An exploration in forensic evidence in a cloud computing environment

Jason E James, Indiana State University, jason.james@indstate.edu

Abstract

Cloud computing is the next step forward in Information Technology and has many advantages: resources are available on-demand, in real-time, and in some case will allow you to only pay for the resources that you actually use. There are also several disadvantages that must be considered: there are no standards for cloud deployments, cloud services can be proprietary, and the data is volatile, and the storage can be decentralized. As a result cloud computing has all of the strengths and vulnerabilities that are associated with the Internet. The best features of the cloud also represent the vulnerabilities and complications at the same time. Consequently, the cloud introduces significant challenges for forensic investigations; these include data ownership, multi-tenancy, collection logistics, legalities and jurisdictional complications, compliance with applicable regulations. This study will begin by conducting a review of the existing literature in cloud forensics and discuss the ramifications and impacts of conducting forensic investigations in a cloud-computing environment. Next the study will attempt to gather forensic evidence during an active penetration test that is being conducted on a cloud platform. The results will be gathered and reviewed to determine; what was found, the importance of the evidence, and the legal value that is retained from the gathered evidence.

Keywords: AWS, Cloud, Digital Forensics, NIST, Penetration Test, SLAs, SaaS

Introduction

Cloud computing, by its very nature has all of the strengths and vulnerabilities that are associated with the Internet and introduces significant challenges for forensic investigations; these include: data ownership, multi-tenancy, collection logistics, legalities and jurisdictional complications, compliance with applicable regulations (Almulla, Iraqi, & Jones, 2013). Cloud computing is a rapidly evolving information technology and is almost everywhere now. Digital forensics is a technology that has been traditionally used to collect, examine, analyze, and preserve the integrity of the data on physical hardware such as hard-disk and flash drives (P Mell & Grance, 2014). However, as the worlds of cloud computing and digital forensics collide, the collecting, examining, analyzing, and preserving of the integrity of the data has taken on a whole new challenge for investigators and examiners. This study attempts to expose the problem of performing digital forensics in a cloud implementation by attempting to gather forensic evidence of a penetration test into a Software as a Service (SaaS) environment.

Background

Cloud Computing

The first challenge in cloud forensics is the fact that there is some debate over exactly what the term cloud encompasses though there are many characteristics that have been identified. One study states that utilizing standardized solutions and centralized and shared use of resources define the cloud (Katilu, Franqueira, &
In an attempt to define cloud computing, another study noted that utilizing a resource pool that is accessed through the web and is a metered service is what makes it cloud computing (Freet, Agrawal, John, & Walker, 2015). The National Institute of Standards and Technology (NIST) points out some specific characteristics that are essential for a deployment to be classified as cloud computing: “On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured service” (Mell & Grance, 2011).

In addition to the generalities NIST also gives a formal definition: “Cloud computing is 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.” (Mell & Grance, 2011). To further the definition Mell & Grance define the following service models: “Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS)” and deployment models: “Private cloud, Community cloud, Public cloud, Hybrid cloud” (2011).

Within the above definitions and framework, cloud computing will continue to grow and evolve. Part of this evolution will be to “realize the concept of computing as a utility, just like water, gas, electricity and telephony” (Reilly, Wren, & Berry, 2010). This phenomenon was also noted by Dahbur et al when they state that cloud computing has the potential to revolutionize Information Technology (IT) services with the ability to utilize the services similar to utilities (2011). The advantages of cloud computing are very easy to discern resources are available on-demand, in real-time, and in some case will allow you to only pay for the resources that you actually use.

The best features of the cloud also represent the vulnerabilities and complications at the same time. Mell & Grance specifically illustrate this features noting that the rapid elasticity, multi-tenancy, load balancing and segregation of duties “introduce unique scenarios for digital investigations” (2014). The only protections that cloud customers have from the cloud providers are the Service Level Agreements (SLAs) that are negotiated and signed prior to establishing service. The problem with SLAs was discussed by a study, they found that these agreements concentrate on usability and availability while not dealing with security or forensic concerns (Farina, Scanlon, Le-Khac, & Kechadi, 2015). This thought was best stated by NIST, “The cloud exacerbates many technological, organizational, and legal challenges already faced by digital forensics examiners” (Mell & Grance, 2014).

Digital Forensics

Forensic investigation in the digital world is essentially the same as it is in the material world. The objective is to gather admissible evidence of a criminal act and the preserve and present that evidence in a trial. While traditional digital forensic techniques have been developed and successfully utilized these approaches may not work in the cloud. A new field of forensics will need to be developed to deal with forensic investigations into cloud implementations. To this end NIST has defined cloud computing forensic science as the ability to reconstruct events using proven methods and principles “through identification, collection, preservation, examination, interpretation and reporting of digital evidence” (Mell & Grance, 2014).

In digital forensics the main technique to gather evidence was to unplug the suspect device capture an image then analyse the image, this would prevent the data from becoming contaminated (Farina et al., 2015). This process will not be practical in cloud computing environments, as the data is not being stored on a stand-alone device. Biggs & Vidalis discuss the problem of digital forensics enduring the constant evolution of the landscape while attempting to stay on a level playing field with cybercriminals (2009). Another major hurdle is that certain measures will need to be put in place prior to an incident and this is generally not the
case (Reilly et al., 2010). To support this point, a different team found that “established digital forensic procedures will be invalidated in this new environment” (Grispos, Storer, & Glisson, 2013).

NIST notes that when looking at traditional digital forensics it is common practice for investigators to physically take possession of the devices in question, as a result of the distributed nature of the cloud the devices may be controlled by many separate actors (Mell & Grance, 2014). This separation of control exacerbates any attempts to retrieve relevant evidence in a timely manner. Law enforcement has been working steadily to improve the toolkit that they have at their disposal. Reilly et al found that while computer forensics has been steadily developing tools procedures and practices over the last decade in order to keep up with the evolution of technology “it now faces possibly its greatest challenges in dealing with cloud computing” (2010).

Balancing benefits against challenges was a theme identified by Morioka & Sharbaf when they found that “the innovative nature of cloud computing has created unique challenges in the field of digital forensics” (2016). In their work Reilly et al found that forensic readiness was not thoroughly considered when the cloud technologies were developed and deployed (2010). This was a common theme throughout the work that was reviewed, while customers recognize the advantages of cloud computing they do not think about potential future investigations.

The purpose of a business is to make a profit, being prepared for a forensic investigation does not directly add value to a service which means that there is little advantage to cloud providers to implement this readiness (Sibiya, Venter, & Fogwill, 2015). With no motivation to provide forensic readiness, cloud providers will continue to provide services that can be utilized for nefarious purposes. This threat was discussed by one team that found the problem of scale in the rise of cloud computing creates problems for digital forensics it will also open a “brand new front for cybercrime investigations with the associated challenges” (Ruan, Carthy, Kechadi, & Crosbie, 2011). Cyber criminals can utilize the dynamic nature of the cloud to hide their tracks from investigators. Zargari & Benford find that in order to assist law enforcement with digital forensics, methodologies must be adapted for use in a cloud environment (2012).

While law enforcement and information technology security personnel recognize the need for implementing digital forensics readiness into cloud deployments, cloud providers may have a different point of view. In examining this point one team noted that accountability and privacy are separate ends of a scale and it is imperative to establish an equilibrium. The ability for any forensics measures being deployed to balance these factors and the needs of clients and providers must be developed (Lee, Awad, & Awad, 2015). Data crimes in the cloud will be difficult to track as “the existing forensic tools are not capable of data recovery and thus lot of research was being proposed at this level” (Faheem, Kechadi, & Le Khac, 2011).

**Scope and Methodology**

Discovery and acquisition of evidence in remote, elastic, provider-controlled cloud computing platforms differ from that in traditional digital forensics, and examiners lack appropriate tools for these tasks. While there are many important issues in this new field, this study will focus explicitly on data acquisition. Crimes that target or use cloud computing will undoubtedly emerge in this landscape, and investigators will rely on their existing expertise in tools like Guidance EnCase or Access Data Forensic Toolkit (FTK) unless alternative tools and techniques are provided. In this research study, the scope was limited to FTK and thus will be discussed inclusively (Dykstra, 2012).

Digital forensics for cloud computing bring new technical and legal challenges. Cloud computing makes forensics different, particularly given the remote nature of the evidence, lack of physical access, and trust
required in the integrity and authenticity. While the goals of the forensic examiner are the same as before, the non-conventional difficult problems include forensically sound acquisition of remote data, large data volumes, distributed and elastic data, chain of custody, and data ownership (Dykstra, 2012).

Seizure and acquisition of digital artifacts are the initial steps in the forensic process (Casey, 2004). Two possible scenarios exist: remote investigators could collect forensic evidence themselves from the source, or providers could deliver it. Each scenario requires a different degree of trust in the data returned. Further, each scenario uses different technical implementations to recover the data. For the purposes of this study, the first scenario will be utilized. Given the years of development, the general acceptance by the judicial system, and the collection of expertise in the field, FTK is ideally prepositioned for the cloud forensic challenge (SC Magazine, 2011).

One question that has remained until now, however, was an evaluation of the ability of such tools to acquire and analyze cloud-based evidence. Cloud computing is a broad, generic term with many meanings and definitions. It has infiltrated the vernacular and been bastardized in marketing and media. Cloud computing is an evolution and combination of decades of technology, resulting in a model of convenient, on-demand, elastic, location-independent computing resources.

Though some definitions of cloud computing include popular web-based services such as email and social networking, this study will limit the scope of this project to computing resources that are billed as utilities. More specifically, we use the Infrastructure-as-a-Service (IaaS) model (National Institute of Standards and Technology, 2011). In this model, the consumer has complete control over a guest operating system running in a virtual machine (VM). The provider retains control and responsibility for the hypervisor (HV) down to the physical hardware in the datacenter. Since the Platform-as-a-Service and Software-as-a-Service models are built on IaaS, beginning with IaaS provides a fundamental basis from which to build future work (Dykstra, 2012).

In this study, the researchers for the forensic investigation will have complete control of the cloud target system. The elastic nature of cloud computing makes it possible for a criminal to commit a crime and then immediately destroy the evidence, but that situation is not considered here. While some cases will involve the cloud as the instrument of the crime, others will involve the cloud-hosted service as the target of the crime. The latter is the scope of this research study (Dykstra, 2012).

In this research study, the researchers measure and evaluate the ability of Access Data FTK Imager Lite to remotely acquire forensic evidence from cloud computing and measure their effectiveness. FTK is widely used today, it is benefited from tool expertise in the field, are trusted by the courts, and have a remote acquisition feature that has been targeted at geographically dispersed corporate LANs. The goal then is to evaluate the ability and accuracy of these features to acquire forensic data from cloud computing environments from a remote connection from the host computer. Utilizing a live forensic acquisition tool, FTK Imager Lite the study will gather data from Access Data. The research will look to evaluate the success at gathering evidence and the reliability of the data (Deoyani, 2014).

Scope

The digital forensics investigation of cloud storage services is what practitioners collect and examine evidence from all computing equipment that are possessed by the user to connect a cloud account and include personal computers, laptops, tablets, and mobile phones. However, for purposes of this limited scope study, laptops, which are the commonly used devices, are analyzed. In MAC or Windows operating system, the forensics examiner acquires and preserves the evidences stored in main RAM space as well as
the hard disk evidence data (system files, history files, file timestamps and root folders) (Easwaramoorthy, et. al., 2016).

A forensic examiner gathers user data from a laptop computer and then analyze for traces of a user account or remote connection software present in the collected client data. If a user account is found, the forensic examiner validates whether any evidence regarding username and password exists. If user credentials and any further details exist for connecting to a user storage account are available then a search warrant would need to be issued to the cloud service provider to access and collect the data. Once a forensic examiner has a warrant, they can then login, access, examine and acquire user details stored in the account. If the forensic examiner cannot acquire a warrant, it is only feasible to examine artifacts that are available on the laptop computer (Deoyani, 2014).

The scope of this research is about the evidence left after an Amazon EC2 account is used from a MacBook Pro OS Sierra laptop, which is a Windows Server-2012 R2 64-Bit virtual machine, which uses Microsoft Remote Connection on the MacBook Pro OS Sierra to logon. Amazon’s Elastic Compute Cloud (EC2) service is one of the most popular third-party cloud services, which enables users to flexibly rent computational resources for use by their applications. EC2 provides the ability to run Linux or Windows as guest operating systems within a virtual machine (VM) provided by a version of the Xen hypervisor (Barham et al, 2003). The hypervisor plays the role of a virtual machine monitor and provides isolation between VMs, intermediating access to physical memory and devices.

When first registering with EC2, each user creates an account uniquely specified by its contact e-mail address—and provides credit card information for billing compute and I/O charges. In addition, any user can setup a free tier AWS EC2 accounts for the first year. The new account user will then be asked to enter a valid telephone and a phone will be made to the user’s number and a 4 digit Pin number generated on screen will need entered when the call is received.

Once identity is verified, the user then has to select either a basic, developer, business, or enterprise support plan. For the purposes of this research a free basic account was setup. Once a valid account is setup (which usually takes up to 24 hours), the user can creates one or more VM images and can run one or more copies of these images on Amazon’s network of machines.

One such running image is called an instance, and when the instance is launched, it is assigned to a single physical machine within the EC2 network for its lifetime. As mentioned earlier the scope of this research involved creating a Windows Server-2012 R2 64-Bit virtual machine instance.

In addition, there are three degrees of freedom in specifying the physical infrastructure upon which instances should run. At the time of this writing, Amazon provides four “regions” and along with multiple availability zones within each one, United States, Asia Pacific, Europe, and South America. Each region contains “availability zones” which are meant to specify infrastructures with distinct and independent failure modes. We will limit our subsequent discussion to the Windows instance.

When requesting launch of an instance, a user specifies the region and may choose a specific availability zone, otherwise one is assigned on the user’s behalf. In the case of this research study, an availability zone was assigned (US East N. Virginia region us-east-1a availability zone for the instance created) (Note that we focus on the United States region—in the rest of the paper EC2 implicitly means this region of EC2).

As well, the user can specify an “instance type”, indicating a particular combination of computational power, memory and persistent storage space available to the virtual machine. There are many types of
Windows instance types (Microsoft Windows Server 2012 R2 Base), but the one chosen, t2.micro, was the only option for the free tier and thus was the 64-bit instance used in the study.

Launching an instance, the user specifies the instance type along with a compatible virtual machine image. Once you select the instance type and before you can finalize and launch and instance, you must create a public/private key pair, which will be used later to access the windows server from a remote desktop connection.

With these constraints chosen, virtual machines are placed on available physical servers shared among multiple instances. Each instance is given Internet connectivity via both an external IPv4 address and domain name and an internal private address and domain name. For example, the instance created for the study was assigned an external IP 54.208.251.56, external name ec2-54-208-251-56.compute-1.amazonaws.com, internal IP 172.31.21.80, and internal name ip-172-31-21-80.ec2.internal. Within the cloud, both domain names resolve to the internal IP address; outside the cloud the external name is mapped to the external IP address.

**Methodology**

Amazon EC2 instances created from most Windows Amazon Machine Images (AMIs) enable you to connect using Remote Desktop Protocol (RDP), which enables you to connect to and use your instance in the same way you use a computer sitting in front of you. For the research experiment, we installed an RDP client and since we were using Mac OS X Sierra, we downloaded the Microsoft Remote Desktop app from the Apple App Store as preferred for connecting to a Windows 2012 R2 instance.

In order to perform a forensics analysis, the researchers are using Forensic Toolkit (FTK®) Imager Lite 3.1.1. This is a data preview and imaging tool used to acquire data (evidence) in a forensically sound manner by creating copies of data without making changes to the original evidence. Normally, after you create an image of the data, Forensic Toolkit® (FTK®) is used to perform a thorough forensic examination and create a report of your findings. However, within the scope of our experiment, we used only FTK Imager Lite (AccessData, n.d.).

**Analysis**

For the research experiment, we used a non-cloud based standalone control machine to evaluate the success of the test. The control was a 2012 MacBook Pro laptop with OSx Sierra, a single 750GB disk drive and 8GB RAM. The study will use a VM Ware Fusion 8 installed on the MacBook Pro to run FTK Imager Lite for the analysis. VM Ware Fusion 8 allows the user to run Windows only applications side by side with Mac applications and can easily switch between the MacBook Pro and PC. A copy of FTK Imager Lite was downloaded and copied ran to the PC side and then it was imaged onto an attached external flash drive that contained one folder named “Documents”. Next the machine was connected to the Internet and several jpeg files were downloaded with identifying names and content and then deleted some files on the external hard drive in the lone folder. It was assumed that a criminal and a forensic investigation had commenced and imaged the external hard drive lone folder with FTK Imager Lite.

The research experiment tested the ability of popular forensic tool FTK Imager Lite to collect forensic data remotely in the cloud at the guest OS. Success or failure would be measured by (a) if the tool was able to collect evidence remotely, and (b) how accurately the data compared to those from a standalone control machine.
A single Internet connected device was prepared, the forensic examiner workstation with MacBook Sierra OSx. FTK Imager Lite was installed on a USB drive and the forensic examiner’s workstation USB drive was re-directed so it could be access in the virtual machine over a Remote Desktop Protocol (RDP). In Amazon EC2, we provisioned (as described earlier) a new virtual machine to simulate the target of an investigation which was an Amazon-provided Windows 2012 R2 64-bit image with a single 32GB disk drive and 1.7GB RAM. We configured the Amazon firewall to allow only RDP.

A connection was established to the target machine using RDP (the assumption that is made here is that the computer that is confiscated is imaged and the login username and password is stored automatically within the RDP as most users would not enter them every time) and proceeded to exercise normal behavior of a user adding jpeg files to a virtual machine. We downloaded several jpeg files with identifying names and content to the “Documents” folder and deleted some files. We assumed that a criminal and forensic investigation had commenced and imaged the virtual machine Documents folder with FTK Imager Lite.

We tested using FTK Imager Lite version 3.1.1 and the product was copied over the Remote Desktop connection from the examiner’s workstation and run interactively. FTK Imager Lite does not require installation, and runs self-sufficiently once uncompressed. For this experiment the examiner ran FTK Imager Lite to acquire images of the documents folder.

FTK Imager Lite version 3.1.1 was installed according to the manufacturer’s instructions on the PC side of the MacBook Pro through VM Ware Fusion 8. The image from the virtual machine was then mounted to the PC by way of FTK Imager Lite and compared to the standalone machine image.

**Results**

The use of FTK Imager Lite via USB flash drive copied over the Remote Desktop connection from the examiner’s workstation and run interactively on the virtual machine was a success and we were able to acquire an image of the “Documents” folder remotely save them on the attached USB flash drive. Analyzing these images in FTK Imager Lite compared to the exact same standalone image correctly revealed a timeline of activity, including the downloading of files and that were created. The analysis revealed no unusual artifacts of the virtual environment, nor any apparent anomalies to raise doubt about the integrity of the data.

**Discussion**

A benefit of the examiner using FTK Imager Lite is that no changes to the cloud infrastructure are necessary nor is assistance from the provider required. However, the forensic limitation is not being able to validate the disk images. Usually, in a forensic examination, cryptographic hashes are used to validate that the image of the device under investigation is identical to the original. However, since no hash is available for the original data source, legally, the evidence might not be accepted. In our experiment, we were not able to verify cryptographically that our cloud images were identical to the standalone control because of differences such as different hardware, operating systems, and network configurations but we were able to reconstruct the evidence from a crime and these differences matter.

**Conclusion**

We have shown that one of today’s most widely used forensic tools, FTK Imager Lite, is capable of remote acquisition of data in Amazon EC2. However, it still might not be feasible to generate reliable data and unravel the cloud forensic acquisition issue just yet since legally, it might not stand up in court. One
alternative is the EC2 management dashboard for AWS instances. It might be possible to overcome the layer of trust and still have control over the acquisition even though FTK Imager Lite was successful at providing evidence; the trust aspect is still a big concern. The EC2 management dashboard can be used to overcome this trust for forensic examiners. Windows servers (instances) are managed and controlled through this EC2 management dashboard in which virtual assets are interfaces with the cloud infrastructure.

In Amazon Web Services, this system is called the AWS Management Console. This web-facing system interfaces with the provider’s underlying file system and hypervisor, and is used to provision, start and stop virtual machines. The management plane is particularly attractive because it is user driven. The provider, end users, and law enforcement could download log files, disk images, and packet captures from the AWS Management Console on demand and if a computer crime is committed, chances are the logons to the AWS Management Console already exist on the suspect's captured computer.

While attractive, this solution does require trust in the management plane, a potential vulnerability that does not exist with non-virtualized, physical computers. As a web-facing interface, the management plane opens a new attack surface, which must be protected by the provider. Access to the management plane should be logged and strictly enforced with identity and access management. Future work in this area should focus on the AWS Management Console.

Cloud computing is not going away and is not only the future of technology, but also crime. Our work continues the path of cloud-based forensics and lays the groundwork for future forensic examiners to take further steps in establishing trust in acquiring forensics cloud-based data that could be used for crimes (Dykstra, 2012). This study attempted to expose the problem of performing digital forensics in a cloud implementation by attempting to gather forensic evidence of a penetration test into a Software as a Service (SaaS) environment. As the worlds of cloud computing and digital forensics collide, the collecting, examining, analyzing, and preserving of the integrity of the data has taken on a whole new challenge for investigators and examiners. Multiple challenges to conducting digital forensic investigation in a cloud environment have been identified and explored. As Birk & Wegener found, “evidence acquisition and analysis have to evolve along with the technology… there are no guidelines specific to evidence gathered in the cloud” (2011). Future work must be conducted to attempt to mitigate these challenges, until then the cloud will continue to be an untraceable breeding ground for malicious activities.

References

AccessData. (n.d.) FTK Imager Lite 3.1.1 retrieved November 25, 2016 from AccessData.com:
http://marketing.accessdata.com/ftkimagerlite3.1.1


