of tracing cyber-criminals (KPMG, 2009). Thus, the audit and analytics process is an activity that results in the identification of anomalies (see Fig 2.1). Responses found from audits and analytics range from warnings (less risky) to high security incident level breaches (very risky) depending on the specific nature of the case (Forte, 2009). Therefore, the process of computer auditing and analytics using log files can be considered crucial in identifying anomalies and necessary for providing evidence for internal and external investigations. Nonetheless, during computer auditing, various types of log files can be used depending on which the organisation favours most.

The most common types of log files used, not only in many financial organisations but also in governmental organisations, are the UNIX Syslog, Microsoft Windows NT and the Microsoft Windows Vista (Shebaro *et al*, 2010). Nevertheless, these log file types have a number of log file formats that can be logged on the network devices. The most common log file formats used in digital forensic investigations include the NCSA (National Computational Science Alliance) Common log file format, W3C Extended log file format, Hypertext Transfer Protocol (HTTP).sys log file format, ODBC (Open Database Connectivity) logging, Centralized Binary logging and Microsoft IIS log file format (Das & Turkoglu, 2009). However, in applying log files as an instrument of acquiring digital evidence for court cases, the importance of the standardised procedure is always emphasised.

Shebaro *et al*. (2010) are of the view that well-established security standards and more recent ones regarding computer auditing all imply log file management, a practice that is also written into a number of national and international laws and directives. Viewing and extracting digital evidence from the log files of an organisation's computer networks for court cases is part of log file management (Casey, 2008). Log file management systems allow businesses to correlate and draw information from log files and then to begin to use them as a security asset (Yu, 2000). Traditionally, there are five ways of logging: console logging (for screen only); buffered logging for log files in the RAM; terminal logging through telnet; syslog usually used for central log servers and the AAA (authentication, authorization and accounting) protocol (Tug *et al*, 2006). In all these types of logging, each line of log entry represents a record with the Internet Protocol (IP) address, time and date of visit, accessed object and referenced object (Shebaro *et al*, 2010). A single line in a log file may appear as illustrated in Fig 2.3. The information in Fig 2.3 is described in detail in Table 2.1.

```
172.16.12.127--[10/Apr/2008:12:16:37=0300]"GET/home/images/index_on.gifHTTP/1.1"
404304"http://www.selcuk.edu.tr/""Mozilla/4.0(compatible; MSIE 5.0; Windows 98; DigExt)"
```

Fig 2.3: A single line in log files of computer networks (Own Compilation)

Table 2.1: Log file table adapted from Tug *et al*. (2006)

Related part of the line from log files	Its description
172.16.12.127	Address or DNS
-	RFC931 (or identification)
-	Authuser
[10/April/2008:12:16:37+0300]	Time Stamp
"GET/home/images/index_on.gif HTTP1.1"	HTTP request
404	Status Code
304	Transfer Volume
"http://www.selcuk.edu.tr/"	Referrer URL
"Mozilla/4.0(compartible;MSIE5.0; Windows 98; DigExt)"	User Agent

All log files must have three fundamental requisites for network forensic purposes namely integrity, time-stamping as well as normalisation and data reduction (Forte, 2004a). The integrity requisite stipulates that the logs must be unaltered and not accessible to any tampering or modification by unauthorised operators; the time-stamping requisite requires

that the logs must guarantee certainty as to the date and hour a certain event is registered for correlation purposes; and lastly, the normalisation and data reduction requisite necessitates that the log files must allow correlation tools to extract a piece of information from the source format of the log file that can be correlated with others of a different type without having to violate the integrity of the source of the information (Forte, 2004a).

Reconstruction of activities which takes place in the network devices through log file exploration is an important and challenging process of tracing cyber-criminals (KPMG, 2009). According to Das and Turkoglu (2009), the log file entries have to be retrieved (data acquisition) from the log files and then handled through the pre-processing process. After the pre-processing the data (log entries), the pattern discovery activity follows in which activity techniques from several research areas such as data mining, machine learning, statistics and pattern recognition are employed to discover how different individuals manoeuvred within the computer networks (Das & Turkoglu, 2009). Then, the pattern analysis will be conducted, choosing interesting rules, patterns and statistics. Das and Turkoglu (2009) claim that the network log mining process can be mapped to the log file processes as illustrated in Fig 2.4.

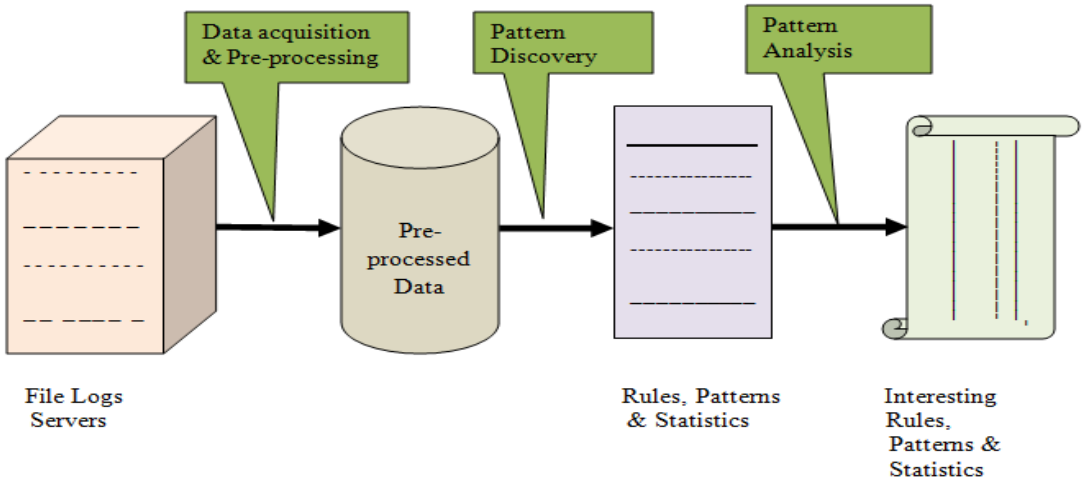


Fig 2.4: Steps in log file processes: (Das & Turkoglu, 2009)

Log mining uses the concept of data mining. Deeptee (2010) agrees with Ayre (2006) that data mining is essentially the extraction of interesting, non-trivial previously unknown and potentially useful patterns or knowledge from huge amounts of data. Therefore, network log mining is a process that refers to the recovery of relevant log entries that can provide substantial proof of intrusion, extraction of knowledge, modelling and analysis of events

recorded in the log files of computer networks as a way of convicting cyber-criminals in a court of law (Casey, 2008; Munk *et al*, 2010; Das & Turkoglu, 2009). The main advantage of data mining inherent in log mining is that, the process aids practitioners by providing them with useful and accurate trends about how people with access to the networks manoeuvre within those networks (Suherman, 2009). Thus, based on the trends, practitioners can direct their attention to the entries with more precision towards tracing the digital footprints of cyber-criminals. As log files of recent network devices can record millions of log entries per day (Tug *et al*, 2006), the technique of data mining in log files assists practitioners in selecting and copying relevant entries within a vast number of log entries. Therefore, basic log file analysis techniques like manual, filtering, summarisation and correlation can be improved through network log mining as the focus is on the relevant log entries only; hence, the analysis process becomes more efficient and effective. A logging infrastructure diagram for a filtering analysis technique is illustrated in Fig 2.5.

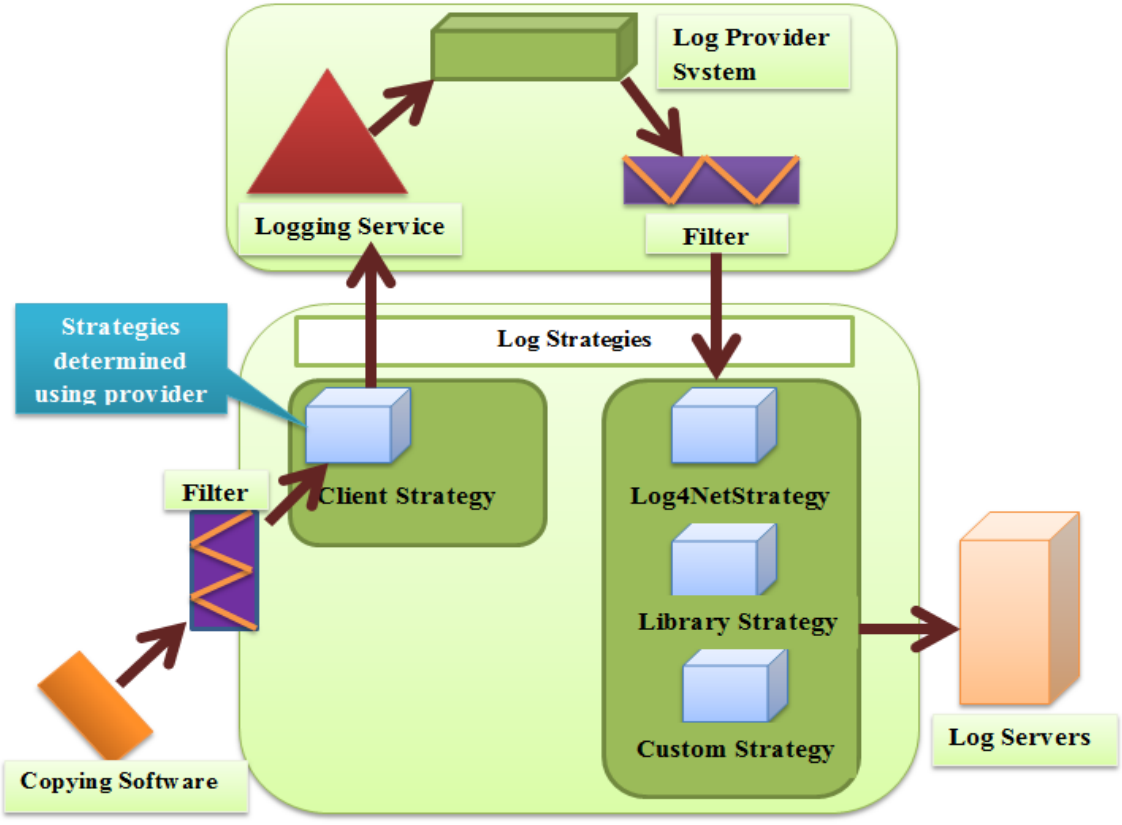


Fig 2.5: Copying log entries: logging infrastructure: (Vaughan, 2008)

After the processes of data acquisition (copying of log entries), pre-processing, pattern discovery and pattern analysis as shown in Fig 2.4, the digital evidence is presented in a

court of law for the prosecution of cyber-criminals. Consequently, the process of network log mining and/or in general forensic investigations has to be conducted according to the laws and regulations of the presiding country as well as international laws. It is important to highlight that in this research project, the words 'log file entries' and 'digital evidence' are used synonymously. The next section discusses the cohesion between network log mining and laws and regulations.

2.3 Network Log Mining alongside Legislation and Regulations

The evidence retrieved from log files is inept unless it is presented in a court of law for the prosecution of criminals. Therefore, the process of retrieving log entries from the log files of computer network devices must be guided by laws and regulations. According to Martha (2010), forensic law utilises forensic science to apply basic science methods and techniques to legal matters. Courts of laws are coming to appreciate that a piece of incontrovertible evidence which conclusively proves that someone has committed a crime, may not appear on paper only but also as electronic data (Jones, 2007b). However, a pervasive lack of knowledge about electronic data, coupled with experience grounded exclusively on paper discovery, makes it difficult for judges to meet the challenge of digital data discovery (Ball, 2005).

2.3.1 General notions of Laws behind Forensic Investigations

Network log mining which follows a general computer forensic procedure: identification, extraction, preservation, interpretation and presentation of computer-related evidence, requires expert skills and experience (Schuster, 2007). However, many who offer their services as computer forensic practitioners have no formal forensic training or certification (Taylor *et al*, 2009); which is not to say they cannot do the job well, but organisations, lawyers and judges in a court of law tend to challenge their credibility (Ball, 2005). Martha (2010) asserts that there are compelling reasons to hire a formally trained and experienced computer forensic specialist as required by laws and regulations. The reason why legislation demands competent practitioners in forensic investigations is because voluminous information that can be used as evidence may be left behind in digital sources by incompetent practitioners (Ball, 2005). Therefore, without using the appropriate procedures, tools, techniques and skills to preserve, examine and extract log entries, this study insinuates that the risk of losing important data is high, rendering findings

inadmissible in a court of law, or even the practitioner being charged with manipulation of evidence.

According to Ball (2005), the cardinal rules of law to be followed in computer forensics can be expressed in five forms: expressed as admissibility which must guide actions done thus, documenting the chain of custody; acquire the evidence without altering or damaging the original; authenticate your copy to be certain it is identical to the source data; analyse the data while retaining its integrity; and lastly anticipate the unexpected. This study proposes that the cardinal rules are designed to facilitate a forensically sound examination of network devices and to enable a forensic examiner to testify in a court of law as to the handling of log files. Lalla and Flowerday (2010) agree with Casey (2008) that forensically sound network forensics have to be conducted under controlled conditions, such that they are fully documented, replicable and verifiable. Correspondingly, a forensically sound model, in this case: the proposed model specifically for network log mining, accentuates no change to the data or evidence on the original log entries, preserving it in pristine condition. In addition, Ball (2005) articulates that the results must be replicable such that any qualified practitioner who completes an examination of the media employing the same tools and methods will secure the same results.

Of the disciplines that comprise Information Assurance (IA), digital forensics, which encompasses network log mining, is perhaps the one most closely defined by legal requirements and one whose growth and evolution is informed and guided by case law, regulatory changes, and the ability of cyber-lawyers and digital forensics practitioners to take the products of forensic tools and processes to a court of law (Ryan & Shpantzer, 2005). Roger, Goldman, Mislan, Wedge and Debrota (2006) say the tension between privacy rights and law enforcement's need to search and seize digital evidence sometimes mirrors, and frequently extends the tensions inherent in rules of evidence. This legal foundation makes forensics tools and techniques for recovery, handling, analysis and preservation of digital evidence unique among the technical arcana of IA (Ryan & Shpantzer, 2005). However, this is in contrast to firewalls, anti-virus, routing, or log servers, among others, where progress is made with much less scrutiny and little guidance from legal scholars but only from well-trained forensic practitioners (Haugdahl, 2007).

In a court of law, in defence cyber-criminal(s) focuses on whether the model used as well as the intrusion detection systems are proven technology and whether the digital evidence was gathered, stored and analysed properly (Ryan & Shpantzer, 2005). If the evidence is insufficient to refer the case to the grand jury and the charges are dropped, the business suing the criminals may lose a multi-million dollar lawsuit (Dagada, Eloff & Venter, 2009). Thus, lack of proper skills and attention to the legal rules surrounding the retrieval (collection) and use of log entries (digital evidence) cannot only make the digital evidence worthless, it can leave the practitioners vulnerable to liability in countersuits (Ryan & Shpantzer, 2005).

However, there is no specific test that can be used to determine whether the digital evidence possesses the requisite scientific validity (Brandel, 2010). In this regard, the Court in Daubert (legal precedent set by the United States Supreme Court in 1993 which defined the criteria for admissibility of expert witness testimony in the Federal Courts) suggested several factors to be considered. Such factors include whether the theories and techniques employed by the scientific practitioner have been tested; whether they have been subjected to peer review and publication; whether the techniques employed by the expert have a known error rate; whether they are subject to standards governing their application; and whether the theories and techniques employed by the expert enjoy widespread acceptance (Ryan & Shpantzer, 2005). Nonetheless, the Court in Daubert went on to say that testimony may be admissible even where one or more of the factors are unsatisfied (Ryan & Shpantzer, 2005); because the court's admissibility inquiry must focus 'solely' on the expert's principles and the model used but not the conclusion they generate (Walker, 2007).

Forensic evidence from log files proposed for admission in a court of law must satisfy two conditions: it must be relevant and it must be derived by scientific method supported by appropriate validation (Nikkle, 2006). Network forensics is highly technical and therefore grounded in science disciplines namely computer science, mathematics, physics, artificial intelligence and engineering (electrical, mechanical and systems) (Jones, 2007b); however, applying science and engineering in specific investigations is a complex process that requires professional judgement that is sometimes more art than science (Ryan & Shpantzer, 2005). In addition to the restrictions, numerous laws strictly bind network

investigators; but Jones (2007a) says the implementation of some of these laws is sometimes rather disdainful.

2.3.2 Presiding and International Laws governing Network Log Mining

According to Bejtlich (2006), a forensically sound copy of log files is created by a method that does not, in any way, alter the log entries on the logs being duplicated and a forensically-sound duplicate does not contain any data (except known filtered characters) other than that copied from the source logs. The manner used to copy the log entries has to be documented, and should be justified, referring to public law Acts, to the extent applicable (Murr, 2006). In this study, the South African presiding laws supporting network forensics (log mining) which are going to be discussed are the South African Computer Evidence Act, 57 of 1983; Electronic Communications and Transactions Act, 25 of 2002; and Electronic Communications Act, 36 of 2005; not withholding the verity that network log mining is also prominently supported by international laws.

In the context of network log mining, the Computer Evidence Act, 57 of 1983 provides for an authentication for computer generated facts (Adams & Adams, 2000). This Act endorses that the deponent or practitioner must be qualified to give testimony and have knowledge as well as experience in computer networks and examination of records including facts regarding the operation of network devices (KPMG, 2010). According to Adams and Adams (2000) in Act, 57 of 1983, the general rule is that a witness' opinion is deemed to be irrelevant and therefore inadmissible, but the practitioner must state the facts not the inferences drawn from those facts.

The second South African statute which supports network log scrutiny for court cases is the Electronic Communication and Transactions (ECT) Act, 25 of 2002. The ECT Act, 25 of 2002, is claimed by KPMG (2010) to have assailed the Computer Evidence Act, 57 of 1983. According to the South African Government Gazette (2002) the ECT Act, 25 of 2002 provides for the facilitation and regulation of electronic communication and transactions and provides for the development of natural e-strategy for the Republic of South Africa; to promote universal access to electronic communications and transactions and to prevent abuse of information system, for example cyber-crime. KPMG (2010) asserts that the issue of following digital footprints of cyber-criminals on log files as in the

South African ECT Act, 25 of 2002 is supported by some international laws. This demonstrates that the ECT Act, 25 of 2002 incorporates some measures concerning computer related crimes drawn or reflected in international laws.

Another South African enactment which corroborates forensic investigations of computer networks is the Electronic Communications (EC) Act, 36 of 2005. In accordance with network log mining, the EC Act, 36 of 2005 makes new provision for the regulation of electronic communication services and electronic communications network services and provides for the continued existence of the Universal Service Agency and the Universal Service Fund as well as making new provision for the regulation of electronic communications services and electronic communications network service (Universal Services and Access Fund Regulations (USAFR), 2008). In addition, the Independent Communication Authority of South Africa (ICASA) (2006) illustrates that the EC Act, 36 of 2005 also promotes stability in Information and Communication Technology (ICTs), ensures information security and network reliability, promotes research and development of models and methods within the ICT sector, and facilitates the development of interoperable and interconnected electronic networks, that is the provision of forensic investigations contemplated in the Act.

Current South African laws and regulations underpinning concern about tracking the digital footprints of cyber-criminals using evidence from the log files of computer networks as discussed above, is a high quality important national aspect. However, the issue of tracking digital footprints of cyber-criminals in log files is also addressed in international laws. Forte (2009) coincides with Casey (2008) that the subject of rummaging for digital footprints of cyber-criminals within retained log file entries complies with regulations such as Sarbanes-Oxley. Moreover, KPMG (2010) agrees with Gatti, Ogrady and Morgan (1997) that the practice of delving for cyber-criminals through network log mining is a process which is consolidated in international laws, namely the Foreign Corrupt Practices Act (FCPA) and the Bribery Act of the USA. With the verity that the process of network log mining is an activity supported by South African presiding as well as international laws and regulations; the next section will describe how the evidence from log files is received, viewed and scrutinised in a court of law.

2.4 Evidence from log files in a Court of Law

Digital evidence is gaining trust nowadays in most jurisdictions (Dazey, 2009); however, like any other evidence in any crime scenario, the evidence can either make or break an investigation. Therefore, it is important to ensure that all evidence from digital sources is admissible in court; otherwise, it can hurt the case should the court reject any item of evidence (Jones, 2007b). Furthermore, this rejection can portray the practitioners as incompetent. According to Brown (2005), since the items of evidence, in this case the original copy of log file entries, first need to be submitted to court for approval of admission, the correct terminology is "...*artefact of potential evidentiary value*...", rather than evidence. Brown (2005) went on to say an item can only be formally considered evidence once the court accepts it. In court very few people may be familiar with digital forensic practices; therefore, it is necessary for the witnessing practitioner to first teach the audience about the underlying scientific theory (Ryan & Shpantzer, 2005). Walker (2007) stresses that the practitioner may elicit professional opinions regarding the validity of a theory and the reliability of specific tools; however, the judges are not concerned with those but rather the methods, model and principles used during the investigation. A court may accredit a witness if they have the necessary academic qualifications or specifications, that is, specific digital forensic training as well as experience (KPMG, 2009). Many jurisdictions require the theory used by practitioners to meet certain qualifications before being used in court; in general, courts apply the Frye and Daubert tests to determine the validity of the artefacts (Ryan & Shpantzer, 2005). The Frye and Daubert tests have their roots in the United States of America (USA); however, Monteiro and Erbacher (2011) assert that even though founded in the USA, the principles are applicable to other countries. Therefore, it is important to mention that the courts referred to in this section are American in the USA; nevertheless, the ruling can be applied in South Africa as well as the rest of the world.

2.4.1 The Frye Test

A heuristic known as the Frye test controlled the admissibility of expert evidence in a District of Columbia court of Appeals case from 1923 to 1993 (Daubert v. Merrell Dow Pharmaceuticals, 1993). According to Ryan and Shpantzer (2005), the Frye test holds the expert scientific evidence admissible only if the scientific community generally accepts the scientific principles upon which the evidence from the log files is based. The Frye test

states that admissible scientific evidence must be as a result of a theory that has general acceptance in the scientific community; as a consequence, this test results in uniform decisions regarding admissibility (Cheng & Yoon, 2005). Gardner (2000) is of a view that although no law forces courts to apply the Frye test, non-committing cases can easily be appealed. The Frye test has proved to be helpful in many disputed court cases; however, with new developments in the field of digital forensics it may be problematic in some instances (Dejnozka, 2004). In the practical application of this standard, proponents of a particular scientific issue need to provide a number of experts to speak for the validity of the science behind the issue in question (Dejnozka, 2004; Cheng & Yoon, 2005). The disadvantage of the Frye test is that it may not be flexible enough to adapt to new and novel scientific issues (Gardner, 2000); however, the Daubert test was developed and replaced the Frye test.

2.4.2 The Daubert Test

Referring to the Daubert v. Merrell Dow Pharmaceuticals (1993), the Daubert test took the place of Frye test officially in controlling the admissibility of expert evidence in late 1990s. Counsel uses this technique to exclude the testimony of an expert witness who has no specific expertise in the area or used questionable methods to obtain the information (Cheng & Yoon, 2005). According to Dejnozka (2004), courts should only allow expert witnesses if their testimony is relevant to the case, with judges having authority to exclude inappropriate testimony. Before the Daubert test, trial courts preferred to let judges hear evidence offered by both sides; however, in Daubert, a motion is raised before or during the trial to exclude the presentation of unqualified evidence to the jury. Thus, the Daubert test excludes evidence which fails to meet the relevancy and reliability standard (Daubert v. Merrell Dow Pharmaceuticals, 1993).

In Daubert, the Court held that Rule 702 of the Federal Rules of Evidence, adopted in 1973, supplanted Frye. Rule 702 states that: If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise (Daubert v. Merrell Dow Pharmaceuticals, 1993). This implies that the scientific evidence proposed possesses the scientific validity to be considered competent as evidence if it is grounded in the

methods and procedures of science (Ryan & Shpantzer, 2005). Nevertheless, not all of the considerations in Daubert have to be met for evidence to be admitted; however, the majority of the tests must substantially comply with (Cheng & Yoon, 2005). In overall, the Daubert test requires four elements to be fulfilled: the theory is testable; the reliability and error rate has to be known; the extent of general acceptance by the scientific community; and whether the theory (the model) used has been reviewed (Daubert v. Merrell Dow Pharmaceuticals, 1993).

The Daubert ruling necessitates the judge to assess the scientific validity of a model or technique invoked by an expert witness before the trial starts with the intention of deciding beforehand whether the model can be applied to the facts at issue (Amenya, 2004). Amenya (2004) further says the ruling considers whether the technique or model used has been subjected to contemporary review and publication, if there are any standards controlling the technique and whether the relevant scientific community has accepted this technique. These conditions increase the importance as to why the Log File Digital Forensic Model proposed in this research project had to be reviewed by information technology security experts.

The question of applicability and decisional processes to non-scientific expert evidence was addressed by the USA Supreme Court in Kumho Tire Co. v. Carmichael (1999). Kumho Tire extended the Daubert approach to assess the reliability of expert testimony, regardless of whether the proposed testimony was based on scientific principles, engineering principles, or 'other specialized' knowledge (Kumho Tire Co. v. Carmichael, 1999). This avoided the very real problem of ambiguous decisions regarding whether the proposed testimony was rejected because it was scientific but did not satisfy Daubert criteria, or because it was non-scientific and therefore not subject to Daubert analysis but yet was defective in some other way (Kumho Tire Co. v. Carmichael, 1999). In practice, the result is that every expert, including computer forensics practitioners, are now subject to challenge reliability (Ryan & Shpantzer, 2005). Ryan and Shpantzer (2005) are of the view that trial courts and counsel are required to seek indicia of reliability that is reasonably pertinent to the practitioners' field of expertise. Cheng and Yoon (2005) say testing and verification of the model, theories and techniques of digital forensics, peer review, existence of known error rates, articulation of standards for digital forensics

investigations, and differences of opinion among digital forensics practitioners regarding applicability and acceptance of tools and techniques, are all areas that will be probed in such threshold determinations of admissibility under the Daubert test.

According to Daubert v. Merrell Dow Pharmaceuticals (1993), the courts of law in the USA do not require digital evidence to meet stringent foundations more than that required for other types of evidence. This principle can be generalised to other countries and their courts. Generally, introduction of digital evidence, in this case log file entries, is allowed provided the party offering the log entries lays a foundation sufficient to warrant a finding that such information is trustworthy and that the opposing party is given the same opportunity to determine the accuracy of the log file entries and input procedures from the network devices in order to inquire into the accuracy of written business records (Dejnozka, 2004; Cheng & Yoon, 2005). The Daubert v. Merrell Dow Pharmaceuticals, (1993) claims that the arguments that digital evidence is inherently untrustworthy because it can easily and undetectably be modified, have not been readily accepted in a court of law. From the discussion of Frye and Daubert tests this study infers that in most jurisdictions, the Daubert standard has superseded the Frye standard. This section discussed the heuristics on which digital evidence is received and perceived in courts of law in many countries. The next section deliberates on the admissibility of the evidence.

2.4.3 Log files and Legal Admissibility.

Whilst it may be possible to distinguish the original from the copy of a printed document, it is much more difficult to identify the original copy from the duplicated copies of the structures storing log file entries; this complexity complicates the defending argument to show beyond a reasonable doubt that practitioners did not modify the evidence (Dazey & Gryzbowski, 2009). This is why practitioners are encouraged to authenticate the original copy of the log entries soon after retrieving them from log files. According to Fujitsu (2002), if as part of a case, the log file entries are challenged in a court of law, the computer network devices where the logs were copied from as well as the storage structures such as digital evidence bags (DEBs), may be called into question. In addition to that, Fujitsu (2002) stresses that a seemingly acceptable copy of log entries may be discredited by demonstrating that the log entries were not copied or stored in a proper manner or else that the network system was unreliable and not maintained satisfactorily.

Monteiro and Erbacher (2011) state that there are five principles explicit to a true representation of information: duty of care by practitioners; business procedures and processes followed; and technologically and audited trailed evidence which the judge examines before considering any judgement.

Legal admissibility is about deciding whether log entries retrieved would be accepted in a court of law; if however the retrieved log entries do not carry enough evidential weight, the evidence could potentially harm the case (Dazey, 2009). Therefore, practitioners have to assess log entries for evidential weight before a decision is taken. Fujitsu (2002) says the admissibility of evidence from log files will be less readily challenged if practitioners can demonstrate that the theories, model and principles followed operated to support a business function that produced log files as part of that function and created as well as maintained accurate logs, and produced logs whenever anyone entered the computer networks. In a court of law it is therefore crucial for practitioners to prove that: the log entries are accurate, that they are a complete and unaltered representation of the information; the log entries are authentic, that is they are what they purport to be; the log entries have not been tampered with (believable); and that the log entries are stored in a system (e.g. DEBs) that has been secured throughout the document's lifetime (Dazey & Gryzbowski, 2009). If these factors cannot be proven, then the evidential weight of the record is reduced and the case may be rejected.

Dazey and Gryzbowski (2009) go on to say that there are no firm rules for determining whether a digital record is legally admissible but it is possible to maximise its evidential weight by following stipulated guidance. Nowadays, due to international laws which support the use of log files to curb cyber-crime in financial organisations, the Computer Security Institute NetSec (CSI NetSec) (2009) says the practitioners must surrender a log file high-performance architecture of the victim organisation to the judges in order to enlighten them about how the log files are used in the organisation as well as how the information is accessed by different users of the organisation's computer networks. Fig 2.6 shows an example of log file high-performance architecture.

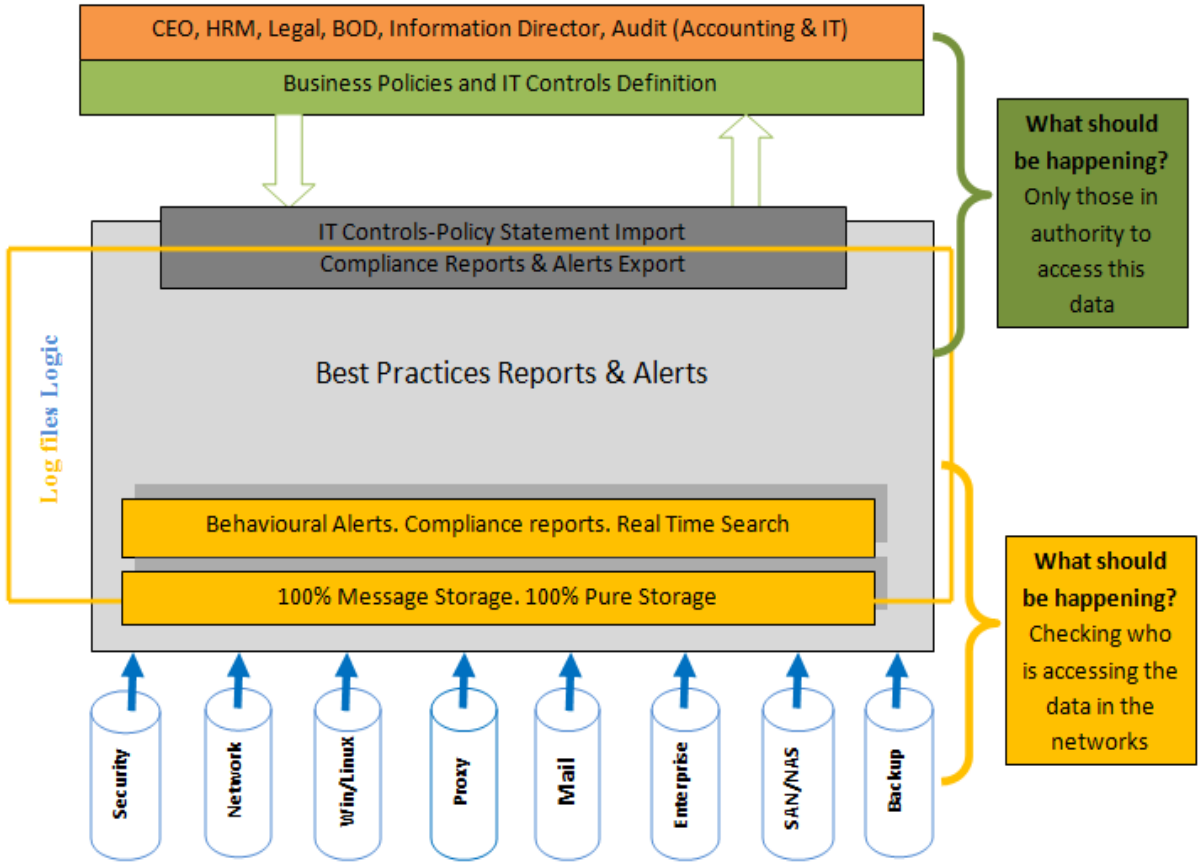


Fig 2.6: Log File High-Performance Architecture: (CIS NetSec, 2009)

According to Monteiro and Erbacher (2011), after a network attack there could be evidence in the log files throughout the entire organisation; this makes it unfeasible to take the systems with potential evidence off-line as in traditional computer forensics. Thus, removal of network devices is usually discouraged as they are functional in the everyday life of a business, unless law enforcement ordered such action (Roger *et al*, 2006). This implies that the evidence from log files needs to have a degree of certainty attached to it in order to make it credible, and thus for it to be legally admissible in a court of law (Monteiro & Erbacher, 2011). A mapping of certainty levels to log files, as adapted from a certainty model developed by Casey (2004b) is presented in Table 2.2.

Table 2.2: Certainty of Log Files, adapted from Casey's certainty model (2004b)

Level	Level Confidence	Relationship to Log Files
C_0	Erroneous/Incorrect	Retrieval errors while copying log entries from the log files; slurred images and not copying relevant log entries.
C_1	Highly Uncertain	Unreliable/patchy log file entries with manipulated log entries.
C_2	Somewhat Uncertain	In the event of an attack, the only evidence that is available is the log entries from computer network devices.
C_3	Possible	Variations in evidence from different log file formats (Unix Syslog, Microsoft Windows XP and Microsoft Windows Vista).
C_4	Probable	Log file entries and authentication traces that are stored and transmitted in digital evidence bags (DEBs) classified to have this level of certainty.
C_5	Almost Certain	This level of certainty specifies log entries to tamperproof and asserts a match between independent sources of log files.
C_6	Certain	If authentication traces were validated at every system/network devices that they were generated on and more importantly at immediate stages in the process of collecting log entries from computer networks.

This study, to this point, has unveiled the issue of how evidence from log files is prospected, received and reviewed in a court of law. The next section discusses practical examples where log files from computer networks were used to solve some disputes in certain organisations.

2.5 Network Forensics Case Studies

Cyber-crime can take place at any time to any organization or individual as long as there is information stored in a computer which is linked to other computers through the Internet. Therefore, it is important for business organisations to be prepared to curb the problem of cyber-crime in case it ensues at any point in time of the business' life. Case studies below illustrate some tangible instances.

Case Study One

In a Trojan defence case performed in tandem with a UK police force, 7Safe Information Security Services (7SISS) was able to dismiss the defendant's allegations that malicious software on his computer was responsible for the illegal images found on it. This was achieved by 7SISS analysing the log files of the machine and locating all malicious software present and active, then investigating each separate instance of the malicious

software. Utilising network log files experience, 7SISS checked traffic flows to and from the computer to investigate and verify results. During the investigation, keystroke logger output text log file entries were discovered. The log entries recorded the defendant's own keystrokes, including the input of incriminating web site addresses and his instant messaging conversations. The defendant was found guilty of all charges and was given a custodial sentence (Kent, 2010).

Case Study Two

In a malicious hacking investigation, 7SISS was commissioned to investigate a claim by a company who suspected that they had been hacked into by a third party. Investigators were able to determine through log file investigations that the company's computer networks indeed were compromised and that 'back door' recent access software had been installed in a number of their systems. Further analysis revealed that keystroke loggers had also been deployed to harvest information. The third part had a remote access to the company's servers and workstations with administrative privileges. 7SISS was able to trace the identity of the third party hackers after examining the log entries and it was found to be a rival company intent on stealing the intellectual property of that company. The expert witness's testimony resulted in civil action being successfully pursued by the client who was able to recover damages from the offending third party (Kent, 2010).

Case Study Three

After finding pornography downloaded on its network server and a number of individual office computers, one of the clients of the Data Recovery Services began to build a case for employee dismissal. Data Recovery Services was hired to locate any deleted files and verify certain illicit and non-work related contents of the hard drives in question. Forensic technicians were able to locate spy software, illegal file-sharing software, pornography, and information pertaining to a personal side business in the log files of the routers. Both the Chief Executive Officer and the network administrator were dismissed as a result of the investigation (Data Recovery Services, 2010).

Case Study Four

After being sued for negligence, a client of Data Recover Services was about to settle a multi-million dollar suit and re-write the entire software package because the plaintiff was

charging: installation of the software in question had permanently damaged and erased existing files; and files to a specific software application critical to running his business including irreplaceable data were not recoverable by any means. Data Recovery Services was able to restructure and reformat all the files needed for the claimant's specific software application and reprogram data. Using experience of log file examination and forensic in computer network analysis applications, Data Recovery Services discovered that the software installation had not caused the data loss and determined the plaintiff had manually erased the alleged lost data. When shown the evidence the plaintiff dropped the suit and was promptly counter sued (Data Recovery Services, 2010).

Case Study Five

A family-owned product manufacturer and designer on the verge of being bought for many millions of dollars found most of its designs missing after the departure of key managers and designers. A program used for deep file destruction had been implemented to destroy both product designs and evidence of the procedure itself. An external consultant was brought in to recover designs and overwrites evidence. Burgess Forensics was brought in and discovered remnants of file destruction utility in the log files of the computer networks and data patterns consistent with sabotage by the same utility. Suspects admitted the use of the utility and were arrested (Burgess Forensics, 2009).

Case Study Six

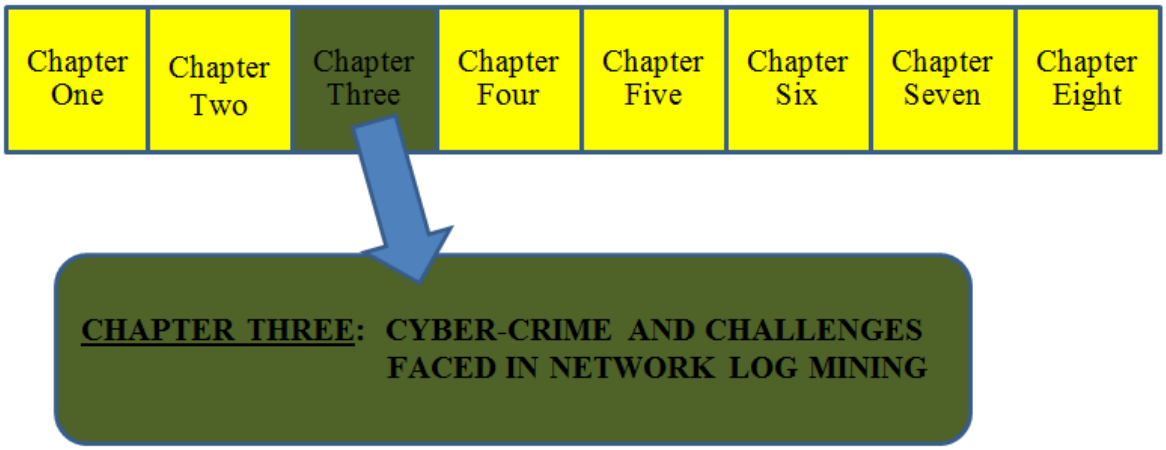
One of the most prominent cybercrime cases involved the hacker Alistair Peterson. Peterson was a South African computer scientist who headed an elaborate online bank-hacking syndicate. When he was caught in February 2008, he had already accumulated R17 million by defrauding businesses, trust funds banks and corporate accounts. Peterson entered an imploration bargain with the Scorpions Legal practitioners. Part of the plea bargain was that he works with the Council for Scientific and Industrial Research (CSIR) to develop an anti-virus to prevent further attacks by Regger.W32, a virus programme originally created by him. This crime was not solved by means of forensic practices because not enough evidence from the log files of computer networks of the hacked organisations was available to convict him; however, Peterson confessed afterwards that he enjoyed the forensic part of computers, looking for holes in systems that could be plugged by cyber-crime fighters to prevent criminal access (Rondganger, 2008).

From the case studies examined, this study suggests that cyber-criminals can be traced and convicted through the use of log files in computer networks as long as the process is conducted within laws and regulations. In addition, misunderstandings which can happen within some organisations involving the use of computers and their networks can be solved through the analysis of log files of the network devices. Nevertheless, even though evidence from log files can convict cyber-criminals and solve some disputes within organisations; there should be enough evidence to provide legitimate proof; if the evidence from the log files is insufficient it may be difficult to pursue a guilty verdict of the accused individual(s).

2.6 Conclusion

This chapter revealed the importance of network log mining in business organisations. Then, the relationship between the process of network log mining and legislation and regulations was discussed. Network log mining, however, is a process supported by both presiding as well as international legislation and regulations; for example in South Africa there is the Computer Evidence Act, 57 of 1983; the Electronic Communications and Transactions (ECT) Act, 25 of 2002; and the Electronic Communications Act, 36 of 2005. International laws such as the Sarbanes-Oxley Act; the Foreign Corrupt Practices Act (FCPA) and the Bribery Act all from the USA support the gathering of digital evidence from log files for court cases. In a court of law, the admissibility of evidence from log files follows the Fry Test and the Daubert Test heuristics. Without following the standards and laws, the evidence will be null and void, resulting in failure to convict criminals. Lastly, the case studies involving network log mining were discussed in a way to prove that the process of acquiring log files from computer networks is extremely important nowadays in the lifetime of almost all business organisations.

As network log mining is vital for tracing the digital footprints of cyber-criminals who have intruded into an organisation's computer networks, it is important to understand what cyber-crime really is and the difficulties or challenges confronting practitioners who undertake the investigation of cyber-crime in business organisations. The following chapter will discuss more on what cyber-crime is and the challenges faced by practitioners in network log mining.



3.1 Introduction

In chapter 2, the importance of network log mining, the relationship between network log mining and legislation, as well as how digital evidence is viewed or perceived in a court of law were discussed. According to Forte (2009), cyber-crime is one of the main reasons why business organizations today appreciate the process of network log mining. Combating cyber-crime and obtaining successful convictions remains a challenge for present legislation around the world due to the fast moving pace of technology (Saboohi, 2008). In terms of financial losses, organisations are losing billions of dollars annually (Saboohi, 2008). In 1998, a loss of US$ 11 billion in revenue was suffered in the United States of America alone and in 2007, 89% of businesses were the target of cyber-crime in the United Kingdom, which resulted in a loss of Great Britain Pound (GBP) £2.4 billion (Saboohi, 2008). Therefore, if left ignored, cyber-crime can potentially affect the expansion of electronic commerce worldwide (KPMG, 2009). This is the reason why this study asserts that there should be a proper way of retrieving digital footprints of cyber-criminals from log files of organisations' computer network devices for court cases. In this chapter, the discussion rests mainly upon definition of cyber-crime and the challenges confronting digital forensic investigations in network log mining as well as lastly the supporting IT security standards.

3.2 Cybercrime

According to Vezina (2008), cyber-crime is an offense where the Internet or computers are used as a medium to commit crime. In addition, the South African ECT Act, 25 of 2002 describes cyber-crime as when a person accesses data to which he/she is not authorized to do so (South African Government Gazette, 2002). The South African Government Gazette

(2002) also states that cyber-crime can further be defined as the interception or interference with data stored in a computer, computer related extortion, fraud and forgery, and also attempting, aiding or supporting the above. Deriving from the definitions given, one can conclude that cyber-crime is a form of offense which involves interference of someone's data or information using computers without the permission of the owner.

The concept of cyber-crime is not radically different from the concept of conventional crime because both involve the breaching of laws and regulations of a state (Pati, 2003). However, conventional crime is committed physically (Zvekic, 2009) while cyber-crime is committed in virtual space (Chizoba, 2005). Chizoba (2005) goes on to say that crime committed in virtual space is a law-breaking activity through local and global computer networks. Saboohi (2008) is of the view that cyber-crime generally falls into three categories which are: crimes where a computer is the target, for example hacking; crimes where the computers are the medium by which criminal operations are executed, for example Internet fraud; and crimes where the use of a computer is incidental to criminal acts, for example storing information on a computer about drug trafficking. Some of the primary characteristics of cybercrime are discussed in the next section.

3.2.1 Features of Cybercrime

Anonymity is the major characteristic of cyber-crime (Shinder, 2002; Brenner, 2004). Nowadays organisations spend considerable amounts on IT Security to protect their networks and software from external and internal threats of hackers, viruses and fraud of which the source is usually mysterious (KPMG, 2009). Unfortunately, often when an organisation is confronted with a cyber-attack, they do not know where to start, what to do, or even whom to consult (Saboohi, 2008). According to Shinder (2002), cyber-crime spans not only within a country but also international boundaries. In 2001 the 10th United Nations Congress on the Prevention of Crime and Treatment of Offenders workshop which was devoted to the issues of crimes related to computers and their networks, was conducted (Shinder, 2002). From this congress, Shinder (2002) claims that the major features of cyber-crime were established and these included: illegal access of a computer system; damage to computer data or programs; computer sabotage; illegal interception of communication (network intrusion); and computer espionage. This study concedes that some of the features of cyber-crime listed at this congress are activities currently

happening in many organisations on a regular basis, but the lack of forensic knowledge and laws make people obliviously commit cyber-crime and subsequently, organisations paying no attention to such activities.

Adding to the features mentioned above, Brenner (2004) and Chizoba (2005) mention other features of cyber-crime as they compare the attributes of cyber-crime and conventional crime. Brenner (2004) notes that, unlike conventional crime, cyber-crime does not require physical proximity between the victim and the criminal; thus, only a computer linked to the Internet allows a criminal to make attacks. Additionally, Brenner (2004) discovered that in cyber-crime little effort can lead to significant distraction because cyber-crime requires few resources relative to potential damage. Furthermore, cyber-crime can be committed instantly with potentially more than one crime committed at a time (Chizoba, 2005). Moreover, it is difficult to identify the offender and offence pattern in cyber-crime as compared to conventional crime (Brenner, 2004). Therefore, the features of cyber-crime discussed above show that cyber-crime is a complicated type of crime which is difficult to combat; hence, digital forensic practitioners must have a well-defined standardized procedure of hunting-down cyber-criminals until these criminals are convicted in a court of law. Conversely, even though people understand that cyber-crime is illegal, there are various reasons why people continue committing this type of offense.

3.2.2 Reasons behind Cyber-crime Incidents

There is a range of cyber-crimes which vary from being least harmful to more dangerous (Chizoba, 2005). Some individuals commit cyber-crime for the sake of recognition and this is usually associated with people who want to be known in society; hence, these people do not mean to hurt anyone in particular but they just want to be distinguished (Chizoba, 2005). In addition, simple negligence can give rise to criminal activities, for example saving a password on an official computer, using official data in a public place or even storing data without protecting it, resulting in anyone gaining access of the computer to log on and viewing confidential data (Sentor, 2007). Cyber-crime is also associated with individuals who want to make money illegally; usually this group of people is motivated by greed and are career cyber-criminals who tamper with data on the net or system especially e-commerce and e-banking data information with the sole aim of committing fraud and deceiving unsuspecting customers (Chizoba, 2005).

Furthermore, Sentor (2007) asserts that complex codes of operating systems that can be decoded or manipulated to gain access to a computer system are sometimes utilized by cyber-criminals; thus, some criminals who are skilled in computer programming can exploit weaknesses in the codes of the security systems and operating systems to gain access to the private information of business organisations. Moreover, cyber-crime can be committed by individuals who are fighting a cause they believe in resulting in threats and more often damage that may affect the recipients adversely (Chizoba, 2005); this is the most dangerous of all reasons for cyber-crime (KPMG, 2009). Chizoba (2005) goes on to say that such individuals are cyber-terrorists who believe that they are fighting a just cause; therefore, they do not mind who or what they destroy in their quest to reach their ultimate goals. Finally, insufficient evidence to bind cyber-criminals by laws and regulations provides the space for individuals to commit the crime without fear; unfortunately, there are so many ways to hide the trail of a cyber-crime and little evidence to convict the criminals in a court of law (Sentor, 2007). In summary, to successfully convict cyber-criminals, there has to be an orthodox trailing technique which can be followed when gathering evidence from computer network devices in order to find physical evidence to use in a court of law. Cyber-crime is the main reason why business organisations delve into network log mining.

According to Seth (2010), cyber-crime statistics are intensifying significantly per year and it is an extremely challenging task to present a critique on the subject of combating cyber-crime in law enforcement and in most judiciaries of the world. However, some organisations are training IT and forensic practitioners, specifically computer auditing courses, so as to relegate the problem of cyber-crime in most business organisations (KPMG, 2009). Even though academia, scientists, lawyers, governments and private organisations are putting a great deal of effort into curbing cyber-crime in businesses (KPMG, 2009), there are some challenges that forensic practitioners are facing in the process of network log mining.

3.3 Challenges faced in Network Log Mining

According to Jones (2007a), computer network forensic investigations are exciting; nonetheless, it is an incredibly challenging field. Its landscape is constantly changing, thus, new ideas and skills are introduced into the technical aspect of network forensics on a daily basis. Cohen (2010) supports this idea as he says the collection of digital evidence

from log files is not an easy task; extra care, new ideas, experience and skills need to be applied. Network forensic investigations, which involve network log mining, are constrained by technical issues bound by the laws within which practitioners must operate (Jones, 2007b). Forensic investigation is a practical field used across the globe to enable prosecutions to be mounted on a wide range of criminals (Jones, 2007a). Ayers (2009) asserts that digital forensic practitioners face many challenges during the process. Presumably, these challenges are hindering the effective conviction of cyber-criminals in courts of law. Some of these challenges are revealed in the following sub-sections.

3.3.1 Forensic Tools

General purpose forensic investigation tools such as EnCase and Forensic Tool Kit (FTK), provide a convenient and user-friendly environment for conducting forensic analysis (Ayers, 2009). Before the introduction of these tools, even a simple analysis was difficult and frequently required hexadecimal-level examination of data and manual interpretation of file system structures (Richard & Roussev, 2006). Garcia (2007) and Ayers (2009) agree that most of the existing forensic tools are now industry standard tools for network forensics because they enable read-only access to the contents of the file system using the Graphical User Interface (GUI) based user interface. Analysts are capable of looking through the file contents, searching keywords and employing a range of other analysis techniques (Richard & Roussev, 2006). Although these tools have greatly enhanced the practitioners' ability to analyse digital evidence, the functional limitations of the existing forensic tools have become apparent (Ayers, 2009). Some of the limitations are described below.

Processing Speed

The forensic tools available are struggling to keep pace with modern analysis workloads (Ayers, 2009). For example, recent log files can record 240 million entries per day (Munk *et al*, 2010), which is a very large workload for the forensic tools. The reason is that even when deployed on expensive high end work-stations with multiple processor cores, large amounts of memory and fast disk storage with the ability of a single or even multi-threaded application to quickly process log entries, processing speed is constrained (Ayers, 2009). Also, offloading log files in the memory to an external storage system, for example into a digital evidence bag, which is the approach used by version 2 of FTK, is of marginal

benefit with a single-threaded application (Richard & Roussev, 2006). Garcia (2007) noticed that in some instances network log analysis is conducted urgently as part of a serious crime investigation where there is a risk to the safety of the public; however, practitioners face delays of several hours or even days when processing large volumes of log entries. Therefore, there is an apparent need to increase the speed with which these investigations can be completed (Ayers, 2009). Usually, the drawback of slow processing speed of the forensic tools is congruent with data analysis and input/output bottlenecks.

Data Analysis and Input/Output bottlenecks

Most of the forensic investigation analysis tasks, with the exception of password cracking, are input/output (I/O) bound (Ayers, 2009). The operation of a forensic analysis system will be determined by the speed at which evidence data can be accessed (Garcia; 2007); however, this in turn is influenced by data and algorithm design (Richard & Roussev, 2006). With the available tools, I/O optimisation is usually limited to selection and configuration of the fastest possible disk storage systems (Ayers, 2009). Even if analysts take some caution and spends significant money to install high performance disk systems, Ayers (2009) is of the view that this achieves only a modest improvement in data throughput.

Richard and Roussev (2006) say that most forensic investigation tools use either flat image files, for example those created by the UNIX tool "dd[3]"; and by hardware devices such as the Voom Hardcopy data capture unit, or a flat image encapsulated in a proprietary data format to store metadata and ensure evidential integrity. This introduces significant distractions when a new case is loaded or when recovering from a crash, particularly if large composite log files have been opened (Richard & Roussev, 2006; Garcia, 2007). Furthermore, Ayers (2009) realised that since evidential data (log entries) was effectively retained in its raw format, any fragmentation in the original evidence continued to slow down the parsing of log entries for the duration of the case. The riddle of delays in data analysis and I/O bottlenecks is exacerbated because existing forensic tools cannot be audited (Garcia, 2007).

Auditability

Leading forensic tools, for example EnCase and FTK are closed-source commercial applications (Richard & Roussev, 2006; Ayers, 2009). The ability of practitioners to gain insight into how they work, how the log entries have been captured and the accuracy of that capture is very limited (Ayers, 2009). Most of the time, the investigators depend upon information provided by the software vendors (Garcia, 2007); thus, most existing forensic tools do not provide adequate information about logging and debugging that would help practitioners to validate the operation of the tool with respect to log extraction activities (Ayers, 2009). Usually the practitioners compare the results of different tools to ensure they deliver the same output (Richard & Roussev, 2006; Garcia, 2007; Ayers, 2009). Hence, practitioners are not free to browse the source code for the tool or to even use debuggers to trace the operation of the tool in respect to particular evidence (Ayers, 2009). Therefore, one can conclude that when using existing forensic tools, forensic practitioners confront significant challenges.

Planning and Control of Analysis tasks

Most of the available forensic tools provide poor support for detailed planning of investigative tasks and recording of the detailed results of the investigative process (Garcia, 2007). This problem is enlarged when practitioners cannot audit the tools to improve them so as to store detailed results of the investigations; therefore, most of the tasks are left to investigators who use nothing more than a pen and a notebook (Ayers, 2009). However, the caution taken by practitioners when organising the log analysis process and recording results manually varies widely (Ayers, 2009). Moreover, no matter how careful practitioners are, they are prone to mistakes (Jones, 2007a) since log analysis is a very crucial and difficult process to carry out (Forte, 2009). When the process is undertaken manually, mistakes are inevitable.

Data Abstraction

Today, available digital forensic investigation tools have made it possible to review captured log files in a forensically sound manner (Richard & Roussev, 2006; Ayers, 2009). Nevertheless, Richard and Roussev (2006) went on to say that these tools became persistent around the abstraction of a hierarchical file system. In other words, current forensic tools remain concentrated at a single log entry at a time capturing all entries one-

by-one rather than on a higher level capturing suitable entries in multiple log files at once (Ayers, 2009). Nonetheless, the main objective of network log mining is not to capture all data in the log files but to identify and capture relevant log entries in a quick manner. Therefore, the practitioner's ability to recognise and analyse relevant log entries can be improved if the evidence is presented at a higher level of abstraction which nets more appropriate log entries than the present lower level (Garcia, 2007).

The discussion above shows that most of the existing forensic tools have some weaknesses which need to be addressed so as to improve the process of log mining in computer networks. The processing speed, auditability, data abstraction and data capturing of the tools need to be improved to augment the process of collecting and analysing digital evidence from log files. Otherwise, evidence will continuously exhibit shortfalls which can be utilised by the criminals as they challenge the evidence in a court of law; due to known flaws associated with available current forensic tools. However, not only problems allied with forensic tools are hampering network forensics but also procedures followed during acquisition.

3.3.2 Methods and Procedures

According to KPMG (2009), forensic investigations on computer networks is a task that has to be conducted by well trained personnel because unskilled individuals may end up destroying evidence or conducting investigations in an inappropriate manner rendering any evidence submitted evidentiary weightless. In the field of digital forensic investigations many models that claim to be standardised have been used in different digital forensic scenarios (Sansurooah, 2009); however, the models do not provide a comprehensive procedure for network log mining. Ieong (2006) asserts that there are a number of digital forensic models that have been put forward; nonetheless, there are many discrepancies among them. Thus, the models available have many differences, which usually lead to some dispute among forensic practitioners (Saboohi, 2008).

Leigland and Krings (2004) indicate that most of the existing models that claim to be standard were created ad hoc and therefore do not meet the standards required in a specific digital forensic field or scenario for example in e-mails, databases and network log mining. Thus, forensic models now available are generic hence, they lead to several

disparities (Saboohi, 2008) and confusion often rises on deciding which procedure is best and up to standard to be used in a specific forensic field (Kohn *et al*, 2006). Saboohi (2008) also notes that there are emerging issues of privacy of retrieved evidence and civil liberties (rights and freedom) of practitioners and the prosecuted individuals not addressed in most of existing forensic models. This slows down the process of litigation and prosecution of cyber-criminals in a court of law.

The discussion reveals that there is not yet a standardised model that can be used for network log mining. This is where this research project is positioned; therefore, the investigation into this aspect will contribute to a comprehensive model for forensic acquisition that will be developed as a standardised forensic model to use when salvaging log files in computer networks. Another significant problem which is a counterpart to the lack of a standardised model for network log mining, involves challenges associated with the collection or seizure of evidence from log files of computer network devices.

3.3.3 Acquisition of Digital Evidence

Saboohi (2008) says that the best line of defence is to make use of right policies, communications, procedures and relevant experts. Challenges associated with the acquisition of digital evidence include lack of expertise, IP spoofing, large and complex technologies, different computer installations, and slurred images.

Lack of expertise

One of the major problems in computer network investigations, especially regarding the issue of retrieving evidence from log files, is a shortage of practitioners handling the investigations (Saboohi 2008). Taylor, Haggerty and Gresty (2009) supported this idea as they noted that numerous incompetent and unqualified practitioners have failed to convict criminals in a court of law, leading to organisations failing to trust forensic investigations. Taylor *et al*. (2009) agree with KPMG (2009) that most people, who claim to be forensic experts, do not know the practical ways and are not well-trained without proper qualifications for the task of collecting digital evidence from computer network devices. As a consequence, this usually leads to the manipulation of the evidence. Once this happens, conviction of criminals in a court of law is likely to fail. The problem of lack of expertise is aggravated by the issue of IP spoofing.

IP spoofing

Additionally, criminals have resorted to widespread Internet Protocol (IP) Address spoofing in response to the efficiency of network log mining. IP spoofing denotes generating IP packets with false source IP addresses in order to impersonate other systems or to protect the identity of the sender (Velasco, 2000). Even though forensic practitioners have complete bit for bit log file images or copies of the log files, if the source IP address is spoofed then the evidence is of no practical value (Lessing & Von Solms, 2008). In this scenario, the practitioners have to use unique and rare skills to recover the actual IP addresses (Thomas, 2001). Since practitioners cannot always rely on a suspect's cooperation in supplying a true IP address, the method of tracing digital footprints of cyber-criminals should be adjusted. Relative to the problem of IP spoofing is large and complex technologies as an alternative challenge.

Large and complex technologies

To comply with traditional forensic requirements, all data collected from the log files of computer network devices must be gathered and examined for evidence; however, modern computers consist of huge log files which can record millions of log entries per day (Munk *et al*, 2010). Therefore, these huge and complex technologies, coupled with cyber-crimes becoming more advanced, lead to more complicated and time-consuming digital investigations in computer networks. According to Fei (2007), it is increasingly difficult to use modern tools to locate relevant log entries within the massive volumes of log entries in network devices. Thus, log files are increasing in size and features, complicating a cyber-forensic investigation even further and in some cases, practitioners gather worthless data that takes unnecessary time, while storing takes up valuable spaces (Fei, 2007). This study then insinuates that the increase in intricacy of modern log files in network devices postulates a challenge to the practitioners.

Different computer installations

Despite the challenge of large and complex technologies, Lessing and Von Solms (2008) noted that some problems in the collection of evidence from network devices are due to differences in computer installations. Although there are many common components and aspects, administrators of many business organisations compile their systems to their own specifications. Therefore, it is the practitioner's task to make sure that enough experience

of a wide variety of hardware, software and operating systems is available because the practitioner is likely to encounter any combination of these components, and must be ready to carry out the tasks (Lessing & Von Solms, 2008). In addition, as a result of many possibilities from forensic analysis, the practitioners have to learn the rules of digital evidence acquisition and the effect that specific actions may have on the validity of the evidence (Jones, 2007b). Therefore, deriving from Jones (2007b) this project infers that it is up to the interpretation of the practitioner to analyse the situation and apply the forensic rules in such a way that actions can be justified in a court of law.

Slurred Images

Another problem facing the acquisition of digital evidence from computer networks is the one of slurred images (Jones, 2007b). Similar to when someone takes a photo of a moving object as shown in Fig 3.1; slurred images are the result of copying log entries while some new entries are being added to the log files (Jones, 2007b). This usually results in unclear text readings of log entries which generally are construed as manipulated or false evidence in a court of law. The issue of slurred images can be linked to the dispute of limited amount of evidence gathered in some instances of network log mining. If log entries are somewhat slurred, Ryan and Shpantzer (2005) suggest that such evidence does not provide a complete representation of the original system, and can be clarified as possible data corruption.

Fig 3.1: The difference between proper and slurred images: (Lessing, 2008)

In general, this section revealed that there are not yet proper and reliable conditions as well as many challenges associated with the retrieval of log entries from log files of

computer network devices for the purpose of tracing digital footprints of cyber-criminals for court cases. This increases the risk of evidence rejection in a court of law. The problems associated with acquisition of evidence from log files are exacerbated by the issue of laws and regulations binding the activities conducted during the investigation.

3.3.4 Legislation and Regulations

For digital evidence to be permissible in a court of law, the whole process of seizing evidence from computer network devices until the presentation of the evidence in a court of law must be done according to the laws and regulations of the residing country as well as international laws (Ryan & Shpantzer, 2005). This has been revealed in chapter 2. Jones (2007b) says one of the major problems confronting the field of forensic investigation is that technology and the law do not seem to be in cooperation; instead, they grapple to find common ground. Technology is improving very fast, almost on a daily basis, while laws remain static, or take time to update; as a result digital forensic practitioners remain restricted by antiquated laws (Jones, 2007b). Hailey (2010) supports this view as he says that the legal foundation, which evolves slowly, restricts digital forensic investiagtions; thus, new ideas which evolve daily in the technology field, and which are not in the judiciary system, are neither relevant nor applicable in some aspects of network forensics. Perhaps, new technical developments in the digital forensic investigation field must be updated in the regulatory system. This would ensure that the constant use of old technics is replaced by new ones, avoiding questions in court by criminals who are usually ahead in technology.

Dagada, Eloff and Venter (2009) are of the view that the differences in policy approaches amongst the key role players and countries make it difficult to provide a better information security policy and regulatory model. Even though this challenge materialises at macro level, it manifests itself also at the organisational (micro) level (Buys Incorporation Attorney (BIA), 2009). Whilst these can be regarded as generic policy aspects, the South African legal framework on e-commerce and information security still has some problems because many documents about legal and policy aspects regarding digital forensics in the South African context, do not explain how these aspects are used in the provision of computer forensics in the South African corporate environment (BIA, 2009). The computer forensic related legislation and regulations at macro level as well as

organisations policies are not used absolutely by South African organisations as a way of protecting their information assets from cyber-criminals (Dagada *et al*, 2009).

Moreover, Grobler and Louwrens (2006) point out that, legal and ethical dimensions of digital forensics are very important in business organizations; however, in cyber-space there is no universal or common 'cyber law'. Cyber-crime is not limited to national borders while judiciary systems vary from country to country (Kohn *et al*, 2006); the practitioners have to be familiar with local and international laws, treaty requirements and industry specific legal requirements when preparing for network log mining for court cases produces evidence that will stand up to legal scrutiny in a court of law (Grobler & Louwrens, 2006). However, in some cases forensic practitioners tend to focus on international laws under the assumption that they supercede presiding laws yet in the first instance provide a strong foundation for the case (Hailey, 2010). The issue of laws and regulations with respect to network log mining as well as case studies have been dealt with in chapter 2.

From what has been examined, this study concludes that there is still dispute about how presiding as well as international legislation and regulations impact on network forensics and the forensic field in general. Also, some countries are advanced in technology while others are behind; therefore, it is difficult to have general cyber-laws which govern network forensics because the judiciary of some countries involve recent and advanced technology while in others it takes time to update. Nonetheless, even if the presiding as well as international laws and regulations have been followed throughout the process of network log mining, the evidence must be presented in a court of law to secure a conviction.

3.3.5 Presentation

Advances in technology have led to the use of various digital techniques in the presentation of evidence in the courts (Hailey, 2010). In some cases, digital techniques have allowed the court to gain more valuable information from digital evidence than would otherwise have been possible (Hak, 2003). Basically, digital techniques allow the court to receive evidence previously not possible except with the assistance of digital technology, for example audio enhancement, photograph enhancement, forensic video

analysis and the digital enhancement of latent fingerprints. Even though the technology assists a great deal in forensic investigations, the way the evidence is presented in a court of law has an impact on the outcome (Nikkle, 2006).

According to Ryan and Shpantzer (2005), some people who witness the court case presentations take time to understand forensic ideas and views, and therefore the practitioner has to educate the audience first; otherwise it may not make sense. Thus, the practitioner has to present in simple and clear (layman) language. Hailey (2010) stresses that by only providing copied log file entries as evidence will not convince the judges; the judges need to hear facts drawn from the log entries, in order to have a clear understanding of the case. Additionally, Grabosky, Smith and Dempsey (2001) point out that presentation of digital evidence determined from log files in a court of law can be difficult as a result of exclusion through the use of hearsay and best evidence rules to electronic documents where evidence may not be attainable without understanding the facts behind the case. Furthermore, Perumal (2009) notes that some forensic practitioners are technically competent, but their frailty is on presenting their findings in a convincing way. Therefore, this research project suggests that presentation skills are important in network log mining; hence, practitioners must be trained on how to present facts derived from log files in court where a convincing explanation is needed, and where the judges can question the evidence.

Considering all the problems discussed in this section, network log mining is still a challenge in the present world. Without addressing these challenges, the network log mining process will not be a smooth and easy task to carry out. Therefore, there is a need for improvement in forensic tools to suit the demands of today; standardised models need to be developed for specific fields or scenarios of forensic investigation, for example network log mining and e-mail forensics; practitioners need to be trained in how to professionally acquire digital evidence from digital sources while observing binding laws and also how to competently present the evidence in a court of law. Without these amendments, it will remain an almost insurmountable challenge for practitioners to prosecute and convict cyber-criminals.

Even though the challenges discussed confronting practitioners in network log mining can be addressed, there are some international conventions specifically in the field of IT which practitioners must comply with. As this research project aims to develop a Log File Digital Forensic Model to be used in organisations when computer networks are compromised by cyber-criminals, international IT Security standards and best practices have to be acknowledged to provide credibility to the analysis. The next section discusses more on the IT Security standards that can be applied in network log mining.

3.4 IT Standards and Network Investigations

From the attributes of cyber-crime discussed in section 3.2, it should be noted that cyber-crime is not limited to national borders. Also, taking into account the challenges associated with computer network investigations, the process of network log mining has to be performed in line with international protocols. International protocols can enhance admissibility of the evidence in a court of law anywhere in the world. Table 3.1 shows a comparison of some of the international IT Security standards on different areas of organisations.

Accordingly, well known Information Technology Security Standards include ISO/IEC 27002 (formally known as ISO17799), the SAS70 Type 11, GLBA, PCI DSS, EU Privacy, COBIT 4.1, Common Criteria, GAPP and the GASP. All these IT Security standards can be applied in network investigations (IT Service Strategy, 2008). However, this research project will follow the ISO/IEC 27002 because, as illustrated in Table 3.1, ISO/IEC 27002 is an IT Security standard which has the capability of improving information security when applied in every area of business operation. In Table 3.1, the blank space shows that the IT security standard is not applicable in that area of business operation and one meaning that IT security is applicable. Therefore, this study suggests that ISO/IEC 27002 can be equally nominated as an IT Security standard to apply in network log mining in business organisations.

Table 3.1: IT Security Standards: (IT Service Strategy, 2008)

Enterprise Area	ISO 27002	SAS70 TypeII	GLBA	PCI DSS	EU Privacy	Cobit	Common Criteria	GAPP	GASP
Access control	1	1	1	1	1	1	1	1	1
Application development	1	1			1	1	1		1
Assert management	1	1		1	1				1
Business operations	1	1		1	1	1	1	1	
Comunication	1	1	1	1	1	1	1	1	
Compliance	1	1	1	1	1	1			
Corporate governance	1				1	1			
Customers	1	1	1	1	1	1		1	1
Incident management	1	1	1	1	1	1	1	1	1
IT operations	1	1	1	1	1	1	1	1	1
Outsourcing	1	1		1	1	1	1	1	1
Physical environment	1	1					1		1
Policies & procedures	1	1		1	1	1	1	1	1
Privacy	1	1	1	1	1			1	
Security	1	1		1	1	1	1		1
Score	**15**	**14**	**7**	**12**	**14**	**12**	**10**	**9**	**11**
Percentage	**100%**	**93%**	**47%**	**80%**	**93%**	**80%**	**67%**	**60%**	**73%**

3.4.1 ISO/IEC 27002

The ISO/IEC 27002 was established as a Department of Trade and Industry Code of Practice (DTI CoP) for information security in the United Kingdom in the late 1980s (Hamster, 2006; IT Service Strategy, 2008). After that, the DTI CoP was republished in 1995 by the British Standards Institute (BSI) as the BS7799-1 and later on was revised again in 1999 to significantly strengthen the standard by the International Organisation for Standardization (ISO) and then published as the ISO 17799 (Hamster, 2006). Afterwards, the standard was republished again in 2005 as the ISO/IEC 27002 and this is the name used today (Hamster, 2006). Other versions which contain identical information were published by the South African Standards (SANS) (2006) and Opsecure (2010).

According to SANS (2006), the ISO/IEC 27002 is one of the IT Security standards widely used, recognised and accepted by industry experts. The ISO/IEC 27002 is a security standard that can be used by organizations that need to establish a comprehensive

information security management program or improve its current information security practices (Hamster, 2006). The standard details a comprehensive set of information security control objectives and a best-practice control (Opsecure, 2010). Thus, the ISO/IEC 27002 standard derives security requirements based on assessing risks to the organisation and by adhering to legal, statutory, regulatory and contractual requirements and a particular set of principles, objectives and business requirements (Opsecure, 2010). Therefore, the ISO/IEC 27002 establishes guidelines and general principles for initiating, implementing, maintaining and improving information security management in an organisation (SANS, 2006).

Network log mining is a subsidiary of information security management in business organisations (Forte, 2009), where evidence from log files is gathered for court cases as an information security practice. Therefore, referring to the IT Service Strategy (2008), network investigations have to be conducted following the security control clauses stated in the standard, for example, practising network log mining following well defined policies and procedures; conducting the investigation as a business operation under asset management; and keeping security on the privacy of the evidence to avoid manipulation. The ISO/IEC 27002 suggests that a particular set of established principles stated by laws and guidelines should be observed (Opsecure, 2010). Consequently, this coincides with the claim by Ryan and Shpantzer (2005) that even though the evidence might be valid but extracted without following legislation principles and best practices, the digital evidence may be rejected in a court of law.

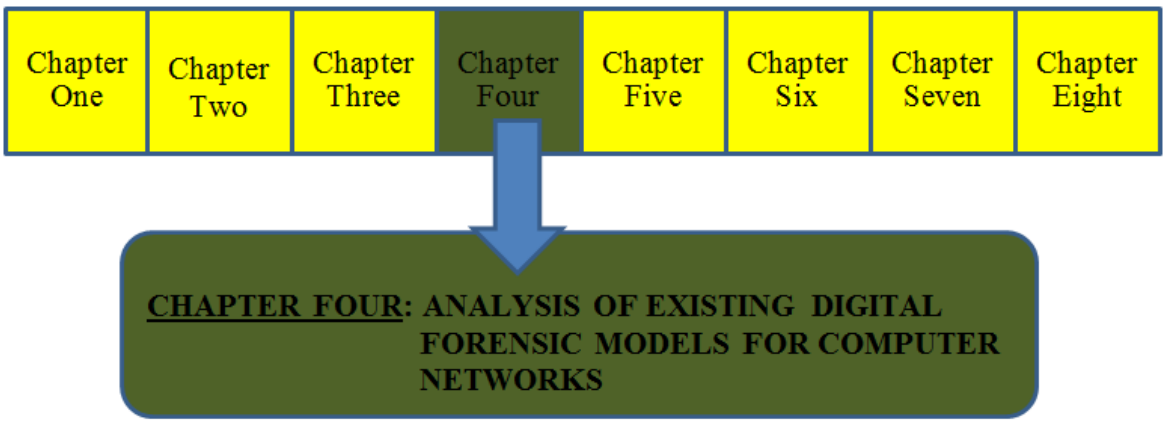
Furthermore, sections 10.5.1 and 10.6.1 of the ISO/IEC 27002 support the process of network forensics with log files as a centre of attention. According to Custer and Vogel (2010) "*section 10.5.1 of the ISO/IEC 27002 says computer networks should be appropriately managed and controlled in order to be protected from threats and to maintain security for the systems and applications using the network through appropriate logging and monitoring of network activities including security-related actions*", p1. Additionally, section 10.6.1 of the ISO/IEC 27002 emphasises that security features, service levels and management requirements for all network services should be identified in reasonable detail and included in a network service agreement, allowing procedures and processes to control/restrict network access and technologies such as digital forensic

investigations and connection controls that can be applied for security of network services (Custer & Vogel, 2010). Therefore, this study asserts that the ISO/IEC 27002 is a suitable IT Security standard to adhere to in network log mining.

3.5 Conclusion

This chapter discussed cyber-crime, focusing on characteristics and features behind cyber-crime. Cyber-crime, which is not limited to national borders, makes tracing the offender difficult. Subsequently, challenges confronted by practitioners in network log mining were discussed in-depth, providing the answer to the first sub-question of the main question of this research project. Challenges fall into five categories namely: forensic tools, methods and procedures, acquisition of evidence, laws and regulations and presentation. Without addressing these challenges, convicting cyber-criminals in a court of law will remain a significant challenge. Finally, the chapter discussed IT Security standards which can be followed in network forensics. Among the mentioned IT Security standards, the ISO/IEC 27002, which is proficient in all areas of operation in business organisations, is emphasised in this research project.

So far the importance of network log mining, the laws and regulations behind network forensics, evidence in a court of law, cyber-crime as well as its characteristics and the challenges faced in network log mining, have been discussed. What needs to be addressed and analysed now are the procedures or models that have been used in network forensics; some of these models are discussed in the next chapter.



4.1 Introduction

A digital forensic investigation is a specialised investigation where the usefulness of the results obtained depends on the procedures, models and techniques used (Selamat, Yusof & Sahib, 2008). Chapter 3 revealed the challenges which forensic practitioners face in the process of forensic investigations on computer networks. One of these challenges is the absence of a standardised model that can be used by forensic practitioners in specific forensic scenarios, for instance network log mining. According to Kohn *et al*. (2006), lack of standardised procedures in the forensic field is one of the main causes for digital forensic evidence collected from digital sources being disputed in a court of law. Therefore, there is a need for a comprehensive model, consisting of several important steps, which have to be used when practitioners retrieve digital evidence from log files of network devices. This chapter examines, compares and contrasts existing digital forensic models identifying commonalities that exist. Also, this chapter will highlight the discrepancies among the existing forensic models.

4.2 Commonalities in Forensic Models for Computer Networks

A number of digital forensic models were proposed in a bid to develop a standardised forensic model. In this chapter, forensic models that can be applied in computer network forensics focusing on evidence from log files are discussed. However, most of the existing models are very broad. One of the forensic models which marked the foundation of most digital forensic models was proposed by Pollitt (1995). However, from this model some individuals began to develop additional forensic models for digital forensic investigations in an attempt to increase security of information assets in people's lives as well as organisations. This study asserts that digital forensic models that can be used towards log

file analysis in computer networks can be grouped into three categories: those models developed by researchers and software engineers; security force based forensic models; and forensic models with laws and regulations as a strong foundation of quality forensic investigation.

4.2.1 Academia and Software Engineering based Forensic Models

In the field of digital forensic investigations, there have been many digital forensic models developed worldwide (Ieong, 2006); and KPMG (2009) suggests that each organisation tends to develop its own model. Some focused on the technical qualities of data collection while others focused on the data analysis portion of the investigation (Ieong, 2006). According to Palmer (2001), the definition of digital forensic investigation was established at a conference by the Digital Forensic Research Working Group (DFRWG). Other definitions for digital forensics have been proposed over the years; however, it is not possible to discuss all the definitions in this research project. According to Reith *et al*. (2002), the unique feature of the DFRWG's model is that it was the first model created by academia (researchers). The steps involved are as depicted in Fig 4.1. Kohn *et al*. (2006) are of the view that DFRWG's digital forensic model was the first generic model which could be used in investigations involving all digital systems; thus, it can also be used when tracing the digital footprints of cyber-criminals recorded in log files in computer networks.

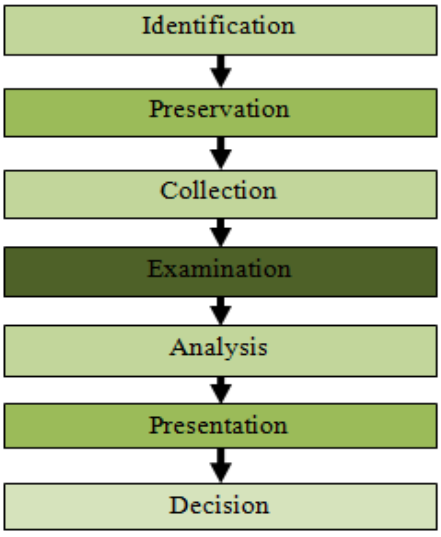


Fig 4.1: DFRWG's Forensic Model: (Palmer, 2001)

The DFRWG's model inspired many researchers and software developers and according to Kohn *et al*. (2006) several models were drawn and are still drawn from this model.

Kruse and Heiser (2001, cited in Kohn *et al*, 2006) summarised the DFRWG's model into a new model commonly known as the three '**As**' which represents Acquire, Authenticate and Analyse. The Acquiring phase involves the processes of Collection, Identification, Storage and Documentaion of digital evidence. The Authenticate phase matches with Examination, while the processes of Analysis, Presentation and Decision were condensed to form the Analyses phase. Therefore, the model proposed by Kruse and Heiser (2001, cited in Kohn *et al*, 2006) inherits many commonalities from the DFRWG's model.

In addition, Reith *et al*. (2002) also drew a forensic model from the DFRWG's model, hence sharing several features. Reith *et al*. (2002) added some components not in the DFRWG's model making it somewhat more comprehensive. The forensic model by Reith *et al*. (2002) is illustrated in Fig 4.2.

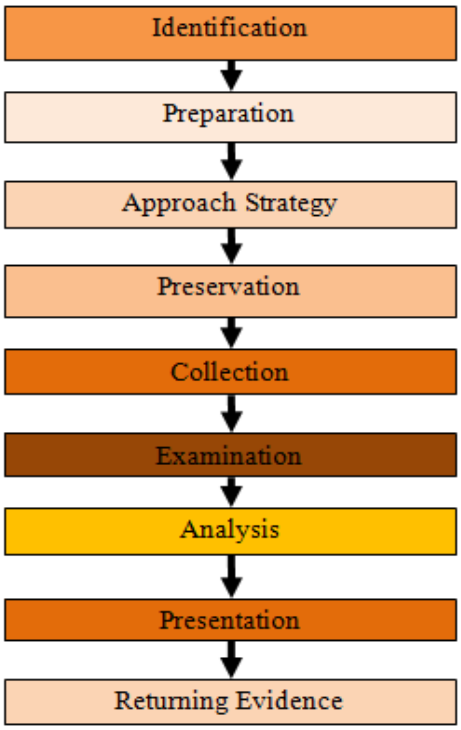


Fig 4.2: The Abstract Digital Forensic Model: (Reith *et al*, 2002)

By comparing Fig 4.1 and 4.2 the additional features are Approach Strategy, Preparation and Returning Evidence phases. Approach Strategy is the process of planning the way the investigations are to be conducted. Preparation of tools and the dynamic formulation of investigative approaches fall under Preparation phase (Ray & Bradford, 2007). The Returning Evidence phase is the process of analysing the final results of the investigation

and discovering the areas which need to be improved in the whole process of digital forensics (Selamat *et al*, 2008).

From the DFRWG's model Carrier and Spafford (2003) developed a forensic model called the Integrated Digital Investigation Process (IDIP). This model resembles the DFRWG's forensic model; however, the IDIP has some new ideas which make the forensic investigation more comprehensive. The model, as illustrated in Fig 4.3, organises the activities into 5 groups with 17 phases (Carrier & Spafford, 2003). The groups are divided into Readiness phase, Deployment phase, Physical Crime Investigation phase, Digital Crime Investigation phase and Review phase. The phases include Pre-incident Preparation, Detection of Incident, Initial Response, Response Strategy, Duplication, Investigation, Secure Measure Implimentation, Network Monitoring, Recovery, Reporting and Follow up. This model was refered to as the most inclusive to date and more complex by Ciarduain (2004) and Perumal (2009).

Unlike the DFRWG's model, Carrier and Spafford (2003) paid more attention to networks and included the step 'Network monitoring'. However, it is not clear as to whether they were discussing mobile networks, computer networks or database networks among others and principally, Carrier and Spafford (2003) did not clearly demonstrate whether they focused on log file examination or other types of digital evidence. Thus, as the model was adopted from a generic model and although focusing on networks, it also has a broad spectrum.

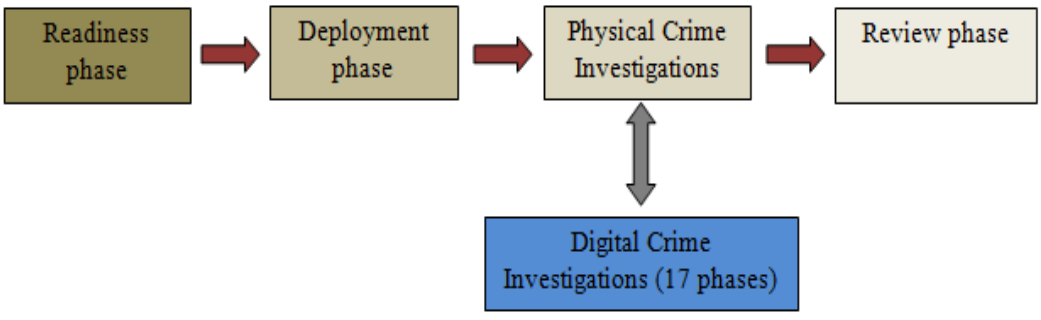


Fig 4.3: Integrated Digital Investigation Model: (Carrier & Spafford, 2003)

The model by Carrier and Spafford was further expanded by Baryamureeba and Tushabe (2004) and they named it the Enhanced Integrated Digital Investigation Process (EIDIP) as shown in Fig 4.4. Baryamureeba and Tushabe (2004) separated the primary (focus on the

computer) and secondary (focus on the physical) crime scenes and depicts the process as an iterative process as opposed to a linear process. As the EIDIP model was derived from the IDIP, it carries some elements from the DFRWG's forensic model and all other models derived from it. The deployment phase of the IDIP model was expanded to include the Physical and Digital crime scene and a new phase was added to Traceback the computer used to commit the offence (Baryamureeba & Tushabe, 2004). Additionally, this model proposes that the reconstruction process under Review phase occurs after all investigations are completed in order to avoid any inconsistencies in future investigations.

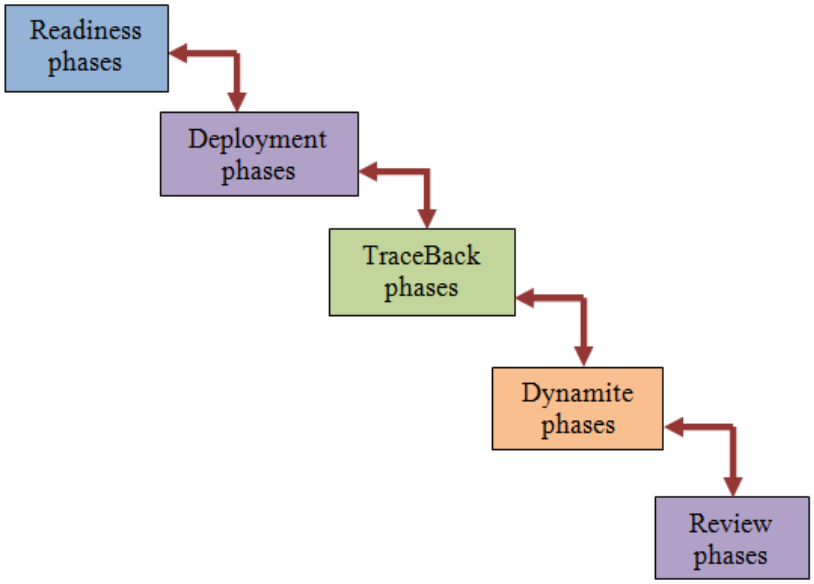


Fig 4.4: Enhanced Digital Investigation Process: (Baryamureeba & Tushabe, 2004)

However, by following this trend, this study concludes that there are many digital forensic models which have their roots in the model proposed by the DFRWG since these models have many commonalities in the naming and sequence of the phases as shown in Table 4.1. Thus, most of the existing digital forensic models created from the perspective of academia and software engineers were created as an improvement of the DFRWG's model. In other models, only the names of the phases or steps differ from those used in the DFRWG's model but identifying the same process for example, in the model developed by Ciardhuáin (2004) the process 'Proof and Defence' is just the same as 'Presentation'. Also, in some cases a few phases in the DFRWG's model were condensed to form one process, or one process from the DFRWG's model was divided to form two or more processes; for example, the process 'Investigation' in a model by Kohn *et al*. (2006) condensed the Identification, Preservation, Collection, Examination, Analysis and

Decision processes from the DFRWG's model. On the other hand, the 'Collection' process in the DFRWG's model was divided into 'Recovery' and 'Investigation' by Carrier and Spafford (2003). Table 4.1 provides a summary of some of the forensic models developed from the perspective of academia and software engineers mapped to the model developed by Digital Forensic Research Work Group.

Table 4.1: Commonalities in widely used forensic models mapped to the DFRWG's model (Own compilation)

Classification Model	Commonalities	Improvements
Pollitt (1995)	Identification; Evaluation; Acquisition (Collection)	Admission
Kruse & Heiser (2001, cited in Kohn *et al*, 2006)	Acquire (Collection); Analyse	Authenticate
Reith *et al.* (2002)	Identification; Collection; Examination; Analysis; Presentation	Preparation; Approach Strategy; Return Evidence
Carrier & Spafford (2003)	Detection (Identification); Recovery (Collection); Investigation (Collection); Duplication (Examination); Reporting (Presentation); Follow-up & Review (Decision)	Readiness; Deployment; Physical Crime; Pre-incident Preparation; Initial Response; Response Strategy; Secure Measure; Network Monitoring
Baryamureeba & Tushabe (2004)	As per Carrier & Spafford's model	TraceBack; Dynamite
Ciardhuain (2004)	Identification; Collection; Examination; Hypothesis (Analysis); Proof & Defense (Presentation); Dissemination (Decision)	Awareness; Authorisation; Planning; Notification
Beebe & Clark (2004)	Data Collection; Analysis; Presentation; Examination	Preparation; Incident Response; Survey; Extract
Casey (2004a)	Recognition (Identification); Preservation; Collection; Classification (Examination); Comparison & Individualism (Analysis)	Preliminary Consideration; Planning; Reconstruction
Kent *et al* (2006)	Collection; Examination; Analysis; Reporting (Presentation)	
Kohn *et al* (2007)	Investigation (Identification, Preservation, Collection, Examination, Analysis, Decision); Presentation	Preparation
Freiling & Schwittay (2007)	Pre-analysis (Identification, Preservation, Collection, Examination); Analysis; Post Analysis (Decision)	
Perumal (2009)	Identification; Analysis; Reconnaissance (Collection, Preservation); Proof & Defense (Presentation)	Diffusion on Information; Planning

In Table 4.1, under the commonalities column, all the steps/phases in brackets are derived from the DFRWG's model and given new names; otherwise the names are just as they are in the DFRWG's model. Thus, the name in the bracket is the one stipulated on the

DFRWG's model while the one outside is a new name with the same meaning. Referring to Table 4.1, this study notes that Reith *et al*. (2002), Carrier and Spafford (2003), Ciardhuáin (2004), Beebe and Clark (2004) as well as Casey (2004a) made some improvements by adding some new ideas onto the DFRWG's model developing models somewhat more comprehensive and up to date. Commonalities appearing in forensic models developed by academia and software developers as illustrated in Table 4.1, give a clear indication that with progress, existing forensic models were created as an improvement on previous models. Under the improvements column, the steps provided indicate new phases added onto the DFRWG's model, making the model more efficient.

From the models discussed, this study concludes that all models adopted from the forensic model developed by the DFRWG also embraced the gene of non-specificity. While the forensic models from the perspective of academia and software engineers are common and widely used (Reith *et al*, 2002; Caloyannides, 2004), security forces also developed some forensic models that could be used when tracing digital footprints of cyber-criminals in the log files of networks. The next section deliberates on the forensic models from the security forces' point of view and demonstrates whether there are significant differences between the forensic models from the perspective of academia and software engineers and those of the security forces.

4.2.2 Security force based Forensic Models

Security force based forensic models were developed particularly to find out which rivals intrude into security force confidential information systems, seeking to steal ideas and to discover the weaknesses and strengths as well as the secrets of the security forces so as to launch an attack (Bejtlich, 2007). In addition, the security force based forensic models were created to be used when the security force undertakes a digital forensic examination when hired by an organisation to trace cyber-criminals. One of the security force based forensic models was introduced by Farmer and Venema (1999, cited in Reith *et al*, 2002). Reith *et al*. (2002) avow that the steps in this model are Secure and Isolate; Record Scene; Systematic Search; Collect and Package; and Maintain Chain of Custody. From the analysis of existing forensic models in this research project, this model is the first one to mention the processes of Secure and Isolate, Record Scene, and Maintain Chain of Custody which are associated with security forces in conventional crimes. However, this

model was focused specifically on Uniplexed Information and Computing System (UNIX) forensic procedures (Reith *et al*, 2002); thus, unlike models from academia and software engineers that can be applied to any operating system; this model can only be used in UNIX computer network devices.

Dittrich and Brezinski (2000, cited in McCombie & Warren, 2003) fostered the model developed by Farmer and Venema (1999, cited in Reith *et al*, 2002) adding some new ideas, generating a new digital forensic model; hence, the models have much in common. The steps introduced in this model are Formulate Plan; Approach and Secure Crime Scene; Document Crime Scene Layout; Search for the Evidence; Retrieve Evidence; and Process Evidence. One can argue that the processes Secure, Search for evidence, and Collect or Retrieve occur in both models. However, according to McCombie and Warren (2003), the model by Dittrich and Brezinski involves the processes of Formulating Plan, Approach and Documentation which makes it more effectively competent in the process of forensic investigations on computer networks than the later model.

Both of the above models were surpassed by a model quoted by Reith *et al*. (2002) as developed by Mandia and Prosise in 2001. This research study concludes that the model by Mandia and Prosise was drawn from both of the later models because the model encompasses the steps incorporated in both models; however, new ideas were introduced making it a more comprehensive forensic model for computer networks as illustrated in Fig 4.5. According to Reith *et al.* (2002), the model by Mandia and Prosise gives the impression that it is the most recent model created mainly for computer networks as it is intended for Cisco routers, UNIX and NT/2000 computer network devices. This model provides a more detailed procedure than those of Farmer and Venema as well as Dettrich and Brezinski in that it addresses Pre-incident Preparation as an explicit step to professionally organise the investigation prior to technical response.

This research project infers that unlike the forensic models developed by academia and software engineering, security based forensic models have networks as the center of attention; however, the models do not stipulate the actual activities to be carried out on the computer networks during network log mining investigations.

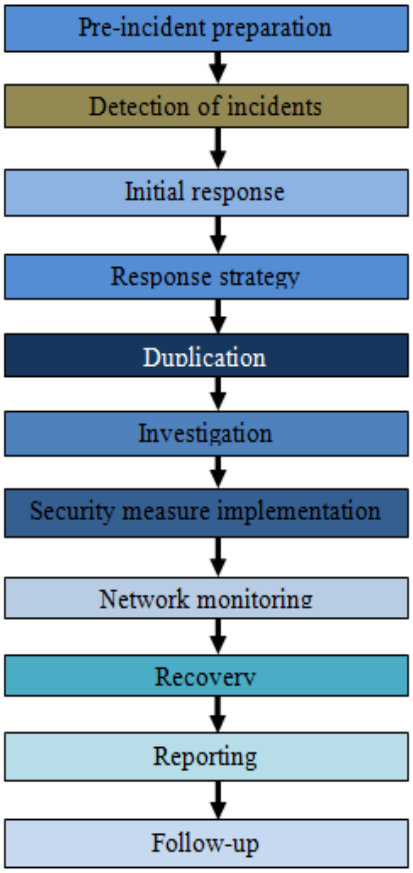


Fig 4.5: Mandia & Prosise Forensic model: (Reith *et al*, 2002)

Considering the models discussed in section 4.2.1 and 4.2.2, some similarities can be found between security force based digital forensic models and those from the perspective of academia and software engineers. The name of the processes differs because of the focus of the models. For example, the phase 'Collection' in the view of academia is just the same as 'Retrive or Recovery' in the security force based forensic models. The major difference is based on the verity that security force based forensic models focus more on computer networks because security forces keep confidential and sometimes secret information in huge databases connected to users through computer networks (Bejtlich, 2007). On the other hand, it has been concluded in section 4.21 that most of the models developed by academia and software engineers are basic and embraced the gene of non-specificity.

Moreover, by closely looking at the examined digital forensic models, one can surmise that Carrier and Spafford (2003) drew some of the phases on their model from the one by Mandia and Prosise (1999, cited in Reith *et al*, 2002). Thus, authors of security force

based forensic models and academia borrowed ideas from one another; therefore, even though terminology and focus of the models is different, commonalities still exist in the models from both perspectives. Nonetheless, Ieong (2006) argues that forensic investigations must be carried out based on laws and regulations. The issue of laws and regulations in forensic investigations is not incorporated into the forensic models of academia and software engineers or the security forces discussed above. Ieong (2006) says conducting forensic investigations without considering legislation and regulations is as good as not conducting the investigations in the first place.

4.2.3 Law based Forensic Models

The last group of digital forensic models that can be utilised in network investigations focusing on log files, are models created with laws and regulations as the basic foundation for quality forensic investigations. Of the models examined, none explicitly state the legal implications of the processes of digital forensics on computer networks although it is essential for the collection and presentation of evidence in a court of law (Cardwell *et al*, 2007). These models have been developed by traditional forensic scientists and this must be addressed as the challenge in the digital forensic discipline. As specified in Chapter 3, technology is improving on a daily basis; therefore, new concepts developed in technology towards digital forensics have to be encompassed in the legislation. Most of the models which incorporated laws and regulations were developed by authors who went on to combine ideas from different perspectives, that is, the academic and engineering perspective as well as the security forces perspective resulting in the development of complex forensic models.

When Kohn *et al*. (2006) introduced the Framework for Digital Forensic Investigation model they mentioned preparation, investigation and presentation as the minimum steps to be undertaken for a forensic investigation to be regarded as complete. Subsequently, Kohn *et al*. (2006) set a legal base as the foundation so that the legal requirements were established right at the start of the forensic investigation and informed each subsequent step or phase as illustrated in Fig 4.6. From the examined digital forensic models in this research project, this is the first digital forensic model to include the aspect of laws and regulations in the process of forensic investigations. However, Ieong (2006) drew a digital

forensic model from the law perspective and examined the relationship between law and digital forensic investigations in detail.

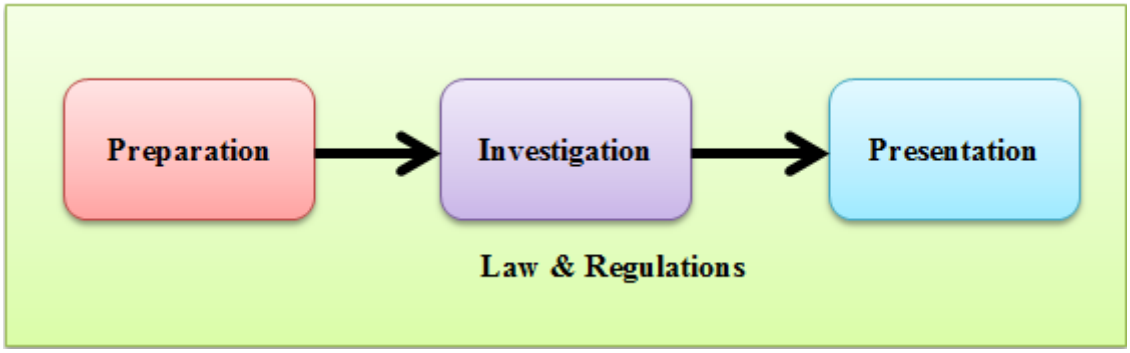


Fig 4.6: Framework for Digital Forensic Investigation model: (Kohn *et al*, 2006)

According to Ieong (2006), digital forensic models developed by technical people, especially researchers and software developers, focus on technological details in capturing digital evidence while forgetting the actual purpose and core concept of digital forensic investigations. Legal practitioners then have difficulty in applying or understanding the processes of forensic investigations on computer networks (Ieong, 2006). Consequently, there is a wide gap between the technical specialist and the legal practitioners. Ieong (2006) states that many lawyers and the judiciary understand that they need to become familiar with digital forensics practices; nonetheless, they consider that the technical procedures and knowledge are difficult for them to learn or even to follow. From this idea, this study concludes that lawyers and the judges find themselves lost in technical aspects of digital forensics without understanding the more advanced principles in digital forensic investigations.

Ieong (2006) developed a model that can be utilised when searching for digital evidence from log files in computer networks. In his FORZA (FORensics ZAchman) model, Ieong (2006) is of a view that fundamental principles of digital forensic investigations are Reconnaissance, Reliability and Relevancy. The FORZA model emphasises laws and regulations rather than the technological aspect. Under the Reconnaissance principle, the forensic practitioners must exhaust different phases in the forensic model employed as well as the tools developed for a particular environment to Collect, Recover, Decode, Discover, Extract, Analyse and Convert data into readable evidence (Ieong, 2006). These processes were referred to (common) in the models by academia and security force based

forensic models. Under the Reliability principle, Ieong (2006) emphasises that the Chain of evidence must be kept and updated from the beginning until the end of the forensic investigation. The Chain of evidence, time and integrity of evidence provide strong non-repudiation features of evidence (Ieong, 2006). The concern of keeping the Chain of evidence was already mentioned by Farmer and Venema (1999, cited in Reith *et al*, 2002), marking another commonality with already examined forensic models. Lastly, in the Relevancy principle, the legal practitioner must advise on what should be collected during the process to reduce time and money spent on the investigation. The FORZA model developed by Ieong (2006) is as illustrated in Fig 4.7.

In addition, Ieong (2006) states that the participants like the system owners, digital forensic investigators, and legal practitioners must be involved as shown in Fig 4.7 and their roles as well as responsibilities should be divided into eight individual roles starting with the case leader. The case leader is the organiser of the whole digital investigation process and leads the case, meeting the system owner/business owner and determining whether the investigation should proceed or not (Ieong, 2006). The case leader will then seek legal advice as to whether to proceed for legal disputes or avoiding a waste of money and time on a case that may fail in a court of law (Ieong, 2006). Subsequently, the case leader then has to interview the system/internal auditor in order to find relevant information about the intrusion of the organisation's computer networks so that he can estimate the scope of the case and extract the security controls design implemented in the computer network system (Ieong, 2006). These duties played of the case leader at this level of the FORZA model are regarded as the 'Preparation' or 'Pre-analysis' phase in other models from academia as well as security forces. This marks a close link or a commonality between what is proposed in the FORZA model to the models already examined.

After acquiring enough information the case leader has to assign or hire a digital forensic specialist to plan the entire operation while using the legal advice to plan the investigation strategy (Ieong, 2006). The digital forensic specialist will provide a defined strategy to the digital forensic investigator. According to Ieong (2006), the digital forensic investigator has to Collect; Extract; Preserve and Store the digital evidence from computer network devices and submits the findings to the digital forensic analyst (see Fig 4.7).

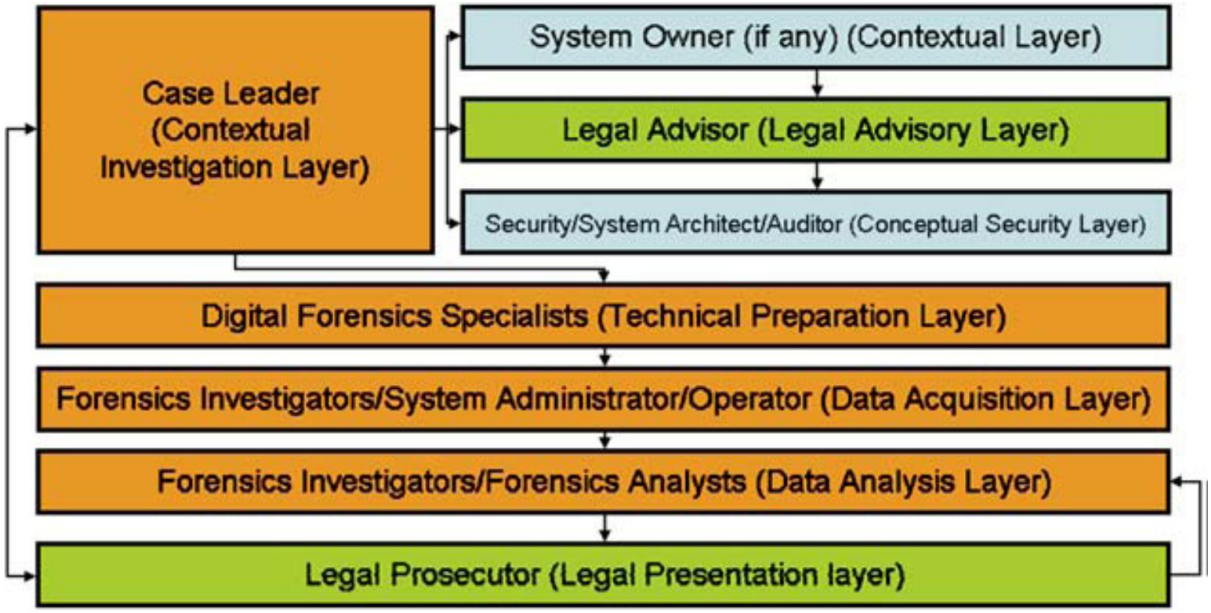


Fig 4.7: Process flow between the roles in digital forensics: (Ieong, 2006)

According to Ieong (2006), the digital forensic analyst then extracts relevant data from the collected evidence and analyses it against the model proposed for investigation. Also, the digital forensic analyst can perform tests to prove or reject the model used and reconstruct the timeline of the case based on the extracted data (Ieong, 2006). Likewise, the roles of the digital forensic specialist, digital forensic investigator and digital forensic analyst at this level of the FORZA model were previously mentioned as Planning, Collect/Extract, Preservation and Analyse in the models designed by academia and software engineers, and security forces marking another commonality between the FORZA model and previously examined forensic models.

Lastly, the prosecutor has to advise the case leader whether the collected evidence is sufficient, relevant, admissible and favourable to the organisation (Ieong, 2006). Also, Ieong (2006) is of the opinion that the prosecutor can suggest to the case leader the most feasible legal system to choose and the prosecutor then leads the case in the litigation process. Similarly, the role of the prosecutor and the case leader at this level of the FORZA model is the same as Presentation and Dissemination phases mentioned in models by academia, software engineers and security forces, marking another common aspect between the FORZA model and already examined forensic models. Therefore, the model by Ieong (2006) emphasises the recognition of laws and regulations at each and every phase of the investigation. However, this study discovered that the FORZA model is not

specific to any field/scenario of digital forensics, that is, it is generic. Through the FORZA model, Ieong (2006) linked together the roles and responsibilities of the participants discussed in Table 4.2. Table 4.2 shows the layers of the FORZA model and the roles and responsibilities of the participants in each layer. Ieong (2006) emphasised that six sets of categories of questions are supposed to be asked during the digital forensic investigations. These questions are in the form of: what (data attributes), why (motivation), how (procedures), who (people), where (location) and when (time).

Table 4.2: FORZA model: (Ieong, 2006)

Layers (Stages)	Description of the participants duties
Contextual Investigations Layer	After receiving a report of compromised digital sources, the case leader determines the motivation of the case, the involved parties, confirms time of incident and verifies the location, finds the nature of event and plans the next procedure.
Contextual Layer	The case leader interviews system owner(s) to understand the nature of the company and the objectives of the affected systems, determines the business and event nature, confirms business and system process model, explores the business geography, determines the business and initial timeline and understands organisation and participants' relationship
Legal Advisory Layer	After obtaining the details of the compromised digital sources, the case leader seeks legal advice to determine legal objectives of the case, legal background and preliminary issues of the case, legal geography and jurisdiction, legal entities and participants and timeframe
Conceptual Security Layer	The case leader then explores further the design of information systems and relevant security control from system owner(s); thus, exploring system control objectives implemented to protect from external attacks, understanding the system information and security control model, collecting the implemented security mechanisms details, exploring the security domain and infrastructure, determining the user and security entity model and determine the security timing and sequencing.
Technical Presentation Layer	The case leader appoints digital specialists to determine and plan the relevant forensic investigation strategy objectives, determine the forensic data model, explore geography location within the forensic data model, draft the forensic entity model, propose a hypothetical forensics event timeline and define the forensic strategy.
Data Acquisition Layer	The system administrator(s) follow the strategies and tasks outlined the forensic specialist, hence understand the forensic acquisition objectives assigned by the forensic specialist, perform on-site forensic data observation, interview participants and identify witness, perform forensic acquisition and seizure procedures, perform site network forensics data acquisition and keep the forensic acquisition timeline and chain of evidence.
Data Analysis Layer	Digital forensic analysts extract relevant information and review it according to the model, reconstruct event data, extract network information, extracting entity accounts information and rebuild the relationship linkage, analyse the extracted data based on forensic analysis procedures and reconstruct the event timeline.
Legal Presentation Layer	After extracting and analysing the information collected from the digital sources, the prosecutor has to discuss with the case leader and the system owner: the legal presentation objectives, attributes, legal presentation procedures, legal jurisdiction location, entities in litigation procedures and timeline of entire event presentation

Another forensic model based on the foundation of laws and regulations which can be applied in network forensics was developed by Lalla and Flowerday (2010). By closely analysing the model by Lalla and Flowerday (2010), one can infer that the model was

derived from the one by Carrier and Spafford (2003) as well as the one by Baryamureeba and Tushabe (2004). However, the model emphasised the aspect of laws and regulations throughout the forensic investigation. The model by Lalla and Flowerday (2010) was developed with e-mails as the center of investigation and the unique aspect of this model is the use of stylometry and data mining in the gathering of evidence to convict criminals. Nevertheless, the model can be applied in computer network investigations because it does not stipulate physical phases which can be depicted as specifically for e-mails forensics; all the phases in the model are the same as those developed by academia and software engineers. Conversely, Lalla and Flowerday (2010) introduced a new important idea, that chain of custody, which is a phase or step in some of the security forces models, is not a phase but an activity which has to be done throughout some phases of the investigation; that is, documenting all the processes conducted during the investigation. In spite of this, this study concludes that the model developed by Lalla and Flowerday (2010) has much in common to those developed by security forces and academia.

Considering all the digital forensic models examined in this section, there are many common attributes that can be engaged in network forensics even though they were developed from different perspectives. Therefore, one can conclude that all forensic models follow a similar methodology. The differences are in terminology and focus depending on one's point of view towards digital forensic investigation processes. Otherwise, the aim of all these forensic models is to provide a standardised method with the archetype following Gathering, Processing, Interpreting and Utilization of digital evidence in a way legally admissible in a court of law.

In this section, the discussion revealed the commonalities existing in most available digital forensic models. However, the next section discusses some of the discrepancies among existing forensic models, supplementing some of the weaknesses established in the reviewed literature with those discovered in this research study.

4.3 Discrepancies on existing Forensic Models

According to Kuhne (2010), all models are imperfect; however, a model must have three important features namely mapping feature, reduction feature and pragmatic feature. This implies that, each and every model has its own weaknesses and strengths, depending on

the perspective from which the user views it. Ray and Bradford (2007) are of the same view and suggest that each model's distinct characteristics give it inherent advantages and disadvantages. Therefore, if one closely examines the existing forensic models shortfalls can always be detected. Reith *et al*. (2002) argue that most digital forensic models have not been tested nor proven to be standardised forensic models; conversely, this contradicts what most authors claim that their models are standardised (Ieong, 2006). In addition, Perumal (2009) asserts that there is no easy or obvious method for testing any proposed digital forensic model.

The models examined attempted to enhance the process of digital forensic investigations by identifying the solidarity of digital technologies and working towards establishing a solid forensic process that applies to many digital technologies (Reith *et al*, 2002). Therefore, most of the proposed forensic models were created without considering presentation of abstract steps that add value to specific forensic scenarios (Reith *et al*, 2002); hence, the phases in those forensic models may be defined as too general. In addition to that, other forensic models are very rigid, meaning that the model can hardly be changed or altered; for example Ray and Bradford (2007) noted that the model developed by the DFRWG is rigid although it is particularly suitable where necessary and investigative activities are well-understood. Perhaps, this is why most of the forensic models developed by academia and software engineers do not really divert from it.

Furthermore, problems associated with existing forensic models involve process redundancies (repetitions), area focus and framework characteristics. For example, the models by Stephenson (2003); Beebe and Clark (2004); as well as one by Freiling and Schwittay (2007) focus on the analysis process. These models do not emphasise other important forensic aspects or phases crucial to investigations. Selamat *et al.* (2008) claim that most existing forensic models do not pay attention to important processes which are supposed to be conducted prior and post technical analysis processes as illustrated in Fig 4.8. In addition, Selamat *et al*. (2008) are of the view that the model by Carrier and Spafford (2003) as well as Roger *et al.* (2006) focused on building a mechanism for quicker forensic examinations; therefore, the models did not address digital forensic procedure in detail.

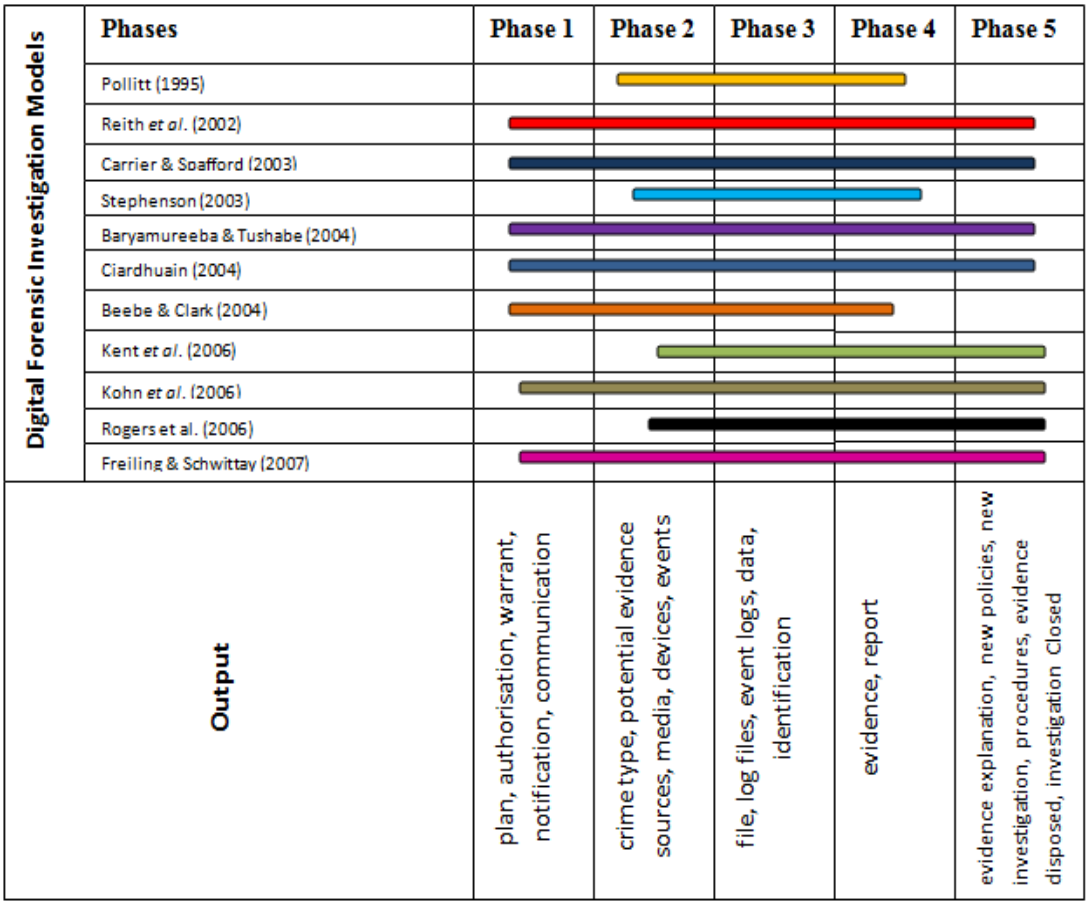


Fig 4.8: Digital Forensic Investigation Framework Map: (Selamat *et al*, 2008)

Another important issue follows that Richard and Roussev (2006) assert that practitioners must ensure that the investigation process can be started and run within a proper procedure keeping the chain of custody throughout the investigation. Nevertheless, most of the existing forensic models examined in this study do not emphasize or include chain of custody in the process except only for the models by Farmer and Venema (1999, cited in Reith *et al*, 2002); Ieong (2006) and the one by Lalla and Flowerday (2010). Furthermore, at the end of the investigation there has to be a process of evaluation and reflection on improvements needed in the investigation procedures used (Selamat *et al*, 2008; Perumal, 2009); unfortunately, most of the digital forensic models examined in this project do not involve this important aspect.

This study alludes that most of the existing forensic models reviewed can be used in forensic investigations focusing purely on personal computers; they do not address the forensic processes in terms of other digital devices such as network devices (routers,

firewalls and switches). Furthermore, most of these models give the impression that they can be used in any forensic investigation because they are broad-based. Yet this study argues that there is a need of forensic models designated for each and every part/scenario of the digital field, for example in the case of this study network log mining.

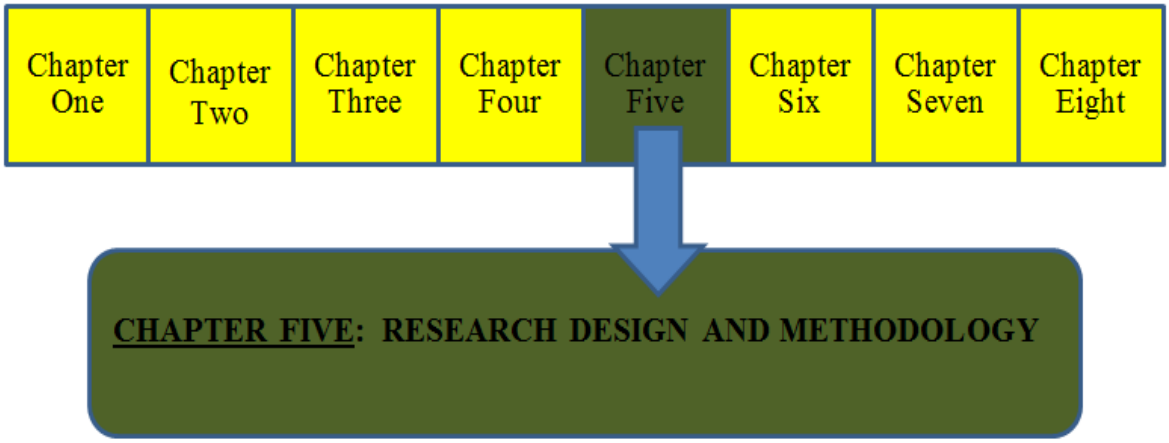
Moreover, most of the existing digital forensic models mention the process of Collection or Recovery of data and the process of Examination but they are not explicit about the methodology of Collecting and Examining digital evidence. Thus, this study argues that the processes of Data recovery and Examination in the existing forensic models do not clearly demonstrate a way that can direct a practitioner on how to conduct network log file investigations. In addition, no model among those reviewed, deliberated on safety measures which can be observed by forensic experts during the forensic investigations bearing in mind that Jones (2007b) claims that it is difficult for a practitioner to finish a forensic investigation without making mistakes.

Therefore, considering all that has been discussed above this study suggests that there is a need for an improvement; that is, developing a model for computer networks, specifically for network log mining.

4.4 Conclusion

This chapter discussed commonalities found in existing forensic models that can be used on computer networks. There have been many digital forensic models developed in an attempt to create a model pertinent to the seizing of digital evidence from digital sources. These models fall into three categories: models developed from the perspective of academia and software engineers; developed from the perspective of security forces; and those with laws and regulation as a foundation of a quality forensic investigation. In this chapter, this research project discovered that although models in each category had some evident commonalities; and some other commonalities emerged in different categories developed from different perspectives. From all the models discussed, the aim of forensic investigations is to Gather, Process, Interpret and Utilize digital evidence in a way that is legally admissible in a court of law.

However, among all the potential forensic models that can be used to trail the digital footprints of cyber-criminals in log files of computer networks discussed in this chapter, there is not yet a model that provides explicit methodology to be followed in network log mining. The discrepancies of existing forensic models have been discussed and general weaknesses when it comes to networks and log files, were revealed. The next chapter will discuss the research design and methodology behind this research project.



5.1 Introduction

This chapter will discuss the research methodology, explaining in detail what has been done towards the creation of the proposed model and its evaluation. According to Myers (1997), a methodology provides a strategy of inquiry that moves from the underlying philosophical assumptions to research data. The process in which this study was conducted is as illustrated in Fig 5.1. In short, towards the development of the proposed model, Design Science (DS) methodology and its research methods were employed. The model was sent to experts for validation, thus, expert reviews. The next section elaborates more on the research design in which this study was projected.

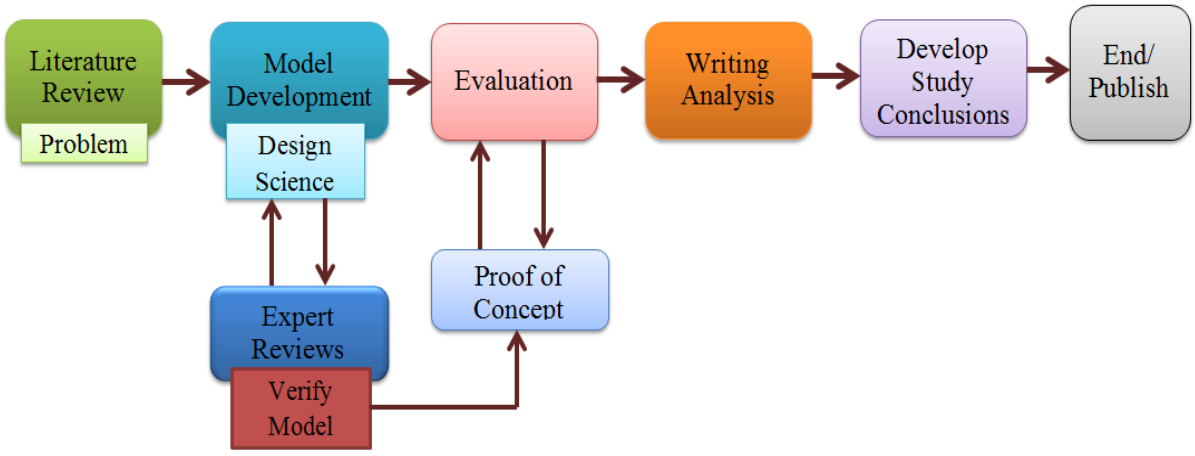


Fig 5.1: Research Process, Adapted from Hevner & March (2003)

5.2 Research Design

A research methodology defines what the research is, how to proceed, how to measure progress and what constitutes success (Pai & Bhattacharya, 2006). Each method changes the manner in which the researcher harvests the data required for the research project.

Myers (1997) asserts that each method relies on different skills, assumptions and research practices. In this research project, an Information Technology (IT) artifact or model entitled ***'A Log File Digital Forensic Model'*** was developed. Research design ensures that the evidence obtained through the research makes it possible to answer the research questions as unambiguously as possible (Abrahams, 2010). Each and every research project follows a philosophical assumption of which Collis and Hussey (2003) suggest can be classified along a continuum of core ontological assumptions. At the extreme ends of the continuum there are positivistic and phenomenological paradigms. Nevertheless, very few research projects would operate at these extremes, but lie in between these two paradigms (Collis & Hussey, 2003).

Rationally, Morgan and Smircich (1980) portray reality as a continuum between the extremes of positivistic and phenomenologist assumptions illustrating the difficulty in cutting a clear distinction between the assumptions. Fig 5.2 shows the continuum of core ontological assumptions from positivistic through to phenomenologist paradigms and postulates the reality where this study falls, that is the "*Reality as a Contextual Field of Information"* category. According to Holden and Lynch (2004), the "*Reality as a Contextual Field of Information*" allows man to be an information processor in order to map context based cybernetic metaphors. Therefore, as this study focuses on proposing a model that can be used when tracing cyber-criminals, it conforms to this reality.

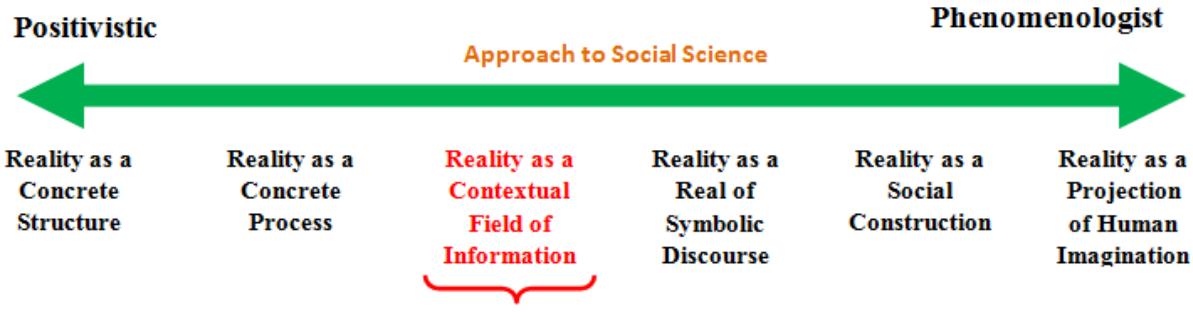


Fig 5.2: Continuum of Core Ontological Assumptions: (Collis & Hussey, 2003)

In the "*Reality as a Contextual Field of Information"*, human beings are engaged in a continual process of interaction and exchange of information, thus, receiving (reading relevant literature), interpreting, and acting on the information received (reviewed literature) and in so doing creating a new pattern of information that influences change in

the field as a whole (Nelson, 2004). Nelson (2004) goes on to say that this field allows for contextual analysis as it is a field of ever-changing form and activity based on the transmission of information. Hence, it is the information which evolves and that an adequate understanding of the process entails grasping the ecological nature of the new context (proposed artifact) as a whole (Boal, Hunt & Jaros, 2004).

Fulfilling the "*Reality as a Contextual Field of Information"* category, the developed model is based on the ideas gathered from the literature. With respect to network log mining, this reality allowed this study to scrutinise the challenges faced in network forensics (chapter 3) as well as forensic models already published (chapter 4). This research project established that there is a need of a forensic model that can be used explicitly in network log mining because most of the available models are nonspecific. However, the "*Reality as a Contextual Field of Information"* allows other individual(s) or researcher(s) to modify the model proposed in this project to develop new information. The next section deliberates on the paradigm employed in the development of the proposed model.

5.3 Design Science

In this research project Design Science methodology and its research methods (data collection and analysis) were followed. Design Science is a research paradigm that attempts to expand human potential by creating ground breaking artifacts (Alturki, Gable & Bandara, 2011); thus, this paradigm supports the development of new digital forensic artifacts such as the one proposed in this study. Hevner and March (2003) proposed a conceptual research framework in order to assist understanding, execution and evaluation of Information Systems research so as to assess each paradigm one against the other in the context of business needs. This paradigm is foundational to the Information Systems (IS) discipline as it is at the confluence of people, business organisations and technology (Wieringa, 2010; Venable 2010). Therefore, referring to the Design Science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the development of new models/artifacts (Kuechler, Park & Vaishnavi, 2009; Sevaldson, 2010). Thus, Design Science can be applied when one intends to improve the effectiveness and efficiency of available digital forensic models to suit a specific task in the field of forensic investigations.

There are two design processes and four design artifacts produced by Design Science in Information Systems. The two design processes are 'build' and 'evaluate' while the artifacts are constructs, models, methods and instantiations (Oates, 2006). Peffers, Tuunanem, Gengler *et al*. (2006) assert that constructs provide the language within which problems and solutions are defined and communicated; models use constructs to represent a real world situation, the design problem and its solution space; methods define processes and provide guidance on how to solve a problem; and instantiations shows that constructs, models and methods can be implemented in a working system as they demonstrate feasibility and enable concrete assessment of the sustainability of an artifact or model towards its intended purpose(s). Artifacts from Design Science extend the boundaries of human problem solving and organisational capabilities by providing intellectual as well as computational tools (Kuechler *et al*, 2009; Alturki *et al*, 2011); hence, Design Science coincides with the Reality as a Contextual Field of Information. Wang and Wang (2010) agree with March and Storey (2008) that the fundamental principle of Design Science research is that, knowledge and understanding of a design problem and its solution are acquired in the building (development) and application of an artifact (model).

Hevner, March, Ram and Park (2004) proposed seven guidelines which were followed in the development of an IT artifact (model) in this research project. The following seven guidelines provides an adaptation of the guidelines introduced by Hevner *et al*. (2004):

1. Through Design Science methodology a conceptual artifact (model) for a specific problem domain was proposed (developed). The absence of a comprehensive model that can be employed specifically for log mining in computer networks was discovered as the problem domain.
2. The model is unique in the sense that it focuses on the area of study (log mining) never before dealt with. This was discovered in the literature review (see chapter 4). The model has a number of steps proposed and added to be followed explitly in network log mining.
3. The model was defined in detail and formally presented (see chapter 6).
4. The roots from which the model was derived were patently discussed and the demonstration of the model granted (see chapter 6).
5. The proposed model was evaluated by IT experts.

6. As part of communication, the opinions of the experts and the outcome of the research were discussed (see chapter 7). Also, the model proposed in this study was published on the Eighth Annual IFIP WG 11.9 International Conference on Digital Forensics in 2012.
7. The proposed model can also be improved by other researchers.

The seven guidelines by Hevner *et al*. (2004) were followed as they provided a universal and clear method of how an IT artifact should be developed. The model proposed in this research project was developed with the expectation of changing the way forensic investigations are conducted during network log mining. The discussion above shows how this study was carried out torwards the development of a proposed model. At this point, there is a need to understand how the model has the potential to be received in the industry. The Diffusion of Innovations (DOI) theory will help explain how the model can be received and adopted in the business.

5.4 The Diffusion of Innovations (DOI) Theory

The DOI theory centres on conditions which increase or decrease the likelihood that a technological idea, product or practice will be adopted in the community (Agarwal, Ahuja, Carter & Gans, 1998). Therefore, the DOI theory helps predict the prospect of acceptance. The DOI theory is a process in which an innovation is communicated through certain channels over a period of time among the members of a social system (Rogers, 1995; Clarke 1999). According to Mettler (2009), innovations either can be categorised as 'product innovations', that is, the engrossment of a new or improved object whose characteristics differ significantly from previous ones; or as 'process innovations', such as the contribution of new or improved practices to enhance production and or delivery of objects or services. The model proposed in this study differs from previous ones in the sense that it endeavours to improve practices by trying to enhance the process of retrieval and analysis of digital evidence from log files in computer networks. As the model proposed is precisely for network log mining, the introduction of this model is designated to improve the practitioners' capabilities.

Agarwal *et al*. (1998) indicate that the DOI theory purports to describe the patterns of adoption, explaining the mechanism and assists in predicting whether and how new

inventions will be successful. It is not always 100% of the targeted group would adopt an innovation; in this case, only few Information Technology Security Experts (ITSEs) may adopt the model proposed in this research project. According to Clarke (1999), in DOI theory the stages of technological innovations are:

i. Knowledge, which is the exposure of the model's existence and understanding of its functions.
ii. Persuasion, which anticipates the forming of a favourable attitude to the model.
iii. Decision, which deliberates the commitment to the adoption of the model.
iv. Implementation, which emphasises putting the model into use; and lastly
v. Confirmation, which provides reinforcement based on positive outcomes of the model.

Rogers (1995) discovered that individuals possess different degrees of willingness to adopt innovations and thus it is generally observed that the portion of the population adopting an innovation is approximately normally distributed over time. Mettler (2009) agrees with Robinson (2009) that breaking this normal distribution into segments leads to the segregation of individuals into five categories: innovators, early adopters, early majority, late majority, and laggards. When the adoption curve is converted to a cumulative percent curve, an **S** curve is generated that represents the rate of adoption of the innovation within the population (Rogers, 1995); this cumulative curve is illustrated in Fig 5.3.

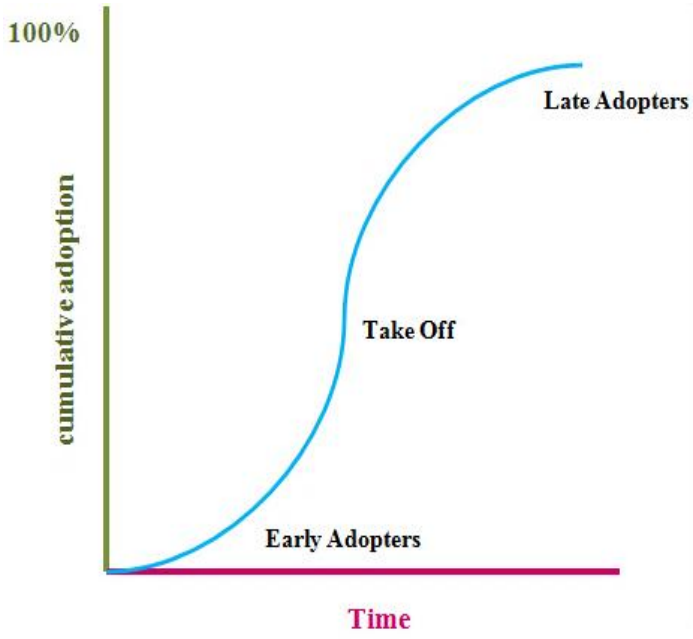


Fig 5.3: Diffusion of Innovations Theory: (Rogers, 1995)

Fig 5.3 shows that very few people (early adopters) adopt the proposed model as soon as it is introduced into the system. However, with time when the model has been used by a few individuals and proved workable, many people may adopt it transiently as depicted by the take off period in Fig 5.3. Nevertheless, there are those individuals (late adopters) who will adopt the model only after it has proved its worth and been used by many in the forensic industry. The rate of adoption of innovations is impacted by five factors namely: relative advantage, compatibility, trialability, observability and complexity (Furneaux, 2010). Furneaux (2010) suggests that the first four factors are generally positively correlated with rate of adoption while the last factor, 'complexity', is generally negatively correlated with the rate of adoption.

Therefore, this research project believes that its product, ***A Log File Digital Forensic Model***, can be adopted in the community as described through the DOI theory. At this point, the research project illustrates how the project was undertaken from the beginning to the development of the model through the discussion of Design Science methodology. Also, it describes the theory (DOI) which will facilitate acknowledgement or reception of the model in the community. The next section deliberates on the investigations which support the development of the proposed model.

5.5 Investigations

As mentioned in section 2.3.2, this research project mentioned three South African presiding legislations as well as some international legislations that scaffold network log mining and digital forensics in general. The mentioned South African legislations are the Computer Evidence Act, 57 of 1983; the Electronic Communications and Transactions (ECT) Act, 25 of 2002; and the Electronic Communications Act, 36 of 2005. However, this section does not discuss these legislations, but deliberates on the supposed drawbacks associated with the legislation. This research project argues that there are some important aspects not emphasised or fully addressed by the South African presiding legislation which supports digital forensic investigations. For example, the legislations do not mention:

i. The measures to consider, to avoid manipulation of the evidence by practitioners.
ii. What has to be done by the individual (practitioner) responsible for the manipulation of the evidence.

iii. How the legislations corroborates with International Best Practices such as ISO/IEC 27002 and COBIT 4.1.

iv. The issue of unqualified practitioners and complicated investigations.

Subsequently, this research project proposed some resolutions to these drawbacks called ***"Forensic Investigation Precautions"***. These precautions are discussed in detail in chapter 6 as part of the discussion on the architecture of the proposed model. The next section discusses how information concerning this research project was obtained.

5.6 Data Collection

According to Egger and Carpi (2008), data collection is a systematic recording of information as a process of gathering and measuring information on variables of interest, in an established fashion that enables one to answer stated research questions, test hypotheses and to evaluate outcomes. Jiaqi, Shunxin, Ming and Jianchuan (2005) assert that a well designated study must be supported by both primary and secondary data. Initially, primary data collection is discussed finishing with the secondary data.

5.6.1 Primary Data

Primary data collection was conducted through the validation of the proposed model. This was conducted through expert reviews. On the validation stage, the model, in a form of a paper, was dispatched to the Information Technology Security Experts (ITSEs) through e-mails. Most of the experts interviewed were local as well as international professors and authors of most referenced papers concerning digital forensics, log mining and IT auditing. Also, some of the experts were individuals working in forensic and/or IT auditing departments of some accounting firms in South Africa. In total, fourteen experts responded; thus, three experts in the first round, five experts in the second round and three experts in the third as well as the fourth round. In each round, the experts were given four weeks to respond. Lastly, the appraisal of the experts review was conducted as the fifth round. Nevertheless, before the model was sent to the experts for validation, a pilot study was conducted; the following section discusses the pilot study conducted prior to the primary data collection.

5.6.1.1 Pilot Study

To ensure that the questions attached to the model provided the highest quality responses, a select number of respondents conducted a pilot study. A pilot study is a small experiment designed to test logistics and gather information prior to a larger research project, in order to improve the quality and efficiency of the study (Altman, Burton, Cuthill, Festing *et al*, 2006). The pilot study took place two weeks before the model was released for the validation process. Pilot participants were randomly selected. Pilot study participants were asked not to answer the questions but to provide the following information:

1. Indicate whether the questions are few, moderate or too many;
2. Please mark any questions you feel are not clear, ambiguous, irrelevant and/or duplicate the endorsement of any other question;
3. The flow of the questions (is the order logical, or would you have preferred a different one, unnecessary questions, provide your suggestions);
4. Indicate whether the instructions were clear; and
5. Any additional comments or recommendations would be appreciated.

There were negative comments about the content and length of the questions in the pilot study. Accordingly, the observed mistakes were rectified soon after the study. Nonetheless, the ethical and administrative questions were understood and accepted by all pilot study participants. Based on the comments of the participants, the instructions and the questions accurately explained the intentions of the researcher. Soon after the pilot study, the model accompanied with questions, was remitted to the ITSEs.

5.6.1.2 Validation Process

During the validation process, both positive and negative comments were received. The opinions of the experts aided in highlighting the weaknesses and strengths of the model. In addition, the feedback from the experts provided advice on whether the proposed model contributed new ideas to the field of knowledge of Information Technology and digital forensics in general and also suggested areas that needed to be improved on the model.

All the experts validated the proposed model focusing on the originality aspect, technical quality, and presentation of the model. In this research project, originality refers to

developing or creating something that no one else has done; that is independent thinking and new ideas in the proposed artefact (Guetzkow, Lamont & Mallard, 2004). The technical quality aspect refers to the degree to which the physically measurable attributes of the developed artefact or model meet professionally acceptable standards (English, 2000). On the presentation aspect the experts checked whether the developed idea (model) was displayed, shown and explained clearly, making the idea easily understandable and known to others and/or to the public (Guetzkow *et al*, 2004).

All experts were urged to provide their opinions of the questions; hence, providing their views about the proposed model within their field of expertise. In providing opinions about the model, the qualitative component of the research methodology was satisfied. At the beginning of the questionnaire, the administration and ethical issues were addressed. It was mentioned that the names of individuals or companies participating would remain anonymous. Therefore, the proposed model was more important than the identity of individuals and/or organisations participating in the validation of the model.

5.6.1.3 Data Analysis

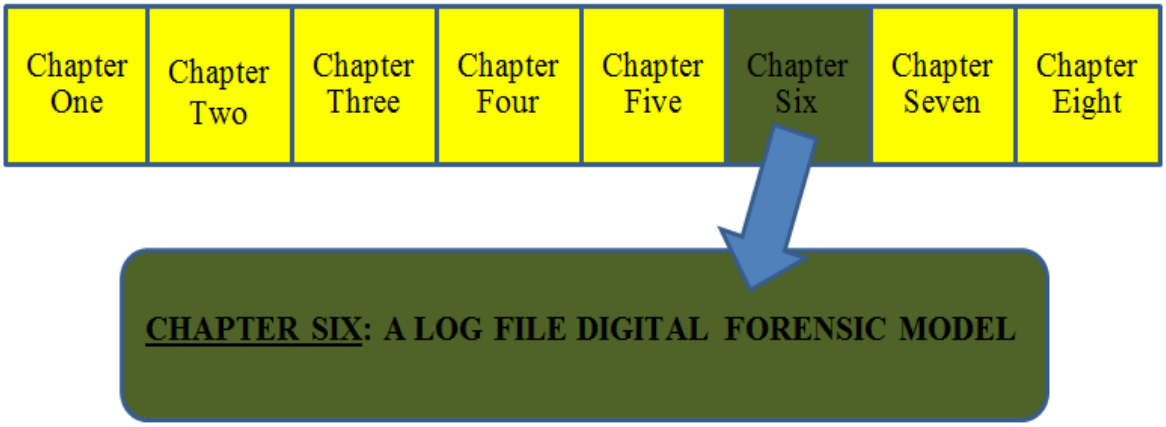
As part of analysis, the qualitative data from the experts was summarised while changes were made according to feedback as an auxiliary phase in improving the proposed model. As the comments of the experts both supported and contrasted the proposed solution, mind mapping technique was used to lap some ideas from the feedback to improve the model. This added to the credibility of the project. Additionally, findings were drawn from the comments of the experts. This marks the end of the primary data discussion and its analysis; the next section contemplates the secondary data.

5.6.2 Secondary Data

In this research study secondary data comprised information collected from various sources. The sources included articles from academic journals, books, conference proceedings, results from previous studies and existing models, reports and websites. Information and Technology is a fast growing field; new ideas are introduced every day; therefore, the most recent data available was included. All secondary data was referenced, ensuring that the primary author was recognized; all references are listed in full at the end of the research project.

5.7 Conclusion

This chapter discusses the research design and methodology followed in this research project. This research project falls under the "*Reality as a Contextual Field of Information*" along the continuum of core ontological assumptions. The Design Science research methodology was discussed as a research method of adoption towards the development of a new artefact called ***A Log File Digital Forensic Model*** intended for use specifically in network log mining. The Diffusion of Innovations (DOI) Theory was introduced as a notion behind the manner in which the proposed model could be received in the community. Then, the chapter discussed the investigations which generated some new ideas from the presiding South African legislation. The data collection procedure followed the primary and secondary data assemblage. Primary data was collected through the validation of the model from expert reviews. The experts provided their opinions and/or views about the proposed model. The views of the experts provided a qualitative data element to this research project. The chapter then deliberated on how the data was analysed. The analysis of all the data collected in this research project will be discussed in chapter 7. The next chapter introduces and discusses the proposed model.



6.1 Introduction

This chapter introduces a new model called: A Log File Digital Forensic Model. Chapter 2 revealed the importance of network log mining and how digital evidence from log files is viewed from a legal perspective. Chapter 3 discusses the challenges confronted in network log mining. Chapter 4 revealed that among the existing forensic models reviewed in this research project, most of them are nonspecific, yet this study argues that there has to be forensic models specifically for different digital forensic scenarios, such as in this case, for network log mining. This chapter introduces a proposed Log File Digital Forensic Model that can be used in the process of network log mining which has an expected adoption explained by the Diffusion of Innovations (DOI) Theory. Firstly, the architecture of the proposed model will be discussed in detail. Then, a demonstration of how the model can be used in practical situations will be discussed. Lastly, the evaluation of the model including the benefits and limitations will be discussed.

6.2 The Proposed Model

This research project insinuates that the proposed Log File Digital Forensic Model improves and provides a standardised and comprehensive process flow which facilitates detailed specification towards a network log mining process in computer networks. The model is drawn from other models that have been proposed all over the world. This model was developed using some ideas borrowed from other digital forensic models such as the one proposed by Mandia and Prosise (2001, as cited in Reith *et al*, 2002) and Ieong (2006). However, some new imperative processes to specifically suit network log mining have been introduced. The principal practices have their origin in the United States of America (USA) which is more advanced in the digital forensic discipline. The Proposed

Log File Digital Forensic Model is a comprehensive analytical procedure that can be useful regardless of the skills applied and the evidence to be recovered. The model is a collection of generalized steps that can be followed by a forensic practitioner as a legal and detailed model when recovering digital footprints of cyber-criminals from the log files of computer network devices. The objectives of the proposed model correctly signify physical procedures that form part of an investigation. The model is as illustrated in Fig 6.1.

The model is broken down into four basic layers, with laws and regulations being the base of the model. Laws and regulations as well as the proposed ***Forensic Investigation Precautions*** (see Fig 6.1) are meant to provide a strong guideline to the forensic investigation procedure. The four main layers of the model are:

- Preparation Layer,
- Discovery Layer,
- Testing Layer and
- Elucidation Layer.

Each layer has unique processes that have to be run and completed before the inchoation of the following layer. The model has a top-down layout, meaning that the layers are run from the top and flow downwards. Thus, based on the proposed model, the investigation starts on the Preparation Layer and finishes with the Elucidation Layer. Processes within each layer are conducted in the direction of the arrows as shown within each layer. If forensic practitioners carry out a log file forensic investigation in computer networks through these four basic layers, this research project proposes that there will be little doubt that the forensic investigation will have been thoroughly conducted. The following paragraphs provide a full description of the proposed model.

6.2.1 Preparation Layer

Thorough preparation is one of the most important factors to be considered before starting technical investigations in the log files. According to KPMG (2009), the impact the investigation may have on the organisation must be acknowledged and the cooperation of the victim organisation must be justified. The investigation may be costly for the organisation. Therefore, before making any progress, the leading practitioner should meet

the leaders of the organisation to discuss the merits and detriments the investigation may cause (Ieong, 2006). The processes in the Preparation Layer are Formulate Approach and Pre-incident Preparation.

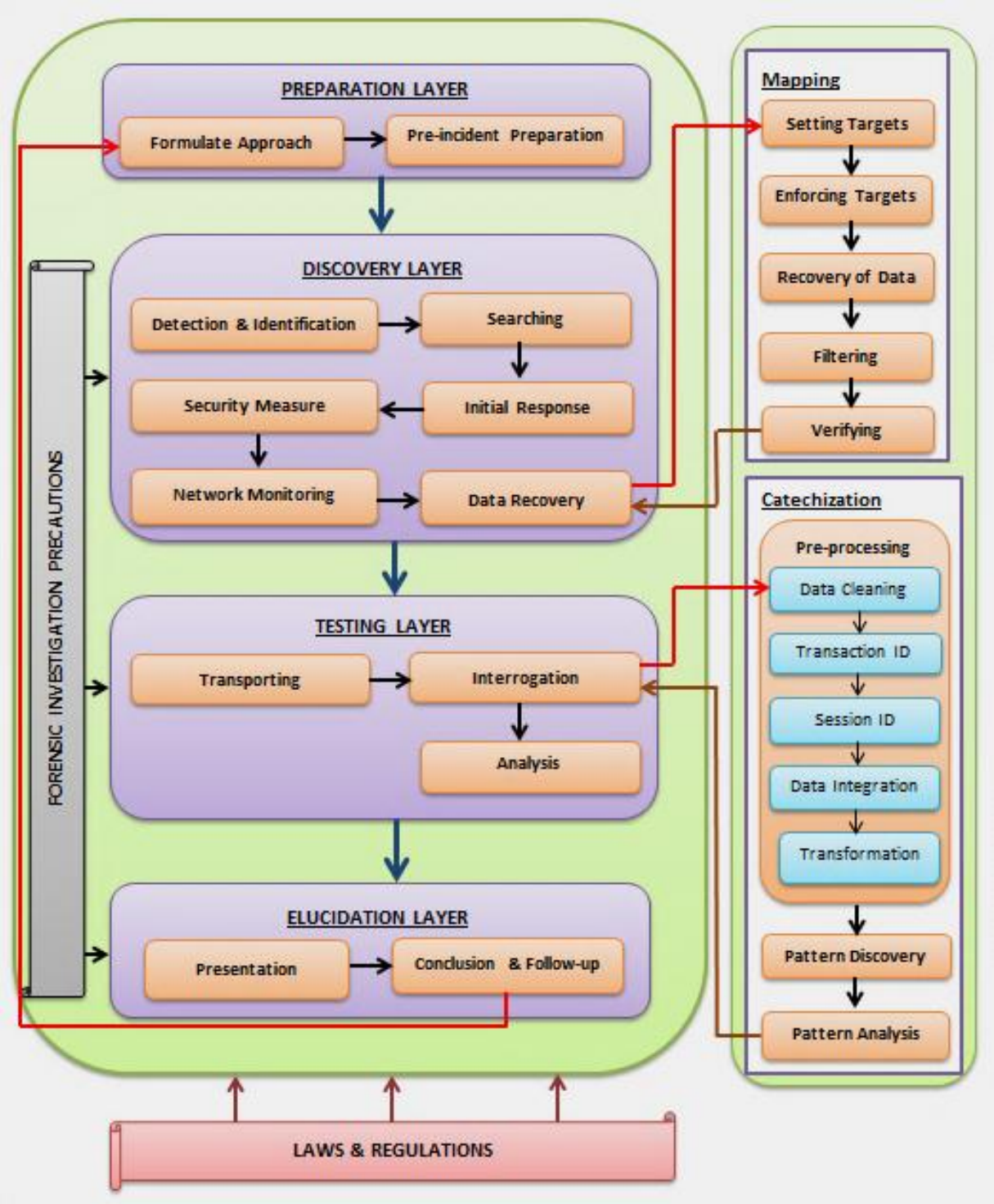


Fig 6.1: A Log File Digital Forensic Model (Own Compilation).

The process of the Formulating Approach is the first aspect to administer in network log mining. Ayers (2009) states that before starting any part of the technical investigation it is important to discuss important considerations which must be brought to the attention of the practitioners as part of the approach. Therefore, conducting network log mining without taking these considerations into account may render the digital evidence insubstantial in a court of law; the considerations are legal and technical operations (Roger *et al*, 2006). This study asserts that as with any other type of investigation, practitioners must make these considerations prior to making any decision about the effective and efficient methods to use.

Legal considerations involve having legal authorization such as legal notice and scope of investigation from the law enforcement (Ramabhadran, 2007). Roger *et al.* (2006) also assert that the leading practitioner must check closely whether the legal notice allows the seizure and removal of the network devices, whether it allows an onsite examination, whether it gave the reporting obligations to the issuing magistrate or judge as well as particular discovery issues present and anticipated, and whether conducting an onsite examination can affect the integrity of the original evidence as well as the expected error rate. Furthermore, authorisation must be obtained in order to acquire the necessary data, and in the South African context the applicable laws as well as international laws must be adhered to from the beginning of the investigation (Dagada *et al*, 2009). The reason for working closely with the legal notice is because judges usually check whether what was done during the investigation corresponds with what is stated on the notice.

Additionally, technical considerations on the legal notice include the type of the case, how critical the time factor is, the skills and abilities of the forensic practitioners as well as the type of technology to be used (Roger *et al*, 2006). Thus, the leading practitioner must make sure that the skills required from the beginning to the end of the investigation and the tools needed for the investigation, are available. An estimated timeline is the time/period the investigation is expected to take and this has to be documented during the formulating approach process as a critical time factor (Garcia, 2007). Moreover, technical considerations seeks to know if the scene is safe and effectively controlled and whether the computer network system(s) in question is to be powered off or will remain running (Ayers, 2009). In addition, the technical skills and knowledge process of the suspect(s), if

anticipated, has to be considered (Roger *et al*, 2006); not withstanding that it is important to check whether the practitioners have proper equipment for both onsite and private laboratory examinations.

For an adequate investigation to begin, the necessary authorisation must be obtained from key individuals of the victim organisation, for example, the system administrator, directors and owners of the organisation during the approach process (Ciardhuáin, 2004). In addition, the organisation should inform the staff members and other participants involved in the investigation concerning its extent. Thus, the legal notice must be provided to all concerned parties notifying them of the forensic investigation (Ramabhadran, 2007). The type of crime reported by the organisation should be taken into consideration because the type of crime affects the approach taken towards the investigation (Richard & Roussev, 2006). For example, the way the investigations should be conducted after cyber-terrorism is different from when illegal websites or documents are posted on the website of an organisation (Chizoba, 2005). Planning the approach is the foundation of a quality forensic investigation on computer networks (Ciardhuáin, 2004). Therefore, the approach technique has to be robust; otherwise if the approach is weak, the results are likely to be weak and not convincing. Hence, an appropriate approach strategy for the investigation should be developed taking into account the nature of the incident and various technical, legal and business factors. After formulating a good approach, then the leading practitioner and the forensic team commence the Pre-incident Preparation.

The process of Pre-incident Preparation is the second and last process in the Preparation Layer. Pre-incident Preparation implies how the investigation is going to proceed. Without proper preparations, Ciardhuain (2004) avows that the investigation is likely to be run haphazardly; as a consequence, this can reduce the weight of the evidence and can be strongly questioned in a court of law. The leading practitioner must have a matrix that quantifies the various possibilities of the crime scene, the suspect and the digital evidence as well as the expertise of various practitioners on the forensic team (Rogers *et al*, 2006). This matrix is used to identify what is already available and what is yet to be established, thus, aiding in determining the necessary supplements. Additionally, during the Pre-incident Preparation practitioners must have an initial understanding of the nature of the crime and prepare the tools and equipment to be used, relevant skills and continuing to

educate oneself about new technologies, building an appropriate team, assigning duties to each personnel, legal coordination and general monitoring approval, and accumulating materials for packing evidence sources (Carrier & Spafford, 2003).

Moreover, during the Pre-incident Preparation process, practitioners should ensure that the victim organisation is really certain of the need for a forensic investigation; explicitly that, its computer networks have been compromised and that laws and regulations have been violated (Stephenson, 2003). Additionally, practitioners must ensure that the operations and infrastructure of the organisation can sustain the forensic investigation (Baryamureeba & Tushabe, 2004). A thorough Pre-incident Preparation process increases the reliability of evidence and minimizes the risks and threats associated with an investigation (Ramabhadran, 2007). Bearing in mind what has been discussed, this study purports that the purpose of Pre-incident Preparation is to allow the leading practitioner and the forensic team as well as the victim organisation to be ready for a forensic investigation which is greater than merely focusing on evidence recovery; however, KPMG (2009) says this is very costly. The Pre-incident Preparation process marks the end of the Preparation Layer and the investigators must then delve into the Discovery Layer.

6.2.2 Discovery Layer

The discovery layer is the second layer in this model in which vital tests are carried out on the suspected routers, firewalls network management systems and switches in order to identify the ***digital footprints*** of the cyber-criminals. In other words, practitioners have to discover what exactly transpired in an organisation's computer networks and elicit the evidence. Processes associated with this layer are Detection and Identification, Searching Evidence, Initial Response, Security Measure Implementation, Network Monitoring and Data Recovery.

Under the Detection and Identification process the forensic practitioners must make use of adequate tools and equipment to detect the intrusions of the private networks before rushing into the log files of network devices. This is done to confirm the intrusion of cyber-criminals into the organisation's computer networks (Baryamureeba & Tushabe, 2004). Specialized tools, equipment and techniques must be used (Forte, 2004a) and carried out by either using the signature based intrusion detection system or the anomaly

based intrusion detection system (Jajodia, 2008). Signature based intrusion detection systems rely on pattern-matching techniques; they contain a database of signatures of known attacks and try to match these signatures with the analyzed data. A match indicates computer networks intrusions (Jajodia, 2008). On the other hand, anomaly based detection systems first build a statistical model describing the normal network traffic, and then flag any behaviour that significantly deviates from the model as an attack (Jajodia, 2008). The anomaly based detection analyses the payload data in the transport layer (network layer 4), and/or considers packets headers in a network layer (network layer 3) of the Open Standard Interconnection (OSI) model, a standard model for networking protocols. Once the Detection and Identification process is completed, a warrant specifically for the detected incident will have to be obtained from the law enforcement before proceeding with the investigations (KPMG, 2009). After the Detection and Identification process, the leading practitioner and the forensic team should begin to search for the evidence.

The process of Searching Evidence involves activities such as Evaluating Crime Scene, Formulating Relevant Search Plan, and Searching Potential Sources (Ciardhuáin, 2004). In this process, the leading examiner must not only evaluate the electronic equipments at the scene to determine whether any expert assistance is required in processing the scene but also identify people in the scene and conduct preliminary interviews (Ramabhadran, 2007). Furthermore, the owner(s) and users of the electronic devices or system administrator(s) are supposed to provide valuable information like the purpose of the system, security schemes, various applications present in the network devices, user names, passwords and encryption details if available (Grand & Carrier, 2004). If it is important to search for some information or items not listed on the warrant, then appropriate changes must be documented or else a new warrant must be obtained; failure to do so will result in questionable evidence in a court of law (Ramabhadran, 2007). The systematic search for evidence is a process where investigators survey the physical and virtual crime scenes, generating the best methods to pursuit the evidence (Reith *et al*, 2002). After Searching for the evidence, the leading practitioner has to meet the leaders of the organisation for Initial Response.

With the outcome of both the Detection and Identification process as well as Searching Evidence process, the leading practitioner has to relay the information from these

processes to the owners and leaders of the organisation. Under the Initial Response process, the leading examiner has to reveal the results of the latter processes verifying whether the claim made by the system administrator(s) of the victim organisation is valid or not (Carrier & Spafford, 2003). If the claim is not true then the investigation will have to end at this process; but, if it is true then the owners or leaders of the organisation have to whether acknowledge the continuation of the investigation, or not (Carrier & Spafford, 2003). This depends on the nature of the case; if the case can be resolved within the organisation then there is no need to take it further to a court of law (Ramabhadran, 2007). Thus, on the initial response process the owner(s) of the organisation have to be made aware of their claim and also whether the source of intrusion is within the organisation or outside. As the investigation moves forward then the practitioners engage the Security Measure Implementation process.

When permission has been granted to continue with the investigations the practitioners have to push through the Security Measure Implementation. This process is required in some cases especially when practitioners involved are part of the victim organisation's internal incident response team (Bejtlich, 2007). The practitioners have to advise the administrator(s) of the organisation on how to improve the security of the organisation's computer networks (Reith *et al*, 2002). Giordano and Maciag (2002) state that practitioners are required to help by explaining the weaknesses or vulnerabilities used by criminals to enter the organisation's computer networks and how the security on the computer network devices can be improved. The Security Measure Implementation should be conducted just after the Initial Response process so that the weaknesses discovered during the Detection and Identification as well as Searching process can be resolved before moving forward with the forensic investigations (Reith *et al*, 2002). When the Implementation Measure process is complete the practitioners with the help of the administrators of the organisation have to monitor the computer networks to provide favourable conditions for the Data Recovery process.

The Network Monitoring process follows the Security Measure Implementation and this involves reducing the amount of network traffic flow within the organisation; this is a crime scene protocol followed in any crime arena (Reith *et al*, 2002). The network monitoring process involves the securing of incident site, hence maintaining the integrity

of log files in the network system devices (Carrier & Spafford, 2003). Network monitoring involves rechanneling the flow of network traffic through other local networks to avoid significant changes in the log files of the suspected network devices (Ramabhadran, 2007). The goal of network monitoring is to reduce the rate at which the log files in the network devices are updated by new transactions taking place on the networks; hence, reducing the chances of making slurred images when the log file entries are copied during evidence retrieval while new transactions are updated (Jones, 2007b). The log files of an organisation can record about 240 million entries per day (Munk *et al*, 2010), so to capture the proper metrics including intrusions and disruptions, the network system has to be restricted. The main aim of the Network Monitoring process is to reduce the rate at which new entries are recorded on the log files of the computer network devices. This reduces the tiresome work of searching for relevant entries amongst millions of entries in the log files during the Data Recovery process.

In the Data Recovery process, forensic practitioners have to rummage for the digital footprints of the cyber-criminals in the log files in the computer network devices. This task is one of the most crucial tasks in network log mining; therefore, special skills are required to perform the duty (Jones, 2007b). If the organisation has a log file server(s) then the practitioners have to focus on the server(s); however, if not then they have to check the log files of individual suspected network devices (Das & Turkoglu, 2009). The processes within Data Recovery process mark the first distinctive aspect of this model, specifically for network log mining, from other proposed forensic models. The Data Recovery process involves some activities under the umbrella name Mapping; these activities are Setting Targets, Enforcing Targets, Recovery of Data, Filtering and Verifying. According to Casey (2008) the Data Recovery process is a very tiresome manual job.

The Setting Targets activity in network log mining requires practitioners to document what they are supposed to accomplish during the process of Recovery of Data (Forte, 2004a). In addition, Forte (2004a) emphasizes that all steps must be numbered such that if there is a change in personnel the tasks will still be carried out in the same sequence. After setting the targets, the practitioners must Enforce the Targets. By Enforcing Targets the practitioners have to ensure that all documented aspects are achievable (Forte, 2004a). When satisfied that the targets can be achieved, practitioners are required to discover those

batches of relevant log entries which can provide significant evidence about the cyber-crime from the log files (Shebaro *et al*, 2010). Once these batches are found, they must be copied into new directories or folders. The evidence must not be modified, manipulated or destroyed and Ryan and Shpantzer (2005) are of the view that if to some extent modification occurred, the rate of modification must be negligible and has to be documented. Appropriate forensic tools must be used in the recovery of digital evidence to ensure the admissibility of evidence in a court of law (Roger *et al*, 2006). The integrity and authenticity of the recovered evidence must be ensured through mechanisms such as hashing, write protection and non-repudiation (Ramabhadran, 2007). The process of looking for batches of relevant log entries and the copying or imaging of those batches is the Recovery of Data activity.

When the relevant log file entries are copied to respective directories, the directories can then be filtered into safe repositories, for example digital evidence bags (DEBs) (Schuster, 2007). A DEB is software used to collect information for digital forensic examination; thus, DEBs are universal containers for digital evidence from any source (Pladna, 2008). Pladna (2008) goes on to say that all possible sources of digital evidence must be identified and tagged correctly before packaging. The Verification activity is conducted last and this involves rechecking the log files to verify that all relevant log entries have been copied (Forte, 2009). If the practitioners are not satisfied with the retrieved log entries the process of Data Recovery can be restarted before moving forward with the investigation. When this happens, the practitioners have to chronologically move through each activity once again; hence, beginning with the Setting Targets and moving towards the Verification activity. Data Recovery process is the last process in the Discovery Layer.

6.2.3 Testing Layer

The Testing Layer comes just after the Discovery Layer and consists of Transportation and Storage processes, Interrogation process and Analysis of the digital evidence. According to Forte (2009), algorithms and techniques from several research areas and skills are needed in the processes associated with this layer. According to Ramabhadran (2007), the digital evidence will be verified in order to find out the intentions, missions and the actions of the cyber-criminals within an organisation's computer networks.

When the practitioners are satisfied that the mapping process has been conducted comprehensively, the filtered evidence and seized devices (if any) have to be transported to a forensic laboratory for safe keeping, Interrogation as well as analysis of the log entries (Baryamureeba & Tushabe, 2004). Proper safety measures must be maintained because the evidence (filtered log file entries) in the digital evidence bags (DEBs) can be destroyed while in transit due to shock, excessive pressure, humidity and temperature (Ramabhadran, 2007). Plastics should not be used to cover the seized network devices or the DEBs because they cause static electricity; however envelopes may be used (Ramabhadran, 2007). Thus, during the transportation process the integrity of the evidence must be ensured so that it remains valid because filtered evidence has to be presented in raw form to the judges in a court of law (Das & Turkoglu, 2009). Furthermore, the DEBs and seized devices must always be kept in a conducive environment or place where there is no or less electromagnetic radiation, dust, heat or moisture (Ramabhadran, 2007). Moreover, to make sure that the evidence is well preserved and authentic, unauthorized people are not allowed contact with the evidence throughout the investigation (Jones, 2007a).

After Transportation, the Interrogation process begins. The Interrogation process involves examining the contents of the collected log entries by forensic practitioners so as to procure important information relevant to proving the case (Freiling & Schwittay, 2007). This is a process involving major tasks and marks another exceptional aspect of this particular model from other proposed forensic models. During the interrogation process an in-depth exploration of the filtered log files is carried out. This involves the application of specialised digital forensic tools and techniques used to gather evidence and to further scrutinise the log file entries (Arthur & Venter, 2005; Tug *et al*, 2006). At this point a number of back-ups of the filtered evidence must be created before analyzing the collected log entries. The process of interrogation must make the evidence visible, explaining its originality (source computer) and significance (the mission of the cyber-criminals). Huge volumes (batches) of collected data need to be divided into small sizes and form for better analysis (Das & Turkoglu, 2009). The Interrogation process, which is associated with the Catechization activity, is broken down into three processes namely: Pre-processing activity, Pattern Discovery activity and Pattern analysis activity (see Fig 6.1). According to Gilder and Vigilante (2001), Catechization in computer related tasks is a unique process

of meta-cognition, learning, gaining new skills and knowledge about how events took place. The purpose of the Pre-processing activity is to offer a structural, reliable and integrated data source for Pattern Discovery (Das & Turkoglu, 2009). The Pre-processing task consists of Data Cleaning, Transaction Identification, Session Identification, Data Integration and lastly Transformation (see Fig 6.1).

All the batches of log files filtered into the DEBs are in raw form. Therefore, not all of the log entries are valid for pattern analysis; thus, only the entries that carry relevant information must be retained (Das & Turkoglu, 2009). Hence, the Data Cleaning activity eliminates irrelevant entries from the access log files, which include entries that have 'error' or 'failure', some access records generated automatically by a search engine agent, requests for picture files associated with requests for particular pages, and entries with unsuccessful HTTP status codes; only successful codes must be considered (Munk *et al*, 2010). After Data Cleaning the Transaction Identification activity begins. The aim of Transaction Identification is to create meaningful clusters of references for each individual who accessed the organisation's networks (Das & Turkoglu, 2009). After Data Cleaning, the log entries must be partitioned into logical clusters using one or a series of Transaction Identification modules (Munk *et al*, 2010). The module can be a merge or a divide module and both types of modules take a transaction list and possibly some parameters as input; the output is a transaction list that has been operated by a function in the module in the same format as input (Das & Turkoglu, 2009).

Subsequently, the Session Identification activity trails the Transaction Identification. A session can be described as a group of activities performed by a user from the moment he enters the computer networks to the moment he leaves (Das & Turkoglu, 2009). Munk *et al*. (2010) aver that identification of a session on networks helps to identify how the cyber-criminals manoeuvred within an organisation's computer networks. Therefore, Session Identification is the process of segmenting the access log of each log file entry into individual access sessions (Munk *et al*, 2010). A session includes the used IP address (source address), user identification (ID) and the Uniform Resource Locators (URLs) of the accessed data, and access time (timestamps) (Das & Turkoglu, 2009). Thus, Session Identification helps to establish whether the computer used was from an Internet cafe, a computer classroom or a home based computer (Munk *et al*, 2010). Data Integration

follows identification of all relevant transactions and sessions. According to Munk *et al*. (2010), Data Integration, also known as Data Fusion, is the activity of combining the matching Transactions and Sessions. After the matching process in Data Fusion then the Transformation of the combined data commences. The Transformation activity involves making sure that the whole Pre-processing activity has been conducted and rechecked thoroughly and that there are no errors (Das & Turkoglu, 2009), thus, providing a good foundation for the Pattern Discovery and Pattern Analysis activities. The Transformation activity marks the end of the Pre-processing activity.

The Pattern Discovery activity follows the Pre-processing activity. The Pattern Discovery is one of the key processes in log file forensic examinations (Das & Turkoglu, 2009). Das and Turkoglu (2009) state that Pattern Discovery includes the algorithms and techniques from several research areas such as Data Mining, Machine Learning (artificial intelligence), Statistics and Pattern recognition. In addition, Pattern Discovery involves a thorough search for passwords used to enter the organisation's computer networks, finding unusual hidden files or directories accessed, file extension and signature mismatches (Munk *et al*, 2010). The techniques such as statistical analysis, association rules, clustering, classification, sequential pattern, dependency modelling and path analysis are used to analyse the pre-processed log file data in a process of discovering patterns of the movements of criminals within the networks (Das & Turkoglu, 2009). Ieong (2006) says recovering relevant ASCII as well as non-ASCII data are some of the major steps performed during the Interrogation process. Moreover, Das and Turkoglu (2009) suggest that in computer auditing, the Pattern Discovery allows forensic practitioners to determine the path used by criminals as they navigated within the organisation's computer networks.

The last process in the Interrogation process is Pattern Analysis. Pattern Analysis, also called Reconstruction, has two major activities, Resolution and BackTracing. The aim of Resolution Activity is to extract the interesting rules, patterns or statistics from the Pattern Discovery process by eliminating the irrelative rules or statistics (Das & Turkoglu, 2009; Munk *et al*, 2010). According to Das and Turkoglu (2009), in the Resolution activity the Pattern Analysis process of log file examination is one of providing tools to facilitate the transformation of information into useful knowledge. When the forensic practitioners have enough knowledge of the crime from the Resolution activity, then the BackTrace process

begins. BackTracing is the process of using a retrieved source IP address acquired in Session Identification to trace back through the Internet Service Providers (ISP) to the source computer (Baryamureeba & Tushabe, 2004). Thus, the goal of network traceback search is to allow determination of the source of attack traffic, so that a particular host used by an individual to initiate an attack can be identified, and real world investigative techniques used to locate the person responsible (Buchholz & Tjaden, 2007). Therefore, BackTracing is a process of reconstructing the activities of criminals in the organisation's computer networks focusing on retrograding completion of records on the path followed by the criminals (Munk *et al*, 2010). The BackTracing process also allows practitioners to ascertain the password used on the used computer, the user ID on the used computer, by referring to the timestamps on the log files (Forte, 2004a; Casey, 2008). Furthermore, it is important to prove that the evidence has not been altered by the forensic specialist after the Interrogation process; therefore, the hashing techniques such as md5 and SHA1 must be used for mathematical authentication of original evidence just after the Pattern Discovery process (Rogers *et al*, 2006). If the forensic practitioners are not convinced of the results from the process of Pattern Analysis then the process of Interrogation can be restarted (see Fig 6.1).

The Analysis process comes after the Interrogation process and involves technical reviews which have to be conducted by forensic practitioners on the patterns discovered and analysed. The major activity in the analysis process is the correlation task. Correlation of events on log files involves identifying relationships between fragments of data, analyzing hidden data and determining the significance of the information obtained from the log files from both the source computer and the filtered log files (Forte, 2004a). Reconstructing the event data based on the extracted data and arriving at appropriate conclusions are some of the activities performed during the correlation activity (Kent *et al*, 2006). User IDs, passwords as well as usernames from both source logs and filtered logs have to be correlated to see if there are some resemblances with respect to time. According to Forte (2004a) the most important issue in correlation of the events on log files is the timestamp interpretation.

Timestamps provide proof of when the criminal offense occurred referring to both the source log files and the filtered log files (Casey, 2008). Buchholz and Tjaden (2007) are of

the view that timestamps allow reconstruction of events obtained from multiple; geographically separate sources each with its own clock. A useful technique for event reconstruction using timestamps is "***time-lining***"; thus, depending on the source events ***time-lining*** can provide a practitioner with a detailed sequence of the events that took place on the system (Buchholz & Tjaden, 2007). Casey (2008) agrees with Buchholz and Tjaden (2007) that to make sure that the original time the source computer was utilized correlates with the time recorded on the log files obtained from the victim organisation; the Coordinated Universal Time (UTC) must be used. The UTC is a high-precision atomic time standard based on the earth's rotation rather than the passage of seconds which may differ quite considerably from 'real' time, be misconfigured to be used in the wrong time zone, manipulated arbitrarily on the source computer when launching an attack, and/or may run fast or slow (*clock skew*) (Buchholz & Tjaden, 2007; Casey, 2008). Timestamps also allows the events on log files to be recorded, analysed and potentially correlated regardless of the difference of the time zones of the source log files and the filtered log files (Forte, 2004b; Casey, 2008). In other words, timestamps are important so that cause-time relationships may be established for an investigation (Buchholz & Tjaden, 2007).

In addition to the correlation of events on log files, the U.S. Department of Justice (USDOJ) recommend timeframe analysis, hidden data analysis, application analysis and file analysis of the extracted data (U.S. Department of Justice, 2004). The results of the Analysis process should be completely and accurately documented as they are used in a court of law (Roger *et al*, 2006). However, in most cases when the analysis process is complete a concrete hypothesis is established and it is most likely that the suspect is clearly identified by the data and information from this process (Ciardhuáin, 2004). Ciardhuáin (2004) adds that at this stage most suspects will either plead guilty to the charges or accept a deal due to the incriminating evidence. The Analysis process marks the end of the Testing Layer (see Fig 6.1).

6.2.4 Elucidation Layer

The Elucidation Layer trails the Testing Layer and is the last layer in the Proposed Log File Digital Forensic Model. This layer explains how the processes were run throughout the whole forensic investigation. The explanation must be very simple and clear, considering that some people take time to comprehend digital forensic ideas and views.

The processes in this layer are only Presentation process and Conclusion and Follow-up process.

After Data Recovering, Interrogation and Analyzing processes, the results must be presented in a court of law not only before law enforcement officials but also technical experts, legal experts and corporate management (Grand & Carrier, 2004). Perumal (2009) suggests that it should be possible to confirm or discard the allegations regarding the particular crime or suspicious incident. The individual results of each of the previous processes may not be sufficient to arrive at a proper conclusion about the crime; therefore, Presentation of these results in a court of law gives a clear picture to the audience (Giordano & Maciag, 2002). The results of the Interrogation and Analysis processes must be reviewed in their entirety to obtain a complete picture (Ramabhadran, 2007). When presenting the evidence in a court of law opposing theories will also be presented and therefore, there is a need to provide substantiated exhibits of the events that occurred as well as support of the theory or model used (Ciardhuáin, 2004). Ciardhuáin (2004) says digital forensic practitioners must prove the legitimacy of their theory. This research project asserts that this is where the benefits of having a well-designed and accepted digital forensic model enables either the conviction of a suspect or the discharge of an innocent individual. In a court of law the presented model will be challenged using opposing models and theories, and their supporting evidence will also be presented (Ciardhuain, 2004). The successful challenge of the model used will result in the revaluation and construction of a better model; each model will be supported the evidence presented and be based on existing laws and legislation (Lalla & Flowerday, 2010).

When presenting the evidence a report comprising of an in-depth abstract of the various steps in the whole process of the investigation and the findings must be provided (Freiling & Schwittay, 2007). In many cases, the leading practitioner provides expert testimony in a court of law and all complex terms associated with digital forensic investigation processes need to be explained in layman's terminology (Jones, 2007b). The expertise and knowledge of the forensic practitioner(s), the model adopted, tools and techniques used are all likely to be challenged in a court of law (Ramabhadran, 2007). In addition, Kent, Chevalier, Grance and Dang (2006) suggests that along with the report, supporting materials like copies of the DEBs containing the filtered raw log file entries, a chain of

custody document and printouts of various items of evidence should also be submitted to the judges. When the court case has ended whether in a positive way or in the negative way, then the Conclusion and Follow up process begins.

The Conclusion and Follow-up process involves reviewing all the steps in the investigation and identifying areas which need improvement (Carrier & Spafford, 2003). As part of the Conclusion and Follow-up process, the results and their subsequent interpretation can be used to further refine the Data Recovery, Interrogation and Analysis processes of the network log mining investigations (Freiling & Schwittay, 2007). The Conclusion and Follow-up process involves the distribution of information in order to provide a basis for future investigations (Ayers, 2009); therefore, the conclusions made in a court of law about the evidence as well as the model used, influences future investigations (Ciardhuáin, 2004). The Conclusion and Follow-up process marks the end of the Elucidation Layer as well as the whole process of network log mining (see Fig 6.1).

However, mere running of the processes from the Preparation Layer to the Elucidation Layer as explained above may cause disputes in a court of law. Accordingly, laws and regulations must be involved and observed in forensic investigations; if the laws and regulations are not followed then the court may dismiss the allegation (Ryan & Shpantzer, 2005). Nevertheless, in addition to the laws and regulations, this research project proposes some forensic precautions. As mentioned in chapter 5, section 5.5, this research project argues that there are some important aspects not emphasised or fully addressed by South African residing legislation towards network log mining and digital forensics in general. Therefore, this study proposed ***Forensic Investigation Precautions*** in an effort to augment some of the aspects not addressed in the legislation. These precautions are included as part of the forensic investigation. The issue of forensic precautions is one of the exclusive points projected in this research project. The forensic investigation precautions proposed in this research project are discussed in the next section.

6.2.5 Forensic Investigation Precautions

The forensic investigation precautions were created to augment and enhance the seizure of digital evidence when applying the proposed log file digital forensic model in computer networks. When using this proposed model, all precautions listed must be adhered to from

the Discovery Layer through to the Elucidation Layer where most of the technical and illuminating tasks are run. Some of the precautions are aspects already known in the field of digital forensics but are not emphasised either as part of any forensic model or in the presiding laws and regulations. Hence, further research resulted in the development of the precautions to intensify the process of the forensic investigation. This study therefore included these precautions as part of the proposed model. The proposed ***Forensic Investigation Precautions*** involve avoiding experiments on original evidence; accountability for any change; observation of available IT best practices; not exceeding one's knowledge; maintenance of a chain of custody (documentation) and utmost good faith.

1. Avoid experiments on original evidence

During network log mining, the evidence can be manipulated in the course of the investigation. During the Interrogation process, much care must be taken to avoid manipulation of the evidence, especially avoiding tests on the original copy of the log entries (Ayers, 2009). More than one test may be conducted and the original copy has to be submitted to the judges in a court of law (McCombie & Warren, 2003). When the condition of the evidence on the original copy is changed, it may be strongly challenged and can be rejected in a court of law because the jury may consider the evidence biased (Ryan & Shpantzer, 2005). If the evidence is manipulated all the efforts and money already spent in the investigation will have been wasted (Beebe & Clark, 2005). Hence, considering the running costs involved in an investigation this study asserts that practitioners must always avoid tests on original evidence.

2. Accountability for any change

Even though practitioners may avoid experimenting on original evidence, Jones (2007b) argues that it is difficult to finish the whole process without making errors because most of the tasks conducted in forensic investigations are done manually. Nonetheless, Ryan and Shpantzer (2005) are of the view that the degree of manipulation (error rate) if any, even if insignificant must be documented by the responsible investigator. The name of the responsible practitioner must be documented as well as the time and date on which the alteration of evidence took place in the form of year, month, day, hours and minutes (McCombie & Warren, 2003). This helps in the Presentation of evidence in the court as it

is important to show the time and date the evidence was altered as well as the error rate. In addition, Ramabhadran (2007) asserts that the responsible practitioner must indicate the state from which the evidence changed to its netmare state.

3. Observation of available IT Best Practice

IT best practices provide clear comprehensive guidelines on how computer security can be applied in business organisations in a way that is acceptable in the IT field and in a court of law (Hamster, 2006). Hamster (2006) says that observing best practices when conducting forensic investigations on computer networks, gives the judges confidence that the evidence was acquired in a manner internationally acceptable in the IT discipline. ISO/IEC 27002, COBIT and some other IT security best practices were discussed in the last section of chapter 3. This research project follows only the ISO/IEC 27002; Hamster (2006) claims that following relevant IT best practices when conducting forensic investigations, aids the acknowledgement of the evidence by the judges in a court of law.

4. Not exceeding one's knowledge

This precaution emphasises the capabilities of the practitioners. When conducting forensic investigations, practitioners must not go beyond their knowledge (McCombie & Warren, 2003). This avoids a practitioner carrying out tasks which are outside his/her proficiency, thus, reducing the chances of manipulation of the evidence. Accordingly, a senior practitioner must be requested where investigations become difficult and complicated (McCombie & Warren, 2003). As a result, more experienced practitioners should be available during the investigation so that they can be referred to in complicated assignments. Failure to observe this precaution may render the whole case evidentiary weightless.

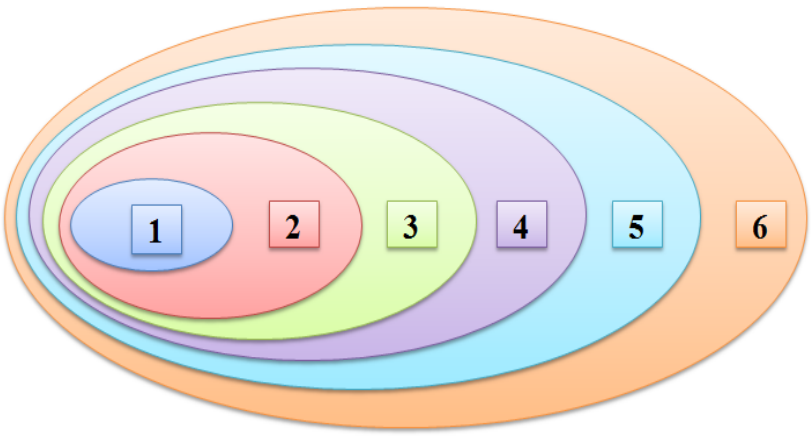
5. Maintain the chain of custody (Documentation)

Another precaution emphasises the recording (documentation) of all the steps followed in the process of network log mining. This precaution is important in the sense that it supports the victim organisation as well as forensic practitioners by showing all steps conducted; hence, supporting them in a court of law (Turner, 2005; Pladna, 2008). Pladna (2008) asserts that without a proper chain of custody, the digital evidence can be strongly questioned in a court of law; consequently, the evidence may not convince the judges and

therefore, can be discounted as true testimony of the raised allegations. Hence, this study emphasises that the chain of evidence and timeline of events must always be updated and consistent.

6. Utmost good faith

This is a precaution which hallmarks all other precautions. This precaution is derived from 'uberrima fides' standard, associated with insurance companies, meaning abundance of honesty or utmost good faith (Hasson, 2011). This precaution stresses the behaviour and attitude of the practitioners. Forensic investigations on computer networks are serious examinations which can lead to an organisation losing millions of dollars if cyber-criminals escape punishment, or conviction of criminals to the benefit of the victim organisation, or an innocent individual suffering negative consequences (Baryamureeba & Tushabe, 2004). Therefore, practitioners must be scrupulous in every aspect of their duty (Jones, 2007a). For example when the evidence has been altered, the responsible practitioner must document it and mention all necessary details while remaining completely objective. Nevertheless, this research project states that the precaution of utmost good faith do not upraise the trustworthiness of practitioners in the court; but emphasises provision of true testimony of their findings for the sake of the victim organisation, involved individuals and the practitioners' reputation. All the precautions discussed in this section can be viewed in an onion form as illustrated in Fig 6.2.



1. **Avoiding tests on the original evidence.**
2. **Accountability for any change.**
3. **Observing IT best practices.**
4. **Not exceeding your knowledge.**
5. **Maintain chain of custody (Documentation).**
6. **Utmost good faith**

Fig 6.2: Forensic Investigation Precautions (Own Compilation)

The most important information portrayed in Fig 6.2 is that all precautions depend on each other; thus, they influence one another in the sense that failure of one affects other precautions negatively. For example, if the practitioners ignorantly experiment on the original evidence, all other precautions will be fruitless even though they were observed. As a consequence, even if a practitioner followed the IT best practices, without exceeding his knowledge, maintaining the chain of custody and being honest, it would be difficult to convince the judiciary if the evidence had been modified during the experiments. It is important to mention that all precautions are of equal importance; the diagram only shows that all the precautions work together towards one goal; that is, reliable, dependable and sound digital evidence from log files of computer networks. Having precautions as part of the forensic model is important; however, laws must be adhered to as stated by Lalla and Flowerday (2010) and Ieong (2006) that without following laws and regulations it is just as good as not conducting the investigations in the first place.

6.2.6 Laws and Regulations

This research project was carried out in South Africa and the model was based on South African investigations, therefore presiding South African laws and regulations were used as a basis of this research project. However, this model can be applied in any country in the world by replacing South African presiding laws. It is important to understand the impression laws and regulations have on the digital forensic investigation process (Ieong, 2006). In the proposed Log File Digital Forensic Model the laws and regulations must be observed at each and every layer; hence, all the processes in all layers must be conducted following the relevant laws and legislation in place in the presiding jurisdiction (local laws) as well as international laws such as the Sarbanes-Oxley Act, Foreign Corruption Practices Act (FCPA) and Bribery Act of the USA. This research project also stresses that when conducting an investigation, laws and regulations must be the first criterion to consider before carrying out any initiation of the forensic investigation, as mentioned in the Formulate Approach process. Without abiding to the laws, criminals may go unpunished and the organisations may bear considerable loses (Baryamureeba & Tushabe, 2004). At this stage, the whole proposed forensic model has been discussed; the next section illustrates the demonstration of how it can be used.

6.3 Typical Scenario

It is importance to illustrate how a newly proposed artefact can be applied in practical situations. The theory of Design Science emboldens the practice of demonstrating a newly proposed artefact (Peffers *et al*, 2006; Hevner *et al*, 2004). When the model if fully instigated; gradually, through the Diffusion of Innovations Theory (DOI), the model will be employed in the industry. Concentration is on an investigation whose indisputable evidence is network intrusion of an organisation. When the system administrator(s) realises that the network of the company has been breached and some confidential information has been accessed or manipulated, action must be taken urgently. When this happens, the system administrator(s) must inform the information manager and the director(s) or owner(s) of the organisation about the issue. Then management must decide whether to take legal action against the offense or not.

If the management agrees to take legal practice, then the information manager will hire practitioners if the organisation does not have its own. Upon receiving this information, the leading practitioner will discuss with the business leaders the impact the investigation may have on the organisation. If management agrees on the prosecution of the cyber-criminals, then the 'preparation' of the investigation begins. Under the preparation process, the practitioners have to formulate approach and prepare necessary activities. When formulating approach, initially laws and regulations have to be considered and the leading practitioner must then obtain a legal note from law enforcement. The practitioners must meet the criteria on the legal note and scope of the investigation; basically the considerations are law and technical based. Also, the main goal of the investigation must be recovering legally admissible evidence without interrupting business processes. Thus, if the legal notice allows the removal of the network devices the leading practitioner must make sure that this minimises the impact on the running of the business.

When a robust approach is achieved, the leading practitioner and the team have to prepare for the investigation. The leader must make sure that tools, equipment, building an appropriate team, assigning duties and accounting materials for packing evidence are ready. Also, the practitioners have to explore the victim organisation and ensure that the operations and infrastructure of the organisation can sustain the investigations.

The team will then detect the intrusion; if an intrusion occurred the leader will have to obtain a warrant of search from law enforcement and search for evidence. When the evidence is found, then the leading practitioner will meet with the management of the organisation. If given consent, then the investigations will continue. Then the system networks have to be restricted to capture the proper metrics including intrusions and disruptions. Subsequently, the practitioners will embark on data recovery which has to be conducted through setting targets, enforcing targets, recovery of data, filtering evidence into folders and serving them in the DEBs, and lastly verifying whether there are no important log entries left uncopied.

If satisfied with the recovered evidence, the DEBs and seized devices (if any) will then be transported to a private laboratory for interrogation. Practitioners must clean the data by removing unnecessary entities, identifying transaction, identifying sessions and integrating matching transactions and sessions. All these are conducted as a pre-processing process. Then, the practitioners have to discover the patterns of how the cyber-criminals manoeuvred within the computer networks. When the patterns are discovered they will be analysed to determine the source of attack and locate the responsible person.

When enough evidence is acquired, the evidence will then be presented in a court of law. The model used will be questioned in the court; therefore, the use of a well-defined and widely accepted model is advantageous. If the model used is defeated, then the model will need to be improved. This marks the end of the log forensic investigations.

6.4 Evaluation of the Proposed Model

Like all other models proposed, the evaluation of this model is of great importance; thus, showing its benefits and its limitations. Also, it is a recommendation of the Design Science theory. Under the evaluation of the model, the distinctiveness of the model as well as the value the proposed model adds to the discipline of digital forensics, specifically in computer networks and academics, has to be illustrated. The proposed log file forensic model is derived from other already existing digital forensic models. It is also very important to mention that there are other models which define the same procedure; however, the model in this research project is unique in the sense that it is specifically for network log mining.

What is crucial about the proposed model is that it explains all the processes of the digital forensic investigation from beginning to end and it stresses laws and regulations as well as forensic precautions. The investigation precautions as well as laws and regulations must be observed since the heart of a forensic investigation is to provide legally admissible evidence in a court of law. From what have been reviewed in the literature most of the existing models do not mention the importance of forensic investigation precautions as well as laws and regulations which are taken to be an absolute aspect of an investigation; however, Ieong (2006) states that to effectively illustrate a forensic process the laws must be explicitly stated and recognised. Additionally, some of the practitioners are aware that they are supposed to follow some of the mentioned proposed precautions during an investigation; nevertheless, there is little mention in previous works (models) about these forensic precautions. Without laws and regulations and forensic precautions, this study states that the investigations will remain questionable in a court of law. The benefits and limitations of the proposed Log File Digital Forensic Model are discussed in following sub-sections.

6.4.1 Benefits

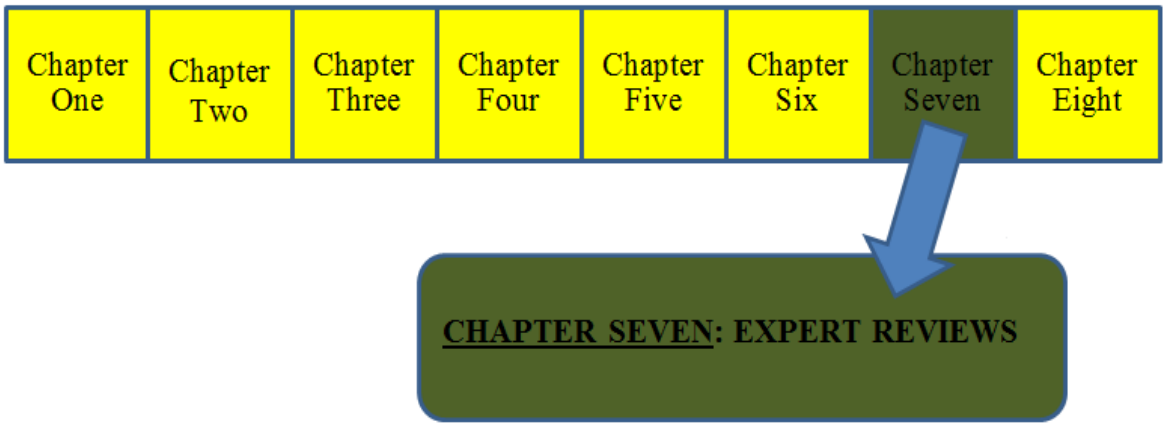
One of the key benefits is that the proposed model clearly designates the forensic precautions as well as laws and regulations as one of the chief characteristics of the model. This is very important as most of the already existing digital forensics identify these two characteristics as minor. In addition, the model clearly defines stages practitioners are supposed to follow in the whole process of network log mining. Thus, the process of recovery is clearly defined showing all the sub-processes specifically for network log mining. Also the interrogation process describes in depth how the process should be conducted exclusively for network log mining. Moreover, the analysis process particularly discusses the issue of correlation, which is typical in log mining. Furthermore, there are also some processes within the proposed model that can be repeated if enough evidence is not acquired. Some researchers, software engineers and security force agencies recognise these features but they were not pointed out in previous forensic models. Lastly, this model allows for expansion by other researchers.

6.4.2 Limitations

The proposed Log File Digital Forensic Model is a high-level conceptual artefact which, similarly to other digital forensic models, is an abstract one that is normalised to accommodate the comprehensive scale of forensic investigations. However, this does not weaken the value that is benefited when using the proposed model. Finally, the laws and precautions are listed as obvious characteristics; nevertheless, there are no conventions that bind the forensic investigators to strictly use or follow the laws and precautions.

6.5 Conclusion

This chapter introduced a new model to be used in computer network investigations when focusing on extraction and analysis of digital evidence from log files of computer networks. When conducting digital forensic investigations the evidence has to be collected in a legally admissible standardized manner. The proposed Log File Digital Forensic Model was discussed in detail; each and every aspect of the model was discussed precisely. Each layer must be completed before commencing processes in another layer. This research project stresses that forensic precautions as well as the law and regulations must be significantly considered in network log mining. Lastly, the benefits and limitations associated with the proposed model were elaborated.



7.1 Introduction

This chapter provides the analysis of the primary data collected during this research project. The analysis handled in this research project is discussed dichotomously. Firstly, the opinions of IT experts towards the proposed model will be analyzed. The opinions of the experts cover the large part of this chapter. Secondly, the appraisal of the experts and the improvements of the model from the experts' opinions will be discussed. It is important to mention that the proposed model was created entirely from the literature reviewed following a Design Science methodology in the creation of the artifact. Also, this study claims that through the Diffusion of Innovations (DOI) theory, after its publication, the model will infiltrate into the industry. The analysis of the expert reviews provides professional views towards the proposed model. The expert review process also helps verify the credibility of the study's findings. Towards the end of the chapter, findings are interpreted from the expert reviews.

7.2 The opinions of IT Experts

As discussed in chapter 5, the technique used for the validation of the proposed Log File Digital Forensic model was Expert Reviews. The model accompanied with its description in the form of a structured paper was sent to various local as well as international IT experts. This was done in recitations. In each round, the model was sent to five experts, three local experts and two international experts. Thus, after sending the model to a certain group of experts, their reviews were used to modify the model and then sent again to another group. This provided the responses from the South African as well as international perspective. The questions asked are as follows:

1. Considering the **originality aspect**, do you think the investigation procedure proposed in this paper adds value to the academic field?
2. Please provide your views on the **technical quality** of the model.
3. What is your opinion on the **presentation** and explanation of the model (procedure)?

In addition, it is important to highlight some of the ethical issues emphasised for the validation process. Some of them are as follows:

1. The information obtained will be used for the sole purpose of this research project.
2. No names of the respondents will be published.

The following paragraphs provide an account of the responses, both negative and positive, about the proposed model.

Round One

Three experts responded in the first round. The first expert remarked that the proposed Log File Digital Forensic model provided clear and interesting information towards the conviction of cyber-criminals. The expert also noted that the model was of high importance as it provided a clear etiquette towards log mining as most of the already published forensic models were too broad and unspecific, hence, not providing a direct way of conducting network log mining investigations. Nevertheless, the expert mentioned that the model had to be published in order to gain more support and critiques of other researchers. Also, the expert noted that the diagram of the model had to be large enough to make sure any reader will understand the summary of the whole model before reading its architecture.

The second expert remarked that the model was presumably a novel process mapped to the network forensic area, failing to prove itself as original. In addition to that, the expert asserted that *"the presence of responsible law enforcement personnel must be mentioned during the extraction of the evidence to keep the original evidence at a trust for future disputes"*. However, the expert mentioned that the presentation of the structure of the model and the explanation of its architecture was logical and provided a standard procedure of conducting network forensics.

The third expert commented that the model presented a step-wise approach to examining log files so as to maintain the integrity of the files and investigating them in a legitimate manner. The expert also noted that the model covered all aspects from the presentation stages to the various aspects of discovery and analysis. Moreover, the expert acknowledged that the authors substantiated their findings using relevant references. Furthermore, the expert mentioned that the model presented a well thought out approach to examining log files in a forensically sound manner.

Round Two

Five reviewers responded in the second round. The fourth expert, who was a forensic expert in a consultant organisation, remarked that even though the presentation of the model and the architecture sounded so realistic, it was difficult to picture the proposed procedure in the industrial forensic practices as the model was too theoretical and overlooked some crucial points which investigators frequently confront in the industry. Nevertheless, the expert pointed out that the issue of forensic precautions sounded important and powerful as legislation in some countries behind in technology focuses on prosecuting criminals after intrusion, stealing of information and invading business privacy, with little support on the activities behind the process of gathering the evidence from digital sources.

The fifth expert did not agree with the proposed model; however, the expert had no faith in most of the existing forensic models. According to the fifth expert, existing forensic models are not realistic as he commented that *"as far as I am concerned all the models projected so far are imagined theories which are hard to put in practice because most of the researchers are not aware of the situation in the industry"*. Nevertheless, the expert pointed out that the issue of forensic precautions was important and could help if one could research more on them and then publish them in a separate paper discussing them in detail. The expert also remarked that the model was very unique and could not be disputed that it was an improvement of old models while involving most of the important aspects around log mining.

The sixth expert noticed that the proposed model encompassed crucial forensic steps that could be followed easily. In addition, the expert expressed that the procedure was well

detailed, and agreed with the fifth expert that the forensic precautions should be discussed in detail in a separate document. The expert also remarked that the model focused on a theme (log files) that had never been discussed in most of the previous digital forensic models and as such provided a procedure specific to log mining. Also, the expert acknowledged that the model provided new idea of the creation of forensic procedures specifically for some forensic scenarios, of which existing models lack.

The seventh expert agreed with the second expert that the model was created from imaginary thoughts brought into the digital forensic field pictured in a hypothetical situation within forensics investigations. The expert was quoted saying *"it can be seen that the model was created from literature without any knowledge of what happens practically, hence, the little practical steps are illustrated on the model"*. In addition to that, the expert mentioned that even though the precautions sounded like a new important idea in the forensic field, there was no guarantee that they would be followed, and if not observed there was nothing that could be done to disqualify the evidence as long as the evidence was collected in the presence of a law enforcement personnel.

The eighth expert remarked that the model was described and discussed in detail. The expert also, acknowledged that the model illustrated the procedure crucial in the forensic industry especially in the auditing, banking and financial institutions. Furthermore, like the first expert, the expert also mentioned that the model was well supported by relevant references. Nevertheless, the expert was of the opinion that like most of the models already proposed, the proposed model lacked originality as it was constructed from existing models which do not have originality. The expert further stated that forensic models are assumed procedures.

Round Three

In the third round three reviewers responded. The ninth expert added that "*I find the presentation very helpful to fully comprehend the depth of the model and its various layers. All of these are well explained, defined and presented in a format which is easily understandable to anyone*". The expert was so impressed by the proposed model and said nothing negative about the procedure. This was one of the few opinions which did not point out any flaws associated with the model.

The tenth expert remarked that the proposed model looked similar to most of the already existing models including the one developed by him but well improved to suit the targeted scenario. The expert was also impressed by the projection of forensic precautions. Additionally, the expert pointed that the proposed model apparently had necessary phases that required an investigation to be considered complete. Also, the expert suggested that the proposed model must be published so that it could contribute to the field of Information Systems and after publication the model should be tested in the industry.

The eleventh expert asserted that the proposed model held crucial information in the field of digital forensics. Additionally, the expert pointed-out that the architecture of the model was clearly defined and well-presented. The expert remarked that the model proved its uniqueness in the sense that it focused on a single aspect, network log mining, and the aspects marked as new on the model were remarkably in line and ensembles of log forensic investigations.

Round Four

In the fourth round also three reviewers responded. The twelfth expert noted that the model was very relevant and of high significance in the field of Information Technology. Also, the expert remarked that the architecture of the model was easily understandable; and all necessary aspects of a complete forensic model were presented. The expert was quoted as saying "*I think this model and its architecture are well-written and covers a topic of interest to the academics*". Nevertheless, the expert mentioned that the originality of the model was low and was not convinced that the model would have a high level of impact on practitioners.

The thirteenth expert pointed-out that the architecture of the model was clearly described in a logical manner. Also, the model was of high significance and relevant in the forensic field. Nonetheless, the expert says the originality of the model was low. The expert also pointed out that the description of the model proposed some presumptions which were hard to prove.

The fourteenth expert commented that the model was relevant towards forensic investigations, with the originality aspect ranked medium. The expert mentioned that the model was described providing a complete forensic procedure. Additionally, the expert pointed out that the architecture of the model was very plausible. However, the expert commented that this model was marginal because it did not provide enough information concerning log mining and correlation of timestamps.

7.3 Analysis of the Expert Reviews

The analysis of the expert reviews was done through mind mapping technique as a fifth round of examining opinions of the experts. Thus, after every round, the model was improved; making it more meticulous, according to the feedback. After the first round, the diagram of the model was improved to appease the apprehension of the first expert. Also, this study encompassed the issue of involving the presence of law enforcement personnel to keep the original evidence in confidence for future disputes as per the concern of the second expert. Some of the experts, such as the fifth and the tenth experts, suggested the publication of the model as well as a separate publication of the proposed ***Forensic Investigation Precautions*** in detail. However, this study intends to publish this model at the end of this research project.

Some of the experts, such as the thirteenth and the fourteenth, emphasised that this study would sufficiently address the issue of detecting and accounting for log/evidence tampering as part of the incident. Additionally, they suggested that sufficient information had to be added indicating those aspects of the model that differed from other models. Furthermore, these experts advised to thoroughly incorporate contributions in timestamp evidence/analysis, the error rate and correlations of timestamps.

From the opinions of the experts encapsulated in section 7.2, this research project concluded that some experts do not have confidence in a model that has not been tested in the industry. This means that, after the publication of the model, it has to pass through an industrial scenario solved already. The model has to secure the same results achieved previously for it to be regarded as a comprehensive model. However, some of the experts do not credit forensic models as genuine, but rather abstract theories.

In summary, the proposed model faced critical evaluation. One of the most important points raised was that the model could not be respected as widely acceptable unless it was published and tested. Nonetheless, most of the experts supported the model as well as the way it was presented. The next section provides the findings which were drawn from expert reviews.

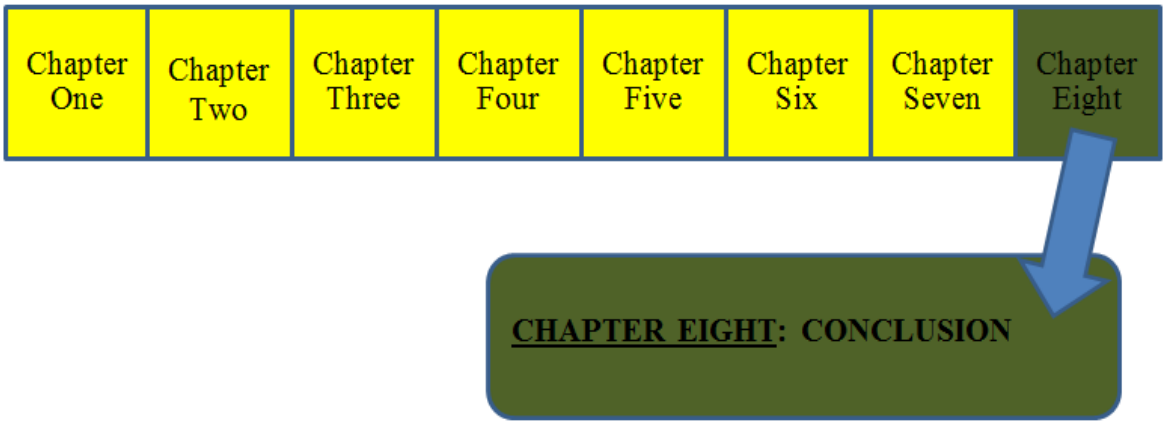
7.4 Findings

The proposed Log File Digital Forensic Model represents a logical way of conducting log forensic investigations on computer networks. From the above assessments, it can be determined that the proposed model provides a useful foundation of comprehending the procedure of network log mining. The procedure and the layout of the proposed Log File Digital Forensic Model was arranged in such a way that it presents what it was developed for. Therefore, according to the reviews the research project met its objectives. Nevertheless, the reviews show that the proposed model has some shortcomings including that there are no directives which force forensic practitioners to abide by forensic precautions. Also, reviews from experts claim that the proposed model cannot be regarded as a widely accepted model until it has been applied in the industry and considered workable. However, all models are firstly conceptual and later tested in the real environment.

From all the opinions of the experts this research project concludes that the model that can be followed in any digital forensic investigation has to be the one that incorporates necessary phases/processes towards a specific scenario of digital investigation, encompassing respective laws and regulations. Also, a model needs to be published and received well in the digital forensic community. Additionally, the model should be applied and proved to be working well in the industry while providing clear and rigid directives which any forensic practitioner can follow. Furthermore, the model must secure the same result when the investigations are repeated. Lastly, the model has to be flexible to be used and fit anywhere in the world. Therefore, the Log File Digital Forensic Model must have above mentioned propertied and has to pass through tests for it to reach a point of being regarded as a widely accepted Log File Digital Forensic Model.

7.5 Conclusion

This chapter provided the analysis of data collected during this research project. The proposed model was entirely developed from the literature reviewed through Design Science methodology. The process of the validation of the model passed through four rounds of expert reviews. After each round the model was improved following the ideas and opinions of the experts. The findings were drawn from the opinions of the experts; and from these findings, the answer to the third research sub-question was answered. The following chapter provides the conclusion of this research project in the form of a summary of the project, future research and concluding remarks.



8.1 Introduction

This chapter serves as the conclusion of this research project which introduced a digital forensic model called A Log File Digital Forensic Model in chapter 6. The model was developed after this research project discovered that cyber-criminals were utilizing computer networks through the Internet to invade or breach into other organisations. Cyber-criminals practise such activities knowing that they can evade the law because of their anonymity. Network forensics is an evolution of typical digital forensics whereby evidence is gathered from network devices such as firewalls, switches and routers by IT security and network practitioners, gathering evidence for court cases. Log mining is a fruition of network forensics where forensic practitioners have to follow the digital footprints of cyber-criminals until they reach the used device(s) and find the particulars of the individual(s) who used the machine at the time of the incident. The previous chapter discussed experts reviews gathered during the research towards the establishment of the proposed model.

8.2 Discovery

The literature examined in this research project shows that network log mining is an important activity vital to most of today's business organisations. Chapter 2 defined network log mining and discussed its significance in a number of organisations. In addition, network log mining in respect of applicable laws and regulations was discussed relative to South African presiding laws and regulations which support digital forensic investigations. The presiding laws and regulations discussed were the South African Computer Evidence Act, 57 of 1983; Electronic Communications and Transactions Act, 25 of 2002; and Electronic Communications Act, 36 of 2005, not withholding

international laws and regulations which also support network log mining such as Sarbanes-Oxley Act, Foreign Corrupt Practices Act (FCPA) and the Bribery Act of the USA.

Furthermore, chapter 2 explained how the evidence from log files is viewed in a court of law. Two heuristics namely the Frye Test and the Daubert Test are usually applied in a court of law when judging crimes which involve evidence from digital sources. The Daubert test has become more important than the Frye test and is the one commonly used in most courts of law throughout the world today. Lastly, case studies concerning log mining and network forensics showing the significance of the study, were discussed towards the end of chapter 2.

Challenges confronted by forensic practitioners were deliberated in chapter 3. The challenges were discussed under five categories namely: forensic tools; methods and procedures; acquisition of digital evidence; laws and regulations; and presentation of evidence. This research project asserts that unless these challenges are resolved, it will remain difficult generate evidence admissible without dispute in a court of law through digital forensics. This study falls under methods and procedures category as it proposes a model specifically for use in network log mining.

IT security standards were highlighted in order to make it clear that forensic investigations involving network log mining must be conducted in accordance with Information Technology Security Standards (ITSSs). The standards include ISO/IEC 27002, the SAS70 Type II, GLBA, PCI DSS, EU Privacy, CobiT, Common Criteria, Generally Accepted Privacy Principles (GAPP) and the Generally Accepted Security Principles (GASP). This study followed the ISO/IEC 27002 because it is a security standard that has the capability of improving information security when applied in every area of business organisation. Section 10.5.1 and 10.6.1 of the ISO/IEC 27002 support the process of network log mining.

In chapter 4 this study discovered many digital forensic models that have been proposed throughout the world; however, most of them are too generic. The models are broad-based not clearly illustrating the procedure which practitioners should follow when log mining

digital evidence from log files of computer network devices. Even though most of the models were discovered to be generic, it was also discovered that the existing models can be classified into three categories: those developed by academia and software engineering people; security force based forensic models; and forensic models with laws and regulations as a strong foundation. Some commonalities were identified and this research project concluded that authors of the examined forensic models borrowed ideas from different models from distinct categories.

8.3 Research Methodology

Design Science research methodology and its methods were followed in the development of the proposed model. The process of the development of the proposed model falls under the "*Reality as a Contextual Field of Information"* category along the continuum of core ontological assumptions. This reality allows continual process of interaction and exchange information where man is an information processor in order to map context based on cybernetic metaphors. While the model was being developed, this research project relied on the Diffusion of Innovations Theory as a premise on which the model could be received in the society or industry.

The proposed model was created entirely from the literature reviewed and critical thought. After the model was developed, it was then sent to ITSEs for validation. The opinions of the experts were summarised, analysed and the findings were derived from those opinions. The first round of validation was responded by three experts. The second round was responded by five experts; the third round was responded by three experts; and lastly, the fourth round was responded by three experts. In total, fourteen of twenty experts responded. All the experts provided their critical reviews about the proposed Log File Digital Forensic Model. All the suggestions, comments and feedback provided by the experts were considered and used accordingly. The feedback was mostly favourable in terms of relevance of information, presentation of the model, suitable references and architecture of the model. The validation of the model provided the qualitative primary data aspect towards this research project.

The secondary data collection included literature acquired from the Internet, journal articles, proceedings of conferences, reports, books and past research projects. The

literature reviewed determined the problem and research objectives. The secondary data collection stage provided the body of knowledge in the discipline of Information Systems. It is important to mention that the model developed in this study was published (see Appendix C).

8.4 Discussion

The proposed Log File Digital Forensic Model was presented and an in-depth discussion of the architecture of the model was provided thereafter. The model is a product of the Design Science approach towards digital forensic investigations. The merits and drawbacks of the model were discussed after the demonstration of the model in chapter 6. The criticism of the model was largely on the aspect of reality as some of the experts suggested that the model is conceptual and may not reflect actual steps followed in the industry. Nevertheless, this research project intended to develop a high-level procedure that can be followed in network log mining.

The research objective and the sub-questions defined in chapter 1 of this research project were met. Three sub-questions were developed in order to satisfy the research objective. The research objective is:

- To produce a standardised model that will aid digital forensic practitioners in the retrieval of log entries from the log files of computer networks (network log mining). Subsequently, it would assist with the correlation of log entries in a legally sound manner to ensure that the evidence was admissible in a court of law.

The main research question and how each of the sub-questions were met is as follows:

- What can be done to improve the process of recovering log file entries from an organisation's computer networks, in order to provide legally admissible evidence in a court of law?

i. What are the challenges faced in computer network investigations with respect to the recovery of log file entries from computer networks and the presentation of legally admissible evidence?

Resolution: This question was deliberated in chapter 3 as the challenges facing forensic practitioners in network log mining were identified. These were grouped into forensic

tools, methods and procedures, acquisition of evidence, laws and regulations, and presentation of evidence.

ii. What commonalities exist in models that have been used for log mining in computer network forensic investigations?

Resolution: This question was discussed in chapter 4 where the models, processes and frameworks currently followed in network investigations were presented, scrutinised and criticised. The models were grouped in three categories; those developed from the perspective of academia and software engineering; those developed from the perspective of security forces; and lastly, models with laws and regulations as the base of quality investigations. As these three categories of models were deliberated, the commonalities in these models were exposed.

iii. What procedures can be followed to ensure comprehensive digital evidence is collected correctly when conducting forensic log mining in computer networks?

Resolution: This question was answered in chapter 7 after the analysis of the opinions of the experts towards the proposed Log File Digital Forensic Model. The study developed this model with the intentions of producing a quality procedure and acquiring important ideas about the model from experts. Through their opinions, the experts revealed the qualities of a model required to ensure comprehensive digital evidence when conducting network log mining. Findings were drawn from those opinions and the answer to this question was derived from the findings.

8.5 Future Research

The findings, limitations and scope discussed above suggest a number of possible extensions to this research project. These include enactment and application of the Log File Digital Forensic Model in a practical and genuine log mining forensic case, preferably a case already dealt with to verify if the appropriate record of the procedure was accomplished. This approach would highlight weaknesses associated with the proposed Log File Digital Forensic Model so that unnecessary steps and mismatches can be revealed to conceal unanticipated variances that can arise in the future. This would be in accordance with expert opinions that the proposed Log File Digital Forensic Model should be applied in an industrial situation. In addition, there is no directive(s) that obliges forensic

practitioners to follow proposed forensic precautions; therefore, in future, possible fostering of these directives has to be developed otherwise the practitioners may run the investigations without following the necessary precautions.

8.6 Concluding Remarks

In summary, the objective of the study was achieved. This research project was situated in the context of log mining investigation, which is not discrete from traditional digital forensic investigation procedures, but it went on to extrapolate segments which are projected to be followed specifically in network log mining. This project recognised the nonexistence of correlation in the digital forensic process contained in the scope of network log mining. A number of researchers, software engineers and IT security forces concentrated on the issue of normalisation of the digital forensic procedures in many countries world-wide as illustrated in the analysis of existing forensic models. This research project attempted to add knowledge to current literature on the normalisation of the process of digital forensic investigations. This research project emphasised the need for forensic investigations to follow a repeatable and common practice using a chain of custody that will stand up to legal scrutiny.

REFERENCES

Abrahams, L. (2010). *Research Design and Methodology*. Retrieved November 24, 2010, from Wits University Graduate School of Public and Development: http://link.wits.ac.za

Adams & Adams. (2000). *Evidence-the basic revisited.* Retrieved October 6, 2010, from Adams & Adams: http://www.adamsadams.com

Agarwal, R., Ahuja, M., Carter, P. E., & Gans, M. (1998). Early and Late Adopters of IT Innovations: Extensions to Innovation Diffusion Theory. *Proceedings of the DIGIT Conference* (pp. 1-18). Florida: Florida State University.

Altman, D., Burton, N., Cuthill, I., Festing, M., Hutton, J., Playle, L. (2006). *Why do a pilot study?* Retrieved November 24, 2011, from National Centre for the Replacement, Refinement and Reduction of Animals in Research: http://www.nc3rs.org.uk

Alturki, A., Gable, G. G., & Bandara, W. (2011). Develping an IS-Impact decision tool: A literature based design science roadmap. *European Conference on Information Systems (ECIS).* Helsink, Finland : ICT and Sustainable Service Development.

Amenya, M. (2004). *Recovering, examining and presenting computer forensic evidence in court.* Retrieved June 17, 2010, from www.csam.montclair.edu

Arthur, K. K., & Venter, H. S. (2005). An Investigation into Computer Forensic Tools. *Information and Computer Security Architectures (ICSA) Research Group.* Pretoria: ISSA. University of Pretoria.

Ayers, D. (2009). A second generation computer forensic analysis system. *Digital Investigations*, 6(1), 34-42.

Ayre, L. B. (2006). Data Mining for Information Professionals. San Diego, California, USA.

Ball, C. (2005). *Six articles on Computer forensics for Lawyers.* New York, Mondgomery Texas, USA: Craig Ball.

Baryamureeba, V., & Tushabe, F. (2004). The Enhanced Digital Investigation Process Model. Kampala, Uganda.

BBB Accredited Business. (2010). *Computer and Digital Forensic Bachelor's Degree*. Retrieved July 08, 2010, from DegreeDirectory.org: http://degreedirectory.org

Beebe, N. L., & Clark, J. G. (2004). A Hierarchical, Objective-Based Framework for the Digital Investigation Process. *Digital Investigations*, 2(2), 146-166.

Bejtlich, R. (2006). *Forensically sound evidence.* Retrieved June 18, 2010, from TaoSecurity: http://taosecurity.blogspot.com/2006/08/forensically-sound-evidence.html

Bejtlich, R. (2007). *Network Security Monitoring History.* Retrieved July 15, 2010, from TaoSecurity: http://taosecurity.blogspot.com/2007/04/network-security-monitoring-history.html

Boal, K. B., Hunt, J. G., & Jaros, S. J. (2004). Order is free: On the ontological status of organisations. *Debating Organisations: Point/Counterpoint in Organisational Studies*. Lubbock, Texas, USA: Texas Tech University.

Boeckeler, M. (2004). *Overview of Security Issues facing Computer Users.* Retrieved March 3, 2010, from http://www.sans.org

Bosley Group. (2010). *What is mind mapping?* Retrieved July 21, 2010, from MINDMAPPER: http://www.mindmapperusa.com

Brandel, M. (2010). *Legal Hold Software: Features and Evaluations Criteria.* Retrieved May 12, 2010, from http://www.csoonline.com

Brenner, S. (2004). Toward a criminal law for cyberspace: product liability and other issues. *Journal of technology law and policy*, 5(1), 9-17.

Brown, C. L. (2005). *Computer evidence.* Retrieved April 6, 2010, from Charles River Media: http://www.charlesriver.com

Buchholz, F., & Tjaden, B. (2007). A brief study of time. *Digital Investigations,* 4(1), 31-42

Burgess Forensics. (2009). *Case Studies.* Retrieved August 21, 2010, from Computer Forensics, Electronic Discovery & Expert Witness: http://burgessforensics.com

Buys Incorporation Attorneys (BIA). (2009, March 11). *Email disclaimers explained.* Retrieved March 8, 2010, from http://www.buys.co.za/

Caloyannides, M. A. (2004). *Privacy Protection and Computer Forensics.* Norwood: U.S. Library of Congress.

Cardwell, K., Clinton, T., Cohen, T., Collins, E., Cornell, J. J., Cross, M., et al. (2007). *Best Damn Cybercrime and Digital Forensics book period.* United States of America: Syngress Publishing.

Carrier, B., & Spafford, E. H. (2003). Getting Physical with the Digital Investigation Process. *International Journal of Digital Evidence,* 2(2), 121-130.

Carrier, B. D., & Spafford, E. H. (2006). *Categories of digital investigation analysis techniques based on the computer history model.* Retrieved April 4, 2010, from http://www.dfrws.org

Casey, E. (2004a). The need for knowledge sharing and standardization. *Digital Investigation,* 4(2), 1-2.

Casey, E. (2004b). *Digital evidence and computer crime: forensic science, computers and the internet.* UK: Academic Press.

Casey, D. (2008). Turning log files into a security asset. *Network Security,* 2(2), 4-7.

Chaula, J. A., Yngstrom, L., & Kowalski, S. (2005). A framework for evaluation of Information System Security. *ISSA Workshop* (pp. 1-11). Pretoria: ICSA.

Cheng, E. K., & Yoon, A. H. (2005). Does Frye or Daubert Matter? A study of scientific admissibility standards. Alexandria, Virginia, USA.

Chizoba, O. M. (2005). *Cyber crime.* Retrieved July 14, 2010, from www.takingitglobal.org

Ciardhuain, S. O. (2004). An Extended Model of Cybercrime Investigations. *International Journal of Digital Evidence*, 3(1), 1-22.

Clarke, R. (1999). *A Primer in Diffusion of Innovations Theory.* Retrieved April 29, 2010, from http://www.rogerclarke.com/SOS/InnDiff.html

Cohen, F. (2010). *Moving Toward A Science Of Digital Forensic Evidence Examination.* Retrieved 07 August, 2010, from http://www.experts.com/ShowArticle.aspx?Articleid=423

Collis, J., & Hussey, R. (2003). *Business Research: a practical guide for undergraduate and postgraduate students. (2nd Ed).* New York: Palgrave Macmillan.

Computer Security Institute NetSec (CSI NetSec) . (2009). The Logs & The Law. What is admissible in a court of law? San Diego, Califonia, USA

Congdon, J., & Dunham, A. (1999). *Defining the Beginning: The Importance of Research Design.* Retrieved April 18, 2010, from http://www.iucn-mtsg.org

Custer, W.L., & Vogel, V. (2010). *Communications and Operations Management (ISO27002).* Retrieved June 22, 2011, from http://privacy.med.miami.edu/

Dagada, R., Eloff, M. M., & Venter, L. M. (2009). Too many laws but very little progress! Is South African highly acclaimed information security legislation redundant? *ICSA* (pp. 1-20). Pretoria: ISSA.

Das, R., & Turkoglu, I. (2009). Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. *Expert Systems with Applications*, 36(3), 6635-6644.

Data Recovery Services. (2010). *Computer Forensic Case Studies.* Retrieved September 3, 2010, from Where technology meets the courtroom: http://www.datarecovery.net

Daubert v. Merrell Dow Pharmaceuticals, Inc 1993. 509 U.S. 579, 589. (n.d.). *U.S. Supreme Court.* Retrieved June 14, 2010, from FindLaw: http://caselaw.lp.findlaw.com/

Dazey, A. (2009). *Electronic records and legal admissibility :Version 1.* Retrieved April 18, 2010, from Records Management Section: University of Udinburgh: http://www.recordsmanagement.ed.ac.uk

Dazey, A., & Gryzbowski, A. (2009). *Electronic records and legal admissibility: Version 8.* Retrieved April 18, 2010, from Records Management Section: University of Edinburgh: http://www.recordsmanagement.ed.ac.uk

Deeptee, A. (2011). *Advantages of Data Mining*. Retrieved November 16, 2011, from Buzzle.com-Intelligent Life on the Web: http.buzzle.com/articles/advantages-of-data-mining.html

Dejnozka, J. (2004). Daubert vs. Frye:Logical Empiricism and Reliable Science. Orange, Caroline, USA.

Egger, A. E., & Carpi, A. (2008). Data: Analysis and Interpretaion. *Visionlearaning*, 1(2), 154-161.

English, L. (2000). *Assessing Information product specification and Information architecture quality: Assuming quality of Information Definition and Information Models.* Information Impact International, Inc.

Fei, B. (2007). *Data Visualisation in Digital Forensics.* Retrieved April 17, 2010, from http://upetd.up.ac.za

Freiling, F., & Schwittay, B. (2007). A Common Process Model for Incident Response and Computer Forensic. *Proceedings of Conference on IT Incident Management and IT Forensics*, (pp. 1-21). Germany.

Forte, D. V. (2004a). The Art of log correlation: Part 1. Tools and techniques for correlating events and log files. *Computer Fraud and Security*, 1(6), 7-11.

Forte, D. V. (2004b). The Art of log correlation: Part 2. Tools and techniques for correlating events and log files. *Computer Fraud and Security*, 1(7), 15-16.

Forte, D. V. (2009). The importance of log files in security incidence. *Network Security*, 3(7), 18-20.

Fujitsu. (2002). The Legal Admissibility of Electronic Documents: An Executive Briefing Paper. Slough, Berks, England.

Furneaux, B. (2010, May). *Diffusions of Innovations Theory*. Retrieved July 23, 2010, from http://www.fsc.yorku.ca

Gabel, M. (2007). *What is Design Science.* Retrieved July 7, 2010, from http://designsciencelab.org

Garcia, G. (2007). Forensic physical memory analysis: an overview of tools and techniques. *Seminar on Network Security* (pp. 1-6). Limongarcia: Publications.

Gardner, R. (2000). *Notification of Judgement: Kilgore v. Boyd (U.S.).* Retrieved June 17, 2010, from http://www.fact.on.ca/Info/pas/pasnote.htm

Gatti, M. M., Ogrady, C. R., & Morgan, O. F. (1997). Foreign Corrupt Practices Act. Los Angeles, Califonia, USA.

Gilder, G., & Vigilante, R. (2001). *Stop everything. Its techno-horror*. Retrieved April 01, 2011, from Kurzweil Accelerating Intelligence: http://www.kurzweilai.net

Giordano, J., & Maciag, C. (2002). Cyber Forensics: A Military Operations Perspective. *International Journal of Digital Evidence*, 1(1), 1-13.

Grabosky, P. N., Smith, R. G., & Dempsey, G. (2001). *Electronic theft: unlawful acquisition in cyberspace.* Cambridge: University Press.

Grand, J., & Carrier, B. D. (2004). A hardware-based memory acquisition procedure for digital investigation. *Journal-Digital Investigation,* 1(1), 50-60.

Grobler, C. P., & Louwrens, B. (2006). Digital Forensics: A Multi-Dimensional Discipline. *ISSA Workshop* (pp. 1-12). Pretoria: ICSA.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in Humanities and the Social Sciences? *American Sociological Review,* 69(2), 190-212.

Hailey, S. (2010). *Digital Forensics Certificate.* Retrieved May 12, 2010, from http://infosec.edcc.edu/

Hak, J. W. (2003). *The admissibility of digital evidence in criminal prosecutions.* Retrieved September 22, 2010, from www.khodges.com/digitalphoto/hak.doc

Hamster. (2006). *The ISO17799 Newsletter - News & Views on the ISO/IEC Security Standard.* Retrieved 2010, from ISO17799 News: http://17799-news.the-hamster.com/issue10-news7.htm

Hasson, R. A. (2011). The Doctrine of Uberrima Fides in Insurance Law- A Critical Evaluation. *The Modern Law Review*; 32(6), 615-637.

Haugdahl, S. J. (2007). *Network forensics, methods, requirements and Tools.* Retrieved June 24, 2010, from http://www.bitcricket.com/downloads/Network%20Forensics.pdf

Hershensohn, J. (2005). IT Forensics: The Collection and Presentation of Digital Evidence. *ICSA* (pp. 1-14). Pretoria: ISSA.

Hevner, A. R., & March, S. T. (2003). The Information System Research Cycle. *MIS Quarterly*, 2(1), 111-113.

Hevner, A., March, S., Ram, S., & Park, J. (2004). Design Science in Information Systems Research. *MIS Quarterly,* 28(1), 75-105.

Hofstee, E. (2006). *Literature Review. In Constructing a good Dissertation.* Johannesburg: EPE.

Holden, M. T., & Lynch, P. (2004). Choosing the appropriate Methodology: Understanding Research Philosophy. *Irish Research Council for the Humanities and Social Sciences* (pp. 1-18). Waterford: Waterford Institute of Technology.

Ieong, R. S. (2006). FORZA - Digital forensics investigation framework that incoporate legal issues. *Digital Investigation,* 3(1), 29-36.

Independent Communications Authority of South Africa (ICASA). (2006). Electronic Communication Act, 36 of 2005. *Electronic Communications Act 36 of 2005* . Free State, South Africa: ICASA.

IT Service Strategy. (2008). *Comparison between ISO 27001, SAS70, PCI DSS, Cobit and Common Criteria free download.* Retrieved May 12, 2010, from IT Service Strategy: http://www.itservicestrategy.com/

Jajodia, S. (2008). *Intrusion Detection Systems.* Italy: Springer.

Jiagi, W., Shunxin, Y., Ming, L., & Jianchuan, C. (2005). Mobile data collection methods used in the pavement management system. *Proceedings of the 24th Southern African Transport Conference (SATC)* (pp. 912-920). Pretoria: Document Transportation technologies cc.

Jones, B. R. (2007a). Comment - virtual neighborhood watch: Open source software and community policing against cybercrime. *Journal of Criminal Law & Criminology*, 97(2), 601-629.

Jones, R. (2007b). *Safer Live Forensic Acquisition.* Kent: University of Kent at Canterbury.

Kent, J. (2010). *Computer Forensics Case Studies.* Retrieved June 20, 2010, from 7safe information security: http://www.7safe.com/computer_forensics_case-study.html

Kent, J., & Ghavalas, B. (2005). The unique challenges of collecting corporate evidence. *Digital Investigation*, 2(4), 239-243.

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). *Guide to Integrating Forensic Techniques into Incident Response.* National Institute of Standards and Technology , US Depaartment of Commerce. Gaithersburg: NIST Special Publication.

Kohn, M., Eloff, J., & Oliver, M. (2006). Framework for a Digital Forensic Investigation. *ISSA* (pp. 1-7). Pretoria: Information and Computer security Architectures (ICSA) Research Group.

KPMG. (2009). *E-Commerce and Cybercrime: New Strategies for Managing Risk of Exploitation.* Retrieved March 21, 2010, from http://www.uazuay.edu.ec/

KPMG. (2010). Electronic Discovery Management. *Risk, Compliances and Advisory* . Johannesburg, Gauteng, South Africa: KPMG (Pvt) Limited, SA.

Kruse, W. G., & Heiser, J. G. (2004). *Computer forensics incident response essentials.* USA: Addison Wesley, Pearson Education.

Kuechler, B., Park, E. H., & Vaishnavi, V. (2009). Formalizing Theory Development in IS Design Science Research: Learning from Qualitative Research. *Proceedings of Fifteenth Americas Conference on Information Systems.* San Francisco; California.

Kuhne, T. (2010). *What is a Model*? Darmstadt, Germany: Darmstadt University of Technology.

Kumho Tire CO. V. Carmichael. (1999). Supreme Court of the United States: on writ of centriorari to othe United States court of appeals for the eleventh circuit. Washington, DC, District of Columbia, USA.

Lai, X., Gu, D., Jin, B., Wang, Y., & Li, H. (2011). *Forensics in Telecommunications, Informations, and Multimedia.* Shanghai, China: Springer.

Lalla, H., & Flowerday, S. (2010). Towards a Standardised Digital Forensic Process: E-mail Forensics. *ISSA* (pp. 1-8). Johannesburg: ICSA.

Leigland, R., & Krings, A. W. (2004). A Formalization of Digital Forensics. *International Journal of Digital Evidence, 3(2)*, 1-32.

Lesssing, M. (2008). *Between Life and Death, Problems with live forensics.* Johannesburg: Council of Scientific and Industrial Research (CSIR).

Lessing, M., & Von Solms, B. (2008). *Live Forensic Acquisition as Alternative to Traditional Forensic Processes.* Johannesburg: University of Johannesburg (UJ).

Lim, M. J. (2008). *Computational Intelligence in E-mail Traffic Analysis.* Tasmania: University of Tasmania.

Mandi, J. L. (2010). *Mind Maps.* Retrieved July 22, 2010, from Mind Tools: http://www.mindtools.com/

March, S., & Storey, V. (2008). Design Science in the Information Systems Decipline: An introduction to the special issue on Design Science Research. *MIS Quarterly*, 32(4), 725-730.

Martha, R. (2010). *What is a forensic lawyer?* Retrieved May 14, 2010, from http://www.ehow.com

McCombie, S., & Warren, M. (2003). Computer Forensics: An Issue of Definition. *secau Conferences* (pp. 1-8). Australia: IT Security Investigation and Response, IBM Global Services.

Mettler, T. (2009). *A Design Science Research Perspective on Maturity Models in Information Systems.* Retrieved November 9, 2010, from http://ehealth.iwi.unisg.ch

Monteiro, S. D., & Erbacher, R. F. (2011). An Authentication and Validation Mechanism for Analysing Syslogs Forensically. *Management of Computing and Information Systems, Security and Protection* (pp. 1-10). Logan, UT: Department of Computer Science, UMC.

Morgan, G., & Smircich, L. (1980). The case of Qualitative Research. *The Academy of Management Review*, 5(4), 491-500.

Munk, M., Kapusta, J., & Svec, P. (2010). Data pre-processing evaluation for Web log Mining: Reconstruction of Activities of a web visitor. *Procedia Computer Science*, 1(1), 2273-2280.

Murr, M. (2006). *What is "forensically sound"?* Retrieved May 11, 2010, from Windows Incident Rsponce: http://windowsir.blogspot.com/2006/08/what-is-forensically-sound.html

Myers, M. D. (1997). Qualitative research in information systems. *MIS quarterly*, 21(2), 241-242.

Nelson, M. L. (2004). *The Applicability of i2i as a Supply Chain Management Tool in Facilities Management.* Salford, Britain, UK: University of Salford.

Nikkle, B. (2006). Improving evidence acquisition from live network sources. *The International Journal of Digital Forensics and Incident Response*, 3(2), 1-22.

Oates, J. B. (2006). *Researching Information Systems and Computing.* London: SAGE Publications.

Opsecure. (2010, March 20). *ISO/IEC 27002:2005 (ISO/IEC 17799:2005).* Retrieved July 14, 2010, from http://www.opsecure.co.uk

Pai, M. V., & Bhattacharya, D. (2006). Research Methodology. *The Journal of Obstetrics and Gynecology of India*, 56(6), 479-480.

Palmer, G. (2001). A Road map for Digital Forensic Research-Report from the First Digital Forensics Research Workshop (DFRWS) (Technical Report DTR-T001-01 Final). Utica: Rome Research Site.

Pati, P. (2003). *CYBER CRIME.* Retrieved August 1, 2010, from http://www.naavi.org

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., et al. (2006). *The Design Science Research Process:A Model for Producing and Presenting Information Systems Research*. Retrieved November 9, 2010, from http://iris.nyit.edu

Perumal, S. (2009). Digital Forensic Model Based on Malaysian Investiagtion Process. *IJCSNS International Journal of Computer Science and Network Security*, 9(8), 38-44.

Pladna, B. (2008). *Computer Forensics Procedures, Tools, and Digital Evidence Bags:What they are and who should use them.* Retrieved May 23, 2010, from http://www.infosecwriters.com

Pollitt, M. (1995). Computer Forensics: an Approach to Evidence in Cybercrime. *Proceedings of the National Information System Security Conference*, (pp. 487-491).

Ramabhadran, A. (2007). *Forensic Investigation Process Model for Windows Mobile Devices*. Retrieved May 11, 2009, from http://www.forensicfocus.com

Ray, D., & Bradford, P. G. (2007). *Models of Models: Digital Forensics and Domain-Specific Languages.* Tuscaloosa: University of Alabama.

Reith, M., Carr, C., & Gunsch, G. (2002). An Examination of Digital Forensic Models. *International Journal of Digital Evidence*, 1(3), 1-12.

Richard, G., & Roussev, V. (2006). Digital Forensic Tools:The Next Generation. *Communications Of The ACM*, 49(2), 76-80.

Richard, G., & Roussev, V. (2006). File Systems support for Digital Evidence. In M. Oliver, & S. Shenoi, *Advances in Digital Forensic II* (pp. 29-30). USA: SpringerLink.

Robinson, L. (2009). *A summary of Diffusion of Innovations.* Retrieved April 30, 2010, from http://www.enablingchange.com.au/Summary_Diffusion_Theory.pdf

Roger, K. M., Goldman, J., Mislan, R., Wedge, T., & Debrota, S. (2006). Computer Forensics Field Triage. *Proceedings of Conference on Digital Forensics, Security and Law*, 1(2), 27-40.

Rogers, E. M. (1995). *Diffusion of Innovations Theory. (4th Ed).* New York: Free Press.

Rondganger, L. (2008). *Hacker with a conscience.* Retrieved June 12, 2010, from www.iol.co.za

Ryan, D. J., & Shpantzer, G. (2005, April). *Legal Aspects of Digital Forensics.* Retrieved March 4, 2010, from http://www.danjryan.com/

Saboohi, M. (2008). *Collecting Digital Evidence of Cybercrime.* Islamabad: International Islamic University

Sansurooah, K. (2009). A forensics overview and analysis of USB flash memory devices. *Australian Digital Forensics Conference. 2*, pp. 1-11. Perth Western Australia: Edith Cowan University.

Schuster, A. (2007). Introducing the Microsoft Vista event log file format. *Digital Investigation*, 4(1), 65-72.

Selamat, S., Yusof, R., & Sahib, S. (2008). Mapping Process of Digital Forensic Investigation Framework. *International Journal of Computer Science and Network Security*, 8(10), 163-168.

Sentor, S. (2007). *Cyber Crime and Its Causes.* Retrieved September 2, 2010, from Associated Content: http://www.associatedcontent.com/

Seth, K. (2010). *India – Cyber crimes and the arm of Law – An Indian Perspective.* Retrieved September 2, 2010, from Seth Associates Advocates and Legal Consultants: http://www.sethassociates.com/

Sevaldson, B. (2010). Discussions and Movements in Design Research: A system approach to practice reearch in design. *FORMakademisk*, 3(1), 8-35.

Shebaro, B., Perez-Gonzalez, F., & Crandall, J. R. (2010). Leaving timing-channel fingerprints in hidden log files. *Digital Investigations*, 7(1), 104-113.

Shinder, D. (2002). *Scene of the cybercrime: Computer forensics handbook.* Rockland, Mass: Syngress Media.

Stander, A., Dunnet, A., & Rizzo, J. (2002). A survey of computer crime and security in South Africa. *ICSA* (pp. 1-10). Johannesburg: ISSA.

Stephenson, P. (2002). The Forensic Investigation Steps. *Computer and Fraud Security*, 10(2), 17-19.

South African Government Gazette. (2002). *No. 25 of 2002: Electronic Communications and Transactions Act, 2002.* Retrieved June 16, 2010, from http://www.info.gov.za

South African Standards (SANS). (2006). *ISO17799:2005.* Retrieved June 12, 2010, from Information Security Management: http://www.sans.org

Stephenson, P. (2003). A comprehensive approach to digital incident investigation. *Information Security Technical Report*, 8(2), 42-54.

Suherman, A. (2009). Data Mining. *Using Data Mining in new business organisations.* Columbia, United States of America: Harvard University.

Taylor, M., Haggerty, J., & Gresty, D. (2009). The legal aspects of corporate e-mail investigations. *Computer Law & Security Review*, 25(4), 372–376.

Thomas, R. (2001). Tracking Spoofed IP Addresses Version 2.0. *IP Spoofing* . Columbia, USA: ThomasRobin.

Tug, E., Sakiroglu, M., & Arslan, A. (2006). Automatic discovery of the sequential access from web log data files via a genetic algorithm. *Knowledge Based Systems*, 19(3), 180-186.

Turner, P. (2005). Digital provenance – Interpretation, Verification and Corroboration. *Digital Investigation*, 2(1), 45-49.

U.S Department of Justice (USDOJ). (2004). *Forensic Examination of Digital Evidence: A Guide for Law Enforcement.* USA: National Institute of Justice (NIJ).

Universal Service and Access Fund Regulations (USAFR). (2008). *Electronic Communications Act 36 of 2005.* Retrieved September 7, 2010, from http://www.mangaung.co.za

Van der Poll, H. M. (2004). *Book entries and a new information perspective.* Retrieved May 18, 2010, from http://upetd.up.ac.za

Velasco, V. (2000). *Introduction to IP Spoofing.* Chicago: SANS Institute InfoSec Reading Room.

Venable, J. (2010). Design Science Research Post Hevner *et al*: Criteria, Standards, Guidelines and Expectations. *Global perspectives in Design Science Research in Computer Sciences*, 6105(10), 1-16.

Vezina, R. (2008). *What is Cybercrime?* Retrieved September 1, 2010, from Articlesbase: http://www.articlesbase.com

Vaughan, D. (2008). *Clog: Client Logging WFP Edition*. Retrieved June 27, 2011, from The Code Project: Your Development Resource: http://www.codeproject.com

Walker, C. (2007). *Computer Forensics: Bringing Evidence to Court.* Retrieved November 26, 2010, from http://www.infosecwriters.com/text_resources/pdf/Computer_Forensics_to_Court.pdf

Wang, S., & Wang, H. (2010). Towards innovative design research in Information Systems. *Journal of Computer Information Systems*, 2(7), 11-18.

Wieringa, R. (2010). *Design Science Methodology.* Netherlands: Deutsche Telekom.

Yu, L. (2000). Studying E-Journal user behaviour using log files: The experience of SuperJournal. *Library and Information Science Research* , 22 (3), 311-338.

Zvekic, U. (2009). *Conventional Crime*. Retrieved October 6, 2010, from http://sociologyindex.com

APPENDICES

APPENDIX A: GLOSSARY

Abbreviations and Acronyms

7SISS	7Safe Information Security Service
AAA	Authentication, Authorisation, Accounting protocol
ASCII	American Standard Code for Information Interchange
BOD	Board of Directors
BSI	British Standards Institute
CEO	Chief Executive Officer
CSI NetSec	Computer Security Institute NetSec
CSIR	Council for Scientific and Industrial Research
DEBs	Digital Evidence Bags
DFRWG	Digital Forensics Research Working Group
DOI	Diffusion of Innovations Theory
DS	Design Science
DTI CoP	Department of Trade and Industry Code of Practice
ECT Act	Electronic Communications and Transaction Act
EIDIP	Enhanced Integrated Digital Investigation Process
EU	European Union
FCPA	Foreign Corruption Practices Act
FORZA	FORensic ZAchman
FTK	Forensic Toolkit
GAPP	Generally Accepted Privacy Principles
GASP	Generally Accepted Security Principles
GBP	Great Britain Pound (£)
GLB Act	Gramm-Leach-Bliley Act
GUI	Graphical User Interface
HRM	Human Resource Manager
HTTP	Hypertext Transfer Protocol
IA	Information Assurance
ICSA	Independent Communication Authority of South Africa
ID	Identity
IDIP	Integrated Digital Investigation Process
I/O	Input/Output
IP Address	Internet Protocol Address
IS	Information Systems
ISO/IEC	International Organisation for Standardisation/International Electrotechnical Commission
ISP	Internet Service Providers
IT	Information Technology
ITCs	Information and Communication Technology
ITSEs	Information Technology Security Experts
ITSSs	Information Technology Security Standards
KPMG	Klynveld Peat Marwick Goerdeler (accounting firm)
NCSA	National Computational Science Alliance

ODBC	Open Database Connectivity
OSI	Open System Interconnection
PCI DSS	Payment Card Industry Data Security Standard
RAM	Random Access Memory
SANS	South African Standards
UNIX	Uniplexed Information and Computing System
URL	Uniform Resource Locator
USAFR	Universal Services and Access Fund Regulations
USDOJ	United States Department of Justice
USA	United States of America
UTC	Coordinated Universal Time

APPENDIX B: DEFINITIONS

Important Terms

Data Mining:
Is the process of analysing data and extraction patterns from data from different perspectives and summarising into useful information (Deeptee, 2012).

Design Science:
Design Science is technologically oriented and is essentially a problem solving process that leads to a development of an effective artifact, which can fall into four categories, explicitly constructs, methods, models and implementations (Alturki, Gable & Bandara, 2011).

Electronic Evidence:
Electronic evidence is electronic data which is sought, located, secured, and searched with the intent of using it as evidence in a court of law (Dazey & Gryzbowski, 2009).

Interpretivist:
Assumes that access to reality (given or socially constructed) is only through social constructions such as language, consciousness and shared meanings (Morgan & Smircich, 1980).

IT Artefact:
IT artefacts are defined as constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices) and instantiations (implemented and prototype systems) (Peffers, Tuunanem, Gengler *et al*, 2006).

Model:
A hypothetical description of a complex process. A computer forensic model can be defined as a structure to support a successful forensic investigation.

Positivist:
Assumes that reality is objectively given and can be described by measurable properties which are independent of the observer and the instruments used (Morgan & Smircich, 1980).

Qualitative and Quantitative:
Qualitative refers to relating to or involving comparisons based on qualities. Quantitative refers to relating to the measurement of quantity.

Standardisation:
The condition in which a standard has been successfully established (SANS, 2006).

APPENDIX C: CONFERENCE PAPER

In this research project, a paper was published at the Eighth Annual IFIP WG 11.9 International Conference on Digital Forensics (January 2012). This paper is presented below:

A Log File Digital Forensic Model

Tendai Sanyamahwe
Department of I.S
University of Fort Hare
East London, South Africa
tsanyamahwe@gmail.com

Stephen Flowerday
Department of I.S
University of Fort Hare
East London, South Africa
sflowerday@ufh.ac.za

Paul Tarwireyi
Department of I.S
University of Fort Hare
East London, South Africa
ptarwireyi@ufh.ac.za

Himal Lalla
Department of I.S
University of Fort Hare
East London, South Africa
hlalla@gmail.com

Abstract - **Cyber-criminals utilise the Internet to intrude and launch attacks in different organisations' computer networks. A widely accepted digital forensic procedure would aid in convicting cyber criminals. However, this paper argues that the existing digital forensic models do not facilitate detailed specification towards the extraction of digital footprints based on log files in computer networks. This research paper proposes a digital forensic model that describes the various processes involved in the forensic investigation of computer networks focusing specifically on computer network log mining. Thus, a thorough examination of log files is needed to reveal the hidden actions which criminals perform in an organisation's computer networks. The proposed model emphasises the steps which forensic practitioners can follow on extraction and examination of digital evidence from log files of an organisation's computer networks so that the evidence can be used to convict criminals in a court of law.**

Keywords: ***digital forensic model; network forensics; log files***

I. INTRODUCTION

Network forensics is an evolution of typical digital forensics where the evidence is gathered from network devices such as firewalls, switches, routers and network management systems by IT security practitioners for court cases [1]. Cyber-criminals use computers connected to the Internet to penetrate organisations' computer networks. The attack traffic has to pass through network devices as the computer used to launch the attack and the targeted organisation are usually on different networks [2]. From these network devices the practitioners need to follow the digital footprints of the cyber-criminals until they reach the computer(s) used to commit the crime and further

investigations on those computers may reveal the particulars of the person or machine that was responsible for the attack [3]. If the network devices are configured properly, each attack action might leave digital footprints recorded in the log files of the devices [2]. Log files provide significant evidence in the computer networks when tracing cyber-criminals [3], because they record all the activities which take place in an organisation's computer networks [4]. Therefore log files can be considered an eyewitness to the attack and should be secured if there is a threat that they will be lost before the system is copied [14]. After tracking down the machine used for attacking, the practitioners have to correlate the logs found on both the used computer and those found in the victim computer networks [5]. Log file forensics involves all the procedures involved in trying to find out and prove that a crime has been committed.

The task of following the digital footprints and mining digital evidence from the log files of computer networks must be conducted following a well-defined process in the forensic field [6]. Network log mining is a process of discovering, extracting knowledge and modeling and analysing events recorded in the log files of computer networks as a way of convicting cyber-criminals in a court of law [7]. The rest of this paper is structured in the following units: section 2 discusses the importance of network log mining in security incident prevention, section 3 the proposed model and section 4 the conclusion.

II. NETWORK LOG MINING SECURITY INCIDENTS

A typical organisation has many different network devices which when configured correctly can create or store user log files and system activities [2]. In order to make sense of all the data provided in continuous streams of network devices, dedicated logging infrastructures (log file servers) have been developed to store log files from all connected devices for better management of the logs [2]. Some of the ways of logging include console logging, buffered logging, terminal logging, syslog, Simple Network Management Protocol (SNMP) traps and the Authentication, Authorisation and Accounting (AAA) protocol [8].

When people obtain access to an organisation's website through the Internet, the log files of computer network devices record all the activities each individual performs in the form of Who has done What and When [9]. Log files record each activity executed by the computer network devices and each line represents a record of the IP address, time and date of the visit (timestamps), accessed object and referenced object [3]. Therefore, log files are most important source of digital forensic investigations because they usually connect a certain event to a point in time [4]. Thus, the log files are an excellent source to check the network anomalies for insider threats, e-discovery for data leakages and malicious activities for IT asset misuse [7].

In addition, log files present the problem of unique identification of an intruder into an organisation's networks; hence, they are the source of anonymous data about the user [10]. The goal of using log files is to capture the behavioral patterns of individuals interacting with the organisation's computer networks which are then modeled and analysed by practitioners; thus, it is the method that gives significant evidence about how cyber-criminals navigate in the computer networks [9]. Consulting firms, for example KPMG, Deloitte and Touché, PWC, Ernst and Young and financial institutions such as banks regard log files on computer networks as important evidence to use when tracing cyber-criminals through computer auditing. In applying log files as a process of acquiring digital

evidence for court cases, the importance of a well-defined procedure is always emphasised [10].

Many digital forensics models have been developed around the world and the number of proposed models reveals how complex the process of evidence gathering in digital forensic investigations can be [11][6]. The models reviewed in this study include those developed by the Digital Forensic Research Working Group (DFRWG) [12], Reith, Carr and Gunsch [13], Carrier and Spafford [14], Baryamureeba and Tushabe [15], Ieong [16] as well as Perumal [6]. However, among all the models reviewed, there is no model that illustrates the actual processes or steps that practitioners may follow specifically in network log mining. This paper implies that examined forensic models are generic and practitioners may dispute on how to carry out some tasks within some steps and or phases in the process of log mining. This complicates regulatory compliance issues as practitioners, lawyers and judges may have different opinions about the procedure used, hence, making it difficult to provide legally admissible evidence in a court of law [17].

III. THE PROPOSED MODEL

The proposed model is broken down into four basic layers, with laws and regulations being the base of the model. Also, forensic investigation precautions are part of the model (see Figure 1). The four basic layers are preparation layer, discovery layer, testing layer and elucidation layer. Each layer has unique processes that are supposed to be run and completed before the initiation of the following layer. The model has a top-down layout which makes it simple, easy to follow and easy to comprehend. Processes within each layer follow the direction of the pointing arrows.

A. Preparation Layer

Thorough ***preparation*** is considered as one of the most important aspects before beginning the technical investigations of log files on computer networks. There are two processes involved in the preparation layer, namely formulating the approach and pre-incident preparation.

1) Formulate Approach

The organisation must to be aware [20] of the need for an investigation and ensure that the operations and infrastructure can sustain an investigation [15]. In order for an organisation to be aware of the need, there must be a trigger, which will initiate from an event/s that has compromised data/information and has violated the law. During the ***formulating approach process*** legal notice must be obtained from law enforcement and all concerned parties have to be notified about the forensic investigation [18]. The legal notice contains legal and technical considerations [19]. The impact the investigation may have on the business processes must be established; hence the main goal of the investigation of log files is to recover legally admissible evidence without interrupting business processes [17]; therefore, if the legal notice allows the removal of the network devices, this must be done only when it does not impact on the running of the business [17][19]. This is a risk assessment task performed at business level and a forensic readiness plan with the following goals: Recovering legally admissible evidence without interrupting business processes, keeping the cost of the investigation proportionate to the incident and ensuring that the evidence makes a positive impact on the outcome of legal action. All of these goals define the relationship with the events clearly as it impacts other

steps [37]. Therefore the Approach technique has to be robust so that the results can be convincing [13].

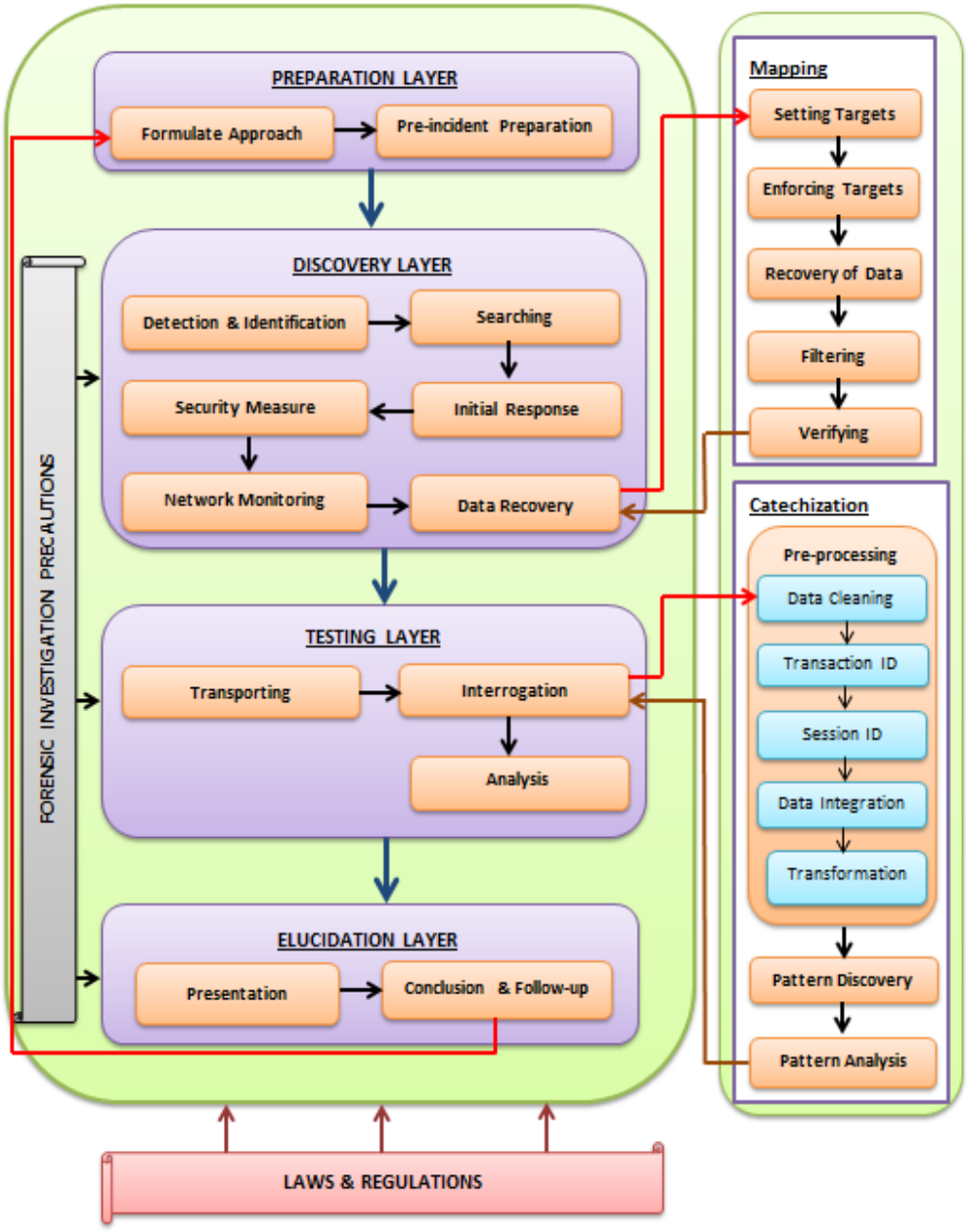


Fig. 1: A log file digital forensic model

2) Pre-incident Preparation

Preparing the direction of the investigation is particularly crucial. Without proper preparation, practitioners will run the investigation unsystematically; this can lead to excessive errors rendering the evidence weightless [20]. Furthermore, during the ***pre-incident preparation process***, an initial understanding of the nature of the crime and activities such as the preparation of tools and equipment; relevant skills and continuing to educate oneself about new technologies; building an appropriate team; assigning duties to each team member; accumulating the materials for packing evidence sources or retrieved log file entries; and legal coordination and general monitoring approval, are important

aspects to be taken into account [14]. Also, having a thorough pre-incident preparation process typically leads to high quality evidence and minimises the risks and threats associated with an investigation [18].

B. Discovery Layer

The ***discovery layer*** is the second layer and as shown in Figure 1, it follows the preparation layer. In this layer vital tests and experiments are carried out on the suspected network devices in order to identify the digital footprints of the criminals [2]. The goal is to provide a link between the suspect and the incident. Processes associated with this layer involve, detection and identification, searching for evidence, initial response, security measure implementation, network monitoring and recovery of data.

1) Detecting and Identification

Before rushing into the log files in the network devices, the forensic practitioners must first confirm the claim of intrusion, by using specialised tools, equipment and techniques [22], of cyber-criminals into the organisation's computer networks [15]. The signature based intrusion detection system or the anomaly based intrusion detection system are used for this task [23]. Signature based intrusion detection systems rely on pattern-matching techniques; they contain a database of signatures of known attacks and try to match these signatures with the analysed data [23]. When a match is found then that suggests that the computer network has been compromised. On the other hand, anomaly based detection systems first build a statistical model describing the normal network traffic, and then flag as an attack any behaviour that significantly deviates from the model [23]. The anomaly based detection analyses the payload data in transport layer and/or considers packets headers in network layer. Once the detection and identification process is completed, a warrant specifically for the detected incident will have to be obtained from the law enforcement before proceeding with the investigations [39].

2) Searching

The ***process of searching*** for the evidence involves evaluation of the crime scene, formulating the relevant search plan and searching potential sources of evidence [20]. The first responder identifies key pieces of evidence that contribute to the hypothesis of what crime has occurred [14]. In the process of searching, the leading forensic practitioner must not only evaluate the electronic equipment at the scene to determine whether any expert assistance is required in processing the scene but also identify people the scene and conduct preliminary interviews [18]. Furthermore, the owner(s) or system administrator(s) of the business have to provide valuable information such as the purpose of the system, security schemes, various applications present in the devices, user names, passwords and encryption details if available [21]. At this point, if it is important to search some information or items not listed on the warrant, then appropriate changes must be documented or else a new warrant must be obtained; failure to do so will result in questionability of the evidence in a court of law [19]. Therefore, documentation is not a specific phase but is required when the evidence is found; however, a chain of custody needs to be established early on in the investigation for the evidence is to be used in a court of law. The systematic search for evidence is a process where practitioners survey the physical and virtual crime scenes, generating the best methods of detecting and identifying the evidence [13].

3) Initial Response

In the ***process of initial response***, the leading practitioner has to discuss with the owners or leaders of the organisation the outcome of what has been detected and identified in the network system [14]. The process of initial response involves the case leader revealing the results of detection and identification process to the leaders or owners of the organisation verifying whether the claim made by the organisation's system administrator(s) about an intrusion is valid or not [24]. If the claim is untrue then the investigation will end at this point, but if the claim is true then the owners or leaders of the organisation have to respond by either giving the practitioners permission to continue with the investigation or not [24]. Once an attack has been detected the initial response is to secure the incident site, which includes closing off the computer from the network, and maintaining the integrity of log files in the system [14].

4) Security Measure Implementation

When permission has been granted to continue with the investigation the practitioners have to start with the ***security measure implementation process***. This process is required in some cases especially when the practitioners involved are part of the victim organisation's internal incident response team. Forensic practitioners advise the administrator(s) of the organisation on how to improve the security of the organisation's computer networks [13]. The practitioners are required to help by explaining the weaknesses or vulnerabilities used by criminals on the computer networks and how the security on the computer network devices can be improved [25]. The process of security measure implementation must be conducted just after the initial response process so that the weaknesses of the organisation's computer networks can be made clear to the administrators before moving forward with the forensic investigation [13].

5) Network Monitoring

The ***network monitoring process*** involves the monitoring and controlling of the flow of network traffic in the organisation's computer network(s) [14]. Thus, network monitoring is the securing of the incident networks, which involves reducing the amount of network traffic in the computer networks. This is a crime scene protocol that has to be followed in any computer crime arena [13]. The network monitoring process involves the securing of the incident site, hence maintaining the integrity of log files in the network system [14]. The goal of network monitoring is to reduce the rate at which the log files in the network devices are updated by new transactions taking place on the networks [26]. The log files of an organisation can record approximately 240 million entries per day [2], so to capture the proper metrics including intrusions and disruptions, the network system has to be restricted [3]. This reduces the tiresome work of searching for relevant entries amongst millions of entries in the log files during the evidence retrieval process [3].

6) Data Recovery

During the ***data recovery process*** practitioners have to rummage for the digital footprints of the cyber-criminals in the log files. If an organisation has a log file server(s) then the forensic practitioners have to focus on that; however, if not then they will have to check the log files of the individual suspected network devices [9]. In this process skilled and experienced personnel are required to perform the duty [26]. The process of evidence retrieval marks the first distinctive aspect of this model, specifically for network log mining, from other proposed forensic models. This process has some sub-processes called

mapping; these sub-processes are: setting targets, enforcing targets, recovery of data, filtering and verifying. The process of evidence retrieval is a very tiresome manual job [2].

Setting targets in network log mining requires practitioners to document what is supposed to be accomplished during the process of retrieving evidence [5]. In addition to that, all the steps must be numbered such that if there is a change in personnel the tasks will still be carried out in the same sequence [22]. After setting the targets, the practitioners must enforce the targets. By enforcing targets the practitioners make sure that all documented aspects are achievable [5]. When satisfied that the targets can be achieved, the practitioners are required to locate the batches of relevant log entries which can provide significant evidence about the crime from the log files.

Once the batches of relevant log file entries are found, they have to be copied into new directories [10]. The process of looking for batches of relevant log entries and the copying or imaging of those batches is called the recovery of data activity [10]. When the relevant log file entries are copied to respective directories, the directories can then be filtered into safe repositories, for example, digital evidence bags (DEBs) [4]. A DEB is software used to collect information for digital forensic examinations; thus, DEBs are universal containers for digital evidence from any source [27]. The verification process is conducted and involves rechecking the log files to verify that all relevant log entries have been copied [7]. If the practitioners are not satisfied that the retrieved log entries can prove, in a technically sound manner, that a crime has been committed, the process of evidence retrieval can be restarted before moving forward with the investigation.

C. Testing Layer

The ***testing layer*** is the third activity which comes just after the discovery layer as illustrated in Fig 1. The testing layer consists of the transportation and storage process, interrogation process and analysis of the digital evidence. The processes in this layer are conducted in private laboratories. The evidence is verified in order to find out the intentions and the actions of the criminals [18]. Also, the processes in this layer must be conducted in the presence of law enforcement agencies.

1) Transportation and Storage

When the practitioners are satisfied that the mapping process has been conducted comprehensively, the filtered evidence and seized devices, if any, have to be ***transported*** to a forensic laboratory for ***safe keeping*** for interrogation as well as analysis of the log entries [15]. This step ensures the integrity of the evidence and reduces the risk of evidence tampering [14]. Proper safety measures must be maintained because the evidence (filtered log file entries) on the DEBs can be destroyed while in transit due to shock, excessive pressure, humidity and temperature [18]. In addition to that, plastics should not be used to cover the seized network devices or the DEBs because they cause static electricity; however envelopes may be used [18]. Thus, during the transportation process the integrity of the evidence must be ensured so that the evidence remains valid because the filtered evidence has to be presented in raw form in a court of law [9]. Furthermore, the DEBs must always be kept in a conducive environment or place where there is no or very little electromagnetic radiation, dust, heat and moisture [18].

2) Interrogation

The ***interrogation process*** involves examining the contents of the retrieved log entries by forensic practitioners so as to acquire important information relevant to proving the case [18]. This is a process involving major tasks and this process marks another particular aspect of this model compared to other proposed forensic models. Most activities in this process are not involved in any of the previous models. During the interrogation process an in-depth exploration of the filtered log files is carried out. This involves the application of specialised digital forensic tools and techniques used to gather evidence and to further scrutinise the log file entries [11][8]. Analysis of such large amounts of data is made manageable by using layers of abstraction and more specifically, analysis of the network abstraction layer which translates the lowest level data from a physical network to the data used by an application [38]. At this point a number of back-ups of the filtered evidence must be created before analysing the collected log entries.

The process of interrogation must make the evidence visible by explaining its originality and significance [15]. Huge volumes of collected data need to be divided into small sizes for better analysis. The interrogation process, which is associated with *catechization* activity, is broken down into three processes namely pre-processing, pattern discovery and pattern analysis [9]. Catechization in computer related tasks is a unique process of self-reflection, learning, gaining new skills and knowledge about how the events took place [28]. The purpose of the pre-processing process is to offer a structural, reliable and integrated data source for pattern discovery [9]. ***Pre-processing tasks*** consist of data cleaning, transaction identification, session identification, data integration, and transformation [9].

All the batches of log files filtered into the DEBs are in raw form. Therefore, not all log entries are valid for pattern analysis; thus, only those entries that carry relevant information must be kept [9]. Hence, the ***data cleaning process*** eliminates irrelevant entries from the access log files, including entries that record 'error' or 'failure'; some access records generated automatically by a search engine agent; requests for picture files associated with requests for particular pages; and entries with unsuccessful HTTP status codes [3]. After data cleaning the ***transaction identification process*** begins. The aim of the transaction identification is to create meaningful clusters of references for each individual who accessed the organisation's networks [9]. The log entries must be partitioned into logical clusters using one or a series of transaction identification modules [3]. The module can be a merge or a divide module and both types of modules take a transaction list and possibly some parameters as input; the output is a transaction list operated by a function in the module in the same format as input [9].

Subsequently, the ***session identification*** process trails the transaction identification. A session can be described as a group of activities performed by a user from the moment he enters the computer networks to the moment he leaves [9]. Identification of sessions on networks helps to identify how the cyber-criminals manoeuvred within an organisation's computer networks [3]. Therefore, session identification is the process of segmenting the access log of each log file entry into individual access sessions [3]. A session includes the used IP address (source address), user ID, the URLs of the accessed data and access time (timestamps) [9]. However, session identification helps to establish whether the computer used was from an Internet cafe, a computer classroom or a home based computer [3];

hence, before tracing the origin of the intrusion, the practitioners first ascertain the environment from which the attack was launched.

Once all relevant transactions and sessions are identified the process of ***data integration*** begins, which is also known as data fusion, and is the process of combining the matching transactions and sessions [3]. After the matching process, transformation of the combined data commences. The ***transformation process*** involves making sure that the whole pre-processing activity is conducted and rechecked thoroughly and that there are no errors, thus, providing a good foundation for the pattern discovery and pattern analysis processes [9]. The transformation process marks the end of the pre-processing process.

The ***pattern discovery*** process follows the pre-processing process. Pattern discovery is one of the key processes in log file forensic examinations [9]. It includes the algorithms and techniques from several research areas such as data mining, machine learning (artificial intelligence), statistics and pattern recognition [9]. In addition, pattern discovery involves a thorough search for passwords used to enter the organisation's computer networks, unusual hidden files or directories accessed, file extension and signature mismatches [3]. Techniques such as statistical analysis, association rules, clustering, classification, sequential pattern, dependency modelling and path analysis are used to analyse the pre-processed log file data to discover the movement patterns of the criminals within the networks [9]. Moreover, in computer auditing, the pattern discovery allows the forensic practitioners to determine how the criminals navigated within the organisation's computer networks [9].

The last process in the interrogation process is ***pattern analysis***. Pattern analysis, also called reconstruction, has two major activities, namely resolution and BackTracing [3]. The aim of the resolution activity is to extract the salient rules, patterns or statistics from the pattern discovery process by eliminating the irrelevant rules or statistics [9][3]. The pattern analysis process of log file examination is one of the providing tools to facilitate the transformation of information into useful knowledge [9].

When the forensic practitioners have enough knowledge about the crime from the resolution activity then the BackTrace process begins. This process uses a retrieved source IP address acquired in session identification to trace back through the Internet Service Providers (ISP) to the source computer [15]. Thus, BackTracing is a process of reconstructing the activities of criminals in the organisation's computer networks focusing on retrograding completion of records on the path followed by the criminals [3]. The BackTracing process also allows the practitioners to ascertain the password and the user ID by referring to the timestamps on the log files of the used computer [5][2]. If the results from the process of pattern analysis do not convince the forensic practitioners then the process of interrogation can be restarted.

3) Analysis

The ***analysis process*** comes after the interrogation process and involves technical reviews which must be conducted by forensic practitioners on the patterns discovered and analysed. The major activity in the analysis process is the correlation task. Correlation of events on log files involves identifying relationships between fragments of data, analysing hidden data and determining the significance of the information obtained from the log files from both the source computer and the filtered log files [5]. Reconstructing the event data

based on the extracted data and arriving at appropriate conclusions are some of the activities performed during the correlation activity [29]. User IDs, passwords as well as usernames from both source logs and filtered logs have to be correlated to see if there are some similarities with respect to time. The most important issue in correlation of events on log files is the timestamp interpretation [22].

Timestamps can provide proof of when the criminal offense took place referring to both the source log files and the filtered log files [2][40]. To ensure that the original time the source computer was utilised correlates with the time recorded on the log files obtained from the victim organisation, the Coordinated Universal Time (UTC) may be used [2]. The UTC is a high-precision atomic time standard based on the earth's rotation rather than the passage of seconds that cyber-criminals can always manipulate on the source computer when launching an attack [2]. Also, this allows the events on log files to be recorded, analysed and potentially correlated regardless of the difference of the time zones of the source log files and the filtered log files [22][2]. In addition to the correlation of events on log files, the U.S Department of Justice recommends timeframe analysis, hidden data analysis, application analysis and file analysis of the extracted data [30]. The results of the analysis process should be completely and accurately documented because they will be used in a court of law [19].

D. Elucidation Layer

The ***Elucidation Layer*** is the fourth and last layer in the proposed log file forensic model (see Figure 1). This layer explains how the processes ran throughout the investigation. The explanation must be very simple and clear, considering that some of the people take time to understand forensic ideas and views. The processes in this layer are the presentation process and the conclusion and follow-up process.

1) Presentation

After the recovering, interrogating and analysing processes, the results must be ***presented*** in a court of law not only before law enforcement officials but also technical and legal experts as well as corporate management [21]. It should be possible to confirm or discard allegations regarding the particular crime or suspicious incident [6]. Therefore, the individual results of each of the previous processes may not be sufficient to reach a convincing conclusion about the crime; hence, presentation of these results in a court of law should provide a clear picture to the audience [25]. The results of the interrogation and the analysis processes must be reviewed in their entirety to elicit a complete picture.

When presenting the evidence in a court of law opposing theories will also be presented and therefore there is a need to provide substantiated exhibits of the events that occurred as well as support for the theory or model of the events that occurred [20]. Therefore, digital forensic practitioners must prove the legitimacy of their theory [20]. This is where the benefits of having a standardised model enable either the conviction of a suspect or the discharge of an innocent individual. However, a successful challenge will result in the revaluation of the evidence in order to construct a better model; each model will be supported by the evidence presented and will be based on existing laws and regulations [31].

A report comprising of an in-depth abstract of the various processes of the investigation and the findings must be provided on the presentation of evidence [31]. The

case leader has to give expert testimony in a court and all the complex terms involved in various stages of the investigation process need to be explained in layman's terminology [26]. The expertise and knowledge of the forensic practitioner(s), the model adopted and the tools and techniques used are all likely to be challenged [18]. Along with the report, supporting materials like the copies of DEBs containing the filtered raw log files entries, the chain of custody document and the printouts of various items of evidence should also be submitted to the jury [29].

2) Conclusion and follow up

The ***conclusion and follow-up process*** involves reviewing all the steps in the investigation and identifying areas which need improvement [14]. The results and their subsequent interpretation can be used for further refining the gathering, interrogation and analysis of network log mining investigations [31]. The conclusion and follow-up process also involves the distribution of information in order to provide a basis for future investigations [32]. Sharing of information provides a basis for future investigations i.e. court precedents. Therefore the information will influence future investigations [20]. Conversely, there are various policies and procedures that need to be followed as prescribed by the victim organisation and the law in order to share information relating to a crime. Certain information may only be made available in the investigating organisation while other information may be more widespread [20]. The conclusion and follow-up process marks the end of the forensic investigation.

However, mere running of the processes from preparation layer to the elucidation layer as explained above may cause disputes in a court of law. Accordingly, laws and regulations must be involved and observed in forensic investigations [33]. Hence it is of the utmost importance that due diligence is applied during the investigative process and that the chain of custody is maintained. If laws and regulations are not observed then the court is likely to dismiss the allegation [33].

Meanwhile, in addition to the laws and regulations, this research paper proposes some forensic precautions. The issue of forensic precautions is one of the exclusive points projected in this model. Figure 1 illustrates the inclusion of forensic precautions as well as laws and regulations in computer network log mining.

E. Forensic Investigation Precautions

The ***forensic investigation precautions*** augment the process of network log mining. The precautions listed have to be observed from the discovery layer through to the elucidation layer because most of the technical and illuminating tasks likely to be associated with questionable activities in a court of law, are run in these layers. This therefore reduces the number and the impact of mistakes made during the investigation. Some of the precautions are aspects already known in the field of digital forensics but not necessarily emphasised as part of a forensic model; hence, this model emphasises that these precautions are part of a digital forensic model to be used in network log mining.

Some of the precautions taken so that innocent individuals do not suffer negative consequences, that companies do not lose millions of dollars and that criminals do not escape punishment [15] are listed below:

- avoid experiments on the original copy of log entries, because more than one experiment can be conducted; meanwhile the copy is needed in a court of law [34][14];
- accountability for any change - if any change occurs on the original copy of the logs the responsible practitioner must document his name and the time of the alteration [34][14];
- observation of available IT best practice i.e. ISO27002, COBIT, COSO and PCI DSS [35][34];
- not exceeding one's knowledge; thus, senior practitioners must be consulted when investigations become difficult and complicated [35];
- maintenance of a chain of custody [36], pervades the entire process, for example response action, command extended and tools used ; and
- utmost good faith [27].

F. Laws and Regulations

Due to the nature of the digital forensic process and the applicable laws, mistakes can be costly and therefore criminals may not be prosecuted. It is important to understand the impression ***laws and regulations*** have on the forensic investigation process [16]. Hence, all the processes in all layers must be conducted according to the relevant laws and legislation in place in the presiding jurisdiction (local laws) as well as international laws such as the Sarbanes-Oxley, Foreign Corrupt Practices Act (FCPA) and the Bribery Act. This paper also stresses that when conducting an investigation, law must be the first criterion to consider, as mentioned in the formulate approach process, before carrying out any initiation of the forensic investigation. Without abiding to the laws and regulations criminals may go unpunished and the organisations may suffer considerable loses [17]. Furthermore the review process allows for further improvements thereby creating a better process of digital forensic examination [15].

IV. CONCLUSION

The proposed log file digital forensic model is homogeneous to accommodate for general network log mining on computer networks. However, the significance of using the model will enhance the trustworthiness and acceptability of the evidence in a court of law. Investigations on log files are useful and accepted by the international law. Hence, this model's center of attention is on the extraction, analysis and correlation of data from log files. Following the presiding as well as international law and regulations as part of forensic investigations is very important; hence, this model seeks to address compliance to the laws and regulations by focusing the foundation of the model on the legal aspects of the investigation. Furthermore, these aspects have been proposed as part of forensic investigation precautions. Precautions should be observed from the moment technical investigations begin until the end of the investigations. The emphasis of every investigation should be grounded in the laws and regulations and therefore every aspect of the digital forensic investigation must be scrutinised for faults and improvements. The focus of network forensics is to provide a link between the suspect and the event by analysing log files for crucial data, such as timestamps, that can be traced back through the use of a pattern analysis to a specific network or computer. Lastly, this information is used in a court of law in order to support a legal case.

V. REFERENCES

[1] BBB Accredited Business, "Computer and Digital Forensic Bachelor's Degree", http://degreedirectory.org, 2010

[2] D. Casey, "Turning log files into a security asset", in Network Security, vol. (2), no. (2), pp 4-7, 2008.

[3] M. Munk, J. Kapusta, and P. Svec, "Data pre-processing evaluation for Web log Mining: Reconstruction of Activities of a web visitor", in Procedia Computer Science, vol. (1) no. (1), pp. 2273-2280, 2010.

[4] A. Schuster, "Introducing the Microsoft Vista event log file format", in Digital Investigation, vol. (4), no. (1), pp. 65-72, 2007.

[5] D. V. Forte, "The Art of log correlation: Part 1. Tools and techniques for correlating events and log files", in Computer Fraud and Security, vol. (1), no. (6), pp 7-11, 2004.

[6] S. Perumal, "Digital Forensic Model Based on Malaysian Investiagtion Process", in IJCSNS International Journal of Computer Science and Network Security, vol. (9), no. (8), pp. 38-44, 2009

[7] D. V. Forte, "The importance of log files in security incidence", in Network Security, vol. (3), no. (7), pp 18-20, 2009

[8] E. Tug, M. Sakiroglu, and A. Arslan, "Automatic discovery of the sequential access from web log data files via a genetic algorithm" in Knowledge Based Systems, vol. (19), no. (3), pp. 180-186, 2006

[9] R. Das, and I. Turkoglu (2009), "Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method" in Expert Systems with Applications, vol.(36), no. (3), pp 6635-6644, 2009.

[10] B. Shebaro, F. Perez-Gonzalez, and J. R. Crandall, "Leaving timing-channel fingerprints in hidden log files", in Digital Investigations, vol. (7), no. (1), pp. 104-113, 2010

[11] K. K. Arthur, and H. S. Venter, "An Investigation into Computer Forensic Tools", in Information and Computer Security Architectures (ICSA) Research Group. Pretoria: ISSA. University of Pretoria, 2005

[12] G. Palmer, "A Roadmap for Digital Forensic Research-Report from the first Digital Forensic Research Workshop Group (DFRWG)", (Technical Report DTR-T001 Final). Air Force Research Laboratory. Utica: Rome Research Site

[13] M. Reith, C. Carr, and G. Gunsch, "An Examination of Digital Forensic Models", in International Journal of Digital Evidence, vol. (1), no. (3), pp. 1-12, 2002

[14] B. Carrier, and E. H. Spafford, "Getting Physical with the Digital Investigation Process", in International Journal of Digital Evidence, vol. (2), no.(2), pp 121-130, 2003

[15] V. Baryamureeba, and F. Tushabe, "The Enhanced Digital Investigation Process Model", http://www.dfrws.org/2004/day1/Tushabe_EIDIP.pdf, 2004

[16] R. S. Ieong, "FORZA - Digital forensics investigation framework that incoporate legal issues", in Digital Investigation, vol. (3), no. (1), pp. 29-36, 2006

[17] KPMG, "E-commerce and cybercrime: New Strategies for Managing Risk of Exploitation". http://www.uazuay.edu.ec, 2009

[18] A. Ramabhadran, "Forensic Investigation Process Model for Windows Mobile Devices", in Security Group - Tata Elxsi . India, 2009

[19] K. M. Roger, J. Goldman, R. Mislan, T. Wedge, and S. Debrota, "Computer Forensics Field Triage", in Digital Forensics, Security and Law, vol. (1), no. (2), pp. 27-40, 2006

[20] S. O. Ciardhuain, "An Extended Model of Cybercrime Investigations", in International Journal of Digital Evidence, vol. (3), no. (1), pp 1-22, 2004.

[21] J. Grand, and B. D. Carrier, "A hardware-based memory acquisition procedure for digital investigation", in Digital Investigation, vol. (1) no. (1), pp. 50-60, 2004

[22] D. V. Forte, "The Art of log correlation: Part 2. Tools and techniques for correlating events and log files", in Computer Fraud and Security, vol. (1), no. (7), pp 15-16, 2004.

[23] S. Jajodia, Intrusion Detection Systems. Italy: Springer, 2008

[24] N. L. Beebe, and J. G. Clark, "A Hierarchical, Objective-Based Framework for the Digital Investigation Process", in Digital Investigations, vol. (2), no. (2), pp 146-166, 2005
[25] J. Giordano, and C. Maciag, "Cyber Forensics: A Military Operations Perspective", in International Journal of Digital Evidence, vol (1) no. (1), pp. 1-13, 2002
[26] B. R. Jones, "Comment - virtual neighborhood watch: Open source software and community policing against cybercrime", in Journal of Criminal Law & Criminology, vol. (97), no. (2), pp. 601-629, 2007
[27] B. Pladna,"Computer Forensics Procedures, Tools, and Digital Evidence Bags: What they are and who should use them", http://www.infosecwriters.com, 2008
[28] G. Gilder, and R. Vigilante, "Stop everything. Its techno-horror". http://www.kurzweilai.net/stop-everythingit-s-techno-horror, 2001
[29] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to Integrating Forensic Techniques into Incident Response", in National Institute of Standards and Technology, US Department of Commerce. Gaithersburg: NIST Special Publication, 2006.
[30] U.S Department of Justice, Forensic Examination of Digital Evidence: A Guide for Law Enforcement. USA: National Institute of Justice (NIJ), 2004
[31] F. Freiling, and, B. Schwittay, "A Common Process Model for Incident Response and Computer Forensic", in Proceedings of Conference on IT Incident Management and IT Forensics, Germany, pp. 1-21,2007
[32] D. Ayers, "A second generation computer forensic analysis system". Digital Investigations, vol (6), no (1), pp 34-42, 2009.
[33] D. J. Ryan, and G. Shpantzer, "Legal Aspects of Digital Forensics", http://www.danjryan.com, 2005
[34] S. McCombie and M. Warren, "Computer Forensics: An Issue of Definition", in 1st Australian Computer, Network & Information Forensics Conference, Perth, Australia, pp. 1-4, 2003
[35] IT Service Strategy, "Comparison between ISO 27001, SAS70, PCI DSS, Cobit and Common Criteria," http://www.itservicestrategy.com, 2008
[36] P. Turner, "Digital provenance – Interpretation, Verification and Corroboration" in Digital Investigation, vol. (2), no. (1), pp 45-49, 2005
[37] Rowlingson, R. A ten step process for forensic readiness. Internation Journal of Digital Evidence , 2 (3), pp 1-28, 2004.
[38] B. Carrier, "Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers", in International Journal of Digital Evidence, vol. (1), no.(4), pp 1-12, 2003
[39] N.L Beebe and J.G Clark, "A hierarchical, objectives-based framework for the digital investigation process". Digital Investigation, vol (2), no (2), pp 147-167, 2005
[40] F. Buchholz and B. Tjaden, "A brief study of time". Digital Investigation, vol (4), no (1), pp 31-42.