Chapter 17
Data Recovery Strategies for Cloud Environments

Theodoros Spyridopoulos
University of Bristol, UK

Vasilios Katos
Democritus University of Thrace, Greece

ABSTRACT

Data acquisition and data recovery are essential to any e-discovery or digital forensic process. However, these two aspects seem to be considerably difficult in a cloud-computing environment. The very nature of the Cloud raises a number of technical and organizational challenges, which renders traditional approaches and tools inapplicable. Resource pooling, rapid elasticity, and geographical distribution of data are only a small part of the Cloud's features that hinder the forensic investigation. At the same time, there is significant absence of forensic readiness in cloud computing policy framework. In this chapter, the authors discuss the challenges pertaining to data acquisition in a cloud environment and discuss possible directions for meeting these challenges by presenting representative cases and sketching acquisition process and scenarios.

INTRODUCTION

The increasing technical and legislative complexity of cloud computing systems has made traditional data acquisition infeasible in principle. Traditional acquisition tools and processes have become inapplicable due to resource sharing among multiple tenants and distributed data in multiple jurisdictions. The very nature of cloud computing is highly coupled with networking operations, thus cloud activities are reflected to network activities and the data identification process requires network-type metadata, however, additional changes might also need to be introduced in the current cloud infrastructure to facilitate data acquisition procedure.

In this chapter, we first highlight the inapplicability of the traditional acquisition tools base on the premise that data acquisition in cloud computing environments must follow live forensics practice. We then move on to discuss acquisition and recovery scenarios.

DOI: 10.4018/978-1-4666-6539-2.ch017

LIVE ACQUISITION IN THE CLOUD

Traditional forensic acquisition often follows the well established 'dead forensics' practices. A number of hard disks and other digital storage media are seized and their images are acquired using imaging software and/or hardware. Prior to the seizure of the suspect's equipment, a search warrant is issued (in most countries) and this warrant must identify a unique physical address. The seized equipment is also tied to the suspect, or at least is demonstrated to have been a part of the alleged offense.

Upon acquisition, the digital artifacts are self-sufficient in that they can be examined and analyzed independently. Although in some cases correlation between the different pieces of evidence is necessary in later stage of the analysis, the examiner can relatively easily interpret the artifacts found on the storage device at the initial stage of the analysis. We take a domestic hard disk as an example. All metadata are contained in the disk, including the partitions and file systems that determine the structure and representation of data in this disk. If this disk contains a virtual system or logical volume arrangements for redundancy and better manageability of the storage resources, the underlying structures are still available in the local system. Finally, the user's presence is normally local. Even when remote access is involved, the network activity metadata and logs are still maintained locally.

However, in a cloud-based storage system most of the assumptions above do not hold. Firstly, it is unlikely that the user will access a cloud storage service locally. Internetworking[1] is a critical enabler for cloud services and this alone is sufficient to put forth the need for live forensics. With respect to the cloud storage itself, metadata plays a crucial role not only in tracking down the evidence but also in protecting the legitimate users' privacy. Without metadata, the cloud storage can be seen as a well-stirred soup of data, thus entropy (uncertainty) will be very high, and the lack of context will make it impossible to associate data and stored objects with users, owners, and underlying activities. In a cloud storage environment, file structure is enforced by the master servers, which orchestrate the distribution, movement, and replication of files on the storage (or chunk) nodes. Similar to the users, master servers are entities different from the chunk nodes. The nodes and the master servers exchange command and control information as well as metadata through the network. Therefore local file system metadata are not sufficient for representing the whole picture of the states, ownership, and history of a data object.

Let us take a typical user activity scenario on a local desktop as an example. The information stored on the hard disk is normally sufficient to answer questions related to the user creating, moving and deleting files, attaching and removing storage devices, installing and removing software, etc. Most of this information is revealed by examining the user or system files such as the registry in Microsoft Windows and various log files. Tools like EnCase and Aftertime are very efficient in creating timelines of the user activity as reflected in the acquired disk images. In order to conduct the same exercise on a cloud storage setting, there is a need for further correlation between the participating entities. These entities are the user's computer, the cloud storage master server, the chunk nodes, and depending on the cloud service model used there may be more entities required as we will explain in the following sections.

Data recovery is one of the most important steps in the forensic process as it takes place in the beginning of the forensic process and heavily affects the quality and effectiveness of all forensic phases in later stage, especially forensic analysis. Although data resilience is one of the main drivers for cloud adoption, the absence of standards and guidelines for forensic acquisition poses significant challenges when it comes to the admissibility of the evidence.

Data are unlikely to get completely lost in a cloud storage environment as recovery processes