Impact of Cloud Computing on Digital Forensic Investigations

Stephen O'shaughnessy, Anthony Keane

Chapter 20

IMPACT OF CLOUD COMPUTING ON DIGITAL FORENSIC INVESTIGATIONS

Stephen O'Shaughnessy and Anthony Keane

Abstract As cloud computing gains a firm foothold as an information technology (IT) business solution, an increasing number of enterprises are considering it as a possible migration route for their IT infrastructures and business operations. The centralization of data in the cloud has not gone unnoticed by criminal elements and, as such, data centers and cloud providers have become targets for attack. Traditional digital forensic methodologies are not well suited to cloud computing environments because of the use of remote storage and virtualization technologies. The task of imaging potential evidence is further complicated by evolving cloud environments and services such as infrastructure as a service (IaaS), software as a service (SaaS) and platform as a service (PaaS). The implementation of forensics as a service (FaaS) appears to be the only workable solution, but until standards are formulated and implemented by service providers, the only option will be to use traditional forensic tools and rely on service level agreements to facilitate the extraction of digital evidence on demand. This paper explores the effect that cloud computing has on traditional digital forensic investigations and proposes some approaches to help improve cloud forensic investigations.

Keywords: Cloud computing, cloud forensics, digital forensic investigations

1. Introduction

Cloud computing is a rapidly evolving technological solution and business model as evidenced by the upsurge in the global adoption of cloud services. While cloud computing has its origins in mainframe computing and shares similarities with traditional Internet hosting, the ways in which cloud services are offered differ considerably. Cloud consumers can avail of self-provisioning, auto scaling and pay-per-use through ser-

vices that offer increased availability, performance and scalability. In this regard, cloud computing is an evolutionary step in the provisioning of services on the Internet, allowing organizations to easily outsource their information technology requirements and pay only for the services they use. Cloud service providers such as Google, Amazon and Microsoft are driving expansion by turning their excess capacity into a business pay-per-use model that offers scalable information technology resources. This expansion is also facilitated by the availability of high-speed broadband connectivity and low-cost access from service providers.

The vast supply of anonymous computing resources in the cloud potentially provides a breeding ground for computer crime. Garfinkel [6] notes that sensitive information such as credit card data and social security numbers stored in the cloud render it an attractive target for thieves. Furthermore, cloud computing resources such as easy-to-use encryption technology and anonymous communication channels reduce the likelihood that the nefarious activities undertaken by criminal elements are intelligible to law enforcement.

The cloud can also be used as an instrument to perpetrate denial-of-service attacks. With the help of the homemade Thunder Clap program costing just six dollars, Bryan and Anderson [4] leveraged ten virtual servers in Amazon's Elastic Compute Cloud (EC2) system to launch the denial-of-service attacks that succeeded in taking their client company off the Internet. The experiment was not detected by Amazon and no mitigation efforts were initiated against the attacks. The software that controlled the attacks was executed via a command placed on a social network.

Due to the lack of cloud-specific methodologies and tools, traditional digital forensic methodologies and tools are being adapted for use in cloud environments. However, existing digital forensic approaches typically assume that the storage media under investigation are completely under the control of investigators. Thus, these approaches do not map well to cloud computing environments. Cloud computing changes the traditional characteristics of how data – and potential evidence – are stored and retrieved. In the cloud, evidence can reside in different geographical locations on servers that are shared by multiple customers and that are under the control of different cloud service providers. This significantly impacts the identification and acquisition of evidence as well as chain of custody. The fundamental task is to ensure the integrity of evidence retrieved from the cloud so that it may be used in legal proceedings.

2. Traditional Forensics vs. Cloud Forensics

Digital forensics is defined as the use of scientifically-derived and proven methods for the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations [13]. 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 [12]. Thus, cloud forensics can be defined as the use of proven methods for the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence from distributed computing systems in a manner that maintains the integrity of the evidence so that it is admissible in a court of law.

A digital forensic process model provides a framework for conducting sound forensic investigations. While there is no digital forensic process model that is suited to all digital forensic investigations, a generic process model can be applied to many different types of digital forensic investigations regardless of the technology that is used. This section clarifies the differences between traditional and cloud forensic investigations by engaging the generic Integrated Digital Investigation Process Model proposed by Carrier and Spafford [5]. The model incorporates five phases: (i) preservation; (ii) survey; (iii) search and collection; (iv) reconstruction; and (v) presentation.

2.1 Preservation Phase

The preservation phase of a traditional digital forensic investigation involves securing the digital crime scene and preserving digital evidence. This includes isolating the computer system from the network, collecting volatile data that could be lost when the system is turned off, and identifying any suspicious processes that are running on the system. Suspect users that are logged into the system should be noted and possibly investigated. Log files often contain valuable evidence and should be secured if there is a possibility that they could be lost before the system is copied.

In the case of a cloud forensic investigation, direct physical preservation is limited to the suspect's machine, if this is available. Any other direct preservation is not possible because the data is stored remotely in

virtual images. An investigator can attempt to preserve data resident in the cloud by serving a legal order to a cloud service provider. However, the investigator must trust the service provider to acquire and preserve the data in a forensically-sound manner using proven digital forensic methods.

2.2 Survey Phase

The goal of the survey phase is to identify the obvious pieces of evidence and to develop an initial theory about the incident. Fragile pieces of evidence such as volatile memory are documented and collected immediately to prevent any possible damage or corruption. Carrier and Spafford [5] discuss a server intrusion case in which an investigator looks for obvious signs of a rootkit installation, analyzes application logs and searches for new configuration files. In a cloud environment, the computer system at the scene can be examined for evidence, but the investigator may not have access to external data because the physical examination of remote servers may not possible.

The level to which an investigator can identify potential evidence in a cloud environment is influenced by the specific cloud service model in use – software as a service (SaaS), infrastructure as a service (IaaS), or platform as a service (PaaS). In the SaaS model, the client retains no control over the underlying infrastructure such as the operating system, applications and servers, with the possible exception of limited user-specific application configuration settings. In this case, the investigator has no easy way of identifying evidence on the server side and has to rely on application logs and system logs obtained from the service provider; this is only possible if the service provider has some form of logging mechanism installed and makes the logs available.

The IaaS model offers the most in terms of evidence available to an investigator. In an IaaS environment, the customer controls the set-up of the virtual instances, as well as the underlying operating system and applications. Therefore, the potential exists for customers to install logging applications to keep track of user activity, which could greatly enhance the quality of forensic investigations – but this is not the norm. Nevertheless, an investigator can access more potential evidence than either of the other two cloud service models, SaaS and PaaS.

In the PaaS model, the customer can develop and deploy applications created using programming languages, libraries, services and tools supported by the provider. The customer does not manage or control the underlying cloud infrastructure, including the network, servers, operating systems and storage, but has control over the deployed applications

and possibly the configuration settings for the application-hosting environment [12]. This significantly hinders the ability of an investigator to identify possible evidence, since this is limited to specific application-level logs, if they exist.

Note that documentation is not considered to be a separate phase in a digital forensic investigation because digital evidence is documented as it is found. The final incident report is created during the presentation phase. Digital evidence must be documented thoroughly. For example, a file is documented using its full path, the file system clusters used by the file and the sectors of the disk on which the file resides; the hash value of the file is also computed to ensure that its integrity can be verified. Chain of custody forms must be created if the evidence could be used in court.

Documentation and chain of custody are difficult tasks in a cloud environment. As mentioned above, the level of evidence available to an investigator can vary, which directly affects how well the evidence can be documented. For example, an investigator who has direct access to a virtual image can document the files found on the image. On the other hand, if the investigator relies on the cloud service provider to recover the files of interest, then the investigator has to trust the service provider to retrieve the evidence in a forensically-sound manner.

2.3 Search and Collection Phase

The search and collection phase involves a thorough analysis of the system for digital evidence. This phase uses the results of the survey phase to determine the types of analysis to be performed. For example, a keyword search can be performed during this phase using the keywords identified from other evidence, or a low-level timeline of file activity can be analyzed to trace user activities.

The search and collection phase consumes the bulk of the time spent in an investigation. Artifacts that are of evidentiary value are collected, usually from some type of digital storage device. The collection method involves taking forensic images of the storage devices so that they can be examined under laboratory conditions. Other collection methods are used to retrieve information stored in volatile memory and live registries. The majority of the search and collection phase in a traditional forensic investigation is conducted at the local level, except, for example, the recovery of network logs that typically reside on a server.

The distributed infrastructure of a cloud environment poses challenges with regard to search and collection. The dispersed nature of data in the cloud means that the forensic investigator has to adapt traditional

methods to the new environment. The investigator must understand how data is stored in the cloud environment and determine how it can be retrieved while maintaining its integrity.

At the local level, evidence can be gathered from the client web browser history; this is because communications between the client and cloud service provider typically use an Internet browser. Other evidence, such as client login credentials for cloud services and instant messaging, must also be extracted and deciphered; this can give the investigator access to previous communications conducted by the client over the Internet. At the network level, it is generally not possible to analyze the traffic because service providers may not provide log data from the network components used by the customer's instances and applications.

If the IaaS service model is used, then it is possible for the investigator to take snapshots of the virtual machine and analyze them in a laboratory as in the case of images taken from a local system. The situation is more complex in the case of PaaS because only application-specific data is available. In the case of SaaS, the investigator can only retrieve limited data such as user-specific application configuration settings. The investigator has to provide a court order that would require the provider to execute the search, collect the data and return it to the investigator. The investigator must assume that the service provider employs trustworthy procedures and tools to execute the search, and reassemble and report the data. A violation of the chain of custody can cause the retrieved data to be inadmissible as evidence in court.

2.4 Reconstruction Phase

The reconstruction phase involves organizing the analysis results from the collected physical and digital evidence to develop a theory for the incident. Data that requires advanced analysis techniques, such as executable file analysis or decryption, is processed and the results are used in this phase. Scientific methods are applied to the evidence to test the incident theory. In some cases, the search phase may be resumed again to obtain additional evidence.

In a cloud forensic investigation, the service provider controls the amount of data released to the investigator; the amount of data released affects incident reconstruction. In addition, the physical disparity of the data can make it difficult to put the data in the correct context and temporal order. This situation is exacerbated by the fact that data is held in different geographic regions and the associated computer clocks may not be synchronized. These problems can negatively impact the credibility of the evidence proffered in court.

2.5 Presentation Phase

The presentation phase is the final phase of a forensic investigation. During this phase, all the physical and digital evidence artifacts are documented and presented to court (or to management). Investigator reports, presentations, supporting documentation, declarations, depositions and testimony are considered in the presentation phase. The documentation supporting each phase in the investigation is of particular importance because it helps establish a verifiable chain of custody.

In a forensic investigation, evidentiary data must remain unchanged and the investigator must be competent and able to present the findings, explaining the relevance and implications of all the actions undertaken during the investigation. Furthermore, strict logs and records should be maintained for every step of the investigation. In a cloud environment, it is difficult, if not impossible, to maintain a strict record of an investigation, especially when evidence resides in multiple locations and is under the control of different entities.

2.6 Shortfalls

The discussion above reveals that certain shortfalls during the various phases of the forensic process model can impact a cloud forensic investigation. This could bring into question the validity of the evidence presented in court. The shortfalls are:

- The inability to preserve a potential crime scene, which can adversely affect the integrity of the data artifacts that are collected.
- The unwillingness or inability on the part of the cloud service provider to provide data such as application logs and network logs.
- The limited access or lack of access to cloud data that can provide incomplete pictures of key events.
- The presence of fragmented data and artifacts whose metadata has been altered.

3. Other Issues

Certain other issues related to conducting digital forensic investigations in the cloud can affect the quality of the evidence retrieved. These, in turn, could affect the credibility and admissibility of the recovered artifacts in a court of law.

3.1 Multi-Tenancy

Multi-tenancy allows multiple clients to share a physical server and use services provided by common cloud computing hardware and software simultaneously. In some cases, multi-tenant infrastructures are a concern because the sharing of resources is extensive, occurs at a very large scale and involves multiple potentially vulnerable interfaces [2]. This resource-sharing environment poses challenges to investigators who have to concern themselves not only with the services used by a single customer, but also the non-customer specific components of a multi-tenant infrastructure and the resources shared with other customers. Shared resources include processors and memory. Cloud service providers are often unwilling to give an investigator access to shared memory because it may contain data belonging to other customers and the release of this data could violate confidentiality and privacy agreements.

3.2 Data Provenance

Data provenance records the ownership and process history of data objects and is, therefore, vital to a digital forensic investigation [11]. The provenance can provide information about who or what created the data object and modified its contents. The degree to which data provenance can be implemented in a cloud environment depends on the type of cloud model. In a SaaS implementation, the ancestry of a data artifact may be difficult to trace because the service provider would not normally give an investigator access to application and system log files. In the case of an account compromise, the customer does not have the ability to identify the data that was leaked or accessed by a malicious entity; this includes data modified or deleted by the malicious entity and data deleted by the service provider (e.g., for storage management reasons).

3.3 Multi-Jurisdictional Issues

Data stored in a cloud environment is often distributed over several locations to promote fault tolerance and efficiency of access. However, data distribution raises the issue of jurisdiction, which can present problems in legal proceedings. According to Garrie [7], a court can only hear a matter if it has jurisdiction over the parties and the subject matter of the action. Moreover, law enforcement agencies can only exercise their powers within their authorized jurisdictions.

The problems are exacerbated when data resides in another country. Confidentiality and privacy laws vary greatly from country to country. For example, some countries have strict laws related to the secrecy of

bank documents and the penalties for violating the laws can include criminal sanctions. In such cases, it may not be possible to retrieve all the evidence pertaining to an incident. Garrie [10] cites jurisdictional issues as a major challenge to conducting cloud forensic investigations.

3.4 Chain of Custody

The establishment of a chain of custody is vital to any forensic investigation [1]. It helps provide a documented history of the investigation, detailing "how the evidence was collected, analyzed and preserved in order to be presented as evidence in court" [16].

In a traditional forensic investigation, the chain of custody begins when an investigator preserves evidence at the scene and ends when the evidence is presented in court or to management. The distributed nature of a cloud computing environment significantly complicates the task of maintaining a proper chain of custody. Evidence must be collected from remote servers in a secure and validated manner in order for it to be presented as evidence. If an investigator is unable to gain direct access to cloud services and hardware, it is necessary to rely on the service provider to create forensic copies of evidence. Nevertheless, the investigator must ensure that the chain of custody is always maintained so that cloud data (including data collected by third parties) can be presented as evidence.

3.5 Service Level Agreements

A service level agreement is a contractual document between a cloud service provider and a cloud customer that defines the terms of use of cloud resources. Most current service level agreements do not incorporate provisions regarding forensic investigations and the recovery of evidence from cloud environments. Some provisions regarding the forensic retrieval of evidence from a cloud environment should be incorporated in the agreements. These include data access during forensic investigations and stipulations regarding investigations in multi-jurisdictional and multi-tenant environments, including legal regulations, confidentiality and privacy [15].

At this time, the terms of service are typically prescribed by the cloud provider and are generally non-negotiable [9]. The customer thus has little or no voice regarding the data that the cloud service provider may and may not disclose. Ultimately, the onus is on the customer to negotiate a suitable service level agreement with the provider that addresses evidence retrieval from cloud environments as well as thorny issues such as multiple jurisdictions, data ownership and the establishment of a chain of custody.

4. Cloud Forensic Solutions

Evidently, there are many unresolved issues pertaining to cloud forensic investigations, all of which are exacerbated by the dynamic, ever-changing nature of cloud environments. This section discusses some solutions that could enhance cloud forensic investigations.

4.1 Forensic Tool Testing

Currently, no cloud-specific forensic tool sets are available, requiring investigators to employ tools that were designed for traditional forensic investigations. Evaluation studies focused on the use of existing tools to acquire evidence from cloud environments would immediately benefit the forensics community as well as stimulate forensic tool refinement and new tool development for cloud environments. For example, tools are needed to perform live analyses of dynamic cloud environments. In many cases, live analysis offers the opportunity to gather valuable information from a running system, such as memory and registry data, but no comprehensive solution exists for cloud environments.

Another important gap concerns datasets for testing tools used in cloud forensic investigations. Yet another issue is the correlation of temporal data in cloud forensic investigations. The cloud customer and the service provider often reside in different time zones, which can produce contradicting metadata, such as the creation, modification and last accessed timestamps of an evidence artifact. Methods for automating the correlation of such data would be very beneficial as they would reduce, if not eliminate, the need to conduct intensive manual investigations.

4.2 Transparency of Cloud Services and Data

The lack of transparency regarding the internal infrastructure of a cloud environment poses challenges in an investigation. While information about the internal workings is valuable in an investigation, service providers may provide little information about the environment in which customer data is stored and processed. This lack of transparency is driven by the need to protect sensitive user data; also, releasing information about the internal infrastructure could expose a cloud service to attack [14]. Furthermore, cloud service providers are often unwilling to release information about their environments because it could be used by competitors, and any negative information released about cloud services or operations could harm the reputation of the service provider [3].

Haeberlen [8] proposes that cloud services should be made accountable to the customer and the provider in that both parties should be able to

check whether or not the cloud services are running as agreed upon by both parties. If a problem occurs, the parties should be able to determine which party is responsible and prove the existence of the problem to a third party such as an arbitrator or a judge. This proposal is beneficial to both parties: the customer can check whether or not the contracted services are actually being provided and the service provider can handle complaints and resolve disputes with more ease.

4.3 Service Level Agreements

Service level agreements must include clear and precise procedural information on how a forensic investigation would be handled by the investigator and by the cloud service provider in the event of a criminal incident. The roles should be clearly defined, and each party should be fully aware of its responsibilities, capabilities and limitations. Furthermore, service level agreements must address the legal implications of conducting an investigation in multi-tenant environments across multiple jurisdictions.

4.4 Forensics-as-a-Service

In a forensics-as-a-service (FaaS) model, the cloud service provider should be responsible for forensic data acquisition or, at the very least, provide support for forensic data acquisition. The service provider is in a position to preserve and collect the data because it controls the cloud infrastructure, not only the virtual machines, but also logging and packet capture mechanisms, and billing records. The service could be implemented by a cloud provider with little change to the existing cloud infrastructure, and it would provide customers with the assurance that high-quality forensic investigations could be conducted.

5. Conclusions

Several challenges exist when conducting forensic investigations in cloud environments. These challenges are posed by the highly dynamic, distributed, multi-jurisdictional and multi-tenant nature of cloud environments. Failure to address these challenges could affect the credibility and admissibility of the recovered digital evidence. Promising solutions include the development of cloud-ready forensic tools and service level agreements with built-in provisions for forensic investigations. However, the most complete solution would be to ensure that service providers implement forensics-as-a-service (FaaS) as a standard offering. This would enable high-quality forensic investigations to be conducted using traditional digital forensic tools under existing service level agreements.

References

[1] Association of Chief Police Officers, Good Practice Guide for Computer-Based Evidence, London, United Kingdom, 2012.

[2] L. Badger, R. Bohn, S. Chu, M. Hogan, F. Liu, V. Kaufmann, J. Mao, J. Messina, K. Mills, A. Sokol, J. Tong, F. Whiteside and D. Leaf, U.S. Government Cloud Computing Technology Roadmap, Volume II, Release 1.0 (Draft), Useful Information for Cloud Adopters, NIST Special Publication 500-293, National Institute of Standards and Technology, Gaithersburg, Maryland, 2011.

[3] D. Birk and C. Wegener, Technical issues of forensic investigations in cloud computing environments, *Proceedings of the Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2011.

[4] D. Bryan and M. Anderson, Cloud computing, A weapon of mass destruction? presented at the *DEFCON 18 Hacking Conference*, 2010.

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

[6] S. Garfinkel, The criminal cloud, *MIT Technology Review*, October 17, 2011.

[7] D. Garrie, Cloud computing and jurisdiction, Part 2: A primer, Law and Forensics, Seattle, Washington (`www.lawandforensics.com/cloud-computing-jurisdiction-part-primer`), 2012.

[8] A. Haeberlen, A case for the accountable cloud, *ACM SIGOPS Operating Systems Review*, vol. 44(2), pp. 52–57, 2010.

[9] W. Jansen and T. Grance, Guidelines on Security and Privacy in Public Cloud Computing, NIST Special Publication 800-144, National Institute of Standards and Technology, Gaithersburg, Maryland, 2011.

[10] S. Liles, M. Rogers and M. Hoebich, A survey of the legal issues facing digital forensic experts, in *Advances in Digital Forensics V*, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 267–276, 2009.

[11] R. Lu, X. Lin, X. Liang and X. Shen, Secure provenance: The essential of bread and butter of data forensics in cloud computing, *Proceedings of the Fifth ACM Conference on Computer and Communications Security*, pp. 282-292, 2010.

[12] P. Mell and T. Grance, The NIST Definition of Cloud Computing, NIST Special Publication 800-145, National Institute of Standards and Technology, Gaithersburg, Maryland, 2011.

[13] G. Palmer, A Road Map for Digital Forensic Research – Report from the First Digital Forensic Research Workshop, DFRWS Technical Report, DTR-T001-01 FINAL, Air Force Research Laboratory, Rome, New York, 2001.

[14] T. Ristenpart, E. Tromer, H. Schacham and S. Savage, Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds, *Proceedings of the Sixteenth ACM Conference on Computer and Communications Security*, pp 199–212, 2009.

[15] K. Ruan, J Carthy, T. Kechadi and M. Crosbie, Cloud forensics, in *Advances in Digital Forensics VII*, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp 35–46, 2011.

[16] J. Vacca, *Computer Forensics: Computer Crime Scene Investigation*, Charles River Media, Hingham, Massachusetts, 2002.