Cloud Forensics - A True Game Changer

by Vassil Roussev, University of New Orleans

and Irfan Ahmed, Virginia Commonwealth University

About the Authors

Vassil Roussev is a Professor of Computer Science and Director of the UNO Cyber Center at the University of New Orleans. He has worked on digital forensics research for over 15 years, and has published over 50 peer-reviewed papers, including a book.

Irfan Ahmed is an Assistant Professor of Computer Science and Director of the Computer Security and Forensics Lab at Virginia Commonwealth University. His research interests are in the areas of digital forensics, malware analysis, system security, and cybersecurity education.

The Challenge

The single most consequential IT development is the continued, and accelerating, movement toward a complete *software-as-a-service* (SaaS) delivery model of computing. The idea of computing provided as a service, or *utility*, has been around since at least the 1960s and, originally, was motivated by the high cost of computer systems at the time. Fifty years later, IT is finally is moving at full speed toward centralized information services of the utility kind in order to extract higher levels of IT efficiency.

Traditional (digital) forensic practice developed since the mid-1980s around the *software as a product* (SaaP) IT delivery model – software is acquired like any physical product and is installed by the owner on a specific computing device, which is the (primary) host for all the application-specific computations being performed. Consequently, the analytical model has been focused on analyzing specific physical devices; investigators work with physical evidence carriers, such as storage media or integrated computer devices (e.g., smartphones). The standalone device is a relatively simple world where it is easy to identify where the computations are performed and where the results/traces are stored. Most of the digital forensic research and development has focused on discovering and recovering every bit of state and historical information left behind by the operating system (OS) and by user applications.

The service-based IT model requires a new approach that does *not* rely on low-level, block-device access to the physical storage medium. Instead, the only authoritative data sources are remote APIs, which provide fine-grain application data structures with well-defined semantics. At one fell swoop, this transition renders the established toolset useless as it can neither acquire the evidence, due to the new APIs, nor can it analyze it due to different semantics.

Case Study – Cloud Drive Services

To illustrate the new forensic landscape, consider cloud drive services, such as *Dropbox, Box,* and *Google Drive*; these represent the cloud analog of the hard drive – the most thoroughly understood source of digital evidence. Yet, as documented in "Cloud forensics–Tool development studies & future outlook", we are largely unprepared to acquire and analyze them forensically:

Partial Replication. The most obvious approach to acquiring a drive is to create a copy of the data as represented on the client. However, there is no guarantee of completeness – devices need not cache all the data on any one of the clients using the drive service.

Revisions. Most drive services provide some form of revision history and it is a feature that users have come to expect; it is also a new source of valuable forensic information. Revisions reside in the cloud and clients rarely tend to only the most recent version in their cache; a client-side acquisition will miss prior revisions and would not even be aware of them.

Cloud-native Artifacts. Digital forensics needs to learn how to deal with a new problem – cloud-native digital artifacts that have no serialized representation on the local filesystem. For example, *Google Docs* documents are stored locally as a link to the document, which can only be edited via a web app. Acquiring an opaque link is not very helpful – it is the content of the document that is of primary interest. It is possible to obtain a (PDF) snapshot of the web app artifact via the service API, but the entire editing history (recorded with millisecond granularity) would be lost.

There are additional technical challenges related to the fact that different services have different APIs and those have different semantics making standardization problematic at this point.

The Opportunity

Integration of Forensic, Security, Audit, and IR Facilities. It is always beneficial for forensics to take advantage of mechanisms that are created in service of other concerns. With the move to cloud IT, the lines between pro-active security monitoring, audit, and incident response (IR) mechanism become blurred as the same mechanisms, such as an execution log, can serve multiple purposes.

build solutions from scratch and, instead, aim at integrating with, and expanding upon, existing systems. The data volume challenge alone points to the need for a lot of processing to take place while relevant data, such as logs, is being generated. Security and IR tools have even tighter time constraints and perform some of the processing (such as indexing) at generation time, so the synergistic opportunity is ripe for the taking.

Domain Knowledge Transfer. As data representations become more abstract and standardized, a large body of data analytics and machine learning methods can be adapted to the forensics domain. Data classification, anomaly detection, pattern recognition, and other common techniques can readily help forensics speed up and automate the analytical process. This will be a major step towards much greater automation and intelligence.

Technology Reuse and Transfer. Forensic processing will ultimately need to move to the cloud – the sheer volume of data being accumulated in the cloud will soon render the traditional acquisition models unaffordable from both performance and cost perspectives. Moving the computation to where the data is offers the only feasible solution, and that requires redesigning and reengineering the *entire* forensics pipeline. It also offers the opportunities to integrate state-of-the-art technologies. For example, it is not possible for a traditional SaaP forensic software to take of advantage of Google's advanced image processing capabilities. However, for a Google Cloud-resident forensic tool, it becomes trivial to do so by employing the relevant API and the problem is reduced to having the budget to afford the capability. As platform capabilities advance, so do the capabilities of the forensic tool at minimal additional effort.

The Ultimate Goal – Full Automation

The ultimate scientific goal of digital forensic research has always been indirectly defined and revolves around the definition of forensics as ``science serving the law''. Thus, improvement of forensic science is understood to mean science serving the law *better* either by providing new sources of evidence, or by improving the quality and trustworthiness of the evidence.

We put forward the argument that, with respect to digital forensics, *automation* needs to be elevated to a first-class concern in order for the discipline to keep pace with IT developments and, thereby, remain relevant to the law. Specifically:

> *The ultimate goal of digital forensic science is to produce methods and tools that enable fully automated and scientifically trustworthy processing of forensic targets and cases.*

In a discipline where analyst expertise is prized above all, this statement may appear untenable; however, a sober view of the status quo and established trends make this goal *both necessary and incrementally feasible*.

Necessity. The present all-human analytical model has very slow capacity growth and is faced with exponential growth in data and in complexity; the latter two -- put together -- make the current human-centric model completely unsustainable. Simply adding more scalable tools (that can hash and index more data per unit of time) will *not* address the central problem of complexity, neither will any improvements in visualization and the user interface

Technology trends are at inflection point and threaten to completely overwhelm our projected forensic capacity. Therefore, they need to be matched by corresponding forensic technologies that are increasingly autonomous and, therefore, scalable and economical.

Feasibility. As the level of abstraction of the acquisition and analysis continues to rise, automated processing is becoming increasingly more plausible. Specifically, most of the data will be accessed via remote APIs which provide well-structured data objects with well-defined semantics. All of the low-level operations, such as device-level acquisition and data and reconstruction – which hands-on human intervention will simply be gone from the processing pipeline. The vast majority of the data will be in the form of logs offering explicit record of events, thereby obviating the need for much of the labor-intensive reverse engineering effort.

In sum, digital forensics is in the early stages of a revolutionary transition that raises difficult challenges, but also offers qualitatively new opportunities to develop intelligent, scalable, and highly automated digital forensic tools. It is worth noting that most IoT forensic investigations will also be resolved in the cloud as IoT systems rely of cloud data services for long-term storage.

References

Irfan Ahmed, Vassil Roussev, “Analysis of Cloud Digital Evidence”, In *Security, Privacy, and Digital Forensics in the Cloud, L. Chen*, and H. Takabi (Eds.), IGI Global, 2018.

Vassil Roussev, Irfan Ahmed, Andres Barreto, Shane McCulley ,Vivek Shanmughan, “Cloud forensics–Tool development studies & future outlook”, Journal of Digital Investigation, Vol 18, Sep 2016, pp.79-95. https://doi.org/10.1016/j.diin.2016.05.001