Evidence and Forensics in the Cloud: Challenges and Future Research Directions

Kim-Kwang Raymond Choo, University of Texas at San Antonio
Christian Esposito and Aniello Castiglione, University of Salerno

Although it is popular with companies and private users, cloud computing can be abused or targeted by criminals. This can range from stealing personal information stored and outsourced to the cloud, to frauds that are more sophisticated, and to attacks that are disruptive, such as compromising a company’s day-to-day operations. Cloud storage services can also be abused by criminals, who use it to store and hide incriminating and illegal materials or to distribute copyright materials.¹

There have been several concerted efforts by cloud service providers to prevent their services from being criminally exploited. For example, Dropbox has implemented a child abuse material detection software, whose details are not publicly available, which allows searching within the files stored on Dropbox to identify breaches of the Terms of Use and Acceptable Use Policy. Similarly, Microsoft’s PhotoDNA is designed to identify child abuse materials from the files stored by companies on their servers, and used in its cloud storage product.

Another commonly seen criminal exploitation of the cloud is to support the execution of large-scale and distributed attacks, for example by compromising some instances of virtual machines within a cloud infrastructure to launch Distributed Denial-of-Service (DDoS) attacks against third-party websites, portals or platforms. In 2012, a group of cyber-criminals exploited the CVE-2014-3120 Elasticsearch 1.1.x vulnerability, in order to compromise virtual machines within Amazon EC2, and launched a UDP based DDoS attack. Predictably, most cloud service providers have platform-wide DDoS protection systems that monitor incoming and outgoing traffic in order to prevent DDoS attack against their platform or to avoid being used to launch such attacks.

A number of other security solutions have been proposed for the cloud in the literature, ranging from access control to crypto primitives to intrusion detection to privacy-preserving, and so forth. Despite the existence and deployment of various security solutions, there will be times where digital investigation is needed. As noted in a previous column,² to successfully prosecute individuals who commit crimes involving digital evidence, one must be able to gather evidence of an incident or crime that has involved cloud servers as well as the client devices that have been used to access the cloud services, a process known as digital forensics (or cloud forensics).
Cloud Forensics

In a cloud forensic investigation, it is necessary to analyze the data flow, commonly at three main stages, data-at-rest on the client device(s), data-in-transit, and data-at-rest on the server(s). Therefore, it is important to conduct static analysis and dynamic (binary code) analysis of apps installed on the client device, analysis of data communication and exfiltration channels and techniques, and investigation and validation of techniques to locate and recover public and private keys, authentication tokens, encrypted blocks, and other data of interest in the network traffic and on the client device and server (e.g. memory dumps). For example, a number of researchers have examined the potential to recover data remnants from client devices, such as Android and iOS devices, that have been used to access cloud services (such as the potential to recover forensic artefacts from an OS X PC after it had been used to access Apple’s iCloud). In a recent investigation of the implementation of the OAuth protocol, a commonly used token-based authentication system in mobile apps, the researchers demonstrated how one can intercept and recover security tokens (e.g. access and refresh tokens used to authenticate the user) from the device’s memory heap. This would allow forensic investigators having obtained the security tokens to access a user account even after the user has changed his/her password-based credentials (depending on the service provider’s implementation).

Data may not initially be in a format appropriate for collection as digital evidence, and as such, it becomes necessary to “decode” the protocol used by the application or operating system for data storage and/or transit. Thus, it is important to conduct a comprehensive, empirical investigation of a range of client devices and cloud servers against existing techniques and commercial and open source digital forensic tools, in order to make a detailed determination of the limitations of existing techniques and forensic tools when collecting data from client devices and cloud servers.

It is expected that such technical investigations will clearly demonstrate the strengths and weaknesses of current techniques and the various forensic tools in terms of their evidential data collection and analysis capabilities. It may also identify types of evidence available on computing devices that forensic investigators would not have otherwise known were available.

However, existing techniques may not be applicable in cloud forensics. For example, investigators may not have physical access to the evidence, and a corrupted insider from the cloud service provider can easily alter the evidence. Roussev et al. also noted that in software as a service (SaaS) forensics, “...the use of traditional forensic tools results in acquisition and analysis [that] is inherently incomplete.”

Infrastructure such as distributed filesystems can support Infrastructure as a Service (IaaS) and other cloud computing environments by providing data fragmentation and distribution, potentially between countries and within datacentres. This results in significant technical, jurisdictional and operational challenges in the collection of evidential data for analysis in both criminal investigations and civil litigation matters. For example, a British barrister and a Senior Policy Advisor and Crown Advocate with UK Government Crown Prosecution Service predicted that the evidence obtained from the cloud will play a more significant role in the foreseeable future.

In addition, as explained by Martini and Choo, investigators must trust the cloud service provider to maintaining trustworthy logs about the cloud activity, and providing reports about the activities of user(s) of interest upon request (e.g. a court order). Zawoad, Dutta and Hasan presented a solution for logging the activities within the cloud, and there have been several concerted efforts by cloud service providers to prevent their services from being criminally exploited.
to ensure the integrity and confidentiality of such logs. Specifically, they propose a Proof of Past Log (PPL) scheme to avoid tampering of the logs after their generation, and to encrypt some crucial information within the logs so as to protect the user’s privacy. The proposed solution also facilitates the presentation of the collected evidence for verification in the court.

Dykstra and Sherman described a method to collect forensic artifacts from Amazon’s EC2 service. They also used Eucalyptus (which operates similarly from a client point of view to EC2) for the purposes of injecting forensic tools into running VMs via the hypervisor layer. Using conventional forensic tools (such as Guidance Software EnCase and AccessData FTK), the authors were successful in collecting evidence from EC2 and Eucalyptus. The level of trust required to execute each of the collection procedures was also reported in the study. In a later work, the same authors contributed a forensic toolkit for the OpenStack cloud platform – FROST. FROST allows a remote user to collect an image of the users’ VMs hosted in OpenStack, and retrieve log events for all API requests made by the user and firewall logs for all of the users’ VMs. FROST is integrated with several OpenStack Dashboard and Compute components.

Martini and Choo presented a four-stage cloud forensic framework, and used it to guide their server and client analysis of the ownCloud private Storage. The authors successfully recovered a range of artifacts, including file data, metadata and authentication credentials. Then they analyzed the server component of ownCloud. In addition to locating a range of metadata and uploaded files (including previous versions), they were able to use the authentication credentials collected from the client to decrypt files stored on the server. This demonstrated the utility of the client followed by server forensic investigation approach. In another work, the same authors designed a process for remote programmatic collection of evidence from an IaaS cloud service, which would provide forensic researchers and practitioners a tool (for instance collecting data via API) to collect evidential data using a repeatable and forensically sound process.

Forensic-by-Design and Forensic-as-a-Service

Ab Rahman and colleagues proposed an alternative forensic readiness strategy, referred to as forensic-by-design. Conceptually, forensic-by-design is similar to security-by-design and privacy-by-design, where requirements for forensics are integrated into relevant phases of the system development lifecycle, with the objective of developing forensic-ready systems. The utility of such an approach is demonstrated in a later work.

There has also been research into offering forensic-as-a-service. Conceptually, forensic-as-a-service is similar to software-as-a-service where forensic applications and services are being moved to the cloud. For example, Castiglione and colleagues presented a cloud-based methodology to acquire forensic evidence from online services, such as webpages, chats, documents, photos and videos. A cloud-based solution hosts a network trusted service used to acquire evidence for subsequent analysis. Such an acquisition can be undertaken using a HTTPS proxy (capable of recording activities at the network level, such as IP, when an online service is accessed), or a software agent for the collection of information obtained by the targeted online service in a What You See Is What You Get (WYSIWYG) manner.

Along with his colleagues, van Beek proposed a cloud-based approach which allows one to process and investigate the large volume of seized digital materials, typically of a criminal investigation. This was also coined big data forensics by Quick and Choo. Specifically, digital evidence obtained
during the investigation are outsourced to the cloud by creating forensic copies, and later examined using a standard set of tools. Thus, evidence copies can be created and stored in a centralized and accessible location. Fu and his colleagues presented a cloud-based distributed solution for tracing Internet criminals using high-bandwidth sentinels within anonymous networks, such as Tor. This allows the capturing of (criminal) communications for analysis.

HARNESSING TECHNOLOGICAL ADVANCES FOR VARIOUS ASPECTS OF POLICING HAS BEEN A KEY OPERATIONAL OBJECTIVE IN MANY GOVERNMENTS AND LAW ENFORCEMENT AGENCIES. Examples include modernizing communications between field investigators, such as crime scene analysis personal and investigators, forensic laboratories, and the digital archives, using cloud computing. For example, Schiliro and Choo presented a cloud-based interactive constable on patrol system, which allows a law enforcement agency (or any other private sector organization) to deliver the organization's capabilities to the frontline officer via a mobile app. This includes the capability to connect and pull/push information and intelligence from a wide range of public and private databases (for example CCTV systems in a particular city, such as San Antonio), employing data-mining and other big data analytical technologies, and so on.

As cloud and related technologies advance, forensic investigators will find it challenging to keep pace, in the sense of identifying new forensic artifacts. Thus, there is a need for ongoing research into identifying new forensic artefacts in the cloud and related environment (for example multi-cloud and federated cloud, fog computing, edge computing, and Internet of Things, such as Internet of Battlefield Things), considering both data-at-rest and data-in-transit, as well as developing new forensically sound data collection techniques.

Current forensic techniques generally make use of vendor data communication facilities built into the mobile devices (such as iTunes backups for iOS devices) for the purpose of forensic extraction. Often this limits the potential for data extraction. For example, current tools would not be able to collect evidence from devices that are encrypted using strong passwords. Therefore, it is crucial to develop, validate and refine novel evidence-based data collection techniques to obtain evidential data from cloud computing (and other computing) devices in crimes that make use of sophisticated and secure technologies, for example, the use of strong encryption to secure both data-at-rest and data-in-transit, as well as anti-forensic techniques. These novel evidence-based data collection techniques need to be designed to circumvent advanced security features (such as developing low-level exploits and undertaking physical hardware analysis) and obtain evidential data from cloud computing devices, without compromising the evidence's integrity. These techniques will enhance “guardianship” and the “deterrent” effect in policing.

A recent literature survey also shows that there is a need for effective visualization of evidential data for forensic practitioners and investigators, as pointed out by the authors “while many researchers have made progress towards a model for visualizing forensic data, there continue to be gaps in this research area which need to be addressed”.

When designing cloud forensic techniques, it is also important to balance the need for a secure mobile telecommunications system and the rights of individuals to privacy against the need to protect the community from serious and organized crimes and cyber and national security interests. This issue has serious implications on the ability of governments to protect their citizens against serious and organized crimes. However, it remains an under-researched area due to the interdisciplinary challenges specific to cloud (and digital) forensics. Thus, it is important to bring together approaches from different disciplines to address the major contemporary challenges associated with cloud forensics. For instance, to ensure individual privacy, the techniques developed by forensic researchers should focus on individual suspect devices under direct judicial oversight (for example under a search warrant), as opposed to broad spectrum surveillance, such as the NSA incident revealed by Snowden in 2013.

References
1. D. Ó Coileáin, and D. O’mahony, “Accounting and Accountability in Content Distribution


ANIELLO CASTIGLIONE is an adjunct professor at the University of Salerno (Italy) and at the University of Naples "Federico II" (Italy). His research interests include security, communication networks, information forensics and security, applied cryptography. Castiglione has a PhD in computer science from the University of Salerno, Italy. He is member of several associations, including IEEE and ACM. Contact him at castiglione@ieee.org.