Reference Architecture for a Cloud Forensic Readiness System

Lucia De Marco\textsuperscript{1,2,*}, Filomena Ferrucci\textsuperscript{1}, M-Tahar Kechadi\textsuperscript{2}

\textsuperscript{1}Department of Management and Information Technology DISTRA-MIT, University of Salerno, Italy\textsuperscript{1}\textsuperscript{1}
\textsuperscript{2}School of Computer Science and Informatics, University College Dublin, Ireland\textsuperscript{2}

Abstract

The Digital Forensic science is participating to a brand new change represented by the management of incidents in the Cloud Computing Services. Due that the Cloud Computing architecture is uncontrollable because of some specific features, its use to commit crimes is becoming a very critical issue, too. Proactive Cloud Forensics becomes a matter of urgency, due to its capability of collecting critical data before crimes happen, thus saving time and money for the subsequent investigations. In this paper, a proposal for a Cloud Forensic Readiness System is presented. It is conceived as reference architecture, in order to be of general applicability, not technically constrained by any Cloud architecture. The principal aim of this work is to extend our initial proposed Cloud Forensic Readiness System reference architecture, by providing more details and an example of its application by exploiting the Open Stack Cloud Platform.

Keywords: Cloud Computing, Cloud Forensics, Cloud Forensic Readiness System, Cybercrimes, Cyber-Security, Forensic Readiness, Reference Architecture, OpenStack

Received on DD MM YYYY, accepted on DD MM YYYY, published on DD MM YYYY

Copyright © YYYY Author et al., licensed to ICST. This is an open access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/3.0/), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/______________

*Corresponding author. Email: lucia.de-marco@ucdconnect.ie
\textsuperscript{1} Via Giovanni Paolo II, 132 - 84084 – Fisciano, SA, Italy
\textsuperscript{2} Belfield, Dublin 4, Ireland

1. Introduction

Rapid and diverse technological advances of the last decade have influenced all the aspects of our lives, from personal to business. However, this progress has not been always positive; it has originated huge and sophisticated digital crimes and their number is in constant progression. This has led to the development of Digital Forensics (DF) discipline [19], which is becoming more and more significant.

DF allows the development of scientifically derived and proven methods for evidence extraction from digital devices by reconstructing a correct event timeline that is fundamental to the crime cases resolution. Several tools and procedures have been established [2], and should be adapted to follow the technological innovations.

Cloud Forensics (CF) is a new frontier in DF [23], since the Cloud Computing (CC) technology is the current ICT evolution [12]. Cloud technologies are becoming very popular for several reasons and Public Cloud services expenditures are expected to record an annual growth rate of 17.7% from 2011 through 2016, i.e., 210 billion dollars in five years [10].

While the Cloud has made its success from some specific appealing characteristics, nevertheless, the same characteristics have also issues for DF investigations.
Table 1 – Cloud Forensics Main Challenges

<table>
<thead>
<tr>
<th>CF Challenges</th>
<th>Elasticity</th>
<th>Multiple Locations</th>
<th>VM</th>
<th>Broad Network Access</th>
<th>Third Party Services</th>
<th>Cross-Providers</th>
<th>SLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced data access</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of physical control</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of standard</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple log formats</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No timestamps synchronization</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No routing information</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of expertise</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal measures</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi–tenancy</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Multiple jurisdiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 summarises the main CF challenges [3, 13, 21, 23, 25, 30]. The Cloud Services are provisioned on demand and released once done. This property, called elasticity, is guaranteed by involving several distributed servers and Virtual Machines (VMs) to deal with important demands [12].

Another important issue is the heterogeneity of log files and their formats, which differ from one to another. In addition, there is a big issue with the lack of synchronisation between data centres controlled by different providers.

Moreover, the principal question that needs to be answered is how to manage the multi-tenancies in order to preserve other clients’ privacy. For instance, there is lack of legal experience specific about the Cloud features, which determines uncertainty about the measures to undertake in different cases, e.g., when a cross-providers or third parties resources supplying happens. The Service Level Agreements (SLAs) [12] are lacking of such information, indeed they should include appropriate rules in order to clearly assign responsibilities. Finally, the multiple jurisdictions concern adds another level of complexity to the Cloud Forensics, because Cloud infrastructure can be distributed all over the world. Several legal principles are taken into account in order to address this matter [9, 18, 29], but all these concerns still made Forensic investigations very challenging.

An approach for dealing with these issues is the implementation of Digital Forensic Readiness (DFR) [5, 22, 26] into the Cloud. DFR allows computers architectures to collect sensitive and critical information related to digital crimes before they happen, leading to save time and money during the investigations [5].

The paper is structured as follows: Section 2 presents the literature review about Forensic Readiness. Section 3 describes the approach undertaken for our proposal. Section 4 discusses the proposed reference architecture for a Cloud Forensic Readiness System. Section 5 illustrates the system requirements. Section 6 concludes the paper and gives some future research directions.

2. Literature Review

The Digital Forensic Readiness (DFR) is a very active research field that attracted many researchers. DFR combines forensic expertise, hardware, and software engineering.

The initial idea of DFR was proposed in 2001 [26]. A DFR system has two objectives: maximizing an environment’s ability to collect credible digital evidences, and minimizing the cost of Forensics in an incident response. DFR includes data collection activities that concern some components such as RAM, registers, raw disks logs. Other related factors have been also analysed, such as, how the logs will be recorded, what is actually logged, how the Intrusion Detection Systems (IDS) behave, how the Forensic Acquisition actually happens, and what are the procedures to be used for Evidence Handling.

Later, more details about DFR have been given. For instance, the DFR capability can be considered as an important feature, because it can help the digital investigations and can facilitate some other activities concerning data security [11, 22]. The first step is to identify, collect, and store critical data coming from the suspect’s computing infrastructure.

Rowlingson [22] proposed a 10-step DFR approach, necessary for an organization who wants to achieve such capability. In this work, a crucial issue is that many
organisations, as part of their general information security, incident response and crime prevention activities already effectively collect and exploit electronic evidences. The lesson learned is that in order to actually implement DFR only a systematic and pro-active approach is required for the gathering and the preservation of evidences.

In [8] several Network Forensics aspects were discussed. The authors analysed situations where the cyber targets are powerless to attackers and intruders, who are able to exploit and disrupt the networks. The authors stated that Network Forensic Readiness (NFR) can be a good solution, even though no comprehensive organizational implementation approach exists yet. Their main objective is to provide a theoretical framework consisting of various models for implementing NFR in enterprises contexts.

Another proposal concerns DFR for a Wireless Sensors Network [14]. The purpose is to design a prototype for such architecture according to a list of requirements. The issue is that these requirements have not been tested in real wireless sensor network scenarios. Thus, the authors performed demonstrations to show the usability of the list of requirements, and a DFR system was prototyped and implemented as an additional layer of the network architecture.

In [28] the DFR was analysed for Public Key Infrastructure (PKI). A PKI system is a set of hardware, software, people, policies, and procedures needed to create, manage, store, distribute, and revoke digital certificates. These systems are used to implement information systems security services such as authentication and confidentiality. The authors investigated a model together with a set of policies, guidelines and procedures, for implementing a DFR framework. They took into account some requirements for either preserving or improving information security and at the same time not altering the existing business processes of such PKI systems.

In [20] an approach for managing DFR in large organizations was discussed. They reviewed the literature and proposed a novel DRF architecture. The architecture is supported by an early proof-of-concept prototype system to demonstrate its feasibility.

DFR has been analysed also for Cloud Computing [25], and few prototypes were implementing. In [6] some existing forensic tools like EnCase were analysed in the CFR context. The results showed that the Cloud data collected by those tools are unreliable, because some important Cloud features are not taken into account. More efforts are required in order to perform Forensic Readiness than simply tailoring existing tools and procedures. The Cloud technical requirements must be managed also for complying with the existing legal principles concerning the digital evidences.

A remote forensic acquisition suite of tools was proposed in [7]. The suite, called FROST, provides a forensic capability at the IaaS level of OpenStack; an open-source Cloud Computing platform. FROST performs data collection from the CSPs and from the host operating system. The collected data includes virtual machines images, logs collected from the API requests, and the OpenStack firewall logs. This suite is considered also as a way for adapting Forensic Readiness to the Cloud, because it performs the necessary data collection activities.

Finally, [27] presented a manner for achieving Digital Forensic Readiness in the Cloud. It comprehends a remote and central logging facility for accelerating the acquisition of data; the model was also prototyped for Windows platforms.

3. Forensic Readiness Approach

In this paper we present an extension and an in-depth analysis of a Cloud Forensic Readiness System (CFRS) [5]. The main purpose of such a system is to provide a manner for implementing the Forensic Readiness capability in the Cloud. This can be achieved by collecting and monitoring sensitive and critical data; i.e., potential digital evidences. The benefits of the FR capability are saving time and money for the investigations and reducing their direct impact on the Cloud Services activities and performance [5, 22, 26].

The approach for defining a Cloud Forensic Readiness System is to start by reviewing the cloud reference architecture (see Figure 1). This approach is not constrained by any specific Cloud architecture. The main objective is flexibility and customization that can be adopted by different organizations and Cloud Service Providers.

Our proposed approach to Forensic Readiness System does not alter or modify the main functions of an existing Cloud architecture. All the components and subsystems that the CFRS comprehends will interact with the existing Cloud modules, so that they can collect data through appropriate communication channels. All computations and readiness operations will be executed in separated and dedicated CFRS components. More technical details are discussed in Sections 4 and 5.

4. FR System Architecture for the Cloud

Figure 1 represents the main components of our CFRS reference architecture. It includes several modules dedicated to specific operations, connected each other via dedicated Open Virtualization Format (OVF) communication channels. OVF is a standard language [17] suitable for both the design of distributed applications for the Cloud. OVF exploits the XML standard to establish the configuration and the installation parameters. It is capable of creating and distributing Software applications to be executed on different VMs, independently from hypervisors and from CPUs architectures. OVF can be also extended for future VM hypervisors developments, thus its usage is reasonably motivated by its features.
The initial activity of such a CFRS is the collection of the data. In the Cloud, this activity is very critical, because it must be performed on time, with respect to existing laws and privacy policies, and involving the data, that have a forensic value in a court. The identified Forensic Readiness Cloud data are the ones represented in the white boxes of Figure 1. They comprehend both Cloud Services artifacts, and outputs from some existing Cloud monitoring tools. All the considered data are defined as Cloud Computing Common Components [4]; this means that their presence can be considered a “must have” requirement in most of the Cloud architectures.

The Cloud artifacts, in the central white box shown in Figure 1, are composed of the following: the VMs Images and the Single Sign-On logs, which can be found in the SaaS level; the system states and applications logs retrievable from PaaS; while in IaaS the snapshots and the system memory are available.

Other Forensic relevant data are composed of logs; they are shown on the right hand side white box (see Figure 1). The ones coming from Cloud Auditors, where information related to the customers accesses and their operations are collected by the system. The error logs from hypervisors, indicating suspicious events, are considered, too. Also some information about the Cloud Carrier has to be considered. A Cloud Carrier is an intermediate between Cloud Consumers and Providers. The Carrier is responsible for providing connectivity and transport of Cloud services to the Consumers through the network and other access devices [24]. Therefore, some information it stores is suitable for Forensic Investigations; they include network logs, activity logs, access record facility logs, hypervisor events logs and virtual images (see Figure 1).

The monitored data are composed of: the output of Database and File Activity Monitoring, such as URL Filtering, Data Loss Prevention, Digital Rights Management System and Content Discovery System. These tools are all defined in the same way among most CSPs [4].

The Database and File Activity Monitoring tools recognize a data migration when a huge amount of data is pushed into the Cloud or replicated. The Data Loss Prevention facility monitors the data in motion; it also manages policies and rights. The URLs Filtering controls the customers’ connections to the Cloud Services. The Digital Rights Management System is responsible of implementing and monitoring the customers’ rights and restrictions on the data; co-signed in the SLAs and the Contracts between CSPs and Customers. The Content Discovery System includes tools and processes for identifying sensitive information in the cloud storage; it also allows to define customized policies and identifies their violations.

The outputs of the tools, included in the CFRS reference architecture on the left hand side white box, are relevant and necessary for Forensic investigations, because they can be used for reconstructing a reliable events timeline. This operation is essential and very important because it helps to understand what actually happened and which information is involved in a case.

The data collected from the Cloud will be manipulated outside the Cloud architecture. The two CFRS components on the top of Figure 1 are responsible of saving and managing the data, respectively. The Forensic Data Base module is dedicated to the preservation of the potential digital evidences. This module has three different subsystems; each contains Cloud data depending on its type (see Figure 2). The Services Artifacts sub-system is dedicated to the Cloud artifacts, such as, VMs Images, Single Sign-On logs, system states and applications logs, running system memory. Cloud Auditors logs and error logs coming from hypervisors are instead included in the Forensic Log, together with the logs coming from the Cloud Carrier. The Monitored Data sub-system refers to the outputs form the mentioned tools, e.g., Database and File Activity Monitoring, URL Filtering, Data Loss Prevention, Digital Rights Management System and Content Discovery System.

The previously collected and saved data are encrypted, stored, and monitored by dedicated sub-components, i.e., Data Encryption and Data Storage. The Data Management sub-system performs Forensic analysis and knowledge extraction in order to reconstruct a correct and reliable events’ timeline. Finally, the chain of custody (CoC) report [15], which is necessary for cases resolution, is performed by the Chain of Custody sub-system.
Figure 2 shows another view of the same reference architecture, with the purpose of representing the sequence of the Cloud Forensic Readiness System activities.

As mentioned above, the datasets coming from the Cloud environment are collected and stored in Forensic Data Base. In Figure 2 we can see how the Cloud data are saved in the three Forensic Data Base sub-systems, i.e., Monitored Data, Services Artefacts, and Forensic Log.

In order to accomplish the widely adopted British ACPO [1] and American National Institute of Justice [15] guidelines concerning the preservation of the potential digital evidence, the collected data has to be copied. This step is necessary for preserving the original copies when Forensic activities are performed. Subsequently, the same data has to be secured to avoid tampering, and this can be performed through a dedicated Data Encryption subsystem (see Figure 2) where proper digital sign and data securing routines are implemented.

The whole system activities and modules are constantly running and collecting data, for obtaining always up-to-date versions.

All this information is fed to the Intrusion Detection process, in order to analyse when a Cloud incident has happened. It has to implement specific policies, dedicated to manage the suspicious behaviours. These policies might consider the co-signed SLAs clauses [12] because they are necessary for understanding what the violated services requirements are. The Intrusion Detection module strongly collaborates with the Events Alerting one, as it generates alarms of suspicious behaviours. The alarms are different depending on the type of events.

The Data Mining module (Data Management of Figure 1) is responsible for hidden knowledge extraction functions in order to generate the case related digital evidence. The reconstruction of the events timeline takes place also in this module.

The evidences must be treated in the respect of the existing guidelines, best practices, and laws. They are used in court admissibility in order to prosecute a case. For this purpose, proper and dedicated policies and routines are necessary to be implemented in the Preservation of Digital Evidences module.

Some information related to the data, e.g., location, treatment, date, time, time zone, and system component, have to be recorded, in order to maintain a reliable chain of custody, which is necessary for prosecution purposes (see Figure 1).

The proposed Cloud Forensic Readiness System has to be collaborative with the competent bodies involved in the criminal cases management. The CFRS is responsible to interact with them; the system can also have the duty of transmitting the retrieved data belonging to the arisen cases, together with the digital evidences and the chain of custody documents, to proceed with the law enforcement. The competent bodies can be both Incident Response and Law Enforcement, thus dedicated interfaces and communication modules are necessary, as they are represented in Figure 2 with other two components (on the right hand side), if the mentioned competent bodies are different, additional communication modules can be included.

5. System Requirements

Cloud Computing architecture, augmented with FRS, should consider the following requirements. The described artefacts and tools’ outputs exploited by the system must be available. These forensic data are defined as common Cloud features [4] therefore the verification of their presence at the moment of the system installation should be easily achieved.

Another requirement concerns the capability of installing the necessary OVF communication channels. They are responsible of transmitting the mentioned Cloud data to the correct Forensic Data Base component for the collection activity. From an organizational and legal perspective, these communication channels require proper permissions and SLAs clauses for data exchanging. In this way all the involved Cloud actors will be warned about data and personal information transmissions, thus avoiding privacy violation actions. More technically, the installation of the OVF communication channels will involve the Virtual Machines used by the Cloud data centres: proper setup files will be used for setting up actions and for implementing the mentioned procedures.

The proposed CFRS is also capable of providing to the competent investigation bodies, who can be Law Enforcements, Court room people, or Private Investigations units, some reports about the case. They can be in the form of Chain-of-Custody report (CoC) [15], where every single action on the Cloud data and on the digital evidences, is recorded. Several information related
to the actions happened must be recorded; e.g., date, time, time zone of an action; data owner, machine and VM instance, cloud service involved. For each action a description is necessary, together with the author, the copy instance, and an ID on which the action happened. Other report formats may be required, and in this case proper CFRS modules can be added.

### 5.1 A CFRS Example using Open Stack

In this section an example of the proposed CFRS reference architecture is provided. The chosen hosting Cloud Service Provider is the OpenStack project. It provides the ubiquitous open source cloud-computing platform for public and private Clouds, released under the terms of the Apache License [16]. OpenStack was conceived as a joint project between NASA and Rackspace. It is a widely used platform for private Cloud instances; indeed some of its users include many large organizations, e.g., Intel, Argonne National Laboratory, AT&T, Rackspace, and Deutsche Telekom. OpenStack has also APIs compatible with commercial Cloud offerings such as Amazon EC2 and S3.

In Figure 3 shows the OpenStack Cloud architecture. It involves several components, each dedicated to specific operations. The main modules are eight: Nova, the compute platform and cloud controller; Swift, the object storage system; Glance, the service for managing disk images; Keystone, the identity service; Horizon, the web-based dashboard for managing OpenStack services; Neutron, the provider of network services for virtual devices; Cinder, the persistent block storage to running instances; Ceilometer, the monitor and meter of the performances for billing, benchmarking, scalability, and statistical purposes.

The proposed CFRS reference architecture can be deployed in OpenStack as described in Figure 4. The Forensic Data Base module needs to communicate with the eight OpenStack modules through OVF channels, represented by the bold arrows in green. Through these mechanisms the Cloud data will be collected located in the correct Forensic Data Base sub-modules.

The Readiness operations described in Figure 2 will be executed by the dedicated CFRS components, outside the Cloud architecture, without altering the OpenStack services, behaviours, and routines. From these premises the CFRS reference architecture represented in both Figures 1 and 2 can be thought as deployable independently from the running Cloud Services, thus respecting the flow of the events and the customer resources usages.

From the example, we can affirm that the requirements mentioned in Section 5 are respected: the required artefacts and tools exist and the OVF communication channels can be established. The remaining concern is related to the changes to make to the SLAs clauses, for respecting the customers data privacy and the Services Levels guaranteed by the providers.

5.2 Discussion

The implementation and the usage of such a CFRS are very important for multiple purposes. For instance, a Provider can exploit the Forensic Readiness routines to improve customers’ data privacy and its internal security, because major control and monitoring will be performed for protecting critical information [5].

A Cloud organization or a CSP becomes prepared for managing possible incidents, and the manner of gathering enough digital evidences minimizes the effects on the organizations routine operations [22, 26].

The CFRS can also furnish a better way for demonstrating the respect of the Service Level Agreements clauses that are responsible to guarantee the accorded quality of the services.

Moreover, a CFRS adoption can be helpful in order to address some CF challenges described in Section 1. For instance, from a technical perspective, a manner for aligning the multiple log files formats and for synchronizing the several machines timestamps can be implemented in the Data Management sub-system, otherwise reconstructing a correct and reliable events’ timeline could not be feasible.

From an organizational perspective, the CFRS can be considered as a valid approach for assigning the necessary roles and responsibilities for managing the Cloud incidents to Cloud people. For instance, profiles like Investigators and Incident Handlers can be responsible of the information needed and produced by the Readiness Core Module, thus more aware about the security condition of the Cloud system.

Finally, from a legal point of view, the CFRS can highlight the main issues regarding the jurisdictions borders: the proposed system can become the instrument for alerting the proper governments’ offices, helping to address a more general and wide problem.
6. Conclusions and Future Work

Cloud Forensics is an emerging topic with the goal of performing digital investigations in the Cloud architecture. CF is facing with several challenges that derive from the specific CC features, suggesting that important adaptations to investigation procedures and instruments are necessary for the Cloud.

In this paper, an examination of the Cloud Forensic Readiness topic is provided. It is considered as an approach for providing more security to the architecture, even though its main concern is facilitating the criminal investigations and the related cases resolutions. It is meant for detecting the potential sources of digital evidences that can be proactively collected from the Cloud and that will be subsequently manipulated by an external system for producing reliable and admissible digital evidences.

CFRS is designed and its reference architecture is given; its components are illustrated and discussed; also an example of its deployment on an existing Cloud architecture is provided.

Addressing some identified CF challenges should allow us to perform both re-active and pro-active Forensics investigations in the Cloud. In fact, our research direction deals with two main challenges. With the former, a structured process model for conducting re-active Forensics investigations will be provided; it requires a systematic review of the existing investigation process models, in order to understand their common features and phases, and produce a process model tailored for the Cloud. The design of such a model will include a set of phases and possible sub-phases, a set of involved CF actors, and a collection of scenarios for the Cloud crimes. For the latter, which concerns pro-active CF, we will design and prototype a CFRS starting from the presented reference architecture. We will investigate in which part of Cloud architecture our system can be implemented. We will attempt to customize such reference architecture, likely commencing with an open source Cloud platform. We will also investigate whether the existing Cloud data included in our proposal are actually suitable for being manipulated by the OVF data exchange module. We will establish the most suitable knowledge extraction procedures to adopt for Forensics Readiness purposes, i.e., the data mining techniques.

Last but not least, in this research direction, further investigations are necessary. They involve the design of the correct behaviour to be performed by the Events Alerting and the Intrusion Detection system components, included in our proposal. Some definitions of how a Cloud crime can be defined are necessary. This information can be provided by several sources; from the CSPs on the basis of past arisen cases. Also the analysis of some SLAs clauses can be helpful for the same purpose; these clauses might be the ones regarding the corrective measures to undertake in case a service level is not respected or violated. At the light of this information, the literature has to be examined for filling the gaps, and some Cloud crimes definitions, likely in terms of XML configuration files, will be produced. Those data will be integrated into the mentioned CFRS Intrusion Detection component. Proper reactive policies will be designed and included in the Events Alerting module. These formats will be carefully specified because they have to be at a higher level, thus meaning that most of the known Cloud incidents and services violations must be covered.

References


http://www.dmtf.org/standards/ovf


Reference Architecture for a Cloud Forensic Readiness System

Figure 4: OpenStack Cloud Architecture furnished with CFRS components