OOXML File Analysis of the July 22nd Terrorist Manual

Hanno Langweg

OOXML File Analysis of the July 22nd Terrorist Manual

Hanno Langweg

NISlab Norwegian Information Security laboratory, Høgskolen i Gjøvik, Norway
hanno.langweg@hig.no

Abstract. We examine the terrorist manual circulated on the day of the attacks in Oslo and on Utøya island on July 22nd 2011 to find out if the OOXML structure is consistent with claims by the suspect apprehended for the terrorist act, and to determine if there have been additional authors.

Keywords: Document forensics, OOXML

1 Hypothesis

We work under the assumption that the document distributed by the suspect apprehended for the terrorist acts was edited without attempts to evade forensic analysis of the OOXML file. The text [2] may nevertheless contain exaggerations, lies, plagiarized content from internet sources.
We intend to support or refute the following hypotheses:

- *The document was written by more than one author.*
- *The timeline of editing events derived from the document's structure is consistent with the diary in chapter 3.154 (pp. 1,415–1,472).*

2 Method

Data acquisition Our investigation is limited to published sources and we did not obtain evidence collected by law enforcement agencies. Several copies of the document are downloaded from different websites on the internet, and a binary comparison is performed of the files using MD5 checksums. We also compare the MD5 hash values with values published in other places. Hence, we believe that we obtain an unaltered copy of the file.
OOXML [1] is a container for a zipped folder structure; after file acquisition, the document is decompressed and partitioned into several XML files.

Analysis Structural analysis looks at the generated table of contents, document revisions, changes in formatting and language metadata of paragraphs.
Content analysis explores how the text is divided into logical parts, how pictures are used and where they originated, how language is used in different parts of the text, and whether there are inconsistencies in the use of words or described events. The goal is to find changes in style that indicate different authorship.

3 Related Work

General challenges of OOXML and related file formats are treated by [4]. Hiding of additional files inside of OOXML documents is discussed in [5]. Source identification of OOXML documents (based on a reference e.g. to detect copyright infringements) and a tool to aid the examination are demonstrated in [3].

4 Findings

Metadata Metadata for the document was retrieved from files app.xml, core.xml, and settings.xml. There is no record of tracked changes in the document. The metadata provides evidence that the table of contents (containing links to the chapter headings) was changed after its generation. The company name "Grizli777" is according to a web search for the term an indication that the document was edited with an unlicensed copy of Microsoft Office, probably retrieved through a Torrent stream and possibly executable from USB pen drive.
It is plausible that the document was created on 2011-03-07 and the content compiled from several earlier partial documents. The recorded total editing time of 03:27:00 for the eight revisions of the document also indicates that the file received input from other files during its existence.
The document was saved seven times after the initial creation (8 revisions). This prompts the examination of revision ids in the document to find out which modifications were applied to the document after creation and when they occurred.
We observe that the last diary entry in chapter 3.154 (p. 1,472) states 12:51 as time of last writing while the document was saved 13:23, i.e., 32 minutes later. This could owe to clock differences or to further editing after the final diary entry. It is also possible that "12:51" is not the true time.

Revision identifiers The document contains six distinct revision identifiers for paragraphs related to paragraph creation (`rsidR`), two additional identifiers are retrieved for paragraph fragments (runs). This is consistent with the metadata. Content associated with the revisions is shown in table 1.
Including revision identifiers for paragraph marks there are 320 revision ids, i.e., that there have been 320 file save operations over the whole period of creating, composing, and editing content in OOXML files. Since only 12 revisions are mentioned as child nodes of the `w:rsids` element in `settings.xml`, we conclude that the document must have been composed of text stored in several separate DOC or DOCX (OOXML) files.

Image metadata There are 98 pictures in the document. None of the image files contains suspicious metadata or reveals additional information when examined with ExifTool.
Original storage locations comprise 11 folders. The path common to almost all referenced files – `C:\Users\` – reveals that the pictures had been stored on a computer running Microsoft Windows Vista or Windows 7, because user data was first stored in that location in operating systems released after Windows XP. That means that the Word document file containing the pictures was created after ca. November 2006–January 2007 (when Vista was released) or contained almost no pictures before that date. One picture was probably inserted from an external hard drive or a USB memory stick (drive letter `J:`).

Table 1. Revision identifiers and affected paragraphs/runs

Revision	Content
0063635B	Almost whole document, including TOC
00C15193	p. 1, dash before “2011” and space following
000C0B04	p. 12, announcement of movie availability
008D5CFB	p. 12, announcement of movie presentation on youtube.com
000C0B04	p. 12, announcement of movie presentation on veoh.com
008D5CFB	p. 18, empty paragraph (to force page break?)
00967D3B	p. 1,387, fragment: “(a certain degree of national Darwinism)”
00967D3B	p. 1,387, formatting of paragraph on a “future servant class”
00C4654C	Chapter 3.153, p. 1,399, fragment: “: Andrew Berwick”
00207CC4	Chapter 3.154, pp. 1,439–1,472, rest of diary chapter after 2011-03-01
00207CC4	p. 1,472, empty paragraph preceding “Further studies”
00DB706A	p. 1,516, empty paragraph (text removed?)

5 Conclusions

We were not able to retrieve evidence that the document contained parts that exhibited differences compared to the remaining content. Even if A.B. started with an initial version supplied by somebody else and even if that content was provided as a Microsoft Word document, a coherent style or revision id could not be detected for a long sequence of paragraphs in the final version. This might not be surprising after at least 320 discovered revisions of the document that was probably edited over a period of more than four years.
We were not able to find contradictory evidence that the events described in chapter 3.154 must have led to a different document structure.
In this article we discussed the 8 revisions of the final document file; it remains to dissect the document based on all 320 revision identifiers.

References

1. Standard ECMA-376 Office Open XML File Formats. `http://www.ecma-international.org/publications/standards/Ecma-376.htm`, 2006–2011.
2. Anders Behring Breivik. 2083. A European Declaration of Independence. `2083-AEuropeanDeclarationofIndependence.docx`, MD5 checksum 7a74e156aefb45416ea057ce19dfe4e9, 2011.
3. Zhangjie Fu, Xingming Sun, Yuling Liu, and Bo Li. Forensic investigation of OOXML format documents. *Digital Investigation*, 8(1):48 – 55, 2011.
4. Simson L. Garfinkel and James J. Migletz. New XML-Based Files Implications for Forensics. *IEEE Security and Privacy*, 7:38–44, 2009.
5. Bora Park, Jungheum Park, and Sangjin Lee. Data concealment and detection in Microsoft Office 2007 files. *Digital Investigation*, 5(3-4):104 – 114, 2009.