A Federated Distributed Digital Forensic Readiness Model for the Cloud

Renico Koen and Hein Venter
University of Pretoria, South Africa

Renico.koen@gmail.com
hein.venter@up.ac.za

Abstract: Digital forensics in modern, cloud-based, microservice-based applications are complicated by multiple layers of abstraction, thereby making it difficult to accurately capture and correlate events that occur across these layers due to filtering caused by abstraction. The complexities linked to each layer of abstraction are primarily invisible to subsequent layers. Similarly, software services are often composed of one or more services provided by various service providers across the globe. Investigators are often faced with situations where breaches span over multiple service provider boundaries where not all digital forensic readiness evidence artefacts are captured by the service provider’s forensic readiness processes. Instead, digital evidence artefacts are scattered across multiple service provider domains. This paper presents a novel, federated distributed digital forensic readiness model suitable for use in software-as-service, platform-as-service and infrastructure-as-service provider scenarios. The proposed model enables a service provider to capture and inspect forensic readiness artefacts in environments with various layers of abstraction. More importantly, the model also offers a way to share and access forensic readiness artefacts in a forensically sound manner to ultimately ensure that investigators can obtain a clear view of digital forensic events as they occur between amalgamated services provided by one or more separate service providers.

Keywords: Digital forensic readiness, Digital forensics, Cloud computing, Information silos

1. Introduction

Cloud computing is often called one of the most transformative technologies in recent history due to the way in which services can be composed and consumed (Ruan, et al., 2011). Cloud computing describes highly scalable, on-demand computing resources offered by service providers on a pay-as-you-go basis.

Cloud computing has been mentioned to reduce operational costs and improve operational efficiencies (Armbrust, et al., 2010). With the traditional client/server model of computing, i.e. without a cloud environment, server infrastructure must be purchased and maintained when hosting applications. The maintenance of these servers requires constant action from skilled team members. When calculating the total cost of ownership of traditional hosting models, the argument can be made that many hidden costs are typically present that impact the total cost of ownership of a provisioned service.

Cloud computing is convenient because many of the functions traditionally required for a client/server model are already included as part of the cloud computing package. Consumers end up paying only for services that they consume (Shetty, et al., 2014). Some cloud computing benefits include resilience, reliability, and on-demand access (Shetty, et al., 2014).

However, cloud computing does pose various challenges in terms of forensic readiness. The purpose of this paper is to highlight current digital forensic readiness challenges in cloud environments. A federated model is proposed to help address some of the challenges highlighted.

The rest of this paper is structured as follows: background information is discussed in section 3. Related work is discussed in section 4 and cloud computing challenges are highlighted in section 5. A federated digital forensics model is proposed in section 6. A discussion of the proposal is presented in section 7 while future work is discussed in section 8. The chapter is concluded in section 9.

2. Background

There are three recognised cloud service environments, namely infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and Software-as-a-service (SaaS) (Mell & Grance, 2011) and (Liu, et al., 2011). IaaS is concerned with the provisioning of virtualised infrastructure, such as network facilities, storage facilities and virtual machines (Alenezi, et al., 2017). The infrastructure provided by IaaS is relatively primitive but can be used to build more complex offerings, such as PaaS and SaaS. Villegas et al. (2012) refer to IaaS as the infrastructure resources used by the rest of the service stack to provide services. The customer does not manage or control the underlying hardware but can control the operating system and deployed applications (Mell & Grance, 2011).
The result is that the customer can focus on providing services that are within their fields of expertise while consuming infrastructure services provided by infrastructure specialists.

PaaS is concerned with providing a set of hosted libraries and tools (Alenezi, et al., 2017). The provided platforms allow service companies to create service offerings based on the provided platforms. The provided platforms are hosted on IaaS (typically included as part of the PaaS offering). Widely known examples of PaaS include Microsoft Azure and the Google App Engine (Villegas, et al., 2012) and AWS. Facilities offered by PaaS are limited to the components that have been assembled or developed by the platform provider (Weir, et al., 2018). The consumer does not manage the underlying hardware or operating system environment but can deploy applications and change application hosting settings (Mell & Grance, 2011). As a result, the customer is shielded from most of the technicalities related to hardware and underlying operating system management and securing of the underlying resources.

SaaS can be described as hosted applications (Alenezi, et al., 2017). These applications are typically provided via the user browser or API (Alenezi, et al., 2017). The SaaS offerings typically include PaaS (and, as a result, IaaS) components as part of a managed service. SaaS offers various benefits, such as shorter application implementation time and the owners' elimination of software version management (as the service provider manages version management).

Cloud computing helps organisations deliver services or access online services that are often amalgamations of services provided by various vendors. However, the investigation of security events in service delivery environments is often complicated because multiple service providers may be present, which can potentially be located all over the world and are governed by different legal frameworks.

The next section of this paper discusses current challenges in cloud forensics.

3. Related Work

Current digital forensics challenges can be classified into five categories, namely complexity of low-level data acquired, diversity of data sources, consistency and correlation problem caused by current tooling that attempts to find fragments of evidence and nothing more, and volume and unified timelining issues caused by various evidence sources acquired using non-consistent time zones and time sources (Lillis, et al., 2016). The challenges were discussed by Lilllis et al. (2016), who also highlighted that the Internet of Things would likely worsen existing digital forensics data processing backlogs due to increases in volumes of data. In addition to these categories, there are significant technological challenges to contend with. Technological challenges include the rapidly changing technological landscape, encryption, large volumes of data and the complexity of digital systems and networks (Alenezi, 2023).

Miller et al. (2014) proposed a cloud-based distributed processing platform called Forensicloud in an attempt to address the increasing volume and diversity of data in forensic examinations.

Furthermore, Duijn & Sloot (2015) explore the use of big data and analytics to determine how organised crime groups operate and adapt over time. Analysts may be able to uncover patterns that are hidden or not obvious by analysing data from a diverse set of sources. However, the quality of data sources plays a role in the output produced.

Monteiro et al. (2023) discussed challenges related to forensics in a microservice environment. The author emphasised the importance of forensic-ready microservices and proposed a framework that incorporates game theory to achieve such a goal.

The lack of standardisation, the complexity of cloud environments and the dynamic nature of cloud environments are typical problems in cloud environments (Alenezi, 2023). Hence, cloud service providers must implement their services in a way that ensures that investigators can gather evidence in a forensically sound manner (Simou, et al., 2022). As a result, an organisation's forensic readiness levels should be monitored to ensure readiness should an incident occur.

Shanmugasundaram et al. (2003) proposed a distributed network logging mechanism named ForNet, to assist digital forensics over a wide area network. A component named SynAps acts as an agent that summarises network events. These events are then sent to a server (called Forensic Server) for analysis.

Federici (2013) proposed an architecture called AlmaNebula with the purpose of providing forensics as a service for cloud computing by utilising commodity machines for processing large amounts of data.
Sibiya et al. (2013) discussed security challenges related to cloud computing and digital forensics. A forensic readiness model was introduced that utilises forensics services hosted in the cloud to minimise the amount of time needed to perform forensic investigations. The model consists of various components, such as application forensics, memory forensics, network forensics and computer forensics.

Kebande & Venter (2018) proposed cloud forensic readiness as a service model for collecting and preservice potential digital evidence. Agent-based sets are used to collect information. Collected evidence stored in a forensics database for analysis.

Blockchain-based solutions were also discussed in the literature. Al-Khateeb, et al. (2019) discussed incorporating blockchain into digital incident response systems, specifically focusing on maintaining the chain of custody for digital evidence. The need for chronological documentation of evidence for the court of law was emphasised, which can be achieved using blockchain technologies.

Finally, an architecture proposed by Nanda & Hansen (2016) caters for layers introduced by cloud computing. An additional layer is introduced to the cloud computing environment that ingests forensic data. The architecture, named Forensics-as-a-service (FaaS), obtains forensic information from services running in the IaaS, PaaS and SaaS layers. The layer is operated by an external party.

The proposals discussed in this section focus on similar problem areas as the model proposed in this paper. However, no clear answer is currently documented in literature that helps to address the forensics-related information silos that currently exist between service providers. A study is therefore needed to determine a way to collect and share forensic readiness data between multiple related parties operating within service delivery environments in a secure and privacy-aware manner.

4. Cloud Computing Forensics Challenges

From a forensics perspective, the IaaS, PaaS and SaaS models are of interest as each model can be seen to introduce a layer of abstraction. IaaS provides a way to virtualise and commoditise hardware. As a result, consumers of IaaS interact with virtual representations of hardware but not the actual hardware itself. Similarly, consumers of platform offerings consume libraries hosted on IaaS. The underlying PaaS and IaaS details are rarely visible to the consumer of these platform services. Lastly, the consumer of SaaS services is generally only exposed to the applications that they consume and not to the details of the underlying platform or hosting environment. In addition, a SaaS offering may consist of PaaS and or SaaS offerings provided by multiple vendors. The inclusion of various services provided by external vendors allows companies to focus on their key strengths while outsourcing everything else to external providers.

A hierarchy, therefore, exists where IaaS is at the bottom of the hierarchy and SaaS at the top. The argument can be made that additional layering in the service model creates abstraction. At the top of the hierarchy, the layers of abstraction are so many that a user is unaware of any finer-grained hosting details. From a service provisioning perspective, the layering greatly simplifies complexities. In addition, should a breach be experienced, it becomes more challenging to perform an investigation as not all the details will be visible on all layers.

Consider the example where the entire user database of a SaaS service has been copied by an external party (adversary). The consumers of a SaaS service will typically have access to application-level audit facilities to determine if the source of the breach was a compromised user account. In some cases, an audit facility may help SaaS consumers determine if a compromised user account is to blame for a breach. However, SaaS consumers will typically not have access to any PaaS or IaaS details. For example, should an access control mechanism be broken on the PaaS layer, then it is unlikely that application-layer controls will be able to identify or compensate for such a deficiency, as the application layer has no visibility over the events occurring on the platform layer.

Similarly, should a physical breach be experienced (where a threat actor enters a data centre), then PaaS and SaaS systems are unlikely to have visibility over such events. An investigation focussing on SaaS controls, as a result, will be inconclusive. Similarly, an investigation focusing on SaaS and PaaS controls will also not produce conclusive evidence needed to identify the source of such a breach beyond a reasonable doubt. Effectively the forensic readiness capabilities of service providers form silos that are only accessible to the service providers but not to service consumers. As a result, it is argued that the layers of abstraction offered by cloud computing help for rapid service provisioning but are a burden to security and forensics.
The problems described can be formulated into a research question as follows: How can a model be defined that can help parties share information in a forensically sound manner in a cloud environment that would facilitate information sharing among service providers in a controlled manner?

The following section presents a model to help the facilitation of forensic investigations in distributed cloud environments, where multiple parties are involved that currently collect forensic data in silos.

5. Model

The lack of information sharing between parties related to an incident can potentially also have a negative impact on the incident investigation process. Each provider of a service in a cloud environment will have a view of events that transpired while providing a service to clients. An infrastructure provider will, for example, have visibility on network events, hardware failures, and physical access control breaches (among other things). Although tracked and monitored by the infrastructure provider, this information will rarely be shared with consumers of IaaS, PaaS and SaaS services. There can be many reasons for this, including privacy reasons; similarly, there will be situations in which forensic events captured by PaaS and SaaS providers are not shared with consumers of those services. Yaqoob et al. (2019) also mentioned that privacy-aware processes are generally lacking in forensic practices. The result of this is that not all information required for full transparency will ever be present unless there is a way for all the various parties to share forensically relevant information in a forensically sound manner while preserving the privacy of other tenants or users that may be involved.

A potential model for distributed forensic readiness is for each service provider in the cloud computing value chain to provide a forensic service that can be queried by upstream consumers in the SaaS value chain. Having access to forensically sound information in a controlled manner may help to reduce the information absence issue currently present in cloud environments. In addition, such a service can potentially be funded by service providers by charging service or subscription fees as an incentive.

Obtaining relevant forensic information from service providers would ensure that information not currently available for forensic investigations may, as a result, be accessible to investigators. In addition, it may also be possible to determine forensic readiness coverage of services provided internally and by external vendors, thereby helping security teams better understand their exposure to risk. Should it be possible to quantify forensic information completeness, then it may also be possible to use such a gauge to determine forensic risk (which insurers may later use to determine information security insurance premiums).

A potential problem with distributed forensic readiness would be sharing information while protecting confidential or personal information. Consider the example of two services that have been included as part of a cloud service offering, and that the service providers are in different countries. Should an incident occur and one service provider requests forensic records from the other provider, and the forensic records contain personal information, then personal information would be transferred across borders. This can be problematic if proper consent is not in place to allow this. Similar problems may also exist in terms of data governance. As a result, approval controls may be needed to ensure that distributed forensic information requests are reviewed and approved on demand as requested by external parties. Appropriate controls are needed to protect the information requested, and as a result, more steps will be needed to obtain potentially sensitive information as compared to non-sensitive information. The assumption to date has been that each service provider will operate their own forensic readiness processes in isolation from other service providers.

Responsible service providers practising forensic readiness will have controls in place to ensure that they capture all relevant information in a forensically sound manner to ensure that it will later be possible to perform forensic investigations using the information captured by their forensic readiness processes. Assuming that one cloud service consists of multiple amalgamated services provided by multiple distinct parties and forensic readiness is practised by all parties involved, then it may be beneficial for parties also involved to share forensic readiness information to ensure that a holistic picture is provided of an incident that occurred.

For example, the forensic readiness system employed by an IaaS provider will be of great use to investigators investigating a database leak of a SaaS solution if the SaaS solution provider has reason to believe that a threat actor with physical access may have been responsible for the data leak. Without sharing information, it would not be possible for a SaaS service provider to accurately determine the incident’s root cause, as not all the pieces of information required will be present. By combining the pieces of forensic readiness information captured by all parties, it would be possible to obtain a clearer picture of events that occur between service providers, thereby providing a holistic picture of what happened after combining evidence from various sources.
Considering that forensic readiness information may contain personal information and considering the complexities involved with the transfer of personal information across borders, storage of personal information and processing of personal information, it can be argued that service providers are likely to retain their forensic readiness information within forensic readiness silos; in other words, service providers will capture forensic readiness information that is available to them. However, due to legal and privacy requirements, forensic readiness information will not be shared among service providers, thereby decreasing holistic visibility over forensic events.

Assuming that this model will not change anytime soon, it can be argued that there may be value in temporarily correlating forensic readiness data captured by various service providers in an attempt to understand the actions of threat actors that span across service provider domains.

A legal basis should exist for an investigator to request access to forensic services provided by other vendors. Once a vendor manages to establish a legal basis, and the legal basis is verified by the service provider with the forensic readiness artefacts of interest, then a structured discussion can occur between the two parties. An investigator will need access to very specific pieces of information, typically events that occurred within a specified timeframe. Considering that service providers are likely to service one or more clients using the same software or hardware, it can be argued that multiple layers of visibility exist, namely:

- Events that occurred for only the tenant, for example, an unauthorised user who managed to log into the supplied service application;
- Events that occurred for all tenants, for example, tenant downtime due to a failure of one of the key services relied upon by the SaaS solution;
- All events that occurred, irrespective of the tenant, for example, hardware failures on the IaaS layer, would impact all layers on top of the infrastructure layer.

A request for information for events that occurred between a specific timeframe for the tenant itself on a remote service provider will be less sensitive compared to a request for information for all events that occurred for all tenants on the same remote service provider due to the fact that sensitive information related to parties that are external to an investigation in question may be exposed to investigators. However, there may be no other option in cases where suspicion exists that multi-tenancy issues on a service provider led to a system compromise of a service hosted by a different provider.

As a result, in addition to a legal basis, consent also needs to be in place to allow investigators to provide information linked to various tenants or their systems to external investigators. In cases where consent cannot be obtained from the relevant parties involved, then some information cannot be provided to investigators. However, it will be possible to calculate a completeness score based on the consent obtained from multiple tenants involved and the amount of data represented by the tenants for which consent was obtained. Without obtaining full consent from all tenants, one can argue that there is a possibility that some pieces of a puzzle will be excluded from an investigation.

Figure 1 provides a visual depiction of components within a distributed digital forensics framework. The diagram shows a multi-tiered architecture where upper-tiered components make use of components located in lower-layered tiers. Each of the components is discussed in the following subsections.

![Figure 1: Federated distributed digital forensics components](image-url)
5.1 Layer A.1: Policy Engine

The policy engine is the glue that holds together all underlying components in the federated model. The policy engine orchestrates all requests made in the federated environment and received in the federated environment. Predefined policies govern how external parties can access forensic artefacts, how long external parties may retain acquired evidence, controls that should be in place to protect the evidence, labels that should be applied and how requests made to external parties should be handled. An example of a policy may require a police case number, justification and special privilege levels, and manual approval before anonymised evidence can be released to a requestor. Conceptually, hundreds of policies can be defined that govern how data can be shared, anonymised and retained.

The definition of policies can potentially be done in a mark-up language, such as YAML, XML or XACML (OASIS, 2013). Since working in a federated environment, it should also be possible for a policy to reference policies applied to other federated resources.

Due to the complexities involved, defined policies will have to be validated before being applied. For federated policies referenced from remote locations, a similar validation process will have to be followed to ensure that the defined policy is valid, and that the policy is valid within the context of the referenced policy.

5.2 Layer B.1: Request Management

In a federated model, authorised parties can request forensic evidence sources relevant to the investigations performed by these parties. Requests made and received need to be validated and fulfilled in a manner that ensures nonrepudiation, confidentiality, and the integrity of data.

Request queuing occurs on this layer. It may often happen that multiple requests are submitted at once or that requests are submitted that may require manual interaction by one or more parties. As a result, requests are attached to manual and automated workflows. As part of the workflow process, requests may also include processing performed by other layers. In cases where additional permissions are requested or de-anonymised data is required, an escalation workflow can be kicked off to obtain the necessary permissions. For such a request, the number of approvers, as well as the level of authority of the approvers, will be dictated by policies managed by the policy engine.

Identity management, authentication, authorisation and platform audit logging are also handled in the request management layer.

5.3 Layer B.2: Federated Alerting

In cases where cooperation agreements are in place, it should also be possible to send alerts to federated parties when local incident detection techniques show that an event that includes one or more external service providers may be present. Alerts can be structured by considering the appropriate labels, retention and anonymisation requirements. Ultimately, alerts are structured and sent to the request management layer to process as workflow items. As a result, alerts will undergo various processing steps to ensure that alerts being delivered are in line with defined policies. This may also imply that different policies are applied for different recipients. For example, different rules can potentially be applied to recipients in Europe compared to recipients in South Africa due to differences in legal requirements.

5.4 Layer B.3: Anonymisation

Data requested by federated parties may contain sensitive information, such as keys or personal information. The sensitivity of the data in question is determined by the labels assigned. More sensitive information will require stricter controls and approvals to be in place when compared to requests containing less sensitive information. Under certain conditions, it should be possible to obtain personal information under certain conditions, but under most conditions, performing incident investigation using anonymised data would be fine up to the point where personal information may be required to identify a threat actor uniquely. By performing requests in a stepwise fashion, as described, it would be possible to perform investigations in a systematic manner while keeping the information exposed to federated parties to a minimum.

5.5 Layer C.1: Labels

Data captured in the evidence storage facility needs to be labelled to facilitate automated decision-making and anonymisation processes. Evidence labelled as containing sensitive information or personal information will require additional permissions and processing, such as anonymisation or encryption (depending on the data in
question). As part of a forensic readiness process, forensic data under the control of the service provider can be labelled according to sensitivity. Federated evidence received from external parties will contain labels, as defined by the external parties for the evidence at hand.

Sensitivity labels can be used to apply the appropriate level of anonymisation or apply the proper level of protection based on the labels being applied. In cases where local policies reference federated policies, it may, in some cases, also be required to map remote labels to local labels to ensure consistency in processing. In cases where data is received in a federated context, the labels applied by the originator should be respected.

5.6 Layer C.2: Retention

Evidence collected and received by external parties should be retained as long as is required for legal purposes. Data retention periods may often be influenced by the data labels. For example, personal information retention legal requirements may differ from the legal requirements for financial data. Similarly, federated evidence received from external parties should be kept only as long as needed to conclude an investigation and as long as permitted by relevant laws. In cases where data is received or retrieved in a federated context, the retention policies of the originator should be respected.

5.7 Layer C.3: Incident Detection

Evidence collected from various data sources needs to be correlated to detect digital incidents. The incident detection will include data collected by the service provider but can also include data obtained through the federation.

For the purposes of this discussion, an incident can be seen as any event that may be of interest to an investigator. The incident detection layer defines what an incident looks like, given the various evidence sources at hand. Minor transformations can also be defined and applied as part of the incident detection process. Defined detection definitions are to be stored in a mark-up file.

5.8 Layer D.1: Forensic Processes

Evidence captured by the service provider needs to be stored in a forensically sound manner with controls in place to ensure adequate protection against misuse. Forensic processes are concerned with the capture and preservation of forensic evidence in a forensically sound manner.

There are some technicalities related to the processing of forensic evidence obtained as part of a federated process. Software vendors in the SaaS supply chain can individually collect and preserve evidence as part of their digital forensic readiness processes. Captured artefacts can be stored within the solution provider’s perimeter (referred to as local evidence storage for the purposes of this discussion). Alternatively, captured information can potentially be stored in the same storage facility used by various service providers in the SaaS value chain (such as AWS CloudTrail).

In both cases, accessed by different parties, potentially in different countries, it implies that data processing agreements need to be established among parties and that the consent of data subjects will have to be in place to perform such actions. Ultimately, such policies will be enforced by components located in higher-up tiers within the model.

The next section presents a discussion on the proposed model.

6. Discussion

The purpose of the proposed model is to help improve information sharing in environments where the sharing of information is structured in ways that respect data privacy laws. This is done through a collaborative framework that dictates the conditions under which data can be accessed.

Conceptually, the proposed model can be applied to local investigations as well as investigations that may span across organisational boundaries. The model was proposed, assuming that various forensic readiness controls that capture data on various service delivery layers will be present. The implementation and operation of controls in an organisation typically incur costs. As can be expected, organisations will have finite budgets allocated to forensic readiness, thereby implying that some forensic readiness controls will be prioritised. In contrast, others will receive little to no attention. Controls that are prioritised are typically decided through a risk assessment based on business risk. Controls that address important business risks are typically prioritised over controls that address less critical risks. In addition, there will also be risks that an organisation may not be aware of. As a result, it can be expected that controls will provide a degree of coverage, locally and remotely.
Future work will explore ways to determine the coverage of local data and remote forensic data more accurately (through federation).

It is also important to mention that organisations will adapt existing information security policies to ensure that cooperation with other organisations and law enforcement agencies is allowed.

In terms of the policy engine discussed, policies may also reference federated policies. In addition, in some cases (such as labelling), mappings between compatible labels will also have to be performed. In cases where clear misalignments exist between policies, a policy can potentially be defined that favours local policies over remote policies (or visa-versa) or requests manual intervention from an approver.

Access to forensic artefacts locally and remotely needs to be logged in a forensically accepted manner. It should be noted that federated requests should be logged in a way that provides remote parties with auditing visibility over actions taken on federated data provided to ensure accountability and traceability.

The next session will present a discussion on future work.

### 7. Future Work

A model to facilitate information sharing among service delivery stakeholders was discussed in this paper. Future research work will include the implementation of the model and a study of the effectiveness of the implementation of the model in real-world scenarios.

### 8. Conclusion

This paper describes a current issue that exists in cloud computing due to various layers of abstraction present. Service providers focus on the provision of services that include their core expertise. All other aspects of a service that is not included are typically outsourced to third parties. This can include hosting and non-core services that are included as part of a service delivery offering.

The problem identified focused on the forensic readiness silos that exist between organisations where virtually no forensic readiness information is shared among service stakeholders. The net result of this is that investigators often only have access to partial information when performing investigations. Ultimately, the lack of information can lead to inconclusive investigations due to lack of information or evidence.

A model was proposed that allows investigators to share information in a structured manner. Furthermore, the model also allows inter-organisational sharing of forensic data in a structured and privacy-aware manner to ultimately help to easy forensic investigations that may span across organisational boundaries.

There are a few discussion points that will be addressed in future research, such as the estimation of local and remote forensic readiness completeness. The implementation of a prototype will also be performed to illustrate the concept in action.

### References


