Forensic evidence isolation in clouds

by

Waldo Delport

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

November 2013

Forensic evidence isolation in clouds.

by

Waldo Delport
E-mail: wdelport@cs.up.ac.za

Abstract

Cloud computing is gaining acceptance and also increasing in popularity. Organisations often rely on cloud resources as an effective replacement for their 'in-house' computer systems. In the cloud, virtual resources are provided from a larger pool of resources, these resources being available to multiple different clients.

When something suspicious happens within a digital environment, a digital forensic investigation may be conducted to gather information about the event. When conducting such an investigation digital forensic procedures are followed. These procedures involve the steps to be followed to aid in the successful completion of the investigation. One of the possible steps that may be followed involves isolating possible evidence in order to protect it from contamination and tampering.

Clouds may provide a multi-tenancy solution across multiple geographical locations. When conducting an investigation into physical equipment the equipment may be isolated. This may be done, for example, by placing a cell phone in a Faraday bag in order to block signals or unplugging a computer's network cable to stop the computer from either sending or receiving of network traffic. However, in the cloud it may not be applicable to isolate the equipment of the cloud because of the multi-tenancy and geographically separated nature of the cloud. There is currently little research available on how isolation can be accomplished inside the cloud environment.

This dissertation aims at addressing the need for isolation on the cloud by creating new methods and techniques that may be incorporated into an investigation in order to isolate cloud resources. Isolation can be achieved by moving the unnecessary evidence to a different location and retaining the required evidence or by moving the required evidence in such a manner that the evidence would not be contaminated. If isolated evidence were to be moved to a digital forensic laboratory, the question arises as to whether it would be possible to create such a laboratory on the cloud utilise the benefits of cloud computing and enable the investigation to be conducted on the cloud without moving the isolated evidence from the cloud. The dissertation will develop various models of isolation. These models are then tested in experimental conditions. The experiments were conducted on Nimbula Director 1.0.3 and VMware vSphere 5.0.

The models were successfully applied in the experiments. It was found that investigations could benefit from the use of the proposed models for isolation. However, the experiments also highlighted that some of the models are not applicable or that a combination should be used. The experiments also indicated that the methods to be used would depend on the circumstances of the investigation. A preliminary "cloud laboratory" was designed and described in terms of which a digital forensic laboratory can be created on the cloud resources, thus enabling an investigation to be conducted inside the cloud environment.

Keywords : Digital forensics, cloud computing, digital forensics process, isolation.

Supervisor : Prof. M. S. Olivier,
Department : Department of Computer Science
Degree : Magister Scientia

Acknowledgments

I would like to acknowledge the following:

- My supervisor, Prof. Martin Olivier, for his professional supervision. Thank you for your wise words and for helping me to complete this journey. Without your guidance, this would not have been possible - you kept me focused and enabled me to achieve this goal.
- My wife. Thank you for walking this path with me and for being there throughout the process. Thank you for all the editing and support during the late nights.
- My mother. Thank you for all the support and love, so needed when conducting research.
- The members of the ICSA research group for their assistance and ideas. A special thanks to Kamil Reddy, Pedro de Souza, Dirk Ras and Francois Mouton.
- Micheal Köhn for guiding me towards a master's degree, and for all the support and wise words.
- The Department of Computer Science for making available the resources I required to conduct my research.
- My employer, EPI-USE Labs, for understanding and accepting that a research student is not always fully rested and needs time for meetings. Also thank you for the use of some of your resources for testing.
- My friends for all the support and wise words which helped me complete this dissertation.

Contents

List of Figures v

List of Tables vi

1 Introduction 1
- 1.1 Problem Statement . . . 3
- 1.2 Research Methodology . . . 4
- 1.3 Layout . . . 4

2 Digital Forensics 6
- 2.1 Introduction . . . 6
- 2.2 History of Digital Forensics . . . 7
- 2.3 Defining Digital Forensics . . . 7
- 2.4 Digital Evidence . . . 8
- 2.5 Live and Dead Forensics . . . 9
- 2.6 Categories of Forensics . . . 9
  - 2.6.1 Computer Forensics . . . 10
  - 2.6.2 Network Forensics . . . 11
  - 2.6.3 Database Forensics . . . 12
  - 2.6.4 Mobile Forensics . . . 13
- 2.7 The Digital Forensic Examination Process . . . 14
  - 2.7.1 Cohen's Model . . . 14
  - 2.7.2 National Institute of Justice Model . . . 17
- 2.8 Evidence Protection . . . 18
- 2.9 Tools . . . 19
- 2.10 Conclusion . . . 21

3 Cloud Computing 22
- 3.1 Introduction . . . 22
- 3.2 History . . . 22
- 3.3 Characteristics . . . 26

3.4 Service Models . . . 27
3.5 Deployment Models . . . 29
3.6 Current Infrastructure as a Service (IaaS) Implementation . . . 30
3.6.1 Amazon Elastic Compute Cloud . . . 30
3.6.2 VMware vSphere . . . 33
3.6.3 Xen . . . 34
3.6.4 Nimbula Director . . . 34
3.7 Current Research on Cloud Computing and the Concerns of Cloud Computing . . . 35
3.8 Value Added from Cloud Computing . . . 36
3.9 General layout of a Cloud . . . 37
3.10 Conclusion . . . 37

4 Cloud Forensics **38**
4.1 Introduction . . . 38
4.2 The correlation between Cloud Forensics and Other Digital forensics (DF) Subfields. . . . 39
4.2.1 Computer Forensics . . . 39
4.2.2 Network Forensics . . . 40
4.2.3 Database Forensics . . . 41
4.2.4 Mobile Forensics . . . 41
4.3 Development of Cloud Forensics . . . 42
4.4 Different Cloud Models . . . 43
4.5 Considerations for Cloud Forensics . . . 44
4.6 Problems Faced by Cloud Forensics . . . 47
4.6.1 Distributed nature of the cloud . . . 48
4.6.2 Multiple time zones . . . 48
4.6.3 Multiple jurisdictions . . . 48
4.6.4 Limited knowledge on cloud computing . . . 49
4.6.5 Evidence acquisition . . . 49
4.6.6 Virtual storage . . . 50
4.6.7 Deleted data . . . 50
4.7 Benefits of Cloud Computing . . . 51
4.8 Conclusion . . . 52

5 Isolating a Single Instance **53**
5.1 Introduction . . . 53
5.2 Isolation of a Crime Scene in a Cloud . . . 54
5.2.1 Instance Relocation . . . 55
5.2.2 Server Farming . . . 57
5.2.3 Failover . . . 58

5.2.4 Address relocation . . . 59
5.2.5 Sandboxing . . . 60
5.2.6 Man in the Middle . . . 61
5.2.7 Let's Hope for the Best . . . 63
5.3 Conditions for Isolation . . . 64
5.4 Testing Conditions for Isolation . . . 67
5.4.1 Locating an Instance . . . 67
5.4.2 Blocking Communication . . . 69
5.4.3 Gathering Possible Evidence . . . 70
5.4.4 Separation . . . 71
5.5 Distributed Instance System (DiS) . . . 71
5.6 Isolation of Cooperating Suspect Instances . . . 72
5.7 Conclusion . . . 74

6 Cloud Isolation **76**
6.1 Introduction . . . 76
6.2 Cloud Separation . . . 77
6.3 Cloud Separation on Different Types of Clouds . . . 82
6.4 Experimentation . . . 84
6.5 Isolation Accomplishments . . . 87
6.6 Conclusion . . . 87

7 Forensic Laboratory in the Cloud **89**
7.1 Introduction . . . 89
7.2 Digital Forensic Laboratory . . . 90
7.3 Requirements and History of the Forensic Laboratory . . . 92
7.4 Forensics Laboratory on the Network . . . 93
7.5 Forensic Laboratory in the Cloud . . . 96
7.5.1 Template Repository . . . 97
7.5.2 Layout . . . 98
7.5.3 Security Measures . . . 100
7.5.4 Administration . . . 102
7.5.5 Challenges . . . 103
7.5.6 Advantages Gained from Cloud Computing . . . 104
7.6 Conclusion . . . 105

8 Conclusion **107**
8.1 Introduction . . . 107
8.2 Derived publications . . . 107
8.3 Future Work . . . 108
8.4 Summary of work . . . 108

Bibliography 109

Appendix 120

A Abbreviations 120

List of Figures

3.1 Deployment Models 31

5.1 The ifconfig Command on a Node 68
5.2 The ifconfig Command on the Instance 68

6.1 Moving an instance from one cloud to another 78
6.2 Moving an instance from one cloud to another using an external cloud . . 79
6.3 Creating Two Clouds from a Single Cloud 80
6.4 Creating two Clouds on One Network 81
6.5 Creating Two Sub-Clouds 82

7.1 Example Layout of a Virtual Environment 101

List of Tables

5.1 Analysed Network Traffic. 74

6.1 Summary of applicability of separation methods. 86

Chapter 1

Introduction

Cloud computing is a fast growing industry and is rapidly becoming part of most enterprises [1]. Cloud computing builds on the advances which have been made in both the networking industry and in virtualisation [2]. In the cloud, virtual resources are provided from a larger pool of resources [1]. The cloud itself may consist of multiple types of resources and is capable of serving multiple clients. One of the most common uses of a cloud is to provide virtual servers; these virtual servers are known as "instances"' [3]. Cloud computing enables a service provider to provide virtual resources over the network [4] while also enabling the service provider to provide a flexible, cost-effective and on-demand infrastructure to its clients, thus freeing the clients from running their own infrastructure [1].

There has been a rapid increase in digital attacks over the last few years [5]. These attacks form part of an incident. Computer incidents may be categorised into two types. In the first type a computer is used directly in the incident, for example hacking. In the second type, although a computer is not part of the incident, the computer may contain some evidence of the incident, such as details concerning the incident in E-mails [6]. Despite the fact that cloud computing is still a relatively new phenomenon, it may also fall victim to various forms of digital incidents. These incidents may include a wrapping attack, malware-injection attack, flooding attack and/or browser attacks [7, 8].

The cloud industry is responding to such attacks by improving the security on clouds in an attempt to stop them [9]. When an attack does happen and the security mechanisms fail to stop the attack, then it is essential that the events be explored in order to gather information about the attack itself and the impact of the attack. Digital foren-

sics (DF) may be used to help answer the "who, what, when, where and how questions" arising from an event [10, 11]. When an event is explored a digital investigation may be required, this in known as a digital forensic investigation (DFI) [12]. When conducting such an investigation, a digital forensic procedure (DFP) is followed [13]. The DFP assists in gathering admissible evidence from the investigation. The DFP involves a number of steps which, when followed, should aid in the successful completion of the investigation.

Digital forensics consists of a number of sub fields, including network forensics and computer forensics, however, there is a need to combine these fields and also to define them more effectively [14]. One of the problems with specific approaches to information gathering in the digital world is the fact that the digital world is continuously evolving. This is no more apparent than in the development of new types of computer related technology. Software vendors are constantly producing new versions of software and also different types of software and, thus, the techniques for the forensic analysis of the older versions may not be acceptable to the newer versions of the software.

Digital evidence is one of the outcomes of an investigation [6]. Digital evidence may be used to assist in an investigation by providing answering to the "who", "what", "when", "where" and "how" questions. When working with digital evidence it is essential that certain procedures be followed otherwise the evidence may lose its value. This brings the argument back to the problem of ever-changing software as, if a new version of software is released, then the older investigation procedures would have to have been proven credible for the new software version.

The process of isolation is being used in the field of digital forensics, although the term "isolation" is not often used in this context. During an investigation Faraday bags and write blockers are used to protect possible evidence [15, 16]. This isolation enhances the admissibility of possible evidence.

Certain complexities are introduced when conducting an investigation in the cloud environment. These complexities that are introduced by cloud computing include the following: the location of evidence may be unknown, the evidence may be distributed over several instances inside the cloud environment, there is no known procedure to isolate either instances or evidence on a cloud, there is no known procedure in terms of which to investigate a part of the cloud and possible evidence must be moved from the cloud environment to a DFI laboratory in order to conduct an investigation.

Instances may be situated at any location on the cloud, while the cloud itself may be

located over several geographic locations. In order to limit the scope of an investigation instances can be isolated. However, there are currently no known methods with which to isolate instances inside the cloud. In the case of multiple instances working together to achieve one goal the relevant DF evidence may be distributed over multiple instances and it may be impossible to gather all the evidence related to the investigation in question. The related instances must then be isolated in order to gather possible evidence from these instances. An investigation follows procedures and policies that have been designed in such a way as to protect possible evidence. However, there is currently no known confirmation that an instance was isolated successfully.

In order to enhance the admissibility of digital evidence it is essential that proven methods be followed so as to complete the digital forensic examination successfully [13, 17]. When new versions of programs are released the software-related investigation methods need to be updated and proven to be valid as regarding the new versions of the software. The introduction of cloud computing has given rise to a problem in that most of the methods used in software-related investigations has been invalidated. In a normal investigation related computer hardware is identified, then confiscated and taken to a DFI laboratory. In the laboratory an analysis is conducted using accepted methods on the hardware in order to gather possible evidence from it. Both the multi-tenancy nature of cloud computing and the vast amount of user data on the cloud prohibits the confiscation of related hardware and this, in turn, means that there is no way in which gather the possible evidence for analysis.

1.1 Problem Statement

The focus of this study is the introduction of new methods and techniques that may be incorporated into a digital forensic investigation (DFI), thus making it possible to use previously defined methods for forensic analysis during a cloud investigation. With the main focus on investigating the feasibility of isolation inside the cloud environment

The study provides the theoretical methods and techniques in order to test them inside a cloud environment. It is envisaged that the outcome of this study may be used to implement DFI products for the various cloud offerings. The study is also limited to the basic cloud model, and no specialised cloud implementations are considered.

1.2 Research Methodology

The study introduced models that were tested using experiments. In addition, the study also proposed methods with which to conduct a DFI on the cloud. These methods form part of a conceptual model enabling a DFI to be conducted inside the cloud environment.

The proposed models are tested in experimental conditions with each of the experimental setup being explained once used. The experimental environments are fully controlled and are manipulated in a controlled manner only. Throughout the study multiple experiments are used in order to conduct the study. The experiments assisted both in providing theoretical knowledge about the previously stated problems and testing the proposed solutions. The experiments are conducted on VMware and Nimbula, these cloud providers were selected because VMware is a prominent cloud provider while Nimbula is a new cloud provider [18, 19].

A literature survey is conducted to enable the researcher to gain information about the fields of digital forensics, cloud computing and cloud forensics. The literature survey is used to explore the existing state of cloud forensics and to help with the path of the current research. The literature survey also serves as a basis for proposing methods with which to conduct a DFI inside the cloud environment.

1.3 Layout

This study consists of eight chapters. *Chapter 1*, the current chapter, serves as an introduction to the study and also discusses the research focus of the study.

Chapter 2 introduces the field of digital forensics, the way in which a digital investigation should be conducted and what the outcome of a digital investigation should be. This chapter will lay a basis for understanding digital forensics. The chapter starts by discussing the history of digital forensics. The various categories of digital forensics are then explained. The chapter then discusses the ways in which evidence may be protected and elaborates on the current tools that may be used in a DFI.

Chapter 3 introduces the concept of cloud computing and the path which was taken to develop cloud computing. The characteristics of cloud computing are discussed and used to describe the concept of cloud computing. The chapter then introduces the various

different deployment and service models are introduced. These are the different forms in which cloud computing may be delivered. The current deployments of cloud computing are also discussed. The chapter then examines the current research on cloud computing.

Chapter 4 explores the field of cloud forensics. This discussion includes the correlation between cloud forensics and other digital forensic fields. The chapter then examines the current work that is being conducted on cloud forensics and how this work influences cloud computing. The chapter also looks into the problems faced by cloud forensics.

Chapter 5 introduces the notion of isolating a single instance. The chapter starts by explaining the reasons why isolation is important and how isolation can aid a DFI. The chapter then introduces the techniques which can be used to isolate instances, including, but not limited to, instance relocation, server farming, address relocation, failover, sandboxing, man in the middle (MITM) and let's hope for the best (LHFTB). The chapter then discusses the conditions for isolation. It is essential that these conditions be met in order to verify that the isolation has been completed successfully. The chapter then provides the experiential results of the isolation conditions.

Chapter 6 focuses on isolating a group of instances or, in other words, a part of the cloud. This refers to the notion of cloud separation. The term cloud separation is introduced and explained, as well as the methods used to conduct cloud separation. The chapter includes the experimental results of cloud separation.

Chapter 7 introduces a DFI laboratory in the cloud. The chapter starts by explaining what a DF laboratory is and the factors that must be taken into account when a DF laboratory is created. The chapter then examines the current work on a laboratory on the network. The chapter also introduces the concept of a laboratory on the cloud and discusses all the considerations, internal workings and advantages of a laboratory on the cloud.

Chapter 8 summarises the work conducted in this study and proposes future work. Chapter 8 is followed by the bibliography, a list of abbreviations and a list of derived publications.

Chapter 2

Digital Forensics

2.1 Introduction

Digital forensics is often used by judiciary systems or in internal investigations to help provide answers to the "who", "what", "when", "where" and "how" questions of an investigation [10, 11]. When an investigation is in progress it is essential to uncover the motivation behind the crime or the object of the investigation. Such motivations include: where the incident happened, the circumstances which lead to the incident, who committed the incident and the time frame of the crime. Without such this information it would be difficult to prove a criminal case in a court of law.

The outcome of a digital forensic examination may be used either to prove or disprove a hypothesis in a court of law. A hypothesis is a statement that is made that will either contradict or support the reason or partial reason for the case [13]. In order to be of use in a court of law it is essential that the digital forensic process be based on acceptable theories and methods. The core components of a digital investigation include the gathering and examination of digital evidence.

In order to conduct a digital investigation the examiners need expertise in a wide variety of fields in the digital world [20]. However, the skills and techniques that are required to conduct a digital investigation are constantly changed, and need to be in line with current developments. It is, thus, incumbent on examiners to update their knowledge on a regular basis or they will not be able to gather meaningful evidence. This chapter will explain digital forensics and also discuss a model for a digital forensic

examination. The phases that will be used in the digital forensic examination will also be defined.

2.2 History of Digital Forensics

One of the first attacks on a computer took place in 1968 when a computer was damaged by gunfire [21]. Such an incident may be classified as physical damage and may therefore be dealt with by the normal procedures of the law. Most of the first crimes that involved computers were physical damage and, thus, it was not necessary to conduct an investigation involving the software.

The first real computer crime categories were introduced in 1978 [21]. The Florida Computer Crimes Act stresses that all companies, branches of government, and any other application of electronic documentation were vulnerable to document forgery, misuse of computer systems and the loss or theft of information [22]. The focus of this act was on the importance of the confidentiality, integrity, and availability of electronic information. However, nationwide legislation in the United States of America was only introduced later.

In the 1980s and 1990s the need for specialised teams to investigate computer related crimes were realised, as the number of computer related crimes was rising rapidly. The escalating crime rate and the lack of expertise in the field meant that the normal forensic offices were not able to handle computer-related crimes and, thus, the field of digital forensics started to grow. However, it was ad hoc and also not unified. The first attempt to standardise digital forensics was the Scientific Working Group on Digital Evidence (SWGDE) which published "Best practices for Computer Forensics" [23]. This marked the beginning of more attempts to standardise the field [24].

2.3 Defining Digital Forensics

In order to define digital forensics, it is necessary first to understand the term "forensics". Forensics is a formal and proven approach to the gathering of evidence and the processing of a crime scene and is sometimes used in a court of law [25]. Thus, based on this definition it is clear that digital forensics is based on substantial scientific methodologies

and techniques which may be used to assist in a court of law if necessary. Although in some cases the outcome may not be used in a court of law, the methods and techniques followed must, nevertheless, still be both creditable and scientifically based.

The digital forensic examination process, including the gathering of information about a specific event, happens after the actual event has occurred [6]. The process of digital forensics is time-consuming and it is usually not a trivial process, the digital forensic examiners require skills in programming and security implementation on various operating systems, the ability to use different operating systems and knowledge of security models [20]. The field of digital forensics makes use of several methods and procedures that have been proven useful and reliable in the non-digital environment. These methods have been adapted so that they may be applied in a digital environment [26].

2.4 Digital Evidence

In the main, digital evidence is used in legal proceedings in the form of information that has been gathered during an investigation [27]. Digital evidence may be regarded as containing concrete, established facts and it may be used as testimony in a court of law. In the most basic sense digital evidence may be seen as evidence that is in a digital form [13]. In the next sections digital evidence will be referred to merely as evidence because this study is set in the digital forensic realm.

Digital evidence has the following characteristics: The evidence can easily be copied and verified to constitute an exact copy of the original evidence at any stage. The evidence can easily be manipulated and altered, thus, resulting in the loss of integrity and value as regards to the case. The evidence can come from a wide variety of sources; including all the files saved on a computer medium, applications on a computer and even in locations that are unknown to the current user [6, 11].

The amount and sources of evidence are always increasing because of new devices that are able to store and create evidence. Many of these new devices can be connected to a computer [28].

2.5 Live and Dead Forensics

The standard computer forensic process uses a static or "dead" analysis [26, 29]. In a dead analysis, the system is powered off, usually by removing the power cord. As soon as the examination team has taken charge of the powered off system, images are made of the storage mediums and an analysis can then be conducted on these images. One disadvantage of a dead analysis is the fact that some information may be lost because the information is in a buffer, the RAM, the network or other live sources. However, an advantage of a dead analysis is that the environment is non-volatile and this makes it easier to capture the evidence and prove the authenticity of the evidence.

The alternative computer forensic process is a "live" analysis. The computer is kept running and evidence is gathered from the computer in the environment from entities on or around the system [26, 30, 29]. The possible sources of live forensic evidence may include, but are not limited to, the running network, working peripherals, the RAM and running applications. Specialised tools are usually used in a live analysis in order to protect the evidence that is being gathered. One problem with a live analysis is that the evidence may unintentionally be either, destroyed or modified and, once the evidence has been damaged, it may be impossible to recapture it.

In some cases a combination of live and dead analysis has been used [26, 29]. The live sources are gathered first, after which a dead analysis process is followed. Thus, a decision needs to be made as to which of the sources will be captured live and those which will be analysed using a dead analysis. In some cases the decision of which type of analysis will be followed is at the discretion of the lead investigator [31]. This decision is usually based on external events, for example, if there are programs running on the computer that are deleting files it may be prudent to switch off the computer before evidence is lost.

2.6 Categories of Forensics

The field of Digital forensics (DF) may be divided into sub categories. These categories unite all the tools and methods that are used for specific circumstances. The categories include, but are not limited to: Computer Forensics, Network Forensics, Database Forensics, Mobile Forensics and Cloud Forensics [12, 32, 15, 33, 34, 35, 13, 36, 37]. These categories are not mutually exclusive and, in some cases multiple categories are combined

as a basis for a DFI. In other cases there may be a main category which includes some methods from other categories. Each sub category will now be discussed.

2.6.1 Computer Forensics

Computer forensics is a sub field of digital forensics and is related to the forensics of computer components and their content [11]. The field of computer forensics attempts to narrow the search for evidence down to the computer itself, the content on the computer and the devices attached to the computer. Computer forensics is a relatively new field and was introduced in 1991 by the International Association of Computer Specialists [38], However, Federal Bureau of Investigation [39] has made claims that it existed prior to its introduction in 1991.

A computer forensics process is used when a computer has been used in some way in a crime. There have been arguments that computer forensics will expose only unskilled people who do not know how to hide or protect their information. On the other hand, a skilled person may protect their information in such a manner that the investigation will not succeed in its attempt to gather evidence of an incident [26]. The counter argument is that, in the non-digital environment, there are also people who are capable of concealing their involvement in a crime but, because of procedures and skilled examiners, it is, nevertheless, still possible to gather the evidence. This leads to the conclusion that it is essential that a digital crime be investigated in a proper manner otherwise the investigation will be futile.

Possible sources of evidence include files on the computer, applications on the computer, the operating system and also hidden information. The hidden information includes unallocated space and deleted files.

Files may be categorised as user files, system files and log files. User-created files are generated by the users when they use some form of application. System files are created by the operating system and the user may not even know of these files. On the other hand, log files are created by the system and contain information about what has happened on the system, including information about user actions. User files are files that contain user data or which are important to the user in some manner. The system files are used by the operating system to function and store the settings for the operating system as specified by the user. Files, created by the user or files that contain user information, are important to an investigation. However, most system files are not

important as they are generic to the operating system. The operating system files may also contain some information about devices that were used on the system. Log files may be important to an investigation as they contain information about the actions of the users.

Encrypted files are a sub category of user files and are usually created by the user to either protect of hide information. A possible deduction that may be made from finding encrypted information is that the user may, perhaps, possibly be attempting to hide information. The decrypting of encrypted information can be problematic and one problem is that investigators in certain countries may not ask for the key required to decrypt the data [40, 41]. Another problem is that breaking the encryption may be difficult or even impossible.

A second sub category of user files is steganography files. A steganography file can be classified as a file that contains other files or information [42]. These are files that are created by using stenographic methods. However, the methods used to detect steganography may be time-consuming and it is always possible that no information may be found.

Computer forensics evolved from the procedures and tools used by law enforcement and security companies [43]. There is a wide variety of tools available to aid in a computer forensic investigation. These tools help to protect the evidence from contamination. Contamination is avoided by collecting the evidence using proven methods and by keeping a time log of the actions that were conducted.

2.6.2 Network Forensics

Network forensics was introduced to assist in computer forensics in instances in which the computers are connected by a network. The evidence of a network forensic investigation is collected from the data which is sent over the physical network. Such a network includes at least two computers [11]. One method of gathering possible evidence is by capturing and analysing network traffic, while other sources of evidence are logs from servers, the browser settings of the user and router information. Network Forensics can be done “live”. However, the problem with live network forensics is that significant hardware resources may be required to conduct an investigation on large networks [32].

The network forensic field often relies on software programs to aid in an investigation [32]. Network traffic consists of data packets, with these packets being generated by

various sources on the network. The packets also belong to different protocols and may, therefore, be used to send and receive e-mail and for web-page content and ping information. The network capturing tools assists in grouping those packets together that belong to the same protocol. The tools also generally reconstruct the traffic flow of the protocol. The tools may then assist in separating those data packets that are unrelated to the case in question. The time frame for the investigation, related protocols and related IP address are generally known to narrow down the search space for evidence regarding the investigation.

Network forensics may be done as a "dead" analysis. The problem with dead forensics, however, is that the primary source of evidence, the network data, is volatile thus lost when the power is out leaving only the secondary sources [29]. The secondary sources of evidence that may be gathered from a dead network DFI include the data and log files on computers and network hardware. This problem is, however, exacerbated by the fact that not all network devices keep network logs. Most home and office network switches will not keep any logs, although the more advanced switches do keep a form of logs [44]. Accordingly, the evidence that may be gathered from a dead network forensic investigation may be limited. A dead network DFI is, thus, not in the best interests of the admissibility of the evidence and a live analysis may be more suited to network forensics. Hoverer, the live forensic team may also be faced with various challenges, for example, most modern network devices such as switches will direct the network traffic to the receiver only. Possible solutions to this problem include network cards in promiscuous mode, using a form of port mirroring where all the traffic on a switch is sent to a monitoring station or using the firewall to direct all traffic to the intended destination, as well as to the forensic computer. Nevertheless, this redirection may also not receive all network traffic as some traffic is internal to a system.

The next sub section will focus on the investigation that is conducted on databases.

2.6.3 Database Forensics

Database forensics consists of a forensic examination of the content of a database and the metadata of the database [37]. Database forensics is of assistance in investigations that are conducted on databases. The metadata is used to describe the content of the database and how the content maps to one field in the database. The information held in a database may contain what was changed by the suspect and it may also contain information about how it was changed.

One source of possible digital forensic evidence is referred to as "redo logs" [45]. A backup of the database is captured after which a log entry is made for every change on the database. These changes are either to the data inside the database or to the metadata. This is done to ensure that, in the case of database loss, the database is restored to its backup, while redo logs are used to remake all the changes up to the point of failure. The redo logs may also contain information about what was done to the database at a certain point in time.

Database forensics may also be done in either a "live" or a "dead" environment [37]. When conducting dead forensics the content and metadata are used to recreate a representation of the database. However, the problem with dead forensics is that some metadata and the last active representation of the database may be lost in the process.

When doing live forensics the most current view of the database may be used. However, the problem then arises that the system may be infected and the view given when interacting with the database may not be the actual data in the database. This is done to distract the investigators on the system while the content is being changed.

An introduction to mobile forensics now follows.

2.6.4 Mobile Forensics

The field of mobile forensics is related to computerised devices that are mobile. A mobile device possesses some processing capacity and limited available storage, while some mobile devices have a form of network connectivity. Possible evidence may be stored on the device itself in the form of logs, other user interaction storage, data on the device and the network traffic [46, 33]. For example; a modern mobile phone may have the following possible evidence sources: data contained inside messages sent and received by the device, the call logs of the phone, the contact list on the phone, the web history inside the browsers, e-mails on the phone, the applications installed on the phone, the data of the applications and GPS location information. There are also more sources of possible forensic evidence.

Mobile devices may be linked to a service provider with this service provider linking the device to the external environment [33]. All communication from and to mobile devices will pass through the service provider's network. Service providers keep logs for billing purpose [47]. In an investigation the logs gathered may possibly be obtained from the service provider to assist in the investigation.

Mobile forensics combines, among others, network and computer forensics. Computer forensics is used to gather possible evidence from the device itself although some of the methods may need to be adapted for mobile devices. For example, it is not possible to remove most of the on-board storage of devices and this must, thus, be imaged using the device. Some of the tools from network forensics may be used to gather evidence, although, some of the network traffic used by mobile devices is not catered for in network forensics. Network forensics can be used as a basis on which to build methods.

Mobile forensics can be done in either a "live" or a "dead" environment [46]. In dead forensics the active network data is lost. However, when conducting a live investigation it is not always possible to use the device itself to aid the investigation. The reason for that is because the device is not powerful enough to run some of the examination tools. Nevertheless, specialised tools for live and dead mobile forensic examination have been developed. These tools usually run on a computer and the device is connected to this computer. The computer is then able to access all the files on the device and also monitor all the connections made by the device.

2.7 The Digital Forensic Examination Process

In order to obtain admissible evidence it is essential that a well-defined forensic process be followed. Cohen [13] proposed a model for the digital forensic examination that consists of the following seven phases namely, identification, collection, transportation, storage, examination and traces, presentation and destruction. There are, however, other possible models. A prominent DFP has been defined by the United States of America's National Institute of Justice (NIJ) [12]. The phases include collection, examination, analysis and reporting. The models all share certain similarities, but also differ from one another in certain ways [13, 17]. This study aims add providing methods that enable these methods to be used inside the cloud environment. following sub section will explain the model proposed by Cohen.

2.7.1 Cohen's Model

If something is to be used it must be identified. However, in order for it to be identified it must first be found. When examiners visit a crime scene there are various sources from which to obtain evidence. During the identification phase, possible evidence is

identified. However, problems are often encountered during this phase, including: the vast amount of possible evidence that may exist and which will need to be identified. Another problem involves evidence that is not clearly visible and, if it is not identified during this stage, may be lost forever.

Once the evidence has been identified it needs to be collected. The evidence collected may be used wither in a court of law or during an internal investigation. It is, thus, essential that the integrity of the evidence be preserved throughout the collection process. In recent cases the need to retain the original evidence at the location has increased. However, the evidence must still be duplicated in order to preserve its original form. The need to copy the evidence and not to remove it saves the company concerned additional costs. A digital investigation should not harm either the person or institution involved.

The evidence collected must then be transported. Although the evidence is collected at the scene of the crime, the rest of the digital forensic examination process takes place at a different location. Accordingly, the evidence is moved to an examination laboratory where there is the equipment required to conduct a digital forensic examination. The normal way in which to ensure that the integrity of the evidence is maintained is to copy the evidence, keep the original evidence in a safe place, and move the copy elsewhere.

The digital forensic process may be a time-consuming process. During the examination itself and also once the examination has been concluded, the evidence must be stored in such a manner that it will not degrade and become inadmissible. The evidence also needs to be protected against natural disaster and sabotage. Storage mediums degrade over time and, thus, it is essential that certain procedures be followed in order to protect the evidence.

The examination and traces phase consists of the following four sub categories, namely, analysis, interpretation, attribution and reconstruction [13]. The examination phase endeavours to explain the route of evidence, from its creation to its current state. The final step, reconstruction, involves trying to create the same output from the original evidence using output from the analysis, interpretation and attribution phases.

During *analysis* the evidence is worked through to discover the origin and possible significance of the current case. The redundant evidence is removed and the remaining evidence is checked for abnormalities to help with the identification of potentially useful evidence. This helps to exclude evidence that is not useful to the current investigation. During this phase the examiner will also searches for evidence that is either hidden or was not apparent after collection. This type of evidence has usually been hidden on

purpose and it often contains valuable information. The main was in which a suspect may endeavour to hide evidence includes deleting files, placing files in places that are difficult to find and using steganography, encryption or other transformation methods.

The *interpretation* phase reveals the facts behind the evidence. These facts are usually either about traces or about events pertaining to the evidence. The facts about the evidence will include where the evidence came from, the history of the evidence and the value of the evidence to the case. The facts gathered will either confirm or refute the consistency of the evidence. The challenge throughout this process involves bearing alternatives in mind. The number of facts that need to be uncovered by the end of the investigation is usually defined by the amount of time and/or by financial constraints. During the interpretation phase redundancy may be helpful because it relies on more than one fact to prove either a trace or event of the evidence.

Once all the facts are known, these facts need to be mapped to a non-digital environment. This is known as *attribution*. Attribution is often difficult and may even be impossible. Actions on a computer are linked either to a user or to system processes. Authentication is the method used to identify a user on a computer system. A user usually needs to be identified before the use of the computer system is permitted. This information is kept and may be used to link a user to an action. The system process of attribution is used to map the applications, system functions and other sources on the system to which user had access prior to manipulating the evidence.

During the three phases of analysis, interpretation and attribution a set of hypotheses is formulated. These hypotheses are tested during the *reconstruction phase*. If the examiners are not able to prove the hypothesis by recreating the circumstances, then hypothesis is flawed and cannot be used. During reconstruction it is essential that the environment be controlled and that it is as close as possible to the original environment. In other words, the reconstruction department requires a wide variety of different types of equipment and software to attempt the recreation of a wide variety of environments.

The next step involves presenting the findings of the investigation. The presentation of these findings can take various forms. A report containing the outline of the examination process and the evidence that was found may be compiled. Such a report is vital in a court of law as it would prove the validity of the processes that the examiners followed. In some cases the examiners may have to testify in a court of law. The report and the content of their testimony will contain summaries of the previous phases. If a presentation contains faults or inaccuracies this would have a negative effect on the evidence that was gathered, possibly to the point where evidence would be declared inadmissible.

The last step in the digital forensic examination involves demolition or the return of the evidence that was collected. The period during which the evidence must be kept will be determined by the court and may range from immediate destruction to seventy years after the case. The time period is influenced by various factors, including data sensitivity and case severity. After such period the evidence which was collected may either be destroyed or returned to the where it came from. However, all copies of the evidence will be destroyed after the period has expired.

Although not previously mentioned, *documentation* is an ongoing process that must form part of all the phases of the digital examination [12, 13]. Documentation is one of the main ways in which the integrity of the evidence may be safeguarded. The documentation should include, at the least the evidence number, a description of the evidence and also the locations where the evidence was collected. The documentation should also include the processes that were followed in identifying, retrieving, storing, and transporting the evidence and must also mention the chain of custody when the examination was in progress. There have been several cases in which the outcome of a case was influenced by the documentation.

2.7.2 National Institute of Justice Model

This section expands on the NIJ model for a forensic investigation.

The *collection* phase involves the process of searching and recognising electronic evidence. The first problem in the collection phase of a digital investigation involves the vast amounts of possible evidence may exist and that would need to be collected. Another problem includes evidence that is not clearly visible as evidence and, if not collected at this stage, may be lost forever. The evidence that is collected will, in most cases, is used in either a court of law or during internal investigations. If evidence is to be considered useful in a court of law it is essential that the integrity of the evidence be preserved while the evidence is being gathered.

During the *examination* phase the evidence that was gathered during the collection phase is worked through in order to discover the origin of the evidence as well as the possible significance of the evidence. During this phase the examiner will have to search for evidence that either is hidden or was not apparent after collection. During the examination it is imperative to continue with documentation, in other words, the current state of the evidence should be added to the previous documentation. The last major

step in the examination phase involves separating the useful evidence from the redundant evidence.

The next step in the digital investigation is the *analysis* of the evidence. The output of the examination phase is used to determine the value and significance of the evidence as regards the case in question.

The last step involves compiling a *report* containing the outline of the examination process and the evidence that was found. This report is vital in a court of law as it would prove the validity of the processes which the examiners followed.

It is clear that the two models include the same set of underlying steps, although, Cohen's model comprises more steps than the NIJ's model. However, this enables a more systematic flow of events.

2.8 Evidence Protection

During the collection phase the evidence is gathered while, during the examination phase, the evidence is examined. In order to protect the evidence from contamination and tampering the evidence collected is often copied [13]. This copying process also ensures that the original evidence is intact after the investigation. Thus, the examiners work on the copied versions of the evidence and not on the original evidence. In order to prove that the copied evidence is identical to the original evidence a hash function may be used [38]. A hash function takes a variable length input and creates a fixed size value that maps the input to an index in an array [48]. This means that a hash value is created for the original evidence and this value is then stored. Later in the investigation the hash value of the copied evidence can be calculated and, if the hash value is the same, the copied evidence and the original evidence are identical. The two best known hashing algorithms are message-digest algorithm (MD5) and secure hash algorithm (SHA).

The MD5 algorithm is one of the most popular hashing algorithms [49]. It was introduced by Ron Rivest to replace MD4. The MD5 algorithm receives as input, a buffer of any length, and as output, returns a fixed 128-bit value. The MD5 algorithm divides the input into 512 bit blocks. The aim of the MD5 algorithm is to uniquely identify the file [13]. However, studies have shown that MD5 is vulnerable to collision attacks. A hash collision occurs when two different inputs give the same hash value [49]. A collision attack may be defined as the process of finding the input that will result in

a specific hash value. Thus, MD5 on its own is not an effective method to prove that a file has not been tampered with as the hash value may have been fabricated.

The National Institute of Justice (NIJ) introduced the SHA algorithm to replace the MD5 algorithm [50]. The SHA algorithm has several versions, including SHA-1, SHA-256, SHA-384 and SHA-512. However, SHA-1 is the most popular algorithm [51, 50]. The output of the SHA-1 algorithm is a 160-bit value. The algorithm also divides the input into 512 bit blocks. Nevertheless, although the output of SHA-1 is larger than MD5, it is also vulnerable to collision attacks.

MD5 and SHA-1 are the most widely used hashing algorithms but both are vulnerable to collision attacks. There are, however, stronger hashing algorithms with larger outputs, but they are rarely used. One solution to the problem of collision attacks is to use a combination of hashes, where both the MD5 and the SHA-1 are calculated and given. This is helpful because the methods involved on collision attacks vary [49, 51, 50]. Providing both hashes may enhance the admissibility of the evidences.

2.9 Tools

In the case of a computer crime, forensic tools are extremely useful and have been widely used in various DFI cases. However, it is imperative that the tools comply with a certain number of criteria, including the necessity of the tool using processes that are both acceptable and reliable. One way of determining the credibility of a tool, is to refer to past cases and, specifically, to whether the tool has been used in a previous case. A credible tool will ensure that the integrity of the evidence is maintained and that the evidence will be admissible in court. Forensic tools differ in their intended use: some are used in order to gather evidence while others may assist in the investigation process. After this study it should be possible to use these tools inside the cloud, one study aims at resurrecting cloud resources for a standard DF investigation [94]. Combined with this study the tools listed below might still be used in a cloud investigation.

The tools that may assist the process of an investigation are usually used throughout the investigation, from gathering the evidence to reporting the findings [11]. The tools also provide sub-tools or methods with which to identify possible evidence. The evidence identified may then be worked through and relevant evidence can be extracted using the tools. The tools provide the required functionality with which to gather the evidence in

a way that will protect the admissibility of the evidence. The evidence collected may also be analysed using the tools. The tools also enable the keeping of records of the actions taken by each investigator during the investigation. Thus, the tools create an audit trail in respect of each piece of evidence found and the interactions that may have occurred. This audit trail may then be presented, if required. The two best known tools are EnCase Forensic and Forensic ToolKit.

The first EnCase Forensic was developed by Guidance Software Inc. and is one of the most commonly used forensic process tools [52, 11]. EnCase Forensic provides a tool that is dedicated to gather evidence from smart devices, RAM, various file types and different hidden files. It has a built-in MD5 functionality which means that, as evidence is gather or logged into EnCase, a MD5 value is computed and, as the evidence is used, the MD5 is compared to ensure that the evidence is unmodified. In addition, EnCase will create reports containing all the evidence and interactions with the evidence. The reports are created in a format that is accepted in courts.

The second Forensic ToolKit or FTK was created by AccessData Group [53]. It provides tools that are useful throughout an investigation and may function as a focal point during an investigation. The tools are able to analyse the Windows registry, crack the passwords of several applications and creates a report at the end of the case. Such a report has been used in some court cases. As EnCase does, it also provides tools with which to protect evidence. Although both EnCase and FTK may prove useful in an investigation they are not always the only tools used by forensic examiners.

As part of an investigation the examiners may image the hard drives, memory cards or RAM. This is done by using tools that are specialised in creating images. The Unix DD is one example of an imaging tool [54]. DD creates bit-wise copies of RAM or any peripherals with memory and the copy may then be sent to an attached device. DD is also able to compute hash values of the images once completed. There are also other tools to create images, for example ByteBack [55]. ByteBack has the same functionality as DD but it also has a basic GUI to aid in its functions. ByteBack may then be run as an operating system in order to create hard drive images in a dead forensic investigation or it may run as an application, with a small memory footprint, in live forensics.

It may be extremely difficult, if not impossible to gather live evidence while conducting network forensics. Accordingly, in order to solve the problem, network sniffer software may be used. This software analyses the packets that are sent over the network. A well-known example is Snort [56]. Snort provides tool sets with which to capture and analyse live network traffic. The network traffic may be captured for later reanalysis.

Other tools that are able to capture network traffic are Wireshark and tcpdump [57, 58]. Tcpdump may be used to capture network traffic and is a prominent open source network capture tool. Wireshark is able to analyse the network traffic that either it or tcpdump captured previously. It is also able to conduct a live analysis of network traffic. The difference between the Wireshark and Snort is that Wireshark is a packet sniffer while Snort is an intrusion prevention and detection system that is also able to analyse network traffic but which also has other non-forensic purposes.

2.10 Conclusion

Digital forensics is an important field and is continuously changing. There are sub fields in digital forensics that were created especially to assist in certain aspects of digital forensics. The process that is followed in a digital investigation is important as it will determine the outcome of the digital investigation to be used in court.

This chapter served as introduction to the basic digital forensic process, including the activities involved in the process. Two main models which may be used to conduct an investigation were explained. The sub fields of digital forensics were expanded on and tools that may be used in an investigation were discussed.

In the next chapter cloud computing will be introduced. In addition, the chapter will discuss on the history of cloud computing and also its current implementations.

Chapter 3

Cloud Computing

3.1 Introduction

"Cloud computing" is a relatively old term but its use has been adopted increasingly over the last few years [1]. Cloud computing builds on different forms of distributed computing and links distributed computing with virtualisation. Cloud computing enables a service provider to provide flexible, cost-effective and on-demand resources to its clients instead of the clients having to run their own infrastructure in order to gain access to these resources.

3.2 History

The basis of cloud computing may be attributed to various distributed computing technologies [4]. These technologies include distributed computing, cluster computing, peer-to-peer computing, mainframe computer, network computing, client-server model, utility computing and autonomic computing.

In distributed computing multiple computing entities form one resource. These entities constitute a form of computing unit and they are connected through a network. All the entities work together to attain one global goal [59]. Forms of distributed computing include, but are not limited to, grid computing, clusters computing and peer-to-peer computing.

One of the forms of distributed computing technology is grid computing [4]. In grid computing multiple resources are combined to form one logical resource. These resources may be located at different geographic locations and may consist of different architectures, connected by a network. The combined resources work together to solve one problem or to achieve one goal. Such problems require a vast number of resources dedicated to solving them.

A computer cluster may be defined as a type of grid computing. In a cluster the resources are usually more closely located than in grid computing [60]. The cluster resources are connected through a high bandwidth network.

In peer-to-peer computing a combination of resources form a decentralised network [61]. Each resource will provide both disk space and processing power to the peer network. Each resource will receive various tasks to complete and, once the tasks have been completed, the results are sent to the correct destination. There are two types of peer-to-peer networks. In the first type, there is no controlling resources and all the resources work together to manage the network. On the other hand, the second type has a controlling resource that dictates what happens on the network. A peer-to-peer network is created dynamically and is maintained dynamically. If a new resource is added to the peer-to-peer network, information will be transferred to the resources. If, on the other hand, a resource is removed from the network, the information will be re-sent to a different peer.

A mainframe computer is a single, large computer which is operated by a single company [62]. Mainframes are designed for high I/O throughput and reliability. Mainframe computers have a specialised operating systems designed specifically for them. Mainframes may also be used to host multiple virtual machines which may, in turn, be used as servers. This virtualisation may be seen as the starting point of running virtual servers instead of actual servers.

Network computing may also be regarded as a predecessor of cloud computing. Although it is not directly related to cloud computing, certain notions still persist from network computing. In network computing users are provided with a thin client. This client may have few of its own capabilities, for example, it will not have CPU, RAM or storage [63]. However, it may be able to display information on a screen and receive user input using a keyboard and mouse. Some thin clients may also support contain peripherals such as printers. The thin clients are all connected to some form of network. Other clients are disk-less clients. This type of client has no secondary storage but does have its own CPU and RAM. All the clients will communicate with a server running

on the network. This server is responsible for storage and most of the processing. In the case of a thin client the server will do all processing and storage. The thin client operating system is a virtual machine running on a server. The hardware of the thin clients interacts with the running virtual machine. These mechanisms may be regarded as the roots of Infrastructure as a Service (IaaS) of cloud computing where a virtual machine is running in the cloud and the users interact with this virtual machine. IaaS will be elaborated upon later in this chapter.

The client-server model divides a computer system into two separate entities [67]. These entities comprise the server on the side of the service provider and a device on the side of client. The server will carry out most of the complex calculations, data retrieval and storage while, the client device is used to display information and the interaction with the server side. Websites are another example of a common use of the client-server models. The web server gathers the information and interacts with databases while the browser of the client displays the website.

Utility computing enables a service provider to sell resources to users [68]. The users will receive resources on demand and will pay for the use of these resources. These resources may include storage, computational power or services. The costs involved vary per service provider but most follow a pay as you use model.

In autonomic computing the technology maintains itself [71]. The system will manage the lower level operations while the user manages an abstracted system. Most of the self-maintaining behaviour follows some form of guide that would have been created by the users. This guide tells the system how to respond to certain events. Other events may be directed by a guide which is created by the system creators. This would be done to ensure that the system continues to function. The system is responsible for the following four categories of self-maintenance, namely, configuration, healing, optimising and protecting. The system will configure itself when the environment changes in order to continue functioning. In the case in which a printer is added, the printer will be configured by the system as soon as it is detected and it will be used after configuration. An automated system must be able to heal itself. If an entity in the system fails the system must respond to this failure by creating a work-around and continuing to function with the least amount of interference to the service. The system must also be able to optimise itself. If the system detects that one part of the entity is overloaded, the system should self-adjust the workload between all the available entities. The system must protect against unauthorised access. The users will have set up policies to guide access to the system. The system must then enforce these policies and block unauthorised

access.

Combining the client-server model, utility computing, virtualisation and autonomic computing will allow a service provider to provide virtual resources on demand to users. The system providing the resources is responsible for processing, storage and maintenance while the user interacts with the resource using a thin client.

Other important milestones includes the introduction of virtualization, creation of a hypervisor and the growth of the term cloud.

The term "virtualization" means creating an abstraction of something [69] and, thus, a virtual resource is created instead of a physical one. The resources include, but are not limited to, operating systems, RAM, CPU's, hard drives and computer networks [18]. Virtualisation enables a single host to run multiple virtual machines with each running their own operating system. Each virtual machine may have a different software and hardware configuration.

The term cloud has been in existence for some time and has been used in various contexts. The term was first used by telecommunication providers to describe virtual private networks on their network [64]. One of the other technologies which was also referred to as the cloud was the X.25 protocol, which was introduced and used in the 1970s [65]. The X.25 protocol used the internet as the cloud and created a secure network in the cloud between endpoints. In most modern references the term "cloud" refers to cloud computing. Cloud computing may be defined as a distributed computing architecture which provides flexible, cost effective and on-demand infrastructure to users over some form of network by using virtualisation in order to create virtual resources on the abstracted hardware [66].

A hypervisor is an important part of cloud computing [18, 70]. It allows physical hardware to support multiple client operating systems which run on it [69]. One of the uses of a hypervisor involves providing the functionality required to run virtual machines on the personal computer of a user. There are two types of hypervisors. The first type on the hardware itself and supports multiple operating systems while the second runs on top of an operating system and then allows multiple operating systems to run on top of it. The first type is used in cloud computing to provide resources while the second is largely used by operating system users providing virtual machines. The hypervisor is responsible for providing the virtual operating system with the hardware required for the virtual operating system to function. Thus, the hypervisor is in control of both the hardware scheduling and the virtual operating system maintenance.

3.3 Characteristics

This section lists characteristics that for the purpose of this study defines the cloud. These characteristics were chosen based on 'real world' clouds, there is no focus on DF for the characteristics chosen. Individually these characteristics may also describe other related technologies but, together, they form the characteristics of cloud computing [66]. These characteristics were defined, by the National Institute of Standards and Technology, as on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.

The users of the cloud service can be provided with resources on demand. The user has to specify the resources that are required and the system will provide them. The resources are provided by the cloud operating system, thus making it self-service. Accordingly, therefore the staff members of the service provider do not play a role in his process.

All interactions between the users and the cloud resources provided are done through the network. The client will typically use a thin client to access the resources. Most of the cloud resources may be accessed by using a website. If the resource is a virtual machine, then access may be gained by using a remote desktop connection or other, similar mechanisms.

One service provider may serve multiple clients. All the resources of the service provider are pooled in order to better serve the users. The client has little control over the specific location of their resources. For example, the resources may be located in different countries. However, users may be able to specify the country or data centre for their resources.

The users can easily modify the resources. For example, if the resources need either to grow or shrink this can be done in a short space of time. The user needs only to specify the new requirements and the system will accommodate these new requirements. This characteristic is known as rapid elasticity.

Cloud computing provides the mechanism with which to measure service on the cloud. This includes low-level measurements, for example, the amount of ram used, high-level measurements, for example, the number of users logged in. The measurements differ per service rendered and should provide the most feedback for the specific service. These measurements may be used by both the users and the providers in order to improve the services and to minimise costs.

There are also other characteristics of cloud computing, namely, security, cost reduction, reliable and maintainable.

Improved security is essential in the cloud [36]. Service providers will protect their users by providing resources, including time, money, and effort. Service providers undertake to protect the Confidentiality, Integrity and Availability (CIA) of their users. However, there are drawbacks concerning these security mechanisms and the users may struggle to obtain logs about information access and, thus, they may opt for a private cloud. Security will be discussed in more detail later in this chapter, as it may also be classified as a risk.

One of the aims of cloud computing is to reduce overall costs, including infrastructure, software, administrators, and utility costs [72]. The service providers bear most of the costs while the client pays only for the service rendered. This means that there may be a cost reduction for most companies.

Clouds are inherently reliable and easy to maintain. They are designed to work when nodes fail or when new nodes are added. The design of the cloud creates a state in which the service provider is able to guarantee a high availability rate of the cloud. In addition the cloud operating system provides mechanisms with which to maintain the underlying hardware while the users of the cloud are almost never affected by this maintenance.

The next section will focus on the various service models of cloud computing.

3.4 Service Models

There are three types of cloud computing service models, namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [73]. The various service models were introduced in order to meet the needs of different clients. For example, some clients wanted to move their entire infrastructure to the cloud while other clients merely wanted to access a service on-line.

The first service model is *Infrastructure as a Service*. The users of a cloud infrastructure are provided with a virtual computer which may be interacted with, usually via the Internet [3]. This virtual computer must be setup and maintained by the user. The virtual computer may also be referred to as an instance. Usually, an instance may be accessed from anywhere in the world, depending on the security setup. The instance may

be a small instance, used by a single user to store backups of files, or it may be a server running the website and database of a company. A user pays the service provider only for services rendered and there is no initial large investment on the part of the users. If the requirements of the user changes in terms of computational power or storage space, it is an easy process to change the scope of the instance to accommodate the new requirements of the user. If a new instance is required it is a simple process to start up an instance. The service provider is responsible for maintaining the CIA of the instances on a hardware level [74] while the user is responsible for protecting the CIA on a higher level, for example, the content of files and the operating system.

An example of IaaS is EC2 from Amazon [75]. The users of an EC2 instance are able to choose from multiple operating systems, including a wide variety of software that is already installed on the instance. Several aspects influences the cost of an IaaS instance, including the number of CPUs, the size of the RAM, the network bandwidth allowed and the storage type. The storage is dependent on the size, the speed and the lifetime of the storage. The storage may be direct attached storage a slower form of storage or the instance may use a Storage Area Network (SAN) for faster storage. There will be more details on IaaS implementations in the discussion later in the chapter.

The second service model is *Platform as a Service*, in terms of which the user is provided with a platform that is maintained by the cloud service provider [73]. The platform is an instance that was created by the service provider and with a specific focus. The service provider may create a default platform for a web server and the user must then configure the application on the platform. The service provider may also provide the necessary tools with which the user may build successfully upon the platform.

Examples of PaaS include Google's app engine and Microsoft Azure [73]. The users are provided with an API with which to develop and deploy software on their platforms. The API is used by the service providers to enable the users to create software with which to communicate with the cloud. The service providers would provision a system with a pre-configured database and operating system in order to ensure greater convenience for the user.

The last service model is *Software as a Service*, in term of which software is made available through the use of clouds. The application and the data of the application are seen as the resources on the cloud [76]. The users pay in order to gain access to an application that may be customised according to their requirements. The most common billing scheme is month to month with users paying a monthly fee for the service [77].

The obligatory fees involve factors such as the number of users and the amount of data. The user has no concerns in respect of the underlying hardware and software supporting the application of interest.

SalesForce CRM is an example of SaaS [78].F SalesForce provides a wide variety of services to its users. Its main focus is on the software required to connect the customers and the employees of the organisation concerned. One of the main uses of SaaS is business applications [77]. An HR system such as PaySpace is one example of a business application [79]. SaaS is typically interacted with by using a thin-client, the most popular of which the web browser. The services may be accessed from any location as long as internet access is available. It is incumbent on the service provider must protect to CIA of the user's data.

The next section will focus on the way in which the cloud may be deployed.

3.5 Deployment Models

There are four deployment models for clouds, namely, public, private, community and hybrid [66]. The driver behind the various deployment models is the end-user. There are users who prefer the cloud to be located on their premises and they have full control whereas other users merely want to access the cloud [36]. In order to meet the various user demands the four deployment models may be used.

In a *public* cloud, the infrastructure is owned by a cloud service provider [66]. The service provider will sell the resources of the cloud to other companies and to the public and, thus, the service provider is also responsible for managing the cloud.

In a *private* cloud, the cloud infrastructure is for the exclusive use of one company only [66]. Thus, the company owns the cloud and uses the resources. The cloud infrastructure may be located on the property of the company or it may be located elsewhere. In the case of the private cloud, the company, or a contracted company, is responsible for maintaining the cloud.

If the cloud infrastructure is used by several companies, it may be regarding as a *community* cloud [66]. The companies own the cloud and use the resources collectively and, thus, they form a community with a shared interest. The cloud infrastructure may be located on the property of one of the companies or it may be located elsewhere. Either

the companies themselves, or a contracted company, would, therefore, be responsible for maintaining the cloud.

A *hybrid* is a combination of at least two of the above mentioned models [66]. Each of the models used is still a separate entity in the hybrid cloud with forms of technology binding the entities together. This is normally used for load balancing.

Figure 3.1 depicts an example of a cloud deployment. In this figure company X and company Y collaboratively maintain a cloud, making it a community cloud that is running the cloud operating system Xen. Company X also has one private cloud running in-house software on VMware [18]. In order to help with availability, company X also uses resources on an external cloud, RackSpace [80]. Their external and internal clouds are connected by using Aeolus [81]. Aeolus manages cloud instances and also assists in the movement of instances between different cloud vendors. This, in terns, enables load balancing because, when the private cloud runs out of resources, instances will be sent to the external cloud. Once the internal cloud has the required resources, the instances are pulled from the external cloud to the internal cloud.

3.6 Current IaaS Implementation

This section will focus on a subset of the cloud implementations which are currently available and, thus, it will provide some required background and terminology for the remaining sections of the study. The available cloud implementations that will be discussed include Amazon Elastic Compute Cloud, VMware vSphere, Xen and Nimbula Director.

3.6.1 Amazon Elastic Compute Cloud

The Amazon Elastic Compute Cloud is a service provided by Amazon and known as EC2 [75]. It is a paradigm example of a public cloud. It provides a public IaaS infrastructure. EC2 uses a data store of instances where users save an instance and are able to launch instances from the data store as many times as required. Amazon calls this data store Amazon Machine Images (AMI). The AMI holds over a 1000 different images from which users may choose. Users that create images for the AMI may decide on the availability to

other users, or which images available for a certain selection of people only, for example the employees of a company.

An EC2 user chooses an image and then creates a virtual machine on the cloud with the initial state, replicating the image. The sizing and pricing is based on a number of factors, including, the number of computational units, RAM, size of storage and type of storage. Amazon provides a set of default sizes and configurations form which the user may choose. The user is then billed on a per hour basis for the amount of time for which

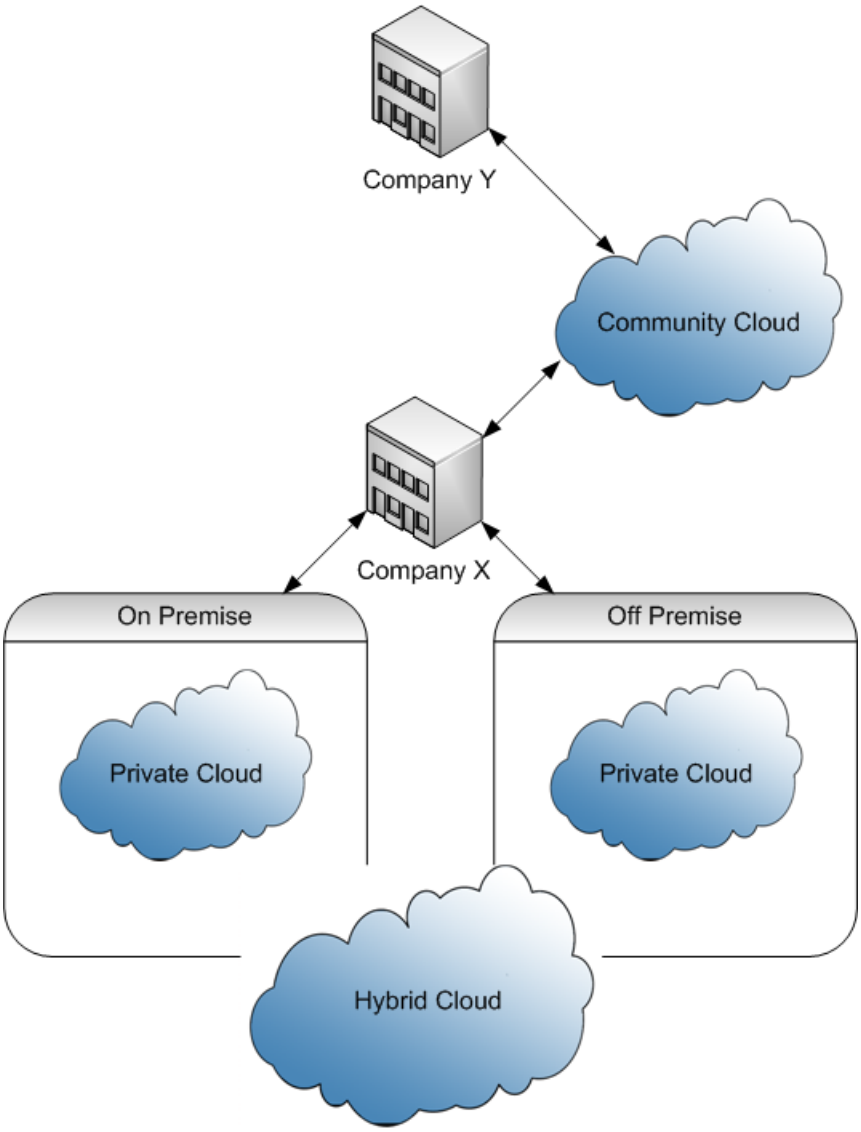

Figure 3.1: Deployment Models

the instance is running.

Following is the salient features of the Amazon cloud:

- In order to provide persistent storage Amazon offers the Amazon Elastic Block Store. This enables the storage of the instance to continue to exist after the instance has been stopped. Thus, after the instance has been stopped a new instance may be started by using this storage. This storage has its own cost, which is separate from that of the running cost of the instance.
- In order to minimise latency and other network problems EC2 uses multiple locations with which to host its clouds. When a new instance is created a location may be chosen which is close the client. Currently EC2 has eight locations.
- Amazon provides Elastic IP Addresses. In other words, an account will be assigned an IP address or a range of addresses. This address space belongs to the users. When removed and new users are created, they will have an IP address from this address space. This, in turn enables easier integration on the user's side.
- By using Amazon's Virtual Private cloud, a company may connect to their instances using a VPN.
- Amazon Cloud Watch is a monitoring tool that provides the user with feedback on the instances, including performance details.
- EC2 provides auto scaling functionality through the use of monitoring tools. The instances may be scaled up or down depending on the demand. If a high load is demanded, new resources can be created and, when few resources are used, the resources may be minimised. However, this can happen only within the boundaries configured by the user.
- Elastic load balancing provides server farm like functionality [82]. The system will divide requests amongst the instances of the user.
- Some users require a large amount of computing power and, therefore, Amazon introduced high performance computing clusters to enable this. A cluster is provided with a large amount of processing power on both CPU and GPU.
- Users are able to import instances to EC2 using VM Import.

3.6.2 VMware vSphere

VMware is a provider of private IaaS cloud infrastructure [18]. VMware provides cloud computing through a combination of VMware ESX and VMware vSphere. ESX is the hypervisor providing the virtualisation and it enables instances to be run on the node's hardware. VMware vShere is the cloud operating system that is in control of the cloud utilising ESX. vShere runs in an instance on top of ESX. They are one of the older companies in the cloud industry and have been operating for over 11 years. vSphere is being used by over 170000 users, most of which are on the Fortune 500 Companies. This means that VMware vSphere is one of the biggest private IaaS cloud providers in the industry.

VMware maintains that their strategy is not to buy new hardware, but to convert existing infrastructure to an IaaS cloud. This means that companies have a lower initial cost. The infrastructure may then be updated as needed. VMware maintains that hardware utilisation can be improved by almost 80% and that operating cost may be reduced by up to 30% [83]. This, in turn, makes cloud computing feasible for the private sector by allowing the private sector to adapt IaaS for its infrastructure. VMWare also provides the functionality to outsource instances to the public domain on public clouds, thus enabling companies to move some of their instances to external service providers.

VMware vSphere may be interacted with by using a client application, an API and a web interface. The API enables the programmers of the users of vShere to automate the creation and allocation of instances on the cloud as needed. It is also possible to use vSphere with vSphere vMotion. The latter allows instances to be moved between nodes. If a node needs to be maintained, then instances may be moved from it without any interruption in the service to the instance.

Another feature provided is vSphere High Availability in terms of which instances may be restarted if the node on which the instances is running fails. When vSphere High Availability is used with vSphere Data Recovery then all data is backed up and the instances are able to restart with either little or no data loss. vSphere also enables CPUs and RAM to be added while the device is running. vSphere also enables a network card, virtual hard drives and other devices to be both added and removed while the instance is running. The users are able to define business rules in terms of which certain processes and applications get higher priority on the host to enable the faster completion of a process and application that are important to the user.

3.6.3 Xen

As compares to proprietary VMware, Xen is an open source project for the provision of private IaaS clouds [70]. Xen provides the infrastructure to several public cloud service providers, including Amazon Web Services, Rackspace Hosting and Linode [70]. In order to deliver cloud computing Xen combines Xen Hypervisor and Xen Cloud Platform. The hypervisor runs between the node and the instances while the Xen cloud platform is the cloud operating system which controls the cloud. Similar to VMware the Xen cloud platform is enterprise ready.

Amazon EC2 uses the Xen hypervisor as the hypervisor for its cloud. Amazon builds an in-house management services on top of the hypervisor. Xen offers most of the features and advantages that VMware offers, including the ability to live migrate instances.

3.6.4 Nimbula Director

Nimbula Director is yet another private IaaS cloud provider [19]. This private cloud may then be rented out to make it a public cloud or else multiple companies can work together and, thus, it may form a community cloud. Nimbula can interact with EC2, thus making it possible to create a hybrid cloud. Nimbula Director is extremely user-friendly. Users are able to upload template images while any number of instances may be launched from that template image. When adding a new node, little interaction is required on the part of from the user, enabling the new node to become part of the cloud. A cloud running Nimbula Director is easy to maintain and to use while Ninbula Director will maintain the cloud in the best working condition.

The interaction with Nimbula may be done using either a web interface or an API [19]. The API is useful for the automation of the overall cloud system. Nimbula claims that, its authentication policies and policy based network isolation mean that its cloud is one of the most secure in the industry [19]. Each user is assigned roles and can perform certain actions only and interact with a certain part of the system only. This also applies to instances that can also only access and do as is prescribed in their policy.

Nimbula is also working on integration with VMware [84]. This, in turn, will enhance the security and automation provided by Nimbula to VMware. Nimbula has been awarded prizes for its innovation [85], thus substantiating its claims that it is innovative and that its product, Nimbula Director, is a competitive product as regards IaaS.

3.7 Current Research on Cloud Computing and the Concerns of Cloud Computing

One major field of future research in cloud computing is the security of cloud computing and the concerns that may arise from cloud computing [9].

Almorsy is conducting research on a cloud security framework [86]. This framework focuses on the security management of a multi-tenant environment. Almorsy's study addresses the following concerns: the security requirements needed to protect multiple users in the cloud, the security controls that may be implemented to manage the service adoption risks and authority, who is the supplier of security controls and what the metric is that should be used to measure the security in the cloud.

Other research that is being conducted is focusing on the challenges facing cloud computing and the benefits of cloud computing [87]. These challenges include data storage technologies, data management technologies and programming models. The storage should be fast and reliable but still affordable. In addition, new distributed technologies need to be developed and tested to confirm their use with cloud computing. Multiple users are able to access the same data storage on the cloud and, thus, it is essential that the data management allows for fast and effective data response time. Cloud computing relies on a form of distributed and parallel technologies. Developers need to be made aware of programming methodologies and programming techniques to ensure the efficient use of the cloud resources. The researcher has also identified other challenges faced by cloud computing as data security and privacy policies, models for data access and storage, and standards for resource accessing.

Chow examined the security concerns of cloud computing [9]. This author conducted research into concerns regarding security in the cloud and identified the following main categories, namely, traditional security, availability and third-party data control. Traditional security includes attacks that may happen to physical computer equipment. These attacks include different forms of intrusion of either the computers or the networks. However, such attacks may also happen to the instances running on the cloud or on the infrastructure of the provider. It is, thus, essential that the users of cloud computing ensure that their resources are protected from an attack on the infrastructure of the provider while the cloud providers must ensure proper authentication and authorisation throughout the cloud. Chow also identified the risk involved in normal forensic techniques where all the equipment has been seized.

Chow also identified availability as a category of concern [9]. Availability is one of the characteristics of cloud computing. Cloud providers maintain that availability is one of the advantages of switching to the cloud. A single point of failure is usually avoided but, in cloud computing there may be a stronger release on a single point of failures and this will need to be resolved.

The final concern that Cow identified was that of third-party data control [9]. The regulation of data may be complex and it will need to be studied in detail. The exact owner of the data is not always clear in cloud computing. Also, during an auditing process it may be difficult to distinguish between the cloud provider and the user data. It is, thus, essential that the users of the cloud fully understand all the implications of the Service Level Agreement as some data rights may be lost if the data is in the cloud.

In addition, Chow indicated certain new problem areas that may arise from cloud computing [9]. These problem areas include: cheap data and data analysis, the cost-effective defence of availability, increased authentication demands and mash-up authorisation. Data is easily gathered and kept using the cloud. For example, Google use its data which it gathers from the cloud for advertising. However, the problem is that attackers may gain access to large data sets and, because the data is on the cloud, the attackers have the resources to extract meaning from the data.

In order to enable users to defend their data on the cloud they may have to pay unnecessary bills [9]. Cloud computing often relies on thin clients and, as stated previously, the thin devices require a better authentication method as thin clients are usually not very secure. The need for external authentication will also become necessary as instances on the cloud communicate with other entities on other clouds and, thus, authentication between these entities should be required.

3.8 Value Added from Cloud Computing

The value that may be added from cloud computing is significant, but primarily as regards small, medium and micro enterprises (SMME) [72]. These enterprises espouse a survivalist methodology [88]. They are usually started without large start-up investments and with a specific knowledge base. Cloud computing enables such enterprises to gain access to servers without the initial start-up cost while they incur no maintenance costs in terms of hardware. As the businesses grow, their infrastructure may easily be changed to

adapt to the growth. Another added benefit is that the enterprise does not need most of the knowledge which is required to build the infrastructure, as the cloud service provider is responsible for building the infrastructure.

3.9 General layout of a Cloud

To aid the reader in later chapters this section will give an example layout of a cloud. A cloud root is a data center, this is all the resources of the provider/owner of the cloud in one specific location. e.g. one building. A data center is then divided into cluster, a cluster is a logical grouping of resources. One cluster might be all the financing departments resources, another cluster all external clients and the last all internal resources. A cluster consists amongst others of nodes and instances, where a node is a server providing the resources and a instance the virtual resource.

3.10 Conclusion

Cloud computing is growing and it is believed that it will become a billion dollar industry [35]. The main reason for this is that some of the largest IT related companies have implemented or are implementing cloud computing solutions, including, Google, Microsoft, IBM and Amazon [74, 3]. These companies maintain that they will provide CIA to their customers by using various techniques.

The focus of this chapter was to introduce the concept of cloud computing. The chapter provided a brief history of cloud computing and both the service and development models. The current implementations of importance for this study were introduced and the present state of cloud research was explained.

The next chapter will focus on cloud forensics as well as the growth and current state of cloud forensics. The chapter will also discuss the correlation between cloud forensics and other digital forensics sub-fields.

Chapter 4

Cloud Forensics

4.1 Introduction

It is possible that the cloud environment may complicate the process of conducting a digital forensic investigation (DFI) [35, 89]. The dynamic environment of the cloud is just one of the problems faced by a forensic investigator. Nevertheless, cloud forensics evolved as the need for a new subfield of DF became apparent.

Cloud forensics refers to the digital forensics that is applied to cloud computing [35]. In view of the fact that as a cloud runs on a network and consists of network equipment it follows that cloud forensics will entail network forensics. However, cloud forensics also entails computer forensics as a cloud consists of nodes that are computers. Instances on the cloud are also acceptable to computer forensics and this, in turn, means that cloud forensics links computer and network forensics. Cloud computing may include most of the subcategories of digital forensics but this does not mean it is not a sub field of digital forensics.

The field of cloud forensics is ever evolving as new methods and problems are being discovered [3]. The field is still characterised by certain unknown factors and the research conducted on cloud forensics aims to identify the caveats and propose viable solutions to the problems which have been identified. The cloud forensics field is expanding to a point at which it may be used to conduct a successful investigation in the cloud environment.

4.2 The correlation between Cloud Forensics and Other Digital forensics (DF) Subfields.

Cloud forensics may be regarded as a new subfield of Digital forensics (DF), although it does include other DF subfields [3, 36, 35]. As stated in chapter 2, DF incorporates a combination of proven methods with which to gather evidence. The methods used in some of the digital forensic subfields may be modified so as to adapt it to the cloud environment or else cloud specific steps may precede existing methods in order to conduct such methods in the cloud environment.

The DF categories that are applicable to the cloud environment include, among others, computer forensics, network forensics, database forensics and mobile forensics. In the cloud there are nodes and instances, which constitute forms of computers and which are acceptable to computer forensics. The nodes and instances may be connected through a network and network forensics may be used to gather network evidence. Cloud operating systems incorporate some form of database to store information about the cloud and activities on the cloud. These databases may provide information when an investigation is being conducted. Another example of database forensics being used on an instance may be where there is a database running. Mobile forensics is another subfield of DF. As regards mobile forensics the cloud is able to provide applications that serve mobile devices, thus, mobile forensics on the cloud and the devices can be used. The remainder of this section will explain each subfield in more detail with an IaaS cloud being used.

4.2.1 Computer Forensics

As stated above the cloud consists of nodes and instances and these are both forms of computers with the nodes being physical computers and the instances virtual computers within a virtual environment. The nodes contain information about the service provider and possible information about the clients of the service provider [90]. The information of the clients resides mainly on instances.

Computer forensic methods may be applied to both nodes and instances, although there are certain restrictions that are caused by the virtual environment. These restrictions include the fact that it is not possible to take an instance apart and nor is it possible to remove its hard drive for analysis [12]. Tools or methods which are specific to virtual

environments are required in order to analyse instances inside the cloud [89]. Although normal computer forensic methods may be applied to the nodes, the problem is that multiple instances may reside on a single node and it may not be possible to gain access to all the instances on a node. In addition, it should be taken into account that the node operating system may contain little or no useful information. Also, the operating system is used to maintain the cloud and clients and operators of the cloud have limited direct interaction with the node itself.

4.2.2 Network Forensics

Networking is one of the basic building blocks of cloud computing [3, 90]. Most of the communication and management in a cloud happens through the network. The network may be divided into the physical network and the virtual network. The physical network connects the actual hardware which includes the nodes, network storage and internet connection [3, 90]. There may be multiple virtual networks inside the cloud environment. The virtual network connects the instances to each other and also connects the virtual network to the physical network [91]. The virtual network is logically separated by the cloud operating system and communication between networks is not possible without a virtual connection. The separation is done to enable isolated networks to be created.

The networks, both physical and virtual, may provide information when network forensics is being conducted [3, 90, 89]. If there is a need to gather evidence from one client's network on the cloud, then as stated in chapter 2, a network forensic tool may be added to that network. Once the tool has been added it may be used in the same way as it would be used on a physical network. When capturing information from the physical network of the cloud there are certain factors that need to be taken into account, namely, data volume on the network, information owners and encryption [3, 90, 89]. On the physical network of a cloud there may be vast amounts of data being sent over the network and, in some cases, thousands of nodes and instances may be using the network. The vast amount of data possible may complicate the successful completion of a live investigation. Yet another problem is information ownership. In a public cloud the data may belong to the various clients of the service provider and, therefore, it is essential that the traffic be filtered in such a way that relevant evidence only is collected. The last consideration is that, on the virtual network, a VLAN with encryption may be used. This means that the traffic flowing over the physical network is encrypted and it may not be possible to analyse.

The next subsection will describe the correlation between database forensics and cloud forensics.

4.2.3 Database Forensics

Inside the cloud environment databases may be found, either on the instances or on the nodes [90, 36]. Databases have two purposes that may be useful inside the cloud. Firstly, there are the clients' databases that contain information related to the clients on their instances and, secondly, there are the databases which are used by the cloud operating system [36]. The database containing client information may be a possible source of client related evidence. For example, clients may host their payroll solution on the cloud. Thus, if there were a suspicion that someone may have suspiciously changed their salary an investigation would be conducted on the database of the client. The client should have full control over their database.

The second purpose of the databases inside the cloud involves the use of the cloud operating system. The cloud operating system may store configuration information, log trails, audit information and usage statics [18]. Thus, the information inside the cloud operating database may be used to assist in an investigation when the cloud environment is being investigated. The information that may be gathered from the database might include changes made to instances and by whom such changes were made. The information may also be able to point out changes in the normal behaviour of the cloud.

4.2.4 Mobile Forensics

The cloud environment provides a variety of applications to the mobile computing market [92, 36]. These applications consist of two parts; one of which is on the mobile device and the other which is on the cloud. The part on the cloud may be an application which is running and which uses some form of database. The device communicates with this application using some form of network. This means that, on the cloud, network and database forensics can be used in situations in which mobile forensics can be used on the device. In order to conduct an investigation in which both the cloud and a device are involved multiple forensic fields would have to be used. The cloud would store most of the information inside the application database while the mobile device would be

responsible for the presentation of the data. The mobile device may be regarded as a thin client (See chapter 3). In this case most of the forensic evidence would be located on the cloud.

When conducting an investigation inside the cloud multiple forensic subfields may be required [36]. The subfields required may be determined by taking into account the demand for the subfield in question. Using the subfield by combining methods from the fields may reduce the complexity of an investigation as proven methods would simply be reused. Cloud forensics itself is still in the developing stage and, thus, the next section will expand upon the development of cloud forensics.

4.3 Development of Cloud Forensics

The growth of the cloud and the number of users on the cloud (see chapter 3) have resulted in the need for forensics. As seen in chapter 2 the need for digital forensics arises when computers are used. The new realm of cloud computing has meant that concerns of digital forensics have shifted into the cloud environment. As discussed in chapter 3 the issue of multi-tenants inside the cloud environment has also strengthened the need for digital forensics inside the cloud. The users of the cloud require their data to be secured and their resources to be protected from unauthorised access and thus, digital forensics and forensic readiness will help with ensuring the CIA of users on the cloud.

The field of cloud forensics is relatively new and is a valid topic for further research [1, 3, 93, 35, 90, 94]. The research is in the process of defining the field of cloud forensics. One researcher has defined cloud forensics as part of network forensics [35]. There are, however, also other opinions and Reilly argues that cloud computing uses models which were adapted for digital forensics [95].

Ruan et al proposed one model for cloud forensics [35]. In the article they suggested the following three dimensions of cloud forensics, namely, organisational, legal and technical. The technical dimension consists of the tools and techniques which are used to conduct an investigation on the cloud. The purpose of these tools and techniques is to enable an investigation to be conducted inside the virtualised, and in some cases, distrusted environment. The investigation is conducted in order to collect forensic data which can then possibly serve as evidence. As part of this dimension the tools and techniques may be used to form part of a proactive forensic toolkit.

The organisational dimension is related to companies using and/or providing the cloud resources. In addition, the organisational dimension also includes users of the cloud, either external or internal to the provider. The users include the users of the cloud as well as the IT professionals maintaining the cloud. The users may also include those individuals handling the incident and who start an investigation as well as the legal advisors of both the users and providers of the cloud.

The legal dimension is related to the judicial system of the country in which the cloud resides or to the multi-jurisdiction systems if the users of the cloud and the cloud itself fall under different judicial systems [35]. The service level agreement (SLA) between the users and the providers is also relevant in this context, if the cloud provider is providing resources to an external user or client, then an SLA will govern the services rendered.

There are some advantages to cloud computing that investigators may use to assist them in a forensics investigation. One such advantage is the fact that all the data may be centralised in a single data centre [95]. The cloud provider may possibly backup data on the cloud and this, in turn would make it easier for investigators to gain access to previous versions of the data. The cloud provider may also hash encode the data which may enhance the admissibility of the evidence gathered from the cloud. The clouds resources may also be used to assist in the investigation by providing an instance which may be used for password cracking. The investigators may also be able to create clones of live running systems without any interference to the running instance [18].

There are however, also disadvantages involved in conducting an investigation in the cloud, one of which is the issue of location inside the cloud. The acquisition phase of an investigation may be tedious because the exact location of data is not always known and it may be difficult to determine.

4.4 Different Cloud Models

As stated in chapter 3 the cloud may be divided into various deployment and service models. These models may, in turn, all have different effects on an investigation on the cloud. When conducting an investigation on an IaaS it must be borne that the cloud service provider is responsible for the underlying hardware and the hypervisor while the client is responsible for the operating system, applications on the system, and all the data on the system. This means that the cloud provider and the client should work

together if the investigation inside the cloud environment is to be successful. On the other hand, when an investigation on a public IaaS cloud is conducted the customer is external to the provider and may even be in a different judicial system. This implies that there are certain factors that should be borne in mind during the investigation if the investigation is to be brought to a successful conclusion. The most important factor in respect of private IaaS is that companies are replacing their in-house servers with virtual in-house servers. This means that cloud infrastructure is on the company's premise and, therefore, the company itself is responsible for conducting an investigation when needed.

In a PaaS model, the cloud provider provides the hardware, hypervisor, operating system as well as some applications for the client. The client then modifies the applications and may even add more applications and data. When an investigation on a private PaaS cloud is conducted, the provider and customer may be located in the same judicial system. However, when an investigation on a public PaaS cloud is conducted, the judicial system where of the customer and the provider are located may differ. The main difference between an investigation on a public PaaS cloud and an investigation on a public IaaS cloud involves the responsibilities of the provider and the customers with the provider in a PaaS being assigned more responsibilities than the provider of an IaaS cloud. Nevertheless, as with IaaS it is essential that the client and the provider work together to ensure that the investigation is completed successfully.

The SaaS service model provides applications to the users. This, in turn, implies that the cloud providers maintain the hardware, hypervisor, operating system and applications while the clients use the applications and may also store data on the cloud. The client owns the data he/she stores on the cloud only and, thus, in the event of an investigation, the providers of the cloud will be responsible for most of the investigation. However, the CIA of the clients' data must still be protected and if the client reports an erroneous event, the client is entirely dependent on the provider to investigate the problem. The client also has little to no say over the investigatory process.

4.5 Considerations for Cloud Forensics

Current literature about cloud forensics is scarce, this section provides a overview of the current literature about considerations for cloud forensics. In a paper by Ruan et al [93] the terms and ideas which should be included in an SLA if the cloud is to be prepared for an investigation are introduced. The SLA determines what the client may expect of the

service provider and vice versa. An SLA is usually used in a public cloud environment and may be specific to each customer. The points, terms and ideas included in the SLA form the basis for the issues related to both the customer and the provider that need to be taken into account during an investigation. The remainder of this section will expand upon such issues in a public cloud. These include access to relevant data, logs from provider and client, location of data, virtual resources of the cloud, personnel required, cloud maintenance and auditing.

During an investigation it is essential that the investigators have access to all relevant data in order to gather the required evidence [93]. However, in the cloud there may be some restrictions as regards data access and thus, there should be some agreement in respect of on access to the data on the cloud. Such data includes, but is not limited to, encrypted data on the cloud, logs from the customer and the provider, the physical location of logs, the physical location of data and the access to virtual disks. Some cloud providers offer their clients the option of encrypting data on the cloud and customers may install tools with which to encrypt the data. The functionality and cooperation between customer and provider should exist contractually so as to enable access to the encrypted data and the keys if there is a justifiable reason for such cooperation. This, in turn, would simplify the investigation process as regards encrypted data.

The customer and the provider may keep logs of both the systems and of the components on the system. As discussed in chapter 1, these logs may be used for a multitude of reasons in a DFI. It is essential that the customer and the provider decide on the reasons for and methods of cooperation between them. These logs must be kept safely in the event of failure of the system and, thus, precautions such as off-site storage need to be considered. The customer and provider should also agree on a retention period for the logs and the methods of drawing up these logs.

As discussed in chapter 2 the location of data inside the cloud is mostly unknown. However, in order to conduct an investigation, the location of data may be required, as well as access to the data. The providers may be able to support this functionality so as to enable an investigation to be conducted.

On the majority of cloud platforms the users may create a form of virtual disk. These disks may contain valuable resources and it is, thus, incumbent on both the customer and provider to determine how and whether these disks would be made available during an investigation. If the information is required in terms of a warrant, the laws of the country in which the data resides would determine the availability of the disks.

When preparing and conducting an investigation various personnel members may be involved [93]. It is, thus, essential to take the following factors in respect of people management into account, namely, the responsibilities of the investigating team, the certification required by team members and collaboration between teams and external partners. Both the customer and the provider need to determine the investigatory roles required on their side and make provision to full these roles. The team members should all be qualified to a certain level which would be agreed upon so that specific requirements may be satisfied. The teams would have to cooperate during the investigation. This cooperation may be planned and agreed upon. The appropriateness as regards the need for an external investigating examiner should also be discussed. The investigation might also be required by law and, thus, cooperation with the law enforcement agencies may have to be decided upon by both the customer and the provider.

The customer may choose the country in which the hosted data will reside and, if the client is offered a choice by the provider, the client may choose a location where legislation is in agreement with the laws that pertain to the customer's location [93]. If, however, the customer is not offered a choice, then the provider may disclose the location of the hosted data to the customers. The provider may be open about the way in which data of multiple customers will be separated logically, but simultaneously hosted together. In the event of an investigation the data may be separated. Another consideration is data ownership. It is imperative that the ownership of data on the cloud be transparent. The customer and provider may also decide on the methods to be used in event of migration between providers. The customer should have the right to change providers and should not be limited to one provider only. The customer and provider should also ensure that the other party is notified timeously about any investigation that being either planned or conducted.

The process of maintaining the cloud may be time-consuming. However, it is essential that this be carried out with due care. Some forensic considerations need to be incorporated into the cloud operation process. These considerations include an agreement on proactive forensic gathering, the deletion and recovery of data, the tools that may be used, incident repose, and recovery and time synchronisation [93]. When the cloud operation is planned the process of proactive forensic gathering must also be planned, the sources of the evidence determined, and the necessary tools configured with which to capture the forensic data. Where and how the data collected should be stored may also have to be decided. The procedures in terms of which to delete data on the cloud should be planned. The cloud data may be deleted on the clients side, but still be kept on the providers side. The client and customer should decide on the time period of retention

and whether instant deletion is possible.

As discussed in chapter 1 there are various tools available which may assist the investigation. However, these tools need to be configured if they ate to work accurately inside the cloud environment. The customer and provider also need to decide on the tools that may be used under certain conditions [93]. A process plan must be formulated in the case of an incident, the necessary steps need to be identified, and roles assigned to personnel members. This would enable the customer and provider to respond to and recover from an incident. The final operational consideration is time. The customer and provider need to decide on a time synchronisation strategy that may be used during an investigation as events from both the customer and provider may be correlated on a timeline.

The last considerations involve auditing [93]. The customer and provider may have to review the SLA at regular intervals to confirm its compliance with any investigation that may be needed. This section described a few of the issues that the customer and the provider may hate to take into account in cooperation with each other.

4.6 Problems Faced by Cloud Forensics

Currently there are various problems that may hinder the successful completion of an investigation on the cloud [96]. It is, thus, essential that these problems be addressed if an investigation on the cloud it so be brought to a successful conclusion. These problems include the distributed nature of data on the cloud, for example, the data may not be contained in one location, different time zones in which the data may reside with the logs perhaps not correlating if the time zones are not managed correctly and, if an investigation covers multiple judicial systems, it may be impossible to gather admissible evidence from all the regions. Another problem involves data acquisition as it may not be possible to follow normal forensic methods in order to gather physical drives while it may also be impossible to retrieve deleted data in the volatile environment of the cloud. The last problem involves is the terms and knowledge of cloud computing that are still new and may be lacking.

4.6.1 Distributed nature of the cloud

The first problem faced by cloud forensics involves distributed nature of the cloud. It is not possible to acquire and seize the node of the service provider as part of a "dead" forensic investigation [96]. In addition, the nodes may also not contain the information required by the investigation team. The nodes may contain the running information of instances only and the persistent storage may be on external storage device such as a SAN. If the investigation is being conducted in order to collect information about a single user there are several factors that need to be taken into account. These factors include the location of the storage, whether the storage of the node is external or internal to the node. Another factor that must be taken into account involves the number of instances. For example, if a user has multiple instances they may not all reside on the same nodes. In order to gather all the evidence related to the client it may be necessary to acquire multiple devices from the service provider. This is not accessible if the CIA of other users is taken into account.

4.6.2 Multiple time zones

A single client or user may own systems over multiple time zones with the systems being located in different data centres across the globe. When conducting an investigation timestamps are made of the evidence in order to enhance the admissibility of the evidence. Logs use time to order events and, thus, the time in logs may also be used to correlate events. This means that, in order to conduct a successful investigation, the time between data centre systems should be synchronised. However, time synchronisation may be an arduous task in a single data centre while a multi time zone and multi data centre may really complicate the process of time synchronisation. When a client uses more than one service provider this may also complicate the time synchronisation as the various service providers may not use the same synchronisation techniques. Thus, this means the client is responsibility to ensure that the time is synchronised between data centres and between service providers.

4.6.3 Multiple jurisdictions

When conducting an investigation it is essential that certain processes be followed. Some of these processes may be dictated by the judicial system and, if the investigation involves

multiple systems with different requirements, it may prove difficult to accommodate all the requirements [96]. The problem may be exacerbated if, for example, a customer owns systems in two different jurisdictions, an investigation may be conducted in one of the jurisdictions and a further investigation may be required in the other jurisdiction. This second investigation may then require evidence from the precious investigation but, because the latter was conducted in another jurisdiction the evidence from that investigation may be inadmissible. If, however, the investigation is conducted internally by a company on their private cloud which spans over several jurisdictions, but there is no intention to use the evidence in a court of law, then the multiple jurisdictions would not be a problem.

4.6.4 Limited knowledge on cloud computing

The final problem does not involve the investigation on the cloud bur arises in the event of the forensic investigator having to testify [96]. The concern is that, although an unimpeachable process was followed, communicating the findings to the jury or panel may be problematic. The jury or panel may have limited computer knowledge and also, because cloud commuting is a relatively new concept, the investigator may not be able to convey either the details of the process followed or the reasons for the processes. One of the tasks of an investigator is to convey an understanding of clouds and the processes followed in conducting an investigation inside the cloud environment. If the jury or panel does not understand the concepts because of their limited knowledge of the field, then the evidence may be considered to be inadmissible.

4.6.5 Evidence acquisition

The next problem involves the acquisition of evidence. When conducting an investigation on computers the hard drives are removed then; using a write blocker, an image is made of the hard drives [96]. The drive blocker is used to ensure a write free read. This, in turn, enhances the admissibility of evidence as no evidence will be either changed or destroyed during the investigation process. The image will be used during the remainder of the investigation. The write blocker is a physical device which is placed between the hard drive and the forensic computer.

4.6.6 Virtual storage

The virtual storage of instances makes it difficult to isolate and gather the virtual disk. The process of acquiring physical disks may not be applicable as multiple clients may share a single resource. The other problem involves the amount of data that may be retrieved. If all the storage devices are gathering data, then, the amount of data collected may make it impossible to complete the investigation on time and within budget constraints. This, in turn, means that the existing procedure used to gather evidence may not be applicable and it may be necessary to device and implement new methods in order to acquire only the virtual disks which are required for the purposes of the investigation.

4.6.7 Deleted data

Another problem involves deleted data. In computer investigations deleted data may serve as a possible source of evidence. Thus, during a computer investigation the investigators may search for deleted files as a suspect may have deleted possible evidence and, thus, have tampered with the evidence [97]. As stated in chapter 3 the cloud environment is volatile. This, in turn, means that, when data has been deleted, the chances of it being recoverable are low. Although data is backed up and replicated in some cloud environments, the service provider may also delete the data from backups as part of a SLA agreement. In addition the main storage area may be used by multiple instances and, thus, the deleted data may easily be overwritten. The processes which are used to recover deleted data from drives may not be applicable in respect of on cloud storage. As previously stated, it is not possible to remove the hard drive for the purposes of examination. Another problem involves the amount of data. Cloud storage may reach several terabytes and also contain multiple files. Multiple instances may reside on a physical storage unit in the cloud, while, during normal day-to-day operation multiple deleted files may be generated. Apart from the vast amount of data that prohibits the recovery of deleted files there is also the problem of file systems. Many cloud providers create their own form of file system to sustain the high demand inside the cloud environment. This, in turn, means that it may not be possible to recover deleted files on the cloud.

4.7 Benefits of Cloud Computing

Despite the fact that the process of conducting an investigation in the cloud may be complicated by some of the cloud attributes as discussed above, the cloud itself may nevertheless also assist in the investigation process. When conducting an investigation the cloud resources may assist in the forensics investigation. The cloud may include methods with which to store the current state of an instance. In the cloud environment, backups are made of the instances and/or the data on the cloud while data inside the cloud is kept together. In addition, the service providers may also assist in and preparing for an investigation.

The first benefit offered by cloud computing is the fact that the cloud may provide its own resources. For example, there may be extremely powerful underlying hardware on the cloud infrastructure. In addition an investigation may require a degree of some processing in order to examine the evidence. The clouds own infrastructure may be used for some pre-processing for analysis with regards to the current investigation [98].

In the cloud it is possible to create an exact copy of a running VM. This functionally may be achieved by the use of either clones or snapshots [99]. A clone creates a new VM from a current VM while a snapshot is a backup that was created. A snapshot may also include the RAM of a system. This functionality is provided by amongst others, VMware, Xen and Nimbula. A clone is an exact copy of the VM to a new VM and it will copy all content of the drive exactly [99]. This copy may then be used as a normal VM while the location of the copy may be specified during the clone procedure. A snapshot is used for short-term backup, where only the difference is saved. This, in turn, means that when changes occur on the system, only the changes will be saved. This clone or snapshot may then be used as a starting point for an investigation while the original VM remains untouched. This also means that the original suspicious environment may be monitored without interference in order to create investigation images.

In the cloud environment backups may be made of the environment so as to comply with the SLA. These backups may also be used when conducting an investigation. If the investigators are granted access to the backups then they have access to some of the data before the investigation has even commenced. This may prove useful when there is a possibility that data has either been corrupted or tampered with. In addition, the investigators are able to construct a timeline or event trace using the backups.

The data on the cloud is kept close together, usually in the same data-warehouse.

However, as stated previously, this may cause problems although there are also certain advantages [98]. The advantages include a narrower search space, and high-speed storage. The investigators would have some indication as to where the data is and this, in turn, would narrow the search and enable a more focused investigation. The high-speed storage also helps to speed up the investigation as it makes possible significantly faster data copying and replication.

Another possible advantage is cooperation from service providers and the service being forensically prepared. Cloud providers and their customers are contractually bound by a SLA. There is a movement toward including digital forensic readiness in the SLA and this, in turn, would mean that both the service provider and the customers would be prepared to conduct an investigation when required and that the investigation would have a sound starting point in view if the rules and regulations laid out in the SLA [93].

4.8 Conclusion

This chapter introduced the current state of cloud forensics. This chapter focused on comparing to the other fields of digital forensics. The reference to the other DF fields was explained in terms of cloud computing. For reader knowledge the current problems in cloud forensics were discussed as were the advantages of the cloud concerning in respect of forensics. The factors that should be taken into account when conducting an investigation inside the cloud environment were also introduced.

The next chapter will introduce the notion of isolating a single instance inside the cloud environment. The chapter will also explain what it means to isolate an instance. The methods used to isolate an instance will be tested and evaluated and the results discussed.

Chapter 5

Isolating a Single Instance

5.1 Introduction

The term "isolate" is defined as the state of being identified and then separated from others [100]. Thus, isolation involves the process of isolating. In a "real world" forensic process the crime scene is isolated [101]. This isolation helps to protect the possible evidence from both contamination and loss of continuity. In order to enhance the admissibility of the evidence a crime scene is divided into separate sections so as to facilitate the isolation. These sections may be entered by authorised personnel in an authorised manner only. During a "real world" forensic process a path is laid out at the scene of the crime and which the personnel may use to walk in and around the crime scene. A log is kept of where the personnel members are and what they are doing.

In the realm of digital forensics the process of isolation is also used although the terminology used to describe isolation differs. For example, seized cell phones are placed inside a Faraday bag [15]. This, in turn, prevents any communication between the cell phone and the outside world, a form of isolation. When conducting hard drive forensics, a write blocker is used to enable a write free read [16]. The write free read protects possible evidence from contamination. Thus, as stated above, the isolation process is being used in the field of digital forensics although the term "isolation" is not often used in this context.

In a cloud an instance must be isolated when it becomes apparent that an incident has happened on that particular instance. This isolation helps to preserve the integrity

of the evidence collected from the instance. However, one of attributes of clouds may contribute to one of the problems associated with preserving the integrity of an instance [3]. In a cloud, the data from one instance may share the storage of multiple instances and it may also not be in a constant place in the cloud. If the integrity of the evidence is to be preserved, then it is essential that the location of the evidence on the cloud be known while the evidence must be protected from tampering and contamination. Another problem is the fact that other instances on the same node may belong to other users. Users should, at the least, be able to expect the availability and privacy of their instance provided from the service provider [9]. Accordingly, it is imperative that the digital forensic process be conducted in such a manner that it will not result in the loss of privacy of other instances while the availability of the instance must be affected as little as possible.

Gathering evidence is one of the aims of an investigation. If there is any suspicion that the evidence has been rendered invalid by any means then it will not be able to serve as admissible evidence. In order to ensure the admissibility of the evidence, the evidence must be protected from both contamination and tampering. Thus, the need for isolation in the cloud environment becomes apparent when the issue of the admissibility of the evidence is taken into account.

5.2 Isolation of a Crime Scene in a Cloud

As discussed in the chapter on digital forensics, in order to isolate an incident the cloud instance needs to be isolated. The node instance is effectively placed in a controlled environment so as to enable a controlled investigation to be conducted. This study introduces possible techniques which may be used to isolate these cloud instances in order to facilitate an investigation. These techniques were created from other fields in computer science and should provide enough diversity for testing purposes. The techniques include, but are not limited to, Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, Man in the Middle (MITM) and Let's Hope for the Best (LHFTB). This list is the fist attempt at providing techniques for a DFI, other studies might add to this list. A discussion of each of these techniques will follow later in this section. This discussion will include a description of each of the techniques, the advantages and disadvantages of using the techniques and the visibility of the techniques.

The methods for clearing a node include either moving the suspicious instance to

another node or moving the uninvolved instances to other nodes. The CIA of the other instances is protected when the suspicious instances is moved. However, this may result in the loss of possible evidence. When an instance is moved data may be lost or the instance might realise it is being moved and, therefore, tamper with the evidence. Thus, to protect the evidence the other instances are moved from the node. However, care must be taken when moving the instances in order to protect their CIA.

When isolating a cloud instance the investigator must consider whether a live or dead analysis is applicable. The techniques that are suited to each type of analysis may differ. When conducting a live forensics analysis the aim is to stop the instance from tampering with evidence. On the other hand, if a dead analysis is chosen the other instances must be protected from the possibility of a power outage. Thus, a decision has to be made regarding what losses and risks may be acceptable before commencing with an investigation in a cloud.

5.2.1 Instance Relocation

Instance relocation is the first technique for isolation that will be discussed. Instance relocation may be defined as the movement of an instance inside the cloud and which is done by moving the instance from one node to another. This may be done either manually or automatically. When done manually the administrators of the cloud will usually move the instance by some means while automatic relocation is done by the cloud operating system. The instance may be moved in one of three possible ways, namely, the existing instance may be ended and a new instance created, a new instance may be created in which case the old instances should be destroyed once the new instance has been created, or the instance may be logically moved. The latter entails the data being moved from one node to another without the instance being destroyed.

In order to move an instance the instance must be divided into three moveable units. These units include the data on the secondary storage, the content of the virtual memory, for example swap memory and the running processes.

Manual Instance Relocation

When an instance is moved manually it is the responsibility of an administrator or investigator to move the instance. The possible methods that may be used to relocate

an instance manually are a subset of the methods discussed above. The existing instance may either be ended and a new instance created, or a new instance may be created and the old instance destroyed. When an existing instance is ended all of the units must be protected or saved. This may be done by a variety of available methods. The storage may be copied to an image file using tools including dd_rescue [54]. The content of the virtual memory may also be written to files by using dd_rescue. Once all the files have been created the original instance is removed and a new instance created. The new instance will receive all the content of the old instance. The new instance may be created with the same network address as the old instance but it must be created on a different node. However, a possible problem may be encountered with the running processes because it is difficult to store processes in such a manner that they may be restored to a new instance. The other method involves creating a new instance and moving all of the units to the new instance, after which the old instances are removed. Once the new instance has been created the storage content may be moved to the new instance. The running process may be moved using which have been designed to move processes between computers [102]. Some of the methods for moving processes as proposed by Milojíčić have been tested and proven to be valid. However, the virtual memory is more difficult to move and care must be taken during the moving process. Virtual memory is difficult to move because, while the instance is being used, the virtual memory is in a constant state of change. Once the new instance has all the units of the original instance then the original instance is removed. The new instance must then be set so that it has the same network address as the old instance in order to receive the network traffic.

Automatic Instance Relocation

The cloud operating system may move the instance by using an automatic instance relocation technique. The methods used to move instances are implemented by the creators of the cloud operating system. The creators must ensure that the method is proved to be credible and that it is reliable. The reasons for an instance to be moved by the system include, but are not limited to, the administrator, or investigator, requesting the system to move an instance or for load balancing purposes. The administrator, or investigator, may request the system to move an instance for the purposes of an investigation. Other possible reasons may include a conflict of interest between instances on a single node. In addition, the load balancing functionality may be implemented in cloud operating systems. When the system notices that instances on a node are extremely resource intensive, and other nodes have adequate resources available, the system may

endeavour to balance the load of the nodes. This functionality may be used by an investigator. The investigator might force an instance to be resource intensive after which the system will move the instance away from the node.

The above mentioned instance relocation techniques enable a node to be cleared for an investigation. The relocation techniques may be controlled and monitored and the service provider may prove to its customers that the investigation is protecting their CIA. The cloud operating system manufacturers may also implement reliable methods so as to enable an investigation on their cloud system to be conducted successfully.

If the instances are moved in a manner that violates their CIA the service provider may be adversely affected the customers may experience either downtime of their instance or a loss of data. The customers may then leave the service provider or the provider may be held liable to reimburse the financial costs of the down time. If the cloud operating system moves the instances it may be difficult for the investigator to prove that they were using reliable methods. This aids assurance from the manufacturers to be involved in an investigation.

It may be challenging to implement the instance movement techniques. As was discovered through experimentation the storage media may easily be copied but although it is no easy task to send the storage data to the new instance while ensuring that instance continues to run. The hard drives would have been copied fully while the process of overwriting system files may result in the failure of the new instance. The process may be moved if the operating system of the instance supports such functionality.

5.2.2 Server Farming

A server farm is a multi-node system that forms one logical resource [82]. In web-server farms the web-site is split over two or more nodes. The user interacting with the website sees the functionality of a single server only. In the farm multiple nodes are used to serve the website. The server farm uses some form of routing to route requests between nodes and users. The server farms use distribution technologies to enable this service. This distribution aids in the quality of service of the website. There is no single point of failure and, in the case of a node failure the router will stop sending requests to that node.

In a cloud multiple instances may be created that form one logical resource over multiple nodes. The load for the logical instance is spread over the actual instances.

When a single node fails the remaining instances will continue to function and this, in turn, enables examiners to terminate instances on the same node and to isolate the suspicious instance on a node. It is essential that small server farms of the uninvolved instances are created at the start of the investigation. These small server farms may be created by adding just one instance to the farm. This means that the farm will contain two instances. Once the server farm is fully operational the original instance may be removed.

To enable server farming on clouds the creator of the cloud operating system must implement such server farming and, thus, the cloud infrastructure must provide for the rerouting of network traffic. However, the process of creating a server farm for the sole purpose of an investigation may place an unnecessary load on the cloud. If the cloud provides the functionality to provide availability to its clients, the HA may be used to assist an investigation.

Although server farming may be resource expensive it may also assist the service providers to manage their client CIA. The instances may be removed from the node without loss of availability.

The server farming technique relies on the cooperation of cloud operating system creators. If the implementation of the server farming technique is wrong the investigation can result in the loss of the CIA of other users on the cloud.

5.2.3 Failover

In a failover environment there is at least one server replicating another server [103, 104]. The replicating server is commonly known as the backup server. If the primary server fails the backup server will take over. This, in turn, means that all the data and processes of the primary server are replicated on the backup server. Failover was introduced to provide high availability for websites. In 1999 E-Bay lost an estimated 5 million dollars when their servers failed [105]. If failover technology had been implemented this problem could have been averted. Failover may be provided in several ways, including client-based failover, DNS-based failover and IP-address take over [103]. In client-based failover the client is aware of both the primary server and the backup server. If the primary server is unresponsive then the client communicates with the backup server. When using DNS-based failover the DNS server redirects traffic to the backup server when the primary

server fails. In IP-address take over the backup server takes over the IP-address of the primary server when the backup server notices that the primary server has failed.

In order to implement failover an adaptation of the IP-address take over will be used. The original instance is replicated by creating a backup instance. Once the original instance has been killed the backup instance will take over the IP of the original instance. The method used to replicate the instance is at the discretion of the investigation team. As with instance relocation, to replicate the same units, namely, secondary storage, the content of the virtual memory and the running processes.

The failover technique will result in virtually no loss of the availability of the instance. The failover may be implemented by the investigation team. There is almost no reliance on the cloud operating system manufacturers. In addition, this technique also does not make use of extensive resources of the cloud.

5.2.4 Address relocation

Address relocation involves relocating network traffic to another computer. The network traffic is redirected by either the router or the DNS server to another computer. In general, network packets are sent to specific IP addresses. However, the computer which has the IP address may be unavailable and the packet is then sent to another computer without the sender being aware of the change. The rerouting mechanism also makes it appear as if the packets that are returned to the sender were sent from the original receiver [103]. Address relocation may be regarded as a special case of the DNS-based fail-over method. A backup server is maintained and, when it has been detected that the main computer has failed, the traffic is routed to the backup server.

In order to apply address relocation inside a cloud environment, a replica instance must be created from the uninvolved instance. Once they replica instances have been created the cloud's internal network DNS server, or another method, must be used to redirect all traffic to the new instance. If it is not possible to change the cloud's DNS server then an extra instance will have to be used. This instance will serve as a middle layer DNS server and will redirect traffic to the replicated instances. The instance may also be used to interact with multiple instances but it is controlled by the administrator of the system and not by the system itself. The top level DNS can be configured when an investigation is in progress in order to redirect all the traffic to a replica which has been created. The primary instance may then be removed.

The switch-over time from a primary instance to a replica instance may be insignificant if the replication has been implemented correctly. However, if the replication is incorrect, then the address relocation will be inefficient. The replication ensures that the CIA is intact. This method may contribute to the complexity of the investigation as a result of the introduction of unnecessary DNS servers. In addition, the service providers may argue that this technique is wasteful with regard to cloud resources.

5.2.5 Sandboxing

In program security a sandbox is a controlled environment in which a program may execute [106, 107, 108]. It is not possible for a program to escape the sandbox and it cannot affect other programs outside of the sandbox. A sandbox is used to stop malicious programs from harming other programs on the same computer by limiting the interactions between the programs. A sandbox is created by software which controls the interaction of the program with other programs.

In terms of a cloud, an instance may be isolated by placing it in a sandbox. The sandbox will prevent the instance from interacting with other instances and, thus, the other instances will be protected from possible harm. In order to enable this functionality two approaches may be followed. Firstly, the cloud operating system may launch a sandbox application or the investigator may launch an application on the instance. This application will monitor all communication channels. It creates a virtual box around the instance. The instance is able to operate freely inside the box but it will not be able to operate outside of the box. This application will run on the network of the instance. Networking enables an instance to communicate with the rest of the cloud. The sandbox application will monitor and block network traffic as needed.

The sandbox technique assists the service provider to protect the CIA of the other instances. The other instances are protected while the investigation is being conducted while the instance that is being investigated is boxed in and will continue to operate as normal.

Information may, however, be lost while the instance is sandboxed. Software in the instance may detect that the instance has been placed in a sandbox and try to tamper with possible evidence. In addition, it may be difficult to block the network traffic in such a manner that may be proved to be accepted in the field of DF.

Accepted DF techniques may help the service provider as regards the CIA of other instances but loss of evidence may occur. The instance may be sandboxed while the other instances are moved from the node. Once the other instances have been removed an investigation may be carried out. The investigation may be conducted in either a live or a dead manner. Once all other instances have been moved from the node, a decision may be made as to which method is preferred. The sandbox may increase the possibility conducting a live forensic investigation as the instance is kept in a controlled environment.

5.2.6 Man in the Middle

MITM is another possible technique for isolation is. The term MITM may be used in network security to describe a Man in the Middle Attack (MITMA) [106]. A MITMA is a combination of potential threats to computer security. These threats include interception, interruption, modification and fabrication. Interception involves another entity gaining access to an asset. The interception is usually unknown to both the sender and the receiver. The asset is delivered to the receiver and a copy is sent to the entity. Interruption occurs where an asset is lost. The asset may be blocked, deleted or destroyed in any other way. Modification happens when the asset is altered in some way. The receiver acquires a changed version of the asset. Fabrication is defined as the creation of a new asset. The sender sends the original asset and the receiver receives the asset that has been created by the entity. A MITMA occurs when the entity places itself between the sender and the receiver. The entity receives all the assets from the sender and sends assets to the receiver. The assets are vulnerable to interception, interruption, modification and fabrication.

In order to permit a MITM to be used in clouds to assist in an investigation an entity will be created that exists between the cloud instance and the hardware of the cloud. This entity may either be part of, or use the virtualisation software of the cloud. The data travelling from the instance to the hardware and from the hardware to the instance may be analysed. The hardware includes, but is not limited to, the network, CPU, RAM and hard drive. This, in turn, enables a forensic process to be conducted on all the data being used in an instance. Such a forensic process will be a live forensic investigation.

The entity may be kept inactive when there are no suspicious activities on an instance. This minimises the computation cost of digital forensic readiness. However, when there is suspicious activity on an instance the MITM entity may be activated. Once activated

the MITM entity will analyse all the actions of the instance as well as the data travelling to and from the instance. In addition, it will stop the instance from deleting data on the storage media and RAM. The MITM entity will allow an investigator to access the resources of the instance without the instance being aware of the analysis. The investigator will also be able to observe the actions which the instance is performing, or is trying to perform. In order to enable the MITM to exist between the instance and hardware the MITM must be added by the creators of the cloud software or by a user company. In addition, to ensure admissibility of the evidence the MITM must be implemented using proven methods.

An advantage of this method is that the instance does not know it is being analysed. This, in turn, may prevent the instance from destroying evidence and also from taking part in any suspicious activity. Other advantages include that the fact that the instance is able to function as expected while other instances will not be affected by the investigation. The technique also assists in the protection of other instances. The instances that are being investigated may logically be blocked from communicating with other instances.

A potential problem is the implementation of the technique. There is a reliance on the cloud operating system manufactures and the cloud operating system manufacturers may not be willing to add this functionality. In order to enable a company to add the functionality the software must be reverse engineered. Once the software has been reverse engineered the MITM must be added. However, there are problems as regards both of these approaches. The cloud operating system creators may make the functionality available to its own employees only or they may create a substandard functionality. The admissibility of evidence may be compromised as a result of poor implementation. In addition, reverse engineering is a problem in itself. Most software packages' term of use permits reverse engineering of the software. However, this implies that the company using a MITM which they added may be sued. There is also the problem that it may be a challenge to prove the correct implementation of the MITM.

These techniques have the potential to assist in an investigation in clouds if the cloud manufacturers agree to implement a reliable and proven MITM functionality in their software. In addition the MITM may also be used together with other techniques. These other techniques may clear the node of instances and a controlled live forensic process may then be followed on the instance.

5.2.7 Let's Hope for the Best

The usual procedure is followed when an investigation is being conducted using the LHFTB technique [26]. In other words, the node is turned off and taken to a controlled environment. Images of the hard drives of the node are then created and analysed. However, a potential problem is that a node may contain multiple instances while the hard drives of the node may contain multiple virtual hard drives. The investigator must also know how the cloud operating systems stores information. Information from other instances may not be used and this would this violate the CIA of the other users. It may also be difficult to piece together the original virtual hard drive and credible evidence may be lost.

A possible advantage of LHFTB is that a suspicious instance receives no warning when an investigation is being conducted and this, in turn, decreases the possibility of possible evidence being contaminated.

As stated above there is the potential problem that a single node may contain multiple instances and these instances may be lost. This would be in contravention of the agreement between the service provider and the client while the clients who are not even involved lose the availability of their instances. Another problem is that running information may be lost. The information in RAM and the network is lost and cannot be used.

It is suggested that this technique should not be used on its own and that it rather be used in combination with other techniques. Firstly a MITM should be started on the instance that needs to be investigated. The RAM and other information may be acquired from the MITM. Other instances should then be moved from the node. The MITM will also assist in the instance moving process while protecting the instances that are being moved and keeping the instance which is being investigated in a controlled environment. The power must then be removed and images made. This would ensure an investigation that is both controlled and monitored.

The next section will introduce the conditions for isolation which may be used to prove that an instance has been isolated successfully by using either the methods explained above or other methods.

5.3 Conditions for Isolation

One of the main goals of isolation is to protect the evidence and, thus, it is vital that isolation be carried out in a systematic and controlled manner in order to assist in protecting of the evidence. This, in turn, entails following a DFP. One of the methods that may assist in ensuring the admissibility of evidence is by confirming certain conditions are met when conducting a DFP.

After careful consideration and experimentation the following list of conditions for isolation was composed. These conditions where selected as they best protect the evidence and form part of current DFP processes. These conditions will also protect the CIA of other resources on the cloud. It is contended that these conditions be met if an instance is to be successfully isolated. The list is as follows:

- The instance‘s physical location is known (location).
- The instance is protected from outside interference (incoming blocking).
- The instance is blocked from communicating with the outside world (outgoing blocking).
- Possible evidence from the instance may be gathered (collection).
- The possible evidence is not contaminated by the isolation process (non-contamination).
- Information unrelated to the incident is not part of the isolation (separation).

The term for each condition is provided in brackets. The rest of this section will describe the conditions in greater detail.

In order to know the physical location of an instance, it must be known where in the cloud the instance is. This, in turn, means finding the node on which the instance resides.

The instance must be protected from other instances and also from other external sources. In the cloud, all interaction with an instance is through the network and, thus, this means that the network connection must be blocked. This will assist in safeguarding against contamination of evidence and also any tampering with the evidence.

The instance may suspect irregular activity and it may then either send messages or tamper with the evidence. It is, thus, essential that the instance be blocked from sending messages over the network. In addition, the instance must also be blocked from tampering with the files and other information on it.

It must be possible to collect all potential evidence from the instance. This potential evidence may include running programs, information in the swap space and information on the hard drive. It must also still be possible to collect all potential evidence after the instance has been isolated on the node.

If the evidence on the instance is contaminated then the isolation process may be deemed to have failed as the aim of the isolation process is to protect the evidence. Thus, if isolation does not protect the evidence there is no reason for it.

The data from multiple instances can be contained on a node. However, the isolation must be conducted in such a manner that unrelated information is excluded from the isolation but related information is isolated. This, in turn, protects the confidentiality of the information.

The previous section introduced and explained the concept of isolation. Clouds differ in terms of both deployment and service models. These differences have an impact on isolation. This section will focus on explaining isolation in terms if deployment and the various service models. Some problems regarding each model will also be introduced. This study focuses on private and public deployment methods. It can be contended that both hybrid and community clouds may fit into either public and private clouds for the purposes of the discussions.

As regards the *deployment methods* of clouds, the differences between them pertain to where the data is located, and who owns the data. The location and owner of the data are important considerations in the isolation process.

In terms of conducting an investigation on a *private* cloud all the information on the cloud belongs to a company. If that company is conducting the investigation, then the issue of confidentiality is less important than when an external company is conducting the investigation or the external company is responsible for the investigation. The company whose cloud is being investigated would be concerned about ensuring the confidentiality of the cloud and it is essential that only information related to the incident be found. Thus, all other information should be kept confidential. In a case in which the company itself is conducting the investigation, then the first five conditions of isolation must be

met, although the last condition of separation is not of much importance. However, when an external company is conducting the investigation then all six of the conditions are important.

On a *public* cloud, the data on the cloud belongs to different users and these users may even be in different countries. This may create a problem of jurisdictions for the cloud as not all the information may be under the same legal system. On a public cloud the users pay the service provider to ensure that their information is kept safe and available. Thus, service providers are responsible for protecting the confidentiality and availability of their clients' resources. The party responsible for the onset of the investigation would normally be an external source. This external source is, thus, interested in finding evidence and this may not correspond with the concerns of the service provider. In a public cloud, isolation is required to protect the other users of the cloud and this means that all six of the isolation conditions must be met.

The various *service models* of clouds separate the data into different layers inside the cloud. Each layer entails separate conditions for isolation.

In an *IaaS* cloud, each instance on the cloud may belong to a different user while each instance may be seen as a virtual machine. The isolation of an instance on an IaaS service model cloud involves isolating the entire instance. This is done because each instance may be unique. Another reason is that each instance is an unknown computer. The instance may either destroy evidence on itself or it may contaminate other instances on the cloud. The aim of the isolation is, thus, to protect evidence on the instance and also to protect other instances from contamination.

In a *PaaS*, cloud the underlying platform of each instance is known. The differences between instances are the software installed on the instance and the data stored on the instance. In order to isolate the instance, both the applications running on the instance and the data of the instance should be isolated. In view of the fact that the underlying platform is known, most of the capabilities of the instance will be known and/or may be controlled.

The only differences between *SaaS* instances are the setup of the application and the data stored on the instance. Possible evidence may be found in the settings and the data. This means that the data and settings are the only parts of the instance that needs to be isolated. All of the capabilities of the instance are known.

The differences in cloud deployment and service models in terms of isolation have

been explained above. However, clouds may also differ in terms of the implementation of the service models.

5.4 Testing Conditions for Isolation

5.4.1 Locating an Instance

The first condition for isolation involves locating the instance. In order to locate an instance, the node on which the instance is running needs to be found. One method of doing this is using the cloud management software as the software may provide the functionality to locate an instance. Another method involves subnetting in terms of which a sub network is created for each node. The subnetwork is then used to trace the instance to a node.

The section will use subnetting. For the purposes of this experiment Nimbula Director will be used to run the experimental tests [19]. Nimbula Director is a cloud operating system which provides IaaS in a private network. In the experiment an instance was launched on the cloud. The node on which the instance resided then had to be found.

The IP of the instance, how the subnetworking is done and the wire address was all known. The IP was provided by Nimbula when the instance was started. For the purpose of this experiment the IP was 10.128.0.10. A subnetwork (Subnet) was created by Nimbula for each instance group on the cloud. The wire address was the network connections address and it had the lowest IP in the network [2].

In order to find the instances using the known information, the link between the networks of the instance and the nodes had to be found. All the network information of all the nodes and the network information of the instance were taken into account. As a starting point, the `ifconfig` command was run on each node.

Figure 5.1 depicts a part of the `ifconfig` output of a node in the cloud. The figure also shows the IP address of one of the network interfaces on the node.

In Figure 5.2, the `ifconfig` output of an instance running on the cloud is presented. The figure shows the network interface IP address and other network attributes.

As illustrated in Figure 5.2, the subnet mask for the instance is 255.255.255.252, thus `11111111.11111111.11111111.11111100`

[2]. The netmask uses the first 30 bits as the network ID. The remaining two bits may be used as the possible address. This is the cause of the possible availability of three IP addresses on the instance's private network.

In Figure 5.2 the IP address of the instance is 10.128.0.10,
and the broadcast address of the instances network is 10.128.0.11. It is possible to infer the network address when all the address bits are set to zero and then adding one to the

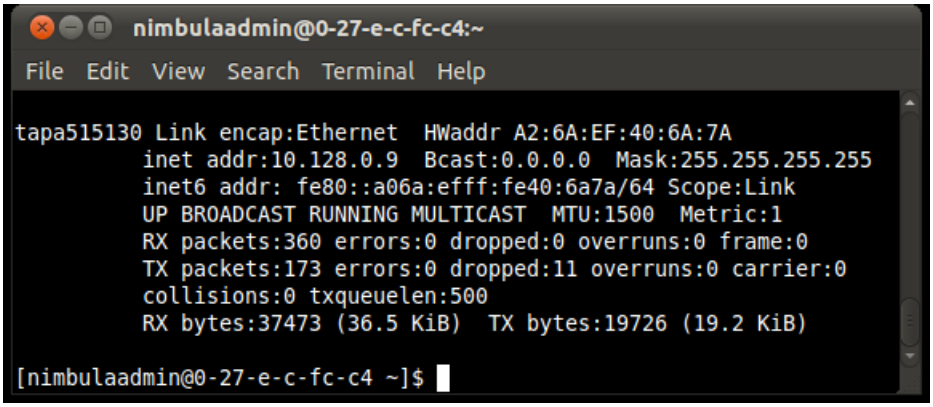

Figure 5.1: The ifconfig Command on a Node

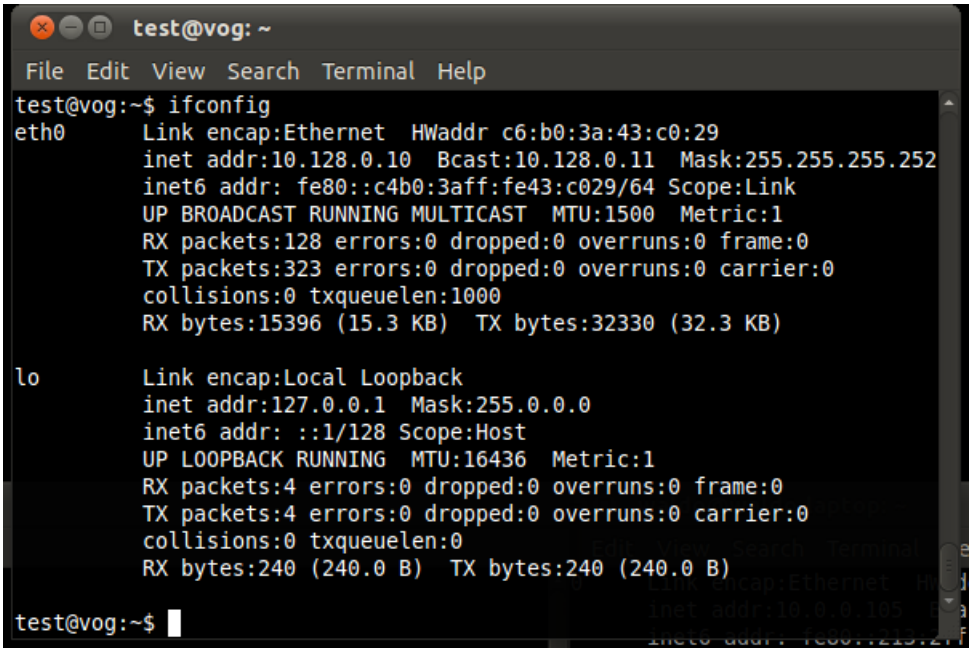

Figure 5.2: The ifconfig Command on the Instance

address [2]. The network address was inferred as 10.128.0.9. This meant that the wire address was the first address in the subnet.

In Figure 5.1, the IP address of one of the devices of the node is 10.128.0.9. This corresponds with the wire address of the instance. It was, thus, logically inferred that the instance was located on the node with a device which had the same address as a wire address of an instance. Thus, in order to find an instance, it is necessary to find a node with a network device that has an IP one lower than the IP of the instance.

In the experiment, the location of the instance was discovered by accessing nodes and an instance. The experimental outcome indicated that it is possible to find an instance without accessing the instance. This, in turn, enables the search to avoid detection by the instance. This would help to protect the possible evidence from possible contamination.

5.4.2 Blocking Communication

The first condition involved locating the instance. The location has thus, been discovered. Conditions two and three involve blocking ingoing and outgoing communication. In the cloud most communication is via the network [66]. In order to protect the instances from outside interference and to block communication with the outside world it is possible to stop the network. Multiple methods exist to block network communication. One option involves bringing down the network while another option is to use software to block the network traffic, for example a firewall.

In-order to bring down a network on a Linux computer the command `ifconfig eth0 down` may be used. The command `ifconfig tap8d6cb50 down` was run on a node with a network interface to an instance. This stopped the device from either sending or receiving data. Any attempt to establish a ssh session into the instance would fail, because no network traffic to the device was allowed. This method also prevents the instance from communicating with the outside world. However, this may not be ideal as all communication is stopped and the cloud operating system and the investigators would not have access to the instance.

One of the other options is to use software. For example, a firewall may be used. The firewall can be implemented in one of two ways. The first would involve a firewall on the node itself while the other would run on the cloud itself. If communication is blocked using the firewall on the node the communication on a specific network device

must be stopped with the network device being the virtual connection to the instance as described in section 5.4.1. The other option involves a single large firewall to control the network on the cloud. The instance IP is given to the firewall in order to block all communication to and from the device. Using a central firewall may make it easier both to maintain and to setup the blocking of communication. An added advantage of using a firewall is that the firewall may be implemented in such a manner that it would enable the investigation to gather evidence from the traffic that has been blocked, as it could be logged.

There is also other software that may be used to block communication, for example, the cloud operating system. However, this was not implemented in the cloud operating system which was used for the current experiments. The administrator may indicate that a specific instance is blocked and the system is then responsible for the traffic blocking. The system may also provide certain information. This information may include the other instances with which the suspect instance is communicating and the kind of communication used.

The methods used to block communication generally block all network traffic and may include using a piece of software to block specific communication. This, in turn, would assist in a "live" investigation. All web traffic may be blocked as it is not required in an investigation. The communication left open may include ssh connections. However, the problem is that there are unmonitored connections left open and, thus, these connections need to be monitored so as prevent unauthorised use. In order to do this a specialised firewall may be used or some specific software to allow certain connections from the DFI team but also block all other communication.

5.4.3 Gathering Possible Evidence

According to conditions four and five it must still be possible to gather evidence and the evidence must not be contaminated. When carrying out isolation, it is essential that the instances be protected against contamination. When finding the instances, there is no direct communication or interaction with the instances. The communication blocking is done in such a way that it will minimise interference with the instance. The precautions taken in the previous steps would assist in protecting the instances from contamination from either external sources or the instance itself.

One of the outcomes of an investigation is the evidence which would have been

collected. If the instances were protected from contamination, the possible evidence on the instances would also be protected from contamination. However, the process involved in gathering possible evidence from the isolated instances is outside of the scope of this study although other studies have been conducted which focus on this issue [94].

5.4.4 Separation

In order to create a clean crime scene it is imperative that the node contain the suspect instance only. This may be done by moving the instance to a clean node or by moving the other instances from the node. Moving the other instances has the advantage that the suspect instance is unaware of the moving process. Possible techniques that may be used to move instances were discussed in section 5.2. In some cases the cloud operating system itself may be used to move the instances.

5.5 Distributed Instance System (DiS)

The previous section focused on isolating a single instance. This section will focus on introducing an environment in which there are multiple suspect instances. An implementation involves is using an IaaS cloud to build a type of server farm.

In cloud computing availability is regarded as a necessity [66]. One model of cloud computing which provides availability is a Distributed Instance System (DiS). In a DiS, multiple instances on the cloud are combined to form one logical resource. This combination aids in protecting the availability of the resource. When one of the instances malfunctions, it is discarded and a new instance may be started to take its place. This clearly implies that the instances are dispensable.

In the cloud, the DiS may be located over several nodes. This also enables the service provider and the client to provide on-going availability. If a node fails, this would have little or no effect on the overall DiS. The focus in section 5.6, is on isolation in a DiS environment.

5.6 Isolation of Cooperating Suspect Instances

In a DiS infrastructure it may also be the case that there are multiple suspect instances and, thus, the possible evidence can then be distributed over multiple instances. The problem may be explained in terms of a basic investigation problem [13]. The problem is the question of when to stop gathering evidence. The instances used in the investigation would form a set of instances. In a DiS with multiple suspect instances there are three options for the size of the set used. One option is to gather evidence from all the instances while the second option is to gather evidence from one instance only. The third case is a middle ground between the first two options.

When all the instances are used in the set all the instances have to be isolated. However, in order to isolate all the instances all the instances need to be found. If one instance is missed then data set is incomplete. A possible problem with finding all the instances is the time required to find them all, particularly as the time that is used to find all the instances may be used by the suspect instances to contaminate evidence. When finding the instances a subset of the instances may discover proof that an investigation is being carried out and this may cause the instance to respond harmfully towards any possible evidence. Another problem is the order in which the instances are isolated. The instances need to be isolated in a manner that would not raise suspicion amongst the other suspect instances.

The second option involves finding one instance on the cloud only. Thus, there is no investigation for other suspect instance and also no search is conducted for them. In other words, the other suspect instances are left untouched by the investigation. However, there are a few problems with this method. One of the problems is the amount of evidence. Would sufficient evidence be found on a single instance? If the DiS is set up in such a way that instances mirror each other, it may be assumed that enough evidence would be gathered. On the other hand, if the DiS is set up in a distributed manner and each instance contains a small subset of the evidence only, then the total amount of evidence that may be gathered may be less than the small subset of evidence available. This, in turn, may cause evidence loss and an insufficient amount of evidence may be gathered. In a DiS, the loss of an instance is expected. When an instance is isolated, the other suspect instances may ignore the loss and continue as normal.

The last option involves finding some middle ground between the two previous options, where a subset of the instances only is found. However, the problem is the number of instances that may be found and the speed at which they may be found. If too few

of the instances are found, then limited evidence may be gathered. However, if too many instances are sought then resources may be wasted. If the instances are isolated one-by-one, this may be detected by other suspect instances and they may react by contaminating the evidence. There is no answer to the number of instances that need to be isolated or how many resources should be dedicated to the process. This varies per investigation and also according to the investigation time frame itself.

When conducting an investigation, it must be decided whether a "*live*" or "*dead*" forensic investigation will be conducted [26]. On a cloud four options are available. These include the "*live*" and the "*dead*" investigations, as mentioned, as well as a new option, namely, the "*halfdead*". In terms of the latter the instance is dead but the node is alive. The node is used to gather evidence and is trusted. The fourth option is where the node is dead but the instance is alive, "*resurrected*". In other words where the node has been killed but the instances were restated in a new controlled environment.

When conducting an investigation on a DiS cloud a combination of "*live*" and "*dead*" forensics may be used, namely, *Community Live*. This is done by isolating some instances and monitoring others. The instances that are isolated are subsets of the suspect instances. These isolated instances are used in a "*dead*" forensic manner. In order to gather "*live*" forensic evidence the network traffic of some of the other instances is monitored. The RAM content and running processes may also be used. Thus, there are five types of forensic investigations that may be conducted on a DiS cloud.

There are, however, problems with the *Community Live* method. Some of the problems have been mentioned previously but are also applicable in this context. If too few instances are isolated there may not be sufficient "*dead*" forensic evidence. On the other hand if too many instances are isolated, it might be noticed by other suspect instances and they may contaminate evidence. This means the "*live*" forensics may lead to inadmissible evidence. The combination of both "*live*" and "*dead*" forensic evidence may prove a vital resource as regards a DFI in a DiS cloud.

Table 5.1 presents an analysis of the network traffic that was captured in an experiment. In this experiment five instances were setup to work together. They send traffic to a controlling instance and the controlling instance responds with an answer. The captured data was retrieved from the node on which the first suspect instance was located. The suspect instance has the IP 10.128.0.26. The analysis shows that the network traffic may be used to detect the instances that are working together. In Table 5.1, the second row shows that 10.128.0.22 sent 44051 to 10.128.0.26 and that
10.128.0.26 sent 36955 bytes back to 10.128.0.22. The other rows may be read in the

Address A	Address B	Packets	Bytes	Bytes A->B	Bytes A<-B	Duration
10.128.0.22	10.128.0.26	1046	81006	44051	36955	303.5644
10.128.0.26	10.128.0.34	1100	85200	38900	46300	294.145
10.128.0.26	10.128.0.30	1102	85332	38900	46432	292.0971
10.128.0.18	10.128.0.26	1122	86904	47226	39678	298.7616

Table 5.1: Analysed Network Traffic.

same manner. After the analysis it was possible to conclude that the instances with IP's 10.128.0.{18,22,30 and 34} were working together.

5.7 Conclusion

Possible techniques which may be used to isolate an instance on a cloud were introduced and discussed in this chapter. These techniques include Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, MITM and LHFTB.

It emerged from the discussion that no single technique provided a perfect solution. However, the techniques may be combined to provide a feasible method with which to isolate a cloud instance. The differences between some of the techniques are small and may actually be regarded as non-existent. The differences between the techniques allow them to be used in different environments.

Six conditions were introduced. It was argued that these conditions need to be met in order to ensure to the successful isolation of an instance.

The conditions were then explained and the methods involved in ensuring that each condition is met were discussed. Some experimental data was also given about the condition outcomes.

The DiS environment was considered. This environment involves the use of multiple instances to form one logical resource in order to assist in the safeguarding of resources availability.

The problems that may arise in a DiS environment during the process of isolating were explained. Solutions were suggested for the number of instances that need to be isolated in a DiS cloud with multiple suspect instances. The "*live*" and "*dead*" forensics

model was expanded to include three other methods, namely "*halfdead*", "*resurrected*" and "*Community Live*".

Out of this chapter is was clear that methods are needed to isolate a sub part of the cloud, this is why the next chapter will focus on cloud separation.

Chapter 6

Cloud Isolation

6.1 Introduction

The previous chapter discussed isolating single instances. The conditions that need to be met to ensure the successful isolation of an instance were listed and arguments offered as to why these conditions need to be met if instances are to be successfully isolated. The conditions may also be applied to the isolation of a subpart of the cloud.

Gathering evidence is one of the aims of an investigation. If there is any suspicion that the evidence may be invalid for any reason then the evidence will not be accepted as admissible evidence. In order to ensure the admissibility of the evidence it is essential that the evidence be protected against contamination and tampering. The need for isolation in the cloud environment becomes apparent when the issue of admissibility of evidence is taken into account.

In order to isolate a cloud a subpart of the cloud must be isolated. This is done in order to keep the isolated part of the cloud in a cloud environment. This chapter will focus not on isolating a single instance or a small subset of instances, but rather on isolating a part of the cloud. This sub-cloud will have the normal functionality of a cloud. The aim of separating the subpart of the cloud from the cloud is to link together cooperating instances and exclude unrelated instances. The separation also enhances the admissibility of the evidence. Once the cloud has been separated the investigation focuses on the isolated part of the cloud without any disruption of service to the other clients of the cloud provider.

6.2 Cloud Separation

It may be argued that cloud separation may be regarded as a part of an investigation on clouds since, as stated above the isolation process may enhance the admissibility of the evidence. The cloud separation forms part of the collection phase of an investigation. The aim of the separation process is to prepare the cloud for an investigation. It may be argued that the conditions for isolation, as stated in section 5.3, must be met if the separation is to be regarded as successful. After careful consideration and while devising the conditions as discussed above the notion of cloud separation was discovered. Cloud separation may be regarded as comprising of moving the instances and dividing the cloud.

Moving the instance involves relocating the instance from one node to another. This process of moving the instances should involve moving all the instances involved to a certain part of the cloud and all non-related instances to another part of the cloud. This movement is the start of the isolation process that was explained in section 5.2. The movement may involve one of several methods. Firstly, the instances may be moved directly from one cloud to another. Secondly, the instances may be to an external cloud and from there to the other cloud. The third option is to move the instance to an external cloud, then move it to one or more other external clouds and, finally, to move it to the other cloud. The fourth option involves using to use the cloud operating system to move the instances while the final option involves identifying the nodes which contain suspect instances but not moving them.

The aim behind the division of the cloud is to complete the isolation. There are several ways in which this division may be done. The first option involves separating the nodes by creating two separate networks from one network the second option involves creating two virtual networks on one logical network while the third option involves creating subclouds inside the actual cloud. The last option involves using the movement methods to move the instances to a cloud which has been dedicated to the purposes of the investigation. Together the movement and the division methods constitute cloud separation. This means that different cloud movement and division methods may be used together in different combinations depending on the specific requirements. The remainder of this section will elaborate on the movement and division methods.

There are two different methods which may be applied when moving instances from one cloud to another. The first of these methods involves mirroring an external cloud and then sending instances from the main cloud to the external cloud. Thus, the external

cloud is, in fact, not an external cloud but a controlled cloud that was set up to accept instances. Figure 6.1 depicts this process. Some of the cloud operating systems have the functionality to send instances from their cloud to another cloud and this functionality would be used to transfer the instances. However, the way in which the instances are sent must be known and also what is required while sending the instances. This would make it possible to mirror one of these external clouds and to receive instances from the main cloud. The advantage of this method is that instances are able to observe this movement as a normal cloud operation activity. In term of the second method the cloud operating system allows the sending and receiving of instances and, thus, the cloud operating system is used to aid in the movement of the instances. VMware is an example of a cloud operating system that is able to send and receive instances [18].

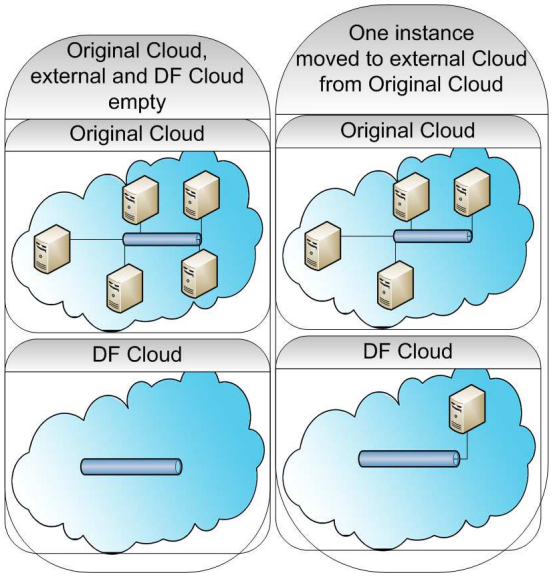

Figure 6.1: Moving an instance from one cloud to another

When using an external cloud during the process of moving instances, the same methods as suggested above, may be used. The instances are sent to an external cloud. This external cloud is able to accept instances from the main cloud and it can be assessed by the investigation team. Once the instance is on the external cloud it is then sent to the controlled cloud. This external cloud may either be hosted by other companies or may be another cloud which is owned by a company. Figure 6.2 depicts the steps involved in this process. Automated methods may be used to move the instances. However, if there are no automated methods available, one of the methods which were proposed to move may instance can be used.

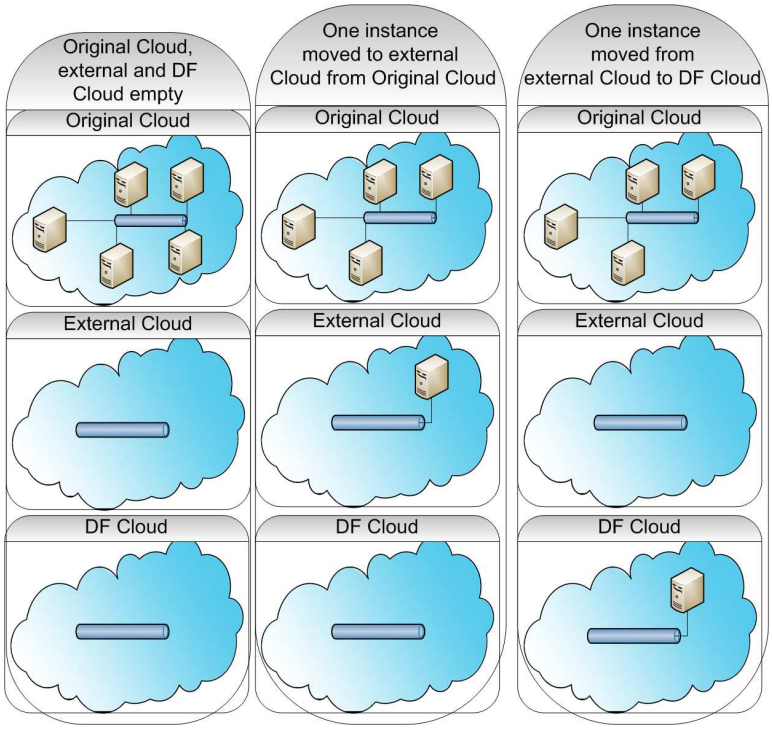

Figure 6.2: Moving an instance from one cloud to another using an external cloud

The option in terms of which multiple external clouds are used is the same as mentioned above although multiple external clouds exist between the two clouds. This method may be used when no middle ground exists between the main cloud and the other cloud. The external clouds are used to link the two clouds.

The cloud operating system may also be used to move instances. Some cloud operating systems provide the functionality required to migrate instances while they are running between nodes. The last option in terms of which the nodes are identified only would be used if there were no methods available in terms of which to move the instances, or if the only instances on the node were suspicious instances.

The first option when dividing the cloud involves using the self-healing characteristic of clouds that would then be used to create two clouds. If a node or nodes malfunction in Nimbula, then the cloud itself will continue to operate. In terms of this option the first step is to identify the nodes that must form part of the new cloud, while the second step involves moving all the non-related nodes from these clouds. The next step will be explained by means of the following example. If the cloud has six nodes and three

of these nodes must be moved to the new cloud, then the process is as follows: Two switches are connected to each another. The one switch has all the nodes connected to it. One by one the network wires/VLANs of the suspect nodes are moved systematically to the other switch. Once all the suspect nodes are connected to the other switch the connection between the two switches is broken. The cloud operating system will then create a new cloud using its self-healing ability. The process is illustrated in figure 6.3.

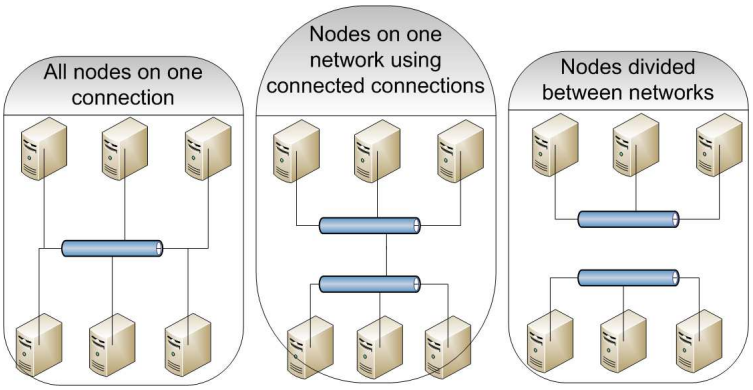

Figure 6.3: Creating Two Clouds from a Single Cloud

The second option for cloud division involves creating two clouds on one network. A high-level overview of this process is presented in figure 6.4. This option may be separated based either on knowing which node belongs to which cloud or not knowing which node belongs to which cloud. Each cloud runs its own cloud operating system that controls it. There are various methods which may be used to create two clouds on the same network. One option is to create separate subnet masks for each instance of the cloud [2]. This would enable the installation and operation of each cloud on a separate subnet mask. A possible alternative option is to use a cloud operating system that enables the selection of the controlling node and, by using this strategy, a new master is set up on the network and some nodes are assigned to it.

The third option involves dividing the cloud so as to create sub-clouds. The cloud is logically divided into separate parts although the same cloud operating system controls these separate parts. The sub-cloud is a fully functional cloud which merely runs on the main cloud although it is interacted with as if it were a normal cloud. Some cloud operating systems allow for the creation of sub-clouds inside the cloud itself. This functionality is used to sell a cloud to the service provider's customers. To enable the creation of sub-clouds inside the cloud itself sub-cloud is created on the main cloud and

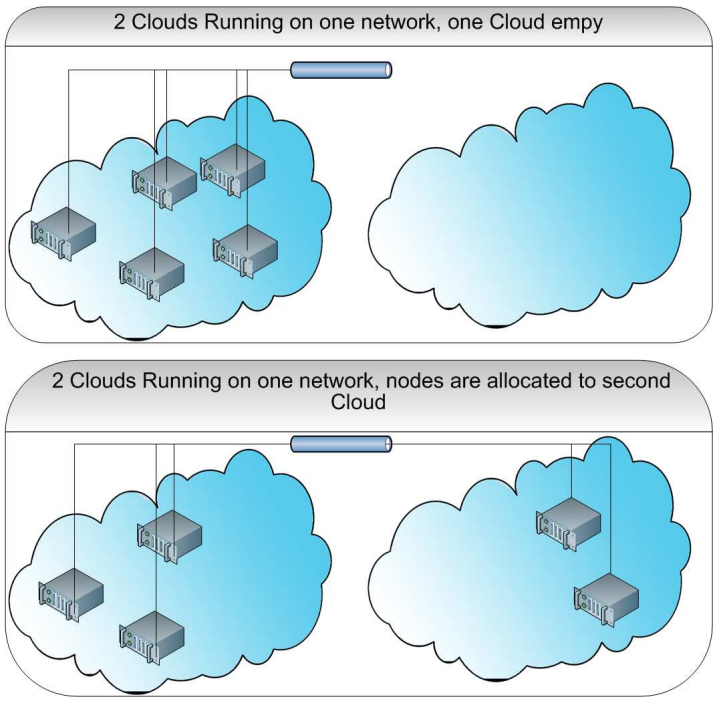

Figure 6.4: Creating two Clouds on One Network

then instances are moved to the sub-cloud. This sub-cloud is implemented on the same hardware as the base cloud. The moving functionality is provided by the cloud operating system. Figure 6.5 depicts the main clouds hardware and the virtual which are clouds created on that hardware.

The last option to divide the cloud involves using any of the instance movement methods to move instances to an already divided cloud. This cloud may be a cloud which has been prepared to conduct cloud forensics. The cloud may also be located either at the premises of the cloud providers or on an external cloud which is at the premises of the investigation team.

While completing the separation it is essential that all the steps be documented as part of the investigation as this creates an audit trail which may be used to prove the viability of the methods which were used.

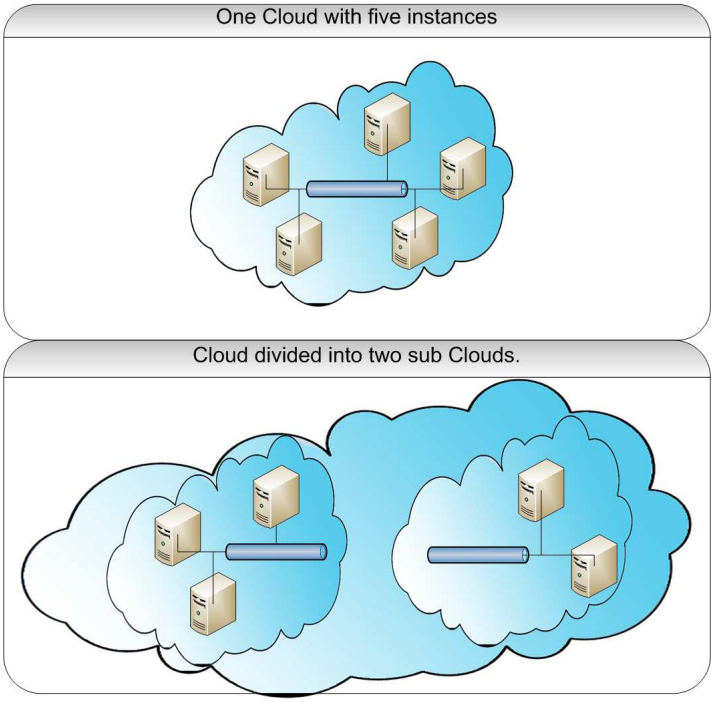

Figure 6.5: Creating Two Sub-Clouds

6.3 Cloud Separation on Different Types of Clouds

The previous section discussed methods for cloud separation. As stated in chapter 3 clouds may be divided into different service and deployment models, all of which have different impacts on isolation. However, certain issues considerations need to be taken into account when conducting cloud separation for the different models. One such consideration is the Confidentiality, Integrity and Availability (CIA) of instances.

One difference between the service models is who owns what part of the instance. The instance may usually be divided into the hardware, the hypervisor, the operating system, applications and data. In an IaaS model, the service provider is responsible for the hardware and hypervisor whereas the client is responsible for the operating system, applications and data. If the service provider requests the investigation then the cooperation of the client is also required in order to collect evidence from the operating system, applications and the other data residing on the system. However, if the client requests an investigation then the cooperation of the service provider is needed in order to collect evidence from the hardware. The client and service provider are both responsible for the

availability of the system. It is possible for clients to have multiple instances working together without the knowledge of the service provider.

In a PaaS model the service provider is responsible for the hardware, hypervisor, operating system and some of the applications while the client is responsible for the other applications and the data on the system. The service provider must ensure that high availability is maintained. The client may provide evidence from their own applications and stored data. It is possible, but more unlikely than in IaaS, that he client would have cooperating instances.

In a SaaS model the hardware, hypervisor, operating system and all the applications are provided by the service provider while configurations of the applications and data on the system are the responsibility of the client. The service provider must ensure the high availability of the systems. Clients are responsible for their data only. It is very unlikely that a client would have cooperating instances.

When conducting cloud separation on a public deployment model the cloud service providers are responsible for protecting the CIA of their clients. When separating a part of the cloud it must be confirmed that only the data related to the suspicious instances is separated. In addition, the separation must also protect the admissibility of the evidence. However, all unrelated instances should not be affected by the separation and they must, thus, remain available. If the service provider is conducting the investigation then the service provider must protect the privacy of its clients and inform its clients of the investigation. On the other hand, if an external company is conducting the investigation the company must protect the privacy of both the service provider and its clients.

When conducting cloud separation on a private development model all the data should belong to one company. The separation is conducted in order to protect the admissibility of the evidence. If the owners of the cloud are responsible for conducting the investigation, then the main focus is not on protecting the privacy of the information. However, if an external company is responsible for conducting the investigation then the separation should also protect the privacy of the owner's data. The owner is responsible for making a decision on the importance of the availability of the cloud.

It may be argued that cloud separation is valid for both IaaS and PaaS models. Cloud separation may constitute an integral part of an investigation on a public cloud but it may also be important in a DFI on a private cloud.

6.4 Experimentation

This section presents the results of the experiment. The experiment was limited to the dividing methods. Moving an instance may be done by the cloud operating system, thus, making it part of normal cloud operation or, if there is no such functionality, by using one of the methods proposed in the previous chapter. For the purposes of the experiment the following two dividing methods were tested, namely, creating two clouds using the network hardware and creating sub-clouds. The experiment used VMware and Nimbula Director. This was done in order to be able to compare the feasibility of the methods. As stated in the introduction chapter, VMware and Nimbula provides a better coverage for both public and private cloud computing.

In order to create sub-clouds more than one layer of abstraction is needed. In the experiment VMware was used to create the sub-clouds. There were two base nodes running VMware and which are known as ESXi hosts. These nodes have Intel I5 processors and 4GB DDR3 memory. On each of the hosts two other ESXi hosts were created. The virtualised ESXi hosts were used to form a cloud on each main host. A vCenter management instance was created on each virtual cloud. VCenter is used to control the cloud [18]. Two instances running CentOS 6.0 minimal were also created on the virtual cloud. Thus, the layout was vCenter running on Windows 2008 server and two CentOS minimal instances running on two ESXi hosts. The ESXi hosts were running as virtual machines on a base ESXi host.

The setup and maintenance of this experiment were relatively easy. The clouds were stable and there were no obvious problems with the recursive virtualisation. In the testing environment there was some loss of performance. This occurred because some of the resources were used to run the other virtual hosts and, also, there were two controlling layers.

In order to test the performance loss a single sub-cloud was created on a Dell PowerEdge R710 with two I7 processors and 97GB of memory. The performance decrease was not significant on this node. The performance drawback should not be significant on most of the powerful infrastructure used by the majority of cloud providers. However, the software on the instances may notice the loss in performance on the node and they may institute self-defence mechanisms. While this may happen the performance on the cloud environment is, in fact, inherently unstable because of the over commitment of resources that is part of most cloud environments [18].

VMware also assist in creating sub-clouds. VMware allows the movement of instances from the main cloud to the virtual cloud and from the virtual cloud to the main cloud. It is incumbent on the user to link vCenter of the virtual cloud to the underlying infrastructure. This allows the user to move instances between the layers of virtualisation. However, the drawback is that there is a connection created from the virtual cloud to the underlying cloud and this may be used in order to tamper with evidence. Nevertheless, the advantage is that it is possible to create a virtual at a later stage and instances may be moved to this virtual cloud from the main cloud. Once a digital investigation is required the instances may be moved to sub-clouds, one for uninvolved instances and one for suspect instances. If instances are no longer suspect in the suspect cloud then it may be moved to the other cloud.

An experiment using Nimbula director was conducted in order to create sub-clouds. The cloud consisted of three nodes with two virtual nodes being created on each of the three nodes. It was possible to access each of the sub-clouds separately. However, the problem was that the virtual node should be on a virtual network and, if not on a virtual network, then they are not able to communicate with the other virtual nodes. It was possible to create instances on the virtual nodes. Nevertheless, because of a limitation in Nimbula instances could not be moved directly from the main cloud to the sub-cloud and it was necessary to use the movement methods proposed in the previous chapter to move the instance to the sub-cloud.

The next experiment created two clouds using network hardware. This experiment was conducted by using both Nimbula director and six nodes with 4GB of RAM and I5 processors. The experiment was conducted as described in section 5.3. Access was lost to the control centre of Nimbula on the one part although the instances continued to run. A possible problem may arise with the control centre holding information about all running instances. If the cloud is broken up the control centre loses communication with the other instances that are running on the other part of the cloud these instances will show as being in an error state. However, the instances may then be "deleted" from the control centre as they are not applicable to it. Nevertheless, the problem is ongoing because the clouds cannot be joined at a later stage. There were two control centres, each running with its one instance. In the experimental conditions it appeared impossible to join the cloud back together. Although connection was lost with the control centre the cloud continued to function, thus proving that the self-healing characteristics of Nimbula were intact.

The last experiment was conducted using VMware to create two clouds using the

	VMware	Nimbula	SAN	DAS
Cloud separation using sub-Clouds	✓	X	✓	✓
Cloud separation using Network Hardware	✓	X	X	✓

Table 6.1: Summary of applicability of separation methods.

network hardware. The same procedures were followed as for Nimbula. The experiment was successful although a few problems did arise and configuration changes were needed. These problems lie in vCentre assuming that host failure occurred and then trying to re-launch some of lost instances. This happened because high availability was enabled on the cluster an, the function of high availability (HA) is to recover lost instances. These instances were stored on storage shared between the nodes. However, other instances were located on direct attached storage, thus, HA is not available for them. It was also necessary to create a new vCenter on the other part of the cloud because there was no management over the new cluster.

It may be argued on the basis of the experimentation that the method in terms of which a sub-cloud is created using network hardware is not advisable considering re-setup is required to put the cloud together again. Accordingly, this method is not recommended. On the other hand, the other experiment revealed that using sub-clouds for cloud separation is a more reliable method. Table 6.1 presents a summary of the experiments.

The following lessons were learned from the experiment. Performance is affected on less powerful clouds, HA needs to be turned off before starting with cloud separation and it may be impossible to recombine the cloud after the investigation.

Another possible general problem that must be taken into account with all methods used for cloud separation is where the storage of the instance is located. As a basic example, the storage may either be on a SAN or the direct-attached storage (DAS). Creating sub-clouds when using a SAN is not recommended, as connection to the SAN may also be lost although creating a sub-cloud may still be possible when using a SAN with a dedicated network because the nodes can still communicate with it. Both methods tested are applicable when using DAS. Another problem with SANs is that multiple instances share the resource, although this problem may be averted by using a SAN dedicated for the storage of suspect instances.

Another problem is the IP address of the instance. When moving an instance the IP of that instance should be constant in order to correlate the IP with the network evidence

collected. In the experiments the instances had static IPs which does not change if the instances are moved. If a dedicated firewall is used to assign the IP then the IP should stay the same even if the instances move. When the IP of the instance is managed by the node on which it is residing then the IP may change. If the instances are moved in order to assist in the correlation of evidence then the IP before and after the move must be noted.

6.5 Isolation Accomplishments

Thus far the study has isolated wither single instances or a group of cooperating instances. This comprises the first part of this study in terms of which an incident inside the cloud was isolated, thus allowing the investigation team to conduct an investigation in a certain part of the cloud only and not in the whole cloud. The isolation agglomerates all the related data of the investigation, making it possible for the investigators to collect admissible evidence.

6.6 Conclusion

As cloud computing grows it will become easier for individuals to create DiS resources. If the DiS resource is used to commit a crime, then it is imperative that there are methods which may be used to conduct an investigation on the DiS without disrupting the other users of the cloud.

The chapter discussed the notion of cloud separation. Cloud separation consists of moving instances and dividing the cloud. The chapter also discussed possible methods to move instances around in the cloud as-well as move instances out of the cloud. The methods that may be used to divide the cloud were also explained.

The researcher conducted experiments on the division methods. It was discovered that the methods used depend on the circumstances of the investigation. In addition it was observed that the method that uses the network hardware to create two clouds may not be a desirable and that the method used to create sub-cloud may be a valid choice.

This chapter completed isolation in the cloud, now once isolated can we move the instances into a digital forensic laboratory. The next chapter will introduce the concept

of a digital forensic laboratory inside the cloud environment.

Chapter 7

Forensic Laboratory in the Cloud

7.1 Introduction

In the previous two chapters instances were isolated as part of the collection phase of an investigation. In a forensic process the isolated instances are used in order to gather forensic data. The investigation may use either a "live" or a "dead" forensic analysis, after which the data captured would be moved to a forensic laboratory. This laboratory may be external to the cloud provider. Once the forensic data of the instances has been captured and moved either to a forensic environment or the forensic laboratory then the collection and transportation phases are completed. The remainder of the forensic process may then be commenced.

It is essential that the processes of a digital forensic laboratory be planned if the laboratory is to be both effective and viable [109]. The type of laboratory and the target of the forensic investigation need to be determined before a laboratory is started. The type of forensic laboratory will determine the equipment required and the staff skills.

This chapter will introduce the concept of a "laboratory inside the cloud" in terms of which a digital forensic laboratory will be created and used inside the cloud environment. The isolated instances may then be kept inside the cloud environment, thus, minimising the movement of the isolated instances. In addition, the laboratory itself may also benefit from the advantages of cloud computing. The cloud laboratory may then be incorporated into an investigation which focuses on cloud forensics. This chapter will also discuss the

requirements of digital forensic laboratories, examine a digital forensic laboratory on the network and introduce a digital forensic laboratory on the cloud.

7.2 Digital Forensic Laboratory

A forensic laboratory is used in the examination and traces phase of Cohen's forensic model [13] and also in the analysis and examination phases of the NIJ model for a forensic investigation [12]. The laboratory should possess all the tools required to conduct a successful investigation. In addition, the laboratory would also need trained personnel in order to conduct the investigation. It is, thus, important that forensic laboratories have the required tools and personnel in place in order to obtain admissible evidence [110]. The tools needed in a forensic laboratory are discussed in the section on tools in chapter 2, although, the laboratory may also utilise other tools to assist in investigations.

The digital forensics laboratory must use procedures and policies to protect the evidence collected [110] and, thus, the laboratory will have prescribed which procedures describe the processes that must be followed in order to gather different forms of evidence. These procedures will be used in conjunction with the policies of the laboratory. The policies will describe the roles of the various staff members and also what should happen if a role switch is required during the investigation. There will also be documentation describing the conditions and procedures in the event of non-standardised method being required in order to conclude the investigation.

The staff in a forensic laboratory also has an important function with the staff members fulfilling different roles. Possible roles in a forensic laboratory include laboratory manager, reception officer, triage officer, imaging officer and analyst [109]. The laboratory manager is responsible for all the operations of the laboratory as well as the financial planning of the laboratory. The reception officer is the contact point of the laboratory. He or she will contact clients and the clients will contact them. The reception officer is also the contact point for external companies or parties that need either communication or cooperation with the laboratory. The triage officer is responsible for accepting or denying cases and for assigning priorities to the cases accepted. The imaging officer is responsible for creating the forensic images that will be used in a case and also for proving the validity of the images. The analyst conducts the investigation. However, various, analysts may be used on different cases as their skill sets would differ. Thus, the required skills would be assigned according to the requirements of each case.

It is essential that the digital forensic laboratory meet various high level requirements to ensure the admissibility of evidence [110]. These requirements include physical and logical security, equipment up keeping, and staff protection.

The first requirement is security. A digital forensic laboratory may contain possible incriminating evidence. This means that the content of a laboratory may be valuable and, thus, the need for physical security [109]. It is vital that the laboratory be secured against intrusion and, therefore, physical security measures such as alarm systems and burglar bars must be installed. In addition the laboratory must to be protected against fire, therefore, there must be some form of fire detection system and, if possible, a suppression system. The laboratory also needs to be secured on a logical level. The logical level includes securing the network, securing the operating systems and enabling access control to all equipment. The equipment, work stations and security systems should be time synchronised to ensure consistency of logs and equipment. One final aspect is a form of backup or redundant power to ensure the protection of the evidence while the evidence is being used and to ensure that the physical and logical security remains intact.

All equipment must be calibrated on a regular basis as this will ensure that the equipment is functioning properly and that, if required, there would be proof that the equipment is fully functional and accurate [110]. The equipment also needs to be updated to ensure that the laboratory is capable of providing admissible evidence.

The last concern is the staff members themselves [110]. One of the concerns of a laboratory is the health and safety of the staff. The physical evidence may be contaminated and, thus, the staff may be exposed to various diseases. As mentioned previously it is also vital that the staff members constantly update their skills.

There is another category of laboratory that forms part of the reconstruction phase. These laboratories have a different set of tools and requirements to those already discussed. The reconstruction laboratory is used to recreate the environment required to test the hypothesis that would have need formulated during previous phases [110]. The focus in the reconstruction laboratory is not on obtaining admissible evidence but rather on the ability to recreate the original environment of the investigation and to prove the hypothesis in a well-documented and repeatable manner. A reconstruction laboratory will, thus require the tools with which to recreate a wide variety of environments. These tools range from different computer components, different operating systems to different applications. Different hardware and software from different time periods may be required as well. The personnel of a reconstruction laboratory are also trained to cre-

ate different environments in a documented manner. The reconstruction laboratory will also have methods and procedures in place to enable the automated recording of the reconstruction process. One of the problems faced by reconstruction laboratories is the associated cost as it may become expensive to have all the tools required with which to recreate all the environments needed.

7.3 Requirements and History of the Forensic Laboratory

Standards in the laboratory not only protect the admissibility of the evidence, but also facilitate the understanding and communication of the process followed by the laboratory [110].

In the previous section it was stated that laboratory equipment needs to be calibrated and, thus, the laboratories which conduct the calibration and testing must also be certified. This certification enhances the acceptability of the equipment. This accreditation for forensic laboratories was introduced as early as 1947 [109]. The first major step in the field of forensic laboratory accreditation was in 1977 at the first international laboratory accreditation conference. [110]. At this conference the need for accreditation was apparent and, therefore the international organisation for standardisation (ISO) and the International Electro-Technical Commission (IEC) created a joint document ISO/IEC Guide 25. This document was the first to describe requirements for a laboratory. This document was created in 1987. In the 1990s the ISO introduced the ISO 9000 series. In 1999 the document was reviewed and converted to the ISO/IEC 17025. The review improved the proposed processes, added continuous improvement and included a component in respect of communication with the clients. A new version of the document is the ISO/IEC 27037, which focuses specifically on digital evidence. This document contains added focus on multi-jurisdiction systems and the digital forensic process also included specifications for the private sector.

The ISO/IEC 17025 indicates the basis for a digital forensic laboratory while the sections on digital forensics are applicable to this study [110]. The ISO 17025 highlights the issue of documentation which, as was stated in chapter 2, is a key factor in the forensics process. The ISO 17025 also highlights the need for staff training, the use of proven procedures and the calibration of equipment. The document does not, however,

provide specific implementation guides but rather gives an overview of and considerations in respect of forensic laboratories.

The ISO 17025 also highlights some challenges which a laboratory may face. These challenges include the issue of training as, there is a need for better training as well as specialised training courses to help improve the existing knowledge about forensics and also provide specialists in the field of forensic laboratories. The second challenge includes proving the validity of the tools used. As stated in chapter 2 it may be extremely challenging to prove the acceptability of a tool which was used in forensics. Some aspects to consider as regards the acceptability of a tool including whether the tools is fulfilling its purpose and also the tool should do the job for which it is intended and nothing else. There is also the issue of false-positives and false-negatives with the tool reporting a wrong result. The number and severity of false-positives and false-negatives should be considered when taking the validity of a tool into account.

7.4 Forensics Laboratory on the Network

In the paper by Craiger et al the authors proposed a digital forensic laboratory on the network [111]. The laboratory was designed to address the costs of laboratories as well as the lack of training and the backlog at forensic laboratories. It was suggested that the costs could be reduced by a laboratory on the network because the infrastructure could be shared between companies with the laboratory also providing a platform from which to conduct training and on-going exercise. Craiger et al also assumed that, if the laboratories were more accessible, more cases could be conducted at the same time, thus, resolving the backlog that currently exists at the law enforcement departments.

The network laboratory should provide the investigators with the software and hardware required to conduct the investigation [111]. This includes the investigation hardware, storage locations for all the data related to the case, the ability to keep the evidence safe and admissible after the case, the software needed to gather evidence and the software required to manage the case. In order to enhance security the network laboratory uses role based access control. This means that the type of access that is granted to an individual will depend on the specific role which a person has in the investigations. This may be explained in terms of the following examples. The managers would have full control of all cases they are managing, examiners would have access to cases to which they are assigned, and attorneys would have limited access to the reports and results of

certain cases.

The architects of the network laboratories formulated the following requirements. The laboratory must be accessible over the network, preferably the internet [111]. There must be security measures in place, including protecting any communication by form of encryption, authentication and access control and, lastly, the physical machines need to be secured. The laboratory must also implement acceptable case management software. All the data related to cases must be stored. This may be terabytes of data that must be stored but it must also be stored securely. In addition, the software must make it possible to have a full case of custody history for each piece of evidence while all the software and hardware used in the network laboratories must be acceptable in a court of law [111].

As stated in the requirements for a laboratory on the network, it is essential that the issue of security be addressed. This is a challenge for a laboratory on the network as in a normal laboratory the network of the forensic examination would not be connected to external accessible networks. However, with the laboratory on the network, there is a communication channel between the examination network and external networks [111]. Multiple machines are also combined to form the network laboratory and these machines must all comply with all the rules and prerequisites in order to be accepted in a court of law.

The network laboratory adds new roles to the management of laboratories [111]. There should be dedicated staff allocated to maintaining the network laboratory while existing management roles need to be adapted for the network laboratory. In a laboratory the manager would have administrative access to all the computers involved in the investigation. However, in the network laboratory, the manager may have partial access to the computers only, with these computers being maintained by other employees. The management of the network laboratory may be a cumbersome task as it is essential that the administrators possess an extremely sound technical understanding of the systems as well as a thorough understanding of the law. Without proper knowledge on the part of the administrators, all the cases conducted on the network laboratory may retrieve inadmissible evidence.

In order to enhance the security of the network laboratory, Craiger et al implemented NSA security guides to create a secure environment [111]. They also employed logging to ensure have an audit trail of all the action taken on the laboratory. The logging may be done over a dedicated network for logging.

One of the main benefits of the network laboratory is cost reduction. The first reason for the cost reduction is the fact that fewer resources are used, including, less storage and fewer tools with which to conduct an investigation. The second reason is the reduction in duplicate jobs/tasks for example tool verification [111].

A laboratory on the network faces certain challenges, including performance, security, management and transparency [111]. The laboratory on the network is also limited by the speed of the network. In addition, once the evidence is in the laboratory there must be no performance loss and/or performance gain despite the fact that the evidence still has to be transferred into the laboratory over the network. The size of possible evidence may be in terabytes. Transferring one terabyte over a 100 MBit LAN network may take over a day although, with internet speeds lower than 100 Mbit, it may take even longer to transfer all the evidence required to the laboratory.

The creators of the network laboratory have made attempts to address security concerns on the network laboratory by using authentication mechanisms for systems and access control on the network and by confining users to a sandbox [111]. The laboratory would make use of multi-factor authentication in terms of which the user must provide more than one form of user authentication in order to gain access to the system. For example: the authentication mechanism may be both a password and a security token. The token has to be presented at the connection station and the password must be used when resources are accessed. Thus, if one or both of these forms of user authentication are either incomplete or incorrect the authentication will fail.

The network poses a security vulnerability for which the creators of the network laboratory suggest the use of a VPN and IP bases filtering to enhance the security of the network laboratory [111]. The VPN would use dedicated hardware to establish a connection into the network laboratory with this VPN enabling two-factor authentication in terms of which two forms of authentication are required. As mentioned above, this may be a password and a security token. The second possible solution to network vulnerability is the use of IP-based filtering, a “white list” or trusted list of IPs that is used to block all non-listed IPs. This should allow only registered and trusted users to access the network. These two proposed solutions may be used in conjunction to provide the best chance of preventing intrusions.

The last solution as regards addressing security concerns in the network laboratory is the sandboxing of users [111]. In terms of this approach, each user is given a workspace. This workspace contains all the data required by the user. A user is not able to access the workspace of another user as each workspace is protected by multi-factor authentication

which uses LDAP to control access. This ensures that an authenticated user only gain accesses to their data if a user is compromised the data of the other users is still protected.

The last challenge is transparency. It is vital that, the end user experience the environment as both accessible and easy to use [111]. The security systems implemented may mean that the network laboratory is difficult to access and also to use. Nevertheless, the network laboratory should be easy to use and easy to access by legitimate users. In addition, the environment should not become a burden for the examiners but it should contribute to the efficiency of examiners conducting an investigation. The creators of the network laboratory should create an environment in which examiners feel that they are working on their own workstations in their own offices.

Craiger et al propose the use of virtual machines with data stored on the SAN [111]. The reasons for using virtual hardware include creating logical separation for users and cost reduction while the use of standardised hardware decreases deployment time and allows for load balancing.

One of the issues to be taken into account in a network laboratory is the type of storage [111]. Craiger et al chose a SAN based on its performance and also the storage capacity. The SAN provides fast IO throughput to multiple users while offering adequate space for an investigation. The SAN itself may be partitioned into various logical partitions, each for a different end use. The use may include VM storage, case data and so forth. The logical separation assisted in the access control of role bases with access being restricted to certain logical parts of the SAN.

Although there are challenges in respect of the network laboratory, Craiger et al argue in favour and support the importance of the network laboratory [111]. A network laboratory may reduce the costs of an investigation and also enhance the admissibility of the evidence.

7.5 Forensic Laboratory in the Cloud

The cloud laboratory would be hosted on a cloud. This may be either a private or a public hosting depending on the requirements for the laboratory. The cloud laboratory would be designed independent from a specific cloud operating system, and it would be built on common cloud characteristics. The cloud laboratory would also use some of few starting points indicated by Craiger et al [111].

In order to introduce the notion of a laboratory in the cloud, the discussion will be divided into template repository, layout, security measures, administration, challenges and advantages. Templates are preconfigured instances that perform a specific task. The cloud laboratory would have a variety of templates with which to conduct an investigation while the layout is the logical layout of both the cloud laboratory and the required instances while the security measures are those measures that would be implemented in the laboratory in order to protect the evidence. Administration is a central aspect of any laboratory and the cloud laboratory would require unique skills in its administration. The challenges faced by the cloud laboratory will also be discussed but the last section will elaborate on the advantages that may be gained from cloud computing and the advantages of the cloud laboratory.

7.5.1 Template Repository

In the cloud a template is a virtual machine that would have been configured to perform a certain task and form the basis on which to build. For example, a template may be a Windows 2008 R2 server virtual machine, which may be used as a base for other uses, such as a mail server. In the cloud it is possible to clone a virtual machine that is to create a copy of the original. Thus enables for the cloning of a template.

The templates may be preconfigured to perform certain tasks of the investigation. In other words, when a specific task needs to be performed, the template is cloned and the new virtual machine is used for that task. The advantages of using templates include the fact that it is possible for a template to be preconfigured and tested by experts as regards the task it is required to perform. In addition, the template may be tested to confirm whether it is capable of collecting admissible evidence, while it would also enhance rapid deployment. The use of templates also reduces the cost of the investigation because the experts are required to setup the template once only.

For example, a template for an investigation may be an instance which has been created to conduct stenographic analysis. This template would consist of the tools required and installed and preconfigured, to conduct the stenographic analysis without the investigator being present to configure the software. Another example of a template for an investigation would be an instance that has had some form of investigation management tool installed and preconfigured. This configuration should enable a successful investigation to be conducted and where the minimum requirements for a successful investigation

are determined for all investigation cases and built into the template, thus forcing the investigators to meet the minimum requirements.

Part of setting up a template involves creating template groups. Group templates are combined and preconfigured to perform certain tasks involved in an investigation. The templates group form a logical landscape created in order to conduct an investigation. This landscape should possess all the tools required to conduct the investigation. For example, a landscape may include both the steganography analysis template and the management template examples and, thus, there is an analysis instance and a management instance. There are different landscapes to meet the needs of an investigation. For example, in a basic landscape there would be at least one analysis instance and one management instance. However, this may increase to n-number of analysis instances and m-number of management instances. In a landscape all the connections between the instances would be set up and tested and a whole landscape could be provisioned and used for an investigation. The landscapes may also be edited. If there were no landscape to meet the requirements of a certain investigation, then the most closely matching landscape could be deployed and customised to meet the requirements of the investigation.

7.5.2 Layout

The laboratory on the cloud would use the cloud infrastructure as a platform. However, this means that this study will not address the cloud layer in detail. However, chapter 3 does contain a discussion of the cloud infrastructure. Nevertheless, the layer above the cloud layer will be discussed in detail for the purposes of this study. The cloud laboratory may be hosted on any public provider or on a private cloud, depending on the needs of the investigation. The laboratory should also be independent from the cloud operating systems. Figure 7.1 depicts a layout of a cloud laboratory.

The cloud laboratory consists of the following three basic parts, namely, the user part, the management part and the analysis part. The user part consists of the users access to the laboratory over some form of network, the management part consists of the instances that maintain and manage the cloud while the analysis part consists of the instances which are used to conduct and manage the investigation.

Users of the cloud laboratory would access the laboratory by using their own personal computer through some form of network. These users of the laboratory would have

access to the cloud laboratory by using an Open-VPN connection or any other client server based on a VPN connection. The user is required to have SSL certificates as well as a parse phrase and access card in order to open a VPN connection. The users would include the examiners, the managers, and any other persons involved in the investigation. The VPN is the first level of access for users. Thereafter, the users are able to access the analysis machines but not the management machines. The analysis machines are accessed by either a webgui (if the users have one) or through a remote desktop connection.

All access to and from the cloud laboratory would go through a firewall. The firewall would control the first level access to the laboratory and the VPN, and would have a whitelist of assessors. This whitelist would contain the IPs of the authorised users to the system, and only users on the list will be allowed to setup a VPN connection. The firewall is also responsible for keeping the network inside the laboratory secure and isolated.

Inside the cloud laboratory there are two networks; one for the management of the machines and one for the analysis machines. These two networks are kept separate in the interest of the security of the cloud laboratory. The networks are logical as they run on a shared infrastructure by using either VLANS, virtual cloud networks or a combination of both. The firewall would be responsible for the traffic flow between the two networks.

On the management network there would a LDAP server, a logging server and a file share server. The LDAP server is used to authenticate users of the laboratory. The users are required to provide an username, passphrase and a key card when accessing resources on the analysis network. The LDAP server will use SSL certificates to encrypt LDAP traffic. All the resources on the cloud laboratory would use the LDAP server for authentication, and once passed, the user would gain access to the resource requested.

The logging server is used to log all activity on the laboratory and, thus, every action which an user performs on any analysis machine is logged. In additions, all the connection requests and all authentications are also logged. The logging is done on a separate server to protect the logs from both tampering and contamination. The logs may be used as an audit trail to prove the admissibility of the evidences. At the very least the logs should indicate from what resource the message is, whom the user is, a time stamp, what the action was and any resources involved, including evidence from the case.

The file server would have all the files of the case as well as the evidence and data. No direct access to the file server is possible and all file requests must go through a virtual

file system. The virtual file system, on a written request, would write a new version abbreviated file to disk and, thus, this means that the original file is never overwritten. On a read request the latest version of the file is returned. There may also be special read requests that return a specific version of the file. This would assist in the protection of the evidence as an evidence trail is created and the original evidence is kept intact. This works in a similar way to a software development versioning system where all versions of the source code are kept and it is possible to go back to older versions. There is also a backup server that will keep backups of the files. However, this backup server should be on a remote site while the connection between the backup server and the file server should be a dedicated VPN connection.

The analysis side would have at least one analysis instance and one management instance. Figure 7.1 depicts a setup with a file system analyser, a steganography analyser and a case management machine. The investigators would use these machines in order to conduct the investigation.

The layout described may be launched from a template group where all the connections between the management and analysis side have been set up and all that needs to happen is that users need to be added for the laboratory to function.

7.5.3 Security Measures

The research proposes that the cloud laboratory has security features built into its design. These security features include the use of SSL certificates, a VPN, a whitelist, multi-factor authentication, a protected file system and user privileges.

The VPN assists in the security process as it encrypts all the network traffic to and from the cloud while it blocks unauthorised access to the cloud. The VPN will also make use of the white list to allow only certain IPs past the VPN connection while the firewall will also use the whitelist to confirm that the IPs have been blocked.

The use of SSL certificates ensures that only users with a certificate and the passphrase are able to access the cloud. The certificates are in the form of public private key certificates. Certificates will also be used by the LDAP server. Each server that requests communication with the LDAP server would have to pass a TLS/SSL phase, thus allowing the LDAP server to encrypt all communication. The TLS/SSL negotiation also aids in blocking untrusted instances from authenticating against the LDAP server.

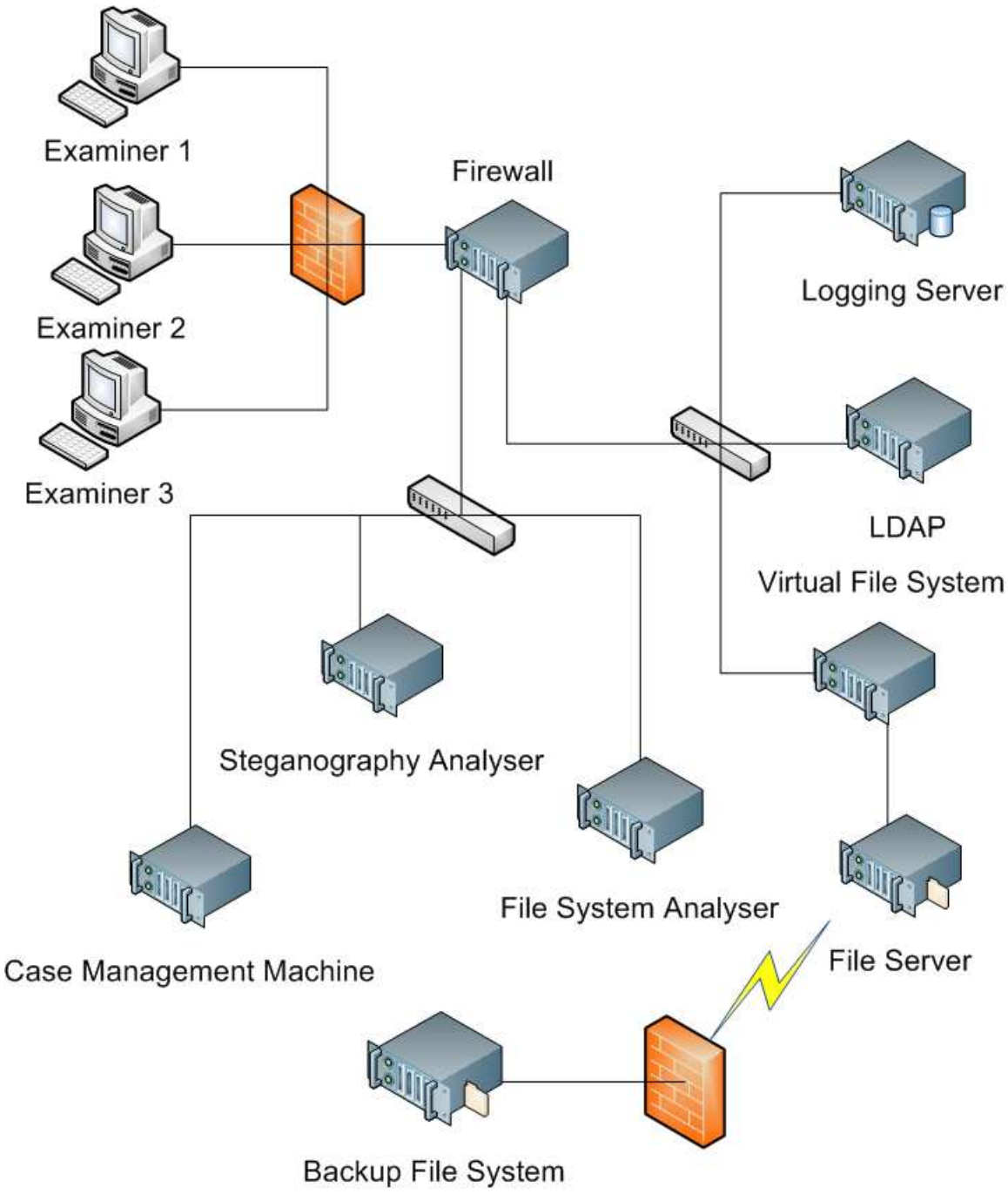

Figure 7.1: Example Layout of a Virtual Environment

The multi-factor authentication forms part of the blocking of unauthorised access. Each user would be assigned certificates, an username, one or more passphrases and a key token. The VPN, in turn, would use the certificates, a passphrase and the key token. When accessing a resource the username, a passphrase and the key token is required. The key token may be in the form of a card scanner, fingerprint reader or any other token mechanism which is used by the specific implementation. The multi-factor authentication assists in ensuring that authorised users only gain access and, thus, it helps to prevent a possible attack.

The next security measure is the protected file system. The file system that stores all the data and evidence related to the case is protected by, firstly, the provision of a virtual file system and, thus, no direct interaction is granted to the actual file system. The manner in which versions of all the files are kept also provides a form of security as there is always a trail of the evidence while it is also possible to go back to previous versions. In addition, it is not possible to delete evidence from the file system and new versions may be saved although it is not possible either to overwrite or delete the original.

The last security feature is privileges. This means that users each have certain privileges assigned to them. These privileges determine what the user is able to do on the cloud laboratory, including the resources the user is able to access in terms of analysis instances and case data. It is possible that not all users would have access to all the data of the laboratory. For example: the file system experts may access the file system through analysing instances and all the data related to file systems, but they may not access the steganography analysing instance and data related to steganography. The administrators and the case managers also have privileges in terms of which they may perform only those specific tasks that they are allowed to perform. For example, one case manager may create users but may not be allowed to access the case data.

7.5.4 Administration

When the cloud laboratory is created it would have no case data and no users. In a newly created laboratory there would be basic accounts and these would, include an administrator and a case manager. When the laboratory is created the administrator and case managers would have one-time passes and they would receive their access tokens. Personnel would log in with the one-time passes and access tokens and they would then be required to create their accounts. After the initial accounts have been created new users could be added to the laboratory. New users would also receive a one-time passes

and will be required to setup their accounts. All authorised personnel would first have to receive their access tokens before being able to access the systems.

The administrator is responsible for creating user certificates for the new users, adding the key cards for these new users to the system and adding users to the LDAP server. However, the administrator is able to can only create the users and maintain the lists and it is the case manager who is responsible for activating the users' accounts. After the administrator has created a new account it is in a disabled state until the case manager has marked it as active. The case manager may also disable an account. The case manager may also create other case managers with either the same or fewer privileges. The case manager may also change the privileges of system users to allow them to carry out different tasks, but the case manager may change only those users whom he/she is authorised to change, otherwise the system would block him/her.

The SSL certificates are created on the LDAP machine by using a tool such as OPENSSL. The certificates are then uploaded to the firewall for VPN access. The administrator and case manager are responsible for the distribution of the certificates.

The administrator is also responsible for adding entries to the white list of the firewall while, the case manager is responsible for activating and disabling them.

7.5.5 Challenges

There are certain challenges associated with the cloud laboratory. These challenges include administration, network speed, security, provider trust and initial setup time. The first challenge is administration. It is essential that the administrators are qualified in the new environment while they may also require initial training. It is difficult to determine the workload of the administrators as the cloud laboratory is a dynamic environment. For example, one week there may only be one cloud laboratory to maintain with three instances and the next week there may be ten cloud laboratories with 50 instances to maintain.

The next problem is network speed. Most of the large file transfers are limited to a network speed or, if over the internet, the internet speed. The interactions with the instances would require a form of network connectivity. If an investigator has limited access to network connectivity, then access to the laboratory may be slow and interrupted.

As does the laboratory on a network, the cloud laboratory also faces security challenges. These security challenges arise mainly because all the data and evidence are closely located in one environment and, if data is contaminated, it becomes inadmissible. However, the security risks may be minimised with the use of proper policies, procedures and a planned implementation.

The next challenge is the issue of trustworthiness of the cloud provider. If the cloud provider is not trustworthy, then the evidence gathered by using the cloud may be inadmissible. It is essential that the cloud providers and cloud laboratory providers work together to ensure that the environment is trusted and it may be proved to be acceptable.

The last challenge is the initial setup time. It is a time-consuming task to create all the templates and template groups for the cloud laboratory. However, once the cloud laboratory has been setup, then the time required to launch it is relatively short. Thus, it is usually only the initial setup that may be time consuming.

7.5.6 Advantages Gained from Cloud Computing

In the section on the network laboratory cost reduction was mentioned as the main advantage to be gained from moving to the virtual hardware. However, although this is important, the list of advantages may be expanded to include the rapid expansion of resources, immediate reaction, training environments, high availability and redundancy.

In an investigation the computing requirements may change as the task of the investigation changes. For example: it takes more computing power to crack the password of an encrypted password as compared to the computing power required to open and read a text file. However, the cloud allows for the expansion and shrinking of the virtual environment. In other words, when more resources are required these extra resources may be assigned and used and, when fewer resources are required, the extraneous may be removed in order to meet the new requirements. This also has an impact on cost as the laboratory pays only for the resources that are required in order to complete the tasks of the investigation.

When an investigation requires a laboratory and a new laboratory must be created, this may be in a relatively short time. However, when a physical laboratory needs to be set up a building is required, the hardware must be setup and the software installed and

configured. In the cloud the virtual environment must be created, but this takes less time than it does to set up a physical laboratory. The cloud laboratory may also be created anywhere with cloud hardware and, with cloud computing being a global phenomenon it should be possible to create a cloud laboratory anywhere in the world. This means that a new laboratory may be created almost anywhere in the word in a relatively short time and this, in turn, means a fast reaction to the need to conduct an investigation.

It may be difficult to provide training inside a laboratory as there are often no spare resources available. In addition, it is not possible to use resources which are used for an active investigation for training because the evidence may be contaminated. However, the cloud laboratory helps to overcome this challenge as it is possible to create and use a training environment without contaminating the evidence of an active investigation. Another advantage in respect for training is that the environment may be changed or recreated to provide training of a different form inside the laboratory. For example: the first setup may be to conduct a simple file investigation but it may then be expanded to a steganography analysis. In other words, this building characteristic enables a laboratory to be created from templates and, thus, enable the trainee to fulfil different roles at various stages of the investigation. For example, in one template only the raw, gathered data is present and the trainee has to gather evidence. However, in the other template the investigation has been completed and the trainee needs to document what has been done. Thus, the templates facilitate the creation of a predictable and known environment for training.

When an investigation is being conducted it may be argued that high availability is required. As discussed in chapter 3, one of the basic properties of cloud computing is the guarantee of high availability. Such high availability enables the investigators to conduct the investigation without interruption. Another property of cloud computing that may assist in an investigation is redundancy. Redundancy helps to prevent data loss and this, in turn, helps to ensure admissible evidence even if there was a loss of hardware at the cloud provider.

7.6 Conclusion

Although there has been progress made on the network laboratory the cloud laboratory may also be a vital tool in some cases. The cloud laboratory may be used by small and medium sized investigation companies and law enforcement offices which do have the

capital required to set up their own on site laboratories and keep them updated.

Chapters 5 and 6 introduced the notion of isolating an instance inside the cloud environment, while this chapter introduced the notion of a cloud digital forensic laboratory which may be used to conduct an investigation using the cloud infrastructure. The cloud laboratory may also be used to investigate instances that were isolated or the part of the cloud that was isolated. This means that the instances never have to leave the cloud environment and may merely be moved to the cloud laboratory. This would increase the response time of an investigation on the cloud.

The next chapter is the final chapter of the study and it serves as a conclusion to the study.

Chapter 8

Conclusion

8.1 Introduction

This study aimed at providing models which should assist in the isolation of evidence inside the cloud environment. To the best of the authors knowledge this is the fist attempt at isolation inside the cloud. The study created various models and techniques that can be used to achieve isolation inside the cloud. The study also introduced conditions for isolation that can be used as a guide as to whether the solution was successful. These models were tested empirically during experiments using two different cloud providers.

8.2 Derived publications

The following paper where produced as part of the contribution of the study.

- W. Delport, M. S. Olivier, and M. Köhn, "Isolating a cloud instance for a Digital Forensic Investigation." in *Information Security South Africa Conference (ISSA)*, 2011. (Based on chapter 5 of this dissertation.)
- W. Delport and M. S. Olivier, "Isolating instances in cloud forensics", in *Advances in Digital Forensics VIII*, ser. IFIP Advances in Information and Communication Technology, G. Peterson and S. Shenoi, Eds. Springer Berlin Heidelberg, 2012, vol. 383, pp. 187 - 200. (Based on chapter 5 of this dissertation.)

- W. Delport and M. S. Olivier, "Cloud separation: Stuck inside the cloud" in *Trust, Privacy and Security in Digital Business*, ser. Lecture Notes in Computer Science, S. Fischer-Hbner, S. Katsikas, and G. Quirchmayr, Eds. Springer Berlin Heidelberg, 2012, vol. 7449, pp. 36 - 49. (Based on chapter 6 of this dissertation.)

8.3 Future Work

Possible future work may include testing the methods for cloud separation and instance isolation on more cloud operating systems in order to better test all the methods and discover any possible pitfalls. If it were discovered that the methods do not work on all platforms then further research would need to be conducted to discover other methods that would work on specific platforms. There is also a need to investigate the performance loss when an investigation on a cloud is conducted.

The cloud laboratory also needs to undergo further development implemented and tested in order to confirm its viability.

8.4 Summary of work

The first experiment conducted involved isolating either a single instance or small set of instances. The aim of the isolation was to protect the admissibility of possible evidence. The isolation techniques included Instance Relocation, Server Farming, Address Relocation, Failover, Sandboxing, Man in the Middle (MITM) and Lets Hope for the Best (LHFTB). One of the experiments found that a single technique on its own is not sufficient and that a combination of techniques should be used. The chapter then introduced conditions for isolation. Meeting these conditions should ensure the success of the isolation process. The conditions included the physical location of the instance must be known (Location); the instance is protected from outside interference (Incoming Blocking); the instance is blocked from communicating with the outside world (Outgoing Blocking); possible evidence from the instance may be gathered (Collection); the possible evidence is not contaminated by the isolation process (Non-Contamination) and information not related to the incident is not part of the isolation (Separation). These conditions were then tested in an experiment to determine their validity. The experiment revealed that all the conditions were valid. "Live" and "dead" forensic models as well

as "*halfdead*", "*resurrected*" and "*Community Live*" models were also addressed. The complete experiment is discussed in chapter 5.

In contrast to isolating a single instance a part of the cloud may also be isolated with cloud separation being used to isolate a sub part of the cloud. Cloud separation was introduced in order to isolate DiS landscapes where multiple instances formed one logical resource. The notion of cloud separation was divided into two aspects, namely, moving instances and dividing the cloud. Moving the instance involves relocating the instance from one node to another while a division of the cloud is done in order to complete the isolation. The movement methods include moving instances from one cloud to another directly, moving the instances to an external cloud and then from there to the other cloud, moving the instance to an external cloud, then moving it to one or more other external clouds and finally moving it to the other cloud, using the cloud operating system to move the instances and lastly, merely identifying the nodes which contain suspect instances but not moving the instances. The cloud may be divided by separating the nodes by creating two separate networks from one network, creating two virtual networks on one logical network, creating sub-clouds inside the actual cloud and, lastly, using the movement methods to move the instances to a cloud dedicated for the investigation. These movement and division methods were tested and it was determined that the investigator would have to choose the most appropriate method for isolation at the start of an investigation. This is as a result of the fact that there is no single best combination and how well the methods function depends on the environment. Chapter 6 contains a full account of the result of the experiments.

Once isolated the instances may be moved to a laboratory. The study introduced the notion of a laboratory inside the cloud and utilising the advantages of the cloud. There is also the added advantage of the cloud laboratory that all instances may be kept inside the cloud environment and the investigation may be conducted inside the cloud environment. The cloud laboratory also makes it possible to provide laboratories for non-cloud related investigations. The cloud laboratory builds on the characteristics of cloud computing, introducing the notion of Software as a Service (SaaS) laboratory. Chapter 7 contains a complete description of the proposed cloud laboratory.

Bibliography

[1] M. Vouk, "Cloud computing: Issues, research and implementations," in *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on*, June 2008, pp. 31 – 40.

[2] D. Barrett and T. King, *Computer Networking Illuminated*, B. Parrish and R. Taylor, Eds. Jones & Bartlett Publishers, 2005.

[3] S. Biggs and S. Vidalis, "Cloud computing: The impact on digital forensic investigations," in *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference*, November 2009, pp. 1 – 6.

[4] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop, 2008. GCE '08*, November 2008, pp. 1 –10.

[5] K. Nance, H. Armstrong, and C. Armstrong, "Digital forensics: Defining an education agenda," in *43rd Hawaii International Conference on System Sciences (HICSS)*, January 2010, pp. 1 – 10.

[6] F. Mirza, "Looking for digital evidence in Windows," in *International Symposium on Biometrics and Security Technologies*, April 2008, pp. 1 – 7.

[7] A. Singh and M. Shrivastava, "Overview of attacks on cloud computing," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 1, pp. 321 – 323, 2012.

[8] K. Zunnurhain and S. V. Vrbsky, "Security attacks and solutions in clouds," in *Proceedings of the 1st International Conference on Cloud Computing*, 2012, pp. 145 – 156.

[9] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J. Molina, "Controlling data in the cloud: outsourcing computation without outsourcing control," in *Proceedings of the 2009 ACM workshop on Cloud computing security*. New York, NY: ACM, 2009, pp. 85 – 90.

[10] G. Palmer, "A road map for digital forensic research, Report from the first digital forensic research workshop (DFRWS)," First Digital Forensic Research Workshop, Utica, New York, Tech. Rep., August 2001.

[11] B. Fei, "Data visualisation in digtial forensics," Master's Dissertation, University of Pretoria, 2007.

[12] J. Ashcroft, *Electronic Crime Scene Investigation: A Guide for First Responders*, Technical Working Group for Electronic Crime Scene Investigation, July 2001.

[13] F. Cohen, *Digital Forensic Evidence Examination*, 2nd ed. Livermore, CA: Fred Cohen & Associates, February 2010.

[14] K. Nance, B. Hay, and M. Bishop, "Digital forensics: Defining a research agenda," in *42nd Hawaii International Conference on System Sciences*, January 2009, pp. 1 – 6.

[15] N. Lim and A. Khoo, "Forensics of computers and handheld devices: Identical or fraternal twins?" *Commun. ACM*, vol. 52, pp. 132 – 135, June 2009.

[16] J. R. Lyle, "A strategy for testing hardware write block devices," *Digital Investigation*, vol. 3, Supplement, no. 0, pp. 3 – 9, 2006, the Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06).

[17] *Electronic Crime Scene Investigation: A Guide for First Responders*, 1st ed., U.S. Department of Justice, July 2001, NCJ Number: NCJ 187736.

[18] vmware, "VMware Inc," Computer Program, 2011, vSphere 5.0. [Online]. Available: http://www.vmware.com/

[19] Nimbula, "Nimbula Director," Computer Program, version 1.0.3. [Online]. Available: http://nimbula.com/products/overview/

[20] R. Koen, "The development of an open-source forensics platform," Master's Dissertation, University of Pretoria, 2009.

[21] S. Bosworth, M. E. Kabay, and E. Whyne, Eds., *Computer Security Handbook*. New York: Wiley, March 2009.

[22] *XLVI Chapter 815 Computer-related Crimes (Computer Crimes Act)*, Florida Statutes, United States of America, 1988.

[23] SWGDE, "Best practices for computer forensics," Scientific Working Group on Digital Evidence, Tech. Rep. 2.1, 2006.

[24] S. Yngvar and S. F. Mjolsne, "Digital forensics research," *Telektronikk*, vol. 1, pp. 92 – 97, 2005.

[25] oxforddictionaries.com, "Forensic," http://www.oxforddictionaries.com/definition/forensic, Oxford Dictionary, Last Accessed: July 31, 2013. [Online]. Available: http://oxforddictionaries.com/

[26] M. A. Caloyannides, N. Memon, and W. Venema, "Digital forensics," *Security Privacy, IEEE*, vol. 7, no. 2, pp. 16 – 17, March 2009.

[27] oxforddictionaries.com, "Evidence," http://www.oxforddictionaries.com/definition/evidence, oxford Dictionary, Last Accessed: July 31, 2013. [Online]. Available: http://www.oxforddictionaries.com

[28] E. Casey, "Digital dust: Evidence in every nook and cranny," *Digital Investigation*, vol. 6, no. 3-4, pp. 93 – 94, 2010.

[29] S. Mrdovic, A. Huseinovic, and E. Zajko, "Combining static and live digital forensic analysis in virtual environment," in *Information, Communication and Automation Technologies, 2009. ICAT 2009. XXII International Symposium on*, 2009, pp. 1 –6.

[30] F. Law, K. Chow, M. Kwan, and P. Lai, "Consistency issue on live systems forensics," in *Future Generation Communication and Networking (FGCN 2007)*, vol. 2, 2007, pp. 136 –140.

[31] J. R. Vacca and K. Rudolph, *System Forensics, Investigation, and Response*, L. J. Goodrich, Ed. Jones & Bartlett Learning, 2010.

[32] V. Corey, C. Peterman, S. Shearin, M. Greenberg, and J. Van Bokkelen, "Network forensics analysis," *Internet Computing, IEEE*, vol. 6, no. 6, pp. 60 – 66, November 2002.

[33] B. Mellars, "Forensic examination of mobile phones," *Digital Investigation*, vol. 1, no. 4, pp. 266 – 272, 2004.

[34] G. G. Richard, III and V. Roussev, "Next-generation digital forensics," *Commun. ACM*, vol. 49, pp. 76–80, February 2006.

[35] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, "Cloud forensics: An overview," *IFIP International Conference on Digital Forensics*, vol. 7, pp. 35 – 49, 2011.

[36] T. V. Lillard, *Digital Forensics for Network, Internet, and Cloud Computing: A Forensic Evidence Guide for Moving Targets and Data*, J. Murray, Ed. Syngress Publishing, 2010.

[37] M. S. Olivier, "On metadata context in database forensics," *Digital Investigation*, vol. 5, no. 34, pp. 115 – 123, 2009.

[38] K. Arthur, "Considerations towards the development of a forensic evidence management system," Master's Dissertation, University of Pretoria, 2010.

[39] C. M. Whitcomb, "An Historical Perspective of Digital Evidence: A Forensic Scientists View," in *International Journal of Digital Evidence*, vol. 1.1, 2002, pp. 7 – 15.

[40] J. R. E. Blackburn, "Order granting application under the all writs requiring defendant fricosu to assist in the execution of previously issued search warrants," In the United States District Court for the District of Colorado, January 2012, criminal Case No. 10-cr-00509-REB-02.

[41] D. Kravets. (2012, January) Judge orders defendant to decrypt laptop. Online. Accessed: 23 July 2012. [Online]. Available: http://www.wired.com/threatlevel/2012/01/judge-orders-laptop-decryption/

[42] M.-A. Fouche and M. S. Olivier, "Steganographic techniques for hiding data in swf files," in *IFIP Int. Conf. Digital Forensics'11*, 2011, pp. 245–255.

[43] P. Dixon, "An overview of computer forensics," *Potentials, IEEE*, vol. 24, no. 5, pp. 7 – 10, December 2005.

[44] H. Kim, T. Benson, A. Akella, and N. Feamster, "The evolution of network configuration: A tale of two campuses," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '11. New York, NY, USA: ACM, 2011, pp. 499–514.

[45] K. Thomas, "Redo and rollback," in *Expert Oracle*. A-Press, 2005, pp. 157–196.

[46] S. Raghav and A. Saxena, "Mobile forensics: Guidelines and challenges in data preservation and acquisition," in *Research and Development (SCOReD), 2009 IEEE Student Conference on*, nov. 2009, pp. 5 –8.

[47] W. Jansen and R. Ayers, *Guidelines on Cell Phone Forensics*, National Institute of Standards and Technology, May 2007.

[48] M. A. Weiss, *Data Structures and Algorihm Analysis in Java*, 2nd ed., M. Hirsch, Ed. Greg Tohin, 2007.

[49] F. Liu and Y. Liu, "Avalanche of MD5," *International Conference on Energy Systems and Electrical Power (ESEP 2011)*, vol. 13, pp. 237 – 246, 2011.

[50] M. Zeghida, B. Bouallegue, A. Baganne, M. Machhout, and R. Tourki, "A reconfigurable implementation of the new secure hash algorithm," in *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*, April 2007, pp. 281 –285.

[51] J. Yajima, T. Iwasaki, Y. Naito, Y. Sasaki, T. Shimoyama, N. Kunihiro, and K. Ohta, "A strict evaluation method on the number of conditions for the sha-1 collision search," in *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '08. New York, NY, USA: ACM, 2008, pp. 10–20.

[52] "Guidance software - EnCase Forensic," Computer Program, vertion 7. [Online]. Available: http://www.guidancesoftware.com/forensic.htm

[53] "Access data - FTK," Computer Program. [Online]. Available: http://accessdata.com/products/computer-forensics/ftk

[54] K. Garloff, "dd_rescue," Computer Program, vertion 1.23. [Online]. Available: http://www.garloff.de/kurt/linux/ddrescue/

[55] M. Solomon, D. Barrett, N. Broom, and K. Rudolph, *Computer Forensics Jump-Start*, W. G. Krus, Ed. Wiley, 2004.

[56] "Sourcefire - snort," Computer Program. [Online]. Available: http://www.snort.org/

[57] "Wireshark Foundation - Wireshark," Computer Program. [Online]. Available: http://www.wireshark.org/

[58] "Tcpdump/Libpcap - tcpdump," Computer Program. [Online]. Available: http://www.tcpdump.org/

[59] V. S. Sunderam, "Pvm: A framework for parallel distributed computing," *Concurrency: Practice and Experience*, vol. 2, no. 4, pp. 315–339, 1990.

[60] M. Isard, V. Prabhakaran, J. Currey, U. Wieder, K. Talwar, and A. Goldberg, "Quincy: Fair scheduling for distributed computing clusters," in *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*. ACM, 2009, pp. 261 – 276.

[61] D. S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, and Z. Xu, "Peer-to-peer computing," HP Laboratories Palo Alto, Tech. Rep. HPL-2002-57, 2002.

[62] M. Ebbers, W. O'Brien, and B. Ogden, *Introduction to the New Mainframe: z/OS Basics*, T. Barthel, E. Buslovich, and A. Schwab, Eds. IBM, International Technical Support Organization, 2006.

[63] T. Richardson, Q. Stafford-Fraser, K. R. Wood, and A. Hopper, "Virtual network computing," *Internet Computing, IEEE*, vol. 2, no. 1, pp. 33 – 38, 1998.

[64] J. Day, *Patterns in Network Architecture, A Return to Fundamentals*, K. Gettman, Ed. Prentice Hall, 2008.

[65] K. L. Cohen and R. P. Levy, "X.25 implementation the untold story," in *Proceedings of the Symposium on Communications Architectures & Protocols*, ser. SIGCOMM '83. New York, NY, USA: ACM, 1983, pp. 60–64.

[66] P. Mell and T. Grance, "The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technolog," National Institute of Standards and Technology, Tech. Rep., 2011.

[67] G. Reese, *Database Programming with JDBC and Java*, second edition ed., A. Oram, Ed. O'Reilly Media, 2000.

[68] J. Bhattacharya and S. Vashistha, "Utility computing-based framework for e-governance," in *Proceedings of the 2nd International Conference on Theory and*

Practice of Electronic Governance, ser. ICEGOV '08. New York, NY, USA: ACM, 2008, pp. 303–309.

[69] T. G. Group, "Virtualization in education," IBM Global Education, Tech. Rep., 2007.

[70] Xen, "Xen," Computer Program. [Online]. Available: http://www.xen.org/products/cloudxen.html

[71] S. Hassan, D. Al-Jumeily, and A. Hussain, "Autonomic Computing Paradigm to Support System's Development," in *Developments in eSystems Engineering (DESE), 2009 Second International Conference on*, December 2009, pp. 273 –278.

[72] G. Reese, *Cloud Application Architectures: Building Applications and Infrastrucure in the Cloud*, 1st ed., A. Oram, Ed. O'Reilly Media, 2009.

[73] C. Binnig, D. Kossmann, T. Kraska, and S. Loesing, "How is the weather tomorrow?: Towards a benchmark for the cloud," in *Proceedings of the Second International Workshop on Testing Database Systems*, ser. DBTest '09. New York, NY, USA: ACM, 2009, pp. 1 – 9.

[74] R. Lu, X. Lin, X. Liang, and X. S. Shen, "Secure provenance: The essential of bread and butter of data forensics in cloud computing," in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '10. New York, NY, USA: ACM, 2010, pp. 282–292.

[75] Amazon, "Amazon elastic compute cloud (amazon ec2)," Online Service. [Online]. Available: http://aws.amazon.com/ec2/

[76] Nitu, "Configurability in SaaS (software as a service) applications," in *Proceedings of the 2nd India Software Engineering Conference*, ser. ISEC '09. New York, NY, USA: ACM, 2009, pp. 19 – 26.

[77] E. Knorr and G. Gruman, "What cloud computing really means," *InfoWorld*, vol. cloud computing, p. 34031, 2008.

[78] salesforce.com, "salesforce.com," Online Service. [Online]. Available: http://www.salesforce.com/eu/?ir=1

[79] PaySpace, "Payspace," Online Service, Insight IT Group. [Online]. Available: http://www.payspace.co.za/

[80] R. Hosting, "Rackspace hosting," Online Service. [Online]. Available: http://www.rackspace.co.za/

[81] Aeolus, "Aeolus," Computer Program. [Online]. Available: http://aeolusproject.org/about.html

[82] E. Casalicchio and S. Tucci, "Static and dynamic scheduling algorithms for scalable web server farm," in *Parallel and Distributed Processing, 2001. Proceedings. Ninth Euromicro Workshop on*, 2001, pp. 369 –376.

[83] vmware. (2011, July) Vmware vsphere, the best platform for cloud infrastructures. [Online]. Available: http://www.vmware.com/files/pdf/press-kit/vmw-vsphere-cloud-infrastructure.pdf

[84] A. Rowland. (2011, October) Nimbula announces auto-scaling and auto-discovery of resources capabilities for vmware cloud foundry that it will demo live at vmworld. Press Release. Nimbula. Last Accessed: July 31, 2013.

[85] ——. (2011, December) Nimbula wins the overall most innovative cloud platform award at the up cloud computing conference. Press Release. Nimbula. Last Accessed: July 31, 2013.

[86] M. Almorsy, J. Grundy, and A. S. Ibrahim, "Collaboration-based cloud computing security management framework," in *IEEE 4th International Conference on Cloud Computing*, 2011.

[87] Songjie, J. Yao, and C. Wu, "Cloud computing and its key techniques," in *International Conference on Electronic & Mechanical Engineering and Information Technology*, 2011.

[88] M. M. Khosa, *Infrastructure Mandate for Change 1994-1999*, M. M. Khosa, Ed. Human Sciences Research Council, 2001.

[89] M. Taylor, J. Haggerty, D. Gresty, and R. Hegarty, "Digital evidence in cloud computing systems," *Computer Law and Security Review*, vol. 26, no. 3, pp. 304 – 308, 2010.

[90] C. Yan, "Cybercrime forensic system in cloud computing," in *Image Analysis and Signal Processing (IASP), 2011 International Conference on*, oct. 2011, pp. 612 –615.

[91] A. Amies, C. F. Wu, G. C. Wang, and M. Criveti, "Networking on the cloud, explore tools and concepts for networking on ibm smartcloud enterprise," IBM, Tech. Rep., 2012.

[92] G. Huerta-Canepa and D. Lee, "A virtual cloud computing provider for mobile devices," in *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*, ser. MCS '10. New York, NY, USA: ACM, 2010, pp. 6:1 – 6:5.

[93] K. Ruan, J. James, J. Carthy, and T. Kechadi, "Cloud forensics: Key terms for the Service Level Agreement," *Eighth Annual IFIP WG 11.9 International Conference on Digital Forensics (In Press)*, vol. 8, 2012.

[94] D. Ras and M. S. Olivier, "Finding file fragments in the cloud," in *Advances in Digital Forensics VIII*. Springer, 2012, pp. 169 – 185.

[95] D. Reilly, C. Wren, and T. Berry, "Cloud computing: Forensic challenges for law enforcement," in *International Conference for Internet Technology and Secured Transactions (ICITST)*, nov. 2010, pp. 1 –7.

[96] G. Grispos, W. B. Glisson, and T. Storer, "Calm before the storm: The emerging challenges of cloud computing in digital forensics," *International Journal of Digital Crime and Forensics*, vol. 4, Issue: 2, pp. 28 – 48, 2012.

[97] A. Yasinsac and Y. Manzano, "Policies to enhance computer and network forensics," in *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security*, 2001, pp. 289 – 295.

[98] D. Reilly, C. Wren, and T. Berry, "Cloud computing: Pros and cons for computer forensic investigations," *International Journal Multimedia and Image Processing (IJMIP)*, vol. 1, pp. 23 – 43, 2011.

[99] D. Birk and C. Wegener, "Technical issues of forensic investigations in cloud computing environments," in *Systematic Approaches to Digital Forensic Engineering (SADFE), 2011 IEEE Sixth International Workshop on*, May 2011, pp. 1 –10.

[100] oxforddictionaries.com, "Isolate," http://oxforddictionaries.com/definition/isolate, Oxford Dictionary, Last Accessed: July 31, 2013. [Online]. Available: http://oxforddictionaries.com/

[101] P. White, *Crime Scene to Court: The Essentials of Forensic Science*, 3rd ed., P. White, Ed. Royal Society of Chemistry, 2010.

[102] D. S. Milojíčić, F. Douglis, Y. Paindaveine, R. Wheeler, and S. Zhou, "Process migration," *ACM Comput. Surv.*, vol. 32, pp. 241–299, September 2000.

[103] K. Singh and H. Schulzrinne, "Failover, load sharing and server architecture in sip telephony," *Computer Communications*, vol. 30, no. 5, pp. 927 – 942, 2007, advances in Computer Communications Networks.

[104] I. Kuzminykh, "Failover and load sharing in sip -based ip telephony," in *Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2008 Proceedings of International Conference on*, February 2008, pp. 420 – 422.

[105] R. Zhang, T. Abdelzaher, and J. Stankovic, "Efficient tcp connection failover in web server clusters," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 2, march 2004, pp. 1219 – 1228 vol.2.

[106] C. P. Pfleeger and S. L. Pfleeger, *Security in Computing*, 4th ed. Prentice Hall, 2006.

[107] C. Greamo and A. Ghosh, "Sandboxing and virtualization: modern tools for vombating malware," *Security Privacy, IEEE*, vol. 9, no. 2, pp. 79 –82, April 2011.

[108] M. Smith, T. Friese, M. Engel, and B. Freisleben, "Countering security threats in service-oriented on-demand grid computing using sandboxing and trusted computing techniques," *Journal of Parallel and Distributed Computing*, vol. 66, no. 9, pp. 1189 – 1204, 2006, security in grid and distributed systems.

[109] A. Jones and C. Valli, *Building a Digital Forensic Laboratory: Establishing and Managing a Successful Facility*, P. Chester and M. Cater, Eds. Butterworth-Heinemann, 2008.

[110] J. Beckett and J. Slay, "Scientific underpinnings and background to standards and accreditation in digital forensics," *Digital Investigation*, vol. 8, no. 2, pp. 114 – 121, 2011.

[111] P. Craiger, P. Burke, C. Marberry, and M. Pollitt, "A Virtual Digital Forensics Laboratory," in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi, Eds., vol. 7. Springer Berlin Heidelberg, 2008, pp. 357 – 365.

Appendix A

Abbreviations

AMI	Amazon Machine Images
CIA	Confidentiality, integrity and availability
DAS	Direct-attached storage
DF	Digital forensics
DFI	Digital forensic investigation
DFP	Digital forensic procedure
DiS	Distributed instance system
HA	High availability
IaaS	Infrastructure as a Service
ISO	International Organization for Standardization
LHFTB	Let's hope for the best
MD5	Message-digest algorithm
MITM	Man in the middle
MITMA	Man in the middle attack
NIJ	National Institute of Justice
PaaS	Platform as a Service
SaaS	Software as a Service
SAN	Storage area network
SHA	Secure hash algorithm
SLA	Service level agreement
SMME	Small, medium and micro enterprises
SWGDE	Scientific Working Group on Digital Evidence