Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

A survey on issues in cloud forensics with an experiment on time consumption

K. H. A. Hettige [1, *], M. S. D. Fernando [2]

[1]*Department of Computing, Rajarata University of SriLanka, Anuradhapura, Sri Lanka*
[2]*Department of Computer Science and Engineering, University of Moratuwa, Moratuwa, Sri Lanka*

ARTICLE INFO

Article history:
Received 3 August 2021
Received in revised form
12 November 2021
Accepted 19 December 2021

Keywords:
Digital forensics
Cloud forensics
Most common issues
Time complexity
Autopsy tool

ABSTRACT

Forensic investigations on cloud platforms are an oft-discussed topic in current digital forensics. Significant growth in cloud platforms is expected in the coming decade. With such growth, cloud forensic investigations may require substantial changes in their approach. The paper surveys the most mentioned issues in cloud forensic literature. It is followed by a description of some of our current work aimed at solving those issues. The first issue that we tried to analyze was the issue of the trustworthiness of the evidence. We identified that the trustworthiness of the Cloud Service Providers is hardly discussed in the literature. Based on previous publications on similar issues on standalone computers, we provided an algorithm as an initial answer to the issue. The algorithm checks for the integrity of the evidence which will be affected in a tampering attempt. The next issue that we considered was time-taken for analysis (time complexity of forensic tools). While the issue has been indicated many times in the literature, we did not find many detailed experiments conducted with tools to observe the processing time over data source size. Therefore, the paper includes the results of an experiment that was performed using an Autopsy forensic tool to measure the time complexity of its operation with a number of source files with increasing sizes. Results indicated that the analyzing times usually increased with the size of the source file and that it might become unmanageable with increasing sizes.

1. Introduction

The idea of cloud computing emerged in the 1950s or 1960s, with one Berkeley (University of California at Berkeley) publication identifying cloud computing as a new term for the long-held dream of computing as a utility, which later emerged as a commercial reality. However, it took some decades before the technology was advanced enough to be used in a meaningful manner.

When looking for the term Cloud Computing, one needs to be careful since it seems to have become a marketing term rather than a term that describes a specific technology (Martini and Choo, 2012; Simmon, 2018). To complicate the matter even further, the technology changes faster than the definitions and standards created. Thus the result has been multiple standards that have certain similarities and technologies that do not match standards. However, despite not having a proper definition or being a mature technology, cloud computing has a large user base that ranges from private individuals to large business organizations. Gartner has predicted that $ 677 Billion will be spent worldwide on cloud services from 2013 through 2016. All cloud services are expected to have significant growth in the upcoming decade (Martini and Choo, 2012).

Various organizations have created different definitions for Cloud Computing which have different acceptance levels. According to the definition produced by the United States (US) government's National Institute of Standards and Technology (NIST), a Cloud should have five essential characteristics.

- The consumer can provision resources on-demand automatically without human intervention from the service provider.
- Broad network access.

* Corresponding Author.
Email Address: akalanka@as.rjt.ac.lk (K. H. A. Hettige)

https://doi.org/10.21833/ijaas.2022.03.003
Corresponding author's ORCID profile:
https://orcid.org/0000-0002-8471-0821
2313-626X/© 2022 The Authors. Published by IASE.

- Provider pools resources that serve multiple clients.
- Capabilities (resources) can be elastically provisioned or released with demand.
- Resource usage is measured, monitored, controlled, and reported.

NIST also identifies three service models as Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), and Cloud Infrastructure as a Service (IaaS). It also identifies four deployment models as Private Cloud, Community Cloud, Public Cloud, and Hybrid Cloud (Liu et al., 2011).

Many authors agree or follow the NIST guidelines, and most, at a minimum, agree with the above-given characteristics (Martini and Choo, 2012). However, similar to any cutting-edge technology, cloud computing has its fair share of concerns. For example, one concern related to cloud computing is the degree of support for the forensic investigation process in situations where cloud services are misused. In addition, due to the various increases in the usage of cloud computing explained above, the complexity of the forensic process is expected to increase significantly (Reilly et al., 2011; Zawoad and Hasan, 2013).

This paper aims to explore the issues in cloud forensics and analyze these issues. The remainder of the paper is organized into the following sections: Forensic process, cloud forensics problems overview with selected publications, a summary of the most frequently identified issues, an analysis of the issues, and our approach and our current work. It includes an experiment that we ran with a forensic tool. Finally, we conclude the paper with our expected future work.

2. Forensic process

The 1980s saw the introduction of the earliest digital forensic models, and the number of digital forensic models has increased significantly since the 2000s. Digital forensic processes usually consist of identifying, labeling, recording, obtaining, isolating, examining, analyzing data/evidence, and reporting/presenting the conclusions. There might be differences in how each step is performed from one digital forensic process model to another, but there are many common steps in most models (Pollitt, 2007; Yusoff et al., 2011).

Due to the number of digital forensic process models available, this paper will use the NIST's guidelines to explain forensic activities without resorting to one specific process model. NIST guidelines themselves do not represent a silver bullet solution to forensic investigations. However, it provides instructions independent of any particular forensic process model to be used in digital forensic investigations. In a way, we can say that the NIST model encompasses many other models and treats each step in equal detail (Kent et al., 2006; Pollitt, 2007).

NIST defines four basic steps in a forensic process, namely: Collection, Examination, Analysis, and Reporting (Kent et al., 2006). Out of these three steps, the Collection step requires special attention in cloud computing due to the volatile nature of cloud platforms.

3. Cloud forensics problems overview

Considering cloud computing environments, it has been proposed that traditional digital forensic models and techniques might not be highly suited for usage in cloud computing environments (Dykstra and Sherman, 2011; Reilly et al., 2011; Grispos et al., 2012; Martini and Choo, 2012; Zawoad and Hasan, 2013). These authors have identified a number of challenges or potential problems and have indicated solutions for them. Out of them, certain challenges are more legal in their nature. In our study, those will be excluded, and the focus will be on the problems related to the technical process.

3.1. Selected publications

In their paper, Reilly et al. (2011) considered the advantages and disadvantages of cloud computing on digital forensics. As for disadvantages, many technical and legal challenges are discussed, and among them, the challenges which are relevant to this paper are as follows (Reilly et al., 2011):

- Identifying evidence locations and acquiring them: It is almost impossible for outside investigators to access the cloud hardware physically and to acquire the data. Therefore, the process did not conform to existing guidelines. Which in turn might lead to the admissibility of the evidence.
- Reliability of evidence received as Virtual Machines (VM) instead of a bit-wise copy: The authors assessed that VM data could be altered if they are booted again during the investigation.
- The difficulty of finding all the evidence hinders building a sequence of events and completing a timeline: The Authors noted that certain data such as registry entries could not be obtained from the cloud.
- Lack of proper software tools for the forensic process on clouds

Dykstra and Sherman (2011) described a number of issues faced by cloud computing through two case studies. These case studies consider the potential behavior of criminals who use clouds to commit crimes, the actions of the investigators, and the problems. The first study was concerned with a scenario where a crime is committed using a cloud platform, while the 2nd study discussed a scenario where criminals target the cloud platform. Some of the chosen issues have been listed below (Dykstra and Sherman, 2011):

- Time taken for analysis (even for an automated tool) will be considerable. Even in traditional

situations, it has been estimated that 2 Terabytes (TB) of data takes 85 hours of processing time: With the sizes and number of investigations, the workload will be increased for investigators.
- When data is from multiple physical locations, certain metadata such as timestamps and IP addresses in each one of them will be different: This will cause problems when creating a timeline of the events.
- Current tools are inadequate to handle the above situations.
- How privacy, access control, and evidence collection will be affected when conducting evidence collection in a cloud platform: In traditional forensics, usually, the whole device will be isolated. In a cloud server, if we isolate the device, that might include VMs that belong to other clients.
- Questions regarding Cloud Service Provider (CSP) trustworthiness and chain-of-custody of the evidence will be raised: Most of the time, the evidence collections will be done by an employee of a CSP who will not be under the supervision of investigators.

Martini and Choo (2012) pointed out multiple observations by various parties who express those current methods, guidelines, and processes are inadequate or outdated. It also identifies the need for more research in forensics on cloud platforms. The paper notes the following issues which exist in cloud computing platforms that would adversely affect the digital forensic process:

- Data in the cloud are often physically distributed among different servers and data centers. Hence the evidence might be distributed among thousands of servers.
- Due to the on-demand and virtualized nature of the cloud platforms, most potential forensic artifacts (deleted files, temporary files, etc.) might get overwritten. By its nature, cloud platforms quickly reuse any unused resources such as memory and storage spaces.
- Almost impossible to physically access the hardware and acquire the evidence. It would be unlikely to actually remove and preserve the hardware, as in the case of a standalone computer.
- Investigators will have to depend on the CSP for evidence gathering, and it would affect the existing chain-of-custody rules. Investigators will not have the power to verify the CSP's process used for the evidence acquisition.

The paper also suggests an iterative forensic framework for cloud computing that combines certain steps of NIST and McKemmish frameworks. In the proposed framework, the first three steps have been given emphasis, and they can iterate multiple times until all the necessary evidence has been identified satisfactorily:

- Evidence source identification and preservation of evidence sources.
- Collection of evidence.
- Examination and analysis of evidence.
- Reporting and presentation.

Iteration has been introduced considering the difficulty of identifying and preserving the evidence in a timely manner on cloud computing platforms. So authors note that all evidence sources may not be identified initially. If any evidence sources are identified during the collection and examination and analysis of evidence, then those evidence sources will also be identified, preserved, and collected for examination (Martini and Choo, 2012).

Grispos et al. (2012) have identified a number of challenges and solutions in forensics on cloud computing. These challenges were categorized under their relationship to the phases in a DFRWS Investigative process (DIP) model. A chosen set of challenges relevant to the paper has been given below:

- Distributed, virtualized, and volatile storage: This will make evidence identification and collection more difficult. Volatility will increase the possibility of evidence destruction and will make the evidence collection a time-sensitive matter. A delay would increase the chance of the destruction of evidence.
- Imaging all physical media in a cloud is impractical; partial imaging may face legal challenges: A single cloud server will contain many VMs, and imaging everything will take too much time and affect privacy laws. Imaging only a part of can lead to the charge that the investigation did not consider all scenarios.
- Evidence from multiple time zones will contain different timestamps, which will affect building a timeline of events.
- Acquisition of physical media from providers is cumbersome, arduous, and time-consuming and will have to depend on the cloud provider.
- Data is inherently volatile: Cloud platforms increase this volatility with transferring of VMs and data within different data centers. When a VM is deleted, the allocated memory space is reallocated automatically.
- Events may occur on many different platforms (traceability and event reconstruction issues due to incomplete data).

The paper further explains certain specific technical challenges which exist. Some of them which are related to the topic of this paper, have been listed down (Grispos et al., 2012):

- Virtualized data storages distribute the data into different physical servers. This decision is not based on the number of cloud consumers who are using the cloud system. This would cause the data to be distributed into multiple places, and identifying and preserving all the evidence would become very difficult.

- The amount of data that would be collected for analysis would also be very large considering the above distribution. Consider that in a traditional environment, the entire HDD (Hard Disk Drive) of a computer would be confiscated if there was suspicion that evidence exists, but this is not possible to do in a cloud computing environment.
- Once the data gets deleted by a user (or a cloud consumer) or when a cloud instance closes, then supposedly unused storage space will be reutilized and overwritten.
- When creating a timeline of the events, special consideration is required if different files existed in different time zones.

Zawoad and Hasan (2013) described the challenges, approaches, and open problems in cloud forensics. Among the observations they made by considering the available literature, the following ones were chosen as relevant for this paper (Zawoad and Hasan, 2013):

- Inability to physically access the evidence sources such as the servers. In situations where servers are distributed in different countries, local legislation may prevent foreign investigators from accessing them.
- Data which exist in the virtual machines (VM) are volatile in the sense that the data would be deleted if the VM shuts down.
- Crucial evidence such as logs is distributed both physically and logically. Even in a single physical server, different tiers and layers contain their own logs. All these logs need to be aggregated to get the full picture.
- Limitations in forensic tools will adversely affect the success of an investigation. Considering that the cloud involves vast amounts of data, tools support for investigations is a must.
- Investigators will have less control over the computer system and depend on the CSP for the evidence collection. Investigators also have to trust the CSP.

Quick and Choo (2014) surveyed existing literature to identify the impact of data volumes on digital forensics. Based on their data from 1999 to 2014, both the number of cases and the average data volume analyzed per case were on the increase. The authors also researched the performance of existing digital forensic tools and found a few. They concluded that investigators would have trouble with the ever-increasing data load. This research reveals an interesting issue regarding cloud forensics, even though the focus of this publication was not limited to cloud platforms (Quick and Choo, 2014).

Pichan et al. (2015) gave a comprehensive, detailed overview of cloud forensic problems which have been discussed throughout the years. After providing an initial background to the problem, they discuss the list of challenges, recommended solutions and include their observations as comments for each challenge and solution. It has to be noted that when listing the challenges, they separated them into eight categories based on the forensic investigation process (Pichan et al., 2015):

- The unknown physical location of evidence
- Control of the platform and evidence is in the control of the CSP
- CSP does not share all available evidence
- Dependence on CSP
- The volatility of data in the cloud
- Duplicate copies of evidence distributed in multiple servers
- Chain of custody issues
- Evidence isolation and collection

Vurukonda and Rao (2016) discussed a range of security issues in cloud computing when storing data. While the publication is not regarding forensics, it considers important points that are relevant to forensic issues. The issues related to cloud forensics are identified below:

- The threat of malicious insiders of the CSP/organization: These types of attacks are very difficult to avoid or investigate for the clients of the cloud platform. Authors note that most organizations simply ignore this risk.
- A cloud user might be able to recover data from another cloud user. This is mainly due to the resource reuse that happens with cloud platforms.

Quick and Choo (2018) introduced methods to handle big forensic data. In the background information, they note how computing resources have become cheaper, thus contributing to the growth of digital media. Using data from the US Federal Bureau of Investigations (FBI), they noted how the number of digital forensic cases, the amount of data to be processed in a single investigation, and the total number of data processed during a year has increased throughout the years. The authors also noted that with growing data loads, forensic imaging and processing times are also increasing. The authors noted that this would increase the time of an investigation, and that will adversely affect the investigations and involved persons/organizations (Quick and Choo, 2018).

Manral et al. (2019) conducted an extensive survey over the published cloud forensic literature. After considering over four hundred and fifty publications, they identified eighty publications that were about cloud forensics. The authors proposed a "cloud computing taxonomy" and systematically categorized the available research. Authors categorized existing research into four categories and identified CSP's control over the hardware infrastructure and consumers' degree of control over the working environment as the most frequently identified challenges (see Table 1 for the summary of issues). It should be noted that some of the issues they identified are not really technical issues of cloud forensics. For example, the authors found the validity of the available forensic solutions as an issue. They felt many solutions require implementation and testing to validate. They also found that some publications focus too much on a particular cloud model or an application making such solutions difficult for other cloud models or applications. The authors also identified the different forensic artifacts with their locations in the cloud stack and identified which cloud stakeholder has access to each of those artifacts. From their findings, the CSPs generally have the most control over the forensic artifacts in the cloud.

Table 1: Most frequently mentioned issues by Manral et al. (2019)

Issue	Frequency
CSPs' control over the hardware infrastructure	Most mentioned challenge
Consumers' degree of control over the working environment	Most mentioned challenge
Focus on particular applications/model	Twelve times
Lack of implementation and evaluation	Seven times
CSP trust/support	Six times
Security weaknesses	Five times

3.2. Summary of issues

The publication by Manral et al. (2019) covered many cloud forensic publications between 2007 and 2018. But it includes publications early as 2003 and late as 2019. Therefore, many of the publications that we discussed were also covered by Manral et al. (2019) in their survey. However, Dykstra and Sherman (2011) and Vurukonda and Rao (2016) were the exceptions. The summaries of the issues mentioned by them can be found in Table 2. From the Trust on CSP/CSP control, Time taken for analysis/data load and Privacy and access control are the issues most mentioned.

Table 2: Overview of selected publications

	Reilly et al. (2011)	Dykstra and Sherman (2011)	Vurukonda and Rao (2016)	Quick and Choo (2018)
Evidence access issues	yes			
Trust on CSP/CSP control		yes	yes	
Multi-tenancy		yes		
Data distribution		yes		
Time taken for analysis/data load (time complexity)		yes		yes
Tool adequacy	yes	yes		
Privacy and access control		yes	yes	
Evidence reliability	yes			
Consumer/investigators control over evidence	yes			

Table 3 provides a summary of the issues from the publications which were already covered by Manral et al. (2019). According to our analysis, Trust on CSP, Dependence on CSP, Data distribution, and Volatile data are the most frequently mentioned issues in those publications.

Considering all three tables, we can identify some of the most frequently mentioned issues in cloud forensics. In the next section, we will go through these issues identifying and clarifying each one. Some issues might be connected to each other, while others might be unavoidable technical issues that cloud platforms have.

Table 3: Overview of already covered publications

	Martini and Choo (2012)	Grispos et al. (2012)	Zawoad and Hasan (2013)	Quick and Choo (2014)	Pichan et al. (2015)
Trust on CSP	yes	yes	yes		yes
Multi-tenancy		yes	yes		
Data distribution	yes	yes	yes		yes
Jurisdiction issues	yes		yes		yes
Dependence on CSP	yes	yes	yes		yes
Physically inaccessible	yes	yes	yes		
Large data sizes		yes	yes	yes	
Volatile data	yes	yes	yes		yes
Heterogeneous logs		yes	yes		yes
Chain of custody	yes		yes		
Tool limitation			yes		yes

3.3. Issues in details

Here we will describe the most frequently mentioned issues in detail. Sometimes, certain issues will be discussed together, considering their interconnection. The most frequently mentioned issues will be discussed first.

1. CSPs' control over the hardware infrastructure: In a cloud platform, the CSP has control over all

hardware and the software infrastructure. This is an unavoidable technical challenge in cloud platforms that do not provide the cloud clients and investigators with a lot of options. This control is clearly described in the NIST cloud computing reference architecture. As indicated in Fig. 1, the service provider (CSP) stays at the bottom of the software layers exerting its control over the whole system. Considering the software and hardware components involved, the CSP has access to the hardware on which the whole system runs and to the underlying virtualization software (indicated by Resource abstraction and Control layer in Fig. 2) on which the individual VMs operate (Liu et al., 2011). Cloud clients or users do not have any access or privilege over the Resource abstraction and Control layer and layer below that (shown in Fig. 2) in any cloud model (SaaS, PaaS, or IaaS). This clearly shows the CSPs' control over the whole cloud platform, including the VMs and the data in them. As a practical deterrence against this overall control, techniques such as data encryption can be used to protect the data against unauthorized access by CSP and agreements can be signed with CSP to make sure that CSPs will respect the client's privacy. But it is very difficult to detect any passive surveillance by the CSP on a cloud client or user.

2. Consumer's or investigator's degree of control over the working environment/Physical inaccessibility/Chain of custody: the cloud consumers or investigators have very limited access to the overall cloud platform. Cloud platforms might be located in foreign countries, and even law enforcement authorities do not get physical access to the cloud hardware. Traditional hardware preservation and acquisition cannot be made in the cloud environment. Manral et al. (2019) emphasized this fact through a detailed description of forensic resources in the cloud platform and the access each stakeholder has over those resources. Some authors have introduced models to increase the access privileges of clients over the cloud platform, but these models require at least the initial cooperation of the CSP.
3. CSP trust/support/dependence: As indicated above, the CSP has full access to the data in a cloud platform. Therefore, clients and investigators have to place their trust in the CSP that no malicious activity originates within the CSP organization (malicious insiders). When gathering evidence, investigators have to depend on the CSP for the collection of certain evidence, and investigators do not have the ability to supervise these activities. Therefore, have to trust the forensic process of the CSP. Vurukonda and Rao (2016) indicated the extreme difficulty of remedying against malicious insiders.
4. Focus on particular applications/models: The authors noted that many cloud forensics solutions focus on a particular cloud model or certain application. This would imply that such solutions might not suit well for other models or applications. While this might be problematic, it might also be unavoidable. Cloud computing is a very vast field with different models and applications. Depending on the model and application, the forensic process that will be followed have to be different. For example, in the SaaS model, the cloud clients will have very little access over the cloud platform, while in the IaaS model, the clients will have much higher accessibility over the cloud platform. Access level will affect the suspect pool for an incident, and it would also determine the amount of evidence that investigators can gather independent of CSP aid. Therefore, it might not be practical to create universally valid solutions across all cloud models and applications within a limited timeframe (all software projects-including attempts to develop forensic solutions-have deadlines). Also, creating universal cloud forensic solutions might result in a generic solution that is not practically usable.
5. Lack of implementation and evaluation of solutions/Tool inadequacy or limitations: Authors surveyed a number of publications that presented solutions that were either not implemented or properly evaluated or both. Without proper implementation and evaluation, solutions cannot be practically used for investigations. This issue inadvertently contributes to the limitations in available tools. While there is only a small number of tools available, the challenges in forensics keep growing. As indicated above, these challenges might be technical, as in the case of data loads. Some might be cultural (or otherwise), as in the case of published concepts without proper implementation or evaluation.
6. Security weaknesses: Authors identified various security weaknesses in the published solutions. With security weaknesses, solutions might lead to attacks on the cloud platform if the solution requires a modification to the cloud platform for better access for investigations. In addition, a solution with a weakness might be exploited by an attacker. If the weaknesses become known, it will raise questions about the validity of the investigation.
7. Time taken for analysis/data loads: The time taken for forensic analysis has been growing over time. In recent times, the sizes of electronic storage media (mainly HDDs) have increased while the price has gone down. This has resulted in more and more being stored in electronic media. This has ultimately resulted in an average increase in the amount of data to be investigated during an investigation. According to the Federal Bureau of Investigations, certain cases required forensic analysis of up to 20 TB of data even back in 2007 (FBI, 2007). According to the data published by the FBI's Computer Analysis Response Team (CART) and the FBI's Regional Computer Forensic Laboratory (RCFL) Program

(which covers a network of 16 regional laboratories), the number of examinations, the total processed data loads, and the average data load per case is all on the rise (as shown in Table 4 and Table 5). These facts are further supported by the data from the Northumbria Police (England) that show gradual growth in data sizes (Irons and Lallie, 2014). With such evidence, it is prudent to assume that the amount of data that needs to be analyzed will continue to grow. As some authors noted, process time bottlenecks have already reached levels where they might hamper investigations and is expected only to grow worse. In addition to the time factor, large amounts of information might confuse or overwhelm the investigators (FBI, 2016; 2013; Quick and Choo, 2018).

8. Privacy and access control/Multi-tenancy: In a cloud platform, multiple client VM can be located inside a single physical server. If we isolate the whole server, it will affect other clients. If the whole server is imaged, then it would contain data from other clients. Additionally, attacks can be launched across VM, making it difficult to be detected from the client-side.
9. Data distribution: Authors assessed that the data of a large VM would be distributed over multiple hardware, and that would create issues when gathering such data. Also, since cloud platforms transfer the VM among different physical servers unbeknownst to the clients, this would spread the metadata related to the VM over many servers. However, in client-side forensics, the CSPs do not provide the metadata of the cloud platform to the cloud clients. Additionally, the underlying infrastructure usually does not affect the VM or the forensic image of the VM. However, during an investigation conducted by the CSP-side, information regarding data distribution would be used.
10. Volatile data: The clients have the ability to shut down their VM in the cloud. This will release the space on the HDD, and the cloud platform will reuse the released space on-demand. This will obviously destroy the evidence in the cloud. However, this scenario will only happen in situations where there is a malicious actor with administrative privileges over the VM. Additionally, another version of this issue is somewhat connected to the previous issue that we discussed. Data in the cloud is distributed, and metadata gets deleted frequently and automatically. Unlike in the traditional scenarios, investigators cannot obtain the HDD and retrieve any deleted data. This increases the volatility of the data in the cloud.

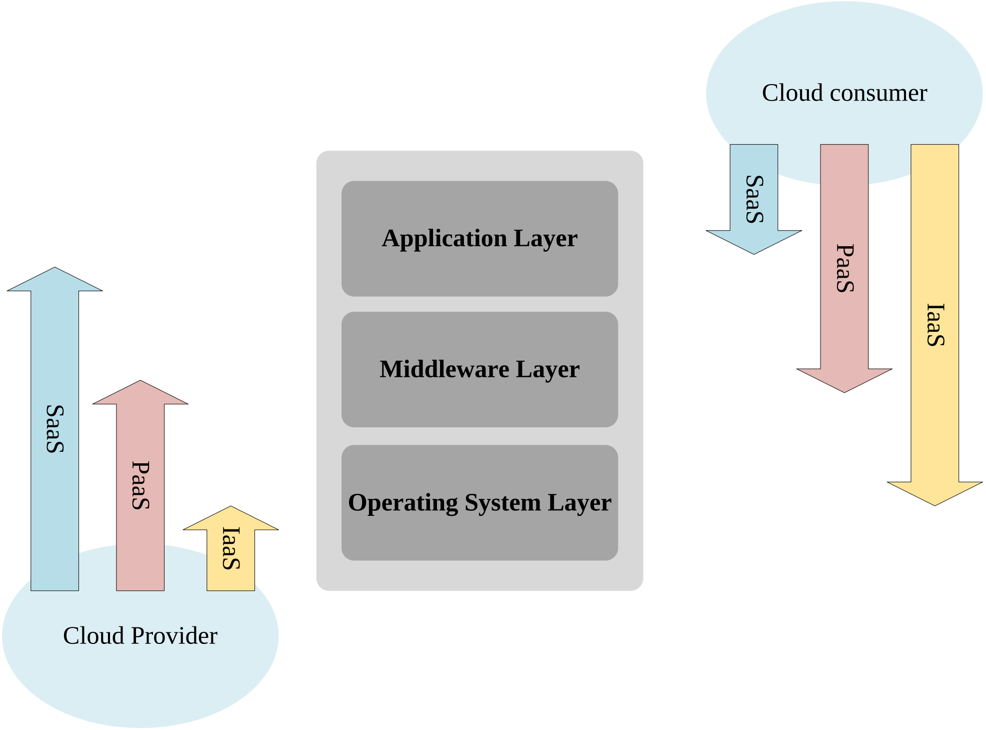

Fig. 1: Scope of control between cloud provider and consumer

Table 4: Comparison of the number of investigations and amount of data processed by the CART

Year	Number of forensic examinations	TB of data processed	Average TB per investigation
1999	2084	17	0.008
2000	3591	39	0.011
2001	5166	119	0.023
2002	5924	358	0.06
2003	6546	782	0.119
2012	10,400	10,500	1.001
2015	7,338	9,770	1.331

Table 5: Comparison of the number of investigations and amount of data processed by the RCFL laboratories

Year	Number of forensic examinations	TB of data processed	Average TB per investigation
2003	987	82	0.083
2004	1304	229	0.18
2005	2977	457	0.15
2006	3633	916	0.25
2007	4634	1288	0.28
2008	4524	1756	0.39
2009	6016	2334	0.39
2010	6564	3086	0.47
2011	7629	4263	0.56
2012	8566	5986	0.7
2013	7273	5973	0.82
2014	6322	5060	0.8
2015	5897	5276	0.895
2016	5229	5667	1.084

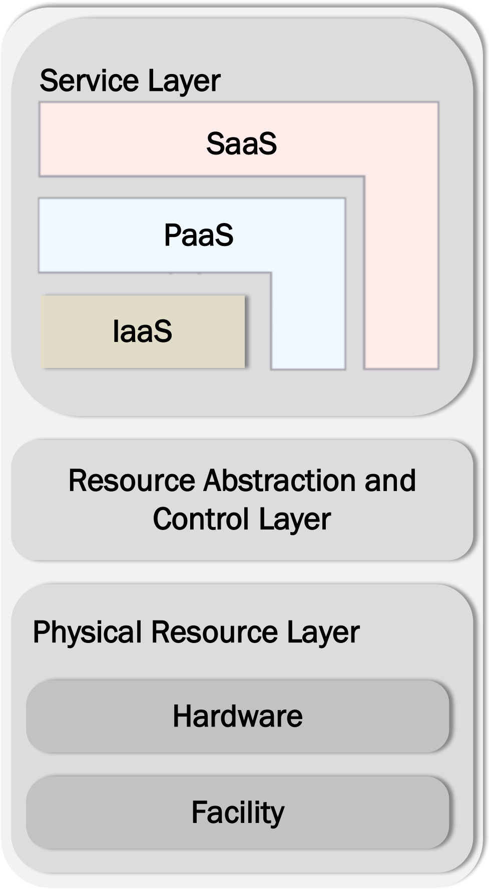

Fig. 2: The component stack of a cloud platform

3.4. Conclusion of issues

Here, we would conclude the discussion about cloud forensics with a few interesting points. Of these ten issues we discussed in length, we identified the following points:

- The first three issues are interconnected. They all arise from the technical nature of cloud platforms that gives more privileges to the administrators of the cloud hardware and hypervisors.
- The fourth one might be an issue, but it might not be solvable, as we discussed there.
- While fifth is a serious issue, solving it requires the solution providers (ones who develop cloud forensic solutions, including researchers) to simply finish their implementations and testing.
- The sixth issue, like the fifth issue, depends on the solution providers. With proper attention to software engineering best practices, it is an issue that can be largely avoided.
- The seventh issue is partly related to the fifth. With more and more valid solutions, the data load issue can be better mitigated.

4. Our approach

As part of our research, we have to explore some of the issues which were mentioned in the previous sections. From them, we would present our current work under two topics. First, we would discuss the first three issues that we discussed earlier. Then we would present you with an experiment we conducted related to the seventh issue.

4.1. CSP reliability assurance

Generally, during a cloud forensic investigation, there is evidence that can be gathered using client privileges. From the cloud service models, the IaaS model allows the most privileges to the client. CSPs provide Application Programming Interfaces (API) for the clients to use for data acquisition purposes (Roussev et al., 2016). Even in that case, the client can request image files of the VM, snapshot of the VM, etc., although the CSP might not provide any authentication data (Dykstra and Sherman, 2012). In both these methods, there is no exact method to ensure that the CSP has not altered the evidence. And despite agreements with the CSP, it would be better if there are ways to assure the integrity of the evidence considering the fact that even a malicious employee on the CSP side would be enough to alter the evidence. Considering the literature, we can identify two main ways to approach this. The first approach has existed for some time, while the second approach is still under research.

- Have some software installed on the Cloud service which will monitor all the activities, including the CSP. This software may generate some kind of a log that can also be provided to a third-party investigator to show that the CSP has acted transparently. Researchers have called this process "forensically enabling the cloud" (Ruan and Carthy, 2012). Many authors have proposed these types of readiness frameworks, but many of them do not identify the threat from the CSP side as a factor. In fact, we only found a single publication that tried to implement a solution that specifically targeted CSP side threats. Zhou et al. (2017) looked at the issue of the threat posed by malicious CSP, considering it as a Service Level Agreement (SLA) violation. They also propose a framework that alerts the client if the CSP is violating the SLA with regard to client privacy. However, in order to implement this framework, it requires the CSP's approval and technical support because the system works below the layers that clients have access to (Zhou et al., 2017).

- Detecting if the disk image has been manipulated: Literature says that creating a forensic disk image through manipulation is difficult, while not impossible. Most will leave some form of evidence, and there is no exact process to detect such evidence. In order to mask the manipulation of the disk image, anti-forensic tools are sometimes used. Detecting evidence for the usage of anti-forensic tools and rigorous checking of timestamps and other artifacts is advised to identify manipulations (Palmbach and Breitinger, 2020; Schneider et al., 2020). Rani and Geethakumari (2021) and Rani and Kumari (2016) discussed detecting anti-forensics in the cloud. Rani and Kumari (2016) called for an anti-forensics identification sub-phase in the evidence analysis phase of a digital forensic investigation, and Rani and Geethakumari (2021) called to identify anti-forensics through a packet analysis framework. However, in both cases, possible attackers are other VM owners or outside parties. Authors do not consider the CSP as a possible threat in their solutions (Rani and Geethakumari, 2021; Rani and Kumari, 2016). We identify another possible approach towards CSP reliability through the research conducted by Jang et al. (2016), Freiling and Hösch (2018), and Schneider et al. (2020), although their studies were not aimed at cloud platforms. Here, the authors researched a way to detect manipulation of various evidence sources. Among the studies, research conducted by Freiling and Hösch (2018) is quite interesting since the experiment was conducted on VM images created using Oracle VM VirtualBox. Therefore, we see the possibility of adapting this method on cloud platforms. One main advantage of this approach is that it does not require the cooperation of the CSP, so it does not depend on the CSP. Also, compared to the other solutions, this approach is a passive approach that does not require any "preparation" of the cloud platform to record incidents. (Freiling and Hösch, 2018; Jang et al., 2016; Palmbach and Breitinger, 2020; Schneider et al., 2020).

Given the fact that CSPs usually do not make modifications to their underlying system, first approach solutions (unless implemented by the CSPs themself) might face practical difficulties when implementing it. The second approach gives an independent solution, but we feel it requires more research on it to validate it to cloud platforms. The second approach does have its weaknesses. The detection effort is centered on finding mistakes made by the individual/s who created the forged evidence. A perfect forgery will go unnoticed during the investigation, and even the detection of an imperfect forgery will depend on the investigator as was shown in Freiling and Hösch (2018) and Schneider et al. (2020).

Considering the experiments that they conducted, failure to detect forgeries was due to human errors. The main reason for that is while there are software tools to extract various forensic data, human investigators are the ones who examine them and decide on the results. With a very large set of information, humans tend to make mistakes. So, one option would be to automate the analysis process to identify suspicious activities so that investigators can decide on them. Following is an example algorithm (in pseudo-code) derived from the publications to check the consistency of the timestamps.

The following algorithm checks for the integrity of files and their entries. The files here include log files, temp files, any DBMS (such as SQLite), and similar files, which usually contain different entries ordered sequentially. It checks whether there is any timestamp made after the evidence collection time. Such events might indicate a manipulation effort after evidence was collected.

```
time $t^{end}$ - the time when the evidence was collected from the cloud.

for each file in evidence source
{
    if( (last modified timestamp in metadata of a file) >= $t^{end}$ )
    {
        Possible integrity failure. Record the name of the current file.
    }
    if(entries are in chronological sequence/temporal order)
    {
        n=1
        for($1^{st}$ entry......$k$-$1^{th}$ entry)//k being the last entry
        {
            if(timestamp of $n^{th}$ entry < timestamp of $n+1^{th}$ entry || timestamp of $n+1^{th}$ entry <= $t^{end}$ )
                Correct;
                n++
            else
                Possible integrity failure. Record the n and n+1 entries as integrity failures.
                n++
        }
    }
    else
    {
        Parse through the data in the file for any timestamps $t^{any}$
        If($t^{any}$ > $t^{end}$)
            Possible integrity failure. Record the name of the file.
    }
}
```

For a complete forgery detection system, we need to develop algorithms to cover all possible cases of forgeries. Automating this forgery detection process would decrease the time taken and would also increase the accuracy and reliability of the process.

4.2. An experiment on handling different data loads on a forensic tool

Next, we would like to discuss an experiment we conducted to assess how a forensic tool works with large datasets and how much time is taken for analyzing the evidence. As surveyed by Quick and Choo (2018), increasing storage sizes are creating a problem for analysis. But we could not find many experiments where the performance of tools was measured against increasing source sizes. Therefore, we conducted an experiment to measure the analysis times over increasing source sizes. We choose the SleuthKit-Autopsy as the tool that we are going to use for forensic analysis. It's an advanced tool with proper user interfaces to visualize the results with the capability to extend its features through plugins. The tests were conducted on an ASUS laptop running Windows 10 with an Intel(R) Core(TM) i-7 CPU 1.8Ghz processor, 16GB RAM, and 128 SSD. For sample data, Virtual Hard Disks (VHD) created using Oracle VM VirtualBox with windows 7 installations were used. There were no other significant data within each VHD, so they were empty other than the Operating System.

The set of tests was performed with Autopsy 4.18 with high-performance settings on Windows. Even on battery power, the computer-operated under high-performance settings (results are shown in Table 6). While the completion time does not linearly increase with the size of the evidence, it showed an overall increase in the time taken.

Table 6: Autopsy experiments with different data sources

size of the evidence	Completion time (hr:mm)
20GB	14:19
60GB	14:00
100GB	20:47
150GB	22:09
200GB	21:44
250GB	22:40
300GB	19:10
400GB	34:55
500GB	33:10

Then a test was performed with VHDs, which contained data. We created a 40GB VHD (same as previous), but we added a 15 GB data bundle (containing video, audio, text, and other files) to the VHD. The Autopsy browser did not manage to finish its analysis in 60hrs when the experiment was terminated. The experiment on the same VHD was repeated with similar results. Next, we inserted the same data bundle into 300 GB VHD and repeated the experiment with Autopsy, and the analysis simply did not finish for 80+hrs, and the experiment was terminated. In all these cases plaso ingest module (that creates a super timeline of all events) showed unchanging zero progress at the progress bar while all other ingest process threads (viewed through Ingest Process Snapshot) remained idle. Next, we tried the experiment with the plaso module disabled since it was indicated as a time-consuming module in certain forums such as GitHub. Autopsy managed to finish the analysis of the 40 GB VHD with a 15 GB data bundle within 13hrs and 36 minutes. This result indicated that previous failures were due to the function of the plaso module.

5. Results discussion

The experiments with Autopsy indicates (Table 6) that forensic analysis, even with tool support, is a time-consuming process. Given the rather small sizes of the samples we used, the times which were taken were significant. And overall, the time consumption increased with the size of the evidence file. As indicated by Quick and Choo (2018), this follows the general trend with forensic tools. Considering we could not find similar experiments, our results will help indicate further avenues for improvements. This further validates the "Time taken for analysis/data loads" issue that we discussed in section 3.3. Another interesting factor that was identified is that the plaso module might consume too much time for practical purposes. Plaso provides the ability to create a super timeline of all events in the evidence source (https://plaso.readthedocs.io). Experiments so far indicate the need for better approaches for forensic analysis.

6. Conclusions and further work

As the next steps, we are hoping to expand the work we discussed in sections 4.1 and 4.2. For the issue of Reliability of the CSP, we are hoping to identify what types of reliability checking mechanisms we can adapt for cloud platforms. We are hoping to validate the solutions that we develop for cloud platforms.

For the issue regarding the time complexity of the tools, we are considering multiple approaches. In 2010, the Netherlands Forensic Institute established a Software-as-a-Service named XIRAF for forensic investigations. This new approach was coined as Digital Forensics as a Service (DFaaS) and was used successfully in both the Netherlands and elsewhere. Since both the data and the tools were on the cloud, this system reduced the storage issues as well the performance issues which existed in standalone workstations (Baar et al., 2014; Beek et al., 2015). Povar et al. (2015), introduced a process for triage forensics on clouds using the Hadoop framework. This framework used the MapReduce model in Hadoop for real-time analysis of an HDD using multiple computing nodes. In this experiment, they used up to 8 nodes and concluded that the overall time could be reduced with an increased number of nodes (Povar et al., 2015). Quick and Choo (2018) introduced a data reduction and data mining framework to handle the issue of ever-increasing data loads. They discussed widely how the data loads increase over time and how they would simply take too much time to process.

Concerning the tests, we conducted, especially the failure of the plaso module, we think the analysis can be sped up by pre-selecting the critical areas of the evidence source for analysis instead of analyzing the whole evidence source. We are currently working on trying to categorize the incidents in the cloud so that we can better identify the critical areas that need different investigations. By doing this, we hope to reduce the time taken for an investigation.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Baar VRB, Beek VHM, and Eijk VEJ (2014). Digital forensics as a service: A game changer. Digital Investigation, 11: S54-S62. https://doi.org/10.1016/j.diin.2014.03.007

Beek VHM, Eijk VEJ, Baar VRB, Ugen M, Bodde JNC, and Siemelink AJ (2015). Digital forensics as a service: Game on. Digital Investigation, 15: 20-38. https://doi.org/10.1016/j.diin.2015.07.004

Dykstra J and Sherman AT (2011). Understanding issues in cloud forensics: Two hypothetical case studies. UMBC Computer Science and Electrical Engineering Department, Baltimore, USA.

Dykstra J and Sherman AT (2012). Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. Digital Investigation, 9: S90-S98. https://doi.org/10.1016/j.diin.2012.05.001

FBI (2007). Regional computer forensics laboratory annual report for fiscal year 2007. Federal Bureau of Investigation, Washington, USA.

FBI (2013). Piecing together digital evidence-the computer analysis response team. Federal Bureau of Investigation, Washington, USA.

FBI (2016). Science and technology branch. Federal Bureau of Investigation, Washington, USA.

Freiling F and Hösch L (2018). Controlled experiments in digital evidence tampering. Digital Investigation, 24: S83-S92. https://doi.org/10.1016/j.diin.2018.01.011

Grispos G, Storer T, and Glisson WB (2012). Calm before the storm: The challenges of cloud computing in digital forensics. International Journal of Digital Crime and Forensics, 4(2): 28-48. https://doi.org/10.4018/jdcf.2012040103

Irons A and Lallie HS (2014). Digital forensics to intelligent forensics. Future Internet, 6(3): 584-596. https://doi.org/10.3390/fi6030584

Jang DI, Ahn GJ, Hwang H, and Kim K (2016). Understanding anti-forensic techniques with timestamp manipulation. In the IEEE 17th International Conference on Information Reuse and Integration, IEEE, Pittsburgh, USA: 609-614. https://doi.org/10.1109/IRI.2016.94

Kent K, Chevalier S, Grance T, and Dang H (2006). Guide to integrating forensic techniques into incident response. NIST Special Publication 800-86, National Institute of Standards and Technology, Gaithersburg, USA. https://doi.org/10.6028/NIST.SP.800-86

Liu F, Tong J, Mao J, Bohn R, Messina J, Badger L, and Leaf D (2011). NIST cloud computing reference architecture. Special Publication 500-292, National Institute of Standards and Technology, Gaithersburg, USA. https://doi.org/10.6028/NIST.SP.500-292

Manral B, Somani G, Choo KK, Conti M, and Gaur MS (2019). A systematic survey on cloud forensics challenges, solutions, and future directions. ACM Computing Surveys, 52(6): 1-38. https://doi.org/10.1145/3361216

Martini B and Choo KKR (2012). An integrated conceptual digital forensic framework for cloud computing. Digital Investigation, 9(2): 71-80. https://doi.org/10.1016/j.diin.2012.07.001

Palmbach D and Breitinger F (2020). Artifacts for detecting timestamp manipulation in NTFS on windows and their reliability. Forensic Science International: Digital Investigation, 32: 300920. https://doi.org/10.1016/j.fsidi.2020.300920

Pichan A, Lazarescu M, and Soh ST (2015). Cloud forensics: Technical challenges, solutions and comparative analysis. Digital Investigation, 13: 38–57. https://doi.org/10.1016/j.diin.2015.03.002

Pollitt MM (2007). An ad hoc review of digital forensic models. In the Second International Workshop on Systematic Approaches to Digital Forensic Engineering, IEEE, Bell Harbor, USA: 43-54. https://doi.org/10.1109/SADFE.2007.3

Povar D, Saibharath and Geethakumari G (2015). Real-time digital forensic triaging for cloud data analysis using MapReduce on Hadoop framework. International Journal of Electronic Security and Digital Forensics, 7(2): 119-133. https://doi.org/10.1504/IJESDF.2015.069602

Quick D and Choo KKR (2014). Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive. Trends and Issues in Crime and Criminal Justice, 480: 1-11.

Quick D and Choo KKR (2018). Big digital forensic data: Volume 1: Data reduction framework and selective imaging. Springer, Berlin, Germany.

Rani DR and Geethakumari G (2021). A framework for the identification of suspicious packets to detect anti-forensic attacks in the cloud environment. Peer-to-Peer Networking and Applications, 14: 2385–2398. https://doi.org/10.1007/s12083-020-00975-6

Rani DR and Kumari GG (2016). A framework for detecting anti-forensics in cloud environment. In the International Conference on Computing, Communication and Automation, IEEE, Greater Noida, India: 1277-1280. https://doi.org/10.1109/CCAA.2016.7813913

Reilly D, Wren C, and Berry T (2011). Cloud computing: Pros and cons for computer forensic investigations. International Journal Multimedia and Image Processing, 1(1): 26-34. https://doi.org/10.20533/ijmip.2042.4647.2011.0004

Roussev V, Barreto A, and Ahmed I (2016). API-based forensic acquisition of cloud drives. In: Peterson G and Shenoi S (Eds.), IFIP International Conference on Digital Forensics: 213-235. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-46279-0_11

Ruan K and Carthy J (2012). Cloud forensic maturity model. In the International Conference on Digital Forensics and Cyber Crime, Springer, Lafayette, USA: 22-41. https://doi.org/10.1007/978-3-642-39891-9_2

Schneider J, Wolf J, and Freiling F (2020). Tampering with digital evidence is hard: The case of main memory images. Forensic Science International: Digital Investigation, 32: 300924. https://doi.org/10.1016/j.fsidi.2020.300924

Simmon E (2018). Evaluation of cloud computing services based on NIST SP 800-145. NIST Special Publication 500-322, National Institute of Standards and Technology, Gaithersburg, USA. https://doi.org/10.6028/NIST.SP.500-322

Vurukonda N and Rao BT (2016). A study on data storage security issues in cloud computing. Procedia Computer Science, 92: 128-135. https://doi.org/10.1016/j.procs.2016.07.335

Yusoff Y, Ismail R, and Hassan Z (2011). Common phases of computer forensics investigation models. International Journal of Computer Science and Information Technology, 3(3): 17-31. https://doi.org/10.5121/ijcsit.2011.3302

Zawoad S and Hasan R (2013). Cloud forensics: A meta-study of challenges, approaches, and open problems. Available online at: https://arxiv.org/abs/1302.6312

Zhou S, Wu L, and Jin C (2017). A privacy-based SLA violation detection model for the security of cloud computing. China Communications, 14(9): 155-165. https://doi.org/10.1109/CC.2017.8068773