An integrated conceptual digital forensic framework for cloud computing

Ben Martini*, Kim-Kwang Raymond Choo

Information Assurance Research Group & Forensic Computing Lab, Advanced Computing Research Centre, School of Computer and Information Science, University of South Australia, Mawson Lakes Campus, SA 5095, Australia

ABSTRACT

Increasing interest in and use of cloud computing services presents both opportunities for criminal exploitation and challenges for law enforcement agencies (LEAs). For example, it is becoming easier for criminals to store incriminating files in the cloud computing environment but it may be extremely difficult for LEAs to seize these files as the latter could potentially be stored overseas. Two of the most widely used and accepted forensic frameworks – McKemmish (1999) and NIST (Kent et al., 2006) – are then reviewed to identify the required changes to current forensic practices needed to successfully conduct cloud computing investigations. We propose an integrated (iterative) conceptual digital forensic framework (based on McKemmish and NIST), which emphasises the differences in the preservation of forensic data and the collection of cloud computing data for forensic purposes. Cloud computing digital forensic issues are discussed within the context of this framework. Finally suggestions for future research are made to further examine this field and provide a library of digital forensic methodologies for the various cloud platforms and deployment models.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Cloud computing has recently become a prevalent technology and is currently one of the main trends in the information and communications technologies (ICT) sector. The 2011 Gartner hype cycle, for example, referred to cloud computing as the “most hyped concept in IT” (Smith, 2011, p.3). “Cloud Computing” has been a trending search on Google since 2009 with continued interest (Google, 2011). Another Gartner report suggested that cloud computing could be a US$149 billion market by 2014 and by 2016 could have 100% penetration in Forbes list of the Global 2000 companies (McGee, 2011). It can be reasonably assumed that many of those top 2000 companies will provide some level of online access via cloud computing to both their internal users and their customers.

Governments worldwide have recognised the advantages of cloud computing for its users, but have acknowledged that there are associated security and privacy risks. The Australian Government, for example, released a strategic direction paper in 2011 to assist government agencies in choosing cloud computing to deliver services where appropriate (AGIMO, 2011). Australia’s Defence Signals Directorate (DSD) has published an authoritative paper for government agencies on the security considerations required when reviewing cloud computing options (Defence Signals Directorate, 2011). The United States (US) government has also placed great emphasis on cloud usage within its agencies. For example, a survey by the United States Government Accountability Office (2010) found that approximately half of the 24 agencies contacted reportedly used cloud computing. These numbers are likely to increase as the US government Federal Cloud Computing strategy has instituted a “Cloud First” strategy to “realize the value of cloud computing by requiring agencies to evaluate safe, secure cloud computing options before making any new investments” (Kundra, 2011, p.2).
Cloud computing presents both challenges and opportunities for government agencies, particularly in law enforcement and national security (CSA, 2010; ENISA, 2011). Cloud computing (like other networked cyber-infrastructure) is subject to frequent attacks by cyber criminals, who may be able to hijack and use them for criminal purposes, hence, adding to the challenge of growing volumes of digital evidence in each specific case under investigation. In addition, cloud services can be used as a launching pad for new attacks or to store and distribute criminal data (e.g. child abuse and terrorism-related materials) by cyber criminals, organised crime groups and politically-motivated actors to avoid the scrutiny of law enforcement and national security agencies (Choo, 2010). The use of cloud computing by criminals (or their victims) means that their devices will be virtualised, geographically distributed and ephemeral, presenting technical and jurisdictional challenges for their identification and seizure by law enforcement and national security agencies. These can impede digital forensic investigators and potentially prevent law enforcement and national security agencies from acquiring digital evidence and analysing digital content forensically in a timely fashion (Garfinkel, 2010). This may also result in the unintended consequence of disrupting the continuity of businesses whose data and information are hosted on the seized hardware by law enforcement agencies (Choo, 2010) – business continuity is one of the key risk areas identified by the Australian Government Information Management Office (AGIMO, 2011).

The lack of recent understanding of the complexity and the challenges in keeping pace with rapid advancement in ICT, including cloud computing, hinders government agencies in criminal and national security investigation. A report by ENISA (2011, p. 51), for example, pointed out the "[d]ifficulties in accessing forensic data [in the cloud computing environment] to determine data linkability and accountability in cases where illegal activities are performed". There is, clearly, an urgent need for law enforcement agencies (LEAs) to adapt and augment technical and procedural digital forensic responses as criminal, state-sponsored or otherwise, who are often early adopters use new technologies in different ways to facilitate crime, both traditional criminal activities (e.g. money laundering) and cyber criminal activities (e.g. unauthorised access and modification of systems and data).

The main contributions of this paper are:

1. Providing a snapshot of the issues faced by forensic investigators with the introduction and increasing usage of cloud computing technologies.
2. Proposing an integrated iterative framework based upon two widely used digital forensics frameworks of McKemmish (1999) and NIST (Kent et al., 2006) – see Section 3.

Section 2 provides a brief discussion and definition of cloud computing, as well as introducing the field of digital forensics using the definitions of McKemmish (1999) and NIST (Kent et al., 2006). Section 3 defines our proposed framework and discusses a range of issues faced by forensic investigators in the context of the four phases of the framework. The last section concludes this paper and discusses potential future directions for research.

2. Forensic analysis in the cloud

2.1. Cloud computing: overview

Cloud computing is an evolving terminology, and as such its strict definition will likely vary over time. This is further complicated by the marketing strategy of branding/labelling web-based technologies as cloud computing, even when there is no clear link between cloud computing and the service. For example many users now refer to services hosted on the internet such as Facebook as “the cloud” (Janjigian, 2011) – a highly contentious use of the term. The (latter) phenomenon is what some may refer to as “cloudwashing” – that is, the use of cloud computing not as a technical description but rather as a marketing term to describe services which have some of the characteristics of a cloud service but often lack the key technical components due to the hype currently surrounding the technology (Smith, 2011). Vaquero et al. (2008, p.51), reviewed over 20 cloud computing definitions and found three key components which were mandatory in a minimal definition: “scalability, pay-per-use utility model and virtualization”; and noted that there are many similarities between cloud computing and “grid computing” (another similar technology which has existed for some time).

In this paper, we use the Australian Government’s (AGIMO, 2011) definition of cloud computing – adopted from the US Government’s National Institute of Standards and Technology (NIST). The latter also appears to be widely used in the literature and has been suggested by some observers to be the “de facto standard” (Grobauer et al., 2011, p.52).

“ICT sourcing and delivery model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (AGIMO, 2011, p.10).

Three common platforms used to describe cloud computing are SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) (Mell and Grance, 2009; AGIMO, 2011; Defence Signals Directorate, 2011).

IaaS (Infrastructure as a Service) is a basic product, which can be offered by a cloud provider. It typically allows a cloud computing user to provision infrastructure (processing time and speed, memory allocations, disk based storage, etc.) on which they can run any software (operating system) of their choosing (Mell and Grance, 2009). In most cases, the user of the cloud service is provided with a virtual machine instance which is configured in terms of CPU, RAM and network access based on the user’s selection in a portal, the user is billed based upon their selections which can be changed at any time (‘pay-as-you-use’ business model).
PaaS (Platform as a Service) provides the user with a platform on which the user can develop and host their applications. Generally this platform consists of an operating system and development tools which are used to develop applications for the platform provided (Mell and Grance, 2009). PaaS services vary in terms of flexibility and billing arrangements depending on provider.

SaaS (Software as a Service) is a popular cloud service where a software environment (e.g. customer relationship management system or email hosting) is provided as a service by the cloud provider. The user of this service generally has no control over the underlying infrastructure or platform (Mell and Grance, 2009).

A range of end users from individuals to organisational entities choose to make use of the various platforms as their needs dictate, although cloud services are often marketed to specific user types. For example, SaaS is generally marketed at end users and small to medium-sized enterprises (SMEs) that may not have the in-house technical expertise to maintain the platform and infrastructure. IaaS is often marketed at larger organisations capable of technically supporting platforms and applications but find cost savings, scalability or other business benefits from having a cloud provider hosting the actual physical infrastructure.

Mell and Grance (2009) defined four deployment models as part of the NIST definition. The two most popular deployment models are the public and private clouds. The choice of the cloud deployment models will almost certainly change the approach used by the digital forensic investigator in acquiring digital evidence. The public cloud model is the most popular model where an organisation such as Amazon makes cloud services available to the general public for a fee. The private cloud model can be used in a large organisation where a single department can build the cloud infrastructure and provide IaaS, PaaS or SaaS services to its internal departments. In a community/managed setting, tenancy can either be single (dedicated) or shared and the IT infrastructure is either managed by the organisation or a third-party cloud service provider. The main difference between hybrid cloud services and other cloud services is that the former “is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability” (Mell and Grance, 2009, p.2).

2.2. Digital forensics: overview

Digital forensics is a relatively new sub-discipline of forensic science when compared to other common forensic science disciplines. Digital forensics has a number of synonyms including forensic computing, computational forensics and computer forensics. To reduce the risk of digital (forensic) evidence being called into question in judicial proceedings, it is important to have a rigorous methodology and set of procedures for conducting forensic investigations and examinations. The development of digital forensics methodologies needs to be built on these sound scientific principles.

McKemmish (1999, p.1) provided one of the first definitions of digital forensics defining it as: “the process of identifying, preserving, analysing and presenting digital evidence in a manner that is legally acceptable”. There are four key elements in McKemmish’s digital forensics framework – namely, the identification, preservation, analysis and presentation of digital evidence.

1. Identification of digital evidence defines the requirement for evidence management, knowing it is present, its location and its type and format. The diversity of devices requiring forensic investigation has grown significantly since 1999. Cloud computing, for example, adds new requirements for investigation. Identification of cloud services used on a seized device may become a critical part of the standard forensic identification process for seized suspect equipment.

2. Preservation is concerned with ensuring evidential data remains unchanged or changed as little as possible. For example if a forensic investigator needs to recover data from a device, it is important that the data is recovered using methods that are forensically sound (e.g. they do not write to the original data source as information such as ‘file last accessed times’ on key system files could be changed if a seized computer is improperly booted before a disk image capture is undertaken).

3. Analysis transforms the bit level data collected in the earlier two phases into evidence presentable to a court of law. In a cloud computing environment, however, the data stored on the cloud (including network data that could be collected when communicating between the client(s) and the cloud and any data stored on the client such as cache data) needs to be analysed.

4. Finally the presentation element is concerned with presenting evidence to the courts in terms of providing expert testimony on the analysis of the evidence. This element is mostly concerned with legal presentation and is generally unchanged regardless of the computing platform being analysed.

Another widely used digital forensics framework is that of NIST (Kent et al., 2006). The four phases and definitions in NIST’s framework share some similarities with McKemmish’s:

1. Collection discusses identifying relevant data, preserving its integrity and acquiring the data;
2. Examination uses automated and manual tools to extract data of interest while ensuring preservation;
3. Analysis is concerned with deriving useful information from the results of the examination; and
4. Reporting is concerned with the preparation and presentation of the forensic analysis (Kent et al., 2006, p.ES-1).

3. Our proposed conceptual framework

A draft report from the National Institute of Standards and Technology (2011, p.64) noted that “little guidance exists on how to acquire and conduct forensics in a cloud
environment” and suggested that existing best practice guidelines still apply to digital forensics in the cloud computing environment. Barrett and Kipper (2010), however, suggested that existing digital forensics methods may not be fit-for-purpose in the cloud computing environment. Other researchers such as Birk and Wegener (2011, p.9) also highlighted that “[c]urrently, guidelines and best practice guides on gathering digital evidence are rare and often outdated. There are no guidelines specific to evidence gathered in the cloud...” and Huber et al. (2011) found little research in this area (e.g. how to extract data from cloud services in a forensically sound manner). Similar observations are echoed by a number of digital forensic practitioners, including the previous Director of US Department of Defence Computer Forensics Laboratory and the previous Chief Scientist at US Air Force Research Laboratory Information Directorate who posited that “[m]ore research is required in the cyber domain, especially in cloud computing, to identify and categorize the unique aspects of where and how digital evidence can be found. End points such as mobile devices add complexity to this domain. Trace evidence can be found on servers, switches, routers, cell phones, etc” (Zatyko and Bay, 2012, p.15).

Data in a cloud storage environment is often distributed physically within data centres. Using traditional methods the procedure to access stored data is to conduct binary searches of the data. In this manner, sources of evidence discovered in the third phase of McKemmish (1999) (described as Extraction and Processing). The four phases of the integrated framework are as follows (see also Table 1):

1. Evidence Source Identification and Preservation: This phase is concerned with identifying sources of evidence in a digital forensics investigation. During the first iteration, sources of evidence identified will likely be a physical device (e.g. desktop computers, laptops and mobile devices). During the second iteration, this phase is concerned with identifying cloud services/providers relevant to the case, possible evidence stored with the cloud provider and processes for preservation of this potential evidence. Regardless of the identified source of evidence, forensic investigators need to ensure the proper preservation of the evidence.

2. Collection: This phase is concerned with the actual capture of the data. There will most certainly be variation in the collection methods in each type of cloud platform and deployment model. For example IaaS part of their collection phase. We, however, suggest that identification of cloud computing use and, more importantly, preservation (in cooperation with the cloud service provider) are the two most critical steps in cloud computing investigations and should be collectively represented – see the first step in our framework. In our framework, we also suggest that the identification and preservation of evidence sources should be conducted simultaneously and as soon as practical (e.g. as soon as the data source is identified, the relevant cloud provider should be contacted to commence preservation).

Although the names and purposes of the phases in our framework are based on McKemmish (1999) and NIST (Kent et al., 2006), the meanings and processes undertaken during the phases may be somewhat different and these differences are explained in the remainder of this paper. The key difference between both frameworks and our integrated framework is the iteration (phase) in our framework – if evidence of cloud computing use is discovered in the third phase – Examination and Analysis, another iteration of the framework will commence simultaneously. The second iteration commences with evidence source identification and preservation via the cloud service provider and if during the examination and analysis of the data collected during this (second) iteration further evidence sources are identified, another (third, etc) iteration would then commence.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Digital forensic framework comparison.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our proposed framework</strong></td>
<td><strong>NIST framework (Kent et al., 2006)</strong></td>
</tr>
<tr>
<td>1. Evidence source identification and preservation</td>
<td>1. Collection</td>
</tr>
<tr>
<td>2. Collection</td>
<td>2. Examination</td>
</tr>
<tr>
<td>3. Examination and analysis</td>
<td>3. Analysis</td>
</tr>
<tr>
<td>4. Reporting and presentation</td>
<td>4. Reporting</td>
</tr>
</tbody>
</table>

*Collection of digital evidence is somewhat embedded in the analysis phase of McKemmish (1999) (described as Extraction and Processing).
Evidence source identification and preservation

This phase is concerned with establishing the information available and its format in order to establish a best preservation of data is of critical importance and hence, the separate collection step.

3. Examination and Analysis: This phase is concerned with the examination and analysis of forensic data. It is during this phase that cloud computing usage would most likely be discovered based upon the examination and analysis of physical devices and this would lead to a second (or more) iteration(s) of the process.

4. Reporting and Presentation: This phase is concerned with legal presentation of the evidence collected. This phase remains very similar to the frameworks of McKemmish and NIST.

3.1. Evidence source identification and preservation

This phase is concerned with establishing the information available and its format in order to establish a best practice method for retrieval of digital evidence. After the identification of the sources of evidence, the immediate preservation of data is of critical importance and hence, identification and preservation of evidence sources are represented as a single phase in the framework.

The cloud computing environment has complicated the identification and preservation of digital evidence (Garfinkel, 2010). For example, an investigator has to first examine other evidence to determine whether cloud services have been used (Beverly et al., 2011), and if so determine which services, how to gain lawful access to the data, what format the data is in, etc. McKemmish (1999) and Kent et al. (2006) placed emphasis on the analysis of physical media (e.g. hard disk drives and USB media) or file operating systems (e.g. Windows, Mac and Linux), which is feasible only if the investigator (and most likely, LEAs) can access and, preferably, seize the hardware for capture. The traditional forensics methodology of seizing hardware to acquire forensic evidence is unlikely to be possible for most LEAs (Reilly et al., 2010), particularly those located outside US, due to the physical location of the cloud computing data centre. In a submission to the House of Representatives Standing Committee on Communications Inquiry into Cyber Crime and its Impact on Consumers, Australian Federal Police’s High Tech Crime Operations noted that “data accessed by an Australian resident from a server hosted overseas may require a Mutual Assistance Request to that foreign jurisdiction” (High Tech Crime Operations, 2009, p.21). Although the investigating (overseas) LEA can formally request for the preservation of the data of interest while court proceedings are undertaken, this may not be presently feasible in a timely fashion using existing legal agreements. In addition, if a suspect user is using a cloud service, the full complement of data stored in the cloud computing environment is unlikely to be stored on the seized hardware. Identification of cloud services used on the seized hardware must be undertaken, which may help to identify and/or locate alternate sources of evidence. For example, artefacts such as cache data and stored login credentials may be stored on the seized hardware and these (artefacts) may contain case relevant evidence including data that leads to the identification of the cloud service providers/services used.

If cloud services are identified, then preservation of data stored in the cloud computing environment must be considered a top priority, and the cloud service provider should be contacted immediately. In the case of a desktop computer or laptop, preservation may be as simple as powering off the hardware. In a cloud computing environment, however, it will most likely require cooperation of the cloud service provider to place a “litigation hold” (or similar ‘freezing’ mechanism) on the account and prevent any further changes to the data. For example, the Australian Cybercrime Legislation Amendment Bill 2011 provides for “preservation notices” available to LEAs (The Parliament of the Commonwealth of Australia, 2011), which may be useful in requesting preservation of communications in cases involving the use of cloud services.

3.2. Collection

While the actual collection of data has not traditionally been a problem for LEAs, it is likely to be a far more involved (and cross border) process when cloud services are used. Most importantly, it may be impossible to physically seize all the servers in a cloud computing data centre due to the amount of hardware involved or due to the data centre being physically located in another jurisdiction. Unless there is a prior legal agreement between the government and cloud service provider, the investigating LEA may not be able to obtain the data from the cloud service provider in a timely manner. When dealing with smaller private data clouds, however, traditional methods for collection such as seizing the equipment for image collection may still be fit-for purpose particularly if the cloud computing data centre is located in the country. In larger public clouds, LEAs would most probably have to rely on the cloud service provider (or an Application Programming Interface (API) provided by the cloud service provider) to provide a copy of the data. This, however, may break the chain of custody requirements that are generally required for forensic evidence to be accepted by a court (Reilly et al., 2010). For example, Western Australia’s Department of the Premier and Cabinet recommended that in order to protect the integrity of the (digital) evidence, the chain of custody register should contain information such as:

- Who collected the evidence;
- How and where the evidence was collected;
- Who took possession of the evidence;
- How was the evidence stored and protected;
- Who accessed the evidence, for what purpose, and the date and time of such access;
- Who removed the evidence, for what purpose, and the date and time of removal and return; and
• Details of any forensic procedures that were carried out (Office of e-Government, 2004, p.15).

Clearly, the above chain of custody requirements would be difficult to achieve in the cloud computing environment. For example depending on the corporate practices and level of cooperation between the LEA and the cloud service provider it may be difficult to determine how the evidence was collected, where the evidence was physically located before it was collected (especially if distributed between data centres) and precisely who handled the evidence before it was handed over to the LEA. Perhaps most importantly it may not be possible for an LEA to verify the forensic procedures used in collecting the evidence to ensure that data from other cloud customers is not mixed during collection and/or modified during or after collection.

Barrett and Kipper (2010) likened the difference between traditional forensics and cloud computing forensics to the difference between traditional hard disk forensics and mobile forensics. For example, a traditional hard disk is a relatively simple device that accepts low level read-and-write commands, whilst a mobile phone is more complicated using high level requests. Cloud computing forensics must take into account both the advantages and disadvantages (discussed below) of “live forensics” – “the means and technique of obtaining artefacts of evidential value from a machine that is running at the time of analysis” (Biggs and Vidalis, 2009, p.2). Live forensics in a typical cloud computing environment may involve exporting data from the target cloud service, for example via an API, while it is still running rather than capturing an image of the entire server which holds the data. In the case of IaaS, this may involve a capture of the virtual disks and current memory; whereas in the case of SaaS, this export would likely be an export of the actual data stored by the application in an appropriate format. Live forensics would also likely include memory and networks data (National Institute of Standards and Technology, 2011), which may not normally be captured during a standard physical hard disk image collection.

The live forensics method, however, raises issues of extraction and preservation for the investigator as LEAs are often separated from the collection process. For example, questions such as whether the cloud service provider is providing all the available data and whether the data remained unchanged (to a level acceptable in a court of law) during the process of collection may not be easily answered. These issues require further research to determine forensically sound best practices and guidelines that would provide LEAs timely access to data stored on cloud services. In many cases, however, live forensics may be the only option available to LEAs. For example, Birk and Wegener (2011) noted that only live forensics could be used to inspect cloud hosted virtual machines which do not have persistent storage or to conduct analysis on virtual machine snapshots. LEAs and the cloud services provider must also be aware of their obligations under the privacy regime in which they operate. In Australia, for example, government agencies and the cloud services providers must ensure they comply with the Privacy Act 1988 (Cth), which regulates how Commonwealth agencies and contracted service providers (regardless whether the providers are based in Australia or overseas) collect, use, disclose, store and destroy the personal information of individuals, and comply with the Information Privacy Principles (IPPs) and National Privacy Principles (NPPs). This could impact the availability of data to LEAs via automated API without human intervention.

Reilly et al. (2010) discussed the role of artefacts (e.g. metadata) in forensic analysis and the (potential) loss of these artefacts when data is collected from a cloud computing environment. If metadata (e.g. creation/modification dates for files, and user ownership logs) were lost during the collection process, this would severely impact the ability of forensic investigators to conduct an investigation to the standard required by the courts of law. Such information, often required to tie a suspect to a file, is especially important in a cloud computing environment where the user is not in physical proximity to the device (in this context, the cloud computing data centre).

Kent et al. (2006) discussed the importance for providers to support digital forensics as part of their information systems planning and the latter is especially relevant to cloud computing providers and users. Their recommendations include centralised auditing and logging (e.g. file access and authentication attempts), data retention (backups, historical activity records, etc.) and file hashing/integrity checking (Kent et al., 2006, p.2–6). All of these considerations could be potential evidence sources in a cloud computing investigation, as well as enhancing an organisation’s forensic readiness. Shields et al. (2011) discussed the potential for settings such as cloud computing to proactively collect data that may be used as digital evidence in the future. While the data storage capacity of many cloud computing environments would make proactive collection of evidence prohibitive, solutions such as the proof-of-concept tool PROOFS (Shields et al., 2011) could be used for the efficient storage of historical data.

“Locard’s Exchange Principle” (as cited in Zatyko and Bay (2012, p.13)) where “every contact leaves a trace…” is relevant to forensic investigations in the cloud computing environment. Although a cloud computing user is unlikely to make physical contact with the cloud servers, their access will likely leave metadata traces on their device (e.g. the laptop that is used to access the cloud service) and generate logs and other data on the servers. The importance of such a data source should not be underestimated as log data, in many cases, may be the only data which can be relied upon for digital forensic investigations on SaaS platforms (Birk and Wegener, 2011).

3.3. Examination and analysis

Once the data source has been identified, preserved and collected, examination and analysis of a copy of the data will be conducted at the laboratory. The copied data can be analysed using standard processes and tools such as Encase® and FTK™. If the data of potential interest is captured from the cloud in a form that is readable by standard forensic tools, then the forensic investigator will only need to conduct analysis on the data collected.
However, if the data was collected in an abstract manner (e.g. data being spread across multiple storage devices using a system such as GlusterFS, or across multiple data centres), then further work will be required to reconstruct the suspect activities from the collected abstract data.

Examination is an important step for data collected from a cloud computing environment as the data is unlikely to be stored and collected in a form which permits immediate forensic analysis. For example, the data may be compressed, encrypted, spread across multiple devices and, therefore, prior to analysing the data, these processes must be reversed as part of this step (Kent et al., 2006). McKemmish (1999) referred to this important step as processing.

It is important to note during a forensic examination that data may also be generated by the cloud service provider in the process of operating the cloud service which could be useful in the analysis of the evidence. For example, Ahmed and Raja (2010) highlighted the importance of using forensic tools to identify unauthorised access to and/or modification of cloud computing users data, and “anti-forensic” behaviour in the cloud computing environment. Metadata and access logs can be modified to remove traces of unauthorised access and other malicious activity. Hence, metadata and other forms of audit data must be properly kept (in accordance with current security and forensic best practices such as those described by Kent et al. (2006)) and made available when requested. Centralised audit logging, for example, allows for the implementation of a more secure and robust logging environment as well as simplified administration should forensic practitioners require access to logs. Otherwise in the absence of auditable metadata for access, evidence rendered from the analysis of the data may not be admissible in a court of law. The Australian Federal Police’s High Tech Crime Operations, for example, noted that “[t]he integrity and admissibility of such data would also depend on the controls and security placed upon it by the third-party vendor” (High Tech Crime Operations, 2009, p.21), and the National Institute of Standards and Technology (2011) also noted that such log information is critical to forensic investigations and should constantly be collected in case it is required.

It is important to understand the type(s) of information or data that can be extracted and/or requested in an investigation involving cloud computing, and hence, further training and tools may be required for forensic investigators to deal with the popular cloud computing data extract formats. For example, in an ideal IaaS case scenario, the cloud computing environment could provide a forensic investigator with an image of the virtual machine for analysis, which would hold all data uploaded by the suspect. However in the case of SaaS, data may only be able to be exported in a proprietary or unstructured manner, which would make examination and analysis difficult unless customised or specific tools are available to read the data format. Hence, translators (utility applications which convert data from the cloud computing exports native format to a format recognisable by forensic tools) may be needed in order to forensically examine cloud computing data (National Institute of Standards and Technology, 2011).

There is, therefore, a need to design tools that would run with proprietary or unstructured cloud format data files or network analysis tools in order to capture the transactions being completed between the cloud service provider and the client in real time.

3.4. Reporting and presentation

This phase is unlikely to require any major changes from current procedure as it is concerned with presenting evidence. Forensic investigators should focus on the technical aspects of forensic investigations, as well as the presentation of evidence to a court. One key consideration for presentation when using cloud computing data as evidence, however, is to ensure that a clear distinction can be drawn between the data generated by cloud service providers in the collection of the case relevant data and data generated/owned by the suspect. This is a critical distinction between undertaking forensic examination in the cloud and traditional digital forensics, as in the former, a number of parties may be involved in collecting evidence (e.g. cloud service providers, Internet Service Providers (ISPs) and Mobile Service Providers (MSPs)).

When a number of parties are involved in the collection of potential evidence, it is critical that chain of custody logs record details such as which individuals have collected/ transferred evidence and any changes that were necessary as part of this process. ISPs and MSPs can provide critical information to a cloud computing investigation regarding ownership of an IP address and after a court order or warrant has been issued, they can assist by capturing network forensics data of a particular internet connection. Connections between the user and the cloud are potential sources of evidence. For example, ISP data can be correlated with cloud service provider logs to link the suspect user to a particular service at a particular point of time (and potentially geo-location). Recognising the importance of expedited preservation of digital evidence, the Australian Government introduced the Cybercrime Legislation Amendment Bill 2011 that includes “a new provisional measure available to enforcement agencies and Australian Security Intelligence Organisation (ASIO) to prevent the destruction of communications” (The Parliament of the Commonwealth of Australia, 2011, p.11) – the latter will be useful for LEAs to ensure that ISP logs are not destroyed for suspect users of cloud services.

Attention must also be paid during all phases to record any evidence that can link a suspect with the data being collected as the issues arising from the lack of physical proximity between the user and the data will need to be mitigated before presentation to court.

4. Conclusion and future work

Few today would challenge the assertion that the era of globalisation has been accompanied by an increase in the sophistication and volume of malicious cyber activities. Criminals, state-sponsored or otherwise, will go to great lengths and constantly seek to exploit new areas, technologies and opportunities (e.g. cloud computing) to manipulate and exploit vulnerabilities and opportunities whether
they be in law enforcement, regulatory, banking, legal, business, economic or online environments. Their only limitation is that of the imagination.

Cloud computing is a business model which presents a range of new issues to digital forensics practitioners and the digital forensics community in general. There is an urgent need for forensic investigators to adapt existing forensic practices and develop an evidence-based forensically-sound framework (and library of digital forensic methodologies for the various cloud platforms and deployment models) that would enable forensic investigators to identify, preserve, collect, examine and analyse data (and/or data remnants or fragments) in the cloud computing environment.

The integrated digital forensic conceptual framework presented in this paper exemplifies some of these issues and provides a basis for further research in this critical area. The framework combines identification and preservation of evidence sources to ensure that preservation of cloud computing data commences as soon as cloud computing use has been identified, this enhancement is required to manage the constantly changing data environment which is inherent in cloud computing. We proposed a separate Collection stage due to the potential difficulties and challenges which may be faced in this increasingly time consuming process (e.g. data distribution both within the data centre in terms of distribution amongst a number of physical devices and distribution of data throughout the world). The iteration of the framework demonstrates one of the key differences in the identification and analysis of evidence sources in a cloud computing environment; the requirement to use client devices such as PCs and phones not only as sources of evidence but as identification sources of cloud computing usage. When cloud computing usage is detected on these devices another iteration of the forensic framework must be undertaken.

While cloud computing introduces new challenges to digital forensics the existing standards and key principles of forensic computing must be maintained. These key forensic techniques are demonstrated in the McKemmish and NIST frameworks and it, therefore, seems most appropriate to build upon existing standards when defining the framework for cloud computing forensic analysis. However the methods of identification, preservation, analysis and presentation will be complicated by the introduction of cloud computing to digital forensics and therefore this field must adapt to the new technological capabilities.

To contribute (1) to a better understanding of the technical challenges and implications of cloud computing environment for digital forensic investigators, and (2) towards the development of the evidence-based forensically-sound framework, two potential future research questions are proposed.

RQ1) What further changes are required to existing forensic frameworks and practices for conducting forensically-sound investigations in a cloud computing environment?

We proposed an integrated (iterative) framework that can be used to conduct digital forensic investigations in cloud computing environment. In our framework, emphasis is placed upon the need for prompt identification and preservation of cloud computing data once cloud computing use is suspected in an investigation. It is noted that it may not be possible to detect cloud computing usage until analysis is completed on client devices. Hence, the framework commences a new iteration starting with the identification and preservation of the cloud computing source once analysis reveals cloud computing use. The cloud computing environment complicates the (previously straightforward) process of collection due to the technical complications (such as distributed file systems), data stored in various jurisdictions, involvement of multiple parties, the proprietary or unstructured data format(s) and lack of physical proximity between the user and the cloud service. The possibility of cloud data migrating between physical locations during its lifetime, as explained by Hay et al. (2011), further complicates the digital forensic process.

While our proposed model serves as the overarching framework for conducting digital forensic investigations in the cloud computing environment, there is also a need to undertake further research to develop a library of digital forensic methodologies that would best suit the various cloud platforms and deployment models. There will most certainly be variation in the way criminal investigation is carried out in each type of cloud platform and deployment model, and hence, LEAs will need to understand the legality around what type of information or data can be extracted and/or requested and what level and type of information will be needed by cloud service providers and the turnaround time from the time of request for data (particularly for cloud service providers based in overseas jurisdictions).

Further research is being conducted in the digital forensics cloud computing environment which will be used to validate the assertions made in this framework. The following phases will be undertaken as part of this research project over a three year period – see Fig. 1.

Phase 1: Conceptual Framework – This paper falls within the framework development phase which consists of reviewing existing frameworks and literature to determine the high level updates required to traditional forensic frameworks to deal with the challenges presented by cloud computing.

Phase 2: Explication Interviews – Explication interviews will build upon the initial work completed in phase 1 by interviewing law enforcement, digital forensics and cloud computing experts and practitioners to determine the pragmatic framework updates required to deal with cloud computing challenges. These interviewees will review the framework developed in phase 1 and make suggestions for improvement to deal with operational issues.

Phase 3: Technical Experiments – Technical experiments will be conducted across a range of cloud computing hardware and software installations to determine the usability of the framework. This will provide practical feedback on the suitability of the various stages in the framework and provide the basis for the best practice (and the library of digital forensic methodologies) recommendations. Both cloud server and client computing
experiments will be conducted to determine the identifiers of cloud computing use and best methods for preservation, collection and analysis.

Server-based experiments can determine the methods of preservation and collection of digital evidence which ensure that all possible metadata and data are preserved and collected using the various methods currently available in cloud server software to achieve these aims. Detailed research into the role of the client (desktop computers, laptops, mobile devices, etc.) and data remnants on the client’s machine (e.g., a SaaS hosted application accessed via the internet would likely leave cache data from the web browser, but an IaaS hosted virtual machine would leave a different set of data on the client that may not be recovered by existing digital forensics tools) is essential when compiling the library of digital forensic methodologies. Findings from the research will help to determine how cloud computing use can be discovered during client analysis and to recovering any cached data stored on the client; and would be particularly important in situations where LEAs are unable to access data stored on the cloud service provider (in other words, the data available on the client may be the only data accessible to the LEAs).

These technical experiments along with the qualitative interviews with digital forensics industry practitioners and cloud computing providers in phase 2 will permit validation of the framework and best practices to provide guidance during the transition to this new and complex technological environment for the digital forensics community.

**Phase 4: Framework Refinement** – Based upon the output of the explication interviews and technical experiments the framework will be refined to become “evidence based”. This could include adding or removing stages in the framework and further defining the stages in the framework as required.

**Phase 5: Validation Interviews** – Validation interviews will be used to ensure the feasibility of the final framework when reviewed by the practitioners who reviewed the framework in phase 2. These interviews will be held in person and via teleconference.

**Phase 6: Finalised Framework** – The finalised framework will be based upon the feedback provided in phase 5.

RQ2) *What are the legal and privacy issues surrounding access to cloud computing data, particularly cross-border legal and privacy issues; and what reforms are required to facilitate access to such data for LEAs?*

Research on cloud computing security and privacy issues in Australia and overseas is still in its infancy, and arguably, a review of legal and privacy issues both within and outside the country (e.g., Australia) needs to be undertaken. For example, it would be necessary to examine the powers of the local LEAs to legally access data stored on overseas cloud services (e.g., messages and files stored in a cloud-based email account) using a suspect or accused person’s username and password sourced from analysis of physical client devices, and examine the approaches undertaken by LEAs in other countries and the way industry practice or other standards for cloud security and privacy are referenced by that enforcement activity.

**Acknowledgements**

The first author is supported by scholarships from both the University of South Australia and the Defence Systems Innovation Centre. The views and opinions expressed in this article are those of the authors alone and not the organisations with whom the authors are or have been associated/supported. The authors thank Eoghan Casey (Editor of this journal) and the anonymous reviewers for providing constructive and generous feedback. Despite their invaluable assistance, any errors remaining in this article are solely attributed to the authors.

**References**

AGIMO. Cloud computing strategic direction paper; opportunities and applicability for use by the Australian government. Canberra: Commonwealth of Australia; 2011.


