Technical Challenges of Forensic Investigations in Cloud Computing Environments

Dominik Birk

January 12, 2011

Abstract

Cloud Computing is arguably one of the most discussed information technology topics in recent times. It presents many promising technological and economical opportunities. However, many customers remain reluctant to move their business IT infrastructure completely to "the Cloud". One of the main concerns of customers is Cloud security and the threat of the unknown. Cloud Service Providers (CSP) encourage this perception by not letting their customers see what is behind their "virtual curtain". A seldomly discussed, but in this regard highly relevant open issue is the ability to perform digital investigations. This continues to fuel insecurity on the sides of both providers and customers. In Cloud Forensics, the lack of physical access to servers constitutes a completely new and disruptive challenge for investigators. Due to the decentralized nature of data processing in the Cloud, traditional approaches to evidence collection and recovery are no longer practical. This paper focuses on the technical aspects of digital forensics in distributed Cloud environments. We contribute by assessing whether it is possible for the customer of Cloud Computing services to perform a traditional digital investigation from a technical standpoint. Furthermore we discuss possible new methodologies helping customers to perform such investigations and discuss future issues.

1 Introduction

Although the Cloud might appear attractive to small as well to large companies, it does not come along without its own unique problems and concerns. Outsourcing sensitive corporate data into the Cloud raises concerns regarding the privacy and security of the data. Security policies, companies main pillar concerning security, cannot be easily deployed into distributed Cloud environments. This situation is further complicated by the unknown physical location of the companie's assets. Normally, if a security incident occurs, the corporate security team wants to be able to perform their own investigation without dependency on third parties. In the Cloud, this is not possible anymore. The CSP obtains all the power over the Cloud environment mainly biasing the way an investigation may be processed.

1.1 Technical Background

According to the NIST [13], Cloud Computing is a model for enabling 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. The new raw definition of Cloud Computing brought several new characteristics such as multi-tenancy, elasticity, pay-as-you-go and reliability. Within this work, the following three models are used in the context of Cloud Computing:

In the *Infrastructure as a Service (IaaS)* model, the customer is using the virtual machine provided by the CSP for installing his own system on it. The system can be used like any other physical computer with a few limitations. However, the additive power over the system comes along with additional security obligations. *Platform as a Service (PaaS)* offerings provide the capability to deploy application packages created using the virtual development environment supported by the CSP. For the efficiency of Software Development Process this service model can be propellent. In the *Software as a Service (SaaS)* model, the customer makes use of a service run by the CSP on a Cloud infrastructure. In most of the cases this service can be accessed through an API for a thin client interface such as a web browser. Closed-source

public SaaS offers such as Amazon S3 and GoogleMail can only be used in the public deployment model leading to further issues concerning security, privacy and the gathering of suitable evidences.

Furthermore, the two main deployment models, private and public Cloud [1] have to be distinguished. Common *public* Clouds are made available to the general public or a large industry group. The corresponding Cloud infrastructure is owned by an organization acting as a CSP and offering services to its customers. In contrast, the *private* cloud is solely operated for an organization but may not provide the scalability and agility of public Cloud offers. The additional notions of *Community Cloud* and *Hybrid Cloud* are not exclusively covered within this work. However, independently from the specific model used, the movement of applications to the Cloud offered as services by a CSP, comes along with limited control for the customer about the application itself, the data pushed into the applications and also about the underlying technical infrastructure.

2 Motivation

With the increasing demand for using the power of the Cloud for processing also sensible information and data, enterprises face the issue of Data and Process Provenance in the Cloud [14]. Digital provenance, meaning meta-data that describes the ancestry or history of a digital object, is a crucial feature for forensic investigations. In combination with a suitable authentication scheme, it provides information about who created and who modified what kind of data in the Cloud. These are essential aspects for digital investigations in distributed environments such as the Cloud.

Unfortunately, the aspect of forensic investigations in such distributed environment has so far been mostly neglected by the research community. Current discussion centers mostly around Cloud Security and Privacy/Data Protection [16]. The impact of forensic investigations on Cloud environments was little noticed. In 2009, the authors of [4] stated that *"[...] to our knowledge, no research has been published on how cloud computing environments affect digital artifacts, and on acquisition logistics and legal issues related to cloud computing environments."*. At the same time, massive investments are being made in Cloud technology. Combined with the fact that information technology increasingly exceeds peoples' private and professional life, thus mirroring more and more of peoples' actions, it becomes apparent that evidence gathered from Cloud environments will be of high significance to litigation or criminal proceedings in the future.

3 Technical Issues

Digital investigations are about control of forensic evidence data. From the technical point of view, this evidence data can be available in three different states: at *rest*, in *motion* or in *execution*.

Data at *rest* is represented by allocated disk space. Whether the data is stored in a database or in a specific file format, it allocates disk space. Furthermore, if a file is deleted, the disk space is de-allocated for the operating system but the data is still accessible since the disk space has not been re-allocated and overwritten. This fact is often exploited by investigators which explore these de-allocated disk space on hard-disks. In case the data is in *motion*, data is transferred from one entity to another e.g. a typical file transfer over a network can be seen as a data in motion scenario. Several encapsulated protocols contain the data each leaving specific traces on systems and network devices which can in return be used by investigators. Data can be loaded into memory and executed as a process. In this case, the data is neither at rest or in motion but in *execution*. On the executing system, process information, machine instruction and allocated/de-allocated data can be analyzed by creating a snapshot of the current system state. This snapshot technology is further discussed in section 3.2.3.

3.1 Sources and Nature of Evidence

Concerning the technical aspects of forensic investigations, the amount of potential evidence available to the investigator strongly diverges between the different Cloud service and deployment models. Independently from the used model, the following three components could act as sources for potential evidential data.

3.1.1 Virtual Cloud Instance

The virtual instance within the Cloud, where i.e. data is stored or processes are handled, provides potential evidence [3, 2, 6]. In most of the cases, it is the place where an incident happened and hence provides a potential starting point for a forensic investigation. The instance can be accessed by both, the CSP and the customer who is running the instance. Snapshots provide a powerful technique for the customer to freeze specific states of the virtual machine. Therefore, virtual instances can be still running which leads to the case of live investigations or can be turned off leading to static image analysis. In SaaS and PaaS scenarios, the ability to access the virtual instance for gathering evidential information is higly limited or simply not possible.

3.1.2 Network Layer

The different ISO/OSI network layers provide several information on protocols and communication between instances within the Cloud as well as with instances outside the Cloud [12, 9, 8]. Unfortunately, ordinary CSP currently do not provide any log data from the network components. This means, that in case of malware infection of an IaaS VM, it will be difficult to get any form of routing information. This situation gets even more complicated in case of PaaS or SaaS. Hence, the situation of forensic evidence is again strongly affected by the support the investigator receives from the customer and the CSP.

3.1.3 Client System

On the system layer of the client, it completely depends on the used model (IaaS, PaaS, SaaS) if and where potential evidence could be extracted. In most of the Cloud scenarios, the browser on the client system is the only application that communicates with the service in the Cloud. This especially holds for SaaS applications which are used and controlled by the web browser. Hence, in an exhaustive forensic investigation, the evidence data gathered from the browser environment [15] should not be omitted.

3.2 Investigations in XaaS Environments

Within this section specific issues of investigations in SaaS, PaaS and IaaS environments will be discussed.

3.2.1 SaaS Environments

Especially in the SaaS model, the customer does not obtain any control of the underlying operating infrastructure such as network, servers, operating systems etc. or even the application that is used. This means that no deeper view into the system and its underlying infrastructure is provided to the customer. Only limited user-specific application configuration settings can be controlled. In a lot of cases this urges the investigator to rely on high-level logs which are eventually provided by the CSP. Given the case that the CSP does not run any logging application, the customer has no opportunity to create any useful evidence by himself. The installation or configuration of any toolkit or logging tool is impossible.

These circumstances do not allow a valid forensic investigation and lead to the assumption that customers of SaaS models do not have any chance to analyze potential incidences. Moreover, evidence data has to be interpreted by an investigator in a proper manner which is hardly be possible due to the lack of circumstantial information. For auditors, this situation does not change: Questions who accessed specific data and information cannot be answered by the customers, if no corresponding logs are available.

Moreover, a lot of SaaS CSP like Google offer Single sign-on (SSO) access control to the complete set of their services. Unfortunately in case of an account compromise, most of the CSP do not offer any possibility for the customer to figure out which data and information has been accessed by the adversary. In private SaaS scenarios this situation is tremendously improved by the fact that the customer and the CSP are probably under the same authority. Hence, logging mechanisms could be implemented which contribute to potential investigations. Additionally, the exact location of the servers and the data is known at any time.

Due to the limited ability of receiving forensic information from the server in SaaS scenarios, the client has to contribute to this process. This can be achieved by implementing *Proofs of Retrievability (POR)* in which a verifier (client) is enabled to determine that a prover (server) possesses a file or data object

without actually downloading it [11]. In [17], the authors introduced a new data integrity verification mechanism for SaaS scenarios which could also be used for forensic purposes.

3.2.2 PaaS Environments

One of the main advantages of this model is that the core application is under the control of the customer. Given these circumstances, the customer obtains theoretically the power to dictate how the application interacts with other dependencies (databases, storage entities etc.). Moreover, depending on the runtime environment, logging mechanisms can be implemented which automatically sign the information and transfer it to a third party storage. Additional encryption could prevent potential eavesdroppers from being able to view log data information on the way to storage server. CSP normally claim that this transfer is encrypted but this statement can hardly be verified. Since the customer has the ability to interact with the platform over a prepared API, system states and specific application logs can be extracted. However potential adversaries, which can compromise the application during runtime, should not be able to alter these log files afterwards which could be realized by push-only mechanisms.

Unfortunately, the customer has no direct control of the underlying runtime environment. As in the SaaS scenario, this strongly depends on the configuration done by the CSP. Concerning the Microsoft Azure platform, the environment is made of an virtualized OS (Microsoft Windows), a webserver (Internet Information Server) and the runtime environment (.NET). Windows Azure Diagnostics, a new feature released in November 2009, gives developers the ability to collect and store a variety of diagnostics data in a highly configurable way.

3.2.3 IaaS Environments

From the forensic point of view, IaaS instances provide much more information that could be used as forensic evidence in case of an incident than the PaaS and SaaS models do [6]. This fact is caused through the ability of the customer to install and set up the image for forensic purposes. Hence, log data and other evidence information could be transferred to other hosts in a frequent manner for providing the ability to perform an investigation if needed.

Snapshots Traditional forensics expect target machines to be powered down to collect an image. This situation completely changed with the advent of the snapshot method which is supported by all popular hypervisors such as Xen, VMware ESX and Hyper-V[1]. Snapshots, also referred to as *forensic image*, of virtual machines provide a powerful tool with which a virtual machine can be cloned by one click including also the running system's memory. This leads to the main benefit that systems hosting crucial business processes do not have to be shutdown for performing a forensic analysis. This could also affect scenarios in which a downtime of a system is not feasible or practical due to existing SLAs.

Due to the fact that the customer is responsible for the security of the virtual instance, the system itself can be prepared for forensic investigation purposes. RFC 3227 [7] contains several best practices for responding to a security incident especially in the case of live investigating systems. According to [3], log data information concerning currently logged users, open ports, running processes, system and registry information etc. should be gathered. These log data should be transferred to an external system mitigating the chance that a maliciously motivated shutdown process destroys the data. Encrypting and signing these log files can be helpful for providing security and integrity of the created log files. Unfortunately, it has to be emphasized that each process such as an encryption process run on the virtual instance, can be controlled by the hypervisor or the CSP respectively. Although this risk can be disregarded in most of the cases, the impact on the security of high security environments is tremendous.

Generally, for an investigator it is important to know if the virtual machine was properly shutdown or is still running [2]. Hence, virtual instances have to be divided into two different categories concerning the forensic analysis of the system: *shutdown (dead)* and *running (live)* virtual systems. In general, live investigations on running virtual machines become more common providing evidence data that is not available on shutdown systems. The technique of live investigation is mostly influenced by the huge amount of evidence data that has to be stored and processed in case of shutdown instances. Nevertheless, it cannot be denied that live investigations change the state of the investigated system and the results

[1] It should be mentioned that proprietary file formats present an issue to investigators [5].

of the investigation may not be repeatable. However, this does no prevent a lot of SMCs from mostly performing live investigations due to the bond of legislation and government-contracting agreements.

Volatile Data Depending on the Cloud offer used, virtual IaaS instances do not have any persistent storage. In the specific case of an AWS EC2 cloud instance, all volatile data is lost if the instance is rebooted or shutdown. Persistent data has to be stored in long time storage environments like Amazon Simple Storage Service (S3) or Amazon Elastic Block Store (EBS).

This situation leads to several issues: In case an adversary compromises a virtual IaaS instance with no persistent storage synchronization, the adversary could shutdown the system leading to a complete loss of volatile data. Additionally, the instance could be abused for sending spam, attack further external and internal targets, join botnets and steal volatile data of the running system. After the attack, the attacker can cancel the contract with the corresponding CSP forcing the virtual machine to shutdown and destroy most of the evidence data which is inevitable for further forensic investigations. This problem mainly results from the unclear situation how CSP handle the termination of customer contracts. In real world scenarios, this process is not transparent for the customer bringing up further questions e.g. does data on virtual systems in the Cloud get exhaustively deleted and how is this done? File deletion is all about control and this used to not be an issue till the advent of Cloud Computing. In current Cloud environments CSP do not offer any verification process providing the ability for the customer to verify that the sensitive data stored on a virtual machine has been deleted exhaustively.

Moreover, an interesting perspective is the case in which the real owner of the image decides to engage in malicious activities through his EC2 machine from a veiled IP address and afterwards claims, someone compromised the password or key pair to his EC2 machine. In a subsequent forensic investigation, it will be difficult to prove the opposite due to the lack of evidences.

Virtual Introspection As expected, even virtual instances in the Cloud get compromised by adversaries as happened to Amazon EC2 instances in the end of 2009[2]. Hence, the ability to determine how defenses in the virtual environment failed and to what extent the affected systems have been compromised is crucial not only for recovering from an incident. Also forensic investigations gain leverage from such information and contribute to resilience against future attacks on the systems.

Virtual Introspection (VI) is the process by which the state of a virtual machine is observed from either the VMM or from some virtual machines other than the one being examined [10]. However, the fact that the VMM has full access to the resources of all VMs represents a significant risk to customers' data. The issue whether VMs can ever be managed by a VMM, while simultaneously being protected from a compromised VMM remains an open research problem.

4 Conclusion

There is no doubt that Cloud Computing has various security benefits for companies which under ordinary circumstances struggle with limited budgets for security resources. However, regarding digital forensics, the loss of control caused by Cloud environments and vendors presents a huge challenge for investigators. Preliminary findings of the computer forensic community in the field of digital forensics have to be revised and adapted to the new environment. Investigators need the possibility of reconstructing the corresponding environment for recreating scenarios and test hypothesizes. In the fast fluctuating world of Cloud Computing, this is not possible anymore.

Furthermore, the question comes up what collateral data the investigator can identify and collect to help him prove or disprove the hypothesis. Is there any knowledge about what the CSP logs and how long he keeps this information? Does the CSP vouch for the integrity of the evidential data? Currently, customers have to accept that evidential artifacts of digital investigations in Cloud environments will be unreliable and incomplete. Considering the concessions made by the CSP through SLAs or other forms of contracts, it is almost impossible for the customer to verify this. Suppose that a customer made a contract with a CSP which guarantees data redundancy for the customer data. How can the customer prove that this agreement is fulfilled? However, these reasons are not the only ones making Cloud Computing a complex issue for digital investigations. The absence of a standards[3] for processes

[2] http://news.cnet.com/8301-1009_3-10413951-83.html
[3] http://cloud-standards.org

within the Cloud and the Cloud in general causes a lot of problems ranging from security, compliance and proper deployment to the question of how an investigation within such an environment shall be processed. When new standards or adjustments to existing standards are needed, as it is the case with Cloud Computing, creating too many standards should be avoided. Standards have to promote innovation and do not inhibit it.

References

[1] Cloud computing: Business benefits with security, governance and assurance perspectives. Technical report, ISACA, 2009.

[2] R. A. Bares. Hiding in a virtual world: using unconventionally installed operating systems. In *ISI'09: Proceedings of the 2009 IEEE international conference on Intelligence and security informatics*, pages 276–284, Piscataway, NJ, USA, 2009. IEEE Press.

[3] D. Barrett and G. Kipper. *Virtualization and Forensics: A Digital Forensic Investigator's Guide to Virtual Environments*. Syngress, 6 2010.

[4] N. Beebe. Digital forensic research: The good, the bad and the unaddressed. *Advances in Digital Forensics V*, pages 17–36, 2009.

[5] D. Bem. Virtual machine for computer forensics - the open source perspective. In E. Huebner and S. Zanero, editors, *Open Source Software for Digital Forensics*, pages 25–42. Springer US, 2010.

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

[7] D. Brezinski and T. Killalea. Guidelines for evidence collection and archiving, 2002.

[8] V. Corey, C. Peterman, S. Shearin, M. Greenberg, and J. Van Bokkelen. Network forensics analysis. *IEEE Internet Computing*, 6(6):60–66, 2002.

[9] EC-Council. *Computer Forensics: Investigating Network Intrusions and Cyber Crime (Ec-Council Press Series: Computer Forensics)*. Course Technology, 1 edition, 9 2009.

[10] B. Hay and K. Nance. Forensics examination of volatile system data using virtual introspection. *SIGOPS Oper. Syst. Rev.*, 42:74–82, April 2008.

[11] A. Juels and B. S. Kaliski. Pors: proofs of retrievability for large files. In *In CCS '07: Proceedings of the 14th ACM conference on Computer and communications security*, pages 584–597. ACM, 2007.

[12] R. Meadows. *Cisco Router and Switch Forensics: Investigating and Analyzing Malicious Network Activity*. Elsevier Science, 1st edition, 4 2009.

[13] P. Mell. Nist.gov - computer security division - computer security resource center, February 2010.

[14] K.-K. Muniswamy-Reddy and M. Seltzer. Provenance as first class cloud data. *SIGOPS Oper. Syst. Rev.*, 43(4):11–16, 2010.

[15] M. T. Pereira. Forensic analysis of the firefox 3 internet history and recovery of deleted sqlite records. *Digital Investigation*, 5(3-4):93–103, 2009.

[16] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage. Hey, you, get off of my cloud! Exploring information leakage in third-party compute clouds. In S. Jha and A. Keromytis, editors, *Proceedings of CCS 2009*, pages 199–212. ACM Press, Nov. 2009.

[17] Y. Shi, K. Zhang, and Q. Li. A new data integrity verification mechanism for saas. In F. Wang, Z. Gong, X. Luo, and J. Lei, editors, *Web Information Systems and Mining*, volume 6318 of *Lecture Notes in Computer Science*, pages 236–243. Springer Berlin / Heidelberg, 2010.