Digital Forensics in the Cloud: The Reliability and Integrity of the Evidence Gathering Process

MARK TANGIWAI MATHEW PIWARI

Bachelor of Information Technology (WINTEC, NZ), Diploma in Applied Computer Systems Engineering (UNITEC, NZ), Diploma in Business Computing (BOP Polytechnic, NZ)

A thesis submitted to the graduate faculty of design and creative technologies
Auckland University of Technology
in partial fulfilment of the
requirements for the degree of
Master of Forensics Information Technology

School of Engineering, Computer and Mathematical Sciences

Auckland, New Zealand
2016

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a University or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

.....

Mark Tangiwai Mathew Piwari

Acknowledgements

This thesis was conducted at the Faculty of Design and Creative Technologies in the school of Computing and Mathematical Sciences at Auckland University of Technology, New Zealand. Support was received from a number of individuals throughout the duration of the thesis although two people took centre stage. First, I would like to thank my partner of 32 years David Robertson, for his unconditional support and resolute belief in my ability to achieve anything I put my mind to.

Second, I would like to thank Dr. Brian Cusack, my thesis supervisor, not only for his unwavering professional support throughout the 4 years of my post graduate studies but also for his personal guidance and mentoring over the past 2 years that has led to this moment; a personal milestone, a warm and fuzzy sense of pride, a time for celebration.

Abstract

Identifying and acquiring data stored in a cloud environment is a complicated and challenging process. Much of the current academic forensic literature focuses on conventional digital forensic principles and meticulous chain of custody processes. Conventional computer forensics focuses upon having physical access to the media that stores the data of potential interest. However, in a cloud computing environment it is often not possible or feasible to access the physical media. The client's data may be stored on virtual servers on physical devices located in numerous data farms across various geographical locations making jurisdictional access also problematic. This research paper identifies the key aspects of cloud computing and analyses the reliability and integrity of the evidence gathering process during a digital investigation in a cloud environment. Case studies are presented in support of the research designed to assess whether existing digital forensics techniques are applicable to cloud investigations. The research examines technical and trust concerns that practitioners and law enforcement agencies (LEA) encounter in acquiring forensic evidence from a cloud.

Research testing involved creating a simulated 'Infrastructure as a Service' (IaaS) cloud environment to evaluate the evidence gathering process between the cloud client and the Cloud Service Provider (CSP). The IaaS cloud environment was created in Microsoft Server 2012 Datacentre, Hyper-V. A Domain Controller was created in Active Directory and populated with user accounts and virtual machines (VMs); client VMs have Microsoft Windows 7 operating system installed. The primary aim of the research is to test the integrity and reliability of evidential data acquired during a digital forensic investigation in a cloud using existing forensic tools, methods and techniques. Research testing was conducted in a controlled home laboratory environment based on an exploratory approach. Microsoft Network Monitor 3.4, Hyper-V SnapShot and Forensic Tool Kit (FTK) were used to capture forensic data along with client and server side log files. Internet Explorer and Firefox were installed on a client-side VM and were used to extract user activity.

The research findings demonstrate that although it may be technically possible to extract forensic evidence from the 'cloud' the investigative process presents significant jurisdictional and chain of custody challenges in the identification and seizure of evidential data by practitioners and law enforcement agencies (LEA) in criminal investigations and by businesses in civil litigation cases. It is also important that the evidential data collected can withstand rigorous scrutiny in a court of law.

Table of Contents

List of Tables

Table of Figures

Chapter 1
Introduction

1.0 BACKGROUND

Digital Forensics is a process of using specified methodologies, techniques and tools to identify, extract and analyse data found in digital media that can be presented as reliable evidence in a court of law (Hogan, Liu, Sokol, & Tong, 2011). The process of extracting data may vary depending on the device or data type being processed. For example, obtaining and analysing data from a conventional computer hard-disk drive requires a different process then obtaining and analysing data across a live network, and different again for cloud base technologies that involve evidence segregation and distributed environments. Regardless of the process, specific forensic procedures must be meticulously followed in order to obtain and preserve viable digital evidence.

The introduction and growth of cloud technology has compelled a re-evaluation of conventional digital forensic investigation methods, techniques and investigative tools used by digital forensic investigators to address cloud security characteristics in an ever-changing and innovative digital landscape. Due to the remote nature of cloud data stores, onshore and/or offshore, traditionally trained digital forensic investigators are faced with technical and legal challenges where conventional methods do not apply. The purpose of this thesis is to analyse and determine the reliability and integrity of data collected during a cloud forensic investigation and to assess the suitability and effectiveness of the methods, tools and techniques used to gather the data. The thesis will consist of two Information Technology (IT) areas; cloud based technologies and digital forensic procedures with a focus on the processes and principles governing the ability to perform digital forensic investigations in a cloud. The creation of a simulated cloud environment with simulated client-side activity to provide empirical data in support of the thesis findings. Also, the vast amount of reviewed evaluative literature provides insight into current and future cloud technologies trends and cloud forensic investigative processes.

The National Institute of Standards and Technology (NIST) has defined cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access

to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Hogan, Liu, Sokol, & Tong, 2011). Authoritative international market research from a number of different sources forecast that the global market for cloud products and services will grow rapidly in the next few years. Since we are in a time of economic constraint, cloud computing has found fertile ground and is seeing substantial global investment. The global research firm International Data Corporation (IDC) forecasts a compound annual growth rate (CAGR) of 50 percent in private cloud services to the year 2016, growing to a total global market value of over US$26 billion (IDC, 2013). According to research conducted by Red Shift Research (Montalbano, 2011) the majority of growth is expected in the private sector rather than the public sector, 42 percent versus 23 percent respectively.

The U.S. federal government certainly has a guarded approach to adopting the cloud paradigm but most government respondents accept that a private, public or hybrid cloud computing environment will play a pivotal role in U.S. federal IT activity over the coming years. The New Zealand Government has also established its own 'Cloud Programme' to develop and deploy a series of all-of-government cloud services with the Department of Internal Affairs leading the way as part of the Government ICT Strategy and Action Plan to 2017 (ICT.govt.nz, 2014). However, moving ICT services from a conventional WAN (Wide Area Network) infrastructure to a 'cloud' poses many challenges; not least of which are data security, integrity and privacy. Since the highly-publicised breach of privacy by the Accident Compensation Corporation (ACC) in 2012 there has been a string of incidences where private information held by government departments has been leaked or made publicly available through administrative errors, such as the Ministry of Social Development, Earthquake Recovery Commission, Ministry of Education, Ministry of Health and more recently the Ministry of Work and Income.

Standards for cloud computing are evolving through private and public organisations like the International Organisation for Standardisation (ISO) and International Electro-Technical Commission (IEC) Joint Technical Committee 1, the European Network and Information Security Agency (ENISA), Organisation for the

Advancement of Structured Information Standards (OASIS), Open Grid Forum (OGF), Institute of Electrical and Electronics Engineers (IEEE), Cloud Security Alliance (CSA), Open Cloud Consortium (OCC) and the Storage Networking Industry Association (SNIA); these organisations and others develop working standards for different aspects of cloud technology. However, there are industry observers who attribute the delay in formalising cloud computing standards to the fragmented approach taken by these organisations (Ortiz, 2011).

It is generally accepted that cloud technology offers cloud subscribers technical and economic advantages, however, despite this potential; customers remain reluctant to move their business IT infrastructure completely to the cloud. One of the main concerns is 'cloud security' and the threat of the unknown. Unwittingly Cloud Service Providers (CSPs) encourage this perception by restricting access to what lies behind the virtual curtain. Security professionals will undoubtedly face complexities and challenges when it comes to addressing key security requirements for cloud computing. There is also a requirement for Enterprise IT Risk Management Framework to be applied in the context of the cloud along with numerous other considerations to be assessed, evaluated and deployed. Managing risk when the information resides out of the enterprises control can be problematic and it is imperative security 'Services Level Agreements' (SLAs) are well defined and agreed upon between the cloud subscriber and the CSP beforehand (Catteddu, 2011).

The complexity of cloud based services introduces a number of unknown parameters and CSPs are cautious about offering guarantees for compliance-ready services and the adoption of those services. CSPs promote a simple and cost effective way in delivering Information Communications and Technology (ICT) services irrespective of jurisdictional borders, this raises questions and challenges in examining compliance with legal frameworks (Guilloteau & Mauree, 2012). A key question to ask when security is breached is how do we access cloud services and capture all relevant data required to carry out a digital forensic investigation? The answer requires the consideration of several sub-aspects as a complete capture of all data related to the event under investigation is not possible. Some data will not be available, some data will be

suspect, and some data will be court ready and can fit into the traditional network forensics model. The challenge for cloud forensic investigators is to recognise the data set for each of the three categories i.e. not available, suspect and court ready. The expansion of data storage capacity in a cloud is also a disadvantage for a digital forensic investigation as it involves an increase in forensic data to analyse.

The general lack of specific tools and limited professional expertise in cloud forensics is of concern, a situation made more challenging when encryption, proliferation of endpoints, multi-jurisdiction and loss of data control are involved. There is a requirement for cloud organisations and cloud subscribers to establish a cloud forensic capability; otherwise, they are likely to face ongoing difficulties when carrying out a cloud forensic investigations i.e. criminal intrusions and major policy violations. Investigators will also face difficulties when collaborating with law enforcement in resource confiscation cases due to limited forensic knowledge and preparation (Thorpe, 2012).

1.1 MOTIVATION

Section 1.0 identified and briefly discussed the background to the chosen research area of cloud technology and the processes of a digital forensics in a cloud. In order to understand the reasoning for the chosen research areas, the motivations of the researcher will be presented and discussed ranging from the rapid growth of cloud technology and its impact on existing digital forensic methods, techniques and tools to the proficiency of the investigators and reliability and integrity of the evidence gathered.

Cloud based services can either be hosted or managed by the user organisation or by one or more third party CSPs. As a consequence the software and data provided to cloud subscribers may be physically stored across many different geographic locations making it difficult to determine the legal framework and procedures that apply to the evidence gathering process (Grispos, Storer, & Glisson, 2012). According to ICT industry leaders cloud technology is regarded as the future of networked computing, for example, the advent of social media services like Twitter, Facebook, and Flickr; services that utilise the cloud as a means of storing and sharing customer data. However, events

such as the Edward Snowden disclosures about the American National Security Agency (NSA) domestic electronic surveillance and the American Federal Communications Commission's (FCC) reconsideration of net neutrality rules, among others, have drawn public attention to the concerns and distrust in the increasing dependence on centralised computing. With the increasing amount of personal data that is stored, shared, and transported via cloud-based services, the need to understand and critically evaluate these interconnected systems has become acute (Sullivan, 2014).

Cloud technology is a relatively new paradigm and the current gatekeepers, CSPs, have yet to standardise procedures on how security breaches are investigated. Jurisdictional considerations and data ownership also influence the investigative process; this complex series of interconnections between CSPs, cloud subscribers and law enforcement agencies provides fertile ground for cybercriminals who look to exploit any opportunity to infiltrate and hack systems (Lillard, Garrison, C.A., & Steele, 2010). For example, criminals may abuse professional anonymous communications systems such as Tor and Anonymizer (anonymous proxy) which were originally designed to protect network users from identity theft and profiling.

Digital evidence is by nature volatile; it can be altered, damaged or deleted through careless handling or improper examination. The evidence can be easily copied and modified, and is difficult to maintain. Safeguards and processes need be in place to document, collect, preserve and analyse digital evidence. In the same way as Deoxyribonucleic Acid (DNA) or fingerprint evidence is concealed digital evidence is also concealed. The US National Institute of Justice (NIJ) published a process model to serve as a guide for first responders. The guide is intended for use by law enforcement and other responders, who have the responsibility for the protection of an electronic crime scene and for the recognition, collection and preservation of digital evidence (Ademu, Imafidon, & Preston, 2011). However, cloud computing introduces new and significant challenges on how evidence is obtained and analysed, and therefore impacts the way cloud-based crimes are prosecuted.

Digital forensic investigators are dependent on CSPs for acquiring cloud evidence and by applying cloud provenance to ensure the integrity of the chain of custody;

investigators will expect CSPs to provide chronological access history of the evidence, how it was obtained, analysed, and preserved. In early 2011, in what was reported as the first public case of a cloud related crime, Sony was the victim of an online data breach that took down the PlayStation Network. Bloomberg News reported that the intruder used Amazon's public cloud to commit the crime (Galante, Kharif, & Alpeyev, 2011). The report also stated that the FBI was investigating the crime, but neither Amazon nor the FBI would comment on whether a search warrant or subpoena had been served. No further information about the case has been made public.

ICT governance is also essential in establishing controls over increasingly complex and integrated systems, services and human resources. ICT compliance controls within a conventional infrastructure are simpler and more distinguishable than that of a cloud environment. Internal infrastructure and services are controlled by the organisation ensuring compliancy through governance i.e. roles and responsibilities are clearly defined; compliance controls are designed and implemented with management approval whilst audit of compliance status can be readily tracked and measured. However, the moment ICT services are migrated to a 'cloud' the organisation effectively loses control on how compliance is implemented and maintained; this is handed over to the CSP. As part of any compliance requirement a gap analysis ought to be undertaken to identify how regulatory, legislative and industry compliance can be designed and implemented from the start. It is imperative that any compliance requirements are validated and certified before migrating to the cloud.

Enterprises considering moving to a cloud environment should consult with digital forensic practitioners to ensure safeguards and processes are in place to combat and to investigate criminal intrusions and policy violations within the cloud. Certified digital forensic practitioners, Digital Forensics Certified Associate (DFCA) and Digital Forensics Certified Practitioner (DFCP), must also adopt and expand their professional capabilities so that enterprises confidently transition to a cloud environment knowing that the digital forensics industry has the capability, competency and associated standards to support a digital forensics investigation. Digital forensic investigators will continue to

rely on existing tools like Guidance EnCase or AccessData Forensic Toolkit (FTK) unless alternative tools and techniques are developed (Dykstra & Sherman, 2012).

In summary, the preceding discussion illustrates that gathering reliable and verifiable cloud based data to be used as evidence in a court of law is not without its unique challenges given the many characteristics of cloud computing. Motivations include the increasing growth of cloud computing services, the associated security, privacy, ownership and legal issues and the potential for increased intentional criminal cyber activity. Furthermore, like all forensic fields, investigative principles for conducting digital forensic investigations in a cloud are dependent on the proficiency and qualifications of investigators, the reliability of the tools and techniques used to gather evidence and the co-operation and skills of CSPs to provide trustworthy records. In conclusion, cloud digital forensic investigations can vary according to the cloud service and deployment model and the location of the evidence, unlike traditional computer forensics where investigators have full access and control over the evidence (e.g., router logs, process logs, and hard disks).

1.2 STRUCTURE OF THE THESIS

The thesis is delivered in a logical sequence that conveys the research carried out. The formal components of the thesis include an abstract, acknowledgements and a table of contents. Additionally, a list of figures and a list of tables are presented.

Chapter 1 provides an introduction to the research project. The thesis topic and associated background is presented including an outline of cloud computing technologies and an overview of the challenges associated with conducting a digital forensic investigation in a cloud. The motivations behind the project identify a need for the proposed research in the chosen area.

Chapter 2 provides an extensive review and discussion of available literature for the topic area in order to build a thorough understanding of the current state of knowledge. Cloud computing standards, cloud types and different service models provide an overview of the technology followed by detailed discussion on cloud security features, associated risks, legal, privacy and jurisdictional concerns and governance. The process

of digital forensics is presented with specific association to cloud investigation techniques and potential evidence resources including Intrusion Detection Systems (IDS). In closing, the problems and characteristics surrounding a cloud digital forensic investigation identifies specific challenges and considerations that become the focus of the research.

Research methodology for the project is critically evaluated in Chapter 3. First, several published similar studies are reviewed in order to be informed on previous research methodologies, as well as to highlight specific areas needed for further potential research. The research questions are then developed in Chapter 4 from the preceding literature discussed in Chapter 2 and the related similar studies in Chapter 3. Each question is also accompanied by a hypothesis; a proposed explanation made on the basis of theoretical information and the gathered knowledge. The research questions provide a goal for the thesis and establish the research requirements needed to determine a resolution for each of the proposed questions. Next, the research model is proposed which outlines four specific phases of research testing divided into Phase 1 and 2 for initial testing and Phases 3 and 4 for stabilised testing. The system architecture, the necessary components and the software and hardware requirements are also discussed to provide information regarding the proposed system design. The data requirements of the research model are then investigated, outlining the data generation, collection, analysis and reporting methodologies that are required for each of the testing phases. The expected outcomes of each phase of research testing are then outlined. The chapter concludes with a consideration of the limitations of the proposed research model establishing the scope of the testing to be conducted.

Chapter 5 reports the findings for each of the research testing phases. First, the variations to the previously proposed data requirements are identified and the subsequent modifications then applied to the proposed methods. The reported test findings are then divided into initial and stabilised testing, with the corresponding four separate phases of testing followed by the analysis of the data gathered. Summing up, the significant and analysed results from the research testing are finally presented in graphical form to visually display the attained findings. Chapter 5 is a discussion of the research findings.

To start with, the research questions developed earlier are revisited and arguments made for and against the associated hypotheses are tabled so that a synopsis of the learnt information and results achieved from the testing phases can be viewed. The research findings are then examined at length; each phase of testing is discussed, as well as an extensive evaluation of the system design developed and implemented for the research testing. Finally, recommendations are suggested based on the outcomes which were discovered during the conducted research.

Chapter 6 concludes the thesis and recommends further areas for study. A conclusion of the research project is presented, stating the most important findings that were achieved and discussing the capabilities of the proposed and tested system design. Limitations of the research are outlined and discussed to identify constraints in the research conducted and findings discovered. Finally, potential future research areas involving Cloud and performing digital forensic investigations complete the chapter. The appendices at the end of the thesis provides additional information regarding the findings; including a full set of results from testing, the hardware and software specifications of the devices used, various configuration files and other log files collected during testing.

Chapter 2

Literature Review

2.0 INTRODUCTION

The main research objective of this chapter is to review the current literature relevant to the study areas introduced in Chapter 1; namely cloud computing technologies and cloud digital forensic investigation processes. Although a definition of cloud computing is presented in Chapter 1 a more detailed description is required in order to fully understand and appreciate the special characteristics and complexities of cloud computing and associated concerns regarding cloud security, controls and governance. Cloud computing also presents unique challenges for digital forensic investigators where conventional methods, techniques and tools used may not be applicable or fit for purpose and therefore it is necessary to understand present-day practises and processes when acquiring, preserving, analysing and reporting cloud digital forensic evidence.

The literature review will not only serve as a fact-finding undertaking but will identify prospective problems and issues from which to derive potential research questions. Literature review resources included MFIT class notes and reference material, relevant online publications/blogs, online journals/magazines, online databases and traditional published reference material. Chapter 2 is structured into seven main sections. Sections 2.1 to 2.5 presents a review of cloud standards, various cloud types/models and their associated services, virtualisation, multi-tenancy, cloud vulnerabilities and security concerns. Issues pertaining to controls, governance, jurisdictional access, privacy and legal considerations are discussed in Sections 2.6 and finally Section 2.7 concludes with an overview of the process of gathering digital forensic evidence in a cloud.

2.1 STANDARDS

Anytime control is surrendered there is a measure of risk added to a situation. For ICT managers, balancing business levels of risk and opportunity is a significant task when

migrating part/all of your ICT services to a CSP. Industry standards, often a risk management safety net, are among the key mechanisms that help mitigate risks; for example, to promote the seamless and secure flow of data. However, cloud industry standards are lagging behind the rapid growth of cloud services especially in the areas of security, privacy and interoperability. Without standards, cloud subscribers will always be uncertain about the risks they are assuming. According to various ICT Industry observers the lack of cloud standards adversely affects the way cloud computing is managed and can delay cloud implementation. A more unified approach to cloud standards may allow for better transparency so that cloud subscribers can assess basic cloud service capabilities before moving part/all of its services to a cloud or switching from one CSP to another (Ortiz, 2011). Incompatibilities in the transition and adoption of cloud computing can be categorised as follows:

1. Technical (Security, Reliability, Scalability, Data Integrity, Performance etc.)
2. Business (Risk Management, Pricing, Expense, Governance, Maintenance, etc.)
3. Semantic (Vendor Lock-In, Portability, Interoperability, etc.)

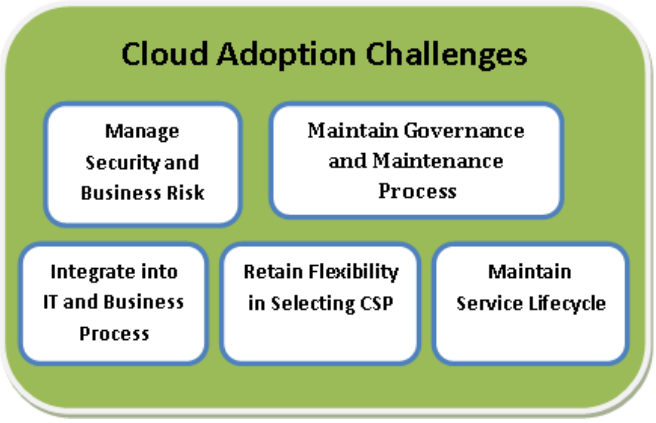

Figure 2.1: Cloud Computing Adoption Challenges (Adapted from International Journal of Cloud Computing and Services ScienceVol.2, No.5, October 2013, pp. 352 ISSN: 2089-3337)

Figure 2.1 identifies management adoption challenges that restrict CSPs and cloud subscribers from realising the full potential of cloud computing. There are also risk elements and adoption/management complexities to consider, such as:

1. How to integrate computer, network and storage services from one or more cloud service providers to cloud subscribers business and IT processes?
2. How to manage security and business continuity risk across many cloud service providers?
3. How to manage the lifecycle of a service in a distributed multi-provider environment in order to satisfy Service Level Agreement (SLA) with customers?
4. How to maintain effective governance and audit processes across integrated data-centres and CSPs?
5. How to adopt or switch to new CSPs?

According to IBM vice president, Angel Luis Diaz, the interoperability between offerings and the portability of services from one provider to another is very important to the subscriber as to maximise their expected return on investment from cloud computing. Moreover, interoperability would keep users from being locked into a single cloud provider (Savage, 2013). Nirlay Kundu, senior manager at Wipro Consulting Services, said in relation to the lack of cloud security standards, “addressing issues such as data privacy and encryption is also hurting wider cloud computing adoption”, and according to Lynda Stadtmueller, Program Director of the Cloud Computing Analysis Service within Stratecast (a division of Frost and Sullivan), "an effective lack of standardisation makes it difficult for buyers to compare and evaluate cloud offerings" (Ortiz, 2011).

The lack of cloud standards is not altogether surprising given that the technology is relatively new and "standards" are generally associated with more established technologies. Some experts go as far to say that due to the newness of the technology it is difficult for anyone organisation to mandate standards. According to Michael Crandell, CEO and founder of cloud computing vendor RightScale, "true interoperability requires the conversion of specific application and service functionality from one cloud to another and this won’t happen without standardisation. For example, there currently is no standardised way to seamlessly convert security requirements and policies across cloud

offerings" (Ortiz, 2011). Winston Bumpus, president of the Distributed Management Task Force (DMTF) from February 1997 to August 2013, an industry-based standards consortium said, "there are challenges to cloud-computing standardisation and to overcome them could determine just how bright cloud computing future will be". While some standards may become permanent, others may become redundant over a period of time (Ortiz, 2011).

A detailed and extensive expose into cloud computing standards, present and future, requires a discussion to take place between various vested industry organisations with many of today's work-in-progress standards, summarised in Table 2.1, based in part on the US National Institute of Standards and Technology's Special Publication 800-145, titled The NIST Definition of Cloud Computing (Draft).

Table 2.1 ***Comparisons - Cloud Computing Standards***

Organisation	Working Group	Standard	Purpose
Distributed Management Task Force (DTMF)		Open Virtualisation Format (OVF)	Establishes a transport mechanism for moving virtual machines from one hosted platform to another
Institute of Electrical and Electronics Engineers (IEEE)	P2301 P2302	P2301: Guide for Cloud Portability and Interoperability Profiles (CPIP) P2302: Standard for Inter-cloud Interoperability and Federation (SIIF)	CPIP: Meta-standard with profiles for existing and in-progress cloud computing standards in areas such as applications, portability, and management. SIIF: Establishes the characteristics necessary to create cloud interoperability and federation.
Open Grid Forum	Open Cloud Computing Interface	Open Cloud Computing Interface (OCCI)	Develop APIs for cloud management tasks. APIs enable interfacing between IaaS cloud implementations.

Organisation for the Advancement of Structured Information Standards (OASIS)	IDC Cloud Technical Committee Symptoms Automation Framework Technical Committee		ID Cloud focuses on security issues such as identity management and vulnerability mitigation. Symptoms Automation Framework establishes communications so that cloud providers understand consumer requirements.
Storage Networking Industry Association	Cloud Storage Initiative	Cloud Storage Initiative Cloud Data Management Interface (CDMI)	Provides standardisation for client interactions with cloud-based storage, cloud data management, and cloud-to-cloud storage interactions.

Notwithstanding the above-mentioned challenges, cloud computing is regarded as the technology of the future, offering high-speed connectivity and rapid deployment across a vast range of devices. However, in the absence of a common framework, the growth and evolution of cloud computing faces challenges similar to the evolution of the Internet in that it requires effort and resources to standardise the medium (Ortiz, 2011).

2.2 CLOUD MODELS

Cloud computing offers a variety of ways for businesses and organisations to increase their ICT capacity and/or functionality without having to add infrastructure, software, and personnel. However, there is no 'one-size-fits-all' cloud solution. There are different cloud models and services that customers can subscribe to depending on their needs. Each CSP provides specific functions that allow subscribers greater or less control over their cloud depending on the cloud type/service. The specific requirements for cloud subscribers' will vary depending on how they intend to use the space, resources and services associated with the cloud (Huth & Cebula, 2011). A cloud model is a way to organise computers so that resources can be quickly orchestrated, provisioned,

implemented and decommissioned, and scaled up or down to provide an on-demand service allocation. The following four deployment cloud models and three cloud services as defined by NIST (Hogan, Liu, Sokol, & Tong, 2011).

2.2.1 Private Cloud

Private Clouds are suited for a specific group or organisation and limits access to just that group. Private clouds comprise of hardware, networks, and software dedicated to a business unit linked to a physical location. Industries such as financial services and health care, government agencies and business units, who process highly confidential information while conforming to regulatory compliance, such as legal departments, human resources, and consumer services, may require higher levels of security afforded by the private cloud. The private cloud model does not benefit from the less hands-on management, nor from the economic advantages that make cloud computing an attractive concept i.e. the costs increase alongside the level of expertise needed (Hogan, Liu, Sokol, & Tong, 2011).

2.2.2 Public Cloud

In contrast to a Private Cloud, Public Clouds can be accessed by any subscriber with an internet connection who has access to the cloud infrastructure and computing resources. A public cloud is owned by the CSP who serve a diverse pool of customers. Examples of public clouds include Amazon Elastic Compute Cloud (EC2), IBM's Blue Cloud, Sun Cloud, Google AppEngine and Windows Azure Services Platform (Hogan, Liu, Sokol, & Tong, 2011).

2.2.3 Hybrid Cloud

A Hybrid Cloud is a combination of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. By utilising “hybrid cloud” architecture, companies and individuals are able to obtain degrees of fault tolerance combined with immediate usability without being entirely dependent on third party services. Hybrid Cloud architecture requires both on-premises resources and off-site (remote) server based cloud infrastructure. Although

hybrid clouds lack the flexibility, security and certainty of in-house applications they provide the flexibility of in-house applications with the fault tolerance and scalability of cloud based services. According to U.S. research company Gartner, the hybrid cloud model will overtake the private cloud model with 50% of large enterprises deploying hybrid cloud models by the end of 2017 (Rivera & van der Meulen, 2013).

2.2.4 Community Cloud

A Community Cloud expands the focus from the single organisation of a private cloud to multiple organisations. An analogy to describe this deployment model; an intranet is to a private cloud as an extranet is to a community cloud. Several organisations within a logical community share and support the cloud infrastructure. The community determines the mission, policy, compliance considerations, and security requirements. Either an external third party or organisations within the community may manage the community cloud. Also, the physical infrastructure of the community cloud may be located within the community (i.e. on-premise) or outside of the community (i.e. off-premise) (Carlton & Zhou, 2011).

2.3 CLOUD SERVICES

A cloud service is any resource that is provided over the Internet. The most common business cloud service resources are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS); sometimes referred to collectively as the SPI model (Mather, Kumaraswamy& Latif, 2009).

2.3.1 Software as a Service (SaaS)

SaaS allows users to run a variety of software applications over the Internet without having possession or management control over the applications (e.g. Salesforce.com, Gmail, and Microsoft Online). The customer is provided with the capability to use the CSP's applications running on a cloud infrastructure. The applications are accessible from client devices through either a thin client interface, such as a web browser (e.g. web-based email), or a program interface (Krutz & Vines, 2010). The customer has no management control over the cloud infrastructure including network, servers, operating

systems, storage, or individual application capabilities. Advantages of the SaaS model include for the customer (Krutz & Vines, 2010):

1. Reduced cost
2. Automatic updates and patch management
3. Compatibility and collaboration: All users will have the same version of software.
4. Global accessibility

Figure 2.2: Software as a Service (SaaS) (from Acclimate Technologies. Retrieved from http://acclimate.com/category/saas/)

2.3.2 Platform as a Service (PaaS)

PaaS model provides the customer with a computing platform that supports the development of their own web-based applications or SaaS applications (e.g. Google App Engine, Force.com and Windows Azure). However, the customer cannot manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage (Krutz & Vines, 2010).

2.3.3 Infrastructure as a Service (IaaS)

IaaS model centres on a delivery of service that provides a predefined, standardised infrastructure that caters specifically for the customer's applications. IaaS providers

manage the transition and hosting of selected applications on their infrastructure. The services provided to the customer include provision processing, storage, networks, and computing resources that allows the customer to deploy and run software; this can include operating systems and applications. However, the customer does not manage or control the underlying cloud infrastructure (Rittinghouse & Ransome, 2010).

2.4 VIRTUALISATION

In the context of network computing virtualisation refers to running multiple operating systems on a single device. While most computers only have one operating system installed, virtualisation software allows a computer to run several operating systems at the same time (Sammes, Antonopoulos& Gillam, 2010).The term virtualisation refers to the abstraction of computer resources (CPU, storage, network, memory, application stack, and database) from applications and end users consuming the service. The abstraction of infrastructure yields the notion of resource democratisation whether infrastructure, applications, or information and provides the capability for pooled resources to be made available and accessible to anyone or anything authorised to utilise them via standardised methods.

Servers that host applications and data on computer networks should deliver seamless, complex tasks with minimal effort using central processing units (CPUs) with multiple processors. Network administrators usually dedicate a server to a specific task as interoperability between many of these tasks is difficult. Although one task per server makes it easier to identify problems in real time and simpler to streamline the network, there are limitations. For example, this type of configuration under-utilises the full capabilities of the CPU processing power and as the computer network expands and becomes more complex, servers will require more physical space. As a consequence data centres can become crowded with multiple racks of servers that consume more power and generate more heat. Server virtualisation can potentially address these concerns simultaneously by using specially designed software to convert one physical device into multiple virtual servers. Each virtual server acts like a unique physical device, capable of running its own operating system (OS). It is possible, although not recommended, to

create enough virtual servers to utilise all of the server's processing power. Virtualisation also allows the portability of virtual servers between physical servers and can increase the overall security of the physical host server (Sammes, Antonopoulos& Gillam, 2010).

Cloud virtualisation is based on separating user applications from the underlying infrastructure. The host operating system provides an abstraction layer for executing a virtual guest operating system. A key aspect of virtualisation is the ‘hypervisor’ also referred to as a 'Virtual Machine Manager' or VMM. The hypervisor is a program that allows multiple operating systems to share a single hardware host. Each operating system appears to have the host's processor, memory, and other resources all to itself. However, the hypervisor is actually controlling the host processor and resources’, allocating what is needed to each operating system while making sure that the guest operating systems or virtual machines (VM) cannot disrupt each other.

Cloud-based systems use para-virtualisation as shown in Figure 2.4, which includes a binary bus between the various virtual machines and a hypervisor that exports a modified copy of the physical hardware. The exported layer has the same architecture as the server hardware with specific modifications that allow the guest OS to perform at near-native speeds. To take advantage of these modified calls small modifications have to be made to the guest OS. For example, modify the guest OS to utilise a hypercall to provide the same functionality expected from the physical hardware. By using the hypercall, the guest OS is significantly more efficient when running in a virtualised environment (Chantry, 2009). Hypercalls communicate directly with the hypervisor and are based on the same concept as a system call. System calls are used by an application to request services from the OS and provide the interface between the application or process and the OS. Hypercalls work the same way, except the hypervisor is used. The hypervisor also provides hypercall interfaces for other kernel operations including memory management and interrupt handling.

Many high performance computing (HPC) applications are only 15–20% efficient, and when running these applications on Cloud-based services it is possible to further increase their overall efficiency. The system will also need to schedule the VMs efficiently and the constituting parts of the application should be placed close together to

reduce communication latency and provide high inter-VM bandwidth. Cloud-based systems can also optimise the use of resources, reduce the amount of electrical power used and provide efficient Green IT computing possibilities (Sammes, 2010).Today, enterprises have deployed virtualisation technologies within data centres in various forms, including OS virtualisation (VMware, Xen), storage virtualisation (NAS, SAN), database virtualisation, and application or software virtualisation (Apache Tomcat, JBoss, Oracle App Server, WebSphere) (Mather, Kumaraswamy, & Latif, 2009).

2.4.1 Full Virtualisation Vs Para-Virtualisation

Full virtualisation is designed to provide total abstraction (completely decoupled) from the underlying hardware by the virtualisation layer. The guest OS is not aware it is being virtualised and requires no modification. Full virtualisation is the only option that requires no hardware or operating system assistance to virtualise sensitive and privileged instructions. The hypervisor translates all operating system instructions on the fly and caches the results for future use, while user level instructions run unmodified at native speed. Full virtualisation can streamline the migration of applications and workloads between different physical systems and helps provide complete isolation between different applications, which helps make this approach highly secure. Microsoft Virtual Server and VMware ESX Server software are examples of full virtualisation (VMware Inc., 2007).

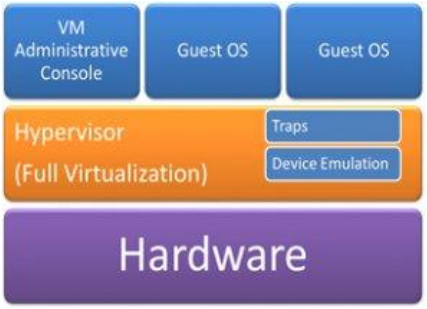

Figure 2.3:Full-Virtualisation (from Geeks Hub. Retrieved from http://www.geeks-hub.com/types-of-server-virtualization/full-virtualization/)

In contrast, para-virtualisation presents each VM with an abstraction of the hardware that is similar but not identical to the underlying physical hardware. Para-virtualisation techniques require modifications to the guest operating systems that are running on the VMs. As a result, the guest operating systems are aware that they are executing on a VM allowing for near-native performance. Para-virtualisation is also based on the hypervisor virtualisation model and eliminates much of the trapping-and-emulation overhead associated with software implemented virtualisation. It requires that the guest operating system be recompiled or modified before installation inside the virtual machine. Para-virtualisation is the primary model used by Xen, which uses a customised Linux kernel to support its administrative environment, known as domain0. Xen can also take advantage of hardware virtualisation to run unmodified versions of operating systems on top of its hypervisor (VMware Inc., 2007).

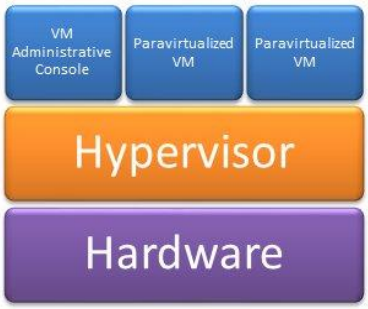

Figure 2.4: Para-Virtualisation (from Geeks Hub. Retrieved from http://www.geeks-hub.com/types-of-server-virtualization/)

2.4.2 When to Virtualise

According to a 2013 VMware sponsored survey by independent US Research Company Forrester Research, 70%of respondents transitioned to virtualisation when it was time to carry out a major hardware refresh as to avoid the cost of upgrading large numbers of physical hardware. Fifty-two percent virtualised their environment when it was time for a major operating system migration such as moving from Windows XP to Windows 7.

Fifty-one percent of respondents virtualised their servers when it was time for a major application license renewal such as Oracle or SAP, where significant savings can be made by consolidating servers. Unplanned system outage is another prompt for companies to move to virtualisation. According to US Software and Solutions Company CA Technologies comprehensive study of businesses in North America and Europe, unplanned outages and downtime was responsible for $26.5 billion in lost revenue.

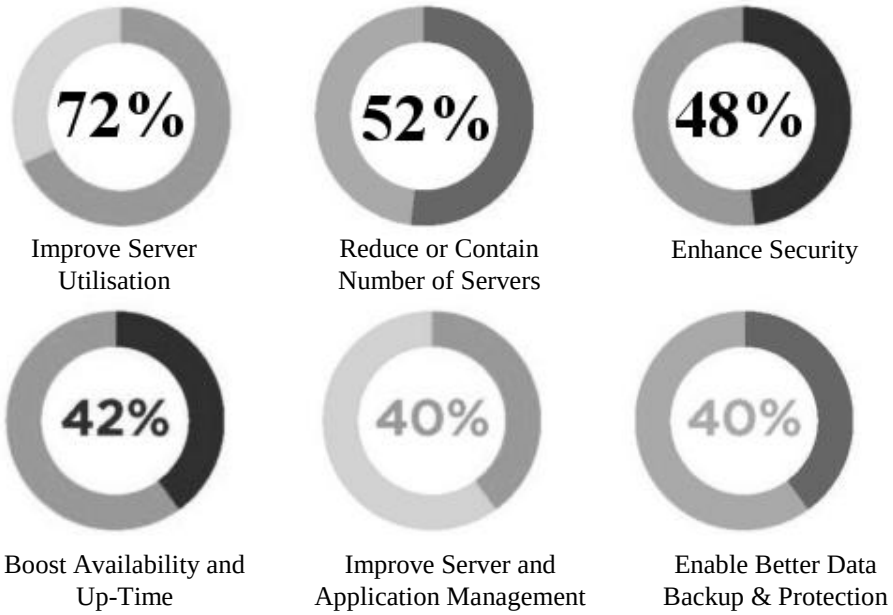

Figure 2.5: Traditional Benefits of Virtualisation (from Beyond Cost Savings: Retrieved from https://www.vmwaregrid.com/peoplelikeyou/cz/NDC/assets/

2.4.3 Virtualised Risks

The benefits of using virtualisation are not without its challenges and risks. Securing a virtualised environment is more complex than a traditional network with a greater need to manage and monitor shared resources, especially when multiple VMs are installed on a single physical device. Virtualised systems are dynamic and flexible making security boundaries difficult to define. If a VM was to be compromised other VM tenants may also be at risk and if not closely monitored can leave the entire virtualised environment vulnerable to cyber-attacks.

Data communication between physical devices on a conventional network and VMs running on a virtualised network share similar security concerns as they both use a shared medium to communicate. In a public cloud model computing resources are shared amongst VM tenants but individual VM clients have no control over the management of how these resources are being shared amongst other VM tenants, a significant security risk. Although some security remedies have been applied to strengthen the security of virtualised systems, there is no assurance of protecting VMs intercommunications on the same server (Szefer, Keller, Lee, & Rexford, 2011).

Widespread adoption of virtualisation necessitates greater production storage, which in turn exacerbates legacy approaches to disk-based protection. Not all backup technologies adequately protect VMs that integrate with APIs such as VMware, vStorage, VADP or Microsoft Volume Shadow Copy Services (VSS). Challenges related to protecting virtualised environments is more challenging in a private cloud architecture, where self-service portals and elastic load monitoring create new virtualised resources dynamically without any IT interaction with much less awareness or automation of backup processes (Buffington, 2014). Virtualisation conceptualises data locality, meaning cloud users cannot identify the exact physical location of their data as VMs can be moved from one machine to another autonomously by the underlying layers. Furthermore, data leakage by exploiting VM or hypervisor vulnerabilities is also a significant risk to virtualisation (Nepal & Pathan, 2014).

VM techniques, such as Xen, VMware and Hyper-V, offer on-demand virtualised IT infrastructures. VM instances use the shared resources on a physical server to deliver business needs. While they are working on the same physical machine and using the shared resources, security threats will become a general problem for all of the VMs. Some threats are common towards all computerised systems; for example, Denial-of-Service but other types of threats are VM specific. The following examples are of specific VM attacks that are inherently destructive for VMs in a virtualised server (Uden, Herrera, Pérez, & Rodríguez, 2012/2013):

1. Shared clipboard attack: as memory is shared among VMs, the attacks through shared clipboard are done by moving clipboard information between malicious programs in VMs of different security realms.
2. Keystroke logging attack: a number of VM technologies provide keystroke logging function and capture screen updates to be transferred across virtual terminals in the VM.
3. Monitoring VMs from an infected host: as all network packets are transferred via a host, certain functions may be compromised:
    a. Full control of VMs such as start, stops, pause, and restart.
    b. Full control and monitoring of resources available to VMs, including CPU, memory, storage, and network usage.
    c. Manipulate shared resources such as adjusting the number of CPUs, memory size, number of virtual disks, and number of virtual network interfaces available.
    d. Full access to monitoring applications running inside the VMs.
    e. Manipulate data stored on VMs virtual disks.
    f. Monitor VMs from another VM. VMs do not have direct access to one another's virtual disks on the host. However, if the VM technology uses a virtual hub or switch to connect the VMs to the host, intruders can use hacking techniques such as Address Resolution Protocol (ARP) poisoning and network packets redirection to redirect packets going to or from other VMs.
    g. VM backdoors, where communication channels are opened between the guests and hosts that can allow intruders to potentially perform malicious operations.

2.4.4 Digital Evidence in a Virtualised Environment

In traditional digital forensics investigations, evidence is distributed across a number of devices such as hard drives, network servers and mobile devices; the analysis of these devices allows the investigator to retrieve information regarding the suspect's activities. An advantage of cloud computing is that multiple applications and servers across

geopolitical locations are able to interact seamlessly to provide the services and applications that a user requires. Cloud computing is independent of hardware and operating system profiles. In addition, the end-user experience requires data to be processed by multiple applications or computers but delivered as if it originates from a single source computer making identification and gathering of evidence in cloud computing more complex. As a consequence, virtualisation software has the potential to render the collection of digital information forensically unsound. Cloud data resides in a virtual instance and closing down the instance for the purpose of a forensic investigation may force other virtual live instances to shut down. Virtualisation by design assumes the hardware management duties of the OS, add to this the growing number of web based applications that assume application management duties from the OS, it is conceivable that over time a disposable OS will be created using a combination of hypervisor functions and Web applications that operate for a single session and completely dismantle when shut down (Barrett & Kipper, 2010).

Accessing software application via a cloud computing system typically writes data logs like registry entries or temporary Internet files to the OS that reside or are stored within the virtual environment but disappear when the user exits the cloud. Virtualisation sanitises resources and the traditional analysis of leftover artefacts could be limited or compromised; this can make cloud virtual digital evidence stored on hard drives unrecoverable. (Taylor, Haggerty, Gresty, & Lamb, 2011). If remote access to a guest VM OS is available a forensic investigator can obtain evidence using forensic tools to capture live data or suspend/terminate the VM and analyse the data offline. Acquisition at this layer requires trust that the guest OS, hypervisor, host OS, underlying hardware, and network connectivity can provide complete and accurate evidence that is free from intentional and accidental tampering or error (Dykstra & Sherman, 2012).

Notwithstanding the above scenarios the amount of potential evidence available to investigators can deviate substantially between the different cloud service and deployment models. However, independent of the model, the following three components could be a source for potential evidential data (Birk, 2011).

2.4.4.1 Virtual Cloud Instance

Typically, this a potential starting point for the investigator. The virtual instance can be accessed by the CSP and the cloud subscriber who is running the instance. Snapshot is a powerful technique used to freeze specific states of the VM while virtual instances are either still running or tuned off i.e. live investigation or static image analysis. However, in a SaaS and PaaS scenario, the ability to access the virtual instance for the purpose of gathering evidential information is limited or simply not possible (Birk, 2011).

2.4.4.2 Network Layer

The different ISO/OSI network layers provide protocol information and communication information between instances within and outside of the cloud. Currently, CSPs do not provide any log data from the network components which means that in a malware infection of an IaaS VM, it would be difficult to obtain routing information; this situation is more complicated in PaaS or SaaS. Hence, the situation of forensic evidence is again strongly affected by the level of support the investigator receives from the cloud subscriber and the CSP (Birk, 2011).

2.4.4.3 Client System

If and where potential evidence could be extracted from the system layer of the client depends on the cloud service model (IaaS, SaaS or PaaS). In most cloud scenarios the browser on the client system is the only application that communicates with the service in the cloud; this is particularly true for SaaS and where an exhaustive forensic investigation of the browser environment is essential (Birk, 2011).

2.5 MULTI-TENANCY

Cloud computing generally includes the principle of multi-tenancy i.e. the ability to use the same software and interface to configure resources and isolate customer-specific traffic and data. In a typical multi-tenancy environment, multiple users will share the same hardware and software applications but do not share or see each other's data while

running on the same operating system. An analogy would be, if you were to provide housing for a number of tenants you could provide either a separate house for each tenant to live in, or provide tenants with individual units within one apartment building. The former would include higher costs, inefficient utilisation of resources, and maintenance would be more complex. The latter would be cheaper, resources like space or air conditioning would be better utilised, and maintenance would be easier to manage and more cost effective. These same principles apply to the use of software. When companies install software on individual employees' PC's or on a dedicated server, it's like providing individual tenants with entire houses; this model requires significant investment, servers and other resources are underutilised, and maintenance is more complicated because servers or installations must be upgraded individually. This approach is not ideal, especially for smaller companies. Alternatively, CSPs can configure multiple users to share a database server and/or applications which will decrease costs, improve utilisation of resources, and streamline maintenance.

Multi-tenancy is not an alternative to virtualisation. A major advantage of multi-tenancy is that all SaaS application users subscribe to a single code-base, and therefore all tenants will benefit equally from any new innovations. Applying updates for a single-tenancy model will only benefit a single tenant. Virtualisation does not change this limitation.

2.6 CLOUD SECURITY

According to a 2013 Cloud Security Alliance (CSI) report, "Cloud computing has simultaneously transformed business and government, and created new security challenges" (Cloud Security, 2013). The report goes onto to say:

> The development of the cloud service model delivers business-supporting technology more efficiently than ever before. The shift from server to service-based thinking is transforming the way technology departments think about, design, and deliver computing technology and applications. Yet these advances have created new security vulnerabilities, including security issues whose full impact is still emerging. (pp. 6).

ICT Security authors, Ronald Krutz and Russell Vines wrote "Security is a principal concern when entrusting an organization's critical information to geographically dispersed cloud platforms not under the direct control of that organization" (Krutz & Vines, 2010, p. 62). Although virtualisation brings its own measure of new challenges, existing issues common in software engineering are transferred to clouds exposing vulnerabilities in APIs, IDEs, and web technologies such as, bad programming approaches in deploying cloud applications or common Cross-Site Scripting (XSS). Subsequently, each cloud service model presents its own challenges that raise concerns about the cloud business model (Krutz & Vines, 2010).

In a recent global survey conducted by Trend Micro(Trend Micro, 2011), IT decision makers indicated which virtualisation and cloud technologies they have deployed or are currently piloting. Worldwide, over half of the companies surveyed have implemented some form of server virtualisation and virtual desktop infrastructure (VDI). Also of those surveyed, 45% are using a public cloud, and 46% are using a private cloud (Table 2.2).

Table 2.2 ***Virtualisation and Cloud Computing Implementation***

% Deployed or Piloting	Total	US	Japan	India	Germany	UK	Canada
Server Virtualisation	59%	70%	58%	51%	61%	68%	47%
VDI	52%	62%	42%	48%	55%	63%	45%
Public Cloud	45%	54%	37%	38%	48%	52%	42%
Private Cloud	46%	56%	34%	42%	54%	51%	42%

2.7 CONCLUSION

The literature review conducted in Chapter 2 provides an overview of the current state of knowledge and of the context of the thesis. It first identifies the fragmentation of industry standards and how cloud standards are lagging behind the rapid growth of cloud services especially in the areas of security, privacy and interoperability. The various cloud models and services on offer are also reviewed, which surmised that there is no

‘one-size-fits-all’ cloud solution. There are different cloud models and services that customers can subscribe to depending on their needs and each CSP provides specific functions that allow subscribers greater or less control over their cloud depending on the cloud type/service. The literature review provides an insight into conventional digital forensic investigation where evidence is distributed across a number of devices such as hard drives, network servers and mobile devices; the analysis of these devices allows the investigator to retrieve information regarding the suspect’s activities. However, cloud virtualisation has the potential to render the collection of digital information forensically unsound. Cloud computing also generally includes the principle of multi-tenancy i.e. the ability to use the same software and interface to configure resources and isolate customer-specific traffic and data. The literature reviewed multi-tenancy, where multiple users share the same hardware and software applications but do not share or see each other’s data while running on the same operating system; multi-tenancy provides significant advantages for the SaaS (Software as a Service) cloud model.

The literature reviewed/discussed and introduced specific areas and identified that there is a need for further research. In particular, security, privacy and legal issues of the investigative process. The current state of knowledge also identified crucial factors that will assist in the design perspectives and developments of a feasible research methodology. It has therefore, been determined that the proposed research will focus on advancing the body of knowledge surrounding digital forensics in the cloud. Specifically, the research will aim to acquire evidential data from the cloud and cloud client by creating a simulated IaaS (Infrastructure as a Service) cloud environment.

Chapter 3 will therefore undertake a review of similar studies relevant to the chosen area of research and together with the literature knowledge, the main research question and associated sub-questions will be derived and the methodology, relevant questions and hypotheses will also be developed.

Chapter 3

Research Methodology

3.0 INTRODUCTION

The primary research objective of Chapter 3 is to formulate a research question and to develop an appropriate methodology and framework for the proposed research. The distributed nature of data in 'cloud' technology, the architectural functionality of virtualisation and restricted physical access to server side digital artefacts are real challenges for investigators and stakeholders where conventional approaches to evidence collection and recovery may not apply and therefore may not stand up to scrutiny in a court of law. In this chapter case studies are presented on 'digital forensics in a cloud' and assess whether existing conventional digital forensics tools and techniques are applicable to cloud forensic investigations and evaluate the reliability and integrity of the data collected. In order to learn from similar studies and experiences conducted by researchers working within the same field, I have included 3 studies that are evaluated in Section 3.1, 3.1.2 and 3.1.3. In conjunction with readings from Chapter 2, these studies are pivotal in forming the research question and hypothesis to be tested. Chapter 4 will outline the main research question and secondary questions with associated hypotheses based on all information gathered.

3.1 REVIEW OF SIMILAR STUDIES

In order to develop the methodology for this research, 3 independent research studies have been sourced and reviewed. There are also a number of references identified in Chapter 2 to facilitate the process of forensic investigation in a cloud environment. The following research studies have been selected for relevance and similarity to the chosen research area including the methodology used and relevant information pertaining to cloud digital forensic investigations.

The first study by Marturana, Tacconi, & Me (2012), shows that it is possible to find evidentiary material about the User/CSP interaction by searching local artefacts. The second study by Martini & Choo (2014), outlines the area of forensics in a cloud distributed file system which includes an in-depth forensic experiment on XtreemFS, a Contrail EU-funded project that covers technical and process issues. The third study by Martini & Choo (2013), digital forensic experiments with the aim of providing forensic researchers and practitioners with an in-depth understanding of the artefacts required to undertake cloud storage forensics (StaaS).

3.2 CASE STUDY 1 - USER AND CSP EVIDENTIARY MATERIAL

Researchers Marturana, Tacconi and Me (2012, p. 111) conducted a case study into digital forensics in cloud computing intended at assessing whether existing digital forensic techniques are applicable for the 'cloud'. The case study was designed according to cloud working principles to show when searching local artefacts, that it is possible to find evidentiary material about the user and the CSP interaction. In this regard, the researchers Marturana, Tacconi, & Me (2012, p.112) selected and analysed document editing and photo sharing SaaS applications, such as Google Documents, Flickr and PicasaWeb to demonstrate that potential evidence may be found in logs and temporary files, internet cache, navigation history, downloads and Web browser cookies. SaaS application Dropbox, together with locally installed storage software were also analysed; a copy of the server data is stored in a synchronised local folder. While connected to the 'cloud' Dropbox checks for file updates and changes made to the local copy to ensure that it reflects the current state of the server data and vice-versa. A copy of data as it exists in the cloud is acquired by simply retrieving data or fragments from local hard drives without the need to access the server directly.

In this regard, the researchers Marturana, Tacconi, & Me (2012, p.113) engineered four test scenarios, whereby a connection to a cloud service and the creation of a user account was established. In each scenario, the researches performed the tests listed in Table 3.1, each test labelled with a unique sequence number and a description of the performed action:

Table 3.2.1 ***User/CSP Interaction Test Scenarios***

Scenario 1 - Dropbox accessed via Web browser:	
Test 1.1:	log onto www.Dropbox.com
Test 1.2:	upload a word document
Test 1.3:	open or download a word document
Test 1.4:	delete a word document
Scenario 2 - Google Documents accessed via Web browser:	
Test 2.1:	log onto docs.google.com
Test 2.2:	create a word document
Test 2.3:	upload a word document
Test 2.4:	open a word document
Test 2.5:	delete a word document
Scenario 3 - PicasaWeb accessed via Web browser:	
Test 3.1:	log onto picasaweb.google.com
Test 3.2:	upload an image file
Test 3.3:	open an image file
Test 3.4:	delete an image file
Scenario 4 - Flickr accessed via Web browser:	
Test 4.1:	log onto flickr.com
Test 4.2:	upload an image file
Test 4.3:	open an image file
Test 4.4:	delete an image file
Scenario 5 - Dropbox client installation with local synched folder:	
Test 5.1:	install Dropbox client software on the local hard drive
Test 5.2:	save a file in the Dropbox local folder
Test 5.3:	open a file in the Dropbox local folder
Test 5.4:	delete a file in the Dropbox local folder

In scenarios 1 to 4, cloud services were tested against the three most popular Web browsers (i.e. MS Internet Explorer, Mozilla Firefox and Google Chrome) on the client side. The browser activity was recorded and analysed (i.e. cache, cookies, navigation history and downloads), network traffic was also captured to recover data fragments of the interaction between the local device and the cloud.

Finally, in Scenario 5, Dropbox services were tested against a set of traditional forensic tools to verify its existence in the list of installed applications or in the list of processes and recover evidentiary material in the local file system through analysing the file access timeline, list of deleted or recently accessed files (Marturana, Tacconi, & Me, 2012).

The researchers Marturana, Tacconi, & Me (2012, p. 113) performed all test twice, the first using live forensics tools on a powered on laptop computer running Windows 7 Home Edition 64 bit and the second with post mortem forensics tools on a physical image of its hard disk.

In Scenarios 1 to 4, cloud services were tested using Web browser versions MS Internet Explorer 8.0.7601.17514, Mozilla Firefox 11.0 and Google Chrome 18.0.1025.168, in association with a number of openly available Nirsoft live forensics tools on a powered-on system. As a cross-check, a post-mortem forensic tool, Internet Evidence Finder v4.0 from JAD software, was used on the physical image of the local hard disk.

Scenario 5 was first analysed using the powered on system using the following openly available Nirsoft live forensic tools:

1. WhatInStartup v1.33
2. RegScanner v1.85
3. CurrProcess v1.13
4. WinPrefetchView v1.10
5. RecentFilesView v1.15
6. SearchMyFile v1.82

The physical image of the local hard drive was searched using Sleuthkit, Autopsy and Log2Timeline.

3.2.1 Summary of Results

The following is a summary of results of the tests carried out by researchers Marturana, Tacconi, & Me (2012). Local folders and Web browsers databases were assigned nicknames by the researchers, for example, the nickname for Cookies is referred to as IE_cookies, History is referred to as IE_history and Cache is referred to as IE_cache. The

use of SmartSniff, a network monitoring utility was also utilised to capture TCP/IP packets that pass through a network adapter, and then viewed as a sequence of conversations between clients and servers. A summary of the results of the 5 test scenarios conducted by researchers Marturana, Tacconi, & Me (2012) are listed in Table 3.2:

Table 3.2.2*User/CSP Interaction Test Scenario Results*

Scenario 1 - Dropbox accessed via Web browser: SmartSniff utility was not always required as the Dropbox server provided a secure HTTPS connection, encrypted via SSL on TCP port 443.		
Test	**Action**	**Result**
Test 1.1	Log onto www.Dropbox.com	Found cookies from www.dropbox.com inIE_cookies, MF_cookies and GC_cookies and traces of the login phase in some HTTP and HTTPS URL in IE_history,MF_history and GC_history, attesting that Dropbox login page was accessed at list once.
Test 1.2	Upload a word document	On uploading a word document on Dropbox server, the dropbox.com/upload URL was saved in IE_history whereas no traces of the file upload were found in MF_history and GC_history.
Test 1.3	Open or download a word document	On opening or downloading a word document, four HTTPS URL reporting the actual filename in URL title were saved in IE_history and a copy of the file was stored in IE_cache; Two HTTPS URL reporting the actual filename in URL title were saved in MF_history and a copy of the file was stored in the \Users\...\AppData\Local\Temp folder; No URL were saved in GC_history whereas a copy of the file was stored in GC_cache.
Test 1.4	Delete a word document	On deleting a word document from Dropbox, no traces of the user-CSP interaction were found locally.
Scenario 2 - Google Documents accessed via Web browser:: SmartSniff utility was not always required as the Google Documents server provided a secure HTTPS connection, encrypted via SSL on TCP port 443		
Test	**Action**	**Result**

Test 2.1	Log onto docs.google.com	Found cookies from account.google.com, and google.com in IE_cookies, MF_cookies and GC_cookies and traces of the login phase in some HTTP and HTTPS URL in IE_history, MF_history and GC_history, attesting that Google Documents logon page was accessed at list once.
Test 2.2	Create a word document	On creating a word doc in Google Documents, an HTTPS URL reporting the actual filename in URL title were saved in MF_history and GC_history whereas Internet Explorer leaved no traces. In IE_cache, MF_cache and GC_cache, we found icons, generic files and JavaScript files used by the browser to interact with the server.
Test 2.3	Upload a word document	On uploading a word document on Google Documents, no traces of the user-CSP interaction were found locally.
Test 2.4	Open a word document	On opening a word document, a URL was saved in MF_history and GC_history, whose title reported the complete name (with extension) of the opened file. No traces were found in IE_history, IE_cache, MF_cache and GC_cache.
Test 2.5	Delete a word document	On deleting a word document on Google Documents, no traces of the user-CSP interaction were found locally.
Scenario 3 - PicasaWeb accessed via Web browser:: SmartSniff utility was not always required as the Google PicasaWeb server provided a secure HTTPS connection, encrypted via SSL on TCP port 443		
Test	**Action**	**Result**
Test 3.1	Log onto picasaweb.google.com	Found cookies from account.google.com and google.com in IE_cookies, MF_cookies and GC_cookies and traces of the login phase in some HTTP and HTTPS URL in IE_history and MF_history, attesting that Google PicasaWeb login page was accessed at list once. With regards to Google Chrome, in particular, two HTTPS URL reporting the actual username in the title were saved in GC_history, attesting that Google PicasaWeb user account was accessed at list once. For each photo album that was created on the server, a cover image was saved in IE_cache, MF_cache and GC_cache.
Test 3.2	Upload an image file	On uploading an image file on Google PicasaWeb, we found an HTTPS URL in IE_history, and MF_history, attesting

		that Google PicasaWeb upload page was accessed at list once. With regards to Google Chrome, in particular, a URL reporting the actual username, as an HTTPS parameter, was saved in GC_history. Finally, for each image that was uploaded on Google PicasaWeb, a correspondent image file was saved in IE_cache, MF_cache and GC_cache.
Test 3.3	Open an image file	On opening an image file from Google PicasaWeb, the research found HTTPS URL in IE_history, MF_history and GC_history, whose title reported the name of the album the photo belongs to. With regards to Google Chrome, in particular, it was possible to find the image in its original dimension in GC_cache.
Test 3.4	Delete an image file	On deleting an image file on Google PicasaWeb, no traces were found in IE_history and MF_history. With regards to Google Chrome, in particular, an HTTPS URL, whose title reported the name of the album the deleted photo belonged to, were saved in GC_history.
Scenario 4 - Flickr accessed via Web browser: With the exception of the authentication stage in which the Flickr server provided a secure HTTPS connection, encrypted via SSL on TCP port 443, the use of SmartSniff to eavesdrop on the connection was regularly utilised as the web connection between the user and cloud server was in the clear.		
Test	**Action**	**Result**
Test 4.1	Log onto flickr.com	Found cookies from flickr.com and yahoo.com in IE_cookies, MF_cookies and GC_cookies, a yahoo account was used to authenticate to Flickr; Traces of the login phase were found in some HTTP URL in IE_history, MF_history and GC_history, attesting that Flickr login page was accessed at list once. Personal images displayed in the home page after the authentication was saved in IE_cache, MF_cache and GC_cache.
Test 4.2	Upload an image file	On uploading an image file on Flickr server, the flickr.com/photos/upload URL was saved in IE_history, MF_history and GC_history whereas a copy the of the album web page the photo in IE_cache andMF_cache were found.
Test 4.3	Open an image file	On opening an image, aURL was saved in IE_history,

		MF_history and GC_history, whose title reported the name (without extension) of the opened file? In IE_cache and GC_cache it was a copy of the opened file whereas, in MF_cache, a copy of a web page pointing to the opened file was saved.
Test 4.4	Delete an image file	On deleting an image file, a copy was stored in IE_cache, MF_cache and GC_cache; With regards to Internet Explorer, two URL were saved in IE_history, whose title reported the partial name (without extension) of the deleted file.
Scenario 5 - Dropbox client installation with local synched folder:		
Test	**Action**	**Result**
Test 5.1	Install Dropbox client software on the local hard drive	Performed both a live analysis of the powered on laptop computer to check for: 1. The presence of Dropbox folders synched with the server, 2. Registry keys attesting installation of the Dropbox client, 3. Dropbox software in the list of installed applications, 4. The presence of Dropbox synchronization process in the list of running processes, 5. A dropbox.pf file in Windows prefetch directory, 6. Attesting that Dropbox was executed at list once, 7. Files recently accessed and Dropbox synchronisation logs, A post mortem analysis of the physical image of the local hard disk to check for: 1. The presence of Dropbox folders synched with the 2. Server and related files, 3. Dropbox synchronisation logs, 4. The timeline of recently opened, modified and deleted file by Dropbox.
Test 5.2	Save a file in the Dropbox local folder	Performed a post mortem analysis of the physical image of the local hard disk to check for: 1. The Dropbox synchronisation logs,

		2. The list of files recently opened, modified and deleted by Dropbox and related timeline.
Test 5.3	Open a file in the Dropbox local folder	Performed a post mortem analysis of the physical image of the local hard disk to check for: 1. The Dropbox synchronisation logs, 2. The list of files recently opened, modified and deleted by Dropbox and related timeline
Test 5.4	Delete a file in the Dropbox local folder	Performed a post mortem analysis of the physical image of the local hard disk to check for: 1. The Dropbox synchronisation logs, 2. The list of files recently opened, modified and deleted by Dropbox and related timeline

All tests scenario's conducted by researchers Marturana, Tacconi, & Me (2012) were successful as it was possible to reconstruct all user activities by performing both live and post mortem analysis.

The case study presented by Marturana, Tacconi & Me (2012) is a viewpoint about the emerging challenges of cloud computing to digital forensics and related countermeasures. The research question addressed by Marturana, Tacconi & Me (2012), "is it possible to analyse cloud environments with traditional digital forensics procedures and how existing techniques, tools and methodologies would cope in cloud a scenario". The researchers subsequently developed a practical case study by conducting a forensic investigation on a cloud environment in which some popular SaaS applications were analysed to demonstrate that, upon sharing files, photos and document in the cloud, evidentiary material may be found in logs and temporary files, saved locally by Web browsers. The researchers Marturana, Tacconi & Me (2012) therefore inspected local folders and Web browsers databases with traditional live and post mortem forensic methods and tools to cross-check the retrieval of potential evidence between the user and CSP.

Marturana, Tacconi & Me (2012) also analysed Dropbox, a popular file sharing SaaS application which may work both as a Web based cloud application or a traditional, locally installed software which stores a copy of the server data in a synched local folder.

The aim was to verify that it was possible to acquire a forensic copy of data as it exists in the cloud by simply retrieving data or data fragments from local hard drives. The outcome of the researchers case study outlined above showed that evidentiary material of the user to cloud interaction from local artefacts can be collected or eavesdrop on network connections without the need to access the cloud server directly. However, the user can delete navigation data upon quitting the browser, which could be considered an anti-cloud forensics technique, the adoption of forensic techniques to retrieve deleted files and timeline creation to overcome or mitigate this problem would be required.

3.3 CASE STUDY 2 - XTREEMFS DISTRIBUTE FILESYSTEM

Researchers Martini and Choo (2014) conducted a case study involving the technical and process issues when collecting evidential data from distributed filesystems in cloud computing environments. Distributed filesystems provide a cost-effective client/server-based application that allows clients to access, process and share data stored on a server as if it were on their own computer. Unfortunately, this technology has the potential to be exploited for illegal purposes. There are various traditional digital forensic practises and methods suited for different cloud computing platforms and deployment models (Martini and Choo, 2014). For example IaaS may provide an export of the virtual hard disk and memory provided to the user while SaaS may only provide a binary export of the data stored on the hosted software environment. It is, therefore, important for the LEA (law enforcement agency) collecting the evidence in one jurisdiction for the use in a criminal prosecution taking place in another jurisdiction to work and cooperate closely with their foreign counterparts to ensure that the methods used in the evidence collection are in full accordance with applicable laws, legal principles and rules of evidence of the jurisdiction in which the evidence is ultimately to be used (United Nations, 2007).

Unsurprisingly, industry professionals have advocated guidelines that focus on cloud digital forensics, a suitable framework for research experiments (IEEE, 2011); (Hogan, Liu, Sokol, & Tong, 2011); (Zatyko & Bay, 2011). In the interim researchers Martini and Choo (2014) used their own previously published cloud forensic framework to conduct an in-depth forensic experiment on XtreemFS, a Contrail EU-funded

project(Contrail Consortium, 2014). The framework is based upon the stages outlined by McKemmish (McKemmish, 1999) and NIST (Hogan, Liu, Sokol, & Tong, 2011) but differs in a number of significant ways. One of the key features of the researcher's framework is that through the client the existence of cloud storage is identified whereby data that is synced/cached on the client is recovered. As such, forensic analysis of the client is carried out before analysis of the server environment.

XtreemFS is an open source example of a general purpose and fault-tolerant distributed and replicated filesystem that can be deployed for cloud and grid infrastructures to support big data initiatives. Researchers Martini and Choo (2014) in this case study examine the technical and process concerns regarding the collection of evidential data from distributed filesystems which are commonly used in cloud computing environments.

Distributed filesystems can potentially support data fragmentation and distribution in cloud computing across the globe within numerous data centres; this presents significant technical and jurisdictional challenges in the identification and seizure of evidential data by law enforcement and national security agencies in criminal investigations as well as by businesses in civil litigation matters (Hooper, Martini, & Choo, 2013). Martini and Choo, in this case study, chose to focus on a single distributed filesystem as this allowed the researchers to conduct an in-depth analysis of the client and server(s) as to understand the potential to collect evidential data as part of a forensic investigation. XtreemFS has received significant attention in the academic community with many researchers choosing to analyse it or implement it as the underlying infrastructure in larger projects. Most commonly, XtreemFS is implemented in cloud computing or grid computing which is considered to be the predecessor to cloud computing (Kielmann, Pierre, & Morin, 2010).

The client can be used to identify the existence of cloud services and to collect data stored by the client. Therefore, forensic analysis of the client is generally carried out before analysis of the server environment. According to researchers Martini and Choo (2014) the following four stages outline the high level processes that a forensic

practitioner should follow when conducting a conventional and DFS (Distributed filesystem) forensic investigation in a cloud computing environment:

1. **Evidence Source Identification and Preservation:** Concerned with identifying sources of evidence in a digital forensics investigation. Initially, sources of evidence such as desktop/laptop computers and mobile devices, will generally be in possession of the suspect. However, in the case of a distributed filesystem used in cloud computing, the filesystem client may only exist on the cloud server devices. Preservation is essential to the integrity of forensic investigations and as such proper preservation techniques must be maintained regardless of the evidence source.
2. **Collection:** The actual capture of the data. There are various methods of evidential data collection suited for the various cloud computing platforms and deployment models. While IaaS may result in the collection of virtual disks and memory, and SaaS may result in an export from the relevant cloud software, the collection of distributed filesystems supporting cloud computing installations is likely to be considerably more involved. Also, if the filesystem is hosted outside of jurisdiction of the investigating LEA (Law Enforcement Agency), appropriate avenues must be taken to legally gain access to the filesystem remotely.
3. **Examination and Analysis:** Concerned with the examination and analysis of forensic data. The examination is essential to gaining a complete understanding of the operating components in the distributed filesystem, while analysis is integral to the reconstruction of the evidence.
4. **Reporting and Presentation:** The legal presentation of the evidence collected and is similar to the frameworks of McKemmish and NIST. In general, the report should include information on all processes, the tools and applications used and any limitations to prevent false conclusions from being reached (Hogan, Liu, Sokol, & Tong, 2011).

3.3.1 XtreemFS Architecture Overview

XtreemFS is a virtual network-provisioned filesystem, which is used to deliver backend storage services for a CSP by providing key services such as replication and

striping. It is one example of a number of products available with similar feature sets (other examples include GlusterFS (Gluster, 2015), BeeGFS (BeeGFS, 2015) and Ceph (Inktank Storage, 2015). It is important to make the distinction between backend and frontend storage systems in a cloud computing environment as both are commonplace. Researchers Martini and Choo (2014) refer to frontend cloud storage systems as cloud storage that is purchased by users to store their personal files. For example Dropbox, Skydrive and Google Drive. Backend cloud storage systems are used by the cloud provider to support IaaS, PaaS or SaaS services. For example a backend storage system would be used to store the virtual machines that are hosted as part of an IaaS cloud or the databases and other files used by a SaaS system. Generally, backend storage is not provided directly to users.

Two main features presented by XtreemFS are striped and replicated filesystem services that are attained using three main components, the Directory Service (DIR), the Metadata and Replica Catalog(s) (MRC) and the Object Storage Device(s) (OSD) (Stender, Berlin, & Reinefeld, 2013). These components work together to provide the virtual filesystem to network users; the various components communicate with each other and with the customer as described below:

1. The Directory Service (DIR) is responsible for maintaining a registry of all services and volumes provided by the XtreemFS service (Stender, Berlin, & Reinefeld, 2013). In this respect all other parts of the XtreemFS architecture (including customers) regularly communicate status and other information to the DIR service which is a possible source of forensic identification information about an XtreemFS instance; this may include the location and number of customers connected to the environment.
2. The Metadata and Replica Catalog (MRC) are responsible for storing and managing the metadata that XtreemFS generates. Stender, Berlin, & Reinefeld, (2013) define metadata as “a collective term for all types of data that need to be managed by a filesystem except for file content”. Metadata forms a critical part of many forensic and civil litigation investigations.

3. The Object Storage Device (OSD) is responsible for storing the actual file data sent by the customers in the XtreemFS instance. A range of structures are used to store the data depending on the striping, replication, and other options selected on a per volume (or per file) basis. The OSD would likely be the focal component of the XtreemFS system for a forensic practitioner as it stores the file content data that a client has added to the virtual filesystem; this makes the OSD a key component for evidence preservation and forensic analysis.

Researchers Martini and Choo (2014) describe how XtreemFS uses the concept of 'volumes' to virtually segregate data. Volumes can be used in a number of ways, including permissions and default policies. For example, replication and striping can be applied to volumes. Volumes are also the primary administrative unit for a customer who is able to mount volumes, set policies on volumes and create snapshots of volumes. According to researchers Martini and Choo (2014) a common implementation is to leave authentication and permissions management to higher level applications. For example, a file sync application using XtreemFS as a backend could use a single volume, with each file owned by a single service user relying on the file sync application server to enforce permissions and provide authentication.

3.3.2 Directory Service (DIR)

The DIR stores the data needed to define and locate the various technical components in an XtreemFS instance. For a practitioner commencing an investigation with an identified XtreemFS instance, according to researchers Martini and Choo this is a logical starting point to determine the components and extent of the XtreemFS installation. Three artefact types of potential value to a forensic investigation exist on the DIR server:

1. **Volatile Environment Metadata:** As the directory service is responsible for maintaining a record of the various components in the XtreemFS environment, a range of environmental metadata of interest should exist. This can include the logical network location, generally an IP address of the various filesystem nodes and unique node identifiers. Other data of interest,

where available, includes data about the individual nodes such as the node type, responsibilities, configuration and ownership information, especially in terms of authentication.

2. **Non-Volatile Environment Metadata:** While in many cases the DIR may store the majority of directory data in volatile storage as it is expected to change, some data may be committed to non-volatile storage. For example, a physical hard-drive; this includes all of the metadata listed above.
3. **Configuration Files:** Configuration files can be an invaluable source of information for a forensic practitioner seeking to gain a better understanding of the operation of the system with a view to collecting evidence from individual components. Configuration information of interest includes network information such as addresses/ports used, authentication information and operational information such as local storage locations and database formats. Researchers Martini and Choo (014) describe how all XtreemFS services provide a HTTP service to present system administrators with status information. The default port for the DIR HTTP status service is 30638 and can be accessed without authentication unless the administrator password is enabled in the configuration file. When the HTTP status service is loaded it provides a range of information that could be of interest to a forensic practitioner, including the following:
    1. **Address Mapping:** Universally unique identifier (UUID) IP address mapping for each network service accessible in the XtreemFS installation. This provides two methods for a forensic practitioner to ensure they preserve and collect all services of interest, generally on different physical devices; including provider IP address, device location and the UUID.
    2. **Service Registry:** A range of registry information for each component listed with the directory service. Specific services like MRC, OSD and volumes, also have specific entries in the registry such as free disk space and total/used RAM. The focal point in the service registry for a forensic practitioner is to determine what each UUID referenced throughout the

system refers to, for example, specific volume, an OSD or an MRC type and UUID (or name) entries.

3. **Configurations:** Two OSD configuration entries consist of the "storage layout" and the "objectdir". These entries specify the location and layout of the objects stored with the individual OSD instance respectively. A practitioner is able to plan for the collection from the OSD if looking to undertake a physical collection. Network monitoring of the services port is another potential source of data for a forensic practitioner in terms of observing the hosts connection (or attempting to connect) to the XtreemFS services. Martini and Choo (2014) present a summary of attributes of common forensic interest to a practitioner as outlined in Table 3.3.1.

Table 3.3.1*Summary of Attributes of Forensic Interest*

Element	Attribute/Key	Rationale
VOLUME	**ID**	The volume ID (UUID) is necessary to manually reconstruct files stored on the OSDs.
	NAME	The volume name is potentially useful to determine the owner of the volume and/or its contents.
DIR	**ID NAME**	The directory ID and name can be useful in reporting. They may also assist in determining provenance.
	UID & GID RIGHTS	The UID and GID owners and their associated rights for the directory may be useful in determining who had access to a directory and data provenance.
	CTIME	The created, accessed and modified times may
	ATIME	be useful in determining when the directories
	MTIME	contents were accessed or modified.
ATTRS (Volume Root)	**Allows Snaps**	Used to determine if snapshots are enabled on
	sp	the volume. Used to confirm the type of striping used, size of individual stripes and potential

		number of stripes created for files in this volume.
FILE	**ID Name**	The file ID and name can be useful in reporting. They may also assist in determining provenance.
	UID & GID RIGHTS	The UID and GID owners and their associated rights for the file may be useful in determining who created/had access to a file.
	CTIME	Times may be useful in determining when files were last added, accessed (when enabled) and changed.
	ATIME	
	MTIME	
XLOC	**PATTERN**	The specific stripe pattern used for an individual file. It should be noted that this may be different to the volume striping pattern.
OSD	**LOCATION**	The UUID of an OSD which stores the stripe(s) of the file. There is generally more than one OSD entry for a file with striping (or replication) enabled.

3.3.3 XtreemFS Client

The XtreemFS client application is a significant utility to a forensic practitioner seeking to extract evidence in a more automated process from the distributed filesystem environment (Martini and Choo, 2014). The various types of metadata collected can be used, combined with the files collected, to create a relatively comprehensive (logical) representation of the data stored by a particular user in the environment. If a practitioner has access to the mounted filesystems on a client either live or based upon logged data or memory captures, this information can be used to simplify the location of the storage nodes in the environment. For example, identify the DIR instance in an XtreemFS environment.

According to Martini and Choo (2014), if a practitioner has access to a client with a mounted XtreemFS volume, they can use the various utilities to interrogate the volume for XtreemFS specific information. For example, xtfsutil is a powerful tool for both gathering information from and managing XtreemFS volumes. In its most basic form, the

command is run with only one parameter: the path to a mount-point or a file or directory under that mount-point. When used in this manner, xtfsutil provides a range of information including items of interest for three types of objects, a volume mount-point, a directory and a file; refer Table 3.3.2:

Table 3.3.2*xtfsutil Key Attributes*

Attribute Name	Applicable Type(s)	Description
XtreemFS file ID	**All**	One of the most important identifiers (discussed further in Examination and analysis section) for locating files on OSDs. A unique file/directory identifier consisting of the volume UUID and file/directory number.
"XtreemFS URL"	**Volume Directory**	The URL used to connect to the volume, including the protocol, hostname/IP address of the DIR, port number and volume name in the following format: [protocol]://[hostname]:[port number.]/[volume name]/[directory name where appropriate]. e.g. pbrpcs://DIR: 32638/xtfsvolume/dirname
Owner	**All**	The name of the POSIX user or certificate (depending on authentication type enabled) which owns the item.
Group	**All**	The name of the POSIX group which owns the item (when X509 is being used, this is derived from the OU listed in the certificate).
Type	**All**	Textual representation of the object type, e.g. volume, directory and file.
Free/Used Space	**Volume**	Free space and used space on the volume delimited by a slash.
Num. Files/Dirs	**Volume**	Number of files and directories on the volume delimited by a slash.
Access Control p.	**Volume**	The access control policy applied to the volume (e.g. POSIX).
OSD Selection p.	**Volume**	Numerical representation of the OSD selection policy.
Replica Selection p.	**Volume**	Textual representation of the OSD selection policy ("default" in our experiments).

Default Striping p.	**Volume Directory**	Textual representation of the volumes default striping policy (as discussed in Examination and analysis section).
Default Repl. p.	**Volume Directory**	Textual representation of the volumes default replication policy ("not set" or "none (not replicated)" in our experiments).
Snapshots Enabled	**Volume Directory**	A textual ("yes" or "no") representation of whether snapshots are enabled on this volume.
Selectable OSDs	**Volume**	A list of UUID, IP addresses and port numbers for the currently selectable OSDs for this volume.
Replicas	**Files**	The number of replicas (e.g. Replica 1, Replica 2, and Replica N) with a number of sub values. These include the "Striping policy" which is a textual representation of the striping policy applicable to this individual file (that may be different from the default striping policy used on the directory or volume). The OSD(s) storing objects for each of the replicas is also listed (e.g. OSD 1, OSD 2, and OSD N) with the UUID, IP address and port noted.

3.3.4 Summary

According to researchers Martini and Choo (2014), collecting evidence from the DIR, MRC and OSD components will require varying levels of access depending on the type of data and acquisition method the practitioner selects. Volatile data such as component status pages will require access to a web browser on the host or access to the (V)LAN on which the status pages are hosted and may require an administrator password (if enabled). Access to non-volatile data such as databases and object stripes may require root or administrator access on the XtreemFS hosts (depending on configuration) if a practitioner is seeking to collect them while the OS is running. However, if a practitioner cannot gain access to an administrative account on the host, then the techniques discussed by Martini and Choo should allow for the manual reconstruction of files of interest.

3.3.5 Conclusion

With the increasing digitalisation of data and use of services such as cloud computing to process, store and disseminate big data, there will be more opportunities for exploitation of large datasets. For example, in corporate or state sponsored espionage, and consequently, the continued development of the digital forensic discipline is more important than ever (Martini and Choo, 2014). An effective investigative process is one that follows well-researched and documented processes, which allow digital forensic practitioners to be able to identify and preserve, collect, examine and analyse electronically stored information from information communication technologies that would be admissible in a court of law (Butler and Choo, 2013; Quick et al., 2014).

In this case study, Martini and Choo conducted an in-depth forensic investigation of XtreemFS, a distributed filesystem that is commonly implemented in cloud computing environments. Findings from the researchers study contributed to a detailed understanding of the both the technical and process issues regarding collection of electronic evidence from distributed filesystems.

The research also highlighted the importance of a forensically sound process such as Martini and Choo's proposed distributed filesystem forensic process (see Figure 3.1) in order to provide clear guidance to digital forensic practitioners in their investigation from evidence source identification and preservation to collection of volatile, non-volatile and network data to examining and analysing the preserved data and reporting and presenting in a court of law.

Future work from Martini and Choo (2014) include validating the researcher’s framework and the proposed process with other similar distributed filesystem products such as GlusterFS, FhGFS and Ceph. Another aspect of future work is to develop forensic processes for cloud/distributed filesystems where APIs can be used for object storage and retrieval, for example Amazon S3, using a similar approach to the one presented by Martini and Choo (2014).

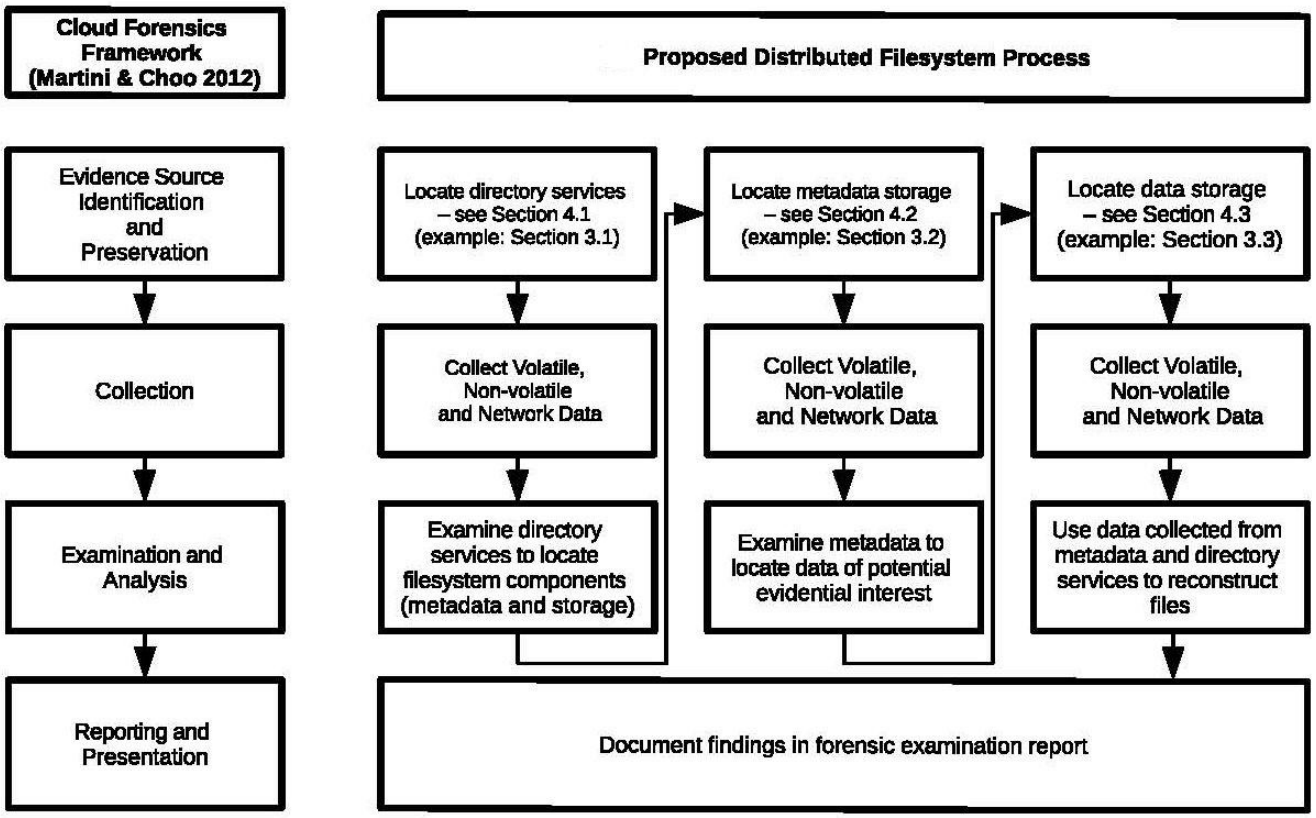

Figure 3.1: Proposed Distributed Filesystem Forensic Process (Retrieve from Elsevier Web site: http://www.elsevier.com/locate/diin)

3.4 CASE STUDY 3 - CLOUD STORAGE FORENSICS: ownCLOUD

Researches Martini & Choo(2013) conducted a case study on the storage as a service (StaaS) cloud computing architecture which is showing significant growth as users adopt the capability to store data in the cloud environment across a range of devices. Using a widely used open source cloud StaaS application, ownCloud, as a case study, Martini and Choo(2013) documented a series of digital forensic experiments with the aim of providing forensic researchers and practitioners with an in-depth understanding of the artefacts required to undertake cloud storage forensics. The researchers experiments focus upon client and server artefacts, which are categories of potential evidential data specified before commencement of the experiments. A number of digital forensic artefacts are found as part of these experiments and are used to support

the selection of artefact categories and provide a technical summary to practitioners of artefact types. Finally, Martini and Choo provide some general guidelines for future forensic analysis on open source StaaS products and recommendations for future work.

3.4.1 ownCloud Overview

For the purposes of forensic analysis, Martini and Choo have separated the ownCloud software package into two related parts, the client software, including the sync clients and the web interface, and the server software running the cloud environment.

The ownCloud server software is primarily coded in PHP to be hosted on a web server. The software to be run on an Apache server on a *nix distribution but installations on other web servers and operating systems do exist. The server uses a database for metadata persistence and offers the administrator the option of using a SQLite database for smaller installations and MySQL for larger installations. By default, files stored in the ownCloud instance are stored relatively unmodified on the server operating system file-system in a subdirectory of the ownCloud application files. Advanced storage features associated with cloud storage such as file clustering for redundancy and scalability are not managed internally by ownCloud. According to researchers Martini and Choo (2013), these features will need to be implemented at the operating system level; using a product such as GlusterFS, XtreemFS or ZFS. The server software can be extended by installing/enabling “Apps” (both internal and third party), which can add features such as server-side encryption, integration with other cloud services/storage providers and additional authentication systems.

The ownCloud client software consists of both a web interface and several client applications. The web interface is standard for this type of cloud StaaS implementation. The default web interface also allows the user to play media files, view images in a gallery and maintain a contact list and calendar. However, these non-file related features are considered out of scope for this research by Martini and Choo (2013).

The desktop sync clients are available for many major operating systems as “binaries” and as source code for compiling manually. The ownCloud sync clients download page advises that “Linux, MacOSX and Windows are built with these sources” under the sources section ("Owncloud Get Started," 2013), based upon this and the

researchers Martini and Choo's observations it is assumed that the core features of the sync client operate equivalently across all operating systems. For researchers Martini and Choo (2013) the ownCloud server is hosted in a CentOS 6 environment with a default Apache, PHP and MySQL setup. The local ext4 file system on CentOS was used to store the uploaded files. The ownCloud desktop sync client was tested in a Windows 7 environment.

3.4.2 Environment Configuration

Martini and Choo (2013) present Tables 3.4.1, 3.4.2 and 3.4.3to represent the environment specifications used in these experiments. Virtualisation was used to implement both the client and server environments. This allowed for efficient data collection (both disk and network based) and in the case of the ownCloud server instance demonstrates a common configuration in many medium/large environments where ownCloud would be found on a virtualised platform.

Table 3.4.1*Environment Specifications (Server Software)*

Server Software Specifications:	
Operating System	CentOS 6.3
Web Server	Apache HTTP Server 2.2.15
Database Server	MySQL 5.1.61
ownCloud Server Applications	Version 4.07

Table 3.4.2*Environment Specifications (Client Software)*

Client Software Specifications:	
Operating System	Windows 7
ownCloud Sync Client	Version 1.05
Web Browsers	Internet Explorer 9, Firefox 15, Chrome 21
iOS Version	5.1.1

iOS ownCloud App Version	2.03

Table 3.4.3*Environment Specifications (Forensic Tools)*

Forensic Tool Specifications:	
Guidance Software EnCase	Guidance Software EnCase
Micro Systemation XRY	Micro Systemation XRY

3.4.3 Client Forensics

Client forensic analysis was conducted on a Windows client using the desktop sync client predominantly and three major web browsers, Microsoft Internet Explorer, Mozilla Firefox and Google Chrome, to access the cloud web interface (Martini and Choo, 2013). Normally in digital forensics research of this nature, "artefacts" are defined before commencement of the research that outline the types of evidence/data the practitioner is looking to recover/present, which can be used to link the suspect(s) to the device and/or cloud services to an alleged offence. In the case of private StaaS, researchers Martini and Choo (2013) are seeking to recover the following artefacts of evidential value from the client:

1. **Sync and File Management Metadata:** Includes logging, database and configuration data stored to facilitate the sync process between client and server. These artefacts can be useful to identify the available evidence for collection from the server environment and used to build a file management history, for example, sync/update times for individual files.
2. **Cached Files:** This artefact describes the files the user has stored on the client device and uploaded to the cloud environment or downloaded from the cloud environment to the client device. In cases where the cloud environment cannot be accessed, cached files may be relied upon as the only evidence available from the cloud environment.

3. **Cloud Service and Authentication Data:** Primarily used by the forensic practitioner to discover StaaS usage and potentially gives the practitioner the opportunity to connect to the cloud computing environment using the user's credentials if no other formal method is feasible. It will commonly consist of an address (DNS, IP, URL, etc.) that identifies which StaaS instance was used and potentially stored credentials, normally the username and password of the user.
4. **Encryption Metadata:** Client encryption metadata could include databases/configurations detailing which files are encrypted and using which algorithm, keys, etc.
5. **Browser Artefacts:** These artefacts can be critical data for a forensic practitioner both in terms of evidence source identification and examination and analysis, as (like cloud service data) it can often be used to identify which StaaS instance the user is communicating with and may also include file metadata often found in URLs.
6. **Mobile Client Artefacts:** With the increasing prevalence and usage of mobile devices, mobile client artefacts may prove an invaluable evidence source for forensic practitioners. The mobile clients may store any combination of the other artefacts.
7. **Network Analysis:** Preliminary network analysis must be conducted to determine the feasibility of collecting StaaS data, with a focus on identification data, via network interception. This evidence source was beyond the scope of Martini and Choo's paper.

3.4.4 Evidence Source Identification & Preservation and Collection

Identification, preservation and collection steps were not followed as part of Martini and Choo's research as the client was setup in a controlled VM environment. During a normal investigation, however, identification would commence with law enforcement identifying electronic devices (PCs, tablets, phones, etc.) that could be of evidential value and seizing these devices for preservation and collection under a search

warrant issued by the court. The devices would then be imaged using the appropriate forensically sound tool depending on the device.

Virtual disk files (VMDK) were provided for examination and analysis as part of this experiment. In a typical law enforcement situation, these steps would likely be part of standard procedures for seizing a client device and the preservation/collection activity of image collection would result in an equivalent physical disk image file.

3.4.5 Client Examination and Analysis

Researchers Martini and Choo (2013) commenced the examination and analysis with an evaluation of the image file system to locate the artefacts. Martini and Choo (2013) found that the ownCloud client sync metadata information is predominantly stored in the "%localappdata%\ownCloud\folders" i.e. C:\Users\[Username]\AppData\ Local\ownCloud\folders directory, in which there are a number of files, named for the sync directories they represent, that also contain the configuration for each sync directory in the "ini" configuration file format.

For this experiment Martini and Choo (2013), created a sync directory named "Pictures". The configuration file for Pictures includes the following directives: "localPath" which describes the location on the client device where the synced data is stored, for example, C:/Users/[Username]/Pictures), "targetPath" which describes the folder name (Pictures) on the cloud service and "backend" and "connection" directives appear to relate to the cloud connection used. Other sync directories configuration files listed the same configuration directives and similar configuration values. File level metadata is stored on the client by the csync library, which forms part of the ownCloud client, in the form of SQLite databases located at "%userprofile%\.csync" named in the format of "csync_statedb_[HASH].db". According to researchers Martini and Choo one of these databases is created for each sync directory setup. In each database is a solitary table named "metadata" which contains an entry for each file in the directory and includes the fields "phash" which is a numerical hash of the filename, "pathlen" which is the length of the filename string, "path" which is the filename string, and "modtime" which is a POSIX timestamp which represents the last modified time of the file.

3.4.5.1 Cached Files

According to researchers Martini and Choo (2013) the ownCloud client keeps copies of all files in synchronised directories on the local disk. The file metadata configurations/databases can be used to locate the files/directories synced to the local client. While the server supports storing multiple versions of files, these do not appear to be synced to the client; refer Table 3.4.2.

3.4.5.2 Cloud Service and Authentication Data

Researchers Martini and Choo (2013) identified the "owncloud.cfg" client configuration file located in "%localappdata%\ownCloud" that contains valuable cloud identification and authentication data. The file is in the "ini" configuration file format and contains the following directives: "URL" which lists the http or https URL for the ownCloud instance synced with this client, "user" which lists the username used to connect to the ownCloud instance in plaintext (if stored), "passwd" which lists the stored password for the ownCloud client stored using base64 encoding (if stored), and "nostoredpassword" which is a Boolean representation of the option to prompt for password at sync application launch. These details are critical to forensic practitioners as it allows them to identify that cloud computing StaaS has been used and the particular cloud computing provider/instance used. According to Martini and Choo (2013), the practitioner can also potentially use the username and password listed in the file to access the cloud server and access all the data available to the user to determine if there is any further evidence stored on the cloud. However, the capacity to legally execute this process is dependent on the statutory authority of the LEA in the jurisdiction where the client is located, a practitioner using this method should also be mindful of the processes required when handling live data especially to ensure no data is inadvertently overwritten.

If law enforcement cannot be specific about where the data is and where the information can be collected, search warrants may need to be created and executed in somewhat of an iterative fashion that supports an analytic discovery process. Information

collected in researchers Martini and Choo's experiment confirms contact with the cloud provider/administrators and to guarantee that cloud data is preserved.

3.4.5.3 Encryption Metadata

The ownCloud server supports encryption of user data however this encryption appears to be handled entirely on the server (Martini and Choo, 2013). The client does not internally support client-side encryption of user data and does not appear to be aware if encryption is enabled on the server. As such no notable encryption metadata is stored on the client in an ownCloud installation.SSL encryption can optionally be used when communicating with the ownCloud server. However, this is dependent on the URL, specifically the use of the http or https scheme, used during the initial client setup.

3.4.5.4 Browser Artefacts

Martini and Choo (2013) used three Internet browsers (Microsoft Internet Explorer, Mozilla Firefox and Google Chrome) to access the ownCloud web interface and perform a number of common operations. For example, download/upload a file, access calendar/contacts. An Internet artefact search by Martini and Choo (2013) revealed artefacts relating to the ownCloud instance from all three browsers. History and downloads records revealed the files which were downloaded from the ownCloud instance and provided information on their original filename, path in the ownCloud profile and initial storage location on the client disk. For example one URLfound in the Chrome downloads list "http://owncloud.local/owncloud/?app=files&getfile= ajax/download.php?files=Chrome.txt&dir=/BrowserUploads" indicates that a file was downloaded which was named Chrome.txt from the Browser Uploads directory. The "?app=" section of the URL changes to reflect the ownCloud "app" which is being accessed, "files", "calendar" and "contacts" were noted values. A request to index.php with the parameter "? logout=true" indicates that the user has gracefully logged out (pressed the logout button on the page).

Page titles can be used to indicate the username of the ownCloud user as it is displayed in the format "[App] | ownCloud ([username])"for example "Files | ownCloud

(johnsmith)". Martini and Choo (2014) found many references to these page titles both within browser artefacts and throughout the unallocated clusters on the disk.

3.4.5.5 Mobile Client Artefacts:

ownCloud has mobile sync clients available for both the iOS and Android platforms. At the time of this research by Martini and Choo (2013), development of mobile apps was on-going. However, a forensic analysis of the current iOS app (version 2.03) was conducted for completeness using the Micro Systemation XRY product. The iOS version of the ownCloud sync client used in the experiments by Martini and Choo (2013) only allowed for the upload of images and videos from the device, as such images and text files (created via the ownCloud web interface) were the primary test files for the mobile sync client experiments. The "App PIN" functionality of the ownCloud app, which permits the user to protect the app separately from the main device with a four digit PIN, was enabled.

Martini and Choo (2013) conducted a "physical" acquisition to allow the practitioner to view the mobile device data partition, extract the actual files stored on the device and potentially recover deleted files under some circumstances. The analysis revealed a number of files of forensic interest below its application root "Documents" directory. The iOS ownCloud client maintains a cache of accessed files in the "Documents/1" directory, according to Martini and Choo (2013) the 1 appears to correlate with the "id" identifier in the users table of the DB.sqlite file. The DB.sqlite file contains a number of tables of forensic interest. The "users" table contains the URL of the ownCloud instance used with the client as well as the username and password of the user stored in plaintext. The "passcode" table contains the "App PIN" used to secure the application (if set). According to researchers Martini and Choo (2013) a mobile client can be of significant value to a forensic practitioner as it not only provides the server details and credentials for the user, potentially allowing the practitioner to connect to the ownCloud instance to collect evidence, but also a cache of files accessed and a list of files stored on the ownCloud instance (for the particular user) at the time of last sync.

3.4.5.6 Network Analysis

Basic analysis of the network communication between the ownCloud client and server was undertaken by Martini and Choo (2013) using packet captures. HTTP traffic was monitored between the client and server to determine that the ownCloud sync client is using the WebDAV protocol to handle file transfers. The ownCloud iOS client appears to use similar WebDAV requests. If plain HTTP is used to establish the connection with the ownCloud server, the content and commands sent and received from the server to the client is readable in plaintext as part of normal HTTP and WebDAV requests. For example, HTTP PUT requests revealed data. Standard HTTP Basic authorisation, comprising the username and password supplied by the user for the ownCloud instance base64 encoded, is sent with each WebDAV request.

3.4.5.7 Reporting and Presentation

Reporting on many of these client artefacts is currently a manual process and detection or heuristics of cloud computing use is not integrated into the major digital forensics analysis products used. While browser artefacts from major browsers were extracted by researchers Martini and Choo (2013) via standard evidence preparation scripts the other artefacts were located manually. If identification of cloud computing usage were to become standard practice for digital forensic investigations this would remain a very time consuming process.

Table 3.4.4*Client Artefact Summary*

Client artefact summary:		
Category	**Artefacts**	**Relevance**
Sync and file management metadata	ownCloud “folders”	Assists the practitioner in determining which client folders are synced with an ownCloud instance.
	File metadata	May assist a practitioner in determining files stored within a synced directory and file modification times.
Cached files	Synced files	Files synced to the client from the ownCloud

		instance appear as regular files. They can be located using the sync and file management metadata.
Cloud service and authentication data	owncloud.cfg	The owncloud.cfg file is one of the key ownCloud artefacts on the client. It allows the forensic practitioner to determine the ownCloud instance which is being used with the sync client and allows the practitioner to collect the users credentials (if stored). If the file has been deleted, a number of avenues for recovery are available including a keyword search of unallocated space, MFT backups and system restore.
Browser artefacts	URL parameters	When using the ownCloud web client, URL parameters (in history, bookmarks, download lists, etc.) can provide a practitioner with a broad range of information (potentially including date and time) on the ownCloud "app" being used, server file names and directories for files downloaded and logoff events.
	Page titles	A keyword search for ownCloud page titles (e.g. "Files | ownCloud (username)") is a key identifier of ownCloud use and may assist a practitioner in determining the ownCloud instance used and ownCloud username if the web client has been used.
Mobile client artefacts	Accessed files	Files which have been accessed on the iOS client appear to be cached locally, and this may allow a practitioner to access files not available on other devices.
	DB.sqlite	The SQLite database used by the ownCloud client stores valuable authentication data and file metadata that relates both to files stored on the device and the server. This may assist a practitioner in gaining access to the ownCloud instance used or in contributing as evidence of files stored and file times.
Network analysis	HTTP/WebDAV artefacts	It was noted that the ownCloud client uses WebDAV over HTTP or HTTPS to facilitate file

		synchronisation between the server and the client. When a non-SSL HTTP connection is setup to communicate between client and server data can be recovered from network captures. HTTP Basic authentication can also be captured in this setup which is another method by which a practitioner can collect a user's ownCloud credentials.

3.4.6 Server Forensics

Server forensic analysis was conducted by researchers Martini and Choo (2013) on a CentOS 6 virtual machine hosting the ownCloud PHP software via an Apache HTTP server and using MySQL as the database backend. For the purposes of Martin and Choo's experiment CentOS, Apache and MySQL were configured using setup defaults where possible or logical selections. Before Commencement of the server forensic analysis, Martini and Choo (2013) defined the "artefacts". In the case of StaaS, the researchers are seeking the following artefacts of evidential value from the server:

1. **Administrative and File Management Metadata:** Data which stores the configuration of the cloud instance and that of individual users within the cloud instance as well as database and configuration files which list the files and data stored by the user on the cloud instance.
2. **Stored Files:** The data uploaded by the user to the cloud instance.
3. **Encryption Metadata:** Data relating to encryption (if enabled) in the cloud instance, specifically data relating to decryption of user data.
4. **Cloud Logging and Authentication Data:** Logging and authentication data associated with transactions made by the user with the cloud instance (files uploaded/downloaded, login events, etc.). The more sensitive the information involved, the more monitoring of log availability is crucial. Log data is also important for incident response so business continuity requirements should be taken into account when reviewing this parameter. Finally, log data is often needed to satisfy corporate data governance and compliance requirements.

3.4.6.1Server Evidence Collection

According to Martini and Choo (2013) there are a number of different collection methods available to a forensic practitioner when collecting evidence from an ownCloud instance. The use of these methods depends on the individual attributes and circumstances of the investigation as well as the resources of the forensic practitioner. Regardless of collection method the objective is to ensure that the maximum possible useable evidence is collected and preserved, the following list of ownCloud server artefacts are recommended for collection:

1. **Data Uploaded by The Suspect:** Artefacts include the main source of evidence in an ownCloud installation (Martini and Choo, 2013). User files uploaded to the ownCloud instance are stored in a directory accessible to the web server defined as part of the initial setup of the application. This data can be located via the web server hosting the ownCloud instance. The ownCloud configuration file ([ownCloud-web-root]/config/config.php) will indicate the location of the ownCloud data directory using the "datadirectory" configuration directive. On a live system the practitioner can use this information to determine if the data is stored on the local device which is hosting the front end software or if the data is being stored on a mounted external or network based storage device as would be common in a cloud computing environment.

   Once the physical location of the data has been determined, the practitioner can make a decision as to the feasibility of taking a bit-stream image of the physical media source or taking a logical copy of the visible data structure; which may be the only practical option if the physical media is too large or complex to acquire in a timely manner.

   The structure of the "datadirectory" varies somewhat depending on the configuration of the ownCloud, however, the users data uploads should be located in the files subdirectory of the user's directory. For example, "[datadirectory]/[username]/files". Martini and Choo (20013) recommend that the "[datadirectory]/[username]" directory be copied in full. The practitioner

must also give consideration to assuring the provenance of the data collected using this method. Due to the shared nature of client devices as well as usernames and passwords, it may not be possible to determine the person responsible for upload or download of the data.

2. **Administrative and File Management Metadata**: ownCloud stores the majority of the file management metadata on the server in the SQL database which would have been collected as part of a typical digital forensic process. Tables prefixed with oc_calendar, oc_contacts and oc_media have not been included for analysis as Martini and Choo decided that they were outside the focus of this paper. However, they are expected to contain the data stored relevant to those applications (media, calendar and contacts).

Table 3.4.5*Server Artefact Summary*

Server artefact summary:		
Category	**Artefacts**	**Relevance**
Administrative and file management metadata	SQL database	The SQL database on the ownCloud server stores a range of data which could be of use to a forensic practitioner. This includes a user list, sharing permissions, encryption configuration and a cache of file system information such as file paths, owner, size, modified types, encryption status, etc.
Stored files	"datadirectory"	The "datadirectory" contains the structure and files uploaded by the user to the ownCloud instance. This is a primary source of evidence for a forensic practitioner.
	File versioning	Within the "datadirectory" is a versions directory which contains past versions of files and potentially deleted files.
Encryption metadata	Blowfish encryption	Encryption can be optionally enabled on an ownCloud instance, when enabled most files uploaded are encrypted (some file types are exempt by default). The encryption key is stored in the

		"encryption.key" file stored in the users "datadirectory" subfolder and encrypted with the user's password. A practitioner can collect the user's password from a number of other artefacts and decrypt the files stored.
Cloud logging and authentication data	Web server logging data	The default logging data stored by the web server (Apache in these experiments) can be of use to a forensic practitioner to determine when a user has communicated with the ownCloud server and the changes made by the user as part of that session. The usefulness of this data was limited when the web client was used. However a large amount of information was available on sync client transactions.

3.4.7 Conclusion and Future Work

The research by Martini and Choo (2013) demonstrates that cloud StaaS provides a significant number of useful artefacts for forensic practitioners in an investigation. Using ownCloud as a case study, Martini and Choo successfully undertook a forensic examination of the client and server components of an ownCloud installation and discussed the relevance of a number of artefacts to a forensic investigation. According to Martini and Choo (2013), the research is the first that provides a holistic discussion on cloud StaaS forensics from both client and server perspectives; previous researchers have focused on either the server (Dykstra & Sherman, 2012) or the client devices (Chunga, Parka, Leea, & Kang, 2012).

Martini and Choo's analysis of the client devices demonstrated that in many cases significant data can be found which links the user to a particular ownCloud instance; providing a forensic trail to the ownCloud server instance even when evidential data on the client may be securely deleted. The client artefacts found in the ownCloud experiments are likely to be common with other open source cloud storage products developed in the future as cloud products mature and develop a common feature set. While individual implementations may vary, practitioners can use the artefacts

discovered in these experiments as a basis for their investigation of the client as a potential evidence source and perhaps more importantly as a link to the cloud computing instance on which other data may be stored. The file metadata and cloud authentication artefacts found are of particular interest in an investigation which heavily involves cloud computing use. These artefacts can be used not only to determine the cloud computing instance used but also provide authentication data potentially allowing an investigator to collect data from the cloud instance directly and help link user actions with the data stored in the cloud computing environment via the use of file metadata such as permissions and timestamps (Martini and Choo,2013).

Martini and Choo's server analysis showed that while collection of data in an environment with one server such as the instance in these experiments may be relatively straightforward, factors such as encryption could complicate investigations significantly. While many practitioners may focus upon collection of the files uploaded by the suspect in the first instance it has been demonstrated that it is important to collect the range of artefacts suggested as they may be required to assist in linking a user with the data stored in the cloud instance, recovering previous data stored by the user in the cloud instance or in decrypting data stored by the user. In many cases, it will not be possible to collect the entire cloud storage instance due to the size and amount of unrelated (other users) data stored on the physical device(s). Consequently, this makes collection of the full range of artefacts critical as once the preservation methods are no longer being applied to the cloud instance, critical data such as encryption keys and metadata may be lost.

The utility of Martini and Choo's iterative cloud forensics framework was demonstrated with client artefacts being used to identify cloud storage usage and being used to decrypt files stored on the server. The iterative nature of the framework suggests that client devices are analysed first to both identify cloud usage and allow practitioners to request preservation by the CSP in a directed manner providing as much information on the data requested to be preserved as possible. Analysis of sync and file management metadata on the client can also help prevent time being spent on investigation of cloud services which are unlikely to be of evidential value.

While it may be possible to preserve an ownCloud instance by disconnecting the environment from the network, this approach is not guaranteed to ensure preservation and will result in potentially significant downtime for all users of the cloud instance. It is instead recommended that StaaS developers integrate preservation technologies directly into the product. In this case, ownCloud could "freeze" a user's account preventing them from making any further changes, after a valid request is received from law enforcement, and provide a forensic practitioner with a package containing the contents of the users files directory, previous versions, encryption key and any relevant metadata and logging information. The provision of this package would not only simplify the extraction of evidence for the practitioner but also ensure minimal downtime for the cloud instance.

Martini and Choo (2013) recommend that future work be continued in this area looking at other cloud StaaS products available including those hosted in the public cloud environment to determine the best practices for forensic extraction and analysis on these platforms. Further work on the potential for network interception as a method of forensic collection should be pursued especially as a method of identification of potential evidence sources.

.

Chapter 4
Research Question and Hypothesis

The literature reviewed in Chapter 3 provided a foundation of written knowledge regarding the reliability and integrity of the evidence collection process during a digital forensic investigation in a ‘cloud’. A significant amount of the literature has been reviewed in relation to the research subject that included the various cloud types and services, industry standards, data security, data privacy/integrity, governance and cloud legal frameworks. The process of conducting a conventional digital forensic investigation was also reviewed in parallel with specific literature relating to the digital forensic processes in a 'cloud'. Additionally, the review of similar research studies (Section 3.1) identifies examples of possible evidence source and the tools and techniques used to gather such evidence within a cloud environment, for example log files, network monitoring tools and digital forensic analysis tools.

The development of the research question was constructed based on the literature reviewed in Chapter 2 and the review of similar research studies in Section 3.1. Obtaining forensic evidence in a cloud although possible comes with caveats that will question the integrity and reliability of the evidence gathering process and whether the evidence is challenged and upheld in a court of law. Caveats such as legal jurisdiction, chain of custody, the time taken to process the evidence, the sources of the evidence and investigator qualifications are areas of the evidence gathering process that are more challenging when conducting a digital forensic investigation in a cloud.

Table 4.1 displays the main research question and the associated hypothesis which has been developed from the information discussed thus far.

Table 4.1 ***Main Research Question and Associated Hypothesis***

Main Question: *What are the processes required to acquire and preserve the reliability and integrity of evidence when conducting a digital forensic investigation in a cloud?*
Asserted Main Hypothesis: Existing systems are designed to acquire and preserve network and user data and are capable of providing viable evidential trails together with sufficient information to support digital forensic investigations in a cloud.

Furthermore, a number of related secondary questions have also been developed. The secondary questions have been devised in order to set out and answer various linked components of the main research question and are set out in Table 4.2.

Table 4.2 ***Secondary Research Questions***

Secondary Question 1: *What are the hardware and software requirements for the successful acquisition of cloud digital evidence for forensic purposes?*
Secondary Question 2: *What are the capabilities of the proposed system design to acquire and preserve digital evidence within a cloud including the qualifications and expertise of those tasked to gather the evidence?*
Secondary Question 3: *What is the effect of monitoring live cloud network traffic in terms of acquiring digital evidence?*
Secondary Question 4: *What are the methodologies, techniques and tools used to conduct digital forensic examination and analysis of the acquired evidence from a cloud?*

Hypotheses have also been developed for each of the secondary questions which have been proposed. Because of the difficulty in composing a succinct hypothesis for each secondary question, a brief synopsis has been given to articulate the reasoning behind each of the informed hypotheses. Table 4.3 displays the hypotheses for each of the secondary research questions presented in Table 4.2.

Table 4.3 ***Secondary Research Questions Associated Hypotheses***

Hypothesis 1: That the hardware configurations used will have the capability to collect and process user VM network traffic. That the software requirements will effectively acquire and preserve forensically-sound digital evidence using repeatable and defensible processes.
Hypothesis 2: That the proposed system will be capable of acquiring enough data to determine and reconstruct certain events, such as user history and event logs. The preservation and integrity of the acquired evidence will be subject to scrutiny given that the person who retrieved the data is unlikely to be the official investigator.
Hypothesis 3: That live cloud evidence can be monitored and analysed without diminishing the performance and acquisition capability of the system.
Hypothesis 4: That examination and analysis of acquiring digital evidence in a cloud can follow conventional methodologies, tools and techniques.

In order to attempt to answer the proposed research questions, validate the asserted hypotheses and conduct research testing in an organised manner a data map was developed. Figure 4.1 presents the research data map outlining the main research question, secondary research questions and the links to the associated research testing phases. Furthermore, each testing phase is also linked to the associated point of data collection achieved from testing. Finally, the findings gathered from the research testing phases and related data collected will be used to aid in determining the asserted hypotheses.

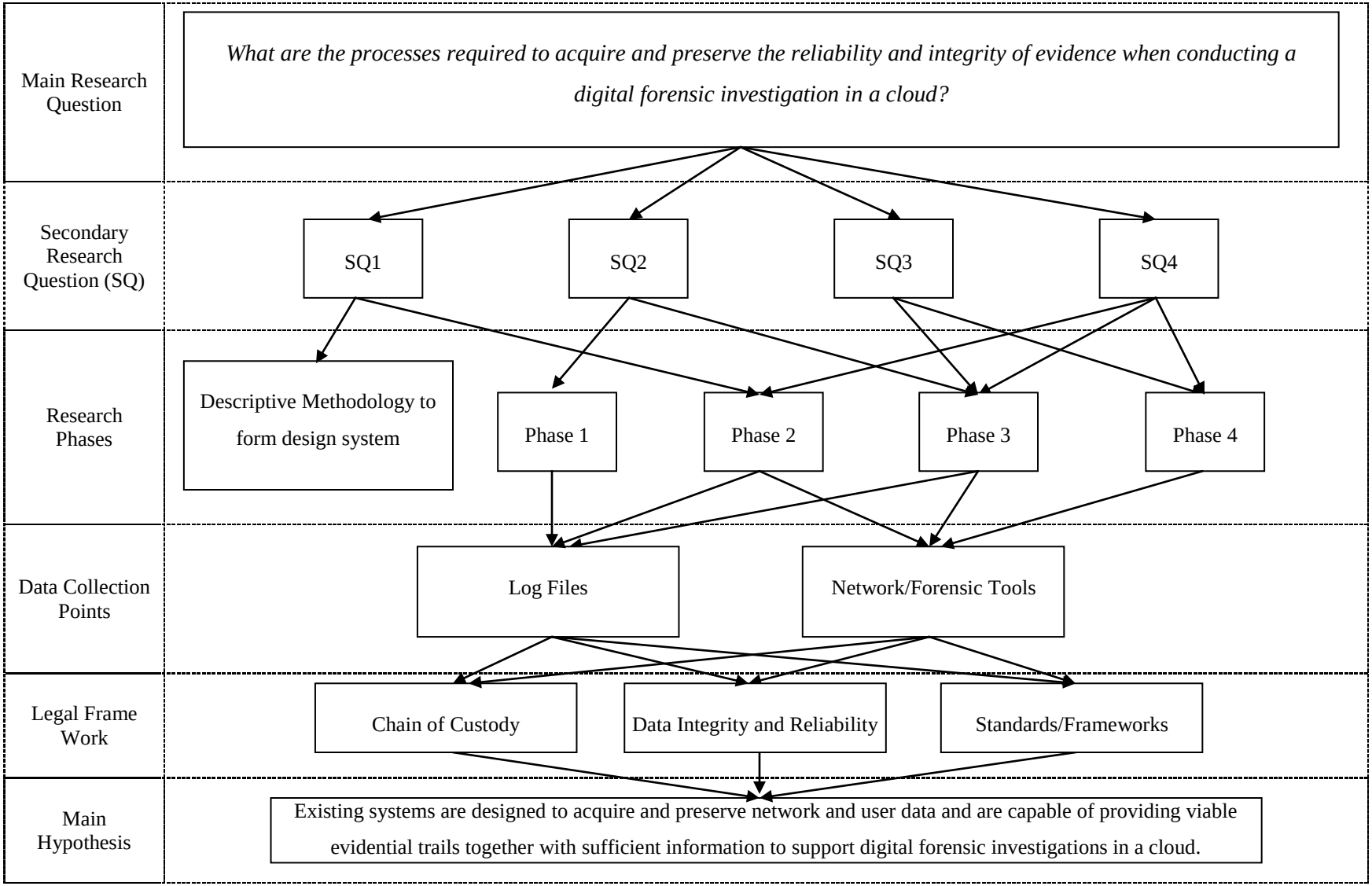

Figure 4.1: Research Data Map.

The complexity of cloud services introduces a number of unknown parameters. Cloud Service Providers (CSPs) are cautious about offering guarantees for compliance-ready services and the adoption of those services. CSPs generally promote a simple and cost effective way in delivering ICT services irrespective of jurisdictional borders, this raises questions and challenges in examining compliance with legal frameworks (Guilloteau & Mauree, 2012).

A key question following a breach in 'cloud' security is, how would the customer access 'cloud' services and capture all relevant data required to carry out an investigation? To help answer this question requires the consideration and co-operation of several subsets; the customer will not only have to consult with the CSP and its stakeholders but will need to consult with its own in-house legal counsel and cloud computing experts.

In a 'cloud' a complete capture of all data related to an event under investigation is not possible, as an alternative a snapshot of the data can be taken but this process has limitations. Some data will not be available, some data will be suspect, and some data will be court ready and can fit into the traditional network forensics model. The challenge for a cloud forensics investigator is to recognise the data set for each of the three categories i.e. not available, suspect and court ready.

The expansion of data storage capacity in a 'cloud' is also a disadvantage for a digital forensic investigation as it involves an increase in forensic data to analyse. Primarily, CSPs only allow remote access to a logical representation of data, rather than the principal physical infrastructure. A further limitation is when the CSP infrastructure is itself virtualised and also leased from another cloud provider. This use of virtualisation affects the privacy of other users of the cloud, whose data may be inadvertently collected during the evidence gathering process. In some jurisdictions, inadvertent access of non-relevant data from a cloud environment may contravene local privacy and/or data protection legislation (Grispos, Storer, & Glisson, 2012).

The general lack of specific tools and limited professional expertise in 'cloud' forensics is also of concern, a situation made more challenging when encryption, proliferation of endpoints, multi-jurisdiction and loss of data control, to name a few, are involved. Cloud organisations, including CSPs and cloud customers, need to establish a

cloud forensic capability; otherwise, they are likely to face ongoing difficulties when carrying out forensic investigations in a 'cloud' i.e. criminal intrusions and major policy violations. Investigators will also face difficulties when collaborating with law enforcement in resource confiscation cases due to limited forensic knowledge and preparation (Ruan K., Kechadi T., 2011).

There are published papers exploring potential difficulties in the process of maintaining the chain of custody in a 'cloud'. Leading private and public organisations like SANS (SysAdmin, Audit, Networking, and Security), ISACA (Information Systems Audit and Control Association) or NIST (National Institute of Standards and Technology) have not yet agreed on a set of recommendations or best practices to follow when there is a security breach inside a 'cloud'. As well, some of the newer cloud data-visualisation tools make excellent forensic and early warning tools for security engineers and security investigators. Some cloud data-visualisation tools work just as well as the traditional tools like NetFlow, but they were never intended to be network forensics tools. Security engineers and IT workers are required to be creative with their current tool sets and make them work in the cloud environment. An additional problem as to how a network forensics investigation can be successful within the cloud is that cloud computing remains an unfamiliar environment for security engineers. The security department should be part of the entire cloud decision process from the architecture to the services and systems that will be hosted by the CSP.

4.1 THE RESEARCH MODEL

The aim of this research is to determine the reliability and integrity of the evidence gathering process when carrying out a digital investigation in a cloud and whether the evidential process will withstand legal scrutiny in a court of law. A theoretical research model is proposed using a design science approach in order to establish a framework for the research to be conducted. Descriptive methodology will be used to conduct fact-finding enquiries to establish the state of affairs in the proposed research area (Kothari, 2004/2006). From this the system architecture and the components needed to construct the system design will be derived. The intention of the research is to implement a Cloud

Forensic Model (CFM) system architecture and to acquire and preserve the reliability and integrity of the captured data. The goal is to capture and record potential digital evidence that will aid digital forensic investigation in a cloud by providing viable digital evidence. It is proposed that the CFM be implemented and reviewed to determine the capabilities of the model and its ability to provide digital evidence of an acceptable standard from within the cloud environment. The implemented CFM will monitor and acquire evidential data between server and the client. The proposed theoretical research model described is illustrated in Figure 4.2.

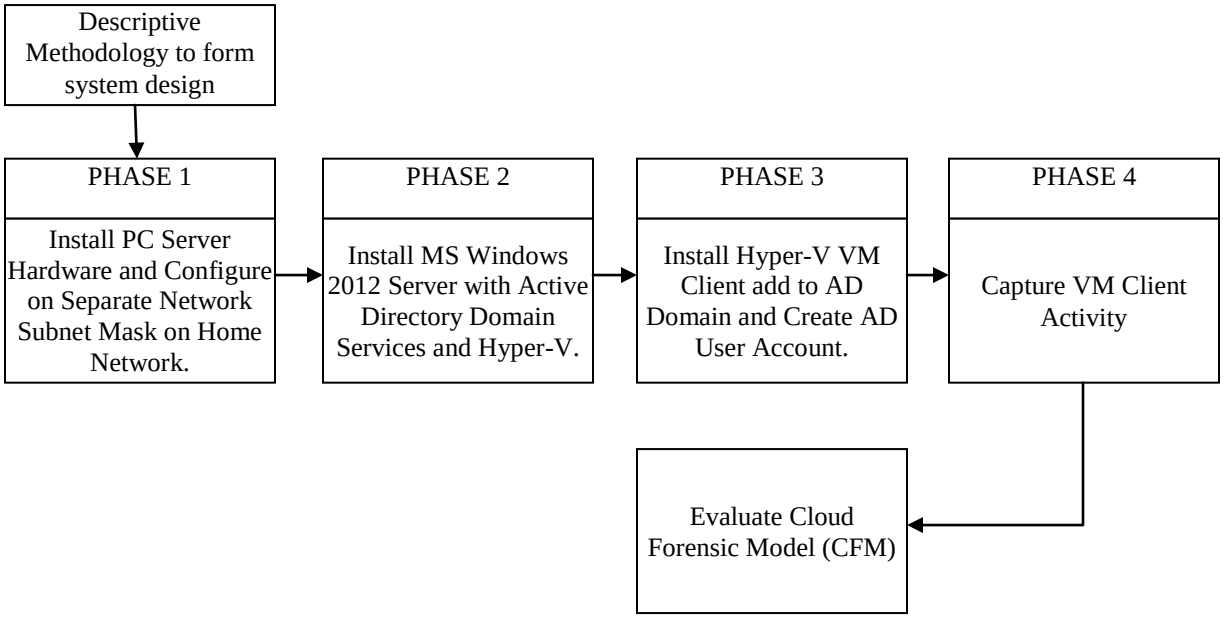

Figure 4.2: Theoretical Research Model

Cloud forensics is the process of identifying, labelling, recording, and acquiring forensic data from a possible source in the 'cloud'. The data source includes client-side artefacts that reside on client premises, and provider-side artefacts that reside on cloud service provider infrastructure. Since the segregation of duties is disparate in different 'cloud' service models, the tools and procedures to collect forensic data are also different. In different deployment models, provider-side artefacts are also different, for example, in

public clouds, provider-side artefacts need to be segregated among multiple tenants, whereas in private clouds, there is no such need (Ruan K., Kechadi T., 2011). As a consequence the research methodology identifies a number of limitations that are outlined and discussed in order to acknowledge the constraints in the proposed research. It is important to identify such limitations in order to correctly evaluate the results obtained and to determine if, or where further areas of research are needed regarding the forensic tools and techniques used in order to maintain data integrity and reliability when carrying out a digital forensic investigation in a cloud environment.

The first limitation to conducting the research was to acknowledge that it was impractical to expect the scope of the research to include all simulated cloud models; for example, Private, Public, Hybrid and Community clouds. A decision was made to conduct the experiment using the Public Cloud model. As a cloud simulation it is also difficult to replicate real world scenarios like data distribution, relocation, compression and resizing policies in the cloud which are far more complex than in a simulated laboratory environment. Also, different cloud providers may use different methods and functions to implement these policies.

The second limitation is that the cloud simulation was installed on a single home networked PC as the university research lab was unable to provide the necessary network configuration/hardware to simulate the research experiment. I utilised a personal home PC to install Microsoft Windows 2012 Server Datacentre and segmented the Microsoft Windows 2012 Server connection from my normal home network by installing and configuring a second Cisco Linksys X2000 router which I was fortunate to have at the time. The MS Windows Server 2012 operated on a separate network subnet mask to that of my normal home network with full Internet access as shown in Figure 4.3 and Figure 4.4.

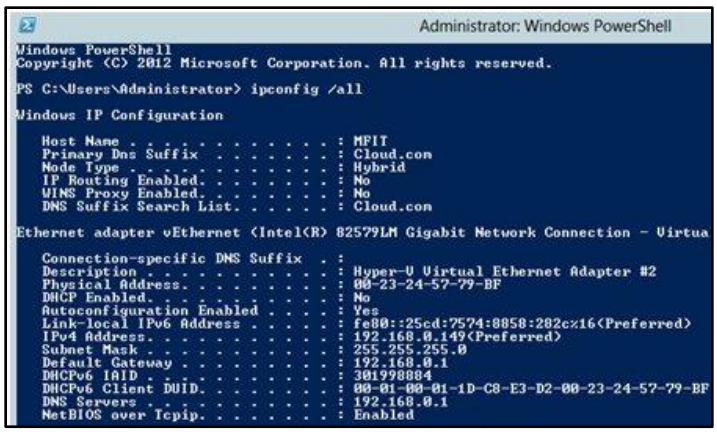

Figure 4.3:MS Windows Server 2012 IP Configuration

Figure 4.3 shows the IP configuration of Microsoft Server 2012 including the static IP address, default gateway and DNS, Host Name (MFIT) and the Primary DNS suffix of Cloud.com. Active Directory Domain Services and Microsoft Hyper-V roles have also been added to the Windows Server 2102.

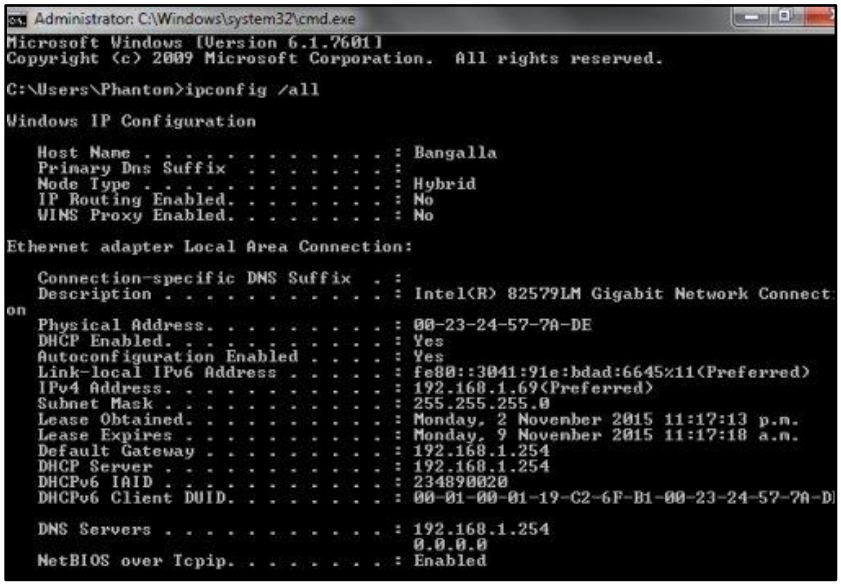

Figure 4.4: Home Network IP Configuration

Figure 4.4 illustrates the segmentation of my home and research laboratory networks with my home network on a different IP subnet mask. My home network IP

address is 192.16.1.69, default gateway of 192.168.2.254 and DNS address of 192.168.1.254. Figure 4.3 also shows the Host Name of ‘Bangalla’ with no Primary DNS suffix as it is not associated with a Domain.

The third limitation is that the researcher assumes that the communication links between the server and the virtual machines are secure and that no data is lost during the investigative process. However, 'in a real world situation' the communication links between VM workstations, the cloud data centres and servers are under constant threat and data may be lost or compromised during the communication stage. The fourth limitation was that due to licensing costs it was not possible to utilise all acceptable industry standard forensic tools, for example, Encase by Guidance Software.

Descriptive methodology requires the delineation of the system design that relates to the architecture and components needed to be implemented into a practical system. The software and hardware components and associated configurations will also be described and discussed in order to present a proposal for the intended system.

Initial testing will be conducted on the proposed system which may necessitate the system to be modified several times in response to learning. Outcomes from testing will also dictate which components in the CFM need to be modified in order to achieve a stable and reliable system architecture.

Testing will involve two phases. Phase One of testing is to implement the required hardware and perform benchmark testing to determine the capabilities of the existing infrastructure. Such information includes the specifications of the server hardware including the CPU and RAM capacities and the operating software of the server and client VM. The testing of the existing hardware/software capabilities is important as it provides a baseline for running a series of network monitoring tools and access to various VM logs. Although researchers have performed similar methodologies for benchmark testing, it is important to ensure that the test network is functioning correctly. Furthermore, with existing cloud capabilities having been determined, the findings of various network monitoring tools and event logs are assured to be accurate.

There are ten logs dedicated to Hyper-V in Microsoft Windows Server 2012 that are available through Event Viewer that are grouped together and listed alphabetically as shown in Table 4.4 (Siron, 2012).

Table 4.4 ***Hyper-V Log Files***

Hyper-V-Config	Contains XML entries that relate to the configuration files that describe individual virtual machines whose names are globally unique identifiers. Files are found under C:\ProgramData\Microsoft\Windows\Hyper-V\Virtual Machines or under VM-specific folders on a Cluster Shared Volume. The most common error is 4096, which indicates that Hyper-V is unable to locate an expected configuration file. It isn't entirely unusual to encounter this error in normal operations, as utilities and operations may move the XML files in a fashion that isn't entirely in sync with the Hyper-V services. It normally doesn't require attention unless it is a persistent error.
Hyper-V-High-Availability	This section contains events related to the interaction of Failover Clustering with Hyper-V. Most of the events here will be informational recording of actions that the Cluster service took on individual VMs. Errors should be very rare and are generally related to the same sort of synchronization issues that cause the Hyper-V-Config 4096 errors.
Hyper-V-Hypervisor	As the name implies, these events are related to the hypervisor itself. Most of the events will be related to the creation and destruction of partitions, which are the temporary container that hold running virtual machines. If there is any sort of problem with Hyper-V itself, especially issues that prevent the service from starting, this is where you'll find out about it.
Hyper-V-Image-Management-Service	The related service is devoted to the handling of VHD files. If any operation involving a virtual hard drive fails, details are logged here.
Hyper-V-Integration	This log tracks events associated with the Integration Services that are installed into virtual machines. Most of the problems reported here can be corrected by re-installing or upgrading the Integration Services components.
Hyper-V-Network	The virtual switch(s) in your deployment will record events here. The first events will be the creation of the virtual networks themselves, as well as pairing of external networks to physical network cards. When a virtual network adapter is created or destroyed in a virtual machine, a matching

	virtual port is created on the virtual switch; the creation/destruction of those ports will be registered here.
Hyper-V-SynthNic	The synthetic network cards in virtual machines will log an event when they start (12582). Look here for clues as to why a network card won't function, such as MAC collisions.
Hyper-V-SynthStor	Virtual storage controller drivers use this log for their events. The most common event is logged by virtual SCSI controllers as they start. The virtual IDE driver is emulated and not synthetic, so it initializes before the VM loads and will not log a matching event. If a drive cannot be attached to the virtual controller port as expected, it will be logged here.
Hyper-V-VMMs	The Virtual Machine Management Service generates these events. Problems with import and export actions will be logged here, as will AVHD merge operations. Host shutdown events will also be tracked in this log. It will also report when it cannot locate the files for a VM. As in other logs, these are likely to be cleaned up once a VM is completely removed.
Hyper-V-Worker	Hyper-V's worker threads log these events. Normally, this is the busiest of all the logs, but most of them are trivial. If you're curious how long that last Live Migration took, this is where you'll find it. Emulated network and storage drivers (as opposed to the synthetic drivers) will create events here.

4.2 DATA VALIDATION TOOLS

As discussed by (Grispos, Storer, & Glisson, 2012) software hashing tools are regularly used in conventional investigations to validate the on-going integrity of data used as evidence. A hash function is an algorithm for converting arbitrary length data strings into fixed length hash values, typically a few hundred bytes in length. Hash functions are designed so that any change in the input data should (with high probability) produce a different output hash value. Hash values can therefore be periodically computed for disk images, files or other data representing forensic evidence to gain assurance that the evidence has not been changed. ("Harvard Law Review," 2014)

Data stored in a cloud can also be subjected to hashing for integrity checking purposes. For example, Amazon Simple Storage System (S3) and Web Services (AWS)

have both implemented MD5 hashing checksums for objects stored in their services (Amazon, 2015). In principle, an investigator can record these checksums to show that any evidence acquired has remained unchanged during the course of the investigation. In addition, this feature may be of future use for investigators wishing to store forensic evidence that they have gathered in a cloud environment.

The use of hashing tools that are implemented, deployed and controlled by cloud providers raises a number of challenges (Grispos, Storer, & Glisson, 2012). The use of external facilities draws the provider into the chain of custody. In addition, the investigator has less opportunity to test and evaluate the hashing features in a cloud, compared with tools developed for use on conventional desktop PCs. Typically; an investigator can use a selection of tools that implement the same hash function to compute a hash for some sample data. Any differences between the results produced can be investigated. However, in a cloud environment, the investigator has only a single implementation (the checksum implementation deployed by the cloud provider) to use. Consequently, the investigator's ability to validate the correctness of their tools is limited.

Data capturing tool FTK and Microsoft Server 2012 Hyper-V Snapshot were used to capture data of a Hyper-V hostCloudPC2 as shown in Figure 4.5 and Figure 4.6. Microsoft Network Monitor 3.4 was used to capture live network packets associated with a test user and CloudPC2

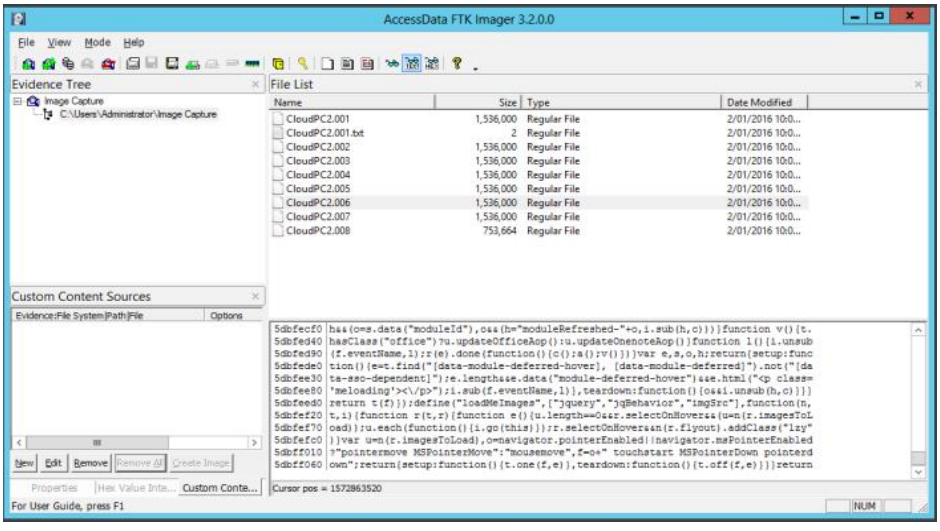

Figure 4.5: FTK Data Capture

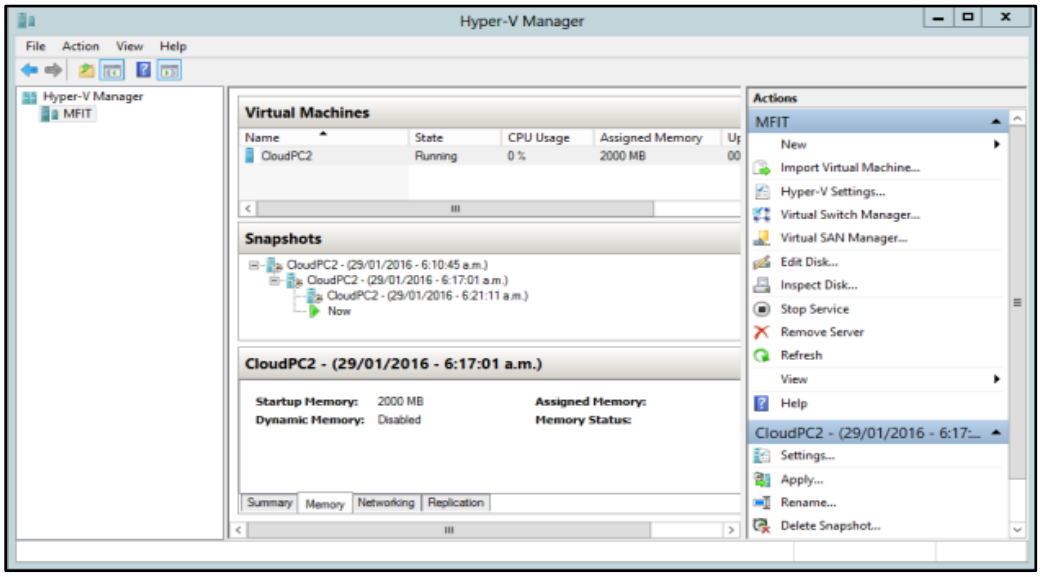

Figure 4.6: SnapShot Capture

4.3 CONCLUSION

Chapter 4 focused on developing the research methodology in evaluating the reliability and integrity of the evidence gathering process when conducting a digital forensic investigation in a cloud environment in accordance with digital forensic investigation compliance principles and standards.

Similar previous studies presented by other researchers were also studied to assist in the development of testing methodologies. The additional information gained from the review of similar studies, coupled with the comprehensive literature review in Chapter 2, was used to develop the research questions, as well as the predicted hypotheses for each question. The proposed research model was then outlined, providing a logical progression of testing phases to be conducted. A descriptive methodology was employed to form the proposed research design, consisting of the design architecture and components. Furthermore, the proposed data requirements and limitations of the proposed research methodology were detailed and discussed.

Chapter 4 has presented an overview of the chosen research methodology. The research data map (see Figure 4.1) provided a graphical diagram of the main and secondary research questions, linking each question to a specific phase of testing and the associated data collection point. The proposed phases of testing presented in Figure 4.2 outlines the phases of research testing needed to address the research questions. The model provides the goals of each phase of testing, involving implementing the system design in a testing environment. Chapter 5 is to report the findings of the experiments that were defined in this chapter.

Chapter 5
Research Findings

5.0 INTRODUCTION

Chapter 4 formulated the research questions based on the problems and issues relevant to the reliability and integrity of the evidence gathering process in a digital investigation in a cloud; from which, the research methodology was established. Relevant studies from previous research were selected for review which guided the proposed research methodology. The research question, sub-questions as well as the research hypotheses were then derived for the selected problem and issues that were identified in the literature review in Chapter 2. The data requirements for the experimentation were presented and the limitations of the proposed research discussed.

Chapter 5 is to report the findings of the research phases defined in Chapter 4. In order to clearly articulate the research findings, various techniques will be used. The outcomes from each independent but consecutive four phases of testing will be reported and analysed to evaluate the purposed research design. The findings from data collection, data processing and data analysis will be presented in Section 5.1, 5.2 and 5.3. The summarised data from each phase will be presented in tabled format in Section 5.2 (initial testing findings) and Section 5.4 (stabilised testing findings). Section 5.4 concludes the chapter.

Section 5.1 will present the previously developed research questions (Section 3.2) in a tabled format. Each question will be answered and discussed in terms of the asserted hypotheses. Arguments will be made for and against the hypotheses and a summary made of the outcome. Following the tabulated questions, the findings of the research will then be discussed in detail in Section 5.2. The chapter concludes with Section 5.2 in which the knowledge gained from the research conducted will be used to develop recommendations from the writer outlining best practices and testing methodologies to further promote digital forensic investigations in the cloud.

5.1 EVIDENCE FOR RESEARCH QUESTIONS ANSWERS

The main question and the following sub-questions were developed from both the literature review (Chapter 2) and the study of similar research cases (Section 3.1). The research questions will now be set out and answered in a table format. The table will be headed by each question asked, followed by the hypothesis as first outlined in the research methodology (Section 3.2). The asserted hypothesis given is a brief theoretical explanation using the knowledge gathered from the literature reviewed at the outset of the research project. The table will then present both the arguments for and the arguments against the hypotheses made, based on the findings of the research testing phases and technical knowledge learned. The arguments for, will be those that find in support of, or prove the hypothesis, while arguments against, will refute or disprove the offered hypothesis. Reference will be made to specific findings to substantiate the statements providing rational reasoning for each argument. At the end of each table, a brief summary of the research question and tested hypothesis will be given in order to accept, reject or found as indeterminate based on the findings achieved.

5.2 RESEARCH QUESTIONS REVIEWED

The main research question was developed to provide a specific goal for the research testing phases and to concentrate testing on a particular area. The main research question is: *What are the processes required to acquire and preserve the reliability and integrity of evidence when conducting a digital forensic investigation in a cloud?*

In order to answer the proposed research question several phases of testing were proposed and conducted. The system design of a Cloud Forensic Model (CFM) was implemented and subjected to various testing to determine the capabilities of a system to acquire and preserve evidential data in a cloud.

Table 5.1 displays the main research question, the associated hypothesis, arguments for and against are made and a summary of the tested hypothesis is given.

5.3 SECONDARY RSEARCH QUESTIONS ASSOCIATED HYPOTHESES

Four secondary research questions were also developed in support of the various elements needed to answer the main research question.

Tables 5.2, 5.3, 5.4 and 5.5 display the secondary research questions, from question one to four respectively. Each table presents the associated hypothesis, the arguments for and against the hypothesis, a summary of points discussed and the significance of the research outcome for each question. A statement of position accepting, rejecting or deeming the hypothesis indeterminate is also given for each question.

Table 5.1 ***Main Research Question and Tested Hypothesis.***

Main Question: *What are the processes required to acquire and preserve the reliability and integrity of evidence when conducting a digital forensic investigation in a cloud?*	
Main Hypothesis: Existing systems are designed to acquire and preserve network and user data and are capable of providing viable evidential trails together with sufficient information to support digital forensic investigations in a cloud.	
ARGUMENT FOR: The acquisition and preservation of evidential data is able to be accomplished by the CFM system design.	**ARGUMENT AGAINST:** Acquisition of data is more difficult to obtain than conventional digital investigations, cooperation from CSP's is required, cloud data may lack key forensic attributes, limited forensic tools to process cloud data, chain of custody is more complex and may be compromised. The implemented CFM provides a single source of evidence derived from network traffic between a user's VM and the cloud for forensic investigation. Additional sources of evidence may be needed to augment and support the evidential trails acquired and preserved by the CFM.

SUMMARY: The CFM was capable of acquiring live network traffic generated on the existing network. Furthermore, the system was also capable of acquiring a snapshot of evidentiary data of the users VM activity and recover various log files. There remain a number of potential issues and limitations pertaining to the system design and architecture as well as the software and hardware used to implement the system design. The arguments made for and against prove the hypothesis to be indeterminate.

Table 5.2 ***Secondary Question 1 and Tested Hypothesis.***

Secondary Question 1: That the hardware configurations used will have the capability to collect and process user VM network traffic. That the software requirements will effectively acquire and preserve forensically-sound digital evidence using repeatable and defensible processes.	
Hypothesis 1: That hardware storage and configurations used are capably sufficient to collect and process VM network traffic. The software requirements will need to include the capability to acquire and preserve VM network traffic using forensically sound applications.	
ARGUMENT FOR: The acquisition and preservation of evidential data is captured by the CFM system design. A cabled LAN link between components is important and affect the CFM system design, therefore, the configuration must accommodate the required bandwidth to transport the captured network traffic data between components. The network monitoring and forensic software applications are important to reliably acquire and preserve VM network traffic. Microsoft Hyper-V SnapShot, Microsoft Network Monitor, Wireshark and Snort were used to capture/monitor live VM network traffic and Forensic Toolkit (FTR) NetSleuth were used to analyse all data collection; therefore, the capability of the implemented system relies heavily on software.	**ARGUMENT AGAINST:** Creating a disk image of the customer cloud environment cannot be achieved without capturing the entire multi-tenancy cloud environment. Closing the customer cloud environment will result in the loss of volatile data.

SUMMARY: The CFM was capable of acquiring live network traffic generated on the existing network. Furthermore, the system was also capable of acquiring a snapshot of evidentiary data of the users VM activity and recover various log files. There remain potential limitations pertaining to the system design and architecture as well as the software and hardware used to implement the system design. The arguments made for and against prove the hypothesis to be indeterminate.

Table 5.3 ***Secondary Question 2 and Tested Hypothesis.***

Secondary Question 2: What are the capabilities of the proposed system design to acquire and preserve digital evidence within a cloud including the qualifications and expertise of those tasked to gather the evidence?	
Hypothesis2: That live cloud evidence can be monitored and analysed without diminishing the performance and acquisition capability of the system.	
ARGUMENT FOR: The CFM is capable of acquiring and preserving a data set of network traffic generated during testing. The data collected from live monitoring of user's network traffic can be presented as credible evidence provided those that are charged with obtaining the evidence are authorised and qualified and that the tools, techniques and methods used are in accordance with digital evidence gathering standards.	**ARGUMENT AGAINST:** Authorised access to evidence stored directly on the cloud is unlikely. Determine the qualifications of those obtaining the evidence on your behalf and whether the evidence gathering process falls within the chain of custody.
SUMMARY: A full data set of network traffic was not always able to be achieved in certain scenarios. However, the significance of the outcome was that enough data was collected to provide information about the activities of the user. Therefore, the hypothesis is proved to be accepted.	

Table 5.4 ***Secondary Question 3 and Tested Hypothesis.***

Secondary Question 3: What is the effect of monitoring live cloud network traffic in terms of acquiring digital evidence?	
Hypothesis 3: That live cloud evidence can be monitored and analysed without diminishing the performance and acquisition capability of the system.	
ARGUMENT FOR: The CFM is capable of acquiring and preserving a data set of network traffic generated during testing. The data collected from live monitoring of user's network traffic can be presented as credible evidence provided those that are charged with obtaining the evidence are authorised and qualified and that the tools, techniques and methods used are in accordance with digital evidence gathering standards.	**ARGUMENT AGAINST:** Authorised access to evidence stored directly on the cloud is unlikely. Determine the qualifications of those obtaining the evidence and whether the chain of custody is compromised during the evidence gathering process.
SUMMARY: A full data set of network traffic was not always able to be achieved in certain scenarios. However, the significance of the outcome was that enough data was collected to provide information about the activities of the user. Therefore, the hypothesis is proved to be accepted.	

Table 5.5 ***Secondary Question 4 and Tested Hypothesis.***

Secondary Question 4: What are the methodologies, techniques and tools used to conduct digital forensic examination and analysis of the acquired evidence from a cloud?	
Hypothesis 4: That examination and analysis of acquiring digital evidence in a cloud can follow conventional methodologies, tools and techniques.	
ARGUMENT FOR: Fundamental methodologies and techniques were used to analyse acquired conventional network traffic in packet capture file format.VM packet capture logs were filtered by frame types and timestamps with tools such	**ARGUMENT AGAINST:** Although fundamental methodologies and techniques may be used to analyse cloud network data, the process of gathering that data has its own unique challenges compared to conventional digital investigations. Obtaining evidence within a cloud environment requires

as Wireshark, Microsoft Network Monitor and Snort.	methodologies and techniques that may have yet to be tested. For example, accepted industry standard tools such as Encase and FTK would normally be used to analyse image data, the results of which would stand up to scrutiny. However, obtaining an image of a user's cloud data can be challenging and not always possible. Therefore, investigators will need to be creative in the way they obtain evidential data within a cloud.

SUMMARY: Although traditional methodologies and techniques can be used to analyse traditional network traffic, there are a number of specific challenges that cloud i.e. VM network packet capture analysis and examination requires. Cloud forensic tools need to be adapted and tested to provide digital forensic investigators the necessary standards to capture and analysis cloud evidential data The arguments made for and against prove that although the hypothesis is possible it remains indeterminate.

5.4 TEST SCENARIO CONFIGURATION

The 'Testing Scenario' involved the identification and recovery of evidential data from a simulated IaaS cloud environment using conventional forensic tools, methods and techniques. At this stage the key consideration is to obtain client-side artefacts relevant to a simulated cloud digital forensic investigation. During the testing scenario the investigator and LEA rely implicitly on the cooperation and trustworthiness of the CSP to provide unaffected evidential data; this may not necessarily be the case in a real-life situation. Chain of custody requirements would be difficult to achieve under this testing scenario as it is difficult to determine how the evidence was obtained, where the evidence was physically located before it was assembled and exactly who handled the evidence before it was handed over to the investigating practitioner or LEA. It may not be possible for the practitioner or LEA to verify the forensic procedures used to collect the evidence. For the purposes of the proof-of-concept, in this particular test scenario, the focus is on the evidence acquired and the tools used to access the data. Later studies could measure the performance and implications of these processes in order to move forward with the development of cloud forensics.

A simple VM Client-Server architecture is adopted to simulate the communication between Microsoft Hyper-V and the Microsoft Server 2012 Datacentre. In order to perform the desired experiment, the following physical hardware/software specifications were used (refer Figure 6.1):

1. Lenovo M93
2. Intel(R) Core(TM) i5-3470T CPU @ 2.90 GHz
3. 16GB Installed Memory
4. 64-bit Operating System, x64-based processor
5. Windows Server 2012 Datacentre
6. Microsoft Hyper-V
7. Microsoft Active Directory
8. Internet Information Services (IIS)

After the 'test' infrastructure was set up, Microsoft Network Monitor and AccessData Forensic Tool Kit (FTK) were installed on the cloud server. Third party software would normally be installed on a separate remote PC/laptop but due to time constraints and limited research resource this was not possible.

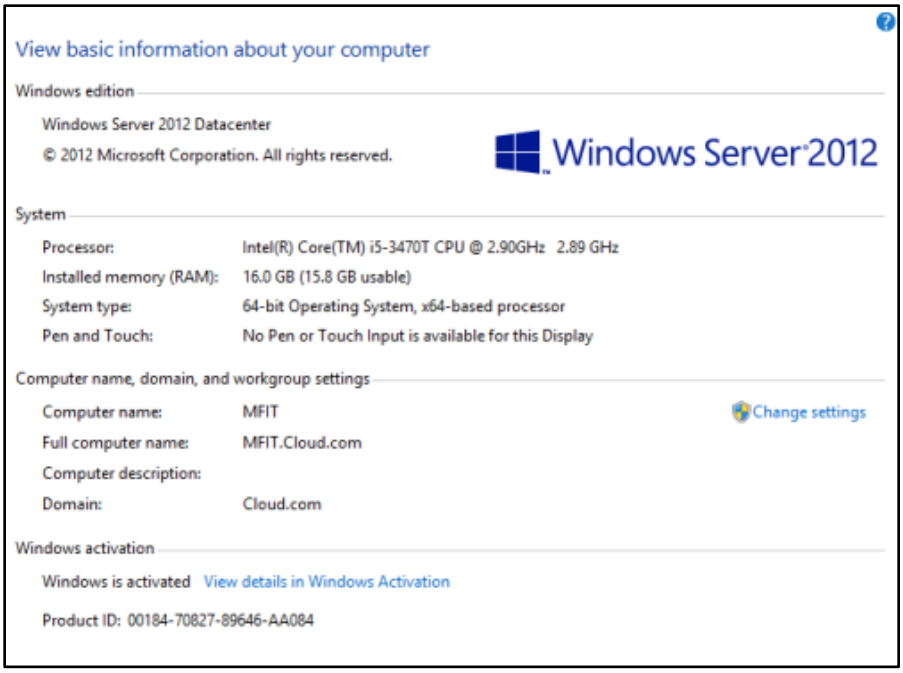

Figure 5.1: Microsoft Windows Server 2012 Specifications

The VM node specifications running on Microsoft Hyper-V (refer Figure 6.2):

1. Windows 7 Professional
2. 2048MB Installed Memory
3. 64-bit Operating System

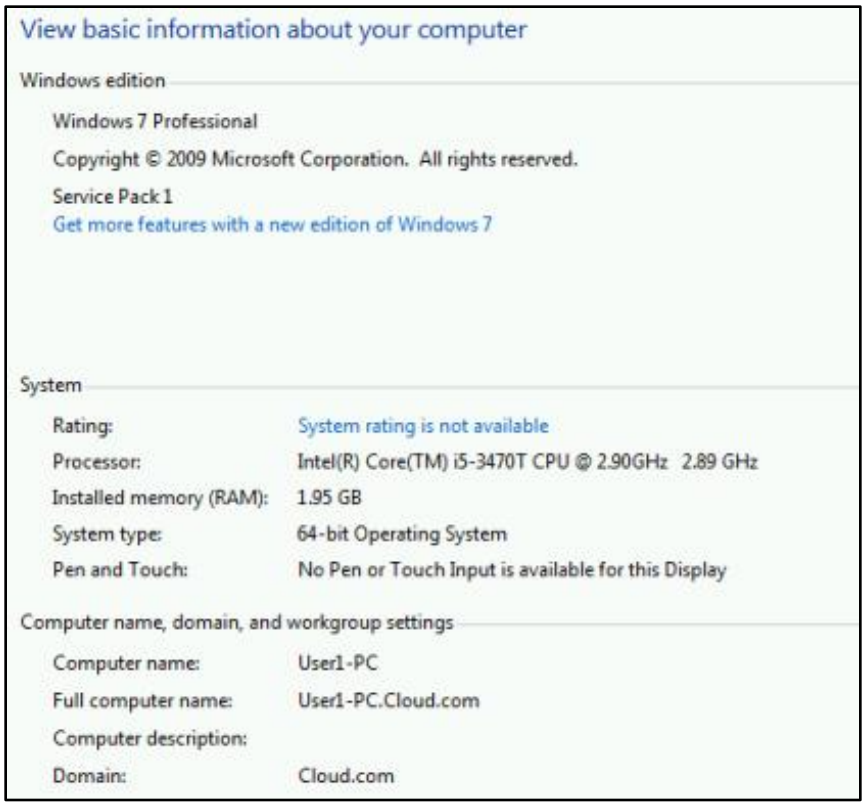

Figure 5.2: MS Windows 7 Professional Specifications

5.4.1 Active Directory

Domain Name Controller titled 'MFIT.Cloud.com' was created with network users accounts set up for Mark Piwari, Brian Cusack and David Robertson (refer Figure 5.3). A single Virtual Machine, USER1-PC, was created using Hyper-V which was then added to the 'MFIT.Cloud.com' Domain (refer Figure 5.4).

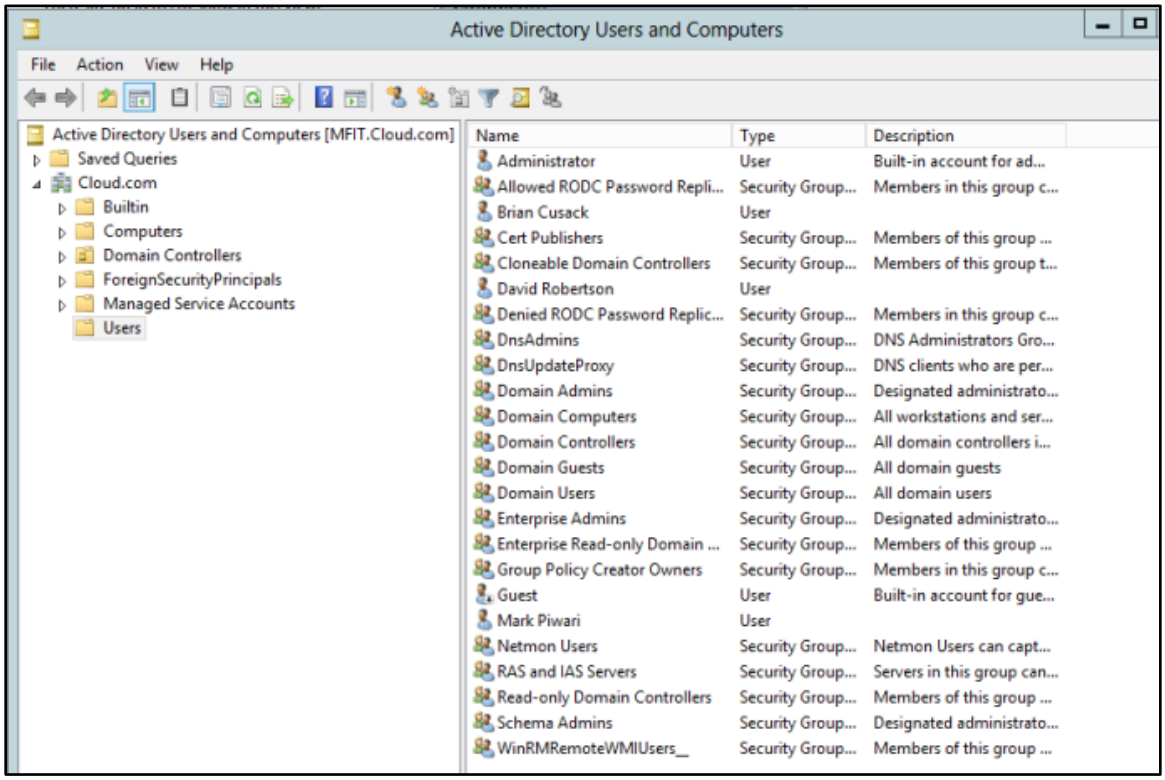

Figure 5.3: Active Directory Users for MFIT.Cloud.com Domain

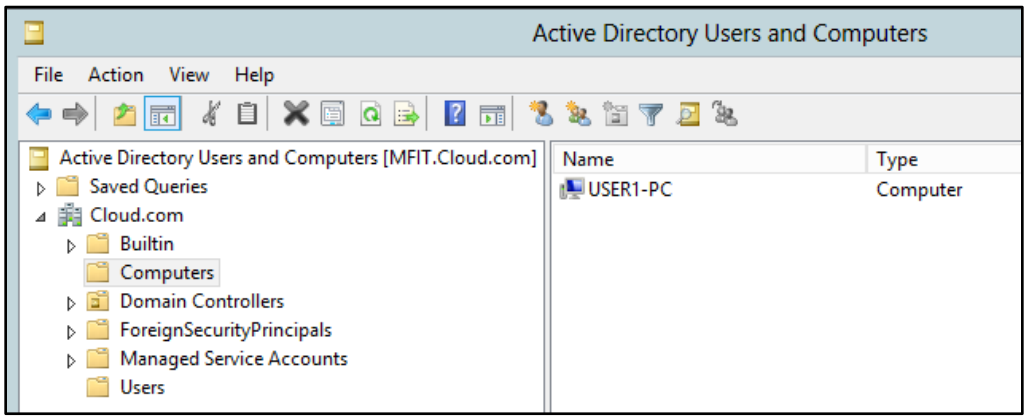

Figure 5.4: Active Directory Computers joined to MFIT.Cloud.com Domain

5.4.2 Microsoft Hyper-V

Among the many new features found in Windows Server 2012, Hyper-V 3.0 is one feature that has gone through a significant change that includes two new options for administrators. The extensible switch extends a virtual network's functionality and replica makes planning for disaster recovery easier for administrators by creating copies of virtual machines. PowerShell v3 has been in focus for much of Microsoft's move toward cloud computing. With over 2,300 cmdlets, a lightweight command that is used in the Windows PowerShell environment, PowerShell v3 makes automation throughout the server the new norm while working with programs such as Active Directory to promote cloud-based servers without the need of a deployment wizard directly on the server.

Hyper-V 3.0 role was installed and configured and virtual hard drive, USER1-PC, was created as shown in Figure 5.5.

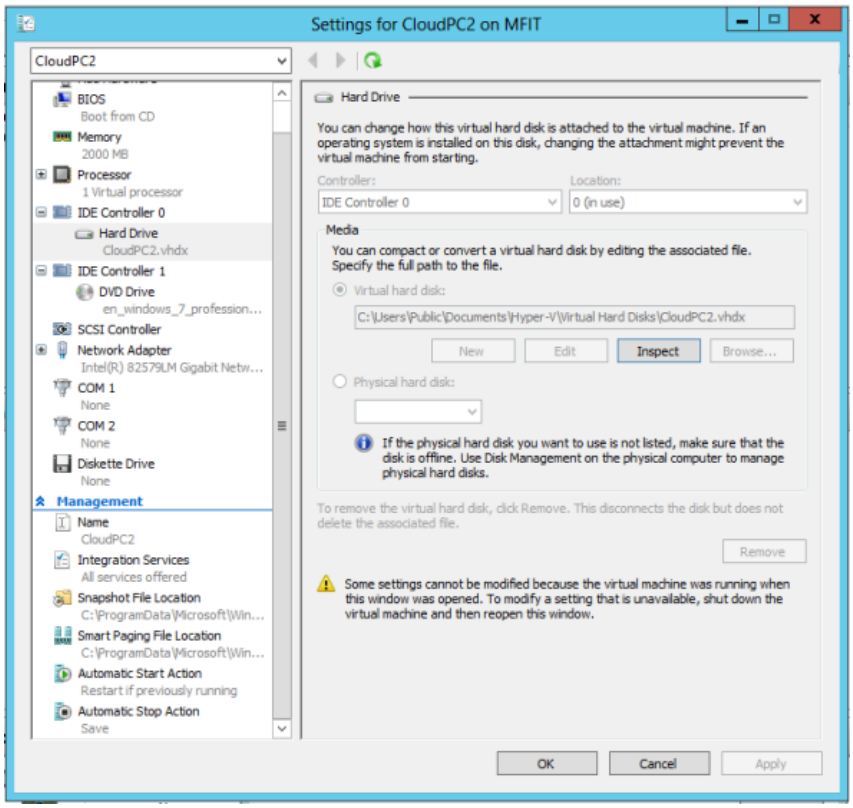

Figure 5.5: Microsoft Hyper-V Configuration

5.5 TEST SCENARIO DATA COLLECTION

The experimental cloud environment was created to test whether evidential data can be extracted from an IaaS cloud without compromising the reliability and integrity of the evidence. With the testing platform in place, including third party software, the process of extracting the data was conducted without the usual challenges that investigator practitioners and/or law enforcement agencies (LEAs) would face when conducting a digital forensic investigation in the cloud. The legal and cross-border jurisdictional hurdles as well as the restrictive access to the data played no part in this research test scenario which is far from being a real-life situation, but more a proof-of-concept study. Throughout the research study continual reference was made to the legal and jurisdictional access to data and how the chain of custody process is more complicated in a cloud environment. For the purpose of this research paper none on the real-life limitations exist and therefore extracting evidential data from the test cloud environment was unimpeded.

Client-side data was therefore extracted with the full co-operation of the CSP who provided a highly qualified technician to perform the data extraction for the user David Robertson. Tools used to acquire the data included Microsoft Network Monitor v3.4, AccessData Forensic Tool Kit (FTK) v3.2, Microsoft Hyper-V 3.0 and various system and event logs. The following are a series of screenshots taken of the data extraction process:

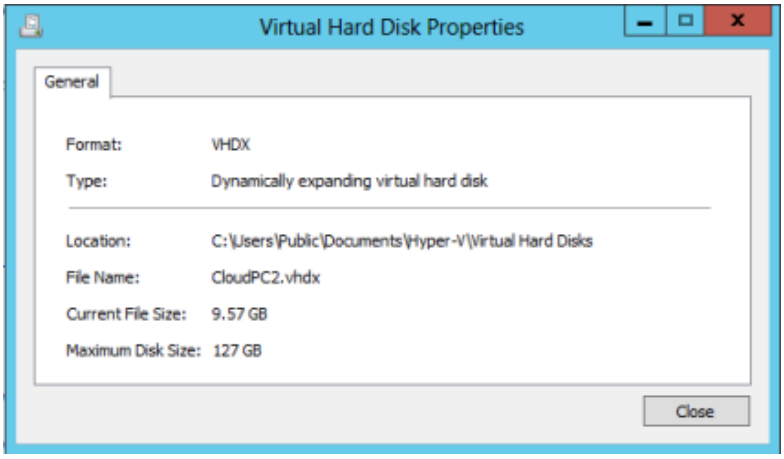

Figure 5.6: User Virtual Machine Image created in Microsoft Hyper-V

A live SnapShot of the client-side VM PC being used by David Robertson was taken using Microsoft Hyper V to capture memory, network and replication details of the PC being used by the targeted user, as shown in Figure 5.7.

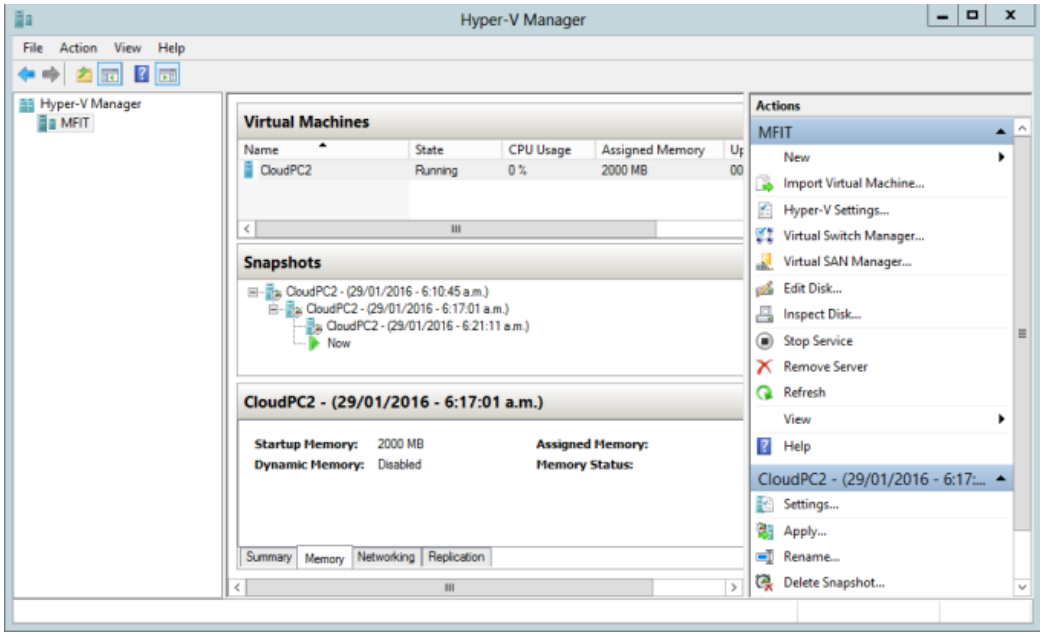

Figure 5.7: Live Hyper-V SnapShot of Client-Side VM

As the research study defines, digital forensics is the identification, collection, preservation, analysis, and interpretation of digital evidence. FTK(Forensic ToolKit) Imager by AccessData is a imaging tool that enables the assessment of electronic evidence to determine if further analysis is warranted. FTK was used to to create a forensic image of David Robertson's VM without changing the original evidence, and hashes for file integrity (refer Figure 5.8).

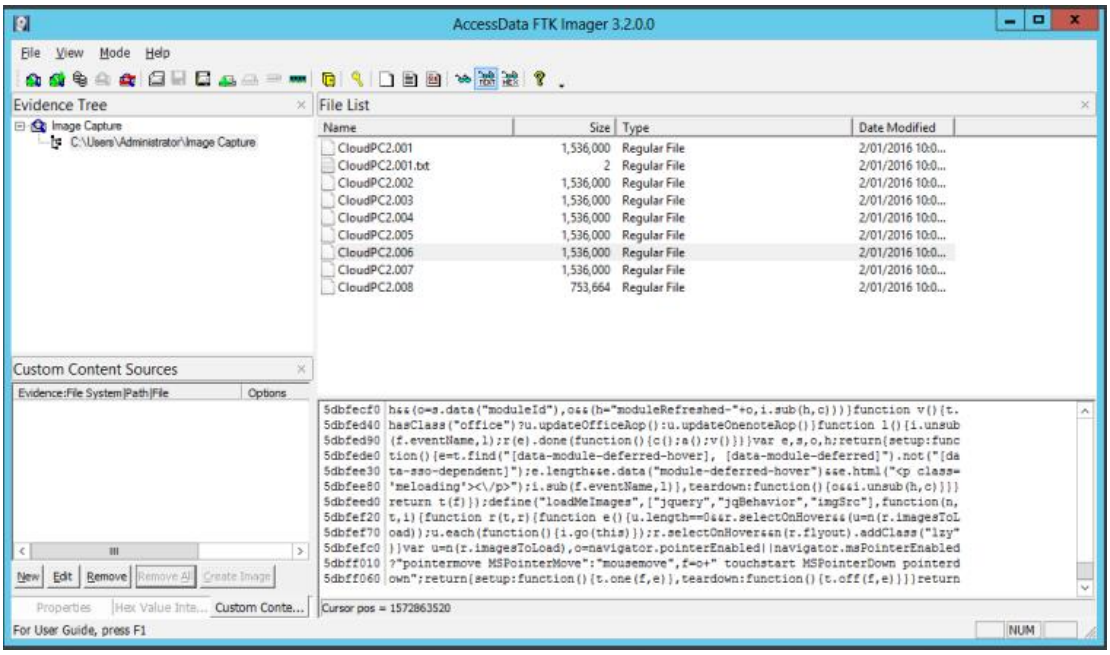

Figure 5.8: FTK Image of Client-Side VM

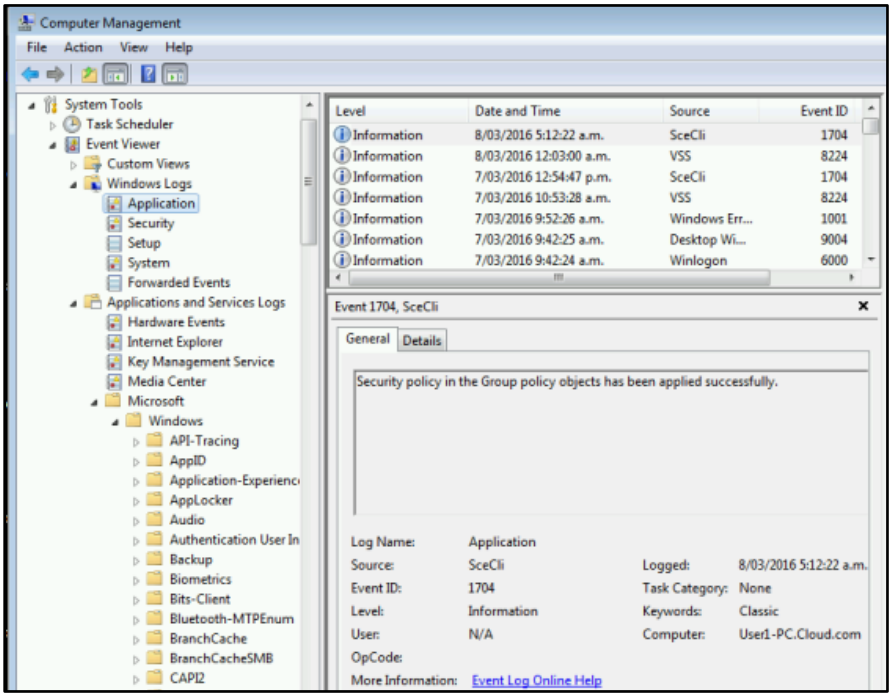

Figure 5.9: Microsoft Event Logs

Windows event Logs as shown come in two categories. The first are the "Windows Logs," and contain familiar logs such as Application, Security, and System logs. They also contain two new logs. Setup Log, that contain events related to application setup and/or as a target log for application installers, and the ForwardedEvents Log, that contain events forwarded by another computer. The "Application and Services Logs" category stores events from a single application or component rather than events that might have system-wide impact. While the Windows logs contains five separate logs (Application, Security, System, Setup, and ForwardedEvents), the Application and Services Logs category contains about 461 separate logs (Golomb, 2013). Figure 5.9 shows the client-side event logs for user David Robertson.

To complete the data capture process browser history artefacts for David Robertson were extracted from the two browsers installed on the client-side VM i.e. Internet Explorer 10 and Firefox 43.01. Along with event viewer logs, cookies, and system log/cache files the process of extracting a comprehensive set of evidential data from the test cloud was relatively straight forward; however, in reality the process would be far more complex and challenging.

5.6 CONCLUSION

Chapter 5 has developed a discussion of the findings from the research testing which was reported, analysed and presented in Chapter 4. The research questions proposed in the research methodology (Section 3.2) have been answered and discussed in terms of the previously asserted hypotheses, and a conclusion reached regarding the validity of the predicted hypotheses. The main research question of the project was centred on the capabilities of a design system to acquire and preserve evidence in VM network traffic in a cloud. Subsequently, a research model was formed (Section 3.3) and a system design prescribed. During research testing the CFM was implemented and a stabilised design was formed. A simulated cloud environment with VM clients was recreated to determine the systems capabilities. The findings discovered that the CFM was able to be implemented using conventional digital forensic techniques associated with readily

available network monitoring tools like Microsoft Network Monitor and built-in Microsoft Server 2012 tools such as Microsoft Hyper-V, and a body of information gained was used to provide recommendations and best practice principles to help safe guard the integrity and reliability of the digital evidence gathering process in a cloud environment.

Chapter 6
Research Findings

Chapter Six presents the final conclusion of the thesis and the research conducted. Therefore, a summary of the research findings (Chapter 4), and subsequent review of the discussion of the findings (Chapter 5) is made. Chapter 6 concludes with a synopsis of the limitations of the conducted research but also identifies opportunities for potential future research within the chosen topic area. A simulated 'cloud' environment was created using Microsoft Server 2012 Hyper-V. A customer domain, MFIT, was created in Microsoft Active Directory and VM's with Microsoft Windows 7 installed were added to the MFIT Domain as the testing environment to be investigated. The popularity of cloud technology for businesses, organisations and government agencies has made significant gains over the past 5 years due mainly in part to cost, scalability and elasticity. However, security, privacy and diminished governance and controls over how information is managed and stored are hurdles that potential customers find difficult to reconcile and trust. The fragmentation of non-binding cloud standards, guidelines and procedures also contributes to the 'buyer beware' argument. If cloud security was breached, how would the customer know and what controls, if any, would the customer have over any subsequent investigation? Cloud service providers who are signatories to the New Zealand Cloud Computing Code of Practise have pledged to notify the customer as soon as possible of any security breach or when data is lost or compromised unless that notification would compromise a criminal investigation (2012, June).

Case studies provided in Chapter 3 and the research model (CFM) used in Chapter 4 demonstrates that digital evidence stored in a cloud can be retrieved using industry standard tools, methods and techniques. However, a major obstacle to a successful investigation relies ultimately on the synchronisation and collaboration between the cloud service provider, jurisdictional agencies and the investigator and the management of the evidence acquired. It is highly unlikely that the investigator will

obtain personal control of a cloud service, but it may be possible to obtain an image of the service's data from the cloud service provider. However, utilising the cloud provider to obtain the image means that the investigator is unable to initially control the chain of custody documentation. A complete chain of custody should identify all individuals who have come in contact with the evidence. Consequently, if the cloud provider is asked to obtain the initial image, then the chain of custody begins with the employees assigned to this task. It is unclear how an investigator would be satisfied that the cloud provider's employees were competent to gather evidence on their behalf. Also, a growing issue in cloud forensics is how to handle large data sets (Clarke, N., & Tryfonas, T., 2011). Extracting evidence from a cloud is likely to involve a much larger data set than what investigators are currently examining. Therefore, how do investigators process cloud evidence effectively and store this evidence safely and securely? Future development in addressing large-scale cloud data processing may include specific hashing techniques, customised information retrieval solutions, random data sampling, parallel processing solutions or a combination of these technologies?

The CFM used in this research paper emulated an IaaS cloud service but each of the cloud services (SaaS, PaaS or IaaS) will require a unique acquisition methodology and each service will provide a unique set of challenges that need to be addressed. For example, in an IaaS environment an acquisition methodology is adopted so that forensic 'images' can be obtained from VM's operating in this environment. To address known environmental and trust issues each cloud environment needs to be configured independently. For example, a private cloud associated with a government agency is going to have more control over the environment than users of a public cloud. Consequently, trust concerns associated with user rights, operating system functionality and capability restrictions can be minimised in a government environment. Conversely, a public cloud allows end-users to utilise multiple operating systems, possess administrative rights, provides minimal audit information, and introduces more complexity and risk into the environment, which makes the forensic acquisition of these environments far more challenging (Grispos, Storer, & Glisson, 2012). Grispos, Storer & Glisson, 2012, p. 18 also pose the following questions:

Can the cloud be suitably stabilised to allow investigators to take an accurate representation of the evidence at a specific point in time? Can datasets be developed to effectively mimic this environment? With the investigator not having complete control of the environment, can the investigator be sure evidence is not in the process of being altered in the cloud at that moment in time?

Table 6.1 ***Summary of Challenges to Digital Forensics in Cloud Environments.***

Phase	Action	Challenge
Identification	Identifying an illicit event	Lack of frameworks
Preservation	Software tools	Lack of specialist tools
	Sufficient storage capacity	Distributed, virtualised and volatile storage; use of cloud services to store evidence
	Chain of custody	Cross-jurisdictional standards, procedures; proprietary technology
	Media imaging	Imaging all physical media in a cloud is impractical; partial imaging may face legal challenges
	Time synchronisation	Evidence from multiple time zones
	Legal authority	Data stored in multiple jurisdictions; limited access to physical media
	Approved methods, software and hardware	Lack of evaluation, certification generally, but particularly in cloud context
	Live acquisition	Acquisition of physical media from providers is cumbersome, onerous and time consuming data is inherently volatile
	Data integrity	Lack of write-blocking or enforced persistence mechanisms for cloud services and data
Examination	Software tools	Lack of tested and certified tools
	Recovery of deleted data	Privacy regulations and mechanisms implemented by providers
	Traceability and event reconstruction	Events may occur on many different platforms
Presentation	Documentation of evidence	Integration of multiple evidence sources in record

	Testimony	Complexity of explaining cloud technology to jury

Table 6.1 (Grispos, Storer, & Glisson, 2012) summarises the challenges faced by investigators in the collection of evidence from a cloud environment. Evidence extraction tools, volatile and persistent memory acquisition software, as used in conventional investigations, on a client computer may not provide the necessary data. Virtualisation of data storage in a cloud also makes it difficult to identify and isolate the segment of one or more physical storage devices owned by a cloud provider that represent the user's data that needs to be gathered for analysis. As discussed in Chapter 2 virtualised data stored on a cloud may be spread across multiple physical devices located in various geographic physical locations with an interface between the virtual storage and the investigator. The use of virtualisation may also impact the privacy of other tenants of the cloud environment, whose data may be inadvertently gathered during the investigation; this may contravene local privacy and/or data protection legislation.

Evidence gathered during a digital forensic investigation can be summarised to explain their conclusions in a number of forms. Evidence may be submitted to a court in the form of a report and an investigator could be asked to provide expert testimony and be subject to cross-examination. Alternatively, the results of an investigation could be used by an organisation to improve their corporate policy and could evolve as a form of documentation for future investigations. In 1993, the United States (US) Courts made a ruling in the case of Daubert v. Merrell, which defined the admissibility of scientific evidence; this admissibility was based upon four criteria as described by O'Connor (Grispos, Storer, & Glisson, 2012):

1. Has the scientific theory or technique been empirically tested?
2. Has the scientific theory or technique been subjected to peer review and publication?
3. What is the known or potential error rate? Every scientific idea has error rates, and these can be estimated with a fair amount of precision. There are known threats to validity and reliability in any tests.

4. Has the theory or technique been accepted as a standard in its scientific community?

The way in which forensic investigations are carried out in a cloud will be subject to the same criteria as outlined in the so-called Daubert principles, if the resulting evidence is to be acceptable in court. The empirical testing of cloud forensic methods may be challenging due to the evolving nature of the technology. Empirical testing of forensic tools typically utilises standard data sets, but these tools need to be developed for cloud forensic methods. There is a clear need to develop a standard evaluation method and data sets for cloud forensics, if results of cloud forensic investigations are to pass the Daubert principles.

This research paper identified the jurisdictional and legal challenges of the evidence gathering process in a cloud and whether the process is robust enough to withstand scrutiny in a court of law. The research explored current and future development of evidence acquisition methods for the cloud and whether the methods are forensically sound to assist digital forensic examiners, law enforcement agencies, and the court to evaluate with confidence the evidence gathering process in a cloud. Although there are processes in place to ensure that best practise and standards are followed during a cloud digital investigation the reliability and integrity of the evidence gathering process is largely based on intrinsic 'trust' and co-operation between all parties involved. A holistic approach that can be thwart with potential complications; highlighting the technological, organisational and legal challenges. However, cloud forensics brings unique opportunities that can significantly advance the efficacy and speed of forensic investigations (Ruan K., Kechadi T., 2011).

6.1 FUTURE RESEARCH

In this research, two digital forensic tools and various system log file have been evaluated for extracting forensic evidence from the cloud client to analyse different artefacts and different evidential sources. Forensic analysis for each tool and comparisons have been performed. For further research testing other digital forensic tools should be included using the same proposed methodology, in order to compare the findings with this

research. Future research could also focuses on other platforms such as Apple Macintosh and Linux. Accessibility to the evidential data on the cloud server is one of the main challenges when conducting a cloud forensic investigation, as the physical device on which the data resides may not be available due to jurisdictional and legal challenges. Also, in cloud forensics, the distributed nature of data processing in the cloud, as described in Section 3.3, represent a serious concern for investigators and stakeholders, as traditional approaches to evidence collection and recovery may no longer be applicable; an opportunity for the CSP to mandate the development of alternative tools to recognise that such an environment is being used and effectively 'fetch' all the evidence from various physical locations. There is also the problem of encryption that needs to be overcome such that acquired evidence is not effectively a forensic ‘image’ of encrypted data (Grispos, Storer, & Glisson, 2012). Cloud service providers such as Amazon, Google and Dropbox are offering alternatives to traditional file storage, email and collaboration solutions. As these options continue to become available, the likelihood of these environments being investigated increases. The modification of existing tools or the development of new tools to capture cloud data is necessary to ensure the reliability and integrity of the data collection process (Grispos, Storer, & Glisson, 2012).

The rapid development in technology associated with an increase in cybercrime in the cloud can, in part, be connected to the widespread availability of low-cost mobile devices such as smartphone's and tablets. These devices provide ubiquitous access to email and file sharing, with onboard specifications and features exceeding traditional desktop/laptop systems and the technology will only continue to improve dramatically over the coming years. Although, the growth of cloud computing comes as no surprise, the versatility of private, public and hybrid cloud environments continue to challenge digital forensic practitioners and LEA's, specifically how to access and isolate evidentiary data using conventional digital forensic methods and techniques. As identified in Section 2.1, the ad-hoc approach to the development and implementation of cloud standards/framework and the fragmented approach to formalising cloud forensic procedures will continue to impede the efficiency and effectiveness of the investigative process for cloud digital forensic practitioners and LEA's. However, future research into

cloud forensics is underway and a new research field, called triage, which is based on Machine Learning theory and has two main applications, 'live' and 'post mortem', that will rank groups of artefacts and quickly identify the most relevant ones from the crime's perspective. Future work is to analyse user artefacts and user CSP interactions with Machine Learning algorithms with the aim of matching well-known crime related patterns (Martini and Choo, 2014).

Evidence legality and industry standards relating to digital forensics in the cloud are central pieces of a larger framework that need to be addressed for future research. Although, industry standards such as the Distributed Management Task Force (DTMF) and the Institute of Electrical, Electronics Engineers (IEEE), NIST and McKemmish provide industry standards for cloud computing there is no international standard model that addresses the complexities of cloud digital forensic investigations. Future research should focus on reviewing all of the current and proposed standards to develop a comprehensive standard that can be applied to digital forensic investigation in the cloud. The standard should address ethical and legal issues such as security, privacy and integrity. The standard should also address jurisdictional legislation and digital forensic investigation laws across international borders.

In conclusion, the 'Reliability and Integrity of the Evidence Gathering Process' during a digital forensic in a cloud can only be fully realised with the cooperation and inclusion of the international digital forensic community in the development and implementation of industry binding standards.

Reference List

(2012, June) Institute of IT Professionals New Zealand. New Zealand Cloud Computing Code of Conduct. National Institute of Standards and Technology. Retrieved February 22, 2014, from http://www.nzcloudcode.org.nz/

2012 Third International Conference on Emerging Intelligent Data, & Web Technologies. (2012, September). *Cloud Forensics: Concepts, Issues, and Challenges* (978-1-4673-1986-7). Bucharest, Romania: IEEE.

Ademu, I. O., Imafidon, C. O., & Preston, D. S. (2011). A New Approach of Digital Forensic Model for Digital Forensic Investigation. *International Journal of Advanced Computer Science and Applications, 2*(12), 175-178.

Alali, F. A., & Yeh, C. L. (2012). Cloud Computing: Overview and Risk Analysis. Journal of Information Systems, 26(2), 13-33.

Amazon. (2015). *Cloud Products*. Retrieved August 21, 2015, from Amazon Web Services Web site: https://aws.amazon.com/products/?nc1=f_cc

Barrett, D., & Kipper, G. (2010). Virtualisation and Forensics - A Digital Forensic Investigator's Guide to Virtual Environments (S. Liles, Ed.). Burlington, USA: Elsevier.

BeeGFS. (2015). *BeeGFS The Parallel Cluster File System*. Retrieved April 6, 2015, from BeeGFS Web site: http://www.beegfs.com/content/

Birk, D. (2011). *Technical Challenges of Forensic Investigations in Cloud Computing Environments*. Retrieved August 23, 2014, from IBM Web site: http:// www.ibm.com/Search/?q=csc2011&co=us&lo=any&ibm-submit.x=0&ibm-submit.y=0&sn=&lang=en&cc=US&en=utf&hpp=

Buffington, J. (2014, June). Why Store Once Federated Duplication Matters to HP — and Should to You, Too. In *Enterprise Strategy Group*. Retrieved August 21, 2014, from Search Storage Web site: http://searchstorage.bitpipe.com/data/ demandEngage.action?resId=1402744421_723

Chantry, D. (2009, January). *Mapping Applications to the Cloud*. Retrieved August 28, 2014, from Microsoft Developer Network Web site: http://msdn.microsoft.com/en-us/library/dd430340.aspx

Chunga, H., Parka, J., Leea, S., & Kang, C. (2012). Digital forensic investigation of cloud storage services. *Elsevier Digital Digest, 9*(2), 81-95. Retrieved February 25, 2015, from ScienceDirect Web site: http://www.sciencedirect.com/science/article/pii/S1742287612000400

Clarke, N., & Tryfonas, T. (2011). *Proceedings of the Sixth International Workshop on Digital Forensics and Incident Analysis (WDFIA 2011)*. Lulu.com

Cloud Computing Standardisation Initiatives: State of Play [Special section]. (2013). *International Journal of Cloud Computing and Services Science, 2*, 351-362.

Cloud Security Alliance. (2013, February). The Notorious NineCloud Computing Top Threats in 2013. In *Top Threats Working Group* (Top Threats Working Group, pp. 6-21). Retrieved October 16, 2014, from Cloud Security Alliance Web site: https://downloads.cloudsecurityalliance.org/.../The_Notorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf

Contrail Consortium. (2014). *Contrail: Open Computing Infrastructures for Elastic Services*. Retrieved February 25, 2015, from Contrail Web site: http://contrail-project.eu/tech

Cuthbertson, S. (2013). Mutual assistance in criminal matters: Cyber world realities. *Elsevier Digital Investigation*, 127-142. Retrieved February 26, 2015, from ELSEVIER Web site: http://www.elsevier.com/online-tools/scopus/SCOPUS-services/integration

Dykstra, J., & Sherman, A. (2011). Understanding Issues in Cloud Forensics: Two Hypothetical Case Studies. *Adfsl Conference on Digital Forensics, Security and Law*, 45-54. Retrieved December 16, 2014, from Retrieved from http://ezproxy.aut.ac.nz/login?url=http://search.proquest.com/docview/884340731?accountid=8440

Dykstra, J., & Sherman, A. (2012). Acquiring Forensic Evidence from Infrastructure-as-a-Service Cloud Computing: Exploring and Evaluating Tools, Trust, and

Techniques. Digital Investigation(9), S90-S98. Retrieved August 12, 2014, from Elsevier Web site: www.elsevier.com/locate/diin

Dykstra, J. (2012, June 26). *Digital Forensics for IaaS Cloud Computing*. Retrieved November 12, 2015, from Digital Forensics and Incident Response Web site: https://digital-forensics.sans.org/summit-archives/2012/digital-forensics-for-iaas-cloud-computing.pdf

Fourth Amendment Search and the Power of the Hash - Second Circuit Creates A Potential "Right To Deletion" of Imaged Hard Drives. (2014). *Harvard Law Review Forum, 128*(2), 743-750. Retrieved from Harvard Law Review Web site: http://harvardlawreview.org/2014/12/united-states-v-ganias/

Galante, J., Kharif, O., & Alpeyev, P. (2011, May 17). In A. Palazzo, Y. Cho, & T. Giles (Eds.), *Sony Network Breach Shows Amazon Cloud's Appeal for Hackers*. Retrieved April 22, 2014, from Bloomberg Web site: http://www.bloomberg.com/news/2011-05-15/sony-attack-shows-amazon-s-cloud-service-lures-hackers-at-pennies-an-hour.html

Gartner IT Glossary. (n.d.). Retrieved October 16, 2014, from Gartner Web site: http://www.gartner.com/it-glossary/multitenancy

Gluster. (2015, March 24). *Write once, read everywhere*. Retrieved April 6, 2015, from Gluster: https://forge.gluster.org/

Golomb, G. (2013, April 24). Uncommon Event Log Analysis for Incident Response and Forensic Investigations [Website]. Available January 27, 2016, from Cylance: https://blog.cylance.com/blog/bid/297047/Uncommon-Event-Log-Analysis-for-Incident-Response-and-Forensic-Investigations

Grispos, G., Storer, T., & Glisson. W.B. (2012). Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics. *International Journal of Digital Crime and Forensics, 4*(2),. Retrieved October 21, 2015, from Cornell University Web site: http://arxiv.org/ftp/arxiv/papers/1410/1410.2123.pdf

Guilloteau, S., & Mauree, V. (2012, March). *Privacy in Cloud Computing*. Retrieved August 23, 2015 from ITU Web site: http://www.itu.int/en/ITU-T/techwatch/Pages/cloud-computing-privacy.aspx

Hogan, M., Liu, F., Sokol, A., & Tong, J. (2011). *NIST Cloud Computing Standards Roadmap – Version 1.0* (Special Publication 500-291, pp. 10-76). Gaithersburg, MD: U. S. Department of Commerce.

Hooper, C., Martini, B., &Choo, K-K. (2013). *Computer Law & Security Review: Cloud computing and its implications for cybercrime investigations in Australia* (2nd ed., Vol. 9). Retrieved February 5, 2014, from ScienceDirect Web site: http://www.sciencedirect.com/science/article/pii/S0267364913000241

Huth, A., & Cebula, J. (2011). The Basics of Cloud Computing. In *US-Cert*. Pittsburgh, PA: Carnegie Mellon University.

ICT.govt.nz. (2014, May 15). *Programmes and Initiatives: Cloud Programme*. Retrieved May 19, 2014, from ICT.govt.nz Web site: http://ict.govt.nz/programmes-and-initiatives/cloud-programme/

IDC. (2013, February 28). *IDC - Press Release*. Retrieved April 20, 2013, from IDC Analyse the Future Web site: http://www.idc.com/getdoc.jsp?containerId=prUS23972413

IEEE. (2011). Technical Issues of Forensic Investigations in Cloud Computing Environments. *Systematic Approaches to Digital Forensic Engineering (safe), 2011 Ieee Sixth International Workshop* (10.1109/SADFE.2011.17), 1-10. Retrieved February 25, 2015, from IEEE Xplore Web site: http://ieeexplore.ieee.org/xpl/articleDetails.jsp? arnumber=6159124

Inktank Storage, Inc. (2015). *CEPH: The Future of Storage*. Retrieved April 6, 2015, from Ceph Web site: http://ceph.com/

Institute of IT Professionals NZ Inc. (2012). *New Zealand Cloud Computing Code of Practice* (Version 1.01). New Zealand: Institute of IT Professionals NZ

Kielmann, T., Pierre, G., & Morin, C. (2010). XtreemOS: a Sound Foundation for Cloud Infrastructure and Federations. In F. Desprez, V. Getov, T. Priol, & R. Yahyapour (Eds.), *Grids, P2P and Services Computing* (pp. 1 - 5). Retrieved February 25, 2015, from SpringerLink Web site: http://link.springer.com/chapter/10.1007%2F978-1-4419-6794-7_1

Kothari, C. R. (2006). *Research Methodology: Methods and Techniques* (2nd ed.). Delhi, India: New Age International (p.3) Limited, Publishers. (Original work published 2004)

Krutz, R., & Vines, R. (2010). A Comprehensive Guide to Secure Cloud Computing. Indianapolis, IN/USA: Wiley Publishing, Inc. (Original work published 2010)

Lawton, G. (2011, January). *Cloud Computing Crime Poses Unique Forensics Challenges*. Retrieved May 8, 2014, from TechTarget Web site: http:// searchcloudcomputing.techtarget.com/feature/Cloud-computing-crime-poses-unique-forensics-challenges

Lillard, V. T., Garrison, G. P. ' Schiller., C.A., & Steele, J. (2010). *Digital Forensics for Network, Internet and Cloud Computing*. Burlington, USA: Elsevier.

Martini, B., & Choo, K-K. (2012). An integrated conceptual digital forensic framework for cloud computing. *Elsevier Digital Digest*, *9*(2), 71-80. Retrieved December 9, 2014, from ScienceDirect Web site: http://www.sciencedirect.com/science/article/pii/S174228761200059X

Martini, B., & Choo, K-K. (2013). Cloud storage forensics: ownCloud as a case study. *Elsevier Digital Investigation*, *10*(4), 287*299. Retrieved December2, 2014, from ScienceDirect Web site: http://www.sciencedirect.com/science/article/pii/S1742287613000911

Martini, B., & Choo, K-K. (2014). Distributed Filesystem Forensics: XtreemFS as a Case Study. *Digital Investigation*, 295-313. Retrieved December 2, 2014, from Elsevier Web site: http://www.elsevier.com/locate/diin

Marturana, F., Tacconi, S., & Me, G. (2012). A case study on digital forensics in the cloud. *Ieee Computer Society*(DOI 10.1109/CyberC.2012.26), 111-116. Retrieved December 2, 2014, from IEEEXplore Web site: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6384935

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud Security and Privacy* (1st ed.) (M. Loukides, Ed.). Sebastopol CA, USA: O'Reilly Media, Inc.

McKemmish, R. (1999, June). What is Forensic Computing? In *Australian Institute of Criminology* (118, pp. 1-6). Retrieved February 26, 2015, from Australian

Institute of Criminology Web site: http://www.aic.gov.au/media_library/publications/tandi_pdf/tandi118.pdf

Montalbano, E. (2011, June 2). *Public Sector Slow To Adopt Cloud Computing*. Retrieved April 18, 2013, from InformationWeek Government Web site: http://www.informationweek.com/government/cloud-saas/public-sector-slow-to-adopt-cloud-comput/229900072

Nepal, S., & Pathan, M. (Eds.). (2014). Security, Privacy and Trust in Cloud Systems. Retrieved August 21, 2014, from Springer Web site: http://www.springer.com/engineering/signals/book/978-3-642-38585-8

Openwall. (n.d.). *Portable PHP password hashing framework*. Retrieved February 25, 2015, from Openwall Web site: http://www.openwall.com/phpass/*ownCloud Get Started*. (2013). Retrieved February 26, 2015, from ownCloud Web site: https://owncloud.org/install/#desktop

Ortiz, S., Jr. (2011, November). The Problem with Cloud-Computing Standardisation. *Technology News* (0018-9162), 13-16.

Pascal Junod. (2012, May 24). *OwnCloud 4.0 and Encryption*. Retrieved February 25, 2015, from Pascal Junod Web site: http://crypto.junod.info/2012/05/24/owncloud-4-0-and-encryption/

Quick, D., & Choo, K. (2013). Dropbox Analysis: Data Remnants on User Machines. *Elsevier: Digital Investigation, 10*(1), 3-18. Retrieved December 15, 2014, from ScienceDirect Web Site: http://www.sciencedirect.com/science/article/pii/S174228761300011X

Rivera, J., & van der Meulen, R. (2013, October 1). *Gartner Says Nearly Half of Large Enterprises Will Have Hybrid Cloud Deployments by the End of 2017*. Retrieved May 19, 2014, from Gartner Newsroom Web site: http://www.gartner.com/newsroom/id/2599315

Ruan K., Kechadi T., & Crosbie M. (2011). Cloud Forensics: An Overview. Dublin, Northern Ireland: Centre for Cybercrime Investigation, University College Dublin. Retrieved May 18, 2013, from Web site:

http://www.cloudforensicsresearch.org/publication/Survey_on_Cloud_Forensics_and_Critical_Criteria_for_Cloud_Forensic_Capability_6th_ADFSL.pdf

Sammes, A. (Series Ed.), & Antonopoulos, N., & Gillam, L. (Vol. Eds.). (2010). *Cloud Computing Principles, Systems and Applications*. Retrieved August 23, 2014, from Springer Web site: http://www.springer.com/computer/communication+networks/book/978-1-84996-240-7

Savage, S. (2013, June). Championing an Open Environment. *IBM Systems*, 1-3. Retrieved July 5, 2014, from IBM Systems Magazine Web site: http://www.ibmsystemsmag.com/linuxonpower/Trends/Open-Source/openstack_foundation/?page=1

Siron, E. (2012, July 24). *Hyper-V Hub*. Retrieved October 12, 2015, from Altaro Web site: http://www.altaro.com/hyper-v/an-overview-of-hyper-v-event-logs/

Stender, J., Berlin, M., & Reinefeld, A. (2013). XtreemFS: A File System for the Cloud. In D. Kyriazis, A. Voulodimos, S. Gogouvitis, & T. Varvarigou (Eds.), *Data Intensive Storage Services for Cloud* (pp. 267-285). Retrieved February 27, 2015, from IGI Global Web site: http://www.igi-global.com/chapter/xtreemfs-file-system-cloud/77442

Sullivan, J. (2014). Vincent Mosco, To the Cloud: Big Data in a Turbulent World [Review of the book *To the Cloud: Big Data in a Turbulent World*]. *International Journal of Communication* (8), 284. Retrieved June 23, 2014, from International Journal of Communication Web site: http://ijoc.org/index.php/ijoc/article/view/3173/1217

Szefer, J., Keller, E., Lee, R. B., & Rexford, J. (2011). Eliminating the Hypervisor Attack Surface for a More Secure cloud (Master's thesis, Princeton University, 2011). *Masters Abstracts*. Retrieved August 21, 2014, from Princeton University Department of Computer Science Web site: https://www.cs.princeton.edu/~jrex/papers/ccs11.pdf

Taylor, M., Haggerty, J., Gresty, D., & Lamb, D. (2011). Forensic investigation of Cloud Computing Systems. *Network Security*, 5. Retrieved August 21, 2014, from

ScienceDirect Web site: http://www.sciencedirect.com/science/article/pii/S1353485811700241

Thethi, N., & Keane, A. (2014). *Digital Forensics Investigations in the Cloud* (2014 IEEE International Advance Computing Conference (IACC), pp. 1475-1480). Retrieved December 15, 2014, from IEEEXplore Web site: http://ieeexplore.ieee.org.ezproxy.aut.ac.nz/xpls/abs_all.jsp?arnumber=6779543&tag=1

Thorpe, S. (2012). An Experimental Survey Towards Engaging Trustable Hypervisor Log Evidence Within a Cloud Forensic Environment. International Journal of Computer Science & Information Technology, 4(6), 125-141.

Trend Micro. (August, 2011). Security Threats To Evolving Data Centres. In Virtualization and Cloud Computing. Retrieved from Trend Micro Web site: http://www.trendmicro.co.nz/cloud-content/us/pdfs/security-intelligence/reports/rpt_security-threats-to-Datacentres.pdf

Uden, L., Herrera, F., Pérez, J., & Rodríguez, J. (Eds.). (2013). *Advances in Intelligent Systems and Computing - 7th International Conference on Knowledge Management in Organizations: Service and Cloud Computing* (172nd ed.). New York: Springer. (Original work published 2012)

United Nations Office on Drugs, & Crime. (February 9, 2007). *Strategy for the period 2008-2011 for the United Nations Office on Drugs and Crime* (E/CN.7/2007/14, E/CN.15/2007/5, p. 5). Retrieved February 25, 2015, from UNODC Web site: http://daccess-dds-ny.un.org/doc/UNDOC/GEN/V07/806/72/PDF/0780672.pdf?OpenElement

VMware Inc. (2007, September). *Understanding Full Virtualisation, Para-Virtualisation and Hardware Assist* (WP-028-PRD-01-01). Palo Alto, CA/USA: VMware Inc.

VMWare. (2015). *VMware Workstation 5.5 What Files Make Up a Virtual Machine?* Retrieved August 23, 2015, from VMWare Web site: https://www.vmware.com/support/ws55/doc/ws_learning_files_in_a_vm.html

Zatyko, K., & Bay, J. (2011, December 14). The Digital Forensics Cyber Exchange Principle. *Forensics on the Scene and in the Lab Magazine*. Retrieved February 25, 2015, from Forensic On The Scene and in The Lab Web site: http://

www.forensicmag.com/articles/2011/12/digital-forensics-cyber-exchange-principle