A Cloud Forensic Readiness Model for Service Level Agreements Management

Lucia De Marco, Filomena Ferrucci, and M-Tahar Kechadi

1School of Computer Science and Informatics, University College Dublin, Ireland
2Department of Management and Information Technology DISTRA MIT University of Salerno, Fisciano, Italy
lucia.de-marco@ucdconnect.ie
fferrucci@unisa.it
tahar.kechadi@ucd.ie

Abstract: Cloud computing is increasingly becoming a target of cyber-criminal attacks. Often the committed crimes violate the Service Level Agreement (SLA) contracts, which must be respected by all the involved parties. Cloud Forensics is a branch of Digital Forensic discipline dealing with crimes involving the Cloud. A manner for leveraging some of the attacks is the provisioning of a Forensic Readiness capability, by performing some activities before the crimes happen. In this paper we introduce a model aimed to represent the management of SLAs through a cloud system.

Keywords: Cloud Forensics Readiness, Service Level Agreements, Cloud Monitoring, SLA Model, SLA Management

1. Introduction

Cloud Forensics (Ruan et. al 2011) is the part of the Digital Forensic Science (Palmer 2011) dealing with the use of scientific approaches to investigate Cloud Computing platforms (Mell et. al 2011). Unfortunately, the current forensic techniques are not suitable for cloud-related crimes due to the challenges arisen by the Cloud physical structure and computing model (Birk et al. 2011, Mishra et al. 2012, NIST 2014, Reilly et. al 2011, Ruan et al. 2013, Reilly et al. 2010) (see Table 1). Cloud services are regulated by Service Level Agreements contracts (SLAs) (Mell et al. 2011), where some constraints among a cloud service provider and its customer(s) are detailed; the contracts are co-signed by the parties and they have legal validity in case of court litigation (Baset 2012, Patel et al. 2009). Because of cloud SLAs importance and contents (CSA 2013), they have been carefully monitored and verified: some attempts have been presented in literature (Czajkowski et al. 2002, Skene et al. 2007, Paschke et al. 2008, Unger et al. 2008, Ishakian et al. 2011, Gosh et al. 2012, EU 2013), also in the forensic community (De Marco et al. 2014), all devoted to better manage SLAs violations. We consider the implementation of a Digital Forensic Readiness capability (Tan 2001, Rowlingson 2004, De Marco et al. 2013) for the cloud as one of the best solutions to render the cloud more reliable. Several case studies can corroborate the issue about SLA automatic management (Amani et al. 2010, Brandic et al. 2010, Maurer et al. 2012, Emeakaroha et al. 2012a, Cedillo et al. 2014), concerning security, law, ethics.

In this paper we present a formal model for SLAs management in cloud forensic readiness. Such model includes a set of rules for input, output, and intermediary computations. This is followed by the design model of cloud forensic readiness system architecture. In the next Section we present some related work about SLAs formal representation and monitoring.

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 Service</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>X</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>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of standard</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple log formats</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No timestamps synchronization</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No routing information</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of investigation expertise</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inappropriate legal measures</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi - tenancy</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Related Work

Several researchers proposed both formal models and some monitoring tools for SLAs. In Czajkowski et al. (2002) a protocol for negotiating SLAs among several actors is designed, which includes different types of SLAs.

In Skene et al. (2007) the SLAs are formalized using set theory, where the concepts of actors, events, parties, and actions requirements are modelled for the purpose of determining the SLAs’ degree of monitor-ability. In Paschke et al. (2008) a framework called Contract Log is presented, which consists of different types of rules
abstracting the contractual contents; they are derivation, reaction, integrity, and deontic rules. All these rules are included in a homogeneous syntax and knowledge base. In Unger et al. (2008) the concepts of Parties, SLA parameters, and service level objectives are used to formalize SLAs for Services Oriented Architecture; the aim is to provide a manner for aggregating more SLAs in a single business process. The same concepts are used in Gosh et al. (2012), but modelled only by tuples. Ishakian et al. (2011) also uses tuples to address the issue of verifying efficient workload co-location of real-time workloads; he applies some transformation rules to compute the target SLA. Moreover, about thirty research projects were granted by the EU commission and their results are reported in (EU 2013). All these projects concern with automating different aspects of SLAs lifecycle. Another formal model utilizing set theory, predicates, and tuples is presented in De Marco et al. (2014) with the purpose of comparing activities changing Cloud resources-status with some SLAs constraints. A framework called DESVI (Emeakaroha et al. 2010) is dedicated to monitor low-level cloud resources metrics, in order to detect whether their values respect the constraints extracted from some SLAs clauses, with the principal goal of detecting QoS violations. In Morshedlou et al. (2014) SLA violations are considered from the user satisfaction point of view; thus, the main goal is to implement proactive resources allocation for reducing the negative impact of SLAs’ violations and improving the level of users’ satisfaction. In order to prevent clauses violations and the related cyber-crimes, Amani et al. (2010) propose to add a phase dedicated to the evaluation of the contract itself to an SLA life cycle. Finally, some other researches are aimed to manage all the SLAs life cycle phases, namely the services requirements definition, the negotiation between providers and customers, the resources brokering, to the matchmaking and the monitoring of the agreed terms and conditions (Comerio et al. 2014, Wieder et al. 2011).

3. SLA Management in Cloud Forensic Readiness

A Cloud Forensic Readiness capability for SLA management (SLACFR) is aimed to record some information about the cloud behaviour with respect to SLAs. Such information is structured as a set of events, responsible of tracking the relationship between an attribute of a cloud entity and a (set of) constraint(s) on that attribute at time t. The input of such capability is composed of both information about some Cloud attributes and some SLA constraints, all represented with formal rules, as we describe in Section 4. The action flow of the capability depicted in Figure 1 begins on the availability of the contract(s) to monitor. The text is properly parsed via information extraction techniques (Grisham 1999) and transformed into a set of formal rules without repetitions. The cloud information is gathered from some services monitoring tools; they can represent network traffic information, resources statuses, file exchange operations tools, resources allocation, and measures facilities. The forensic readiness capability is responsible to record and output some events representing the matching between cloud attributes and the related SLA clauses. In particular, the capability recognizes in a real time manner whether some suspicious events happen: they represent a violation of a contractual constraint, called hypothesis. Such events recording represent some pre-investigative activities.

![Figure 1: SLA Management in Cloud Forensic Readiness](image-url)
4. Formal Model

The SLA Management Cloud Forensic Readiness (SLACFR) formal model is aimed to provide a theoretical approach to represent the management of SLAs contracts for cloud computing services in the context of forensic readiness. It is necessary to mention that a previous draft of the same formalism was discussed in De Marco (et al. 2014), where only a specific scenario has been taken into account. We extend the formal model by structuring the entities in a different manner and by adding some missing information, such as cloud architecture components. The formal model utilizes mathematical formalisms as tuples, set theory, functions (Ben-Ari 1993). Through these formalisms we provide a set of definitions about the entities manipulated by the architecture illustrated in Figures 1 and 3.

4.1 Cloud Computing

A cloud computing architecture (Mell et al. 2011) is object of monitoring activities by the forensic readiness capability discussed in Section 3. Let C be the set of cloud architectures.

\[ C = \{c_1, c_2, c_3, \ldots, c_i\}, \quad i \in N \]

A cloud architecture \( c \) is composed of at least 2 data centres, then \( c = D \), where \( D \) is the set of data centres.

\[ D = \{d_1, d_2, d_3, \ldots, d_j\}, \quad j \in N; \quad |D| > 1 \]

A data centre \( d \) is an entity containing one or more physical and virtual machines; the data centres are connected each other, composing networks. Let \( d \) be a data centre belonging to the set of data centres \( D \), it is described by the tuple

\[ d = (P, V, vmm, N^d); \quad N^d \subseteq CN \]

Where \( P \) is the set of physical machines, \( V \) of virtual machines, \( vmm \) is a Virtual Machine Monitor, and \( N^d \) is the set of data centre \( d \) is connected with. \( N^d \) is subset of the set of all the connections \( CN \). Each element of the set \( N^d \) describes the pairs of data centres connected each other.

\[ CN = \{n_1, n_2, n_3, \ldots, n_k\}, \quad k \in N; \quad n = (d_j, d_j) \]

Lemma

In a cloud computing architecture there exist at least 2 data centres such that \( |CN| \neq 0 \)

Proof

By definition \(|D| > 1\), then at least \( D = \{d_1, d_2\} \), where \( d_1 = (P, V, vmm, N_{i1}; n_1 = (d_1, d_1) \)

\( N_{i1} = \{n_1\}; \quad n_1 = (d_1, d_1) \)

\( N_{i2} = \{n_1\}; \quad n_2 = (d_1, d_1) \)

\( N^d \subseteq CN \) and \( N^d \subseteq CN \) \( \Rightarrow CN = N^d \subseteq CN \Rightarrow CN = \{n_1, n_2\} \Rightarrow |CN| 
eq 0 \cdot \)

Both physical and virtual machines are composed of a set of resources \( R \), either software or hardware.

\[ P = \{R^i / R^i \subseteq R\}; \quad V = \{R^j / R^j \subseteq R\}; \quad R = \{r_1, r_2, r_3, \ldots, r_j\}, \quad j \in N \]

A resource \( r \) is described by a set of attributes \( A \) and their values, e.g., filename, date of creation, size. Let \( a \) be an attribute, it belongs to the set of attributes \( A \).

\[ A = \{a_1, a_2, a_3, \ldots, a_j\}, \quad j \in N \]

\[ r = \{A^i / A^i \subseteq A\} \]

During the execution of a cloud service, the value of a resource attribute is subject to changes via operations. An operation is described by a mathematical tuple composed of a sender \( s \) triggering it, an operation result \( value(r, a) \), an operation resource \( r \), an attribute \( a \), and an operation time \( t_r \). The operation \( value \) is a mathematical function that changes the value of an attribute of a resource at the time \( t-1 \) into another value at the time \( t \). Such value is expressed with a unit measure \( u \). A sender \( s \) is an entity performing operations in the cloud, either a human or a system process, e.g., an Internet session or a Dropbox process. Let \( s \) be a sender, it belongs to the set of senders \( S \).
\[ S = \{ s_1, s_2, s_3, ..., s_n \}, \quad h \in N \]
\[ o = (s, a', \text{value}(a'), u, t_o) \]
\[ \text{value}: a'_{i+1} \rightarrow \text{value}(a') \]

4.2 Service Level Agreement
As aforementioned, SLAs are contractual documents written in natural language composed of a set of information structured as clauses. In our formal representation, an sla \( i \) is an element of the set of slas \( L \). An sla is described by a mathematical tuple composed of a set of hypothesis \( H \), the validity starting time \( t_{\text{start}} \) and the validity ending time \( t_{\text{end}} \).

\[ L = \{ l_1, l_2, l_3, ..., l_p \}, \quad i \in N \]
\[ \text{value}: a_{\text{rt}} \rightarrow \text{value}(a) \]

A hypothesis \( h \) is an element of the set of the hypothesis \( H \). A hypothesis is described by a mathematical tuple composed of the conditioned value \( c_k u \) of the attribute \( a' \), the hypothesis validity starting time \( t_s \) and the hypothesis validity finishing time \( t_f \). A condition \( c \) belongs to the set of conditions \( C \); a value \( k \) is related to the \( a' \) attribute; \( u \) is the unit measure used to express the value \( k \), belonging to the set of unit measures \( U \); the conditioned value \( c_k u \) has to be verified during the time interval \( t_f - t_s \). In case no condition is expressed the default value is set as ‘=’.

\[ H = \{ h_1, h_2, ..., h_j \}, \quad j \in N \]
\[ h = (c_k u, a', t_s, t_f) \]
\[ t_f \geq t_s \]

4.3 Forensic Readiness
A cloud forensic readiness capability \( f \) is an element of the set of forensic readiness capabilities \( F \). A capability is described by a mathematical tuple composed of a set of forensic readiness events \( E \) and a set of SLAs \( L \).

\[ F = \{ f_1, f_2, f_3, ..., f_p \}, \quad p \in N \]
\[ f = (E, L) \]
\[ E \neq \emptyset \]
\[ L \neq \emptyset \]

An event \( e \) belongs to the set of events \( E \). An events is described by a mathematical tuple composed of an operation \( o \), a hypothesis to verify \( h \), and the match time \( t_e \).

\[ E = \{ e_1, e_2, e_3, ..., e_q \}, \quad q \in N \]
\[ e = (o, h, t_e) \]
\[ t_e \geq t_o \]
Given an event \( e \in E \) and a hypothesis \( h \in H \), the event is considered suspicious if the value of the operation \( o \) on the attribute \( a \) of the resource \( r \) at the time \( t \) is different from the conditioned value \( c_k u \) of the related hypothesis \( h \) about the same attribute \( a \) of the same resource \( r \); the validity has to be determined during the correct time interval, namely the time of the operation assigning the value to the attribute \( a \) of the resource \( r \) \( t_s \) has to be included in the time interval of the aforementioned hypothesis \( t_f(h) - t_s(h) \).

\[
\text{value}(a') \neq c_k u \quad t_f(h) \leq t_s \leq t_f(h)
\]

### 4.4 Example

The aim of this section is to illustrate with a conceptual example how an SLA violation is detected by the forensic readiness capability. The SLA clause suitable to help the reader to understand the SLACFR behaviour is derived from the public SLA for Amazon S3 (Amazon 2015):

“Amazon Web Services will use commercially reasonable efforts to make Amazon S3 available with a Monthly Uptime Percentage [...] of at least 99.9% during any monthly billing cycle.”

We assume that dedicated information extraction techniques have been properly designed; they are required to populate the sets defined in Section 4.2 by correctly detecting the textual elements, and producing the following results.

**Resources.** The set of resources is populated as it follows: \( R = \{ \text{AmazonS3}\} \).

**Attributes.** The set of attributes is populated as: \( A = \{ \text{available} \} \).

**Unit Measures.** The unit measure \( u \) is not expressed, but we can correctly store the symbol ‘%’ because it will be necessary during the computation of the match among operations and hypotheses, namely the events. The set of unit measures is populated as: \( U = \{ \% \} \).

**Hypotheses.** The population of this set is a combination of the information already available, i.e., the elements stored in \( R \) and \( A \). By scanning the text using as keywords the elements of \( R \) and \( A \), we obtain that the conditioned value \( c_k u \) of the corresponding hypothesis is ‘\( \geq 99\% \)’. The condition \( c \) of the hypothesis is ‘\( \geq \)’ because in the text the condition is expressed as ‘at least’. Finally, the starting and finishing times are derived by the text; in particular, both the starting and finishing time are calculated through the duration expressed in the text; because no particular hour constraint is written but the month is considered in a billing context, then the procedure assigns midnight as default.

\[
h_1 = (\geq 99\%, \text{available}^{\text{AmazonS3}}, 01/01/2015-00:00, 01/02/2015-00:00)
\]

**Senders, Operations, Events.** SLACFR has access to network channels and the Virtual Machines Manager of the underlying cloud platform. Such information is fed to the dedicated parsers (as depicted in Figure 1), which outputs the set of senders \( S \) and operations \( O \). The element \( IPa_1 \) of the set \( S \) is a short form for \( IP\text{address}_1 \), which we assume is the IP address detected on the monitoring tool communicating with AmazonS3. A set of operations is then translated, and some events are generated by the engine FR Events Generator (see Figure 3).

\[
S = \{ IPa_1 \}
\]
\
\[
O = \{ o_1, o_2, o_3, o_4 \} \\
\quad o_1 = ( IPa_2, \text{available}^{\text{AmazonS3}}, 99\%, 02/01/2015-15:00) \\
\quad o_2 = ( IPa_2, \text{available}^{\text{AmazonS3}}, 98\%, 05/01/2015-17:00) \\
\quad o_3 = ( IPa_2, \text{available}^{\text{AmazonS3}}, 99\%, 10/01/2015-09:00) \\
\quad o_4 = ( IPa_2, \text{available}^{\text{AmazonS3}}, 99\%, 05/02/2015-15:00) \\
\]
\
\[
E = \{ e_1, e_2, e_3 \} \\
\quad e_1 = (o_1, h_1, 02/03/2012-15:00) \\
\quad e_2 = (o_2, h_1, 05/01/2015-17:00) \\
\quad e_3 = (o_3, h_1, 10/01/2015-09:00) \\
\quad e_4 = (o_4, h_1, 05/02/2015-15:00)
\]
The FR Events Generator component is also capable of properly labelling the events as suspicious or not, and to discard them depending on the time they are recorded at. In this case, the system will recognize event $e_z$ as suspicious because the conditioned value of the hypothesis is violated, indeed it is 98% while it is supposed to be 99%; similarly, event $e_x$ is discarded because recorded on a time outside the time interval established in $h_z$, namely after 01/02/2015-00:00.

5. SLACFR Architecture and Requirements

In this section we describe an SLACFR capability model, by highlighting what are the inputs, processes, and outputs of the various components described in Figure 1. The diagram in Figure 3 uses the Business Process Model Notation (BPMN); it has been considered useful due to its capability of differentiating among independent functionalities, represented as business processes, each included in a swim-lane, i.e., the horizontal rectangles; the sub tasks necessary to accomplish a single purpose are located in the swim lanes. In the BPMN model for SLACFR we have three swim-lanes, where the topmost and the bottommost are responsible to compute the input necessary to the middle one, i.e., from the SLAs and the cloud infrastructure, respectively. Since a specific time, in the bottommost swim lane the monitoring tools output are acquired from a cloud platform and translated to obtain the formal rules about senders and operations, properly encrypted in order to respect some forensic best practices (CSA 2013, Grobler et al. 2013). The topmost swim lane, called Formalism Generation, is fed with the SLA documents. An initial engine implements the algorithm described in Figure 2 and transforms the contractual natural-language constraints to rules according to the formal model proposed in Section 4. Possible rules repetitions are reduced by the rules reduction sub-task. In the central swim lane, named Forensic Readiness, the detection of SLA-violation set of events is executed. The module called FR Events Generator is responsible of generating and storing the recorded events as soon as they happen, paying special attention to the suspicious events, responsible of identifying a contractual violation. It is not clear when such CFR business process has to terminate, because suspicious events can be recoded again and again. For the moment, we assume that the capability keeps running until the cloud organization considers it necessary for assuring more reliability to its services.

A system implementing such FR capability is dealing with computer security issues, and very likely managing private users’ data; the main concerns are about the confidentiality of the information themselves. It is worth reminding that a CFRS will be deployed as an additional system, external to the Cloud infrastructure but capable of communicating with the necessary components to acquire the information from the Cloud (De Marco et. al 2013). Some dedicated, encrypted and secure communication channels are necessary at this point. Independently from the chosen technology, a reliable communication protocol must be adopted, allowing encrypting and decrypting the shared information. Moreover, the communication channels endpoints might be secured by sophisticated mechanisms. The system must be reliable in terms of accessibility and data consistency; network connection falls cannot be admitted, as well as not updated or incomplete persistent storages. In terms of performance, the system must be accessible 24h/7, due that crime prevention can happen in every moment of the day, every day; moreover, it should be required to manage big amount of data, and consequently be scalable from this point of view. Nevertheless, the network usage has to avoid channel saturation; indeed a proper medium connection is necessary, for instance optical fibre. The system has to be delivered and packaged in a multi-platform compatible manner, because it is not known at priori what the environment to interface with can be. Last but not least, it has to face with some investigative best practices. For instance, the investigation principles respected by the US National Institute of Justice (USI 2011), by the UK ACPO (2007), the ones about data integrity (SWDGE 2006), and the Forensic Standard ISO IEC 27037 (CSA 2013) are widely adopted by the international community and considered necessary to be respected in case of dealing with potential forensic evidence.
6. Discussion

SLACFR is devoted to implement a forensic readiness capability for SLA-related cloud crimes. The proposed approach concerns the representation of information from both the SLAs and the cloud in a specific formal model (see Section 4). The entities sender, operation, and attribute are necessary to reconstruct events timeline for forensic investigations. The resource attributes included in the operation entity have to match with the resource attributes expressed in the hypothesis to record forensic readiness events. The value of the operations on the attributes is compared with the conditioned value expressed in hypotheses concerning the same attributes. This match has to be performed by the FR Events Generator module of the architecture designed in Figure 3. Very likely, the actual source of cloud information can be represented by some cloud services logs. Such data can be gathered from dedicated logging tools included in the services furnished with a forensic readiness capability. If no logging tool is present, then an alternative information source needs to be identified. Is it worth to mention that log information is not structured in the same manner as expected by the formal model; then log information will be converted by a dedicated FR system component, to be added to the model in Figure 3. In order to calculate an operation resource attribute value, some metrics might be necessary. They compute the measure, necessary to the comparisons module, thus included in the system since its installation. Once the metrics for a resource attribute is applied and the value calculated, then the match with the value to be respected in the contractual constraint can be performed. In case the metrics are established by the contract itself, then such information has to be utilized, and the pre-existing metrics included in the system discarded. In this manner, the formal model demonstrates to be extensible on depending on the necessary entities to represent. Because it is at very high level abstraction, it can be enriched with additional information, such as some metrics or legal principles: specific rules can be added, and dedicated system modules interfaced with the existing ones presented in Figure 3.

7. Conclusion and Future Work

The management of SLA contracts in the context of cloud forensics is a very challenging research topic. Several proposals have been made in literature to guarantee QoS levels with the purpose of monitoring platforms behaviour for resources performances respect. In some cases, the issue is modelled and implemented; in other ones, specific frameworks are proposed to demonstrate how the monitoring is performed. In the future, we intend to prototype the SLACFR system for verifying some specific SLA constraints. The main goal is to demonstrate that SLACFR can really detect the exact time when a suspicious event is happening and recorded. Moreover, we want to test the number of SLA violations detected in a specific time. We intend to simulate an attack to an IaaS Cloud service; this choice is led by the necessity of interacting directly with the physical resources level; moreover, PaaS and SaaS are built on top of IaaS.

The concern of the proposed FR capability prepares pro-actively the cloud for forensic, hence, a crime-reaction module is out of the SLACFR duties; nevertheless it can be considered as an extension or integration, capable of analysing a cloud crime scene with some information already available. We strongly believe that such forensic readiness capability can enhance the security strategies of a cloud platform, such that being considered as a must requirement for it, and very likely become a standard in the future.
References


BPMN - Business Process Model and Notation, [online], http://www.bpmn.org/


