

# Microsoft Office Upload Center Cache Files in Forensic Investigations

Rick van Gorp, Kotaiba Alachkar

*Supervisor:*  
Yonne de Bruijn  
Fox-IT

MSc System and Network Engineering  
University of Amsterdam

February 6, 2018

# Overview - Definition of cache files

- Microsoft Office Cache Files: generated by Microsoft Office Upload Center.
- Path: \Users\\AppData\Local\Microsoft\Office\\OfficeFileCache
- File format list:
  - **FSD-files**: used to store the document
  - **FSF-files**: used as a connecting point between the FSD-file and CentralTable.accdb
  - **CentralTable.accdb**: used to store all metadata regarding the upload process

# Overview (cont.)

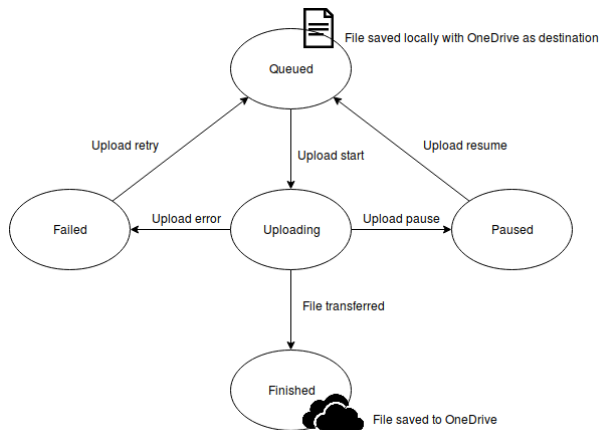


Figure 1: States of cache files during the upload process to OneDrive

# Problem Statement & Research Question

- Only **speculation** on what forensic value the FSD- and FSF- files have
- “**1.2 Billion** Microsoft Office Users and **200 Million** OneDrive users in 2014” <sup>1</sup>

## Research Question

*In what way do the cache files produced by Microsoft Office Upload Center contribute to a forensic investigation?*

---

<sup>1</sup>Microsoft by the Numbers: <https://news.microsoft.com/bythenumbers/planet-office>

- 1 Cloud Hosted Data in Digital Forensics (Slidedeck - 2014):
  - Australian technology company called **Nuix**
  - Briefly described the global contents of **CentralTable.accdb**
- 2 Windows 10 Forensics - OS Evidentiary Artefacts (Slidedeck - 2015):
  - Australian Researcher **Brent Muir**
  - Manually carve document from **FSD-files** but no methodology published

- Generate dataset:
  - cache files in all five states
  - two users on a Windows 7 VM running Microsoft Office 2016
  - .pptx, .docx, and .xlsx to upload: empty, large ( 5MB) and with one line of text (with & without an image)
- Perform statistical analysis: determine what information is available and what not under what circumstances
- Derive unknown file formats: length, offsets, known file headers, number of files, and checksums in data sections

Results outline:

- 1 File description
- 2 Availability of information
- 3 Retrieved data implication

# File Description - FSD-file

- The size of an FSD-file differs depending on the size of a source document

Size input (bytes)	Size FSD-file (bytes)	State FSD-file
11,762	65,536	Queued
11,762	131,072	All, except queued
11,869	65,536	Queued
11,869	131,072	All, except queued
1,163,631	1,245,184	Queued
1,163,631	2,424,832	All, except queued
5,660,169	5,767,168	All, except failed

**Table 1:** Examples of differences between file sizes of input documents and FSD-files per state



- Global file format derived from comparisons
- FSD-file:
  - Magic Header (16 bytes)
  - Unknown data (8176 bytes)
  - Subsection (appearing  $\alpha$  times):
    - Header A (8 bytes)
    - Unknown data ( $\beta$  bytes)
    - Header K (8 bytes)
    - Optional Section Q (appearing  $\gamma$  times)

- Optional Section Q:
  - Header Q (8 bytes)
  - Data (Unknown bytes)
  - End of data header Q - 79 05 (2 bytes)
  
- Data: Contains ZIP-archive headers and image headers
- Implies (part of) Office document is in the FSD-file

- The file format of the FSF-file:

Header (20 bytes)	
$\alpha$ : Data Length (1 byte)	<i>Data</i> : Filename & Terminator(0x05) ( $\alpha$ bytes)
<b>Total Size: 21 bytes + Data Length</b>	

- FSF-file points to FSD-file name: Used by CentralTable to connect metadata in CentralTable to FSD-file

- Microsoft Access database (Date/time unreadable) <sup>2</sup>
- Metadata about documents submitted to Microsoft Upload Center
- It consists of the following tables:
  - 1 MasterFile
  - 2 CacheProperties
  - 3 IncomingEvents
  - 4 OutgoingEvents
  - 5 ServerTarget

---

<sup>2</sup><https://github.com/rickvg/office-cachefiles>

- Table **MasterFile** contains most metadata:
  - Pointer to the FSF-file (*FileEntryFileID*)
  - Name of the file
  - Author of the file
  - E-mail address of uploader
  - Remote location of file (If uploaded)
  - Dates and times: Modified and Uploaded (Server & Local)

# Availability of Information

- Our CentralTable parser shows old rows in table *MasterFile*<sup>3</sup>
- CentralTable: Count of rows per state increases for table *MasterFile*

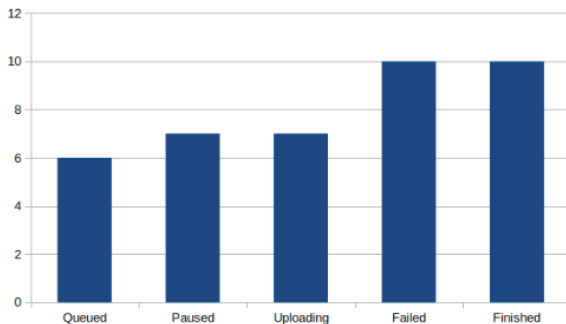


Figure 2: Mean count of rows per state for table *MasterFile*

<sup>3</sup><https://github.com/rickvg/office-cachefiles>

# Availability of Information (contd.)

- Generic changes during state transitions:
  - Tables *MasterFile* and *CacheProperties* change the revision number in column **ColumnRevisionID**
- MasterFile-table changes during state transitions:
  - Most changes
- CacheProperties-table changes during state transitions:
  - No patterns found

- Document recovery from cache files:
  - Manual or Automatic
  - With or without Microsoft Office 2016



- Automatic with Microsoft Office 2016
  - CentralTable requires records for uploading a file
  - Column *FileEntryID* in table *MasterFile* must point to FSF-file GUID
  - Column *FFileSavedToServer* in table *MasterFile* must be set to 0
  - FSF-file can be generated for any FSD-file
- Recover full document including its images and metadata

# Availability of Information - Document Recovery (contd.)

- Manual or automatic without Microsoft Office 2016
- Extraction script for small documents and parts of large documents <sup>4</sup>

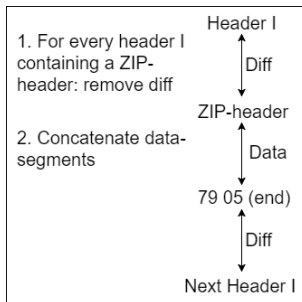


Figure 3: Extraction method for small documents without images

<sup>4</sup><https://github.com/rickvg/office-cachefiles>

- In our research, the retrieved data is divided into two parts:
  - (Parts of) original documents
  - Metadata related to documents
  
- Additional evidence in a forensic investigation <sup>5</sup>

---

<sup>5</sup><http://ieeexplore.ieee.org/document/7379751/>

# Conclusion

- **FSD-file** is used to store the document, **FSF-file** is used as a connecting point between the FSD-file and CentralTable.accdb and **CentralTable.accdb** is used to store all metadata regarding the document
- (Parts of) documents and its own metadata can be retrieved from FSD-files
- Check whether entries in table *MasterFile* have been manipulated (not which)
- The large amount of metadata with(out) the document could be used as additional evidence in a forensic investigation

- Exploring the FSD-file format in more details
- Extending **FSD-files Documents Extractor** script to support large-size documents and documents including images
- Expanding the research to include various **Microsoft Office versions**, **Operating Systems**, and **file-hosting** cloud platforms

Thank you for your attention

Do you have any questions?