Cloud Computing Implications to Digital Forensics
A New Methodology Proposal

Fabio Marturana
marturana@libero.it

Abstract
This paper deals with a novel approach to digital investigations, aimed at optimizing law enforcement’s tasks, concerning digital evidence acquisition, examination, analysis and reporting, and reducing investigation complexity and operational costs. In the face of Internet’s pervasiveness and massive market penetration of high-performing and low-cost handset devices, resulting in a worldwide diffusion of cybercrimes, digital forensics seems indeed to be facing severe challenges which, if not taken seriously, may rapidly jeopardize the achievements of many years of forensic research. Motivated by this, the author proposes a model to perform complex data processing which exploits cloud computing capabilities, on the one hand, and modern mobile handsets capabilities, on the other hand, which can be exploited on the crime scene to perform live and post mortem artifact analysis. The paper provides the reader with the design guidelines, architecture components, interfaces, functional and non-functional requirements of a cloud-based forensic platform implementing the so-called “forensics-as-a-service” delivery strategy. The proposed framework consists of a cloud-based server-side and a client side, running on an Android smartphone or tablet. The client-side, in particular, consists of a collection of open source forensic tools and an Android software application, specifically developed for the purpose.

Author
Fabio Marturana is a PhD student in the University of Rome Tor Vergata’s Computer Science, Systems and Production Department. His research interests include Computer and Mobile Forensics, Information Security and Cloud Computing. He holds a MS in Electronic Engineering from Polytechnic of Turin and co-authored papers on his research topics.

1. Introduction
During a digital investigation, forensic analysts are used to performing queries, indexing data, calculating hashes, extracting features and correlating partial data to narrow the search and solve the case. As long as investigation complexity, on the one hand, and amount of data to analyze, on the other, allowed it, forensic data processing occurred in sequence on stand-alone forensic workstations. Unfortunately, Internet’s pervasiveness and market availability for cheap and sophisticated mobile devices with large storage capacity, have changed the landscape, resulting in an increase of digital investigation complexity
and contributing to the global diffusion of cybercrimes as well. Such crimes, on the one hand, are evolving at an astounding pace, following the same dynamic as the inevitable penetration of computer technology and communication into everyday life. Whilst society is inventing and evolving, at the same time, criminals are deploying a remarkable adaptability in order to derive the greatest benefit from it. Digital forensics, as a consequence, seems to be facing new challenges which, if not taken seriously, may rapidly render the actual forensics techniques obsolete and even not practicable. Current trends in computing and communications technologies have indeed shown growing amounts of disk storage and bandwidth available to ordinary computer users, resulting in significantly larger forensic data, with regards to the ability to process them in a timely manner. Performing common forensic tasks, such as keywords indexing and image thumbnail generation against a captured image, indeed, may take much of the available time before an investigation can even begin. As a consequence, digital forensics practitioners, attempting investigations using a single workstation as a platform, will be very soon completely overwhelmed by backlogs. A possible solution to overcome the issues outlined above is then to develop a new forensic paradigm exploiting capabilities offered by cloud computing. New discussions are indeed emerging, among forensic practitioners, on whether the cloud ecosystem could be adopted or even extended to delivery law enforcement’s forensic tasks “as-a-service”. Cloud computing and Internet’s pervasiveness have indeed radically changed the way information technology services are created, delivered, accessed and managed. Cloud computing has innovated information technology, enabling tasks formerly carried out by well-rounded computers and servers to be performed on a mobile device such as a smartphone. This new service delivery paradigm has the potential to become one of the most transformative developments in the history of computing, so why not use this technology in your favor then? Being available on the cloud and being independent of the device used, indeed, a pool of available resources such as applications, processes and services can be rapidly deployed, scaled and provisioned, on demand. For this reason using cloud computing to deliver on demand forensic services could be an effective new way to support digital investigations, allowing cloud organizations, service providers and customers, to establish forensic capabilities and reducing cloud security risks.

In this regard, the paper’s aims at describing the design guidelines of a secure forensics-as-a-service delivery platform exploiting cloud computing and mobile devices capabilities to support live and post mortem investigations on the crime scene. The platform is designed to provide investigators with a wealth of computing capabilities to conduct a live analysis on the crime scene with the stand-alone client device. The same device may be used to connect to the cloud platform as well, in order to upload forensic data to the server-side for remote processing.

The remainder of the paper is organized as follows: Section 1 describes the proposed platform, the server-side architecture, the functional and non-functional requirements and the client-server interfaces. Section 2 deals with the architecture and the functional requirements of the client-side, running on a mobile device and then conclusion and possible future research directions are discussed in the last two sections of the paper.

2. The Server-Side

To provide a viable solution to the above outlined issues and following on previous work by Marturana et al., the author proposes a model for distributed forensic data processing, called “forensics-as-a-service delivery platform”, aimed at exploiting cloud-computing capabilities. The proposed solution assumes that forensic images captured by the client-side software, may be divided into smaller fragments and uploaded...
in parallel to the server-side platform to manage massive uploads more efficiently. The server-side will be responsible for acknowledging the fragments received and forwarding them to selected servers in the cloud. Exploiting distributed file system’s capability to replicate files, image fragments are then exchanged among the selected servers, and more copies of the captured image are then rebuilt to assure redundancy.

To allow forensic tasks, such as keyword searching, to be split and performed in parallel on different servers, the author have indeed made the assumption that a forensic image can be considered as the natural unit of storage and processing. It is indeed possible to run remote processes on a set of distributed commodity servers capable of caching in RAM and processing image files, allowing complex forensic tasks to be quickly performed in parallel. In this scenario few dependencies place constraints on the distribution of files or require complex distributed synchronization.

\section*{2.1 Service architecture, functional model and interfaces}

The main architectural components of the server-side, summarized in Figure 1, are: a central process, called Orchestrator (O), a number of remote Forensic Processes (FP), a communication interface (i.e. CS-to-SS) between the client-side and the server-side of the platform, for service request and result delivery, and an internal interface (i.e. O-to-FP) for data exchange among the cloud servers.

On the CS-to-SS interface, the Orchestrator will be in charge of:

- Accepting incoming service requests;
- Initiating outgoing communications;
- Aggregating and delivering results;

While on the O-to-FP interface, it will be in charge of:

- Distributing captured image fragments, once uploaded, among available remote Forensic Processes;
- Keeping trace of captured images retained by each remote Forensic Process;
- Distributing processing tasks among available remote Forensic Processes, with regards to retained images;
- Sending flush command to remote Forensic Processes for deleting retained forensic data when needed.

Remote Forensic Processes will be responsible for:

- Receiving, acknowledging and storing captured image fragments;
- Processing captured images;
- Deleting retained images and forensic data upon Orchestrator request;
- Delivering results to the Orchestrator.
2.2 Server-side functional requirements

Forensic data processing requests and output delivery requirements of the proposed forensics-as-a-service delivery platform are described as follows:

- **Upload of a “live” digital media logical acquisition to the server-side.** The author considers here a live forensic scenario in which a logical copy of all the powered-on digital devices (i.e. computers, tablets, mobile phones, smartphones, PDAs etc.), found at the crime scene, is transmitted to the server platform. A secure connection available at the crime scene (e.g. a VPN tunnel upon a Wi-Fi or a 3G data link) shall be used to upload images. Alternative data transfer procedures shall be considered in case the crime scene is not under network coverage, such as the acquisition of a forensic images at the scene and consequent upload, once back at the forensic lab.

- **Upload of a “post mortem” digital media logical or physical acquisition to the server-side.** In this scenario, a post-mortem investigation on digital artifacts is conducted in a forensic lab. Physical and logical content of such digital artifacts shall be acquired and uploaded to the server-side platform. A secure network connection available at the lab shall be used to upload images. Once uploaded, data shall be retained by the platform for the whole duration of the investigation and may be queried at any time.

- **On-demand analysis of the information uploaded to the server-side.** Once uploaded to the platform, it shall be possible to access and process forensic images at any time. The following is a list of possible operations that shall be executed:

  1. *OS information retrieval:* it shall be possible to garner information about the digital artifact’s operating system;
  2. *Deleted file retrieval:* it shall be possible to use carving tools to extract the list of deleted files from unallocated space and slack space;
  3. *File classification by category or extension:* it shall be possible to classify the content of the analyzed image on the basis of statistics on file type, metadata, extension, average dimension etc.;
  4. *Installed application software list retrieval:* it shall be possible to retrieve the list of installed software;
  5. *Web browsing cache, cookies and history repository analysis:* it shall be possible to extract web browsing evidence, concerning navigation cache and history, list of stored cookies, and related to different browsers (e.g. IE, Firefox, Chrome, Opera, Safari etc.);
  6. *Saved chat/IM communication retrieval:* it shall be possible to retrieve evidence concerning saved chat and instant messaging sessions, if stored locally;
  7. *Saved Skype calls, chat and messages:* it shall be possible to retrieve evidence concerning saved Skype calls, chat and messaging sessions, if stored locally;
  8. *Local email database extraction:* it shall be possible to retrieve email databases and extract relevant evidence from the e-mails;
  9. *Digital timeline creation:* it shall be possible to rebuild the sequence of events and actions happened in a specific timeframe;
  10. *Encrypted file retrieval:* it shall be possible to extract encrypted files;
  11. *Content-based indexing:* it shall be possible to index image content to optimize queries;
  12. *Cryptographic hash calculation:* it shall be possible to calculate or verify files’ digest;
13. **Keyword searching**: it shall be possible to search for keywords.

### 2.3 Server-side functional requirements

The present section deals with non-functional requirements that complete the proposal:

- **Platform-independence.** The outlined requirements should be met regardless of the employed machine architecture and operating system.
- **Scalability.** The platform should be able to scale horizontally and the addition of more distributed machines should lead to proportional improvement in forensic tasks execution time.
- **Efficiency.** The extra work performed by the platform to distribute data and queries among its nodes, as well as to collect results, should be negligible compared to the total execution time.
- **Robustness.** Being a distributed system, the service delivery platform includes many components that potentially fail at any times. It should fall on the platform to detect and recover from such exceptional conditions and ensure the same level of confidence in the end result as in the traditional case.
- **Extensibility.** It should be easy to add a new function, as a building block of the Forensics-as-a-Service delivery platform, or replace an existing one. Therefore, writing a processing function from scratch to meet a new requirement should be compliant to the proposed model.
- **Interactivity.** To improve the user interactive experience, in such distributed solution, it should be possible to perform the time-consuming processing in the background (on cloud machines) while allowing operators to issue queries and to view partial results as soon as they become available.
- **Ease of administration.** It should be easy for administrators to operate the distributed system and minimal assumptions should be made about the underlying infrastructure (e.g. operating system services).

### 3. The Client-Side

In this section the author describes a Java implementation of the client-side, developed for the purpose, which integrates with the cloud-based server-side described in section II and implements the requirements discussed later on. The author used some popular forensic Linux-based distributions (e.g. Helix, Knoppix and DEFT Linux) as a reference point for reusing well-known extraction tools. The client-side of the proposed platform, running on a mobile device (e.g. a smartphone or a tablet) with Android operating system, is thought to be portable and very flexible. It is in charge of acquiring, dividing into independent fragments of smaller size and uploading logical and physical images of digital artifacts, in parallel, to the server-side platform, and sending consequently processing requests, based on the acquired data, to the server-side.

#### 3.1 Client-side architecture

The client-side outlined above, in turn, is logically divided into two parts: a target-side, including a set of Windows extraction tools, installed on a data partition of the micro SD card, and an app-side, including an Android App installed on the OS partition, which includes the graphical user interface and the client software. The bidirectional interaction between the app-side and the target-side is straightforward as it is possible to use the micro SD card as a repository that can be used by both the app-side and the target-side. The app-side, indeed, is in charge of configuring (e.g. writing) script files on the micro SD and reading
the data acquired from the target-side. The target-side, in turn, is responsible for reading the configuration scripts, executing them, collecting and writing results on the micro SD. The implemented solution avoids, therefore, complex protocol interactions and synchronization between the parties. The client-side architecture, including the described functional components, is summarized in Figure 2:

### 3.2 Client-side functional requirements

Client-side implementation requirements, with regards to the app-side, have been discussed in the previous section. To summarize, the app-side shall be able to:

- Manage the graphical user interface which interacts with the target-side for live and post-mortem artifact analysis;
- Configure scripts on the micro SD that launch forensic acquisition tools on the target-side;
- Read results written by the target-side on the micro SD and show them via graphical user interface;
- Redirect the target-side acquisition towards an external storage device.

Most of the requirements of the target-side are based on the consideration that a live analysis of powered-on artifacts may be critical during an investigation as it may provide important clues to first responders on the crime scene. In particular, powered-on artifacts contain memory-resident information, such as passwords and encryption keys that will be lost after rebooting. It also happens that artifacts under investigation, such as critical servers, may not be powered-off or afford disruption to service, and must be analyzed with live forensic techniques. In a live forensic scenario, the target-side shall be able to:

- Acquire current system time against an accurate time source;
- Acquire and analyze artifact data in order of volatility (OOV):
  - Physical memory dump, open files, open network connections, swap space;
  - Encrypted file systems where you do not have key to unlock;
  - List of active processes;
  - Windows registry;
  - Temporary file systems;
  - Message digests of gathered evidence.

In a post-mortem forensic scenario, the target-side shall be able to:

- Acquire a logical or physical image of each powered-off artifact;
- Analyze local file systems to recover specific files;
- Recover deleted files from unallocated space and slack space;
- Analyze windows registry to find installed applications, connected USB devices etc.;
- Create a timeline of events;
- Search for hidden data (steganalysis).
Part of such requirements have been already implemented in the current release of the client-side. Details about planned future implementations are provided in the next sections.

4. Conclusion

A traditional digital investigation implies that, using stand-alone forensic workstations, analysts are able to perform various forensic tasks in sequence, against limited datasets extracted from target artifacts, and evaluate correspondent results. Unfortunately this scenario is changing rapidly as the growing size of storage devices, on the one hand, and proliferation of multi-vendor mobile devices, on the other hand, are causing investigation delays and increasing complexity of digital forensics tasks. As a consequence, there is an urgent need to find novel solutions to improve digital investigation effectiveness. To solve the issues outlined above, the paper described an investigative platform for distributed forensic data processing, aimed at taking advantage of both mobility and cloud computing capabilities, called “forensics-as-a-service delivery platform”. The platform consists of a server-side spread across a multitude of commodity servers in the cloud, and a client-side running on a mobile device (e.g. a smartphone or a tablet) with Android operating system, which can be easily deployed on the crime scene to perform live and post mortem forensic tasks. The proposed solution is based on the assumption that each forensic task may be divided into a set of independent subtasks, running in parallel on a multitude of commodity servers in the cloud. It is possible to identify forensic images as the natural units of distribution among the cloud servers since independent file operations, such as keyword searching, may be split and performed in parallel on different servers. In the distributed scenario where few dependencies place constraints on the distribution of files or file fragments, it may be quite easy to reduce delays and increase efficiency by caching each subtasks’ dataset in remote servers’ RAM, as long as the subtask is up and running. The proposed cloud-based platform will take the responsibility of activating remote processing tasks and collecting partial results, allowing complex operations to be quickly performed. The client-side of the proposed platform is very important as well, as it may provide investigators with a wealth of forensic processing capabilities on the crime scene, besides being the access point to the forensic cloud. Architecture components, interfaces, functional and non-functional requirements of both server-side and client-side have been examined to provide the reader with interesting implementation guidelines. A prototype implementation of the client-side has been described in some detail.

5. Future Work

Given the relevance of the topic, a growing interest is emerging among forensic practitioners towards cloud computing implications to digital forensics. Within few years, the demand for processing digital tasks “as-a-service” will probably increase among practitioners and investigators who will experience this new way of performing forensic investigations. Possible future research areas on the subject may be:

- Extending the client-side capabilities to implement: (a) a secure platform for on-demand remote support on the crime scene by lab analysts, (b) live windows registry acquisition and analysis, (c) live windows temporary and log file acquisition and analysis, (d) live and selective file system preview and analysis and (e) live extraction of mounted encrypted partitions content.
• Performing and evaluating benchmarks to compare advantages and drawbacks of delivering forensic support services in house or contract out to a specialized cloud service provider, from the technical, economic, organizational and legal standpoint,

• Performing and evaluating benchmark with traditional digital forensics tools and techniques that may be adopted in similar situations.