云存储服务数字取证调查

董振兴 张 青 陈 龙
(重庆邮电大学计算机取证研究所 重庆 400065)

摘 要 越来越多的用户使用云存储服务来存储数据,但利用云存储服务存储违法信息、盗取公司机密信息等违法案例也逐渐增多。如何提取完整、可靠的证据信息以证明云存储服务访问行为成为一个迫切需要解决的问题。以360云存储服务为例,分析使用浏览器、客户端软件访问云存储后残留痕迹的存储规律性,提出了一种用户行为取证分析方法。该方法通过把日志、历史痕迹等相互关联来重构用户行为时间线,进而分析用户的数据操作行为规律。该方法的取证调查思路、方法也适用于当前广泛使用的其他云存储服务。

关键词 云计算,云存储,数字取证,用户行为分析
中图法分类号 TP309 **文献标识码** A

Digital Forensic Investigation in Cloud Storage

DONG Zhen-xing ZHANG Qing CHEN Long
(Institute of Computer Forensics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract Nowadays, many users utilize the cloud storage service to store or share their data. At the same time, there are an increasing number of illegal cases about preserving illegal information or stealing the company's confidential data through cloud storage service. Collecting the crucial evidences from cloud storage service reliably and completely has become an urgent problem. This paper took 360 cloud storage service as example, analyzed the law of residual data after accessing to the cloud storage through the browser and/or client software, and then presented a forensic analysis method to identify user behaviors. The time line of user's action is reconstructed by combining logs and history data remnants together. Therefore the user behaviors related to the cloud storage service are profiled clearly. These ideas and methods can be applied to other cloud storage services currently used.

Keywords Cloud computing, Cloud storage, Digital forensic, User behavior analysis

1 引言

云计算平台可以给广大用户提供高效服务,但是不法分子也可以在此平台上进行违法活动,这不仅会给企业、个人带来巨大的经济损失,而且也会阻碍云计算的发展。云计算环境主要由大量的分布式异构虚拟计算资源构成,具有数据量大、数据分散存放、服务模型及部署模型多样性等特点,这些复杂的结构和特点给计算机取证工作的开展带来巨大的挑战。为了适应这些取证环境的变化,实现云计算环境下的取证工作成为一个重要的课题,具有非常重要的理论和实用价值。当前国内外学者对云取证的研究主要集中在两个方面:1)在云服务端设计方案事前记录信息并提供获取这些信息的接口,取证人员通过这些接口获得云端记录信息后对用户行为进行分析;2)从用户端提取用户使用云服务所产生的日志、历史痕迹信息,然后对用户行为进行分析。

Shams Zawoad[1]等从云服务提供商出发,设计了一种SecLaaS(Secure Logging-as-a-Service)模型在云端以日志方式记录用户行为信息,在保证用户隐私的条件下为调查人员提供获取日志的接口。Shams Zawoad[2]等还设计了一种PP-DP(Providing Proofs of Past Data Possession)方案在云端记录用户曾经在云中创建或保存过的文件信息,取证人员能够用这些记录信息来证明用户是否曾经在云端保存过某个文件。丁丽萍[3]等提出了ICFF方案,设置了一个专门的取证虚拟机,在保障证据数据的完整性及机密性情况下,能够自动地在虚拟机中抓取证据。Ting Sang[4]提出在云端以日志形式记录用户操作的重要信息,并将日志同步到用户端,通过对比云端和用户端的日志来保证日志信息的完整性。

Darren Quick[5]等关注用户访问Microsoft SkyDrive后在用户设备上残留的痕迹,证明在云计算环境进行取证分析时,使用规范的取证流程模型能降低数字证据在司法程序中受到质疑的风险。Fabio Marturana[6]等针对几种流行的云服务设计了一系列的用户访问行为,来考察在不同情况下从用户设备上能提取到何种有价值的信息。Jason S. Hale[7]等分析了在Windows系统中访问Amazon云后会残留哪些特殊的信息以及收集这些信息的取证工具和方法。Kim-Kwang Raymond Choo[8]等对几种云存储服务测试,来考察文件在上传、下载、储存的过程中文件内容和时间属性的变化情况。Darren Quick[9]等讨论了云环境下数字取证的流程,考察通

本文受国家社会科学基金(14BFX156),重庆市自然科学基金(cstc2011jjA40031),重庆市科委自然科学计划项目(cstc2011jjA1350)资助。
董振兴(1979-)男,讲师,主要研究方向为计算机取证、网络安全;张 青(1988-),男,硕士生,主要研究方向为计算机取证、云安全,E-mail:zhangqing7441@163.com(通信作者);陈 龙(1970-),男,博士,教授,主要研究方向为计算机取证、云安全、网络安全。

过不同方式访问 Dropbox 后，在 Windows 7 和 IOS 系统中残留的数据。

上述研究中，第一类方法在云服务端采取事前取证，以日志方式事前记录用户行为，这需要云服务提供商的配合，并改变现有云设计方案，目前还难以实施。第二类方法从用户端提取证据信息，由于不同的云存储服务各有特点，在本地生成的痕迹也不尽相同，只有充分了解其特点才不会在调查中遗漏一些关键信息。国外云存储服务已有一些研究成果，但尚未有针对国内云存储服务的取证研究工作。本文以当前广泛使用的云存储服务 360 云为例，分析在 Windows 7 中使用浏览器和客户端访问云服务后，在用户设备上残留痕迹的存储规律性。提出一种用户行为取证分析方法，通过把日志、历史痕迹等相互关联来重构用户行为时间线，进而分析用户的数据操作行为规律。

2 360 云存储残留痕迹分析

云存储服务是一种典型的 IaaS，为用户提供存储空间。各种云存储服务除了提供基本的存储功能外，还会提供其它各种功能，如本地文件备份、多个设备间文件同步、文件的版本管理、文件共享等[10]。用户无论是使用浏览器还是客户端作为云服务的接入端，在用户设备上都会残留大量的痕迹信息。分析这些残留痕迹的存储规律性能够帮助调查人员更快地从设备上提取完整、可靠的证据信息，为进一步行为分析提供数据源。

2.1 浏览器残留痕迹

尽管每种浏览器内核和结构设计不同，存储痕迹的方式也不尽相同，但这些痕迹都会以 history、cookies、cache 等方式留下各种记录。通过分析浏览器的历史痕迹，可以知道用户在何时使用浏览器访问过 360 云存储服务、登录的用户名、访问频率、进行了哪些操作以及操作文件的内容等信息，从而帮助调查人员掌握更多的涉案信息。

在 360 云存储取证过程中，浏览器的 history 是一个重要的考虑因素，用户访问 360 云服务后，会生成大量以“http://c51.yunpan.360.cn”开头的 URL 记录，解析浏览器的这些历史记录能获知用户曾经使用浏览器访问过 360 云服务。

登录 360 云服务后，生成的 cookie 所包含的信息也能表明用户访问过云服务，并且其还包含一些额外的信息，将这些信息定义为一个用户行为描述 A={登录账户名，用户名，登录次数，最后登录时间}来描述用户登录 360 云服务的行为。图 1 所示为使用 IE 浏览器登录 360 云服务后，使用 IECacheView 提取的 cookie 信息，这些信息抽象为登录操作 login={登录账户名，用户名，2，2014-5-28 16:37}。

Key	值	网域	变更日期
YUNPAN_USER	1011376002%40qq.com	yunpan.360.cn	2014/6/15 1:46:03
i360loginName	1011376002@qq.com	yunpan.360.cn	2014/6/15 1:46:03
count	2	c51.yunpan.360....	2014/6/15 1:48:03

图 1 登录后生成的 cookie

浏览器的 cache 包含大量取证的关键信息。浏览器访问云存储服务本质上是对网络 API 的请求，云服务器收到请求后，会对用户的请求生成响应。浏览器的 cache 文件实际上是对这些响应文件的存档，这些存档包括访问站点的图片、Flash、JS 脚本、CSS 文件以及一些 html 文件等，对 cache 文件分析能够获取用户使用浏览器访问 360 云服务的详细操作信息。

在对 360 云服务取证调查时，浏览器的 cache 文件中有几个特殊文件能为调查分析提供最有用的信息。特别要注意这些文件的 URL 属性信息，这些 URL 属性是云服务器对用户请求的响应链接，其不仅能作为搜索定位 cache 文件的标识，而且还包含各种信息：文件名、文件 hash、文件大小、操作类型、操作时间等。

当浏览或下载云端文件后，生成的 cache 文件 URL 属性以“http://pXX-X.yunpan.360.cn/intf.php?method=”开头，将 URL 属性包含的信息字段定义为用户行为描述 B={操作类型，文件 hash，文件名，操作时间}。浏览文件的操作方式为“Preview”，下载的操作类型为“Download”，文件 hash 可以用来唯一地标识一个文件，文件名为原文件名的 URL 编码(通过 URL 解码可还原出原文件名)，操作时间为 unix 时间戳(可使用“Dcode v4.02”工具格式化为“MM-DD-YYYY hh:mm:ss”格式)。

分析浏览文件操作产生的 cache 文件，能够得到浏览文件的具体内容信息。对于 txt 文件会直接将文件内容记录到 cache 文件中；对于其他类型的文本文件，则会在 cache 中记录浏览文件的页数以及文件 hash 值，通过这个 hash 值能够在浏览器缓存中找到一个以这个 hash 值命名的 cache 文件，这个 cache 文件记录了所浏览文件的具体内容信息。

使用 IE 浏览器浏览 360 云端文件后，生成的 cache 文件 URL 属性如图 2 所示，图中 method 的值为操作类型，fhash 的值为文件 hash，fname 值为文件名，callback 记录文件的返回数据类型和返回时间，对这些信息解析后获得的信息为 view={Preview，6e553062ce4565dc23e6c598288a8036e42658e，测试文件.docx，2014-10-31 11:10:31}。

```
http://p59-4.yunpan.360.cn/intf.php?meth
od=Preview.getHtmlInfo&fhash=6e553062ce4
565dc230e6c598288a8036e42658e&scid=59&fn
ame=%E6%B5%8B%E8%AF%95%E6%96%87%E4%BB%B6
.docx&pub=0&ck=a2478b928f91f4afa3273b328
bfc1f44&ofmt=jsonp&callback=QWJsonp14147
25030454
```

图 2 浏览文件后生成的 URL

用户上传、删除或重命名文件后，所生成 cache 文件的 URL 属性以“http://s.360.cn/yunpan/webclick.html?u=http%3A%2F%2Fyunpan.360.cn%2Fmy”开头，将包含的信息定义为用户行为描述 C={操作类型，操作时间}，上传、删除或重命名对应的操作类型分别为“upload”、“delete”、“rename”。

使用 IE 浏览器上传文件到 360 云端后，生成的 cache 文件 URL 属性如图 3 所示，可知用户操作行为 upload={upload，2014-10-31 11:07:57}。

```
http://s.360.cn/yunpan/webclick.html?u=ht
tp%3A%2F%2Fyunpan.360.cn%2Fmy&id=3537848.
16511137786162140000.1414724501831.406&but
tonid=Upload&t=1414724877888
```

图 3 上传文件后生成的 URL

2.2 客户端软件残留痕迹

用户安装360客户端后，使用客户端能方便快捷地对云端文件进行查看、上传、下载、编辑等操作。如果用户设置了同步目录，360云客户端会自动将同步目录中的文件上传至云端。在360云存储服务中，删除的文件都会保存到云端的回收站，在取证调查时应该注意对这个位置分析。使用客户端或在同步目录进行操作后，这些操作会以各种形式在本地留下痕迹，包括email地址、客户端软件的安装日期、访问云服务日期、文件同步记录等信息。

分析客户端的安装痕迹、日志等信息，重点考察安装目录中的filecache.db、sync.log和config.ini文件，通过分析能得到用户的账号、在何时访问过360云服务、使用360传输过什么文件、对文件进行了何种处理等信息。表1归纳了在Windows 7系统中安装360客户端软件后，在取证分析中一些重要的文件和路径，其中Path表示"C:\Users\Administrator\AppData"。

表1 Windows 7下的残留痕迹

路径和数据	描述
Path\Roaming\360CloudUI\ sync.log	客户端日志
Path\Roaming\360CloudUI\用户id\filecache.db	本地缓存文件信息
Path\Roaming\360CloudWin2\用户id\ history.dat	客户端上传记录
Path\Roaming\360CloudWin2\sync.log	同步日志
Path\Roaming\360CloudWin2\用户id\ config.ini	用户配置信息
Path\Roaming\360CloudWin2\用户id\filecache.db	同步目录文件信息
Path\Roaming\360CloudWin2\用户id\ history.dat	同步目录文件记录

客户端安装目录(360CloudUI)中的数据库文件filecache.db记录了360云在本地磁盘上的缓存文件信息，而同步目录(360CloudWin2)下的filecache.db文件则记录同步目录下的文件信息，这些文件信息包括存储路径、文件名、文件hash、文件大小、文件创建时间和修改时间等信息。同步目录下的配置文件config.ini记录了用户账户的详细信息，包括用户名、用户id、email等信息。

客户端目录中的sync.log文件记录了用户上次访问360云服务所使用设备的IP地址，通过这个IP能够知道用户曾经在哪些设备上登录过360云服务。客户端和同步目录中的sync.log文件记录了用户使用客户端或在同步目录中的各种操作信息，由于同步目录中的文件同时保存在本地，其sync.log文件不会记录用户浏览和下载文件的操作。

对sync.log日志分析时，我们将从中提取到的信息定义为用户行为描述 *oper*={操作类型，文件名，文件hash，操作时间}。在sync.log中记录用户操作信息的格式有两种模式：一种是缓存模式，这种模式会先在本地生成缓存文件，然后对比云端文件和本地缓存文件来决定客户端如何处理文件；另一种是非缓存模式，不会在本地生成缓存文件，而是直接向云端发起请求。由于同步目录中的文件同时保存在本地，因此只有非缓存模式。

对于缓存模式，sync.log中日志记录的格式以"status 6(ok)->3(CheckCloud)"标识开始，经过"status 3(CheckCloud)-> 4(CheckOffline)"和"status4(CheckOffline)->5(monitor)"两步处理，最后以"status 5(monitor)->6(ok)"结束。在"status 6(ok)->3(CheckCloud)"标识前有一个"[Download]"来记录客户端对缓存的处理方式。

对于非缓存模式，其日志记录的格式以"status 6(ok)->5(monitor)"标识开始，以"status 5(monitor) ->6(ok)"标识结束。在"status 6(ok)->5(monitor)"前有一个"[EVENT]"记录来保存用户的请求类型，根据"[EVENT]"中的信息可以进一步确认用户的操作。

在用户的各种操作中，浏览文件、下载文件操作属于缓存模式，上传文件、删除文件、重命名操作为非缓存模式。修改文件操作在打开文件阶段为缓存模式，编辑文件阶段为非缓存模式。知道日志记录格式的模式后，在sync.log中快速地定位操作的信息范围，然后根据相关标识信息进一步确认用户的具体操作类型以及操作文件的相关信息，这些缓存模式和信息格式也是下一步编写自动化分析工具、提取关键信息、进行关联分析的重要依据。图4所示为sync.log中记录的上传文件操作信息，从中可以解析出用户的操作行为信息oper={上传文件，测试文件.docx，6e553062ce4565dc230 e6c598288a8036e42658e，2014-11-01 10:16:38}。

```
[2014-11-01 10:16:38.558] status 6(ok) -> (monitor)
[2014-11-01 10:16:38.558] [db] Transaction Begin
[2014-11-01 10:16:38.558] [out_upload] [Queue:1]
[new] \测试文件.docx
[2014-11-01 10:16:38.862] [upload][192810392] begin: \
测试文件.docx, size:10258, fhash 6e553062ce4565dc230e6
c598288a8036e42658e
[2014-11-01 10:16:38.862] [req] upload filesize=10258,
\测试文件.docx
[2014-11-01 10:16:39.016] [upload][192810392] have,
new_ver:1, name:\ 测试文件.docx
[2014-11-01 10:16:39.016] status 5(monitor) -> 6(ok)
```

图4 非缓存模式的重命名记录

3 取证过程中的分析方法

取证调查人员对360云存储进行分析时，首先从可能存在残留信息的位置提取出待分析的数据源，这些数据源包括用户设备上的浏览器历史痕迹信息、客户端安装目录中的日志文件和数据库文件等。然后制定针对这些数据源的提取规则，提取"用户行为描述"信息，重构用户操作的行为事件线。

对这些日志及历史痕迹信息进行分析时，先确认用户使用360云操作过什么文件、操作时间，通过建立文件、事件与时间之间的关联关系，抽象出用户的操作行为信息；然后对操作信息做一系列处理，重构用户操作的行为时间线，进而分析用户的数据操作行为规律；最后根据取证目标和用户历史行为的偏差对用户行为做出判断。在取证过程中，有以下几步。

1)确认云端文件及其相关属性。提取用户360云服务的账户信息，访问云服务确定云端存储的文件，提取各个文件的时间属性。在这个过程中注意对云端回收站和"文件时光机"的检查，观察删除的文件和文件的历史版本及其修改时间。

2)确认用户使用客户端的操作信息。从360安装目录中提取客户端和同步目录的日志、数据库等相关文件，统计日志信息中用户上一次登录云服务的IP地址，定位用户访问360云服务的其他设备。从conf.ini中收集用户的账户信息，从filecache.db文件中确定用户传输的文件和本地的缓存文件信息，从sync.log中提取用户操作的详细信息。

3)确定用户使用浏览器的操作信息。提取浏览器的history、cookie、cache等文件，从这些文件中搜索与360云存储

相关的信息，使用一些关键字搜索定位用户对云端文件操作后形成的 cache 文件，提取这些 cache 文件的 URL 属性和其内容信息，抽象出用户的操作信息以及用户请求文件的内容信息。

4)重构用户行为时间线。将上面提取的“用户行为描述”规整于一个数据集中，并归类相似数据、删除重复数据，然后规范化数据，以时间为线索将这些数据顺序排列，补全缺失信息，重构用户行为时间线。

5)分析用户行为。在行为时间线上分析用户行为在不同位置、各个目标、行为意图方面的一些关联和规律，提取用户行为之间的关联特征，为进一步调查分析提供线索。

4 实验及结果

下面通过实验测试来验证本文所提出的分析方法的可行性。为了准确验证用户使用浏览器和客户端软件操作云存储服务后在用户设备上生成的痕迹，创建全新的虚拟机，然后在这个虚拟机上进行各种操作，保证实验不受其他外界因素的干扰。

创建 3 个虚拟机 IE_A、Chrome_B、Client_C，操作系统都为 Window 7 系统，其中 A 安装 IE 浏览器(8.0.7601.17514 版)，B 安装 Chrome 浏览器(33.0.1750.146m 版)，C 安装 360 云存储客户端软件(3.6.0.2210 版)。

实验中事先准备不同格式的 3 份文件，分别为“测试文件.txt”、“测试文件.docx”、测试文件.pdf”。为保证实验最大程度满足现实情况，尽可能在不同的环境下使用浏览器和 360 云客户端软件对这些文件进行各种操作，这些操作包括上传、浏览、修改、下载、重命名、删除，它们是无规律的，但是在操作的同时应该记录所做的操作，然后使用辅助取证工作重建用户行为，通过对比来验证所提辅助取证系统对用户历史数据操作行为的还原率。

实验结果如表 2 所列。从表 2 中可以看到用户在什么时间对什么文件进行了何种操作，基本上还原了用户使用 360 云存储服务进行操作的历史行为记录，但是并不能完全还原出用户的历史行为记录，特别是在浏览器上进行的上传、下载、删除操作，只能知道用户在这个时间做过相应的操作，无法还原出操作的文件信息。

表 2 360 云重建用户行为结果表

ID	时间	文件	Hash	操作	设备
1	2015-03-20 15:51:10	—	—	上传	IE_A
2	2015-03-20 15:52:35	—	—	上传	Chrome_B
3	2015-03-20 15:53:15	测试文件.pdf	7409979c95538aeb42763dac78e769c8562a15ab	上传	Client_C
4	2015-03-20 15:56:38	测试文件.docx	c3e67cd1ed2427806af42a87bfa767eea2e1924e	浏览	IE_A
5	2015-03-20 15:57:20	测试文件.pdf	7409979c95538aeb42763dac78e769c8562a15ab	浏览	Chrome_B
6	2015-03-20 15:57:57	测试文件.txt	9adcb29710e807607b683f62e555c22dc5659713	浏览	Client_C
7	2015-03-20 15:58:37	测试文件.txt	6df73cc169278dd6daab5fe7d6cacb1fed537131	修改	Client_C
9	2015-03-20 15:59:00	测试文件.docx	97570e2e7d0e4cac1b768d7b2d500fd8e511a080	修改	Client_C
10	2015-03-20 15:59:44	—	—	重命名	IE_A
11	2015-03-20 16:00:45	—	—	重命名	Chrome_B
12	2015-03-20 16:01:26	测试文件 1.txt	6df73cc169278dd6daab5fe7d6cacb1fed537131	重命名	Client_C
13	2015-03-20 16:04:08	测试文件 1.pdf	7409979c95538aeb42763dac78e769c8562a15ab	下载	IE_A
14	2015-03-20 16:04:35	测试文件 1.docx	97570e2e7d0e4cac1b768d7b2d500fd8e511a080	下载	Chrome_B
15	2015-03-20 16:05:09	测试文件 1.txt	6df73cc169278dd6daab5fe7d6cacb1fed537131	下载	Client_C
16	2015-03-20 16:05:47	—	—	删除	IE_A
17	2015-03-20 16:06:13	—	—	删除	Chrome_B
18	2015-03-20 16:07:19	测试文件 1.pdf	7409979c95538aeb42763dac78e769c8562a15ab	删除	Client_C

结束语 本文以 360 云为例，分析在 Windows 7 系统上使用客户端软件和浏览器访问云存储服务后，在用户设备上残留的痕迹及其存储方式和存储规律。这些残留痕迹和存储规律有助于取证调查人员更快地从设备上提取完整、可靠的证据信息，大大节约了收集证据的时间和精力。然后提出了一种重构用户操作时间线的方法，将从用户设备上提取到的各种残留痕迹关联，还原用户使用 360 云存储操作的历史，进而分析得到用户的数据操作行为，为进一步调查分析提供线索。虽然本文是以 360 云为分析对象，但是对其它云存储服务，也可以使用文中的思路做相似的分析。本文主要是对 PC 进行分析，而现实中用户可能使用手机等访问云存储服务。下一步将对手机等移动平台做进一步分析。

参 考 文 献

[1] Shams Z, Amit K D, Ragib H. SecLaaS: secure logging-as-a-service for cloud forensics[C]// ASIA CCS'13 Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security Table of Contents. New York: ACM, 2013: 219-230

[2] Zawoad S, Hasan R. I have the proof: providing proofs of past data possession in cloud forensics[C]// Cyber Security. Washington, DC, IEEE, 2012: 75-82

[3] 谢亚龙，丁丽萍，林渝淇，等. ICFF：一种 IaaS 模式下的云取证框架[J]. 通信学报，2013，34(5)：200-206

[4] Sang Ting. A log based approach to make digital forensics easier on cloud computing[C]// 2013 Third International Conference on Intelligent System Design and Engineering Applications (ISDEA). Hong Kong, IEEE, 2013: 91-94

[5] Darren Q, Kim-Kwang R C. Digital droplets: Microsoft SkyDrive forensic data remnants[J]. Future Generation Computer Systems, 2013, 29(6): 1378-1394

[6] Fabio M, Gianluigi M, Simone T. A case study on digital forensics in the cloud[C]// 2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Sanya, IEEE, 2012: 111-116

[7] Hale J S. Amazon cloud drive forensic analysis[J]. Digital Investigation, 2013, 10(3): 259-265

[8] Quick D, Choo K-K R. Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata[J]. Digital Investigation, 2013, 10(3): 266-277

[9] Quick D, Choo K-K R. Dropbox analysis: data remnants on user machines[J]. Digital Investigation, 2013, 10(1): 3-18

[10] Chunga H, Parka J, Leea S, et al. Digital forensic investigation of cloud storage service[J]. Digital Investigation, 2012, 9(2): 81-95