A forensically-enabled IAAS cloud computing architecture

Saad Alqahtany
Plymouth University, saad.alqahtany@plymouth.ac.uk

Nathan Clarke
Edith Cowan University

Steven Furnell
Edith Cowan University

Christoph Reich
Hochschule Furtwangen University

DOI: 10.4225/75/57b3e3a5fb87e

Originally published in the Proceedings of the 12th Australian Digital Forensics Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia.

This Conference Proceeding is posted at Research Online.
A FORENSICALLY-ENABLED IAAS CLOUD COMPUTING ARCHITECTURE

Saad Alqahtany¹, Nathan Clarke¹,², Steven Furnell¹,² & Christoph Reich³
¹Centre for Security, Communications and Network Research, Plymouth University, Plymouth, UK
²Security Research Institute, Edith Cowan University, Perth, Australia
³Information and Media Centre, Hochschule Furtwangen University, Furtwangen, Germany
Saad.alqahtany@plymouth.ac.uk

Abstract

Current cloud architectures do not support digital forensic investigators, nor comply with today’s digital forensics procedures largely due to the dynamic nature of the cloud. Whilst much research has focused upon identifying the problems that are introduced with a cloud-based system, to date there is a significant lack of research on adapting current digital forensic tools and techniques to a cloud environment. Data acquisition is the first and most important process within digital forensics – to ensure data integrity and admissibility. However, access to data and the control of resources in the cloud is still very much provider-dependent and complicated by the very nature of the multi-tenanted operating environment. Thus, investigators have no option but to rely on cloud providers to acquire evidence, assuming they would be willing or are required to by law. Furthermore, the evidence collected by the Cloud Service Providers (CSPs) is still questionable as there is no way to verify the validity of this evidence and whether evidence has already been lost. This paper proposes a forensic acquisition and analysis model that fundamentally shifts responsibility of the data back to the data owner rather than relying upon a third party. In this manner, organisations are free to undertake investigations at will requiring no intervention or cooperation from the cloud provider. The model aims to provide a richer and complete set of admissible evidence than what current CSPs are able to provide.

Keywords
Cloud Computing, Digital Forensics, Cloud Forensics, Cloud Provider, Cloud Customer, IaaS.

INTRODUCTION

Cloud computing is changing the way technology is accessed and the way business is conducted. Cloud computing provides a highly scalable infrastructure and pay as you go services at low cost and on demand computing (Zargari & Benford, 2012). Due to these promising characteristics and financial benefits, cloud computing has become increasingly attractive for both commercial and public entities (Lu et al, 2010). Although significant opportunities are clearly offered by cloud computing for organisations, security issues are ranked as the single greatest challenge of cloud computing (Kuyoro et al, 2011). Issues surrounding the security and subsequent incident management requirements in the cloud are frequently ignored leading to devastating consequences (Josshua, 2012). For example, the Epsilon email system was hacked in 2011. It hosted more than 40 billion e-mails annually for more than 2,000 global brands including three of top ten US banks and financial institutions. As a result of this security breach, all of its customers were also affected (Sherman, 2011).

While security has frequently been an “after-thought” in new technology such as the cloud, digital forensics has historically been an “after-after-thought” (Ruan & Carthy, 2012). When the evidence resides in the cloud, new challenges exist on how to apply current digital forensic procedures. These challenges are novel and unique to the cloud and not encountered in traditional digital systems. This is due to the unique combination of characteristics that cloud computing introduce, including: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service (Mell & Grance, 2011). However, cloud-computing services are as easy and convenient for customers as they are a target for intruder. Intruders know that a single breach can lead to a treasure trove of material and they are skilful to breach various cloud services. For example, the breach of Apple’s iCloud service, copying risqué photos of many celebrities and leaking them online (Steinberg, 2014). Another example is a massive cloud computing breach which occurred in 2011 which affected a number of the largest entertainment companies and cloud service providers. It was the second-largest online data breach in the history of U.S. and it cost the intruders three pennies an hour for renting the Amazon servers that launch the professional and highly sophisticated incursion (Galante et al, 2011).

Conducting a digital investigation without forensics by design is an expensive process. Fannie Mae, The Federal National Mortgage Association, spent approximately 9% of its total annual budget of six million dollars on discovery of electronically stored information for litigation purposes and still failed to meet discovery deadlines.
According to a survey on cloud and eDiscovery disseminated to organisations that are using cloud-based solutions, 26% responded that they do not have an eDiscovery plan in place and 58% responded that they do not even know if a plan exists (Murphy, 2011). This means that in case of litigation and investigation, such organisations will be left scrambling in a reactive manner to collect information from the cloud; leading to greater cost (Ruan, 2013). Nevertheless, acquisition of the evidence in cloud investigations is highly dependent on the cloud service provider. A cloud consumer will only be able to specify the location of their data at a higher level of abstraction. For example, the consumers of IaaS do not manage the underlying cloud physical infrastructure; they have control over the operating systems and limited control over select network components (Loeffler, 2011). The consumer will be in the dark as to where their data is physically located as Cloud service providers (CSP) usually hide the actual physical location for the purpose of data movement and replication (Cruz, 2012). Arguably, it is more cost effective to implement dedicated forensic capabilities as a part of cloud infrastructure (Ruan et al, 2013). However, to do so requires the support of the Cloud Service Provider and the investment in the underpinning architecture. Despite the demand from organisations, to date, no such systems exist to support the routine forensic acquisition and analysis of cloud systems.

This paper proposes a fundamentally different approach to forensic acquisition and analysis within an IaaS service model, giving control for the forensic analysis to the cloud user (who actually owns the data) rather than the cloud provider. Indeed, this research goes one step further in that it is designed to be completely independent of the cloud service provider – requiring no intervention. This reduces the complexity of the acquisition process, the requirement for CSPs active involvement and any modification to the CSP underlying architecture.

The structure of the paper is as follows. Section 2 describes the background literature presenting an overview of data acquisition in the cloud and its current state of the art. The proposed model is then described; following by a brief discussion of it, prior to the conclusion and future work.

DATA ACQUISITION IN THE CLOUD

Data acquisition is regarded as one of the major and key issues when investigating cloud-based incidents (Dykstra & Sherman, 2011). Digital evidence can easily become inadmissible due to a minor deviation or an unexplained operation/procedure. The problem is not merely about acquiring data to perform a forensic investigation but the evidence needs to be collected in a manner to ensure its admissibility in a court of law. As data resides in a cloud provider’s storage, so does the evidence. Thus, when the consumer or law enforcement agency requires evidence, support from CSPs is currently essential in order to obtain access and acquire the evidence – no matter in what state that information might be in. However, the mandate of such support is not defined despite strong support from the user and security community (Ruan & Carthy, 2012). There are, however, forensic related responsibilities which are solely expected from the provider including data ownership, data retention and disposal, facility security, clock synchronisation, audit log and intrusion detection. On the other hand, there are forensic related responsibilities that shared between provider and consumer including audit planning, independent audit, data integrity and segmentation (Ruan & Carthy, 2012). Additionally, the segregation of duties and interfacing capabilities between provider and consumer need to be defined and developed for external investigation when law enforcement is involved (Ruan & Carthy, 2012).

In fact, data retention and disposal is one of the responsibilities that is expected from the CSPs including backup and redundancy mechanisms for data retention and storage. Due to the sheer volume of data, performance requirements and required space of storage, the provider might physically destroy or overwrite storage. If this was the case, the evidence involved in such storage is not likely to be recoverable (Ruan & Carthy, 2012). In spite of the fact that CSPs have been advertising that they have the ability and responsibility to store, acquire and handle cloud-based evidence for the consumer or law enforcement, with its cloud services including 99.9 % uptime, the reality is completely different and their promises seem to have been lacking. For example Amazon’s cloud crash that occurred in 2011 resulted in some of the customer’s data being deleted and Amazon was unable to recover it (Blodget, 2011). However, technically there are many reasons that hamper a CSP from providing the consumer with the desired evidence in a forensically sound manner, in a timely fashion. These include but are not limited to:

- There is a limitation in the volume of backups that the CSP will retain, due to the sheer volume of data and users. This can lead to the sheer impossibility of recovering deleted data or even overwritten data that is deleted by another user.
- CSPs usually hide the data location from customers for data movement and for replication reasons. However, there is decreased access and control at all levels for all consumers. Generally, the consumer has little or no control and has no knowledge where its data physically is located (Dykstra & Sherman,
This effectively removes the opportunity to perform a physical acquisition of the disk – as would be standard practice in computer forensic investigations.

- Resources that are available to cloud consumer are abstracted. In terms of digital investigation there is vital information that the investigator has to understand in order to accurately understand the environment. This information including the cloud architecture, hardware, hypervisor and file system and it is not available to the cloud consumer yet in our today’s cloud architecture (Dykstra & Sherman, 2011).

- In case of an incident occurring, the cloud provider will focus upon restoring the service rather than preserving the evidence and handling it in a forensically sound manner. Furthermore, due to reputational damages, some CSPs may not report the incidence nor cooperate in an investigation.

- Finally, the location uncertainty of the data makes the response time to an e-discovery request extremely challenging (Ruan et al., 2013). In addition, evidence residing in a single CSPs, leading to a potential single point of failure and adversely impacting on acquisition of useful data (Crosbie, 2013).

Fundamentally, the CSP architecture is designed for operational considerations to provide the most effective use of resources in the most economical fashion. As such, they are not designed with forensic acquisition and analysis in mind.

Data collection refers to conducting a physical acquisition of forensic data. The black box nature of the cloud architecture can provide 100% availability of resources but hides the complexities of the underlying architecture. In traditional digital forensics acquisition, investigators have full control over the forensic artefacts including router logs, process logs and hard disks. Thus, investigators have the ability to seize equipment in a forensic manner and perform detailed analysis on the media and data recovered. Furthermore, investigators can take physical custody of the hard disk and conduct a bitwise copy thus maintaining of the integrity of the data. In the context of the cloud, the situation is completely different. Digital investigators are confronted with the cloud due to the distributed nature of the information technology systems. Thus, obtaining access to the computer is difficult and out of reach of digital investigators. The investigator has reduced visibility and control over the forensic artefacts. Currently, relying on the cloud service provider is unavoidable as the required cloud data resides in a virtual instance. Applying standard forensic acquisition procedures to a cloud system simply do not work. For example, the analysis of one virtual instance would require a system shut down and imaging of the physical disks involved. This in itself raises issues of identifying the relevant disks to image with a data centre of thousands, the impact this has on the operational systems of other cloud customers and the infringement upon data owned by other customers. The ability to ensure all data is recovered and integrity of the data is maintained is highly questionable. According to a survey conducted by Ruan et al. (2013) and distributed among 257 respondents that included digital forensic experts and practitioners on cloud forensics, 78% of the respondents strongly agreed that decreased access to and control over forensic data at all levels from the customer side is one of the top challenge for cloud forensics. The type of cloud service model (i.e. IaaS, PaaS and SaaS) does introduce further opportunities for access to data depending upon the access provided. Figure 1 shows customer’s control over different layers in different service model. It is clear that in an IaaS, the cloud consumer has the highest level of control whereas in a SaaS model, the cloud consumer has the least level of control. For example, in IaaS, the consumer can clone the virtual machine image which could be used for forensic analysis (in part) whereas in SaaS, the CSP has effectively complete control to the application log (Zawoad & Hasan, 2013).
In a cloud environment, the evidence could be the image of VM, files stored in cloud storage and logs provided by the cloud providers (Zawoad et al, 2013). Forensic acquisition of the digital artefacts is the essential step in the forensic process. The IaaS model does provide a richer degree of potential evidence than what PaaS and SaaS do. For example, the cloud customer of IaaS model has the ability to install and set up the image of VM for forensic purposes. However, IaaS VM model does not have any persistent storage. This is very risky. For example, in case of VM is rebooted or powered down, all volatile data will be lost (Birk & Wegener, 2011). Furthermore, there is a significant limitation on the backups that the cloud provider will retain (NITS, 2014). The Hypervisor is responsible for enforcing hardware boundaries and routing hardware requests among different VMs. However, it is under the control of CSP. Thus, the customer VM is still under the control of CSP. To tackle such issues, some researchers have proposed their solutions. Birk and Wegener (2011), for example, proposed that the CSP can provide network, process and access logs to customers through a read only API. Furthermore, Dykstra and Sherman (2012) recommended a cloud management plane as it offers the balance between speed and control with trust in a Infrastructure as a service model. Recently, Dykstra and Sherman (2013) developed a set of tools known as Forensics OpenStack tools (FROST). It operates at the cloud management plane instead of interacting with the operating system inside the guest virtual machines. FROST is the first forensics capability to be built into any Infrastructure-as-a-service cloud model (Dykstra & Sherman, 2013). However, FROST is deployed by the CSP. Thus, trust in the CSP is required but not in the guest machine. Furthermore, trust on the cloud infrastructure is required including the hardware, host operating system, hypervisor and cloud employees. FROST also assumes that cloud customer is cooperative and involved in the investigation.

A MODEL FOR FORENSIC ACQUISITION & ANALYSIS IN THE CLOUD

It is evident from the literature that a solution to permit forensic analysis of systems within the cloud has become essential (NITS, 2014). Furthermore, it is clear CSPs have little motivation to provide assistance with incidents, unless forced to do so by law enforcement. Even when they do, their ability to access such data will very much depend upon the current status of the system (i.e. whether the VM is still up and running, whether a backup exists or whether the data has been overwritten). It is imperative organisations remain in control of their data and have the ability to undertake incident analysis/forensic examination of their systems when deemed necessary and in a timely fashion.

Whilst other research has proposed an IaaS solution, it fundamentally relies on the collection and storage of VM images and a key aspect of the approach is to include the CSP as a core part of the solution – largely to ensure cloud management information is also obtained (Dykstra & Sherman, 2013). The proposed approach in this research is to enable the cloud customer to have complete control over the forensic acquisition process, ignoring the data held by the CSP. This is made possible through the implementation of an Agent-based approach that sits on each of the customers VMs and communicates the necessary information to a central Cloud Forensic Acquisition and Analysis System (Cloud FAAS). This use of agents ensures all necessary cloud management data (e.g. VM start time/stop time) are logged. Lower level data, such as physical storage locations for the VM data, which is only accessible via a CSP is not required due to the agent-based acquisition approach. Indeed, through the agents, it is possible to recreate an image of the VM hard drive at any point in time and provide
access to every file in its entirety – going beyond what computer forensics is capable of achieving with partially overwritten files. This results in an increased level of visibility for the forensic investigator and removes any limitations that data carving and fragmentation may introduce.

As illustrated in Figure 2, the model is composed of two major components: the Agent Coordinator, and Cloud FAAS. The Agent Coordinator is responsible for the management of agents that are installed on the individual VM system. Different agents are responsible for various aspects of the acquisition and each can be enabled or disabled depending upon Acquisition Policy that is defined by the cloud customer (from here on in referred to as the organisation), which is housed within the Cloud FAAS. The following agents will be available:

- Non-Volatile Memory Agent – responsible for logically imaging the hard drive associated to the VM
- Volatile Memory Agent – responsible for logically imaging the live memory of the VM
- Network Traffic Agent – responsible for logging and storing network traffic (both egress and ingress)
- Activity Log Agent – acquiring system and application logs

Which agents are necessary will depend upon organisational requirements and the nature/responsibility of the VM. For example, in a 3-tier web application, it would only be necessary to operate the Network Log Agent on the web front end system, as the back-end server will be configured to only communicate with the web server, thus negating the need to replicate the network data store. Furthermore, the Activity Log Agent will only be required in situations where the other three agents are not in use; as such information can be derived from them. The Activity Log Agent will provide high-level log information to an investigator in environments where the overhead and cost of operating the other agents is not deemed necessary.

The Cloud FAAS is the central processing point for forensic data. It provides access to the management information, which defines the forensic acquisition policy for the VMs. The policy is an essential component of the approach that will directly impact its efficiency and financial cost. To ease usability, the policy is based upon a set of standard templates that are derived from server roles – with critical systems having an acquisition policy that monitors all changes across all agents. Less critical systems will incorporate a less granular acquisition approach. Should an organisation desire to, it is possible to modify the template to individualise the policy to directly reflect an organisational risk assessment.
The Agent Coordinator communicates with the Cloud FAAS via a Communication Engine. All communication is undertaken in a cryptographically secure manner – to ensure the confidentiality and integrity of the data in transit. The Agent Coordinator and Agent Manager also include the forensic hashing of all image data (at all levels of data object – complete images to files) to ensure chain of custody and data integrity is maintained throughout the acquisition phase. This information is subsequently stored in the Forensic Image Storage.

The Forensic Manager is responsible for managing the overall system and provides the interface to the forensic investigator. It enables an investigator to select the systems to be analysed and uniquely the timeframe required of interest. The Image Reconstruction module will then take the necessary information from the image repository and reconstruct the image(s). What is reconstructed and to what data granularity will depend upon what had been defined in its policy. Having reconstructed the image, the forensic data will be sent to the analysis component to undergo forensic examination and analysis. As indicated in Figure 2, this analysis will utilise industry de facto tools such as EnCase or FTK. It will also provide a correlation engine and visualisation component so that investigators can understand the relationship and data flows between systems – enabling a higher level of abstraction than individual system analysis would provide.

Acquisition & data handling

The ability to image VMs and transfer the data to a Cloud FAAS will have a huge implication for the underlying network capacity, processing overhead for the agents on the server VMs and on the Cloud FAAS infrastructure. With VM non-volatile storage in range of 100GBs and GBs of network activity, it could become too costly to store such data. Indeed, the introduction on a policy-based approach where different acquisition requirements are placed on different servers is an attempt to mitigate this information overload and reduce it to a manageable yet acceptable standard (forensically) for organisations. As illustrated in Figure 3, the data handling approach devised for this model comprises of two main steps. Initially a forensic image of the non-volatile memory (i.e. the hard drive as seen by the VM) is taken. Operating at this logical layer on a VM means it is not possible to map this data to a physical drive – indeed; the nature of the cloud infrastructure would not permit this. As such, step 2 seeks to operate similar to an incremental backup, recording all file system changes to the drive. To ensure the forensic value of such data, the data clusters of those file changes are also stored. This allows the Image Reconstruction engine to reproduce a forensic image of the drive at any point required – including showing how files have been deleted and overwritten. Importantly, however, as the system stores these files, an investigator would also be able to obtain full access to these deleted files (something that is not possible with normal computer forensic procedures). Rather than requiring a periodic image of the complete drive, the recording of file system changes reduces the volume of data that needs to be communicated and stored.

As shown in Figure 3, at some point in the future it will become necessary to re-image the drive, as the volume and complexity of the file system changes that have taken place since initial imaging is such that a re-image is essential – both for computation and storage requirements. Data retention is defined by the policy but it is envisaged that this will in the region of weeks or months rather than years due to the volume of storage required in the Cloud FAAS. This approach is not devised to replace organisational backup strategies but as a means of investigating incidents in a timely fashion. It is therefore not anticipated such investigations will occur beyond a relatively short (6 month) timeframe.
The frequency of reimaging, the granularity of file system changes, frequency of volatile memory captures and the resolution of the network traffic captures is all defined by the policy. Higher levels of resolution and frequency and lower granularity of data capture will all increase the demands placed upon the Agents and the Cloud FAAS – in particular the Forensic Image Storage. Obviously, anything but the most rigorous policy will have an impact on the forensic value of the resulting data. However, this is no worse than current computer forensics – where the time acquisition takes place can have a direct impact over the quality of the resulting forensic evidence. The transmission of data from the Agent Coordinator to the Cloud FAAS can also be optimised to take advantage of low network usage so to minimise any adverse effects upon the Cloud core operation.

Where a Cloud FAAS operates is largely dependent upon the organisation. They may choose to host it locally to the organisation so that ownership and access to data is as strong and reliable as possible. It could also be hosted within the same CSP as its operational servers – running as a cloud service in its own right. Whilst this is advantageous from a network bandwidth perspective – taking advantage of high bandwidth local area connections, it suffers from a single point of failure should a catastrophic incident occur to the CSP. A Cloud-based deployment more generally would certainly be advantageous from a data processing and forensic analysis perspective. Both of these aspects are computationally very intensive, yet unpredictable as to when they will be required. An elastic and flexible computing environment would allow for this – whether that is a public or private cloud.

CONCLUSION AND FUTURE WORK

It is clear that the cloud is here to stay and is growing with every passing minute. It is also a reality that there is a big concern with regards to data acquisition and its integrity in the cloud environment. There is a need to provide a solution that will ensure organisations remain in control, take the burden/liability off the CSPs and make it easy to acquire the evidence in a forensically sound manner and in a fashion time. The solution proposed in this paper can address the problem. However, further research is required in order to better understand both the technical implications resulting from such a system on the day to day operation of a cloud system and the financial costs.
REFERENCES


