Digital Investigation xxx (2018) 1–16

Contents lists available at ScienceDirect

Digital Investigation

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

Following the breadcrumbs: Timestamp pattern identification for cloud forensics

Shuyuan Mary Ho [a, *], Dayu Kao [b], Wen-Ying Wu [b]

[a] Florida State University, School of Information, Tallahassee, FL, USA
[b] Central Police University, Department of Information Management, Taoyuan City, Taiwan

ARTICLE INFO

Article history:
Received 10 September 2017
Received in revised form
27 October 2017
Accepted 11 December 2017
Available online xxx

Keywords:
Timestamp
Cloud forensics
Behavioral analysis
Pattern identification
File metadata

ABSTRACT

This study explores the challenges of digital forensics investigation in file access, transfer and operations, and identifies file operational and behavioral patterns based on timestamps—in both the standalone as well as interactions between Windows NTFS and Ubuntu Ext4 filesystems. File-based metadata is observed, and timestamps across different cloud access behavioral patterns are compared and validated. As critical metadata information cannot be easily observed, a rigorous iterative approach was implemented to extract hidden, critical file attributes and timestamps. Direct observation and cross-sectional analysis were adopted to analyze timestamps, and to differentiate between patterns based on different types of cloud access operations. Fundamental observation rules and characteristics of file interaction in the cloud environment are derived as behavioral patterns for cloud operations. This study contributes to cloud forensics investigation of data breach incidents where the crime clues, characteristics and evidence of the incidents are collected, identified and analyzed. The results demonstrate the effectiveness of pattern identification for digital forensics across various types of cloud access operations.

Introduction

Cloud technology has become an indispensable channel that facilitates computer-mediated communication in modern society. Unfortunately, data in the cloud storage can be accessed and shared without proper authorization from the data-owner. The law enforcement community has faced difficulties during digital crime investigation in identifying the complex and dynamic operations in cloud storage regardless of the different service types, which can include infrastructure as a service (IaaS), software as a service (SaaS), or platform as a service (PaaS).

The National Institute of Standards and Technology (NIST) has identified a few significant challenges in cloud and digital forensics (NIST, 2014). These challenges include the diversity and complexity of the cloud architecture, as well as data integrity and recovery during data collection, obfuscation strategies and malware. The identity and roles of data owners vs. administrators is also of concern, along with the legal and ethical issues regarding laws and jurisdictions, standards, training and qualifications of forensic investigators as well as cloud providers. In particular, the NISTIR-8006 report pointed out the digital forensic challenge of analyzing important characteristics of timestamps (NIST, 2014).

Lopez et al. (2016) further classified 9 types of computing/digital forensics, including mobile forensics, network forensics, enterprise forensics, systems forensics, proactive forensics, cyber forensics, web forensics, data forensics, and email forensics. Among these varieties of forensics, a wide range of challenges occurring in the cloud computing environment were identified across legal and administrative issues (e.g., lack of standards, lack of international cooperation) as well as challenges identified in the technology domains (encryption, anti-forensics tools, lack of control of the cloud environment, large data volume, log visualization, virtualization, geographical locations, and metadata changes) (Lopez et al. 2016, p. 3). The complexity of the cloud environment challenges forensics practitioners who must apply both new and tailored methods during cloud investigations.

Casey (2011) suggested that forensic examiners (or, investigators) are required to scrutinize digital evidence in the magnetic data of physical medium, and translate this data into a form that humans can interpret. If good judgment and decisions are not applied during an investigation process, digital evidence may be unintentionally altered or destroyed (Williams, 2012).

* Corresponding author.
E-mail addresses: smho@fsu.edu (S.M. Ho), camel@mail.cpu.edu.tw (D. Kao), im821209@mail.cpu.edu.tw (W.-Y. Wu).

https://doi.org/10.1016/j.diin.2017.12.001
1742-2876/

Traces of digital evidence, such as timestamps on a device (whether client machine or a server machine) can be unintentionally left behind by a user during access or an operation, and this evidence can be captured and processed to figure out what happened during a digital crime. But the evidence itself is often quite fragile. If these digital "breadcrumbs" are mishandled in any way, it could cause permanent information loss (Casey, 2010). Information loss occurred due to errors in forensic operations, or incomplete data may be detrimental to any digital forensic investigation. More specifically, the timestamps of file metadata may differ from those of various operations and circumstances. File metadata timestamps can lend significant insight if grouped into clusters in a filesystem. We thus attempt to study this issue of timestamp analysis in file metadata, and hope to identify corresponding and useful countermeasures.

In this study, we take an iterative approach to observing and discussing changes in the timestamps of file metadata. More specifically, Windows NTFS and Linux forensics research is reviewed in Section Literature review. Research design and key experiment component timestamps are discussed in Section Our experiment, where our iterative approach to observe and collect timestamp data is discussed. Section Observation rules describes the generalizable observation rules regarding basic cloud access as part of our research finding. Research limitations, conclusion and future work are discussed in Section Conclusion.

Literature review

Digital evidence is commonly defined as "any data stored or transmitted using a computer that supports or refutes a theory of how an offense occurred. Evidence can also include data that may address critical elements of the offense such as intent or alibi" (Casey 2011, p. 7, Chisum et al., 2010). The identification of an offenders' intent or alibi requires close examination, analysis with reference to digital evidence. Such a close examination of digital evidence requires specific investigative techniques and methods. Thus, digital forensics refers to an examination of "a characteristic of evidence that satisfies its suitability for admission as fact, and its ability to persuade based upon proof (or high statistical confidence)" (Casey 2011, p. 14). Law Enforcement Agencies (LEAs) can then use such evidence with confidence to uncover the truth within court cases.

Buchholz and Eugene (2004) raised the importance of filesystem metadata in digital forensics, discussed the types of information and system log files desirable for investigations, and suggested principles for obtaining this evidence so that it is not lost or destroyed in transit or analysis. Metadata evidence such as timestamps and datestamps are important, but accessing this data can be quite complex (Boyd and Forster, 2004). Boyd and Forster (2004) outlined a checklist for understanding date and time evidence, including structures, formats, time translation, registry information, and browser information (e.g., temporary Internet files in cache and cookies) for examiners to follow so as to avoid making incorrect conclusions.

Consequently, temporal analysis is a critical component in understanding filesystems so as to map behavioral characteristics of the users to their related files (Chow et al., 2007). Casey (2011) illustrated a few cases where timestamps were examined for reconstructing the sequence of events (pp. 355, 544, 778). As a type of metadata, timestamps can be used to reconstruct events and operations pursuant to certain files and folders. In general, timestamps reveal acts of file modification, access, and creation (i.e., MAC time), which can carry significant value in a digital investigation.

However, timestamps can also be easily forged with file time changing tools, or altered inherently by batch operations such as automated tool scanning, or previewing activities (Bang et al., 2011; Cho, 2013). As there are many different sources for time, the correlation of time sources becomes increasingly challenging. Unless time sources can be synchronized, investigators cannot solely rely on one type of timestamp for a particular file to provide evidence for a particular event occurred at the corresponding MAC times (Stevens, 2004). Chow et al. (2007) studied behavioral characteristics of MAC times on an NTFS filesystem, so that the validation basis for temporal analysis in event reconstruction models can be formulated. A set of rules were hypothesized with respect to common operations by end users, such as access, copy, modify, delete, and download. These experiments help LEAs reconstruct online crime scenes (Chow et al. 2007, pp. 3–5). Bang et al. (2009) further categorized these timestamps with MACE (modify, access, create, and entry modified) values.

Regardless of the different types of categorization systems used for documenting time information, different filesystems offer different timestamp behaviors and recording methods. Bang et al. (2009) compared time information formats and details between FAT and NTFS filesystems, such as $STANDARD_INFORMATION property ($SI) and $FILE_NAME property ($FN). This research analyzed the changes in time information of files and folders for different operations across FAT and NTFS filesystems to reconstruct user operation. Bang et al. (2011) further analyzed the changes in timestamp attributes of files or folders resulting from user manipulations under different Windows operation systems in order to deduce user behaviors through a procedure. The Linux kernel, however, only retains the last modification, last inode change, and last access times. Such recording behaviors of Linux filesystems do not allow for a successful recreation of timeline of events (Das et al., 2012).

Moreover, it is often quite difficult to know precisely whether timestamps have been changed during the investigative process of examining timestamps. Cho (2013) analyzed the $LogFile, which is a record with 0x07/0x07 opcode in the data part of Redo/Undo attribute. These timestamps contain past-and-present timestamp data to uncover timestamp forgery in NTFS systems. Furthermore, Cho (2013) studied the difference within timestamp patterns, and proposed a set of rules for detecting timestamp forgery; that is, timestamp forgery can be detected by comparing changes in timestamp patterns made by the file time change tool, as compared to normal file operations. Das et al. (2012) also proposed augmenting the core of pathname lookup operation in the Linux kernel for accurate and authentic preservation of file timestamps for system wide critical files.

Windows filesystems

There are several types of Windows filesystems: FAT (File Allocation Table), NTFS (New Technology Filesystem), exFAT (Extended File Allocation Table), and ReFS (Resilient Filesystem) (Arpaci-Dusseau and Arpaci-Dusseau, 2016). Some patent-protected filesystems have also been designed for specific applications, and generally, a hacker could also create their own to hide their 'secret' data.

Filesystem attributes

The underlying assumption is that attributes of file metadata could be viewed and changed by any authorized systems users. The file metadata has associated information for file/directory data with two states (Casey, 2010): 1) Set or Cleared (similar to On or Off) (Chung, 2014). 2) The file metadata (e.g., timestamps) is used by the operating system and software applications as a way to record

Please cite this article in press as: Ho, S.M., et al., Following the breadcrumbs: Timestamp pattern identification for cloud forensics, Digital Investigation (2018), https://doi.org/10.1016/j.diin.2017.12.001