Cloud Computing: Pros and Cons for Computer Forensic Investigations

Denis Reilly, Chris Wren, Tom Berry
School of Computing and Mathematical
Sciences Liverpool John Moores University, UK

Abstract

Cloud computing is a relatively new concept that offers the potential to deliver scalable elastic services to many. The notion of pay-per use is attractive and in the current global recession hit economy it offers an economic solution to an organizations' IT needs. Computer forensics is a relatively new discipline born out of the increasing use of computing and digital storage devices in criminal acts (both traditional and hi-tech). Computer forensic practices have been around for several decades and early applications of their use can be charted back to law enforcement and military investigations some 30 years ago. In the last decade computer forensics has developed in terms of procedures, practices and tool support to serve the law enforcement community. However, it now faces possibly its greatest challenges in dealing with cloud computing. Through this paper we explore these challenges and suggest some possible solutions.

1. Introduction

The last few decades have witnessed several notable step changes which have shaped future practices in computing and IT. Cloud computing is tipped as the next notable step change, which potentially may change the way in which organizations realize their computing and IT needs. Cloud computing provides an attractive 'pay-per-use' model of computing, which allows organizations to effectively outsource their computing and IT requirements and focus on their core business, paying only for what they use. In the current global economic climate of global recession many organizations incur huge costs in terms of equipment and manpower expenditure keeping large dated legacy systems running. Cloud computing aims to provide a clean effective solution by allowing such organizations to migrate their data to a cloud, which promises high speed access and 99.99% availability, typically provided by trusted household vendors, such as Microsoft, Amazon, Google, Yahoo.

Computer forensics has emerged as a discipline to assist law enforcement agencies in addressing the increasing use of digital storage devices in criminal acts (both traditional and hi-tech). UK police forces have found that computing devices (including mobile phones) feature in many of the day-to-day crimes. Forensic examination of such devices can reveal a wealth of evidence that would otherwise be unavailable using conventional policing methods. Indeed, several high profile murder cases have benefited from digital evidence gathered via a computer forensic examination [1].

Although cloud computing has many benefits to offer, there is still a degree of speculation over its security (or lack of security). More particularly, there are still questions to be answered relating to its ability to support forensic investigations. Through this paper we intend to highlight a number of issues relating to computer forensics in cloud computing and provide our own thoughts on how these issues may hinder or encourage the uptake of cloud computing.

The remainder of this paper is structured as follows: section 2 provides more background on cloud computing and describes the key characteristics and models underlying cloud computing. Section 3 provides background on computer forensics and describes the processes and techniques that are the basis for computer forensics in law enforcement. Section 4 goes on to describe Virtualization, which is an important accompaniment to cloud computing, which enables resources to be shared within clouds. In section 5 we merge the two together and describe the main pros and cons relating to the application of computer forensic procedures within cloud environments. Section 6 draws overall conclusions and expresses our views on how we see the future emerging.

2. Cloud computing – silver lining or dark storm?

Cloud computing intends to realize the concept of computing as a *utility*, just like water, gas, electricity and telephony. It also embodies the desire of computing resources as true services. Software and computing platform and computing infrastructure may all be regarded as services with no concern as to how or from where they are actually provided. The potential of cloud computing has been recognized by major industry players such that the top five software companies by sales revenue all have major cloud offerings [2].

There is still no universal definition of cloud computing, however, there is sufficient literature available in the community that portrays the basic

principles [3,4]. The view taken by several authors is that cloud computing is an extension of cluster computing, or more specifically Cloud Computing = Cluster Computing + Software as a Service [3]. What is relatively clear is that cloud computing is based on five key characteristics, three delivery models, and four deployment models [5].

2.1. Characteristics and models

The five key characteristics are:

- On-demand self-service
- Ubiquitous network access
- Location independent resource pooling
- Rapid elasticity
- Pay per use.

The three delivery models are:

- Software as a Service (SaaS): use of provider's applications over a network
- Cloud Platform as a Service (PaaS): deployment of customer-created applications to a cloud
- Cloud Infrastructure as a Service (IaaS): rental of processing, storage, network capacity, and other fundamental computing resources.

To be considered "cloud" the delivery models must be deployed on top of cloud infrastructure that satisfies the five characteristics.

The four deployment models are:

- Private (internal) cloud: enterprise owned or leased, behind a firewall
- Public (external) cloud: sold to the public, mega-scale infrastructure (e.g. Amazon EC2)
- Hybrid cloud (virtual private cloud): composition of two or more clouds (e.g. Amazon VPC)
- Community cloud: shared infrastructure for specific community (e.g. academic clouds).

2.2. Cloud architecture

Physically a single-site cloud is realized as a datacenter, which consists of:

- Compute nodes (split into racks)
- Switches, connecting the racks
- A network topology, e.g., hierarchical
- Storage (backend) nodes connected to the network
- Front-end for submitting jobs
- Services: physical resource set, software services.

A geographically distributed cloud may consist of multiple such sites (distributed datacenters) and each site perhaps with a different structure and services.

Logically a cloud consists of a front-end and a back-end which are connected through a network. The front end is generally a web browser or any application which is using cloud services. The back end is the network of servers, system and application software and data storage system. Servers are typically organized into server farms to suit specific application software. The back-end is generally a three-tier arrangement (Figure 1), comprising: physical machines and storage, virtual machines and a service level agreement layer (SLA). The SLA is responsible for the monitoring of the service contract to ensure its fulfillment in real-time.

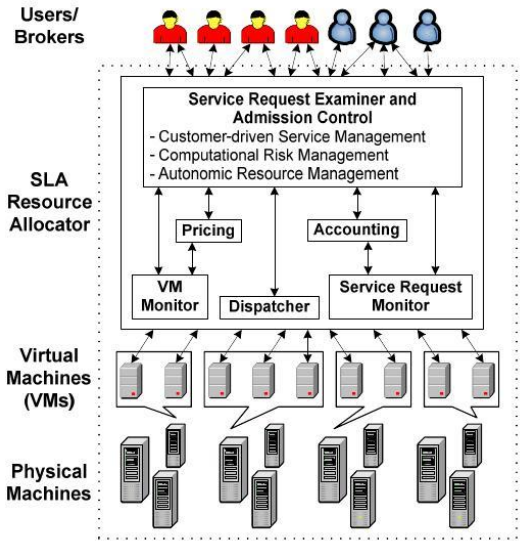

Figure 1. Cloud computing layers [6]

The actual service contract will also detail criteria associated with any computer forensic investigations, such as jurisdiction and data seizure. The jurisdiction covers the local laws that apply to the service provider and consumer. Data seizure covers the seizure of the provider's equipment to capture data and applications belonging to a particular consumer. The contract will also detail how such seizure is likely to affect other consumers that use the same provider. This inability to seize only the data relating to an individual suspect is one of the problems relating to forensic investigations of cloud datacenters that will be discussed further in section 5.

Cloud datacenters operate on the assumption that the demand for resources is not always consistent amongst clients and as a consequence the physical servers are unable to run at their full capacity. To accommodate this server virtualization technique are used. Server virtualization, which is discussed further in section 4, is a method of running multiple independent virtual operating systems on a single physical computer. Through server virtualization

cloud providers can maximize physical resources to maximize the investment in hardware.

Data is central to cloud computing and data security is of utmost importance in cloud datacenters. All data are backed up at multiple locations, which dramatically increases the data storage to multiple times in clouds. This also presents issues to computer forensic investigations. On the one hand data is likely to be mobile as it is moved amongst servers and it may be difficult to seize the original data with multiple copies in existence. On the other hand the availability of backups taken over time may provide useful evidence that would otherwise have been overwritten. These issues, amongst others, are discussed further in section 5.

2.3. Cloud examples

Two popular cloud computing facilities are Amazon Elastic Compute Cloud (EC2) and Google App Engine. Amazon EC2 is part of a set of standalone services which include S3 for storage, EC2 for hosting and the simpleDB database. Google App Engine is an end-to-end service, which combines everything into one package to provide a PaaS facility. With Amazon EC2, users may rent virtual machine instances to run their own software and users can monitor and increase/decrease the number of VMs as demand changes. To use Amazon EC2 users would:

- Create an Amazon Machine Image (AMI): incorporate applications, libraries, data and associated settings
- Upload AMI to Amazon S3
- Use Amazon EC2 web service to configure security and network access
- Choose OS, start AMI instances
- Monitor & control via web interface or APIs.

Google's App Engine allows developers to run their web applications on Google's infrastructure. To do so a user would:

- Download App Engine SDK
- Develop the application locally as a set of python programs
- Register for an application ID
- Submit the application to Google.

Having provided an overview of cloud computing we may now consider computer forensics before we then proceed to consider the two together.

3. Computer forensics – law enforcement perspective

The In this paper we use the term *computer forensics* to refer to the process of investigating computing devices based on off-the-shelf operating systems (Windows, Unix, MacOS) that would typically be found in cloud computing environments. However, in the strict sense *digital forensics* is the more general term to classify the forensic processes applied to a variety of digital devices in order to acquire digital evidence. Digital forensics includes, amongst others: computer forensics, intrusion forensics, network forensics and mobile device forensics.

Computer forensics has its roots data recovery and factors in additional guidelines and procedures designed to create a legal audit trail. Intrusion forensics is a branch of digital forensics that has its roots in intrusion detection and is concerned with attacks or suspicious behaviour directed against computers. Network forensics focuses on the network as the source of possible evidence and involves the monitoring and analysis of network traffic for information gathering. Often a combination of intrusion and network forensics techniques will be used to deal with attacks for which network traffic is significant. Mobile device forensics is concerned with the recovery of evidence from mobile devices, primarily mobile phones, due to the abundance of mobile phones used in conventional crimes. Our consideration of forensic procedures applied to cloud computing is largely concerned with computer forensics with some consideration of intrusion/network forensics.

There is to date no single definition of computer forensics, which is often regarded as more of an art than a science, although several similar definitions are available [7,8,9]. A suitable definition for the purpose of this paper is that according to [7], namely: "The application of computer investigation and analysis techniques to determine potential evidence". This definition suffices as it contains the three important keywords and phrases underlying computer forensics, namely: "computer", "investigation/analysis" and "evidence". Central to computer forensics is evidence, or more particularly digital evidence, which we consider below.

3.1. Digital evidence

Digital evidence is defined by [10] as: "Any information of probative value that is either stored or transmitted in a digital form". Typically, digital evidence may include files stored on a computer hard drive, file fragments or data items stored in memory, digital video or audio, or packets transmitted over a network. Digital evidence presents several challenges over its conventional counterpart and these challenges are born out of the characteristics of

digital evidence:

- Vast quantity of potential evidence: tens of thousands of files in a single computer – let alone a network
- Easily contaminated: rebooting a system may remove vital traces of evidence and contaminating some evidence may contaminate it all
- Crime identification: a crime may not become apparent for months or years (e.g. fraud)
- Vast number of potential suspects: several million Internet users.

Digital evidence (once gathered) must satisfy the same legal requirements as conventional evidence, i.e. it must be:

- Authentic – the evidence must be original and relate to the alleged crime under investigation
- Reliable – the evidence must have been collected using reliable procedures that if necessary could be repeated by an independent party to achieve the same result
- Complete – the evidence may be used to prove guilt as well as innocence
- Believable – the evidence should be convincing to juries and presented in such a way that they can make sense of it
- Admissible – the evidence was collected using procedures that In conform to common law and legislative rules (i.e. Admissible).

The additional problem with digital evidence is that it exists in both a logical context and a physical context. Data is stored physically on storage media (e.g. a hard drive) in blocks or clusters. However, as blocks or clusters are difficult for human users to make sense of data is grouped together in logical constructs such as files and directories, which users are comfortable with. Both the physical and the logical context need to be considered during the acquisition, analysis and presentation of digital evidence. While digital evidence is central to any computer forensic investigation, the procedures used to acquire the evidence and subsequently analyze and present the evidence prove just as important and such procedures are considered further below.

3.2. Dynamic evidence – time-lining

Static digital evidence artefacts alone are often insufficient when it comes to establishing a case for prosecution, what is needed is a collective of related artefacts that effectively facilitate the reconstruction the digital environment in which the alleged crime took place. In line with physical forensic science computer forensics aims to reconstruct a series of events linking a suspect to a crime using available evidence. This can prove very time consuming due to the sheer quantity of potential evidence in a digital environment. Typically an investigation may entail the linking of mobile phone conversations, or reconstructing the sequence of events in hacking attacks between victim, target and intermediates. Emphasis is placed on event-based reconstruction and *time-lining.*

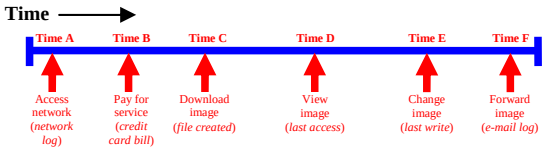

Figure 2. Example time-line of events

Time-lining provides an association of timestamps with each event or data item of interest in order to reconstruct a sequence of events. Time-lining is assisted by the fact that the majority data items are time-stamped. As shown in Figure 2 the sequence of events when an image is downloaded and changed can be time-lined to provide a more complete picture. Time-lining can use time-stamps such as file creation, access, modification times, which when correlated with other information build up time graph of activities that are consistent with non-computer crime events

3.3. Procedures

In addition to digital evidence computer forensics investigations is also characterized by procedures, or more formally *process models*. The process models specify generalized steps that are used to conduct a complete investigation. The steps cover the practical and theoretical aspects of an investigation and most importantly the legal aspects. Currently in the UK, much of the computer forensics work is conducted by law enforcement agencies (Police, UK Border Agency, Customs and Excise) and the process models reflect a law enforcement ethos. Although law enforcement agencies provide the main driving force, the need for computer forensics in the corporate sector is gaining momentum, particularly in the US.

Police forces in the UK adhere to a guide which specifies the principles and procedures that should be followed when dealing with incidents involving digital evidence. The Association of Chief Police Officers (ACPO) Guidelines for Computer Investigations and Electronic Evidence [11] is a

thorough and complete document, which specifies the procedures and steps that officers should take in dealing with a variety of situations associated with computers and digital evidence. By following such guidelines officers can guarantee that any evidence satisfies legal requirements, i.e. it is: authentic, reliable, complete, believable and admissible.

For the purpose of this paper we may regard the computer forensic process according to the six stage model proposed by [12].

- Identification: determine items, components and data possible associated with the allegation or incident; employ triage techniques
- Preservation: ensure evidence integrity or state
- Collection: extract or harvest individual data items or groupings
- Examination: scrutinize data items and their attributes (characteristics)
- Analysis: fuse, correlate and assimilate material to produce reasoned conclusions
- Presentation: report facts in an organized, clear, concise and objective manner.

This six stage process model and several other similar models [13,14] form the basis of the majority of computer forensic investigations. Crucial to any forensic investigation is the preservation of evidence such that the original evidence is not changed in any way. With respect to examination many forensic investigations involve examining a computing device when it is switched off and has no electrical power (so called 'dead' forensics). Occasionally, a 'live' forensic investigation is performed where the computing device is found in a switched on state. Under such a situation vital evidence may be gathered from examining the device's memory and the processes and network connections that are currently active.

3.4. ACPO principles and guidelines

The ACPO principles are stated below and it is essential that computer forensic investigations withhold these principles. Later in section 5 we discuss the issues that cloud computing raises in relation to these principles.

- Principle 1: no action taken by law enforcement agencies or their agents should change data on a computer or storage media which may subsequently be relied on in court
- Principle 2: in circumstances where a person finds it necessary to access original data held on a computer or storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions
- Principle 3: an audit trail or other record of all processes applied to computer-based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same results
- Principle 4: the person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

In addition to the above principles the ACPO guide describes best practices to be followed for each stage in a computer forensic investigation. The best practices range from how to conduct the search and seizure stage of an investigation through to presenting evidence in court and dealing with witness statements. The ACPO search and seizure guidelines are particularly relevant when considering cloud computing as these are the most difficult to satisfy due to the remoteness of cloud datacenters. The search and seizure guidelines describe how investigators should prepare for the search and record all details of the investigation scene and if necessary take photographs and video footage. The guidelines go on to describe how equipment should be seized and 'bagged and tagged' (to avoid tampering). Any seized storage media should then be cloned or imaged, as described below before the analysis can commence. Analysis is usually conducted at the *physical* level were disk partitions are examined and then at the *logical* level on a file-by-file basis. Later in section 5 we consider how such search and seizure would be impractical to conduct during a cloud datacenter.

Analysis of storage media should take place on a bit-by-bit clone or image of the original media (typically a hard disk). It is important that the image is an exact copy of the original media so that it contains deleted files and areas of the media that a normal backup would not copy. Once the image has been taken, both the original and image must then be authenticated, which typically involves computing a checksum for the original and the image at time the image was taken. This authentication may be achieved through a one-way hash function, such as MD5, which can provide a unique hash of a file or a complete disk image and any subsequent modification will alter MD5 signature.

Having considered the characteristics of digital evidence and the procedures to be followed in computer forensics investigations we are already able to see how such investigations will face challenges in a cloud computing environment. To a

large extent the nature of computer forensics relies upon direct access to possible sources of evidence. However, where cloud computing is concerned, such direct access is not possible as the cloud exists as a remote datacenter, typically in another country. However, as we shall see later in section 5, cloud computing does bring several advantages to the table where computer forensics is concerned.

3.5. Computer forensics and cloud computing

As we have seen through this section computer forensics (and its variants) is a rapidly increasing important discipline, which has come about and flourished due to the abundance of computing devices and indeed their uses with crimes, both conventional and hi-tech (digital) crime. If we consider mobile phones, saturation of mobile telephony in a country is generally acknowledged to have been achieved when 82 percent of the population own a mobile phone. In the UK it is reported that over 73 percent of the population own a mobile or have access to a mobile [15]. It is estimated that 6 out of every 10 crimes committed will involve some use of a mobile phone, which may range from road traffic accidents up to the more serious crimes such as murder. With this abundance of digital crime computer forensics has had to formalize, adapt and evolve into a methodology capable of supporting today's law enforcement officers.

Cloud computing is touted as the next major step change in the way that organizations plan, develop and enact their IT strategies. However, where computer forensics is concerned, cloud computing has not been thoroughly considered in terms of its forensic readiness. However, cloud computing has carefully considered security and indeed it was forced to right from the very outset. The reason for this is that security is an essential requirement for any IT application – no individual, or organization wants insecure data and they don't want their personal data exposed to any unauthorized users. On the other hand, computer forensics or forensic readiness is not an essential requirement, it is seen as more of a luxury. This is largely due to a lack of legislation, on a global scale, requiring that any computer installation implements a forensic readiness plan. To a certain extent, one cannot argue with this as security is required at all times, whereas computer forensics is only required when an incident takes place. However, for computer forensics to be successful, it generally requires that certain measures are taken before the actual incident occurs (e.g. some form of logging is enabled). As we shall discuss shortly, certain aspects of the computer forensic process can be applied to cloud computing, but the main stumbling block is the fact that it may be impractical for the computer forensic investigators to get their hand on the physical devices likely to contain digital evidence. This in turn suggests that an alternative or revised computer forensic process needs to be developed to meet the needs of cloud computing investigations.

4. Virtualization

In computing terms *virtualization* is a broad term that refers to the abstraction of computing resources. Virtualization abstracts a physical resource into a virtualized resource that can be shared. A useful analogy is to consider the water supply to a house as a resource:

- Option 1 - have your own well (physical)
- Option 2 - water supplied by water service (virtualized)
- Option 3 - buy bottled water (cloud solution - no long term contract, access water as and when needed).

The main incentive for virtualization is that it enables multiple users to share the same resources but maintains separation based on data or application owner. Within a cloud many resources can be virtualized: servers, storage, software, platform, infrastructure, etc. and for this reason virtualization is used extensively. Server virtualization is the most widely used form of virtualization through technologies such as VMware, and Citrix XenServer. With server virtualization one physical machine is divided into many virtual servers (also called virtual machines or VMs). At the core of such virtualization is the concept of a hypervisor (virtual machine monitor). A hypervisor is a thin software layer that intercepts operating system calls to hardware. Hypervisors typically provide a virtualized CPU and memory facility for the guests they are hosting. In the case of the water supply example, the water company represents the hypervisor by managing the relationship between the physical supply (reservoirs, pumping stations) and the virtual consumers (homes).

From a computer forensics point of view several authors have assessed the advantages/disadvantages of virtual machines in relation to computer forensics investigations [16,17,18]. In general, the findings are mixed, VMs can provide several advantages, for example VMWare provides a snapshot facility, which can be used to provide a 'picture' of your system at the time the snapshot is taken. The snapshot provides an image of the computer's hard drive, which consists of the data on the hard drive, the VMware configuration for that virtual machine and the BIOS configuration. There are also a number

of snapshot files that are created when the snapshot is first taken and these files contain the changes that have occurred to the virtual machine since the snapshot was taken. Thus, over time, the snapshot files will grow as the machine is used more and more. Taken together, the snapshot may seem like an ideal source of potential evidence, however, the use of VM artifacts in court is still questionable. On the downside it is argued that there are notable changes to a VM environment when a VM image is booted into a new environment intended to faithfully re-create the original. Once the image is booted new data may be written thus modifying it. Any image which is known to have undergone change in some way would be challenged in court, which is why the traditional "make a bit-wise copy of the original…" approach is still preferred when it comes to presenting evidence in court.

Although there is a degree of scepticism relating to the use of virtual forensic investigations the potential has been witnessed within the computer forensic community [18]. To a certain extent this has been driven by the upsurge in the use of virtualization within organizations. It is also the case that the use of virtualization may prove beneficial when investigating a suspect system, which itself uses virtualization. However, conventional computer forensic techniques can be used to investigate a suspect system, which does not use virtualization. This may serve as a pointer that cloud forensics should indeed be based on virtualization. Indeed several researchers are active in developing APIs for use with virtualization, which could have benefits for cloud computing. One such project is VIX [17], which considers Virtual Introspection for Xen virtualization. Virtual Introspection makes it possible for the state of a virtual machine to be monitored and examined from a Virtual Machine Monitor (VMM) or other virtual machine. during operation.

We conclude this section with some concrete use cases relating to virtualization in cloud computing. Through virtualization a typical cloud use case could be 40,000 VMs provided by 512 Servers with 1000 users. Such a cloud may typically contain 128TBs of storage across multiple storage technologies and 48TBs of memory. A real cloud example in the use of virtualization is the Amazon EC2 which employs Xen VMs in one of three sizes: small, large or extra large, which relate to EC2 compute units. Each VM instance is sized according to platform (32-bit or 64-bit), the amount of memory available, the amount of instance storage and the number of EC2 compute units.

5. Cloud computing – forensics pros and cons

At the time of writing opinion is somewhat divided as to whether or not cloud computing would assist computer forensics investigations or resist such investigations. We begin this section by generalizing both sides of the argument and then go on to consider the pros and cons in more detail. On the one hand the computer forensic process model would need to change and adopt a different set of procedures to accommodate investigations performed on cloud systems. On the other hand computer forensic investigations could take advantage of the services and resources provided by cloud systems to assist the investigation. In the sections below we elaborate further on these arguments.

5.1. Pros

The main benefit of cloud computing is centralized data, having the data all in the same place assists in forensic readiness**,** which leads to quicker, coordinated response to incidents. With centralized data IaaS providers can build a dedicated forensic server within the cloud, which is ready for use when needed. Other benefits to computer forensics stem from the services and resources that cloud systems can offer, or more precisely the scale and power of these services. Firstly, the availability of potentially peta-bytes of storage and high availability compute intense resources come as a great advantage to the computer forensic investigator. Over a period of time an investigator may amass a number of hard drive images, which could potentially be stored on the cloud, taking advantage of IaaS. Indeed, this approach has been used by [19], who describes how a number of images were transferred to Amazon's S3 using the HTTP/REST API. Secondly, the high availability compute intense resources can be used for compute intense jobs that forensic investigators may need to carry out. For example, forensic investigators may need to crack passwords, encryption keys or examine many images, all of which can be costly in terms of CPU and memory.

Additional benefits include inbuilt hash authentication for authentication of disk images, as mentioned previously. For example, Amazon S3 generates an MD5 hash when an object is stored, which means that it is no longer necessary to generate time-consuming MD5 checksums. In a forensic investigation, various log files can provide a rich source of information. However, logging is often an afterthought and consequently insufficient disk space is allocated and logging is either non-existent

or minimal. The scale of cloud storage implementation means that logging can be performed and tuned to a required level and logs can be made available as required. Modern operating systems offer extended logging in the form of a C2 audit trail. However, this is rarely enabled for fear of performance degradation and log size. With cloud computing enhanced logging can be realized and the granularity of logging can be set accordingly.

A final issue, which is thought of as both a benefit and a drawback is virtualization. As mentioned previously, virtualization is used in clouds to allow multiple users to share the same resources and many resources can be virtualized – software, platform, infrastructure etc. It was also mentioned that forensically sound collection of data involves a bit-by-bit duplicate of a disk image using appropriate software. In live investigations the acquisition of memory images is a more involved time consuming task of freezing memory before removing power form a host to be duplicated. However these mechanisms are not necessary within a virtual environment where disk and memory images can be collected quite easily via snapshot and other administrative functionality. However, this method of acquisition has yet to be proven forensically sound by law enforcement agencies (i.e. ACPO guidelines).

5.2. Cons

The main drawback of cloud computing from a forensic perspective is that of data acquisition – knowing exactly where the data is and actually acquiring the data. The search and seizure procedures used in the conventional computer forensic process are impractical due to evidence being stored in cloud datacenters. It is also difficult if not impossible to maintain a chain of custody relating to the acquisition of the evidence. Essentially, cloud computing means that investigators are unable to conform to the ACPO guide, as it is difficult if not impossible to satisfy ACPO principles. The ACPO guide specifies four basic principles relating to procedures and level of competency required for the handling of evidence. As clouds exist as remote datacenters these principles cannot be satisfied, which consequently makes the ACPO guide redundant, which in turn would cast doubt over the evidence's authenticity, integrity and admissibility in a UK court of law. Overall, there is a general loss of control over the forensic investigation process simply due to the data being stored elsewhere, where it is inaccessible. This in turn hinders crime scene reconstruction as the lack of knowledge of where data is actually stored means that it is difficult to piece together a sequence of events and create a timeline. In addition to data acquisition and loss of control there are several other drawbacks which can hinder the investigation, which are discussed further below.

The loss of important artefacts, which could be potentially crucial evidence. For example registry entries, temporary files and memory may be difficult, if not impossible to access in cloud datacenters (generally due to virtualization). Metadata may also be lost if data is downloaded from a cloud. Metadata such as file creation, modification and access times can provide a useful source of potential evidence to the forensic investigator. Although some cloud systems (Amazon S3) do provide a means to authenticate data (via MD5 checksums), many investigators still prefer to perform their own authentication, rather than rely on cloud hash authentication.

A further shortcoming is the lack of tool support available for dealing with cloud datacenters. Although computer forensics is a relatively new discipline, it has matured to the point where there is sufficient tool support for dealing with conventional localized investigations. Tools such as EnCase, Helix and FTK can be used to assist the forensic investigator with tasks ranging from the initial data acquisition through to providing written documentation, presentable in a court of law.

The final problem stems from the legal/people aspect of computer forensics in that whatever digital evidence is acquired from an investigation it must still be presented to a jury, who will pass judgement on a case. In conventional computer forensics investigators have to present their findings to the jury and this often requires that the investigator needs to explain, using technical jargon, how the evidence was acquired and what exactly the evidence means. This can prove challenging when dealing with conventional localized computer systems, let alone cloud datacenters which may be several thousand miles away, running 40,000 VMs across 512 servers accessed by 1000 tenants of which the accessed is one. This may prove far too much for a jury member to comprehend, give that on average the jury will only have a basic grasp of using a home PC!

6. Conclusions

Through this paper we have considered cloud computing as a notable step change which will affect future practices in computing and IT. We also considered computer forensics as a process used largely by law enforcement agencies to acquire digital evidence associated with some alleged crime or incident. We then went on to discuss how cloud computing will impact on computer forensics investigations and considered both sides of the

argument in terms of pros and cons associated with cloud computing in relation to computer forensics. In conclusion we can say that this impact is yet to be taken up by either party. In other words cloud providers have not yet fully addressed how they will implement forensic readiness. Similarly, forensic investigators have not yet put forward procedures for dealing with cloud investigations. One could argue that the ball is very much in the court of the forensic investigators. However, computer forensics is still an evolving discipline and it has rose to previous challenges in the past. Mobile phones, wireless technology, encryption, and live memory analysis have all presented challenges to computer forensics, yet these have been dealt with by the computer forensic community to expand the armoury of law enforcement agencies. To conclude, the authors are confident that the computer forensic community will rise to the challenge of cloud computing and it will likewise be met with standardized processes and further tools in the computer forensics armoury.

7. References

[1] Summers, C., (2003) Mobile phones – the new fingerprints, BBC News Online, http://news.bbc.co.uk/1/hi/uk/3303637.stm (23 June 2010).

[2] Chow, R., Golle, P., Jakobsson, M., Shi, E., Staddon, J., Masuoka, R., Molina, J., (2009) Controlling data in the cloud: outsourcing computation without outsourcing control, Proceedings of the 2009 ACM Workshop on Cloud Computing Security, Chicago, Illinois, USA, ISBN: 978-1-60558-784-4.

[3] Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., (2010) A view of cloud computing, Communications of the ACM, vol. 53(4), ISSN: 0001-0782.

[4] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I., (2009) Cloud computing and emerging IT platforms: vision, hype and reality for delivering computing as the 5th utility, Future Generation Computer Systems, Vol. 25(6), Elsevier Science Publishers, ISSN: 0167-739X.

[5] NIST (2010) Definition of cloud computing v15, Computer Security Division, Computer Security Resource Center, http://csrc.nist.gov/groups/SNS/cloud-computing/ (18 June 2010).

[6] VIVEK School of ERP, (2010) Cloud Computing, http://acharyavivek.blog.co.in/2010/04/12/cloud-computing-2/ (21 June 2010)

[7] Mohay, G., Anderson, A., Collie, B., De Vel, O., McKemmish, R., (2003) Computer and intrusion forensics, Artech House, Boston, ISBN 1-58053-369-8.

[8] Li, X., Seberry, J., (2003) Forensic computing, Lecture Notes in Computer Science, Springer, vol. 2904/2003, ISBN: 978-3-540-20609-5.

[9] Caloyannides, M., (2001) Privacy protection and computer forensics, Artech House, Boston, ISBN 1-58053-830-4.

[10] Standard Working Group on Digital Evidence (SWGDE) (1999) Digital evidence standards and principles, http://www.swgde.org/documents.html, (20 June 2010).

[11] 7Safe Computer Forensics, (2009) ACPO guidelines for computer investigations and electronic evidence, http://www.7safe.com/electronic_evidence/ (20 June 2010).

[12] Stephenson, P., (2003) Modeling of post Incident root cause analysis, International Journal of Digital Evidence, vol. 2(2).

[13] Reith, M., Carr, C., Gunsch, G., (2002) An examination of digital forensic models, International Journal of Digital Evidence, vol. 1(3).

[14] Carrier, B., Spafford, E., (2003) Getting physical with the digital investigation process, International Journal of Digital Evidence, vol. 2(2).

[15] Institute for Communications, Arbitration and Forensics, (2010) Mobile phone security solutions http://www.security-technologynews.com/article/mobile-phone-security-solutions.html (21 June 2010).

[16] Bem, D., Huebner, E., (2007) Computer forensic analysis in a virtual environment, International Journal of Digital Evidence, vol. 6(2).

[17] Hay, B., Nance, K., (2008) Forensic examination of volatile system data using virtual introspection, ACM SIGOPS Operating Systems Review, vol. 42(3) ISSN: 0163-5980.

[18] Shavers, B., (2006) VMWare as a forensic tool, Forensic Focus, http://www.forensicfocus.com/vmware-forensic-tool, (22 June 2010).

[19] Garfinkel, S., (2007) Commodity grid computing with Amazon's S3 and EC2, login, USENIX, vol. 32(1), pp7-13.

