

# Technical Issues of Forensic Investigations in Cloud Computing Environments

Dominik Birk  
Ruhr-University Bochum  
Horst Goertz Institute for IT Security  
Bochum, Germany  
Email: dominik.birk@rub.de

Christoph Wegener  
Ruhr-University Bochum  
Horst Goertz Institute for IT Security  
Bochum, Germany  
Email: christoph.wegener@rub.de

**Abstract**—Cloud Computing is arguably one of the most discussed information technologies today. 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 their main concerns 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. *Cloud Forensics* constitutes a 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 point of view. Furthermore we discuss possible solutions and possible new methodologies helping customers to perform such investigations.

## I. INTRODUCTION

Although the cloud might appear attractive to small as well as to large companies, it does not come along without its own unique problems. Outsourcing sensitive corporate data into the cloud raises concerns regarding the privacy and security of data. Security policies, companies main pillar concerning security, cannot be easily deployed into distributed, virtualized cloud environments. This situation is further complicated by the unknown physical location of the companies' 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 environment and thus controls the sources of evidence. In the best case, a trusted third party acts as a trustee and guarantees for the trustworthiness of the CSP.

Furthermore, the implementation of the technical architecture and circumstances within cloud computing environments bias the way an investigation may be processed. In detail, evidence data has to be interpreted by an investigator in a

proper manner which is hardly 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.

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 [10]. 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 crucial aspects for digital investigations in distributed environments such as the cloud.

Unfortunately, the aspects of forensic investigations in distributed environment have so far been mostly neglected by the research community. Current discussion centers mostly around security, privacy and data protection issues [35], [9], [12]. The impact of forensic investigations on cloud environments was little noticed albeit mentioned by the authors of [1] in 2009: "[...] 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." This statement is also confirmed by other authors [34], [36], [40] stressing that further research on incident handling, evidence tracking and accountability in cloud environments has to be done.

At the same time, massive investments are being made in cloud technology. Combined with the fact that information technology increasingly transcends 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.

Within this work, we focus the notion of cloud forensics by addressing the technical issues of forensics in all three major cloud service models and consider cross-disciplinary aspects. Moreover, we address the usability of various sources of evidence for investigative purposes and propose potential solutions to the issues from a practical standpoint. This work

We would like to thank the reviewers for the helpful comments and Dennis Heinson (Center for Advanced Security Research Darmstadt - CASED) for the profound discussions regarding the legal aspects of cloud forensics.

should be considered as a surveying discussion of an almost unexplored research area.

The paper is organized as follows: We discuss the related work and the fundamental technical background information of digital forensics, cloud computing and the fault model in section II and III. In section IV, we focus on the technical issues of cloud forensics and discuss the potential sources and nature of digital evidence as well as investigations in XaaS environments including the cross-disciplinary aspects. We conclude in section V.

## II. RELATED WORK

Various works have been published in the field of cloud security and privacy [9], [35], [30] focussing on aspects for protecting data in multi-tenant, virtualized environments. Desired security characteristics for current cloud infrastructures mainly revolve around isolation of multi-tenant platforms [12], security of hypervisors in order to protect virtualized guest systems and secure network infrastructures [32].

Albeit digital provenance, describing the ancestry of digital objects, still remains a challenging issue for cloud environments, several works have already been published in this field [8], [10] contributing to the issues of cloud forensics. Within this context, cryptographic proofs for verifying data integrity mainly in cloud storage offers have been proposed, yet lacking of practical implementations [24], [37], [23].

Traditional computer forensics has already well researched methods for various fields of application [4], [5], [6], [11], [13]. Also the aspects of forensics in virtual systems have been addressed by several works [2], [3], [20] including the notion of virtual introspection [25]. In addition, the NIST already addressed *Web Service Forensics* [22] which has a huge impact on investigation processes in cloud computing environments.

In contrast, the aspects of forensic investigations in cloud environments have mostly been neglected by both the industry and the research community. One of the first papers focusing on this topic was published by Wolthusen [40] after Bebee et al already introduced problems within cloud environments [1]. Wolthusen stressed that there is an inherent strong need for interdisciplinary work linking the requirements and concepts of evidence arising from the legal field to what can be feasibly reconstructed and inferred algorithmically or in an exploratory manner. In 2010, Grobauer et al [36] published a paper discussing the issues of incident response in cloud environments - unfortunately no specific issues and solutions of cloud forensics have been proposed which will be done within this work.

## III. TECHNICAL BACKGROUND

### A. Traditional Digital Forensics

The notion of Digital Forensics is widely known as the practice of identifying, extracting and considering evidence from digital media. Unfortunately, digital evidence is both fragile and volatile and therefore requires the attention of

special personnel and methods in order to ensure that evidence data can be proper isolated and evaluated.

Normally, the process of a digital investigation can be separated into three different steps each having its own specific purpose:

- 1) In the *Securing Phase*, the major intention is the preservation of evidence for analysis. The data has to be collected in a manner that maximizes its integrity. This is normally done by a bitwise copy of the original media. As can be imagined, this represents a huge problem in the field of cloud computing where you never know exactly where your data is and additionally do not have access to any physical hardware. However, the snapshot technology, discussed in section IV-B3, provides a powerful tool to freeze system states and thus makes digital investigations, at least in IaaS scenarios, theoretically possible.
- 2) We refer to the *Analyzing Phase* as the stage in which the data is sifted and combined. It is in this phase that the data from multiple systems or sources is pulled together to create as complete a picture and event reconstruction as possible. Especially in distributed system infrastructures, this means that bits and pieces of data are pulled together for deciphering the real story of what happened and for providing a deeper look into the data.
- 3) Finally, at the end of the examination and analysis of the data, the results of the previous phases will be reprocessed in the *Presentation Phase*. The report, created in this phase, is a compilation of all the documentation and evidence from the analysis stage. The main intention of such a report is that it contains all results, it is complete and clear to understand.

Apparently, the success of these three steps strongly depends on the first stage. If it is not possible to secure the complete set of evidence data, no exhaustive analysis will be possible. However, in real world scenarios often only a subset of the evidence data can be secured by the investigator. In addition, an important definition in the general context of forensics is the notion of a *Chain of Custody*. This chain clarifies how and where evidence is stored and who takes possession of it. Especially for cases which are brought to court it is crucial that the chain of custody is preserved.

### B. Cloud Computing

According to the NIST [16], 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 CSP 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 *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 customer 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, two main deployment models, *private* and *public* cloud have to be distinguished. Common *public* clouds are made available to the general public. The corresponding infrastructure is owned by one organization acting as a CSP and offering services to its customers. In contrast, the *private* cloud is exclusively operated for an organization but may not provide the scalability and agility of public offers. The additional notions of *community* and *hybrid* cloud are not exclusively covered within this work. However, independently from the specific model used, the movement of applications and data to the cloud 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.

### C. Fault Model

Be it an account for a SaaS application, a development environment (PaaS) or a virtual image of an IaaS environment, systems in the cloud can be affected by inconsistencies. Hence, for both customer and CSP it is crucial to have the ability to assign faults to the causing party, even in the presence of Byzantine behavior [33].

Generally, inconsistencies can be caused by the following two reasons:

#### 1) Maliciously Intended Faults

Internal or external adversaries with specific malicious intentions can cause faults on cloud instances or applications. Economic rivals as well as former employees can be the reason for these faults and state a constant threat to customers and CSP. In this model, also a malicious CSP is included albeit he is assumed to be rare in real world scenarios. Additionally, from the technical point of view, the movement of computing power to a virtualized, multi-tenant environment can pose further threads and risks to the systems. One reason for this is that if a single system or service in the cloud is compromised, all other guest systems and even the host system are at risk. Hence, besides the need for further security measures, precautions for potential forensic investigations have to be taken into consideration.

#### 2) Unintentional Faults

Inconsistencies in technical systems or processes in the cloud do not have implicitly to be caused by malicious intent. Internal communication errors or human failures can lead to issues in the services offered to the customer (i.e. loss or modification of data). Although these failures are not caused intentionally, both the CSP and the customer have a strong intention to discover the reasons and deploy corresponding fixes.

## IV. TECHNICAL ISSUES

Digital investigations are about control of forensic evidence data. From the technical standpoint, this 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 harddisks.

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.

In the following sections, we point out the potential sources for evidential data in cloud environments and discuss the technical issues of digital investigations in XaaS environments as well as suggest several solutions to these problems.

### A. 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.

The virtual machine (VM), hosting in most of the cases the server application, provides several pieces of information that could be used by investigators. On the network level, network components can provide information about possible communication channels between different parties involved. The browser on the client, acting often as the user agent for communicating with the cloud, also contains a lot of information that could be used as evidence in a forensic investigation. Independently from the used model, the following three components could act as sources for potential evidential data.

1) *Virtual Cloud Instance*: The VM within the cloud, where i.e. data is stored or processes are handled, contains potential evidence [2], [3]. In most of the cases, it is the place where an incident happened and hence provides a good starting point for a forensic investigation. The VM instance can be accessed by both, the CSP and the customer who is running the instance. Furthermore, virtual introspection techniques [25] provide access to the runtime state of the VM via the hypervisor and snapshot technology supplies a powerful technique for the customer to freeze specific states of the VM. Therefore, virtual instances can be still running during analysis which leads to the case of live investigations [41] 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 highly limited or simply not possible.

2) *Network Layer*: Traditional network forensics is known as the analysis of network traffic logs for tracing events that have occurred in the past. Since the different ISO/OSI network layers provide several information on protocols and communication between instances within as well as with instances outside the cloud [4], [5], [6], network forensics is theoretically also feasible in cloud environments. However in practice, ordinary CSP currently do not provide any log data from the network components used by the customer's instances or applications. For instance, in case of a malware infection of an IaaS VM, it will be difficult for the investigator to get any form of routing information and network log data in general which is crucial for further investigative steps. This situation gets even more complicated in case of PaaS or SaaS. So again, the situation of gathering forensic evidence is strongly affected by the support the investigator receives from the customer and the CSP.

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 scenarios, the user agent (e.g. the web 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. But also in IaaS scenarios, the administration interface is often controlled via the browser. Hence, in an exhaustive forensic investigation, the evidence data gathered from the browser environment [7] should not be omitted.

a) *Browser Forensics*: Generally, the circumstances leading to an investigation have to be differentiated: In ordinary scenarios, the main goal of an investigation of the web browser is to determine if a user has been victim of a crime. In complex SaaS scenarios with high client-server interaction, this constitutes a difficult task. Additionally, customers strongly make use of third-party extensions [17] which can be abused for malicious purposes. Hence, the investigator might want to look for malicious extensions, searches performed, websites visited,

files downloaded, information entered in forms or stored in local HTML5 stores, web-based email contents and persistent browser cookies for gathering potential evidence data. Within this context, it is inevitable to investigate the appearance of malicious JavaScript [18] leading to e.g. unintended AJAX requests and hence modified usage of administration interfaces.

Generally, the web browser contains a lot of electronic evidence data that could be used to give an answer to both of the above questions - even if the private mode is switched on [19].

## B. Investigations in XaaS Environments

Traditional digital forensic methodologies permit investigators to seize equipment and perform detailed analysis on the media and data recovered [11]. In a distributed infrastructure organization like the cloud computing environment, investigators are confronted with an entirely different situation. They have no longer the option of seizing physical data storage. Data and processes of the customer are dispensed over an undisclosed amount of virtual instances, applications and network elements. Hence, it is in question whether preliminary findings of the computer forensic community in the field of digital forensics apparently have to be revised and adapted to the new environment.

Within this section, specific issues of investigations in SaaS, PaaS and IaaS environments will be discussed. In addition, cross-disciplinary issues which affect several environments uniformly, will be taken into consideration. We also suggest potential solutions to the mentioned problems.

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 or 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 contributing to the evidences which can be extracted from the client (see section IV-A3). 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 through the installation of any toolkit or logging tool. These circumstances do not allow a valid forensic investigation and lead to the assumption that customers of SaaS offers do not have any chance to analyze potential incidences.

a) *Data Provenance*: The notion of *Digital Provenance* is known as meta-data that describes the ancestry or history of digital objects. Secure provenance that records ownership and process history of data objects is vital to the success of data forensics in cloud environments, yet it is still a challenging issue today [8]. Albeit data provenance is of high significance also for IaaS and PaaS, it states a huge problem specifically for SaaS-based applications:

Current global acting public SaaS CSP offer Single Sign-On (SSO) access control to the 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. For the victim, this situation can have tremendous impact: If sensitive data has been compromised, it is unclear which data has been leaked and which has not been accessed by the adversary. Additionally, data could be modified or deleted by an external adversary or even by the CSP e.g. due to storage reasons. The customer has no ability to proof otherwise. Secure provenance mechanisms for distributed environments can improve this situation but have not been practically implemented by CSP [10].

*Suggested Solution:*

In private SaaS scenarios this situation is improved by the fact that the customer and the CSP are probably under the same authority. Hence, logging and provenance mechanisms could be implemented which contribute to potential investigations. Additionally, the exact location of the servers and the data is known at any time.

Public SaaS CSP should offer additional interfaces for the purpose of compliance, forensics, operations and security matters to their customers. Through an API, the customers should have the ability to receive specific information such as access, error and event logs that could improve their situation in case of an investigation. Furthermore, due to the limited ability of receiving forensic information from the server and proofing integrity of stored data in SaaS scenarios, the client has to contribute to this process. This could 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 and it can be retrieved unmodified [24]. *Provable Data Possession (PDP)* techniques [37] could be used to verify that an untrusted server possesses the original data without the need for the client to retrieve it. Although these cryptographic proofs have not been implemented by any CSP, the authors of [23] introduced a new data integrity verification mechanism for SaaS scenarios which could also be used for forensic purposes.

2) *PaaS Environments:* One of the main advantages of the PaaS model is that the developed software application is under the control of the customer and except for some CSP, the source code of the application does not have to leave the local development environment. Given these circumstances, the customer obtains theoretically the power to dictate how the application interacts with other dependencies such as databases, storage entities etc. CSP normally claim this transfer is encrypted but this statement can hardly be verified by the customer. 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.

*Suggested Solution:*

Depending on the runtime environment, logging mechanisms could be implemented which automatically sign and encrypt the log information before its transfer to a central logging server under the control of the customer. Additional signing and encrypting could prevent potential eavesdroppers from being able to view and alter log data information on the way to the logging server. Runtime compromise of an PaaS application by adversaries could be monitored by push-only mechanisms for log data presupposing that the needed information to detect such an attack are logged. Increasingly, CSP offering PaaS solutions give developers the ability to collect and store a variety of diagnostics data in a highly configurable way with the help of runtime feature sets [38].

3) *IaaS Environments:* As expected, even virtual instances in the cloud get compromised by adversaries. 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.

From the forensic point of view, IaaS instances do provide much more evidence data usable for potential forensics than PaaS and SaaS models do. This fact is caused through the ability of the customer to install and set up the image for forensic purposes before an incident occurs. Hence, as proposed for PaaS environments, log data and other forensic evidence information could be signed and encrypted before it is transferred to third-party hosts mitigating the chance that a maliciously motivated shutdown process destroys the volatile data.

Although, IaaS environments provide plenty of potential evidence, it has to be emphasized that the customer VM is in the end still under the control of the CSP. He controls the hypervisor which is e.g. responsible for enforcing hardware boundaries and routing hardware requests among different VM. Hence, besides the security responsibilities of the hypervisor, he exerts tremendous control over how customer's VM communicate with the hardware and theoretically can intervene executed processes on the hosted virtual instance through virtual introspection [25]. This could also affect encryption or signing processes executed on the VM and therefore leading to the leakage of the secret key. Although this risk can be disregarded in most of the cases, the impact on the security of high security environments is tremendous.

a) *Snapshot Analysis:* Traditional forensics expect target machines to be powered down to collect an image (*dead virtual instance*). This situation completely changed with the advent of the snapshot technology which is supported by all popular hypervisors such as Xen, VMware ESX and Hyper-V. A snapshot, also referred to as the forensic image of a VM, provides a powerful tool with which a virtual instance can be cloned

by one click including also the running system's memory. Due to the invention of the snapshot technology, systems hosting crucial business processes do not have to be powered down for forensic investigation purposes. The investigator simply creates and loads a snapshot of the target VM for analysis (*live virtual instance*). This behavior is especially important for scenarios in which a downtime of a system is not feasible or practical due to existing SLA. However the information whether the machine is running or has been properly powered down is crucial [3] for the investigation.

Live investigations of running virtual instances become more common providing evidence data that is not available on powered down 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 powered down instances. Nevertheless, if no snapshot of the target VM is used, it cannot be denied that live investigations change the state of the investigated system and the results of the investigation may not be repeatable. Unfortunately, this does not prevent a lot of companies from mostly performing live investigations due to the bond of legislation and government-contracting agreements.

*b) Volatile Data:* Depending on the cloud offer used, IaaS VM do not have any persistent storage. This means in most of the cases all volatile data is lost if the VM is rebooted or powered down. The *on-demand* characteristic of the cloud is one reason for such behavior. Furthermore, with the help of such measures, CSP force their customers to subscribe to further offers for storing data persistently but leading to further costs.

Generally, this situation leads to several issues: In case an adversary compromises a VM with no persistent storage synchronization, the adversary could shutdown the system leading to a complete loss of volatile data, if no further countermeasures are installed. 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 VM to be powered down 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 (see section IV-B4d for further discussion).

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

#### *Suggested Solution:*

As live investigations become more common, the method of *Virtual Introspection (VI)* for live forensics of virtual instances could be helpful [25]. VI is the process by which the state of a virtual machine is observed from either the hypervisor or from some virtual machines other than the one being examined. However, the fact that the hypervisor 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 hypervisor, while simultaneously being protected from a compromised hypervisor remains an open research problem.

The loss of huge amounts of volatile data could be mitigated through frequent data synchronization between the VM and the persistent storage or a non-cloud based storage. However, the loss of volatile data on running systems compromised by an adversary cannot be mitigated, if the CSP does not take precautions. One solution to provide the ability of performing an investigation given the case an instance has been compromised is by providing an API to the customers. In case of a malicious behavior or unintended shutdown of the instance, the customer can read forensic evidence information over the API which stores significant information for a given time.

*4) Cross-Disciplinary Issues:* Besides the specific issues discussed in the previous sections, several cross-disciplinary aspects of forensics in cloud infrastructures have to be considered which count for each single service model alike. These issues are mainly founded in the general concept of cloud computing and do not result from specific service model characteristics. Within this section, we discuss these issues in the context of cloud forensics and propose potential solutions.

*a) Lack of Transparency:* The lack of knowledge about the internal processes, infrastructure and system components make the usage of current cloud computing offers a game of hazard. Customers of cloud services want transparency which is not provided in current real world cloud environments. This is a comprehensible demand due to the fact that in a lot of cases sensible data is computed on services running in the cloud. Without transparency trust is hardly possible. This situation leads to the fact that customers have the legitimate fear of the thread of the unknown. From an economical point of view, the lack of transparency is one of the main reasons why the whole potential of cloud computing is not yet being realized.

Compared with traditional IT outsourcing, cloud computing is peculiar in the fact that physical access to the servers is technically not feasible to customers and investigators alike. It lies in the nature of cloud computing that the exact location of where data is being processed in most of the cases cannot be determined. Consequentially, even determining the applicable body of law that governs and potentially restricts the scope

and proper measures of an investigation is a challenge.

The issue of unknown data location is further enhanced by the technical obfuscation of the underlying infrastructure. The CSP provides almost no information about the system environment in which customer data is stored or processed. This fact has several reasons as adversaries could use technical information about infrastructure and system usage for launching attacks against the CSP or the customers [12] alike. In addition, CSP do not want their customers to see the workload of their service offer. Competitors could use this information for improving their own range of services or use it to harm the reputation of the company.

In the context of cloud forensics, the lack of transparency and trust results in untrustworthy evidence data. The combination of limited access to evidence data and insufficient infrastructure transparency provided by the CSP tremendously exacerbates the ability to perform a digital investigation.

#### *Suggested Solution:*

Unfortunately, building an open, scalable and reusable cloud computing architecture which satisfies the wishes of both customers and CSP still faces challenges in the areas of technology breakthrough and best industry practices. Nevertheless, a long-term trust relationship between customer and CSP can only be established if open-source software frameworks [26] substitute present proprietary cloud platforms. Additionally, CSP have to break the silence and defer to the wishes of customers concerning forensics and security of cloud platforms by providing requested information. The cloud services offered by CSP have to be made accountable [34] meaning that actions are undeniably linked to the node that performed it, systems and applications maintain a record of past actions and evidences of faults can be verified independently by a third party.

*b) Loss of Evidence Data:* Tracking and monitoring user activity is a common process concerning compliance requirements and also contributes to the identification of potential security issues and to future forensic investigations. Depending on the service model, access to relevant log data will be significantly decreased. Although, cloud environments theoretically provide a huge amount of potential evidence data that could be used for an investigation, the CSP mostly decides which amount of evidence data can be accessed by the customer. As discussed in section IV-B, customers of SaaS services obtain almost no ability to gather information of evidence. In other models such as IaaS and PaaS this situation slightly changes. In addition, this is aggravated by the fact that real network and router logs cannot be gathered by the customer for forensic purposes. And even if these data is given to the customer, the difficulty of putting all the evidential data in the correct context still exists [13].

Due to these facts, interruptions in the time-series of the forensic observations, also referred to as missing observations, may occur [15]. Missing observations, as a specific definition for uncertainty, represent an important issue in the discussion

of cloud forensics. In most of the cases, they are the reason for the complexity of proving a hypothesis during an investigation. However, properly speaking, the issue of missing evidence data exists even in ordinary digital and non-digital forensic investigations [14]. They always lead to further problems during the investigation phases because pieces of the whole incident story are missing. From a theoretical point of view, this states a paradox due to the fact that the cloud offers a huge amount of potential forensic data sources.

#### *Suggested Solution:*

CSP should profit from the fact that plenty of evidence data is available in current cloud environments. Network, process and access logs should be provided to customers over a specific read-only API which leads to the fact that customers obtain an improved ability to remodel interruptions in the time-series of potential future investigations. Fortunately, this approach could be applied to all three service models in order to verify and monitor specific actions and processes of services, applications or instances but in some scenarios, tensions between privacy and gathering forensic evidence is caused, since the latter produces detailed records of virtual machines, customers and corresponding user accounts. Furthermore, investigators should be able to handle data evidence from multiple sources which still states a problem for the research community [13]. Digital investigation always postulate the correlation of different sources of forensic evidence for ensuring better results of the investigation. *Data Fusion* methods for collection and correlation of evidence data could be a possible solution to this problem [39].

*c) Compliance Issues:* Companies are forced to stick to various regulations and rules for being able to take part in the global market. This situation gets even more complicated in cloud environments given the dynamic nature of the different service models. Especially in the field of credit card processing, the *Payment Card Industry Data Security Standard (PCI DSS)* [42] as a worldwide information security standard defined by the Payment Card Industry Security Standards Council was established to help the payment card industry to prevent credit card fraud through increased controls around data.

One of the requirements given by the PCI DSS is the strict detachment between the systems processing credit card data and other systems. This means, only one primary function per server should be implemented for preventing functions that require different security levels from co-existing on the same server. In public cloud offers, the implementation of such a requirement is not straightforward due to the multi-tenant host systems. Moreover, until the release of v2.0 of the PCI standard, it was unclear if the VM or the physical hardware is meant to be the system component. In the latest release v2.0 of the PCI standard, the PCI council clarified that system components also include any virtualization components such as virtual machines, virtual switches/routers, virtual

appliances, virtual applications/desktops, and hypervisors. In case virtualization technologies are used, each VM is only allowed to host one specific function. Unfortunately, although the current version of PCI DSS tries to discuss the impact of highly virtualized infrastructures, it is still unclear how the requirements shall be realized in public cloud environments.

The latest release v2.0 of the PCI standard also postulates convenient logging mechanisms and the ability to track user activities on machines computing credit card data. These mechanisms are critical in preventing, detecting, or minimizing the impact of a data compromise. Furthermore, the PCI council stresses that the presence of logs in all environments allows thorough tracking, alerting, and analysis when something goes wrong. In case of a forensic investigation, determining the cause of a compromise is very difficult, if not impossible, without system activity logs. As mentioned before, the customer can only access a limited amount of evidence data and will probably face missing observations in the forensic time-series of the investigation. Hence, achieving compliance with standards like PCI DSS will hardly be possible in public cloud offers which is also caused by the fact that CSP offer from none to generic audit reports instead of answering to the specific policy and compliance requirements of single customers.

*Suggested Solution:*

The solution to leave the cloud and search for an alternative compliant non-cloud service provider is eventually feasible for customers but often comes along with higher costs for the offered service. Hence, another possibility is to force the CSP to adopt to the customer's compliance requirements but as expected, this solution will hardly be accepted by the CSP. Therefore, customers should check their compliance requirements before moving data and processes to the cloud and also figure out, which CSP offers the best service according to the specific customer needs. In return, CSP should offer as much transparency as possible for simplifying these customer steps. In order to verify deployed compliance agreements, a Third Party Auditor (TPA) could be used acting as a trustee between the customer and the CSP. These potential solutions count also for other standards such as the Health Insurance Portability and Accountability Act (HIPAA), ISO2700x, Sarbanes-Oxley Act (SOX) etc.

*d) Secure Data Deletion:* File deletion is all about control and states an intricate problem for customers since 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 data stored in the cloud has been deleted exhaustively. However, the function of exhaustive deletion of data is an important supposition for the storage and processing of sensitive assets. Methods and functions of digital forensics could not only be used by valid investigators but also by malicious intruders for retrieving sensitive data out of compromised cloud instances and applications. Hence, the deletion of evidence in cloud environments is of greater

importance for the customer, the CSP as well as for the adversary.

Unfortunately, several questions still remain unanswered by global acting CSP: How can the customer be sure, that e.g. an email in a SaaS scenario or sensitive data on a VM have substantially been deleted by the CSP and how does he warranty that no traces (e.g. meta-data etc.) of the original asset is still stored?

*Suggested Solution:*

In traditional IT environments, various techniques have been developed to counter *Data Remanence* [28], [29]. In distributed cloud environments, deletion of data is mainly in the hands of the CSP. While this issue can be easily solved in encrypted storage scenarios by throwing away the key, secure deletion of unencrypted data in processing scenarios still remains a huge problem which cannot be solved without the help of the CSP. Generally, one solution could be the usage of a TPA [30] which evaluates the quality of an offered deletion service. However, this still remains an unsatisfying situation for the customers albeit providing an ability to verify the status of dynamic data stored in the cloud on behalf of the cloud client.

Recently, the aspects of so called *Trusted Cloud Computing* have been focused by researchers, combining the notion of *Trusted Computing* with cloud computing [21]. The major aim of such an approach is to provide the customer with the ability to verify confidentiality and integrity of their data and computation. Theoretically, the Trusted Platform Module (TPM) can provide hardware-based verification of hypervisor and integrity of the virtual instance running. With the help of such methods, trusted log files and trusted deletion of data could be theoretically provided to the customer. However, effective and practice-oriented results [27], [31] are still pending.

*e) SLA Verification:* An important sub-aspect of cloud forensics is obtaining evidence of system states concerning the cloud itself. CSP make generous concessions to their customers in their SLAs or other forms of contracts. Considering the distribution of control between CSP and customer, it becomes apparent that it remains almost impossible for the customer to verify the actual performance of these agreements. 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? How could he prove before a court that it was violated?

CSP can be bound to a specific SLA which assures e.g. data availability or backup procedures to the costumers. However, it could be possible that with ordinary cloud environments the SLA assignments cannot be fulfilled. This means, customer data have to be misplaced without the customer's knowledge for fulfilling the SLA assignments. Due to the fact that the customer has no way of knowing where his data is actually located, the CSP can prevent penalties for breach of contract by

resetting the data to another location without the customer's knowledge. This action could cause huge risks to the privacy and the security of customer's data.

Finally, the cloud service used by the customer is just another part of his huge infrastructure chain which he has to trust. Problematically, the different relationships cannot be secured by one overlapping SLA, but require several SLA one for each relationship. For the customer, this situation is not satisfying because the efficient execution of his business depends on too many attributes.

#### *Suggested Solution:*

Third-party auditing is currently considered to be one potential solutions to this problem. This means, a trusted and approved external party audits the security measures provided by the CSP. This is a first step into the right direction, but does not consider specific requirements given by single customers. Cloud customers demand the ability to run their own security audits, ensure that proper security measures are always in place and be able to control their security policies inside their own private cloud [32]. Without further concessions concerning transparency by the CSP, this issue cannot be solved.

*f) Missing Best Practices:* In the context of digital investigations, incidents are discrete computer events that are deterministic in nature and have a temporal causal sequence. This fact is given in ordinary environments as well as in cloud environments. From a theoretical point of view this means that each investigation can be solved if the needed sources and resources are available. Unfortunately, these perfect circumstances are not given in real world scenarios due to the previous discussed problems.

For example, it should be emphasized that evidence data must remain unchanged and the investigator must be competent and later able to give testimony, explaining relevance and implications of all actions. Furthermore, strict logs and records have to be kept for all steps of the investigation. In cloud environments, this is difficult to handle: Evidence data is located in different locations under various controls and hence, the chain of custody is difficult to preserve.

Aside from the mentioned technical issues, the question on how a digital investigation should be conducted in order to maximize the probative value (i.e. credibility) of the evidence. Digital evidence acquisition and analysis have to evolve along with the technology that is their subject. Currently, guidelines and best practice guides on gathering digital evidence are rare and often outdated. There are no guidelines specific to evidence gathered in the cloud, not to speak of precedent that could define legal requirements on data retrieval, handling and storage.

#### *Suggested Solution:*

In order to assure compliance with the issues mentioned above, strict preservation of a chain of custody is essential. It is important that the investigator can prove who created the

data (authenticity) and that the data retrieved was exactly the data supplied by the originator (integrity). Thus, especially for cases which are brought to court it is crucial that the chain of custody is preserved. Establishing a chain of custody means that an investigator can bring uninterrupted proof in form of documentation about who had control over the evidence between collection and presentation in court. In cloud computing environments, this is challenging. In order to be of use in trail, the evidence data goes on a journey from the crime scene, mostly located at the CSP, to court in a validated and secure manner. The question is whether it can be assured that this chain has not been contaminated. For example, a snapshot of a VM that could be used as digital evidence in front of a court, is created under the supervision of the hypervisor run by the CSP. How can this hypervisor be trusted?

## V. CONCLUSION

There is no doubt that cloud computing has various security benefits for SMB 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. It is a fact, that security incidents in cloud environments cross boundaries of responsibility and access and hence, 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, without control and accountability for the customers, this is not possible anymore.

Within this work, we introduced the fundamental issues of forensic investigations in cloud environments. We outlined the current challenges and proposed various potential solutions. Basically, the issues discussed within this work are mainly caused by one reason: The absence of global standards in cloud computing environments 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. The introduction of global standards for processing as well as storing data in cloud environments would simplify potential investigations tremendously. Investigators would know where they have to look for potential evidence and processes of handling evidence would be clear and transparent. However, when new standards or adjustments to existing standards are needed, as it is the case with cloud computing, creating too many standards inhibiting innovation should be avoided.

## REFERENCES

- [1] N. Beebe, *Digital Forensic Research: The Good, the Bad and the Unaddressed*, Advances in Digital Forensics V, 2009
- [2] D. Barrett and G. Kipper, *Virtualization and Forensics: A Digital Forensic Investigator's Guide to Virtual Environments*, Syngress, 2010
- [3] R. Bares, *Hiding in a virtual world: using unconventionally installed operating systems*, in Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics (ISI'09), 2009

- [4] R. Meadows, *Cisco Router and Switch Forensics: Investigating and Analyzing Malicious Network Activity*, Elsevier Science, 2009
- [5] EC-Council, *Computer Forensics: Investigating Network Intrusions and Cyber Crime*, Ec-Council Press Series, 2009
- [6] V. Corey, C. Peterman, S. Shearin, M.S. Greenberg and J. Van Bokkelen, *Network Forensics Analysis*, IEEE Internet Computing Journal, 2002
- [7] M. Pereira, *Forensic Analysis of the Firefox 3 Internet History and Recovery of Deleted SQLite Records*, Digital Investigation Journal, 2009
- [8] L. Rongxing, L. Xiaodong, L. Xiaohui and S. Sherman, *Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing*, in Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS '10), 2010
- [9] L. Kaufman, *Data Security in the World of Cloud Computing*, IEEE Security and Privacy Journal, IEEE Educational Activities Department, 2009
- [10] K. Muniswamy-Reddy and M. Seltzer, *Provenance as First Class Cloud Data*, ACM SIGOPS Operating Systems Review, 2010
- [11] A. Reyes, R. Britton, K. O'Shea and J. Steele, *Cyber Crime Investigations: Bridging the Gaps Between Security Professionals, Law Enforcement, and Prosecutors*, Syngress, 2007
- [12] 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 Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS '09), 2009
- [13] A. Case, A. Cristina, L. Marziale, G. Richard and V. Roussev, *FACE: Automated Digital Evidence Discovery and Correlation*, in Proceedings of the Eighth Annual DFRWS Conference, 2008
- [14] A. Patcha and J.-M. Park, *Network Anomaly Detection with Incomplete Audit Data*, Computer Networks Journal, 2007
- [15] E. Eleazar, *Anomaly Detection over Noisy Data Using Learned Probability Distributions*, in Proceedings of the Seventeenth International Conference on Machine Learning (ICML '00), 2000
- [16] P. Mell and T. Grance, *The NIST Definition of Cloud Computing*, Version 15, 2009
- [17] A. Barth, A. Porter Felt, P. Saxena and A. Boodman *Protecting Browsers from Extension Vulnerabilities*, in Proceedings of the 17th Network and Distributed System Security Symposium (NDSS), 2010
- [18] B. Adida, A. Barth and C. Jackson, *Rootkits for JavaScript Environments*, in Proceedings of the 3rd USENIX Conference on Offensive Technologies, 2009
- [19] G. Aggrawal, E. Bursztein, C. Jackson and D. Boneh, *An Analysis of Private Browsing Modes in Modern Browsers*, in Proceedings of 19th Usenix Security Symposium, 2010
- [20] D. Bem *Virtual Machine for Computer Forensics - the Open Source Perspective*, Open Source Software for Digital Forensics, Springer, 2010
- [21] N. Santos, K. P. Gumjadi and R. Rodrigues, *Towards Trusted Cloud Computing*, in Proceedings of the 2009 Conference on Hot Topics in Cloud Computing (HotCloud'09), 2009
- [22] A. Singhal, M. Gunestas, A. Singhal, D. Wijesekara and D. Gallagher, *Forensics Web Services (FWS)*, NIST Interagency Report Draft, 2010
- [23] Y. Shi, K. Zhang and Q. Li, *A New Data Integrity Verification Mechanism for SaaS*, Web Information Systems and Mining, Spinger LNCS, 2010
- [24] A. Juels and B. Kaliski, *PORs: Proofs of Retrievability for Large Files*, in Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07), 2007
- [25] B. Hay and K. Nance, *Forensics Examination of Volatile System Data using Virtual Introspection*, ACM SIGOPS Operating Systems Review, 2008
- [26] D. Nurmi, R. Wolski, C. Grzegorzcyk, G. Obertelli, S. Soman, L. Youssef and D. Zagorodnov, *The Eucalyptus Open-Source Cloud-Computing System*, in Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID '09), 2009
- [27] Y. Tang, P. Lee, J. Lui and R. Perlman, *FADE: Secure Overlay Cloud Storage with File Assured Deletion*, SecureComm, 2010
- [28] P. Gutmann, *Secure Deletion of Data from Magnetic and Solid-State Memory*, in Proceedings of the 6th Conference on USENIX Security Symposium, Focusing on Applications of Cryptography, 1996
- [29] C. Wright, D. Kleiman and S. Sundhar, *Overwriting Hard Drive Data: The Great Wiping Controversy*, Information Systems Security, LNCS Springer, 2008
- [30] Q. Wang, C. Wang, J. Li, K. Ren and W. Lou, *Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing*, Information Systems Security, LNCS Springer, 2008
- [31] R. Geambasu, T. Kohno, A. Levy and H. Levy, *Vanish: Increasing Data Privacy with Self-Destructing Data*, in Proceedings of the 18th Conference on USENIX Security Symposium, 2009
- [32] R. Chow, P. Golle, M. Jakobsson, R. Shi, J. Staddon, R. Masuoka, and J. Molina, *Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control*, in Proceedings of the 2009 ACM Cloud Computing Security Workshop (CCSW '09), 2009
- [33] L. Lamport, R. Shostak and M. Pease, *The Byzantine Generals Problem*, ACM Transactions on Programming Languages and Systems Journal, 1982
- [34] A. Haeberlen, *A Case for the Accountable Cloud*, in Proceedings of the 3rd ACM SIGOPS International Workshop on Large-Scale Distributed Systems and Middleware (LADIS'09), 2009
- [35] Y. Zhang, A. Juels, A. Oprea and M. Reiter, *HomeAlone: Co-Residency Detection in the Cloud via Side-Channel Analysis*, to be published at Security and Privacy IEEE Symposium, 2011
- [36] B. Grobauer and T. Schreck, *Towards Incident Handling in the Cloud: Challenges and Approaches*, in Proceedings of the 2010 ACM Cloud Computing Security Workshop (CCSW '10), 2010
- [37] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, D. and Song, *Provable Data Possession at Untrusted Stores*, in Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07), 2007
- [38] S. Mehrotra, *Introducing Windows Azure Diagnostics*, <http://blogs.msdn.com/b/sumitm/archive/2009/11/18/introducing-windows-azure-diagnostics.aspx>, 2009
- [39] S. Satpathy, S. Pradhan and B. Ray, *A Digital Investigation Tool based on Data Fusion in Management of Cyber Security Systems*, International Journal of Information Technology and Knowledge Management, 2010
- [40] S.D. Wolthusen, *Overcast: Forensic Discovery in Cloud Environments*, Fifth International Conference on IT Security Incident Management and IT Forensics (IMF '09), 2009
- [41] B. Hay, K. Nance and M. Bishop, *Live Analysis: Progress and Challenges*, IEEE Security & Privacy Journal, 2009
- [42] Payment Card Industry Data Security Standard (PCI DSS), [https://www.pcisecuritystandards.org/security\\_standards/index.php](https://www.pcisecuritystandards.org/security_standards/index.php)