Cloud Computing: The Digital Forensics Challenge

Stander, Adrie; Meyer, Gertruida

Published in:
Proceedings of the Informing Science and Information Technology Education Conference

Proceedings of Informing Science & IT Education Conference (InSITE) 2015

Cite as: Meyer, G., & Stander, A. (2015). Cloud computing: The digital forensics challenge. *Proceedings of Informing Science & IT Education Conference (InSITE) 2015,* 285-299. Retrieved from http://Proceedings.InformingScience.org/InSITE2015/InSITE15p285-299Meyer1562.pdf

Cloud Computing: The Digital Forensics Challenge

Gertruida Meyer and Adrie Stander
University of Cape Town, Cape Town, South Africa

Truiam.1@gmail.com **Adrie.stander@uct.ac.za**

Abstract

Cloud computing has brought many to the world of computers. Unfortunately it also created new opportunities for criminal activities. This paper used a descriptive literature review to determine the issues that are currently influencing digital investigations. The results show that there are many unaddressed issues affecting the identification, preservation and acquisition of evidence in the cloud. Very little research has been done to solve these problems and where research exists, it is still far from being implemented in practice.

Keywords: Cloud Forensics, Digital Forensics, Cybercrime, Cloud Computing.

Introduction

Cloud computing is changing how information services are created and used. Spending on cloud computing is growing at five times the rate of traditional, on premises, information services and technology (Ruan, Carthy, Kechadi & Crosbie, 2011). Cloud computing is an evolution of technology in a model of multiple stakeholders, location independent, elastic, on demand metered supply of computing resources (Dykstra & Sherman, 2012). Resources include networks, servers, storage, processing, applications and services. Cloud computing creates an ability for users to provision and scale computer resource costs efficiently without the complexity of the technology and requirements for any kind of local infrastructure (Grispos, Storer & Glisson, 2012; Zawoad, Dutta & Hasan, 2013).

According to Garfinkel,(Garfinkel, 2010) in the Golden age of digital forensics, digital forensic tools were easily developed for use in investigations due to the fact that encryption was seldom used. The tools allowed for non-technical people to search for data like email messages, deleted files and limited basic file carving. But this period is fading quickly and digital forensics are faced with new technology in operating systems, file formats, growing data sizes, growing, large storage devices and also cloud computing that are challenging all the basic principles of digital forensics (Garfinkel, 2010).

Cloud forensics are the process of digital forensics in the cloud and the traditional methods of digital forensics are challenged by the admissibility of evidence in a court of law due to several issues. Examples of issues are the decentralization of data, segregation of customer data, jurisdictional areas, loss of metadata and the chain of custody.

Editor: Eli Cohen

The objective of this research is to identify the issues that exist in cloud forensics and to create a classification through systematic review of the literature using grounded theory.

Cloud Computing

The National Institute of Standards and Technology's (NIST) defines "cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model is composed of five essential characteristics (on-demand self-service; broad network access; resource pooling; rapid elasticity; measureable service), three service models (Software as a Service; Platform as a Service; Infrastructure as a Service), and four deployment models (private cloud, public cloud, community cloud, hybrid cloud) (Mell & Grance, 2011).

Cloud storage has become a popular option to store data and to access the stored data via a wide range of internet connected devices like laptops, smart phones and tablets. The cloud customer can access the services either through a web browser or client installed software on the device. There are various cloud storage offerings available either for free or a pay-for-service. Gartner in a report highlighted the trend to store personal data in the cloud and predict that this trend will replace PC's as the main personal data storage by 2014 (Quick & Choo, 2013a; Abbadi & Lyle, 2011).

Cloud Forensics

Cloud forensics is the process to retrieve digital evidence from the cloud for investigative purposes. Adversaries use cloud computing in different ways to commit crimes, including storing incriminating evidence like child pornography, launch attacks and crack encryption keys. The adversaries can provision a cloud instance, commit the crime, and immediately de-provision the cloud instance to destroy the evidence. The inaccessibility of data, potential lack of information and unknown provenance of evidence are major concerns for digital forensics in the cloud and can result in a situation where evidence may not be available or where the integrity of the evidence cannot be verified on the systems used for cloud computing (Casey, 2012).

Cloud computing is at risk to be affected by inconsistencies, whether it is maliciously intended faults or unintentional faults. Maliciously intended faults can be caused by external or internal adversaries to cause faults on Cloud Service Provider's (CSP's) servers (instances) or applications. The cloud characteristics of virtualised multi-tenant environments can create greater risks to both the CSP and cloud customer because if an instance is compromised, both the guests and the host are at risk.

The virtual machine where the cloud instance is hosted for the cloud customer may contain potential evidence in most cases where the incident occurred. The network layer in the cloud instance, can also contain potential evidence. On the Cloud customer device, the client used to access the cloud (web browser) is usually the only application that contain evidence (Birk, 2011).

Forensic Issues in Cloud Forensics

Evidence acquisition in cloud computing has been pointed out by several researchers as a major issue in cloud forensics (Dykstra & Sherman, 2012). Research suggested a clear segregation of duties is recommended in collecting evidence. It is unclear who should collect the volatile and non-volatile data for the cloud instance and how it should be collected (Dykstra & Sherman, 2012). There are no clear guidelines on legal issues for acquisition of evidence in cloud compu-

ting environments or decisions about e-discovery regarding remote data acquisitions. (Dykstra & Sherman, 2012)

Research

Objectives

The objective of the research is to determine the issues in Cloud Forensics through a systematic literature review and to identify concepts, patterns and groups and to classify the issues in categories.

Strategy

The strategy for the research is a descriptive literature review using Grounded Theory approach to create a classification of Cloud Forensics.

Scope of Literature Search

The scope of a Descriptive Literature Review is to locate relevant literature through manual and computer searches. As cloud forensics is a recent phenomenon the research focused on the prominent online databases that covers forty four of the IS World's top fifty journals (Levy & Ellis, 2006). The online databases targeted were the ACM Digital Library, IEEE Xplore Digital library, Elsevier (ScienceDirect), General OneFile and ProQuest(ABI/INFORM). The Google Scholar search engine was also used to locate literature. The literature selected was peer reviewed journal articles, conference proceedings and books.

Filtering and Coding Process

The keywords used to search were: cloud computing, digital forensics, computer forensics, and cloud forensics. Google Scholar cross references to cited literature was also used to drill into other literature not necessarily covered by the keywords searched. During the search process 260 articles, conference proceedings and book sections were downloaded.

The process continued to review the abstracts for possible relation to challenges or issues within cloud forensics. If needed the rest of the text was also scanned to determine if articles should be selected and the 260 articles were reduced to 74 articles.

The next step was to go through the 74 articles reading the full text. With Grounded Theory the objective is to use an inductive approach to develop a grounded theory around the core categories that emerges from the selected data. There are several different procedures advocated to collect and analyze data and the same advocates even vary in their procedures between editions of the same book. This research, however only focused on the procedures of Corbin and Strauss (2008).

Concepts emerged from the articles which constituted the Initial Coding process of Grounded Theory. As more articles were reviewed the coding became focused and the concepts were categorized. The 74 articles were reduced to 39 articles which addressed the categories and classification created from the review.

Classification Process

As concepts emerged from the articles the concepts were coded and categorized. The focus was around issues and challenges within digital forensics within cloud computing. Relationships between the categories were identified. The review process continued and the concepts were constantly compared with the already identified codes, categories and relationships. Categories and relationships were re-defined as more concepts emerged and the process of selective coding con-

tinued until conceptual saturation was reached, that is no further concepts, codes, categories and relationships could be identified.

Fourteen categories of issues and challenges were identified in the collection phase of cloud forensics and these fourteen categories were grouped into three processes of the Collection phase of digital forensics: identification of digital evidence; preservation of digital evidence and acquisition of digital evidence.

Table 1.Categories identified during Collection Phase

Digital Investigation Phase	Digital Investigation Process	Issue/Challenges
Collection Phase	Identification	Decentralization of data centres Decentralization of data logs Physical locations unknown or not accessible Specific logging volatile
	Preservation	Inaccessibility to virtual instance Dependency on CSP Metadata/Provenance protection Volatile data Evidence gathering
	Acquisition	Separation of Customer/Multi tenancy Protection of personal Information Layers of Trust Different jurisdictions/Bodies of Law Chain of Custody

Issues and Challenges

Identification

In the collection phase of digital evidence three categories were identified with issues and challenges. The categories are: decentralization of data centres, decentralization of data logs and physical location unknown or not accessible. The process of digital evidence identification is crucial in order to know where there the digital evidence is and how to access the digital evidence. If the digital evidence cannot be identified the first step, the chain of custody fails and there is no digital evidence admissible in a court of law.

Decentralization of data centres

Cloud computing's distributed architecture allows data to be created, stored, processed and distributed over several data centres and physical machines which are globally dispersed and also possibly dispersed into multiple geographical locations and jurisdictions. Data is replicated to other servers to ensure redundancy of data.

The stored data within a data centre is replicated and distributed at physical level and the data can also be fragmented across multiple data centres. The distribution of data depends on the data centres' performance and availability.

Data stored across multiple servers or storage devices complicated the identification of possible digital evidence and the collection of such evidence in cloud computing environments because direct access to remote data centres is not possible and the data centres can possibly be in other countries. The decentralization of data centres in other countries may create jurisdictional challenges during the search and seizure process to locate digital evidence.

References: (Biggs & Vidalis, 2009;Bouchenak , et al, 2013;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Poisel, Malzer & Tjoa, 2013;Quick & Choo, 2013b;Quick & Choo, 2013a;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Baggili, 2013;Trenwith & Venter, 2013;Wolthausen, 2009;Zagari & Smith, 2013;Zagari & Beford, 2012).

Decentralization of data logs

In digital forensics some of the most useful information is stored within log files. In a cloud computing environment, these logs are decentralized, as data stored in the cloud is replicated to multiple server and data centres to confirm redundancy of data. Multiple cloud users' log information may be stored together or can be spread over multiple servers. Cloud architectures consist of several layers and tiers and logs are generated in each tier. All of these layers produce logs that are very valuable for digital forensic investigations.

Even if the cloud customer specifies the location where the data should be stored, log files are decentralized as the replication of the data is decentralized and the cloud customer has very limited access to log files.

For an investigator to access the log files, the investigator needs to identify the locations of the log files and obtain access to these log files. Cloud service providers can assist the investigation by providing the log files, but the chain of custody of the evidence then becomes an issue. The collection of log files from multiple servers / layers and providing the log files securely to investigators have become extremely challenging

References: (Abbadi & Lyle, 2011;Bouchenak , et al, 2013;Grispos, Storer & Glisson, 2012;Ko et al, 2011;Marty, 2011;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Lamb, 2011;Trenwith & Venter, 2013;Zagari & Beford, 2012;Zawoad, Dutta & Hasan, 2013).

Physical locations unknown or not accessible

In digital forensics, access to the physical servers and actual data and to secure the evidential data is crucial to an investigation.

For both the cloud customer and the investigator in cloud environments, physical access to the servers where the data is stored is in most of the cases technically not possible due to the remote location of the data or the fact that the location of the data cannot be determined. The remote location of the data can be distributed over several different geographical locations over different jurisdictions making it difficult to determine which legal framework and procedure to use in the

evidence collection process. The challenge also exists in that there is a lack of physical proximity between the cloud customer and the cloud service provider. The cloud customer or investigator also has no access to the network routers, load balancers and other networking components within the cloud computing environment.

In the case where the locations are distributed over locations in different jurisdictions with different legal requirements, the process to access the data can be too slow due to serious cross border red tape and the adversary may have time to access the data and change or destroy evidential data.

The cloud customer in general has no control or very little control of where the customer's data is stored in the cloud. The cloud customer also has very little knowledge of the actual location or where the data is stored due to fact that the cloud service provider is not transparent about the physical locations of the servers where the data is stored.

Even if the cloud customer has specified in the Service Level Agreement (SLA) with the cloud service provider where the data needs to be stored, the cloud customer has no control where the data is stored and there is no mechanism to verify the actual location of the data.

Cloud computing's nature of virtualization also complicates accessing the physical hardware due to the complexity of identifying the location and to isolate the host servers where the virtual machines are running. The virtual machines can also be spread over different physical servers. Cloud service providers in principle only offer access to a logical representation of the data.

References: (Biggs & Vidalis, 2009;Birk & Wegener, 2011;Bouchenak , et al, 2013;Dillon, Wu & Chang, 2010;Dykstra & Sherman, 2012;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Grobauer & Schreck, 2010;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Ko et al, 2011;Martini & Choo, 2012;Poisel, Malzer & Tjoa, 2013;Quick & Choo, 2013b;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Trenwith & Venter, 2013;Wolthausen, 2009;Zagari & Beford, 2012;Zawoad, Dutta & Hasan, 2013).

Preservation

During the process of the coding and categorizing five categories emerged for the preservation of evidence. The categories are:

- Dependency on Cloud Service Provider (CSP)
- Inaccessibility to Virtual Instances
- Metadata/Provenance protection
- Specific logging volatile
- Volatile data

Dependency on CSP

In a cloud computing environment the cloud service provider has all the power over the environment and therefore controls the source of the evidential data. The process of preserving digital evidence in the cloud highly depends on the support that the investigator receives from the cloud service provider (CSP).

The investigator or law enforcement agency (LEA) currently requires the cooperation of the CSP to collect evidential data. A search warrant is issued to the CSP and the CSP execute the processes of preserving and collecting the evidential data. The investigation does not have initial control

over the chain of custody and there might also be a requirement that the CSP should put a litigation hold on the data to prevent the data from being destroyed or changed prior to the time that the data can be collected.

The CSP can also alter evidential data and logs if they themselves are under investigation. The investigation depends on the honesty of the employee of the CSP to collect the evidential data in a sound manner. The CSP has full control over the evidential data and overall there is a general loss of control over the investigation process.

CSPs have dependencies on other CSPs as data is pushed further back into the cloud computing environment. These dependencies are very dynamic and complex, thus increasing the dependency of the investigation on the CSP. Any break in the chain of dependencies between the CSPs will have a serious impact on the collection of evidential data.

The dependency of the cloud customer on the CSP to provide evidential data opens the need for CSPs to be transparent and to cooperate in investigations in order to support forensic practices and to assist law enforcement. The CSP transparency also includes the reporting of incidents and vulnerabilities of the infrastructure components in a timely manner to the cloud customer.

References: (Birk, 2011;Bouchenak , et al, 2013;Dykstra & Sherman, 2011;Dykstra & Sherman, 2012;Grispos, Storer & Glisson, 2012;Grobauer & Schreck, 2010;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Martini & Choo, 2012;Quick & Choo, 2013a;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Trenwith & Venter, 2013;Watson et al, 2012;Zagari & Smith, 2013;Zagari & Beford, 2012;Zawoad, Dutta & Hasan, 2013).

Inaccessibility to virtual instances

The virtualized nature of cloud computing has an impact on the collection of evidential data due to highly limited or no access to the virtual instance. Even in the cloud service scenario of IaaS, the customer's virtual machine (VM) is controlled by the cloud service provider (CSP) and the CSP is responsible for the hypervisors, network infrastructure right down to the physical hardware of the data centre.

In cloud computing the virtualization of data storage complicates the identification and isolation of physical storage devices where the cloud customer's data may be stored and processed. The virtualized data may be spread over different physical devices, different geographical locations and jurisdictions. In virtualized environments the collection of evidential data may need to occur through the virtualization software which can have an impact on the soundness of the evidence, chain of custody and the admissibility of the evidence in court.

From the CSP there is a lack of transparency for linkages between VMs, the physical host server location and how data is stored in virtual and physical memory. In cloud virtualization there is a general loss of control over the investigation process because the data is stored in different VM's where data is not accessible.

Adversaries can also shut down and terminate VM's used to commit crimes and in the process all evidential data and logs are lost due to the termination of the VM.

References: (Birk, 2011;Dykstra & Sherman, 2012;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Grobauer & Schreck, 2010;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Ko et al, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Wolthausen, 2009;Zagari & Smith, 2013;Zawoad, Dutta & Hasan, 2013).

Metadata/Provenance protection

Metadata, also known as data provenance, is the history of digital objects. Metadata describes ownership and the process history (create, modify and access) of data objects. Metadata is vital to the success of forensic investigations in order to determine the ownership of evidential data (who access the data) and the time-line of evidential data (when data accessed). The uncertainty about metadata and metadata availability are challenges for investigations in the cloud and who and when questions can remain unanswered if the supporting metadata is unavailable.

Re-construction of the correct time-line of events in the process of investigation requires that the correct time and time-zones to be established. This however can be challenging in a cloud environment because the data centres can be decentralized over different time zones. Metadata is lost if the file dates and times cannot be matched after collection.

The question still remains if evidential data is collected from multiple data centres and physical servers, whether the date stamps are consistent over all of the data centres and if the data can be trusted.

Only a few cloud service providers (CSP) have implemented mechanisms to secure metadata in the cloud environment.

References: (Abbadi & Lyle, 2011;Birk, 2011;Casey, 2012;Dykstra & Sherman, 2012;Grispos, Storer & Glisson, 2012;Ko et al, 2011;Martini & Choo, 2012;Quick & Choo, 2013b;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Trenwith & Venter, 2013;Zawoad & Hasan, 2012).

Specific logging volatile

Logs are very useful evidential data in an investigation. Logs include system logs, network logs, firewall logs and router logs. If the cloud service provider (CSP) does not run any logging application then no opportunity exists to collect specific logging information during an investigation. If the CSP does run logging applications, then the logs must be manageable sizes to be useful and to prevent wiping memory on hosting servers. Currently CSPs are not obligated to provide all logs and logs are not reasonably protected by the CSPs.

Another challenge in cloud computing is that that the cloud customer cannot gather network logs or router logs due to the fact that the underlying cloud architecture is under the control of the CSP. Firewall logs may also be volatile as the logs can contain information of other cloud customers accessing the same CSP.

In the case where a cloud service had been compromised, it is rarely feasible to shut down the entire network to preserve evidential data which include specific logging. Specific logging may no longer be available, may be difficult to collect or may be decentralized over data centres and physical locations.

Specific logging is very volatile due to virtualization. As the customer accesses a VM in the cloud, information written to the operating system like logs, registry entries and temporary files are lost as soon as the cloud customer exits the VM.

References: (Abbadi & Lyle, 2011;Birk, 2011;Birk & Wegener, 2011;Casey, 2012;Grobauer & Schreck, 2010;Ko et al, 2011;Martini & Choo, 2012;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Zagari & Beford, 2012;Zawoad & Hasan, 2012;Zawoad, Dutta & Hasan, 2013).

Volatile data

In cloud computing frequently used data may be stored in volatile memory or may be cached in the cloud customer's device during the interaction with the cloud. Volatile data is lost in cloud computing if the virtual machine is powered down or rebooted or if the incorrect preservation process was followed by the CSP in the process of evidence collection on behalf of the investigator.

The virtual instance could be abused by the adversary for sending spam, steal volatile data of the running systems, attack targets (internal or external). After the attack the adversary can cancel the contract with the cloud service provider (CSP) and the VM is powered down and most of the evidential data is lost in the process.

In cloud computing difficulty exist in establishing what data was processed or stored by what software on specific devices. The full complement of the data stored by an adversary in the cloud is not likely to be stored on the devices that might be seized for investigation. The evidential data needs to be preserved in a timely manner to prevent data from being modified or destroyed, but difficulties exist due to cross border red tape with the acquisition process.

The evidential data is more ethereal and dynamic in cloud computing and the question remains if the CSP has provided all available evidential data and whether the data remained unchanged in the collection process. It is not normally possible to go back to the original state of the data due to the dynamic nature of data processing in the cloud.

References: (Birk, 2011;Grispos, Storer & Glisson, 2012;Grobauer & Schreck, 2010;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Martini & Choo, 2012;Poisel, Malzer & Tjoa, 2013;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Trenwith & Venter, 2013;Wolthausen, 2009;Zagari & Beford, 2012).

Acquisition

Chain of custody

The chain of custody is the roadmap that shows how evidence was identified, preserved, acquired, examined and analyzed for the evidence to be admissible in court. The chain of custody refers to identification of devices, the physical control of devices, acquisition of data, whether the devices were running or powered down and also how the evidential data was preserved to prevent any further change to the evidential data. The chain of custody also documents all individuals who were in contact with that data.

In could computing environment the chain of custody of evidence has become an issue for admissibility of evidence in court. The investigator does not have access to the physical servers and cannot document how evidential data was preserved and acquired. The investigator may need to depend on the CSP for the collection of evidence in a highly dynamic environment. The investigator has to rely solely on the word of the CSP that the evidence was required in a sound way and any problem in the collection or any corruption of data can lead to a serious problem in the chain of custody.

References: (Dykstra & Sherman, 2011;Dykstra & Sherman, 2012;Grispos, Storer & Glisson, 2012); 19]; (Martini & Choo, 2012;Poisel, Malzer & Tjoa, 2013;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Trenwith & Venter, 2013;Zagari & Beford, 2012;Zawoad & Hasan, 2012;Zawoad, Dutta & Hasan, 2013).

Different Jurisdictions and bodies of law

The nature of cloud computing with decentralized data centres and virtualization makes it difficult to determine the physical location of the data, let alone the body of law which governs and also restrict the investigation scope.

A search warrant issued by local law may not give an investigator the right to search and seize evidence in the cloud located in a different geographical location under a different jurisdiction even if the cloud is accessible locally through the Internet. The process to access data across borders relies on the formal communication and legal processes between countries with different legal requirements to preserve and acquire data in the process of an investigation. To get the correct legal authorization to acquire data cross border may be very sensitive which gives an adversary ample time to modify or even destroy the data of interest.

Adversaries can use cloud computing environments as havens from where they can launch attacks or use cloud for illegal activities or the store illegal data. They can select clouds in geographical locations where cross border red tape may have huge implications for the acquisition of evidential data.

References: (Biggs & Vidalis, 2009;Blumenthal, 2010;Dykstra & Sherman, 2012;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Ko et al, 2011;Martini & Choo, 2012;Quick & Choo, 2013a;Quick & Choo, 2013b;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty, Gresty & Lamb, 2011;Watson et al, 2012;Wolthausen, 2009;Zagari & Beford, 2012;Zawoad & Hasan, 2012).

Evidence gathering

Data is lost when an adversary cancels a cloud contract and a VM is powered down as a result. Real network and router logs cannot be gathered by the cloud customer for forensic purposes. Evidence is untrustworthy due to the cloud service provider (CSP) involvement in the collection of evidential data. Metadata is lost in the process of evidence gathering and the integrity of evidence data gathered cannot be verified.

Another challenge is that evidence gathering has to be timely to prevent an adversary of destroying or modifying data in the time that the law enforcement agency (LEA) has to wait for search warrants due red tape of cross border investigations due to the distributed nature of cloud computing.

Due to the elastic resources of cloud computing environments and increasing storage capacity of devices and computer systems, there is really no limit to storage capacity and the investigator is faced with the problem of gathering extremely large volumes of data placed in the cloud by cloud customers. The amount of forensic data to be processed is outgrowing the ability to process the data in a timely manner.

The process of the evidence gathering in the cloud computing has become cumbersome and time consuming for the investigator and disruptive for the cloud customer. LEAs are increasingly finding that the data in question are either inaccessible or very difficult to access.

References: (Abbadi & Lyle, 2011;Biggs & Vidalis, 2009;Birk & Wegener, 2011;Birk, 2011;Casey, 2012;Dillon, Wu & Chang, 2010;Dykstra & Sherman, 2011;Dykstra & Sherman, 2012;Garfinkel, 2010;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Grobauer & Schreck, 2010;Hooper, Martini & Choo, 2013;Ko et al, 2011;Martini & Choo, 2012;Poisel, Malzer & Tjoa, 2013;Quick & Choo, 2013a;Quick & Choo, 2013b;Reilly, Wren & Berry, 2011;Reilly, Wren & Berry, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Taylor, Haggerty,

Gresty & Lamb, 2011;Trenwith & Venter, 2013;Wolthausen, 2009;Zagari & Beford, 2012;Zawoad, Dutta & Hasan, 2013).

Layers of trust

Cloud computing environment consists of several different layers and the different cloud services introduce several levels of trust. In SaaS the cloud customer is totally dependent on the cloud service provider (CSP) to provide evidential data from the applications down to the actual servers where data are processed and stored. In IaaS the cloud customer still needs to rely on CSP for evidential data from the physical host servers of the virtual machines (VM) of the cloud customer although the rest of the application layer is under control of the cloud customer. Thus, irrespective of service of SaaS, PaaS or IaaS there is a trust issue present when evidential data are collected by the CSP in the process of an investigation.

The CSP's lack of full transparency and to provide capabilities for tracking and auditing of file access history and metadata from the physical servers and VM contributes to the issue of trust. In the initial collection phase of an investigation the seizure of remote data may be executed by a CSP which require a certain level of trust and depends on the honesty of the CSP employee who is not a certified investigator. The trust issue involves questions of whether all the data was provided and whether the data remained unchanged in the process to be acceptable in court.

For a prosecution, the logs of who had access to the evidence and how the evidence was collected is of high importance to establish if evidence can be trusted and will be admissible in court.

References: (Birk, 2011;Dykstra & Sherman, 2012;Grispos, Storer & Glisson, 2012;Ko et al, 2011;Martini & Choo, 2012;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Lamb, 2011;Zawoad & Hasan, 2012;Zawoad, Dutta & Hasan, 2013).

Protection of personal information

Local privacy and/or data protection legislation differ from jurisdiction to jurisdiction. In cloud computing environments, data can be processed and stored across the globe in different jurisdictions. Data can be stored on a cloud server in a country where privacy laws are not enforced or non-existent.

In cloud computing, appropriate measures and controls should be applied to personal data in the process of investigation and unauthorized individuals may not access personal data. Detailed logs may contain sensitive and private information and access to this information should be controlled and authorized within the investigation team. The cloud service provider is also facing confidentially issues and is thus might be reluctant to provide raw data.

In a cloud computing environment, replication of data can store the same data over different countries which may apply different legislation to protect private information making it difficult to retrieve private and sensitive data.

References: (Blumenthal, 2010;Dykstra & Sherman, 2011;Grispos, Storer & Glisson, 2012;Hay, Nance & Bishop, 2011;Hooper, Martini & Choo, 2013;Ko et al, 2011;Ruan, Carthy, Kechadi & Baggili, 2013;Taylor, Haggerty, Gresty & Hegarty, 2010;Zagari & Smith, 2013;Zawoad, Dutta & Hasan, 2013).

Separation of customers / multi tenancy

The resource pooling nature of cloud computing has the effect that multiple customers can share the same physical server. In the event of an investigation the cloud service provider (CSP) has to assure that the customer or investigator does not have access to other cloud customer's data.

The collection process in an investigation should clearly define the segregation of customer information in the process of acquiring the evidential data and should not breach confidentiality of proprietary information of the tenants sharing the same resources. Detailed logs may show information from other cloud customers that are both private and sensitive and the investigator has to rely solely on the honesty of the CSP that another customer's data was not collected.

The CSP in the process of acquiring evidential data should ensure that other cloud customers who are not the target of the investigation are not impacted due to co-location. Thus the investigator cannot seize the physical server containing the customer's data because the other customers' needs to continue with operations.

References: (Biggs & Vidalis, 2009;Dillon, Wu & Chang, 2010;Dykstra & Sherman, 2011;Grispos, Glisson & Storer, 2013;Grispos, Storer & Glisson, 2012;Hooper, Martini & Choo, 2013;Ko et al, 2011;Martini & Choo, 2012;Ruan, Carthy, Kechadi & Baggili, 2013;Ruan, Carthy, Kechadi & Crosbie, 2011;Taylor, Haggerty, Gresty & Hegarty, 2010;Zagari & Beford, 2012;Zawoad, Dutta & Hasan, 2013).

Conclusion

Cloud forensics are still faced with a multitude of problems that have not been addressed yet by research to find solutions for the issues. Data acquisition in the cloud remains the biggest issue with many and varied problems. Much research is needed to develop procedures and tools that can be used by service providers to extract the data needed by investigators in a forensically sound way.

Service Level Agreements between cloud customer and CSP's can only protect the cloud customer up to a point if a transparent investigation cannot be done.

Legislation to retrieve evidence in a timely manner over borders and third party watchdogs to protect the personal rights of the cloud customer, is needed to protect both the cloud customer as well as the CSP's investment in technology infrastructure and service in the event of an incident. This is however highly unlikely to happen soon, given the differences in legislation in many parts of the world.

Cloud forensics workgroups comprising of forensic specialists; IT specialists, law enforcement agencies; law makers and other cloud parties should address the issues and find solutions that can be applied in any location to stop adversaries from using safe haven countries to commit their crimes in the cloud.

Given the current situation, with very few answers to the problems of legal differences across borders and techniques to ensure that data can be obtained in a forensically sound way, without compromise of the chain of custody, it can be foreseen that it will take a long time to find solutions to the issues highlighted in this paper.

References

Abbadi, I. M., & Lyle, J. (2011). *Challenges for provenance in cloud computing*. TaPP.

Biggs, S., & Vidalis, S. (2009). *Cloud computing: The impact on digital forensic investigations*. In International Conference for nternet Technology and Secured Transactions (pp. 1–6). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5402561

Birk, D. (2011). *Technical challenges of forensic investigations in cloud computing environments*. In Workshop on Cryptography and Security in Clouds (pp. 1–6).

Birk, D., & Wegener, C. (2011). Technical Issues of Forensic Investigations in Cloud Computing Environments. In *Systematic Approaches to Digital Forensic Engineering (SADFE),* 2011 IEEE Sixth International Workshop on, 1–10. doi:10.1109/SADFE.2011.17

Blumenthal, M. S. (2010). Hide and seek in the cloud. *Security & Privacy, 8*(2), 57–58.

Bouchenak, S., Chockler, G., Chockler, H., Gheorghe, G., Santos, N., & Shraer, A. (2013). Verifying cloud services : Present and future. *ACM SIGOPS Operating Systems Review, 47*(2), 6–19.

Casey, E. (2012). Cloud computing and digital forensics. *Digital Investigation, 9*(2), 69–70. doi:10.1016/j.diin.2012.11.001

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research*. Sage.

Dillon, T., Wu, C., & Chang, E. (2010). Cloud computing: Issues and challenges. *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, 27–33. doi:10.1109/AINA.2010.187

Dykstra, J., & Sherman, A. T. (2011). Understanding issues in cloud forensics: Two hypothetical case studies. *Journal of Network Forensics, 3*(1), 19–31.

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. doi:10.1016/j.diin.2012.05.001

Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation, 7, S64–S73*. doi:10.1016/j.diin.2010.05.009

Grispos, G., Glisson, W. B., & Storer, T. (2013). Using smartphones as a proxy for forensic evidence contained in cloud storage services. *2013 46th Hawaii International Conference on System Sciences*, 4910–4919. doi:10.1109/HICSS.2013.592

Grispos, G., Storer, T., & Glisson, W. (2012). Calm before the storm: The challenges of cloud computing in digital forensics. *International Journal of Digital Crime and Forensics, 4*(2), 28–48.

Grobauer, B., & Schreck, T. (2010). Towards incident handling in the cloud. *Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop - CCSW '10, 77*. doi:10.1145/1866835.1866850

Hay, B., Nance, K., & Bishop, M. (2011). Storm clouds rising: Security challenges for IaaS cloud computing. *2011 44th Hawaii International Conference on System Sciences*, 1–7. doi:10.1109/HICSS.2011.386

Hooper, C., Martini, B., & Choo, K. K. R. (2013). Cloud computing and its implications for cybercrime investigations in Australia. *Computer Law & Security Review, 29*(2), 152–163. doi:10.1016/j.clsr.2013.01.006

Ko, R. K., Jagadpramana, P., Mowbray, M., Pearson, S., Kirchberg, M., Liang, Q., & Lee, B. S. (2011). TrustCloud: A framework for accountability and trust in cloud computing. *2011 IEEE World Congress on Services*, 584–588. doi:10.1109/SERVICES.2011.91

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, *9*(1), 181-212.

Martini, B., & Choo, K. K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation, 9*(2), 71–80. doi:10.1016/j.diin.2012.07.001

Marty, R. (2011). Cloud application logging for forensics. *Proceedings of the 2011 ACM Symposium on Applied Computing - SAC '11*, 178. doi:10.1145/1982185.1982226

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing recommendations of the National Institute of Standards and Technology. *NIST Special Publication, 800*(145), 7.

Poisel, R., Malzer, E., & Tjoa, S. (2013). Evidence and cloud computing : The virtual machine introspection approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, (August),* 135–152.

Quick, D., & Choo, K. K. R. (2013a). Dropbox analysis : Data remnants on user machines. *Digital Investigation, 10*(1), 3–18. doi:10.1016/j.diin.2013.02.003

Quick, D., & Choo, K. K. R. (2013b). 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. doi:10.1016/j.diin.2013.07.001

Reilly, D., Wren, C., & Berry, T. (2010). Cloud computing: Forensic challenges for law enforcement. In *International Conference for Internet Technology and Secured Transactions* (pp. 1–7).

Reilly, D., Wren, C., & Berry, T. (2011). Cloud computing : Pros and cons for computer forensic investigations. *International Journal Multimedia and Image Processing, 1*(1), 26–34.

Ruan, K., Carthy, J., Kechadi, T., & Baggili, I. (2013). Cloud forensics definitions and critical criteria for cloud forensic capability: An overview of survey results. *Digital Investigation, 10*(1), 34–43. doi:10.1016/j.diin.2013.02.004

Ruan, K., Carthy, J., Kechadi, T., & Crosbie, M. (2011). Chapter 3 Cloud forensics. In *Advances in digital forensics VII* (pp. 35–46). Berlin Heidelberg: Springer.

Taylor, M., Haggerty, J., Gresty, D., & Hegarty, R. (2010). Digital evidence in cloud computing systems. *Computer Law & Security Review, 26*(3), 304–308. doi:10.1016/j.clsr.2010.03.002

Taylor, M., Haggerty, J., Gresty, D., & Lamb, D. (2011). Forensic investigation of cloud computing systems. *Network Security, 2011*(3), 4–10. doi:10.1016/S1353-4858(11)70024-1

Trenwith, P. M., & Venter, H. S. (2013). Digital forensic readiness in the cloud. *2013 Information Security for South Africa*, 1–5. doi:10.1109/ISSA.2013.6641055

Watson, G. J., Safavi-Naini, R., Alimomeni, M., Locasto, M. E., & Narayan, S. (2012). LoSt : Location Based Storage. In *Proceedings of the 2012 ACM Workshop on Cloud computing security workshop* (pp. 59–69).

Wolthusen, S. D. (2009). Overcast: forensic discovery in cloud environments. *2009 Fifth International Conference on IT Security Incident Management and IT Forensics*, 3–9. doi:10.1109/IMF.2009.21

Zargari, S. A., & Smith, A. (2013). Policing as a service in the cloud. *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, 589–596. doi:10.1109/EIDWT.2013.106

Zargari, S., & Benford, D. (2012). Cloud forensics: Concepts, issues, and challenges. *2012 Third International Conference on Emerging Intelligent Data and Web Technologies,* 236–243. doi:10.1109/EIDWT.2012.44

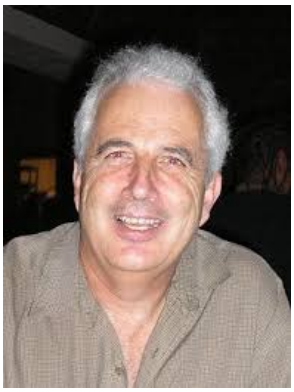
Zawoad, S., & Hasan, R. (2012). I have the proof: Providing proofs of past data possession in cloud forensics. *2012 International Conference on Cyber Security, (SocialInformatics*), 75–82. doi:10.1109/CyberSecurity.2012.17

Zawoad, S., Dutta, A. K., & Hasan, R. (2013). SecLaaS : Secure logging-as-a-service for cloud forensics. In *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications security* (pp. 219–230). ACM.

Biographies

Truia Meyer has a BSc Computer Science degree from the University of Pretoria and BCom Honours in Information Systems from the University of Cape Town.

She has been working in Information Systems sector for 27 years, spending the past 23 years in the fruit export industry. She has been involved in programming, system analysis, project management, business analysis and corporate governance.



Adrie Stander is a senior lecturer in the Department of Information Systems at the University of Cape Town, South Africa, where he is heading the Postgraduate Program in Digital Forensics. He has more than 35 years' experience in the IT industry and has conducted numerous Digital Forensic investigations.