Digital Investigation 11 (2014) 30–42

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Cloud Data Imager: A unified answer to remote acquisition of cloud storage areas

Corrado Federici*

CIRSFID, University of Bologna, via Galliera 3, 40121 Bologna, Italy

ARTICLE INFO

Article history:
Received 6 November 2013
Received in revised form 5 January 2014
Accepted 1 February 2014

Keywords:
Cloud storage
Remote forensic acquisition
Dropbox
Microsoft Skydrive
Google Drive
Cloud computing
Computer forensics
Interoperability
ISO/IEC 27037
Virtual volume

ABSTRACT

The pervasive availability of cheap cloud computing services for data storage, either as persistence layer to applications or as mere object store dedicated to final users, is remarkably increasing the chance that cloud platforms potentially host evidence of criminal activity. Once presented *a proper court order,* cloud providers would be in the best position for extracting relevant data from their platforms in the most reliable and complete way. However, this kind of services are *not so widespread to date and, therefore, the need to adopt* a structured and forensically sound approach calls for innovative weaponry which leverage the data harvesting capabilities offered by the low level program interfaces exposed by providers. This paper describes the concepts and internals of the Cloud Data Imager Library, a mediation layer that offers a read only access to files and metadata of selected remote folders *and currently supports access to Dropbox, Google Drive and Microsoft Skydrive storage facilities*. A demo application has been build on top of the library which allows directory browsing, file content view and imaging of folder trees with export to widespread forensic formats.

1. Introduction

Cloud Computing has made true the long held dream of computers as affordable utilities (Parkhill, 1966) which are charged according to their usage. In this respect, a key role has been played by distributed file systems and object stores, which allowed to reach virtually infinite storage capacity by summing the individual contributes of the disks placed inside commodity servers. Well known solutions exist, either proprietary or open source, that ensure high availability and geographic distributions of data. A side effect of a reliable and cheap storage area is the remarkably increasing chance that it can be used for harboring crime related data, such as credit card numbers, stolen identities or violated credentials. Unfortunately for the digital investigator, distributed architectures may entail difficulties when it comes to rebuild a global picture as files get partitioned in several chunks of configurable size and are scattered among a potentially vast population of participating nodes (Quick and Choo, 2013a). This most probably prevents forensic teams from utilizing write blockers and bit stream copiers because it is hard to detect which of the plethora of nodes hold relevant data without digging into file system internals. But this is regrettably just a part of the story: proprietary technologies, unavailability of the provider to deliver a console with root privileges to third parties or simply lack of jurisdiction help figure out why an on field approach may simply be totally unfeasible. So the natural conclusion should be serving a warrant to cloud providers as, in principle, they would be in the best position to extract relevant data from their platforms. While this approach seems straightforward and rid of troubles, relying on a party that does not natively offer a professional forensic service, requires that a good deal of trust be placed

* Tel.: +39 051 2098771.
E-mail addresses: corrado.federici@unibo.it, corrado.federici@carabinieri.it.

http://dx.doi.org/10.1016/j.diin.2014.02.002
1742-2876/

on procedures and tools used at provider's premises (Dykstra and Sherman, 2012). Data should be delivered to forensic investigators in a well known format, as complete as possible, integrity protected and non repudiable. Consider however the following scenarios where data acquired as a result of a warrant could potentially be deemed unacceptable before a court for lack of reliability or sufficiency:

- a system administrator without a specific forensic background uses an ordinary maintenance script to restore the requested data from a backup. As a result, content gets extracted, but some file metadata are overwritten;
- deleted files are not recovered, even if this was technically possible;
- once packaged, the blob gets delivered without integrity protection codes or it is impossible to uniquely associate it to the provider because of flaws in the chain of custody;
- in case of proprietary templates, raw data is not exported in a well known format and browsing is only possible by means of a viewer program.

Resorting to the scrutiny of a third party appointed as needed to audit and certify the operation might result in additional costs and possibly further delays. Agreeing beforehand on an acceptable strategy for acquisition of data between law enforcement (LE) and provider could translate into delays as well and might need to be redesigned when the counterpart changes. When a provider assisted Forensic As a Service (Dykstra and Sherman, 2012) is not available, a third way may be considered that is secure, officially supported and reduces the point of contacts with the cloud provider so to possibly shorten times and lower costs. Given the self service nature of cloud platform, object storing is also exposed via entry points that usually reproduce all the features available from a web console. A low level interface based on Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) web services enables user created applications to remotely execute operations on folders and files such as download and list. Higher level Software Development Kits (SDK) are often available that wrap Hypertext Transfer Protocol (HTTP) calls and allow a programmer to rely on languages like Java. Reasonable scopes of application include, but are not limited to, technical activities performed during pre-trial hearing with or without the consent of the defendant. In the first case the defendant willingly gives his credentials as he may have interest in taking a trusted snapshot of his cloud stored files without any modifications. In the latter scenario, by performing a forensic analysis of a seized computer law enforcement could have recovered user name and passwords of a storage account (Quick and Choo, 2013a,b) or directly an access token string (AT) so to bypass user authentication, as it might be possible for Dropbox (see Section 6.1.1. and 6.1.2). While the approach of remote acquisitions seems promising, there are some aspects that need to be deepened before blueprinting strategies and tools able to image data in a forensically sound way. First and foremost, forensic best practices, where possible, suggest avoiding alteration of digital evidences (DE) during acquisition. Therefore a read only access to cloud storage areas which mimics the write blocking mechanism applied in traditional bit stream copy of physical mass memories would be beneficial. Indeed, Application Program Interfaces (API) do allow write access: upload, deletion and copy of objects are possible by design. Furthermore, while REST web services seems somehow the 'lingua franca' for interacting programmatically with remote storage, the parameters that need to be specified in the calls may vary greatly from one platform to another and so do the format of returned data. An extra layer which harmonizes the syntactic differences is therefore needed. Not less important is the requirement of protecting the integrity of all the retrieved data and reporting all operations in a detailed log. With this foreword, the paper describes the concepts and internals of the Cloud Data Imager Library (CDI Lib), a mediation layer we developed to offer a read only access to files and metadata of selected remote folders, while presenting a unified front end which masks out the syntactic and functional differences of cloud technologies. We built a desktop application on top of the library which, once instrumented with the necessary credentials, provides functionalities like folder listing with view of present, deleted and shared content, browsing of file revisions, extensive logging and imaging of folder trees with export to widespread forensic formats. CDI Lib currently supports access to three popular storage facilities: Dropbox, Google Drive and Microsoft Skydrive.

The rest of the paper is organized as follows: next paragraph reviews previous and related work. Section 3 gives some background information and Section 4 lists the requirements applicable to cloud forensic software. Section 5 shows the limitations of currently available tools, whereas Section 6 deals with CDI's architecture and functional tests. We draw the conclusions in paragraph 7.

2. Related work

Plenty of work has been developed about discovering traces left on client devices by the interaction with cloud storage platforms. For instance, Chung et al. (2012) have devised a procedure to collect remnants from computer and smartphones accessing, among others, Amazon S3 and Google Docs and found that many artifacts can be recovered by digging into logs, cache files and databases present in a user profile. In two consecutive papers, Quick and Choo, (2013a, 2013b) accomplished a comprehensive analysis concerning traces recoverable in memory and persistent storage of a Windows PCs and Apple iPhone after Dropbox and Microsoft Skydrive services were accessed via browser or client applications. A similar research was accomplished for Amazon Cloud Drive (Hale, 2013). Conversely, procedures and tools for server side acquisition of file content and metadata from a cloud object store appears to deserve a far larger degree of deepening. Quick and Choo (2013c) have explored the possibility of collecting files from a user account of Dropbox, Google Drive and Microsoft Skydrive. As a preliminary consideration, the authors observe that their investigation lacked a suitable