On Digital Forensic Readiness in the Cloud Using a Distributed Agent-Based Solution: Issues and Challenges

[1,2]Victor R. Kebande[*], [1]H.S.Venter[†]

Department of Computer Science, University of Pretoria,

Private Bag X20, Hatfield 0028, Pretoria, South Africa.

Email: vkebande@cs.up.ac.za[*], hventer@cs.up.ac.za[†]

The need to perform digital investigations has over the years led to the exponential growth of the field of Digital Forensics (DF). However, quite a number of challenges face the act of proving – for purposes of Digital Forensic Readiness (DFR) – that an electronic event has occurred in cyberspace. The problem that this research addresses involves the challenges faced when an Agent-Based Solution (ABS) is used in the cloud to extract Potential Digital Evidence (PDE) for DFR purposes. Throughout the paper the authors have modified the functionality of an initially malicious botnet to act as a distributed forensic agent to conduct this process. The paper focuses on the general, technical and operational challenges that are encountered when trying to achieve DFR in the cloud environment. The authors finally propose a contribution by assessing the possible solutions from a general, technical and operational point of view.

Keywords: forensic science, digital forensic readiness, agent-based solution, botnet, cloud, challenges, issues.

1. Introduction

The increasing demand for computing, information dissemination and pervasiveness of the Internet has led to technological advancements in Information Technology (IT) which have enabled organisations to design, build and run their operations efficiently in the cloud environment. A new report by the International Data Corporation (IDC)[1] on the top ten predictions of 2014 indicated that worldwide IT spending would grow 5% per year to $2.1 trillion in 2014. On cloud computing, the IDC predicts that by 2017, 80%+ of new cloud applications will be hosted on six Platform-as-a-Service (PaaS) platforms[1]. These developments show that organisations have become agile with the use of cloud computing. However, this has in turn opened a new horizon for adversaries who have taken advantage of the escalating technology to perform breaches in security systems, and launch cyber threats and attacks.

Carrying out a Digital Forensic Investigation (DFI) process in the cloud during incident response faces multiple challenges. The primary challenge is how the data of vendors and consumers can be protected. Nevertheless, cloud consumers' data residing in the cloud is basically aggregated in multi-tenant environments and these environments are hardly ever shared by Cloud Service Providers (CSPs). According to Birk[2] on a technical aspect of forensic investigation in the cloud, the amount of Potential Digital Evidence (PDE) that can

be available to a DFI strongly diverges between the cloud service and deployment models. This serves as an indicator when compounding the hurdles while trying to achieve Digital Forensic Readiness (DFR) in the cloud environment. The cloud continues to grow enormously with major organisations preferring the usage of Virtual Machines (VMs), which have enabled them to venture into privately owned clouds. Major organisations have moved their applications, systems and data because of the economies of scale and scalability through centrally powerful and effective hosted virtual servers. While this has benefited some organisations by freeing 35 to 50% of operational and infrastructure resources [3], not enough proactive solutions to mitigate potentially inherent security risks have been enforced.

Cloud exploits against cloud consumers are major risks that are expected due to the open nature of the cloud. Moreover, the CSPs cannot mitigate this by leveraging the exploding number of users. In their previous work, Kebande & Venter[4] extrapolated that without modifying the functionality of existing cloud architectures, a modified form of a botnet acting as a distributed Agent-Based Solution (ABS) could be deployed within the cloud environment to forensically capture PDE for purposes of DFR. The captured information was digitally preserved to aid in the reactive process when conducting a DFI. Discounting that, the problems that this paper investigates are the issues and challenges faced when conducting DFR in the cloud environment using an ABS. In this context, an ABS is modified to act as a forensic client implemented in the cloud environment on a Software-as-a-Service (SaaS) platform. Further, the ABS is capable of capturing digital information for DFR purposes. The contribution of this paper is presented in four phases. Firstly we present a highlight of the prototype named Cloud Forensic Readiness Prototype (CFRP). Thereafter, we identify the challenges in the cloud as a result of using an ABS in a Cloud Forensic Readiness (CFR) approach; next, we match all the challenges to the related work on cloud forensic challenges and finally we propose possible high-level solutions to these issues and challenges from a general, technical and operational point of view.

The rest of the sections in the paper are structured as follows: The paper begins by describing the background of cloud computing, DF, DFR, and botnets in Section 2. Thereafter, Section 3 discusses motivation and the scope of the study while Section 4 discusses related work. This is then followed by a discussion on the CFR model and the CFRP prototype in section 5 and 6 respectively. Next, Section 7 discusses the general, technical and operational challenges of the model, as well as the proposed high-level solutions. This is followed by Section 8 that contains discussion of the concept. The paper concludes with Section 9 stating a conclusion and suggesting future work.

2. Background

This section gives an overview of the following topics: DF and DFR, cloud computing, motivation and related work. DF is discussed owing to the fact that whole research is focused on the scientific process of digital investigation. On the other hand, DFR which is a proactive process is discussed to show the need for pre-incident preparation and planning in the DF environment. The cloud is discussed because the whole process occurs within the cloud environment. Finally, related work is discussed to show the approaches that have previously been applied in DFR.

2.1. On digital forensics

At the first Digital Forensics Research Workshop (DFRWS) in 2001, DF as a science was defined by Palmer [6] as "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of Digital Evidence (DE) derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations" [6]. Furthermore, on the maturity of DF, Casey [7] describes DF science as "coming of age", which shows that a maturation has taken place in respect of the concepts and principles used in the field. The prevailing principles behind this are to use scientifically proven methods to maintain the continuity of evidence or the chain of custody.

2.2. On Digital Forensic Readiness

According to ISO/IEC 27043: 2015 [8], DFR is a proactive process that precedes incident detection and involves pre-incident planning within the DF circuit. Its implementation is still in a fledgling stage, but DFR is gaining a lot of interest and much research into the field is still needed. In a 10-step process for forensic readiness, Rowlingson [10] identifies the business requirements for gathering and using DE. He also argues that DFR as presented by Tan [11] has an objective of maximising the environment's capability of collecting digital forensic information, while minimising the cost of the forensic investigation during an incident response [10, 11]. On the policies that facilitate DFR, Yasinsac and Manzano [12] proposed the following: information retention; planning of the response; training; investigation acceleration; prevention of anonymous activities, and protecting the evidence. As a result, the paper in hand is inclined towards Tan's objectives[10] and the policies of Yasinsac and Manzanos[12]. Notwithstanding that, at the heart of security incidents lay a number of standardised guidelines that have been proposed on the preparation of incident response. Table 1 shows various standards and guidelines for security incidents.

Table 1.Proposed standards and guidelines for security incidents

	Proposed Standard	Target Year of Publication	Description
1	**ISO/IEC 270355**	2011	Information Technology-Security Techniques-Information incident management
2	**ISO/IEC 27001**	2013	Information Security Management
3	**ISO/IEC 27042**	2015	Guidelines for the Analysis and Interpretation of Digital Forensic Evidence
4	**ISO/IEC 27043**	2015	Security Techniques-Incident Investigation Principle and Processes
5	**ISO/IEC 27044**	2016	Guidelines for Security Information and Event Management (SIEM)
6	**ISO/IEC 30121**	Committee Draft	Governance of Digital Forensic Risk Framework
7	**ISO/IEC 27050**	Committee Draft	Electronic Discovery

ISO/IEC 27043:2015 standard proposed by Valjaveric and Venter [8] is the umbrella standard for high-level concepts of 27037, 27035, 27040, 27041, 27042, 27050. At the time when this paper was written, ISO/IEC 27043:2015 had already been published as an international standard for information technology, security techniques, incident investigation principles

and processes (see Table 1). In its incident investigative principles and processes, ISO/IEC 27043:2015[8] clearly define readiness through its classes of processes as a process that occurs before incident identification. This is represented as a proactive process of planning and preparing before incidents occur. The depth that exists on information security standards is immense and fortunately ISO/IEC 27043: 2015[8] open the readiness spectrum further. The next section discusses cloud computing.

2.3.On Cloud Computing

Cloud computing has taken the world by storm. The concept of the cloud is basically one of central remote servers that are used to maintain data and applications. This allows vendors, organisations and consumers to access data, applications and systems efficiently on any given computer. The biggest advantages of the cloud are its centralisation of data storage, processing capacity and bandwidth optimisation. All of these occur while the resources are being shared. The National Institute of Standards and Technology[13], NIST, has a standard definition of cloud computing (3) and defines it as "a model for enabling ubiquitous, convenient and on-demand network access to a pool of shared configurable resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [13]. Additionally, cloud computing operates under three service models: Software as a service (SaaS), Platform as a Service (SaaS) and Infrastructure as a Service (IaaS) [13].

While the cloud boasts of its scalability, flexibility and effectiveness, numerous vulnerabilities and potential risks exist that make it susceptible to digital investigations. The presence of insecurity in the cloud is confirmed by Gartner's research[14], which states that "from 2012, 60% of virtualised servers will be less secure than physical servers they replace, dropping to 30% by 2015" [14].

Cloud forensics itself has been coined as a cross discipline of cloud and digital forensics which exists as a subset of network forensics according to Ruan et al[15]**.** On the same note, a number of cloud forensic frameworks have been proposed before that have further been able to be validated using a significant number of cloud solutions. For example, a proposal for an integrated conceptual framework for cloud computing by Martini & Choo[16] emphasized on how digital data is preserved and collected in the cloud for digital forensic purposes. Nevertheless research on artefact collection, acquisition, preservation and timestamp collection from cloud computing devices for forensic purposes has also been highlighted by Quick & Choo [17-26] which has presented better ways of identifying digital remnants in computing devices and cloud environments. Additionally more research on cloud forensics presented by Quick & Choo[19]; Chung et al.[24] ; Marty [25] and Dykstra & Sheman [26] has focused on the technical perspective regarding remote extraction of digital data , forensic logging and ways of collecting digital data from cloud storage products.

Discounting the above, extracting evidentiary data for forensic readiness purposes in the cloud environment without modifying the existing architecture is a challenge in itself, due to the existence of multiple data centres. In order to gain more insight into this, the next two subsections discuss botnets, motivation of the research study and related work.

2.4. Botnets

A botnet or robot network is a generic term that describes a set of scripts written to perform systematic predefined functions. Bot itself is derived from "ro-bot", in this perspective bot represents the set of commands. It is a group or network of computers which have been infected with a type of software which allows an operator to control the infected systems or zombies.

Additionally, botnets constitute a malicious network of infected hosts under the command of a botnet operator over the Internet that works in the following manner: Bots are usually known to spread themselves across the Internet by means of searching unprotected and vulnerable computers. They are then able to infect via malicious electronic mail, through attachments or visited pages in a website. After infection the bot is able to report to the Command and Control (C&C) channel. In this context, the role of the C&C is to allow the botnet operator to communicate and give commands and direct the actions of the botnet. This is done using Hypertext Transfer Protocols (HTTP), Internet Relay Chat (IRC), Peer to Peer (P2P) or any other suitable botnet administration protocols. Through the C&C server the botnet operator is able to send commands to the bot. Thereafter, the botnet operator is able to gain an army of bots that he is able to control from one single point. In fact Ollman[27] highlights that C&C allows a bot agent to receive new instructions and malicious capabilities.

A number of botnets are able to spread by automatically exploiting vulnerabilities, however, others spread through the downloading of malicious attachments. This happens through its scanning of hosts for vulnerable services, installing itself in stealth mode, extracting digital information illegally and reporting back to the botnet operator. Generally the intention of a botnet operator is to collect digital information on a continuous basis. Falliere and Chien [28] refer to the example of Zeus bot, which is a dangerous information-stealing toolkit that was used to steal online credentials, gathering system information, stealing protected storage information and sending it to the command and control server. In sharp contrast to these cynical activities, our research is to comprehensively optimise the usage of botnets in a non-malicious fashion for the purpose of DFR investigations.

3. Motivation and scope

The cloud has not fully adapted to traditional DFI processes because potential evidential data is distributed and there is still lack of standardised processes. In fact the complexity of the cloud has given rise to many open issues like inability to conduct digital investigation due to the difficulty involved while trying to gain physical access and inability of the Law Enforcement Agencies (LEA) to trace the provenance of a digital object. ISO/IEC 27043: 2015[8] highlights the standardised Digital Forensic Readiness Investigation Process (DFRIP) that can be used to conduct DFR in a given organisation without the disruption of any business processes [8], however ISO/IEC 27043 is not focused on the cloud. Notwithstanding that, using an Agent-Based Solution (ABS) to conduct DFR in the cloud has been motivated by the fact that there is no alteration or modification of the of the functionality of existing cloud architecture even though the scope of the study show the manner or capability through which DFR can be conducted in the cloud with a focus on SaaS. At the outset of conducting DFR, the Cloud Service Providers (CSPs) will be assured that the process will save money and time because of lack of having to reprogram the infrastructure time and again.

4. Related Work

In this section we present work that is somewhat related to our research on challenges and DFR approaches. To begin with, a survey on information security incident handling for mitigating risks to confidentiality, integrity and availability in the cloud environment by Rahman & Choo[29] presented different methods of handling incidents in digital forensics and the existing gaps in the cloud environment were identified. Moreover, the authors were able to identify clouds organisational data as one of the challenges encountered while handling security incidents and a comparative summary of different international incident handling models were presented. A final study on the same paper provides a summary of cloud security

challenges. While this research work on cloud forensic challenges was very informative, its major focus was mainly on incident handling mechanisms.

Nevertheless, another paper aimed at addressing cloud forensic technical challenges and solutions by Martini & Choo[30] reviewed various prominent technical publications that were aimed at providing conceptual solutions to Cloud Service Providers (CSPs) with better systems for forensic evidence collection while implementing them in the cloud environment. On the same note a conceptual framework to integrate DF tools into different ways of developing a cyber-physical cloud that was aimed at helping organisation to recover from cyber-physical attacks addressed the following factors: Risk management, forensic readiness, incident handling, laws and regulations, hardware and software requirements and industry specific requirements. The main need for this framework was to point out cloud specific forensic challenges like multi-jurisdictional, multiple versions and data extraction issues as highlighted by Rahman, Glisson, Yang & Choo[31].

Research on challenges in digital forensics by Vincze [32] has identified the following operational challenges: Diversity, scale and cloud resources, digital evidence seizure, privacy, hiring, training and development. In this research the author categorically identifies the major challenges in DF but also acknowledges that there is still more work to be done on the same. Also, Simou et al.[33] has addressed major forensic challenges and issues of the cloud from a review perspective and using a model. The authors were able to categorize identified challenges to the following stages that are applicable to IaaS, PaaS and SaaS: Identification, preservation-collection, Examination-analysis, presentation and uncategorized. Nevertheless, research by Delport, Olivier and Kohn[34] has proposed an isolation of a cloud instance through instance relocation, server farming address relocation, failover and sandboxing in order to prevent contamination, tampering and losing instances of possible digital forensic evidence.

On the same note, after a survey of existing literature on draft NISTIR 8006[5], the NIST cloud computing forensic science working group (NCC FSWG) documented a list of challenges in cloud computing environments [5]. In this document NIST listed 65 challenges related to cloud forensics, based on a normalised formula of four variables. These variables include the following:

- Stakeholders (noun) – this variable identifies the affected stakeholders by the challenge that has been identified. Examples include first responders, investigators and cloud consumers[5].
- Action (verb) – This represents the activities that the stakeholder intends to do, for example gaining access, imaging and decrypting[5].
- Object – This identifies the specific item on which the action is to be performed, for example data, audit logs, evidence and time stamps[5].
- Reason – This refers to the primary challenges that the stakeholder faces so that he can perform the specific action on the object [5].

Throughout this research and based on the normalised formulae, NIST specifically identified the general challenges without proposing any solutions.

Nevertheless, whilst these challenges are documented well, the authors realised that implementing an ABS in the cloud poses many challenges in respect of which we provide our thoughts on possible high-level solutions later in this paper. The next instances of related work do not deal with challenges but present related work on DFR. They are included because they present a DFR approach and we also make use of a DFR approach.

In a Harmonised Digital Forensic Investigation (HDFI) model that was proposed by

Valjarevic and Venter [35], the authors highlight the need for DFR phases before incident detection. The HDFI [35] is a vital comprehensive digital forensic investigation model that has already been published as ISO/IEC 27043: 2015 international standard [8]. We are adopting the forensic readiness processes described in the various classes of digital investigation process[8] to aid with the conducting of DFR in the cloud.

A prototype on DFR in the cloud environment developed by Trenwith and Venter [36] has the following requirements: Identification, collection, transportation, storage and examination. In this prototype there was a collection of potential digital evidence from each virtual machine in the hybrid cloud that was later used as an operating system (OS) Application Programming Interface (API) to conduct a backup. Furthermore, the prototype had a communication channel and implemented encryption, compression and authentication. It shortened the DFI process through data acquisition in a proactive process that involved a remote and a central evidence server – this portrayed the effectiveness of forensic readiness. The study by Trenwith and Venter [36] portrays a mechanism of conducting DFR in which the emphasis lies on identifying and collecting information. Our research attempts to do the same, however it uses an ABS to accomplish this. On the other hand, according to Cohen [37], it is important to identify and preserve relevant log files and audit data. In spite of that, Cohen[37] highlights that all these potential evidence should be linked to the servers used to send, receive, process and store the evidence. This is preferred in situations where PDE might be sought if an incident is detected. The next section gives a discussion on a model for conducting DFR in the cloud environment using an ABS.

5. A Cloud Forensic Readiness Model

In this section, we present a cloud forensic readiness (CFR) model as a block diagram. The model depicts a typical approach to DFR in the cloud environment [4]. Each of the constituting parts of the model is discussed in the subsection that follows.

5.1 Model Overview

The model is based on active monitoring, gathering and retaining of information over the network within a cloud environment. In contrast to that, throughout this novel concept an ABS is used in the cloud environment to harvest digital information, preserve it digitally in preparation for DFR investigations. Initially, we use ABS to denote an optimised tool (a modified version of a botnet) that is used to collect digital evidence, but when it is implemented in the cloud delivery model as Software-as-a-Service (SaaS), we are able to coin it as a new cloud service the term i.e. Agent Based-Solution-as-Service (ABSaaS). It is therefore worth noting that this is one novelty emerging from this research.

This novel concept can only be achieved by deploying the ABS to install in the Virtual Machines(VM) of computers in any cloud environment so as to harvest digital information [4]. By means of this proactive approach, potential evidence is gathered forensically in readiness processes as explained in ISO/IEC 27043: 2015[8]. This process is shown in Figure 1.

The underlying assumption in this model is that *N* numbers of VM within a given cloud are being provided services by the CSPs. Sensitive and critical information can be gathered in and from the cloud through the ABS installed in the *N* number of virtual machines(see figure 1). The gathered information is stored as data that represents PDE. During preservation, the integrity of the collected data is maintained by means of hashing PDE before it is exposed to DF investigators and Law Enforcement Agencies (LEA). This is illustrated by the block

diagram in Figure 1, which shows the composition of the CFR model. Through the gathering of potential evidence, the effort required in performing a digital forensic investigation is reduced because of the availability of captured, preserved and hashed logs as digital evidence. An explanation of the interaction of the model components follows next.

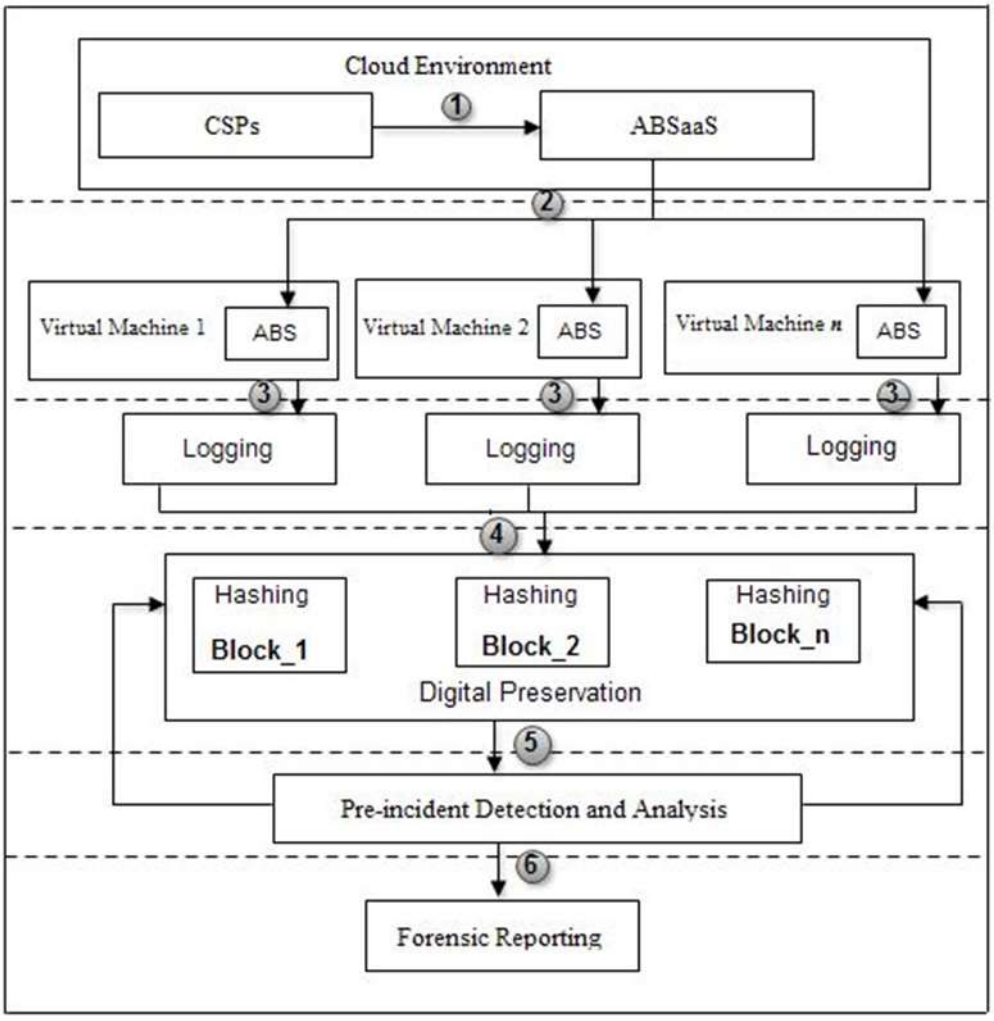

Figure 1. A block diagram for CFR Model

5.2. Operation of the CFR Model

In this section we present a discussion on the interaction between CFR components and analyse the mechanism that the CFR model uses to conduct its operations. The description highlights the proactive process in the cloud environment.

The CFR model shown in Figure 1 considers the cloud environment, which consists of CSPs and the ABSaaS. The CSPs provide virtual services to cloud users whereas the ABSaaS is implemented as a (SaaS) delivery model. This is shown by the arrow pointing from the CSPs to ABSaaS box labelled 1. The ABSaaS is installed in the virtual instances as a service within a network. This has been shown in the next step.

In the step labelled 2, the CFR model encompasses *N* number of VMs. In this step the ABSaaS is installed in the VMs with ABS, as is shown by the arrows pointing inside the VMs. Thereafter, in the step labelled 3, the ABS proactively capture distributed acute traffic in the cloud that moves as possible attack logs in the logging process. The ABS is able to gather both

non-volatile and volatile digital data from target VMs and from the network in the cloud. The captured potential evidence are categorised as monitored data, service artefacts and forensic evidence. All the captured PDE are sent to a centralised digital preservation centre for evidence management in the next step. Collected large-scale evidence according to Kebande and Venter [38] is analysed through an efficient and timely computation that uses MapReduce that is capable of forensically reducing analysis time of large-scale evidence.

In the step labelled 4 (with the arrows pointing downwards), the captured potential evidence is digitally preserved through hashing. Hashing is performed as a block of hashes to potential evidence to maintain the integrity of the data, as shown, using *block 1, block 2….block n*. In this context, a block is a set of captured potential evidence. A comparison of cryptographic hash functions is performed on the forensically logged data to ensure that its integrity is maintained and to help in proving the correctness of the captured forensic data. The significance of digital information preservation is to ensure changes are not made to the gathered potential evidence. PDE to be used for DFR has to be retained in its original form.

This is followed by the step labelled 5. In this context, pre-incident detection is used as a process that prepares possible potential evidence to be used for DF investigation if an incident is detected. On the other hand, pre-incident analysis is a process of extracting exact evidence that may be admissible in a court of law if an incident is identified. During pre-incident analysis, the originality of potential evidence is proved by comparing the hash of the preserved data. This is shown using two-sided arrows pointing to the part labelled step 5.

The final part of the CFR model is the forensic reporting phase. This has been shown by the part labelled 6 with an arrow pointing downwards. Forensic reporting is a crucial stage during which a DF evidence examiner needs to recognise a roadmap of the steps undertaken during the proactive process before evidence is considered as admissible. This is the final stage and whatever is provided in this case has to be used by DF investigators and law enforcement agencies. This is shown with the arrow labelled 6 in Figure 1. According to Cohen [37], potential evidence provided at this stage has to be relevant to the matters that are at hand.

5.3. Summary of the CFR Model

Through the CFR model, an approach of DFR in the cloud is achieved through which critical and sensitive information is captured from the cloud environment by using an ABS.

The capturing of data can be done by modifying the existing cloud infrastructure such that new modules or software code is introduced. However, the latter approach is costly since modification of the cloud infrastructure is then required. Therefore, our approach of introducing an ABS does not require the modification of the functionality of existing cloud infrastructure, while we are still able to capture the required digital data within the cloud using an ABS.

Other research papers [5, 17-26, 30-33] have also identified numerous challenges that are similar to the challenges arising from utilizing ABSs. In other words, the preliminary work did not identify these challenges by using ABSs per se' but rather identified general challenges, of which some appear to be similar than the challenges arising from ABSs [5].

We have systematically done a comparative study with other literature[5, 17-26, 30-33] on cloud forensic challenges and the challenges we identified, as well as proposed high-level solutions. The next section presents technical experiments that the authors conducted using a prototype.

6. Prototype

The previous sections have introduced the reader to the CFR model, operation and a brief summary. In this section we introduce the prototype that the authors developed for purposes of gathering PDE from the cloud environment.

6.1. Defining the prototype

To accomplish the purposes for which the study is designed for, we developed a tool that is able to collect digital forensic evidence from a virtualised environment that may be used during post-event response if a security incident is detected. The prototype has been implemented in Linux-based environment and using Open Peer to Peer (OpenP2P) protocol. The prototype is namely Cloud Forensic Readiness Prototype (CFRP) which uses a Non-Malicious Botnet (NMB) with modified functionalities that operates as a distributed agent-based solution (ABS). The CFRP performs the following functions:

- Pings the server *n* times
- Collect potential digital evidence
- Transmits collected evidence using openP2P
- Record time stamps
- Hash the collected information
- Posts it back to the server
- Updates and deletes the ABS frequently

6.2. Prototype Architecture

The CFRP prototype is built in client-server network architecture, the distributed forensic agent acts as a forensic clients and the C&C is tasked with distributing new commands, receiving and storing information as evidence. The following third party libraries were used to allow efficient implementations: Boost, OpenP2P, Linux headers and Easylogging++. Figure 2 shows the prototype architecture.

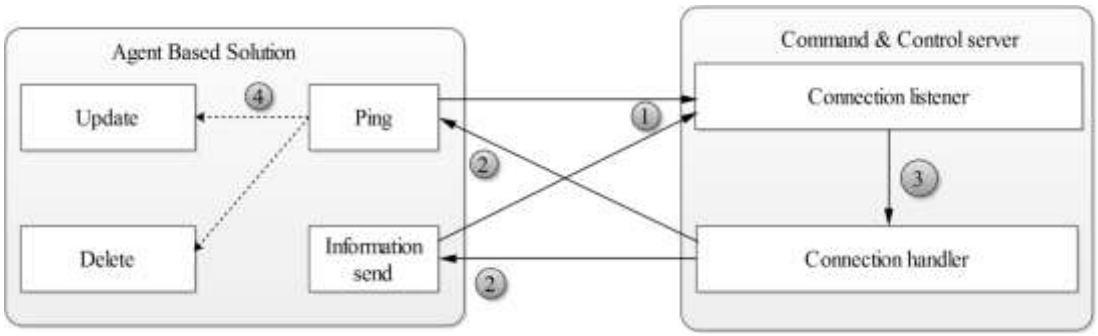

Figure 2. Prototype Architecture

Figure 2 shows the architecture of the CFRP. The ABS connects to the server in the labelled 1, thereafter the server receives a connection which it passes to a separate thread (connection handler) to handle which is labelled as number 3. Next the thread sends any information that is necessary to the ABS in the part labelled 2. After this process, the ABS component that receives this information passes it to another component if at all it's necessary in the section labelled 4.

6.3. Prototype Implementation

The focus of this implementation is to acquire forensic logs as PDE that can be used to support digital forensic investigations from the cloud environment using an ABS. Figure 3 shows the context in which the ABS is able to collect the forensic logs as PDE and post them to the Command and Control (C&C) server for analysis. This implementation has deployed an ABS as a logger in an OwnCloud environment, developed in Linux environment using C++ with the GNU being used in the build-up. The C&C server comprises of two main threads: The main server thread and the receiver thread. The main server thread listens to incoming connections and then passes them to receiver thread. Whenever an ABS is able to ping the server, the server is able to give relevant commands to the ABS which may be updating, deleting or acknowledging the request. On the other hand, the receiver thread performs operations like reading the ABS requests and checking whether the ABS is able to ping the server. The main thread server functionalities and the receiver threads are shown in Figure 3.

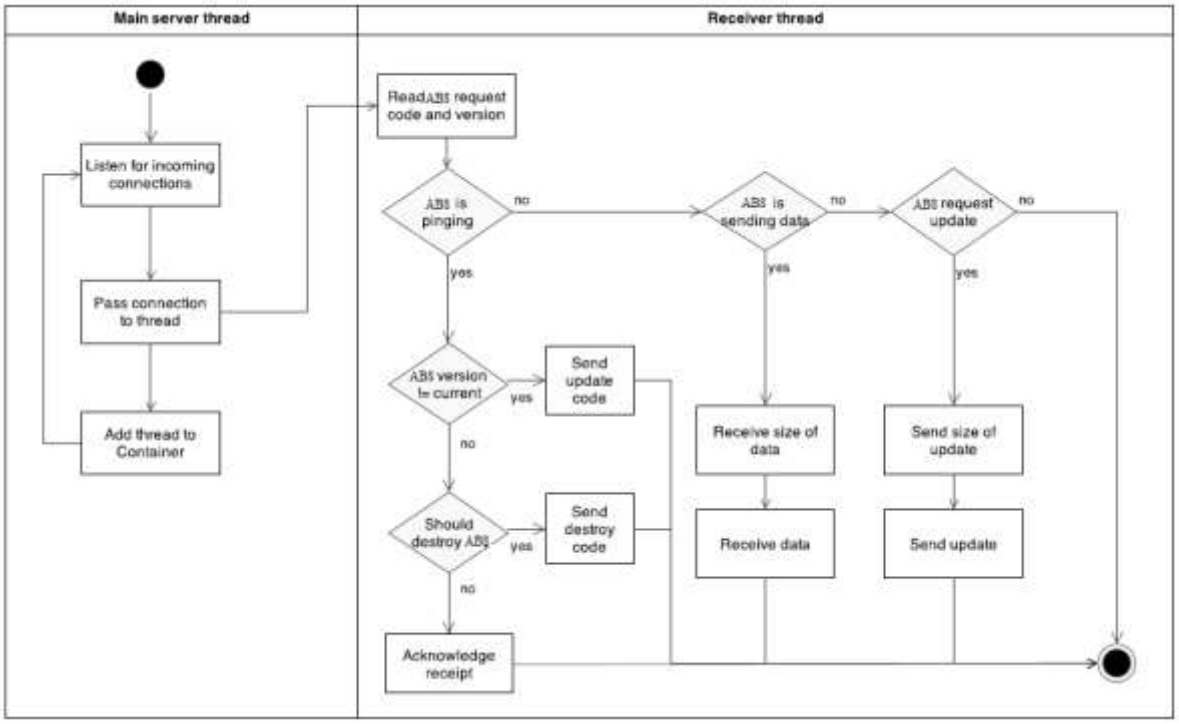

Figure 3. Command and Control server operation flow

Laboratory Environment and implementation

Various components were availed to set-up an environment for deploying the ABS: A forensic server, virtual target, command and control (C&C) center and physical target.

- Server: Comprises of a Linux-based environment on which the operator uses to send periodical pings *n* number of times. The server is directly connected to the internet.
- Virtual target: This is a Linux VM in the OwnCloud environment where ABS command is executed to collect potential digital evidence.

- Command & Control: C&C consist of connection handler and connection listener. It is used to pass commands for ABS execution.
- Physical target: A physical computer where the ABS is simulated.

Modifying the Botnet functionality to act as an ABS

The cloud-based botnet solution that has been modified to act as an ABS. It runs in the owncloud background of the system, gathers digital information and sends it back to the C& C server and it periodically checks additional commands from the server. Importantly, the collected information is digitally preserved and used for digital forensic readiness purposes. The modified botnet that act as an ABS has the following components: main thread, installer thread, ping thread, key-logging thread, update thread, delete thread and send data thread. Figure 4 shows the components flow (labeled 1-7) of the modified botnet that runs as an ABS in a single diagram.

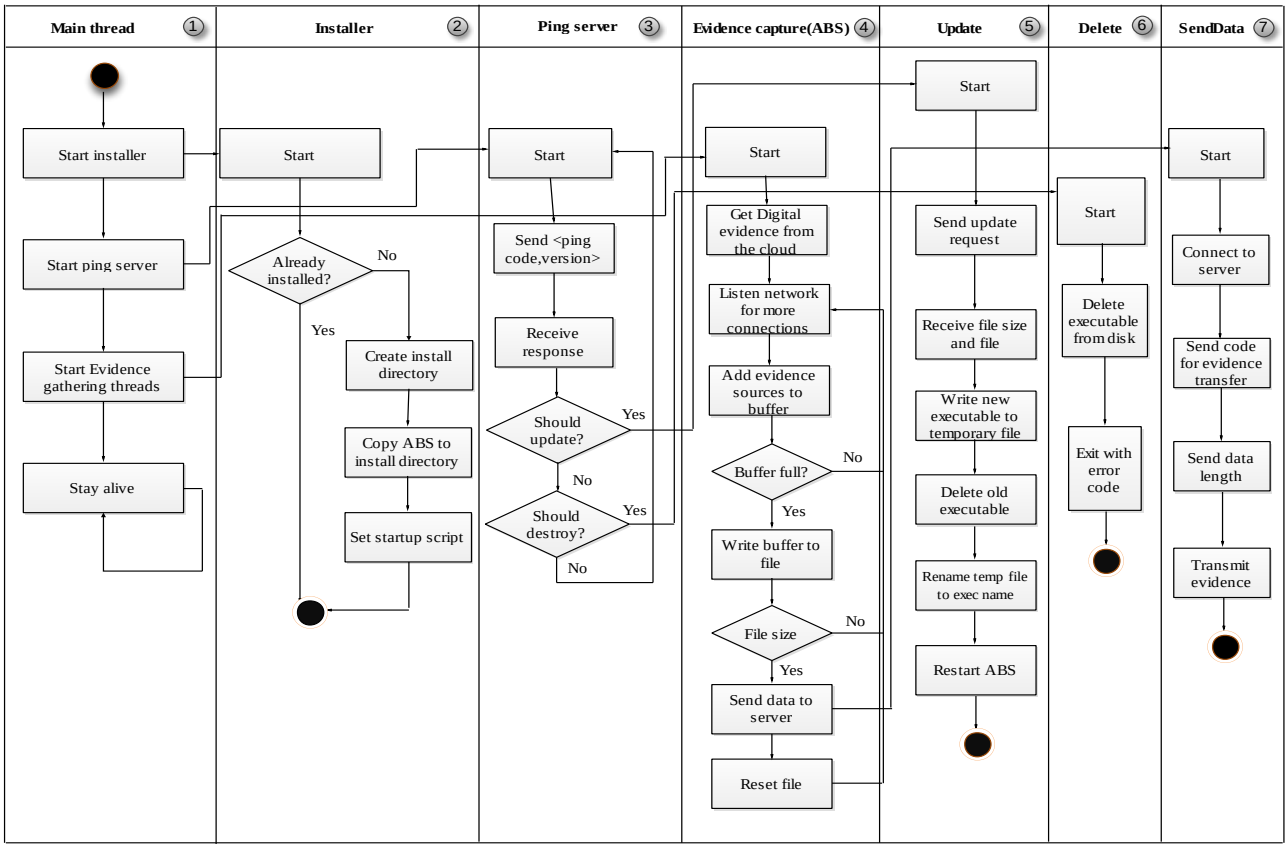

Figure 4. Operational flow of the Agent-Based System

Table 2 shows a summary of each component and its respective description that has been highlighted in Figure 4. The following components that have been labeled from 1-7 have been considered: Main thread, installer, ping server, evidence capture (ABS), update, delete and sendData.

Table 2. A description of Operation flow components of an Evidence gathering agent-based system

	Component	Description
1	**Main Thread**	Starts other threads and keeps the information collecting ABS alive throughout potential evidence collection process.
2	**Installer thread**	Checks if the infection vectors have installed the ABS. If not it installs itself and makes system changes that are needed when the operating system starts up.
3	**Ping thread**	The thread works on a timer which implies that in every *n* minutes the thread contacts the server. When the command is received from the server the ping is able to dispatch to the correct handler for the command sent by the server.
4	**Evidence Capture(ABS)**	Thread starts up when the ABS starts, connects to the network and listens to evidence channels then logs them to a file. After logging and after file reaches a given size the data is transmitted to the server. Thereafter, the file is reset.
5	**Update thread**	Thread starts when the ping is able to receive the update command from the server. It downloads the new version of the bot, installs it and then it is able to restart itself.
6	**Delete thread**	Starts whenever the ping thread receives the destroy command. This command is able to remove the executable from the disk and kill the ABS by exiting with an error code.
7	**SendData thread**	Responsible for transmitting data gathered by other modules to the server. As a parameter, the thread takes in the data length and a pointer to a byte array of data. It transmits the data length followed by the data to the C&C.

6.4.Prototype overview

The CFRP that is implemented in Linux-based environment uses OpenP2P as a communication protocol. For effective communication with the server, the ABS does network requests to the server and is able to collect digital information and send it back to the server. Figure 5 shows technical experiments that were conducted in Linux-based environment using a high-level language. C++ was preferred because of its ability and capacity to operate protocols over network, memory management and efficiency to the CPU.

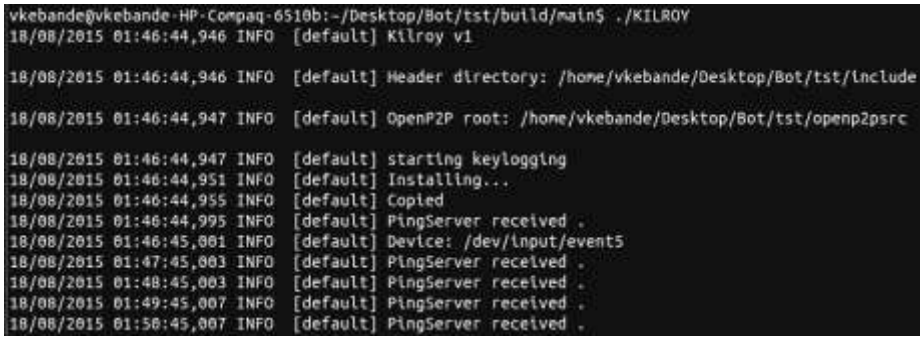

Figure 5. Server receives Ping after every 1 minute

The ABS works on a timer, it sends a ping to the server after every *n* minutes and after every ping data is transmitted and there is an update to the ABS. The update gets started if the ping is able to receive the update command from the server. The main thread of the program is responsible for keeping the ABS alive and starting other threads. The installer thread checks if the ABS is installed and makes changes when the OS starts up. The send thread is responsible for the transmission of data gathered to the server. Figure 4 shows the potential digital evidence that was collected with time stamp 18/08/2015 01:46:44, 956 to timestamp 18/08/2015 01:50:45, 007 after conducting ping transmit and update. An example of a block of captured potential digital evidence is shown in Figure 6.

Figure 6. Sample collected potential digital evidence

Subsequent requests are made to the server in order for the ABS to communicate with the server. This happens when the ABS needs to do network requests to the server. A summary of the requests and replies from the server has been presented in Table 2 below.

Table 3 Structure of the ABS requests.

Code	Purpose	Request structure	Possible server responses
P	Ping	<ping code><ABS version>	<update><destroy><received>
S	Transmit data	<transmit code><ABS version> <module code><data length> <data>	<received
U	Update	<update code><ABS version>	<binary size><new binary>

Table 3 shows the communication protocols that ABS uses while sending requests over the network between the ABS and the server. In this context ABS has been used to show the application with modified functionality. Code *P* in the first column represents the ping thread which works on a timer. The thread wakes up every *n* minutes and contacts the server and once the command is received from the server, the ping thread is able to dispatch a correct handler for a given command that was sent by the server. Once the structure <ping code> that shows the code being pinged and <ABS version> that shows the latest version of the ABS is submitted with the version, the server has an option of returning three possible responses represented as: <update>, <destroy> or <received>. Update, gives an update of the ABS, destroy kills the ABS while received shows the ABS that is accepted.

The next code *S* is for transmitting data that is gathered by the other modules to the server. In this context when represented as parameters S is able to take the data length and a pointer to a byte array of data and then it is able to transmit the data length which is then followed by the data to the Command and Control (C&C) server. Request structure in this context include <transmit code>that shows the code being transmitted <ABS version> which is an 8-bit version code that allows 256 versions of the ABS, <module code> that shows modules that are able to gather data to the server, <data length> which is the size of files uploaded to the server and <data> which represents the upload of gathered data and download of ABS binaries. Finally, the server response in this case is represented as <received>.

The last part of Table 2 is update which has been represented by *U*. Normally this will be started whenever the ping thread receives the update command from the server, it is able to download a newer version of the ABS then install it and then restart itself. The request structure that is represented in this context include: <update code> which is used to provide an update for the new code and <ABS version> that shows the latest ABS version. Finally the server gives the update on the <binary size> and <new binary>.

The next section presents a discussion on the issues and challenges faced when conducting DFR in the cloud using an ABS.

7. Issues and challenges faced when conducting digital forensic readiness in the cloud using an ABS

This section highlights a contribution towards assessing the prevailing issues and challenges when an ABS is used to conduct DFR in the cloud environment. The issues and challenges described in this section emanate from the CFR model.

Based on the comprehensive literature study and the CFR model we are able to identify the existing gaps. We have therefore harmonised the existing gaps between the existing literature and the CFR model by identifying issues and challenges and proposing high-level solutions for each challenge. The challenges are classified into three categories: general, technical and

operational.

Figure 7 illustrates the categories of challenges in a hierarchical structure. A more detailed explanation of the challenges follows for each category by highlighting the sub-challenges and proposing a high-level solution. Each of these challenges is explained in a separate sub-section below.

7.1. General Challenges

The authors provide a brief discussion on the general challenges by breaking down the main challenges into sub-challenges and then proposing a high-level solution to each of the challenges. The challenges arise as a result of conducting DFR using an ABS. The proposed high-level solutions are also discussed in each subsection. A summary of this discussion appears in Table 4. Table 4 is discussed in detail in Section 5.

ABS Obstruction

The existence of an ABS can be affected by the availability of disinfection strategies. As a result, these strategies are able to remove the forensic clients from their functionalities. Agarwal [20] in his research highlighted that in random mitigation strategies that disrupt communication, there exists a method whereby a set of clients are randomly removed from their functionalities. This method offers the possibility of an attack in the form of infiltration. The ABS is infiltrated by many or big fake malicious programs with a view to disrupting communication. Obfuscating the ABS solution from this activity according to Kebande and Venter[40] might prevent the possible infiltration and takedown of the non-malicious activity.

ABS Implementation

The implementation of an ABS involves two factors. Firstly, the botnet has originally been perceived to be of a devious nature and to have bad connotations. Secondly the ABS (modified functionalities of a botnet) collects useful information. Moreover, it is considered malicious because it captures information illegally. However, although the implemented botnet operates in a non-malicious fashion, the presence within the network of possible multiple malicious activities that infiltrate the ABS solution may disrupt communication during a DFR approach. Furthermore, the botnets creators from time to time keep modifying the botnet architecture with the intent of making it more resilient. This has made it very difficult for researchers to implement the botnet with the existing architecture.

It is essential to keep up with the current network architectures as they support a wide range of cloud services and architectures that sustain a broad range of emerging technologies and services.

Increased Number of Devices

The increase in distributed computing devices makes it a challenge to monitor the origin of data objects when an ABS gathers digital information. A potential issue uncovered with the presence of these devices is when data is moving to the cloud. When this happens, direct control of that data is lost. A publication by NIST 800-37[41] highlights the fact that "organisations are accountable for the risk incurred by use of services provided by external providers and address this risk by implementing compensating controls".

Since most devices have gone mobile and ad hoc, locating a specific device is not easy and the distributed nature of most devices poses a real challenge. Furthermore, the increase in

device numbers means an increase in usage, as well as an increase in volume, velocity and variety of data in the cloud, which makes evidence segregation hard. A proper cloud evidence management system for DFR purposes helps in the extraction, mapping and segregation of PDE.

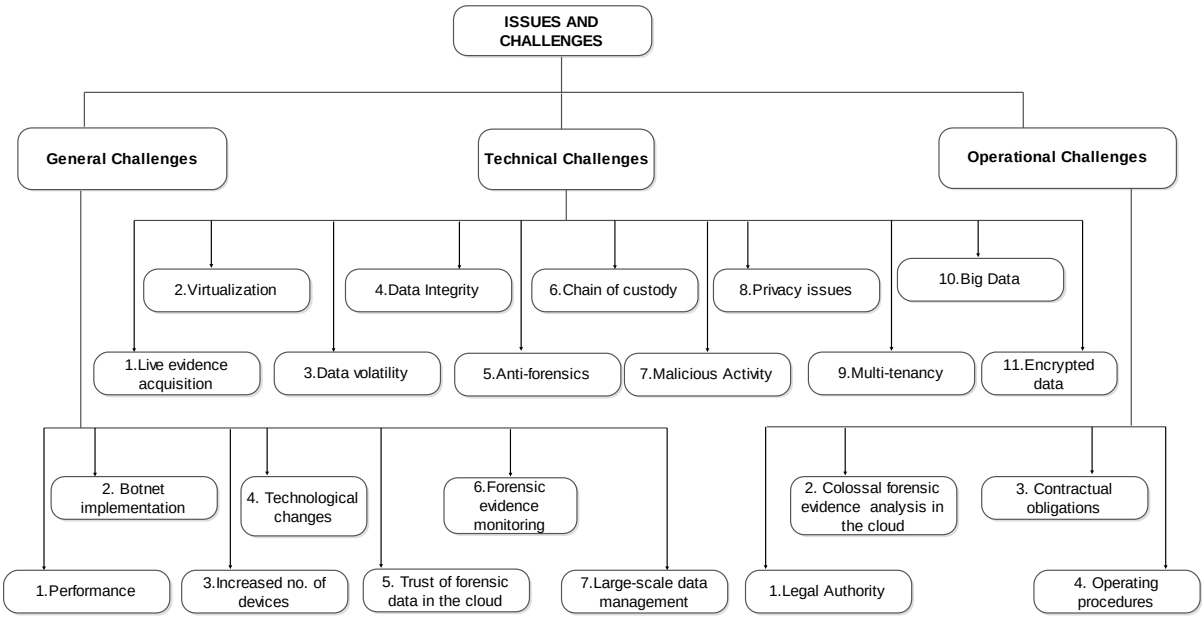

Figure 7. A hierarchical classification of issues and challenges

Technological Changes

The way in which the cloud environment operates is distinctly affected by the implementation of new and upcoming ABS and network technologies. The cloud is embraced as economical and has replaced the act of hosting own resources due to the new technologies that have proved to be efficient, cost effective and flexible [42]. DFR in the cloud is affected by current operational changes and on-demand solutions. This is due to the fact that the current DFIP in the cloud do not match the existing cloud computing characteristics, i.e. the cloud is not adapting to DF processes easily.

These technological changes tend to change the existing cloud-computing architectures in some instances. Furthermore, the change in network interoperability brings about changes in the ABS operation. In solving these predicaments, the new evidence-capturing forensic tools should be at par with the technological changes so that these technological concepts do not affect the confidentiality or authenticity of the forensic evidence being transmitted.

Trust of forensic Data in the Cloud

Conducting DFR in the cloud environment involves users' data gathered and retained by an ABS. For this reason, the chain of trust in the integrity and confidentiality of the data stored in the cloud is not endorsed fully by users, since the cloud is an untrusted execution environment. In this context, trust (as presented by Pearson[43] from a cross disciplinary view) is a psychological state that comprises the intention to accept vulnerability based on positive expectations of the intentions or behaviour of another). The process of capturing PDE is therefore faced by the issue of protecting critical components such as the hypervisor against run-time attacks. A hypervisor in this context is a piece of software that runs or operates the virtual machines. This is the most important key function in securing the PDE collected.

To mitigate this challenge, a chain of trust should be built on the identities, cloud infrastructures and data that are stored in the cloud. This can be done through hardening the hypervisor and VMs so that the risk of run-time attacks on the hypervisors can be reduced. It will also ensure that the privacy and integrity of PDE is maintained.

Large-scale Data Management

There is a plethora of data when gathering PDE from the cloud environment in a non-malicious way by means of an ABS. This is due to increased multi-tenancy and lack of effective scalable data management systems that can manage large-scale PDE. According to Abadi [44], managing the vast volumes of data in the cloud is more difficult and time consuming due to the lack of a timely and efficient back-end data management system in the cloud. This is mainly because data in the cloud comes in huge volumes in heterogeneous environments and it is often replicated across large geographical distances making data management a challenge. Hence, the cloud has to manage a number of resources with different patterns that range from digital maps, images, large files, structured and unstructured data. This is a very tedious exercise of mapping the actual potential evidence that can be admissible in a court of law. Furthermore, large-scale data requires many server instances to speed up data processing.

An ideal cloud potential forensic evidence processing system that is based on MapReduce that can manage the increased scale of generated data in the cloud brings about effectiveness in potential digital evidence management. This makes evidence aggregation and correlation easy through evidence characterisation. Furthermore, this will reduce the need for more server instances required to process more data.

Forensic Evidence Monitoring

Normally digital forensic evidence is aggregated in a multi-tenant environment. This environment is complex for it consists of numerous applications, VMs and hardware. The data is also organised in distributed and scalable fashion. The process of monitoring the cloud-based ABS solution is not very effective because complexity implies an increase of virtual server triggers and an increase of data load, which eventually leads to component overload. If the captured PDE is compromised, evidence may well be contaminated before a digital investigation can be performed. Since evidence monitoring is done on per user basis, a given cloud may have thousands and millions of monitoring tasks that might not be very effective in the long run. Moreover, data centres run a huge number of cloud services, and this may lead to an overlap in the metric being monitored.

By enforcing monitoring-as-a-service (MaaS) enforcement at the application level of the prototype, there can be an opportunity to improve the efficiency of the evidence being monitored. In spite of that, we should ensure that there is no storage of unnecessary forensic evidence. This means based on potential evidence that is collected from different sources, it will be assessed to see if it is relevant or non-relevant.

7.2. Technical Challenges

This section gives an overview of technical challenges encountered when conducting DFR in the cloud. The challenges will be highlighted mainly from a technical point of view and thereafter high-level solutions will be proposed.

A typical approach of cloud forensic readiness begins with the gathering and retention of PDE using an ABS as discussed in Section 3. The PDE collected for DFR purposes might result from hypervisor error logs, network logs, activity logs, virtual images, application logs,

database activities, monitoring, cloud carrier logs, etc. All these comprise forensic logs, service artefacts and monitored data that are contained in a cloud forensic storage database. According to Ananthanarayanan [46] the key infrastructure element of the cloud stack is a storage layer that is able to support large petabytes of data. A number of technical challenges arise as a result of events and log cycles and will be discussed next.

Live Evidence Acquisition

Since the cloud is distributed and volatile, there is no easy way of logically acquiring live PDE in the cloud environment for purposes of DFR after an ABS solution has gathered and retained critical digital information related to crimes. According to Casey[47] acquisition of digital artefacts is the initial step in the forensic process. In this context, if a VM is shut down by an adversary, it becomes impossible to collect forensic evidence because of the VM's volatile nature. Moreover, a challenge of verification arises due to changes in the data. Traditionally, data changes as the systems keep running, which brings about a difference in the evidence collected and the evidence acquired.

Live evidence depends on the system that the suspect is using, but Carrier [48] argues that a suspect's system cannot be trusted. In spite of that, Birk [2] highlights that in overcoming live acquisition, the CSPs can make access potential evidence read-only through an API. On the other hand, proper verification of the integrity of collected evidence should be enforced.

Virtualisation

According to the draft NIST 8006 [5], virtualisation is a simulation of the software or hardware upon which other software's run. A majority of network defence systems are based on physical networks and most data centres support static virtualisation. When the ABS is installed in the virtual instances, monitoring the security threats of VMs becomes difficult as a result.

Hypervisors consequently remain vulnerable to attacks and in order to prevent these, the CSPs must provide a perimeter security as a firewall. This is done to isolate the virtual resource spaces from further potential attacks.

Data Volatility

Once the VM is shut down, PDE is lost. This is because all the data residing in the VM is volatile, making it a difficult task to locate the whereabouts of data. In nature, volatile data tends not to survive when there is power failure. To mitigate this challenge, data must be computational. Zawoad and Hasan [49] highlight the fact that evidence should be stored in a persistent database so that if an adversary attempts to shut down a VM, evidence can still be gathered. Additionally, Dykstra and Sherman [50] suggest that a cloud management that uses an IaaS model can enable evidence gathering when the VM is terminated.

Data Integrity

An ABS solution is able to collect vital data as PDE. On the same note, it is not easy to guarantee the perfection and correctness of essential and critical data that exist as PDE in the cloud environment because PDE streams in from different points at different times. Consequently, it is also not easy to differentiate at different levels which kind of data is essential and which one is not. The integrity of data should be enforced to help to prove in a court of law that the evidence being presented is the same evidence that was captured forensically. This also will increase the chances of admissibility on PDE. In this context, this

is a big challenge for DF investigators and law enforcement agencies, because even if they manage to acquire the necessary evidence, it might not be an easy task to verify the integrity of data and its origin.

According to Grispos et al.,[51] data stored in the cloud should be subjected to hashing, which enables integrity checking. Additionally, ISO/IEC 27037, 10118-2 and ISO/IEC 27043[8,9] stresses on the need to avoid cases of intentional and un-intentional evidence deletion, evidential integrity through using hash functions of all the bits in each media that contain PDE.

Anti-forensics

The draft NIST 8006 [5] , highlights anti-forensics as a set of techniques that are used specifically to prevent or mislead forensic analysts. In this context, these tools frustrate the process of achieving DFR. NIST further categorises anti-forensics as a process of obfuscation and hiding data, as well as the use of malicious codes with the intent of compromising the integrity of PDE. The use of anti-forensics reduces the quality of PDE deliberately by interfering with pre-incident analysis of PDE.

According to Jahankhani and Beqiri [52] to mitigate anti-forensics; computer forensic tools should be improved through improvement of signature analysis and time-stamp analysis. Nevertheless, all forensic tools should be hardened because forensic tools like Encase and FTKs do not check for signatures.

Potential Digital Evidence Handling

Handling and managing the collection of evidence is a daunting task because of the distributed nature of resources and cross-cutting jurisdictional issues. PDE seizures, control, transfer and trails have to be documented systematically according to accepted cross-jurisdictional standards, procedures and technology. Vacca [53] highlights the fact that there has to be a roadmap showing how evidence is collected, analysed and preserved in order for PDE to qualify for admission in a court of law. In the preparation of DFR, the policies regarding retention, collection, planning and evidence acquisition must be documented chronologically. When this approach is not followed, PDE loses quality and may not be admissible. Furthermore, the good practices of the UK's Association of Chief Police Officers (ACPO)[54] describe documentation as a way in which evidence was managed before being presented in a court of law.

Malicious Activity

A cloud-computing platform is a ready target for malicious activity. Protecting an ABS in the cloud from adversaries is necessitated by the existence of numerous threats and attacks because of the pervasiveness of the cloud. VMs are bound to be attacked and this brings a possibility of intentional data tampering, which in the long run makes the cloud platform vulnerable to attacks. The aspect introducing Intrusion Detection Systems (IDS) to monitor the entire network traffic and suspicious behaviour protects an ABS from intentional attacks. Kebande and Venter[55] uncovered a mechanism that was able to detect how malicious botnets or potential malicious activity can be detected in the cloud, namely the Artificial Immune System (AIS).

This was based on the malicious pattern that the botnet traffic uses. Nevertheless, according to Flood and Keane[56], a system should be trained on the possibility of responding to the user interaction. Enforcement of these criteria is a key to protection in the cloud and performing a comprehensive assessment regularly. On the other hand, Claycomb and Nicoll

[57] argue that "transparency into overall information security and management practices, determining security breach notification process and encrypting data in the cloud" are some of the key solutions to the challenge of the presence of malicious activity in the cloud.

Privacy Issues

Gathering and retaining PDE from the cloud by using an ABS poses a huge challenge to cloud consumers, because PDE collected for DFR purposes is not achieved through deliberate planning. Capturing user information within the cloud may have jurisdictional issues because electronic transactions that can lead to disclosure of personal information. This may be treated as a contravention of an individual's right to privacy. A review of the legal perspective on admissibility of DE shows that although the requirements vary across different jurisdictions, some legislation provides exemptions to allow interference with privacy of information, provided that it is a matter of national security or a research activity. These Acts include the Electronic Communications Privacy Act (ECPA), Act of 1986 of the USA; the UK's Association of Chief Police Officers (ACPO) Good Practice Guide for Digital Evidence; the Electronic Communication and Transactions (ECT) Act of South Africa; the Protection of Personal Information (POPI) Act of South Africa and the Stored Communications Act (SCA) of the USA[54, 58, 59, 60, 61, 69].

Moreover, the cases from the United States of America, which include the presentation of digital evidence, are treated under Rule 702 of the federal rules of evidence. This is the rule of evidence that says "If scientific, technical, or other specialised knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise". The Daubert case [62] applied these rules.

Multi-tenancy

A single instance of a software application may have many tenants where an ABS is deployed. However, when multiple VMs tend to share the same physical infrastructure in the cloud environment, computing gets distributed. As a result, potential vulnerabilities arise from the VM technologies used by cloud vendors, which eventually results to a number of security issues while gathering PDE at VM level. A higher chance of evidence contamination is likely because evidence come from different parts of the multi-tenant environment; hence it is left to the DF investigator to prove whether the evidence is strongly associated with the malicious user, while the privacy of other tenants is preserved.

This is because data is shared among multiple computers within multiple locations with numerous tenants and with numerous forensic evidence. Additionally, the process of proving whether the evidence has a relationship with the user in this environment is faced with timestamp inconsistencies and difficulty in assessing forensic evidence. According to Zawoad and Hassan [49], the DF investigator must prove two instances: firstly, that the forensic evidence is not mingled and secondly that the privacy of other users is preserved.

Big Data

Data gathered by an ABS from the cloud comes in vast volumes. This data, which is characterised by velocity, variety, complexity and volume, is a threat when DFR is to be achieved. Big data remains unstructured, and performing pre-incident analysis on this data is a challenge because data comes from different sources. Big data is a focal challenge and it affects almost every activity being conducted within the cloud environment. During a DFR approach, it mainly affects data security: when sensitive, critical data related to crime and

personal information tends to exist within big data, it becomes a big risk.
As a result of increased devices and the large exploding amount of terabytes of digital data that is seized as forensic evidence, Quick & Choo [72, 73] has highlighted that the effects of this leads to large backlogs, increased the size of devices per case, increased number of cases that deals with digital evidence and an increase in the size of data. The authors in this context have gone ahead to propose a digital forensic data reduction and data mining framework that is able to reduce the storage demands while providing a capability of conducting a review on subset data for purposes of intelligence analysis, archival, research and historical review purposes [73].

On the other hand, big data comes in all shapes and sizes, and it contains very heavy schemas that further cause pre-incident analysis problems. Another issue is the need to scale data centres rapidly and the cost of maintaining the Relational Database Management System (RDBMS), which is very high. The authors have proposed a functional architecture that is able to do analysis of large-scale PDE in the cloud using MapReduce [38]. Through this computing is transferred to low commodity cluster computers by mapping and reducing the colossal amount of data using a reducer.

Encrypted Data

Adding an additional layer of security to data may make data unusable because encryption is a strong security mechanism for data protection. This involves making data meaningless to unauthorised users or adversaries. Encryption is embraced because cloud users want protection, compliance and control of their stored data. From a DFR perspective, if encrypted PDE is collected by an ABS, it becomes a task to decrypt the same, due to key management constraints. A major issue regarding data moving to the cloud is how it is handled if there is a breach, as well as the status of data at rest and data in motion. For the purposes of effectiveness in conducting DFR, a data security plan should be laid out in accordance with which data should be encrypted.

Data in motion and data at rest should be encrypted because data protection is a critical aspect of security. Finally, key management should be performed by the CSPs as encryption providers. Logging as forensic evidence collection policies should be controlled by the CSPs and at the same time they need to enforce measures to restrict access to sensitive and critical management tools.

7.3. Operational Challenges

This section first discusses operational challenges and then proposes high-level challenges when conducting DFR in the cloud. Operational challenges consist of the following sub-challenges: legal authority, colossal data analysis, contractual obligations and operating procedures.

Legal Authority

The issue of legal admissibility of DE is encountered when gathering PDE across different jurisdictions. Issues arising from cross-cutting jurisdictions hold a challenge when digital evidence (DE) is collected as part of the DFR process. DE that is stored in multiple jurisdictions cannot be accessed because what is considered legal in one country might be illegal in another. This is because of the lack of proper Internet laws and consistent legislation, which may cause governments to be unwilling to cooperate and thus lead to delays in warrants being served (hence delays in prosecution).

On the concept of legal authority, the Security Techniques Advisory Group (STAG) [43] on lawful interception has highlighted on its ETSI Technical Report (ETR) that a provision on guidance is given to service providers and network operators with lawful interception of telecommunications. Based on this the general requirements that have been highlighted recapitulates that the CSP will intercept, retrieve and store content of communication the entire period and one can monitor or choose to permanently record the results arising from interception. However in order for PDE to be admissible it must satisfy the legal requirements of the particular jurisdiction because the legal requirements for admissibility of digital evidence vary across different jurisdictions.

Nevertheless, according to Cohen [44], most communications are interstate and international when it comes to jurisdiction and cooperative agreements should solve evidence-giving problems. On the other hand, Biggs and Vidalis [42] suggest that in overcoming cross-border challenges, an international unity should be formed to introduce international legislation that can implement globally accepted legislation.

Colossal Forensic Evidence Analysis in the Cloud

The distributed ABS is very versatile, as it captures huge amounts of data with high velocity, volume and complexity. When performing DFR pre-incident analysis and examination, performing analysis of incoming traffic is a challenge because of the mentioned characteristics. The authors of this paper proposed cloud storage with MapReduce, as it can process the collected data by performing computations of the large-scale DE through trace and analysis conducted in parallel[38]. Chen et al. [45] also proposed a cloud-based forensic analysis managing system that could analyse the traffic in the cloud.

Contractual Obligations

Normally, contractual obligations exist between the CSPs and the cloud consumers. During PDE capturing, a dispute is encountered on how the captured logs are to be administered by the client and CSPs. Furthermore, any change in the implementation of Service Level Agree ments (SLAs) may breach the pact between the parties, which might lead to conflict. Mainly, this involves a situation in which CSPs have full control of forensic evidence in the cloud, where customers do not have any control. According to Zawoad and Reilley[68, 70], it is a requirement that PDE that is generated in different layers must be accessible to different stakeholders of the system.

Standard Operating Procedures

There is lack of proper Standard Operating Procedures (SOPs) in the cloud environment on how DFR is conducted. It is difficult to develop tools that can conduct DFR because of cross-platform developments and a lack of standardised infrastructures. According to Aminnezhad[66] conducting DFR processes in the cloud can be extremely challenging when it includes thousands of virtual machines. It may also lead to disruption of service to other users. ISO/IEC 27403[8] was proposed as a standard for security techniques, incident investigation principles and processes. This standard also defines the digital forensic readiness process as a process that can be conducted before incident detection [8]. Additionally, ISO/IEC 27403: 2015[8] presents the main aim of DFR in an organization as "to maximize the potential use of digital evidence while minimizing the cost and preserving the level of information security systems within the organization[10,11]. The next section evaluates the proposed contribution.

8. Discussion

In this research, the inadaptability of DF in the cloud has revealed the challenges faced when conducting DFR in the cloud using an ABS. (See the CFR model as discussed in Section 3.) The paper reports on research that was based on digital forensic gathering and retention mechanisms for purposes of DFR. Based on the existing literature[17-26, 29-24] we conducted a literature review of the research conducted by other researchers on cloud forensic challenges[5] (Table 4) and other well-documented research papers.

Not only were we able to identify challenges as a result of implementing DFR in the cloud by means of an ABS, but we were able to identify other challenges from the existing literatures[17-26, 29-24] that are also attended to in our research study. Based on the current study, we have been able to make a contribution in respect of all the challenges faced in the cloud by proposing high-level solutions. As a result of a review of the related work from authors[17-26, 29-24] and the challenges arising from the ABS implementation, we managed to list all these challenges in a summarised format in Table 4 for possible comparison and contrasting. Challenges that appeared in both the authors[17-26, 29-24] and the model were represented with their respective references, while challenges emanating from the model but not in the related work were represented using (X). Furthermore, the specific high-level solution for each challenge was summarised in the fourth column of Table 3. It is worth noting that in the authors' opinion, the proposed high-level solutions may be enhanced as a result of further research in this context.

If the recommended solutions in this research are adopted fully, it is the authors' opinion that this could significantly facilitate effective proactive processes of pre-incident detection in the cloud environment as is highlighted in ISO/IEC 27403:2015 standard [8]. Digital information retention, correlation and management constitute the most comprehensive way of conducting DFR in the cloud without interfering or modifying the cloud architecture. This can only be possible if the retained potential evidence is correlated, manipulated and managed outside the cloud.

If we explore the challenges from the stakeholders' perspective, the study significantly points out a number of important counter-measures that can help the stakeholders to understand the impacts of DF challenges with respect to future technologies in DFR approaches that are implemented for organisations. This is a big concern to the DF community, especially to the law enforcement agencies and DF investigators.

From a legal point of view, the CFR model's use of an ABS in capturing user information might interfere with an individual's privacy. This was highlighted by the Regulation of Interception Communications and Provision of Communication-Related Information Act (RICA), 70 of 2002[67, 71].However, Section 6(2) of the RICA[67] is a provision that states that "interception can be made for reasons of investigating unauthorised use of that communication system". This can only take place if the investigator has consent from law enforcement authorities.

Table 4.Issues and challenges faced in implementing an ABS in the cloud for purposes of digital forensic readiness

	Category and Identified Challenges	Related work on Identified Challenges	Proposed High-Level Solutions
GENERAL CHALLENGES			
1	ABS Obstruction	xxx	Obfuscate the ABS cloud solution
2	Implementation	xxx	Use current network architecture
3	Increased no. of devices	[72]	Introduce a cloud evidence management system
4	Technological changes	[5][6]	Introduce new evidence-capturing tool, tools at par with new changes
5	Trust	[5][26][33][78]	Built trust based on identities, infrastructure and information
6	Data Management	[22][72]	Develop a cloud forensic readiness management system, it will reduce the need for more server instances
7	ABS Monitoring	[16][19][21][23][25][26][30][32][33][74][75][76][77]	Monitor as-a-service at application level
TECHNICAL CHALLENGES			
1	Live evidence acquisition	[5][30][79][74][77]	Access evidence in read-only through an API
2	Virtualisation	[5][17][18][22][23][33][74]	Provide perimeter security as a firewall
3	Volatile data	[5][17][18][32][33][74][75][76]	Store evidence in a non-persistent form
4	Integrity of collected evidence	[5][19][22][33][74]	Subject data in cloud to hashing and encrypt data at rest and in motion
5	Anti-forensics	[5]	Harden forensic tools by improving signature analysis and time-stamp analysis
6	Potential evidence handling	[5] [16] [17][22][25]	Preserve document evidence as a way of management
7	Malicious activity	[5][78]	Determine security breach notification and encrypt data in the cloud
8	Privacy issues	[5][21][32][33][78]	Handle the issue in accordance with requirements of jurisdictional acts
9	Multi-tenancy	[5][21][33]	Get a digital forensic investigator to prove all instances for consistency
10	Big-data	[33][72]	Use MapReduce for analysis of big data
11	Encrypted data	[5][19][22][33][79]	Get CSPs to handle key management and to have control of logging policies
OPERATIONAL CHALLENGES			
1	Legal authority	[5] [16] [21][26][31][76][79][80]	Formulate international legislation to implement global law
2	Colossal data analysis	[16][17][18][19][22][74][75]	Implement a cloud-based analysis management system that uses MapReduce
3	Contractual obligations	[21][25][29][31][33]	Make it a requirement that evidence generated in different layers be accessible to different stakeholders of the system
4	Standard Operating procedures	[21][31]	Adopt the proposed readiness standards by ISO/IEC 27043

Finally, possible applicability of the aforementioned procedures can be enforced if the CSPs are entrusted to collect evidential information in a forensic readiness process. For each of the contributions mentioned above, the authors believe that DFR as a process will be effective and that the cloud model will still be able to offer state-of-the-art services. The next section concludes this discussion and suggests possible future work.

9. Conclusion and Future Work

The research question that this paper addressed in Section 1 is "what are the issues and challenges when conducting DFR in the cloud using an ABS?" In addressing the research question we identified the challenges and issues from a general, technical and operational point of view (see Section 4).

According to the authors' reflection on the CFR model, an ABS was capturing digital information in the cloud environment, which was a constantly changing environment. The captured PDE was digitally preserved for DFR purposes. The broad concept was to perform evidence computation, forensic processes and evidence manipulations outside the cloud so that the existing architecture of the cloud could not be modified. Furthermore, the authors proposed a high-level solution for each of the contributions.

Research that remains to be conducted as future work involves the development of a prototype that is able to simplify the storage of content data that exist as potential digital evidence. Furthermore, this will enable computation of the large-scale digital evidence captured by the ABS to create effectiveness during the incident response procedures.

Acknowledgement

This work is based on research supported by the National Research Foundation of South Africa (Grant-specific unique reference number UID85794). The Grant holder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF-supported research are those of the author(s) and the NRF accepts no liability whatsoever in this regard. The authors wish to thank the ICSA Research Group of the Department of Computer Science at the University of Pretoria for the support towards coming up with this research.

References

1. IDC predictions 2014. Top 10 predictions: Competing for 2020. Retrieved from http://www.idc.com/prodserv/FourPillars/Cloud/index.jsp.
2. Birk D. Technical challenges of forensic investigations in cloud computing environments. In Workshop on Cryptography and Security in Clouds, pp. 1-6, 2011.
3. Wilcox J. Gartner: Most cios have their head in the cloud. Available at http://betanews.com/2011/01/24/gartner-most-cios-have-their-heads-in-the-clouds/
4. Anonymized
5. Draft NISTIR 8006 NIST Cloud Computing Forensic Science Challenges. Accessed at http://csrc.nist.gov/publications/drafts/nistir-8006/draft_nistir_8006.pdf
6. Palmer G. A road map for digital forensic research. In First Digital Forensic Research Workshop, Utica, New York (pp. 27-30), (2001, August).
7. Casey E. Digital forensics: Coming of age. Digital Investigation 6(1-2): 1-2, 2009.
8. ISO/IEC 27043: (2015). nformation technology -- Security techniques -- Incident investigation principles and processes.[online]-Accessed at

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44407
9. ISO/IEC 27037:(2012). Information technology -- Security techniques -- Guidelines for identification, collection, acquisition and preservation of digital evidence.[online], Accessed at http://www.iso.org/iso/catalogue_detail?csnumber=44381
10. Rowlingson R. A ten step process for forensic readiness. International Journal of Digital Evidence, 2(3), 28, 2004.
11. Tan J. Forensic readiness. Cambridge, MA:@ Stake, 1-23, 2001.
12. Yasinsac A, Manzano Y. Policies to enhance computer and network forensics. In Proceedings of the 2001 IEEE workshop on information assurance and security, pp. 289-295, 2001, June.
13. Mell P, Grance T. The NIST definition of cloud computing (draft). NIST special publication, vol. 800, p. 7, 2011.
14. McDonald N. Gartner Research. Addressing the most common security risks in data center, virtualization projects. Accessed http://bsius.com/media/182447/addressing_the_most_common_s_173434.pdf
15. Ruan, K., Carthy, J., Kechadi, T., & Crosbie, M. (2011). Cloud forensics. InAdvances in digital forensics VII (pp. 35-46). Springer Berlin Heidelberg.
16. Martini, B., & Choo, K. K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), 71-80
17. Quick, D., & Choo, K. K. R.(2013a). Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata?. Digital Investigation, 10(3), 266-277.
18. Quick, D., & Choo, K. K. R.(2013b). Digital droplets: Microsoft SkyDrive forensic data remnants. Future Generation Computer Systems, 29(6), 1378-1394.
19. Quick, D., & Choo, K. K. R. (2013c). Dropbox analysis: Data remnants on user machines. Digital Investigation, 10(1), 3-18.
20. Quick, D., & Choo, K. K. R. (2014a). Google Drive: Forensic analysis of data remnants. Journal of Network and Computer Applications, 40, 179-193.
21. Hooper, C., Martini, B., & Choo, K. K. R. (2013d). Cloud computing and its implications for cybercrime investigations in Australia. Computer Law & Security Review, 29(2), 152-163.
22. Martini, B., & Choo, K. K. R. (2013). Cloud storage forensics: ownCloud as a case study. Digital Investigation, 10(4), 287-299.
23. Martini, B., & Choo, K. K. R. (2014b, September). Remote programmatic vCloud forensics: a six-step collection process and a proof of concept. InTrust, Security and Privacy in Computing and Communications (TrustCom), 2014 IEEE 13th International Conference on (pp. 935-942). IEEE.
24. Chung, H., Park, J., Lee, S., & Kang, C. (2012). Digital forensic investigation of cloud storage services. Digital investigation, 9(2), 81-95.
25. Marty, R. (2011, March). Cloud application logging for forensics. InProceedings of the 2011 ACM Symposium on Applied Computing (pp. 178-184). ACM.
26. Dykstra, J., & Sherman, A. T. (2012). Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. Digital Investigation, 9, S90-S98.
27. Ollmann G. Botnet communication topologies. 2009. Retrieved 30 September 2009.
28. Falliere N, Chien E. Zeus: King of the bots. Symantec Security Response (http://bit. ly/3VyFV1), 2009.
29. Ab Rahman, N. H., & Choo, K. K. R. (2015). A survey of information security incident handling in the cloud. Computers & Security, 49, 45-69.
30. Martini, B., & Choo, K. K. R. (2014c). Cloud forensic technical challenges and solutions: a

snapshot. IEEE Cloud Computing, (4), 20-25.
31. N. H. Ab Rahman, W. B. Glisson, Y. Yang and K. K. R. Choo, "Forensic-by-Design Framework for Cyber-Physical Cloud Systems," in IEEE Cloud Computing, vol. 3, no. 1, pp. 50-59, Jan.-Feb. 2016.doi: 10.1109/MCC.2016.5.
32. Vincze, E. A. (2015). Challenges in digital forensics. Police Practice and Research, 1-12.
33. Simou, S., Kalloniatis, C., Kavakli, E., & Gritzalis, S. (2014d, June). Cloud forensics: identifying the major issues and challenges. In *Advanced Information Systems Engineering* (pp. 271-284). Springer International Publishing.
34. Delport, W., Köhn, M., & Olivier, M. S. (2011, August). Isolating a cloud instance for a digital forensic investigation. In ISSA.
35. Valjarevic A, Venter HS. A Harmonized Process Model for Digital Forensic Investigation Readiness. In Advances in Digital Forensics IX, pp. 67-82, 2013. Springer: Berlin Heidelberg.
36. Trenwith PM, Venter HS. Digital forensic readiness in the cloud. In Information Security for South Africa, 2013 (pp. 1-5). IEEE, 2013, August.
37. Cohen FB. Digital forensic evidence examination. Asp Press, 2009.
38. Kebande VR, Venter, H. S. A Functional Architecture for Cloud Forensic Readiness Large-scale Potential Digital Evidence Analysis. In Proceedings of the 14th European Conference on Cyber Warfare and Security 2015: ECCWS (p. 373). Academic Conferences Limited, 2015.
39. Agarwal S. Performance Analysis of Peer-To-Peer Botnets using "The Storm Botnet" as an Exemplar (Doctoral dissertation, University of Victoria), 2010.
40. Kebande VR, Venter, H.S. Obfuscating a Cloud-Based Botnet Towards Digital Forensic Readiness. In Iccws 2015-The Proceedings of the 10th International Conference on Cyber Warfare and Security (p. 434). Academic Conferences Limited, 2015.
41. NIST SP 800-37, 2010, Guide for Applying the Risk Management Framework to Federal Information Systems, A security Life Cycle Approach.
42. Biggs S, Vidalis S. Cloud computing: The impact on digital forensic investigations. In Internet Technology and Secured Transaction. ICITST 2009. International Conference for (pp. 1-6). IEEE, 2009.
43. Pearson S. Privacy, security and trust in cloud computing. In Privacy and Security for Cloud Computing, pp. 3-42, 2013. Springer: London.
44. Abadi, D. J. (2009). Data management in the cloud: limitations and opportunities. IEEE Data Eng. Bull., 32(1), 3-12.
45. Bousselmi K, Brahmi Z, Gammoud MM. Cloud Services Orchestration: A Comparative Study of Existing Approaches. In International Conference on Advanced Information Networking and Applications Workshops, pp.410-416, 2014.
46. Ananthanarayanan R, K. Gupta, P. Pandey, H. Pucha, P. Sarkar, M. Shah and R. Tewari, Cloud analytics: Do we really need to reinvent the storage stack. In Proceedings of the 1st USENIX Workshop on Hot Topics in Cloud Computing (HOTCLOUD'2009), San Diego, CA, USA, 2009, June.
47. Casey E. Handbook of Computer Crime Investigation. Academic Press. Boston. 2002.
48. Carrier BD. Risks of live digital forensic analysis. Communications of the ACM, 49(2), 56-61, 2006.

49. Zawoad S, Hasan R. Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems. arXiv preprint arXiv:1302.6312, 2013.
50. Dykstra J. Seizing electronic evidence from cloud computing environments. Cybercrime and Cloud Forensics: Applications for Investigation Processes, Hershey, PA: Information Science, 156-185, 2013.
51. Grispos G, Storer T, Glisson W. Calm before the storm: The challenges of cloud computing in digital forensics. International Journal of Digital Crime and Forensics, 4(2), 28-48, 2012.
52. Jahankhani H, Beqir E. Digital evidence manipulation using anti-forensic tools and techniques. Handbook of Electronic Security and Digital Forensics, 411, 2010.
53. Vacca JR. Computer forensics: Computer Crime Scene Investigation. Charles River Media, 20 Downer Avenue, Suite 3, Hingham, MA, 02043, second edition, 2005.
54. ACPO - Association of Chief Police Officers. Good Practice Guide for Computer Based Electronic Evidence, 2007.
55. Kebande VR, Venter HS. A cognitive approach for botnet detection using Artificial Immune System in the cloud, Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2014 Third International Conference on , vol., no., pp.52,57, April 29 2014-May 1 2014
56. Flood J, Keane A. A Proposed Framework for the Active Detection of Security Vulnerabilities in Multi-tenancy Cloud Systems. In Third International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), pp. 231-235. IEEE. 2012, September.
57. Claycomb WR, Nicoll A. Insider threats to cloud computing: Directions for new research challenges. In Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual (pp. 387-394). IEEE, 2012, July.
58. Doyle C. Privacy: An Overview of the Electronic Communications Privacy Act. Congressional Research Service, Library of Congress. 2011.
59. Gereda SL. The Electronic Communications and Transactions Act, (2006).
60. Telecommunications Law in South Africa. The Protection of Personal Information Act, Vol. 581, No 4, 2013.
61. Scolnik A. Protections for electronic communications: The stored communications act and the Fourth Amendment. Fordham L. Rev., 78, 2009, 349 ISBN:
62. Daubert V. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 1993.
63. E. TC-STAG. Security techniques advisory group (stag); definition of user requirements for lawful interception of telecommunications: requirements of the law enforcement agencies, 1996.
64. Cohen F. Challenges to digital forensic evidence. Fred Cohen and Associates, 2008.
65. Chen Z, Han F, J. Cao, X. Jiang, S. Chen. Cloud computing-based forensic analysis for a collaborative network security management system. Tsinghua Science and Technology, 18(1), 40-50, 2013.
66. Aminnezhad A, Dehghantanha A, Abdullah MT, Damshenas M. Cloud Forensics Issues and Opportunities, 2013.
67. Regulation of Interception of Communications and Provision of Communication-related Information Act 70 of 2001.
68. Zawoad S, Hasan R. Digital Forensics in the Cloud. Alabama Univ in Birmingham, 2013.

69. Owen P, Thomas P. An analysis of digital forensic examinations: Mobile devices versus hard disk drives utilising ACPO & NIST guidelines. Digital Investigation, vol. 8, pp. 135-140, 2011.
70. Reilly D, Wren C, Berry T. Cloud computing: Pros and cons for computer forensic investigations. International Journal Multimedia and Image Processing (IJMIP), 1(1), 26-34, 2011.
71. The regulation of interception of communications and provision of communication-related information, 2010. Accessed 10 September 2014.
72. Quick, D., & Choo, K. K. R. (2014e). Impacts of increasing volume of digital forensic data: A survey and future research challenges. Digital Investigation,11(4), 273-294.
73. Quick, D., & Choo, K. K. R. (2014f). Data reduction and data mining framework for digital forensic evidence: storage, intelligence, review and archive. Trends & Issues in Crime and Criminal Justice, 480, 1-11.
74. Alqahtany, S., Clarke, N., Furnell, S., & Reich, C. (2015, April). Cloud Forensics: A Review of Challenges, Solutions and Open Problems. In Cloud Computing (ICCC), 2015 International Conference on (pp. 1-9). IEEE.
75. M. Damshenas, A. Dehghantanha, R. Mahmoud and S. bin Shamsuddin, "Forensics investigation challenges in cloud computing environments," Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on, Kuala Lumpur, 2012, pp. 190-194.
76. Meyer, G., & Stander, A. Cloud Computing: The Digital Forensics Challenge.
77. Sández, M. J. R. (2015). A Review of Technical Problems when Conducting an Investigation in Cloud Based Environments. arXiv preprint arXiv:1508.01053.
78. Daryabar, F., Dehghantanha, A., & Udzir, N. I. (2013). A review on impacts of cloud computing on digital forensics. International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2(2), 77-94.
79. Sibiya, G., Venter, H. S., & Fogwill, T. (2012). Digital forensic framework for a cloud environment.
80. Van Eecke, P. (2011). Cloud Computing Legalissues. *Recuperado de http://www. isaca. org/Groups/Professional-English/cloud-computing/GroupDocuments/DLA_Cloud% 20computing% 20legal% 20issues. pdf*.