http://dx.doi.org/10.3745/KIPSTC.2012.19C.4.215

OOXML형식을 사용하는 MS 파워포인트 파일에 대한 편집 이력 조사 방법

윤 지 혜[†] · 박 정 흠[††] · 이 상 진[†††]

요 약

개인 및 기업의 업무 처리 시 컴퓨터 사용이 일반화됨에 따라 각종 문서 파일들이 디지털 형태로 생성되고 이메일, USB등 다양한 매체를 통해 이동, 복사되고 있다. 이러한 디지털 자료를 면밀히 분석하면 문서 작업 중 발생한 편집 이력을 추적할 수 있다. 이에 관하여 복합문서 파일 형식에 대한 연구는 있었으나 새로운 OOXML형식에 대해서 파일의 작성 과정을 파악하기 위한 저장되지 않은 파일을 찾거나 내부 작성 순서를 추적하고 서로 다른 문서 파일간의 연관성을 분석하는 방법에 대한 연구는 없었다. 향후 OOXML형식 디지털 문서의 사용이 더욱 증가할 것이며, 이러한 편집이력 추적 연구는 문서 파일에 대한 디지털 포렌식 수사에 큰 도움이 될 것이다. 따라서 본 논문은 OOXML형식 문서에 대해서 포렌식 관점을 가지고 문서파일의 내부 작성순서를 분석하고 파일 간 연관성을 추적하는 조사 방법을 제시한다.

키워드 : OOXML, 디지털포렌식, MS파워포인트; 슬라이드식별자, 편집이력

Methods for Investigating of Edit History about MS PowerPoint Files That Using the OOXML Formats

JiHye Youn[†] · JungHeum Park[††] · Sang-jin Lee[†††]

ABSTRACT

Today, individuals and businesses are a lot of paperwork through a computer. So many documents files are creating to digital type. And the digital type files are copied, moved by various media such as USB, E-mail and so on. A careful analysis of these digital materials can be tracked that occurred during the document editing work history. About these research are on the compound document file format, but has not been studied about the new OOXML format that how to analyze linkages between different document files, tracking an internal order, finding unsaved file for identify the process of creating the file. Future, the use of OOXML format digital documents will further increase, these document work history traceability in digital forensic investigation would be a big help. Therefore, this paper on the new OOXML format(has a forensic viewpoint) will show you how to track the internal order and analyze linkages between the files.

Keywords : OOXML, Digital Forensics, Microsoft PowerPoint, Slide Identifier, Edit History

1. 서 론

일반적으로 문서파일 작성 시 Microsoft office를 많이 사용하는데, 2007년 이전 까지 사용되던 문서 파일은 복합 형태로 저장되었다. 따라서 현재까지 문서파일에 대한 연구는 주로 복합 형식을 가지는 Microsoft office의 2003이하 버전 파일들을 다뤘다. 복합문서 형태 파일에 대한 이전 연구는 일반적인 문서 작성 응용프로그램이 별도의 정기적인 작업을 하게 됨으로써 사용자가 모르게 생성되는 데이터가 대부분의 문서 파일에 존재함을 설명하였다[1].

Microsoft office 2007버전이 등장하면서 파일 포맷이 바뀌었는데, Microsoft office의 2003 이하 버전 파일들은 복합문서 파일 형태로 저장하였으나 2007 이상의 버전에서는 Office Open XML(이하 OOXML)파일 형태로 저장하기 시작하였다[3][7]. 일반적으로 기업 및 개인은 문서 작성 시 Microsoft office를 많이 사용하므로 OOXML형태의 파일 또한 비중이 크게 늘어났다. 앞으로도 OOXML형태의 파일은 더 많이 생성될 것이고 OOXML과 관련된 포렌식 관점의

※ 본 연구는 한국연구재단을 통해 교육과학기술부의 바이오연구개발사업으로부터 지원받아 수행되었음(2011-0027732)
† 준 회 원 : 고려대학교 정보경영공학전문대학원 석사과정
†† 준 회 원 : 고려대학교 정보경영공학전문대학원 박사과정
††† 종신회원 : 고려대학교 정보경영공학전문대학원 교수
논문접수 : 2012년 3월 7일
수 정 일 : 1차 2012년 6월 22일
심사완료 : 2012년 6월 27일
* Corresponding Author : Sang-jin Lee(sangjin@korea.ac.kr)

연구가 필요할 것이다. OOXML형태를 가지는 Microsoft 워드 파일을 다룬 포렌식 관점의 이전 연구가 있었으나 파워포인트와는 다르고 파워포인트 파일에 대한 연구는 없었다. 따라서 새롭게 바뀐 OOXML파일 포맷에 대해서 숨겨진 데이터가 있고, 그로 인해 추가적으로 얻을 수 있는 정보들이 있음을 분석할 필요성이 있다.

대부분의 중요한 기업 및 개인의 문서, 발표, 보고서와 같은 파일들은 Microsoft 파워포인트로 많이 작성된다. 파워포인트 작성 시 보통 사용자는 이전의 서식을 유지하기 위해 이전 파일을 복사 및 수정하여 내용을 추가하거나 삭제하는 작업을 하게 되는데, 이러한 작업 특성을 이용하여 파워포인트 파일을 분석하면 서로 다른 문서간의 연관성 혹은 파일 내부 슬라이드간의 연관성을 파악할 수 있다. 이러한 분석은 파워포인트 파일에 대한 편집 이력 추적을 요하는 디지털 포렌식 조사 시 큰 도움이 될 수 있다. 또한 Microsoft 파워포인트로 작업을 하던 중 저장을 하지 않아도 컴퓨터 어딘가에 자동 복구를 위해 저장된 파일이 존재할 가능성이 있으므로 이를 확인하고, 식별자를 조사함으로써 서로 다른 파워포인트 파일의 작성 순서 혹은 파생 여부를 파악하는 것은 중요하다. 따라서 본 논문에서는 Microsoft office의 파워포인트 파일을 대상으로 포렌식 관점에서 의미 있는 데이터를 조사하는 방법을 소개한다.

본 논문은 다음과 같이 구성된다. Ⅱ장에서는 Microsoft office 파일의 응용프로그램에 의해 생성되는 데이터와 관계된 이전 연구를 소개한다. Ⅲ장에서는 Microsoft 파워포인트의 OOXML 파일 형식을 설명한다. Ⅳ장에서는 Microsoft 파워포인트의 포렌식 속성을 설명하고, Ⅴ장에서 Microsoft 파워포인트 파일에 대한 편집이력 조사 방법을 제시한다. 마지막으로 Ⅵ장에서 결론을 내린다.

특별한 언급이 없다면, 이 후 장에서 언급하는 Microsoft 워드, 엑셀, 파워포인트 파일은 Windows 7 환경에서의 Microsoft office 버전 2007, 2010에서 생성된 파일을 의미한다.

2. 관련 연구

컴퓨터 사용이 대중화 되었고 문서도 많이 이용하게 되었지만 일반 사용자의 경우 문서 파일에 숨겨진 데이터가 있다는 사실을 잘 모른다. 이에 대해 2004년 Simon Byers는 Microsoft 워드에 문서 생성자의 조직과 이름, 워드 응용프로그램의 버전과 문서 형식, 문서가 구성된 경로, 프린터 정보, 이메일 헤더 혹은 웹 서버정보 등이 숨겨진 데이터로 존재한다고 설명하였다[1]. 이 연구는 워드파일을 대상으로 조사되었지만 Microsoft 파워포인트에도 숨겨진 데이터가 존재하며, 새롭게 바뀐 OOXML에도 존재한다.

2007년 Aniello Castiglione등은 2003이하 버전의 워드 파일에서 포렌식 관점으로 의미 있는 데이터를 찾고, 복합문서 파일 형식을 기반으로 데이터를 은닉하는 방법을 설명하였다[2]. 이어서 2011년 새로운 OOXML형태를 가지는 Microsoft office 2007 이상의 버전에서 4가지 스테가노그래픽 기술을 제안하고 그에 대한 분석을 소개하였다[5]. OOXML은 Zip기반의 파일이다. Zip파일은 다양한 압축방식을 지정할 수 있는데, 이러한 압축방식의 차이를 이용하여 데이터를 은닉할 수 있다. 또한 Microsoft office 내부에서 사용하는 고유 식별자 값을 변환하여 정보를 은닉하는 방법이 있다. 새로운 Microsoft office에서의 은닉 방식은 OOXML형태의 특성과 구조를 알지 못하면 분석하기 어렵다.

JH Park 등은 Microsoft 파워포인트 2003(한글/영어) 이하 버전에서 '빠르게 저장하기' 기능에 의해 남겨지는 잉여정보 속성과 파일 내에서 사용하는 슬라이드, 개체 식별자 속성을 포렌식 관점으로 분석하였다[4]. 잉여정보 속성을 이용하면 이전 작업 내역이 추적 가능하고, 메타데이터인 식별자 속성을 이용하면 서로 다른 파일 간의 관련성을 파악하기 용이하다.

Microsoft Office 문서가 널리 사용되면서 문서의 저작권과 관련된 분쟁이 많이 발생하고 있다. 2011년 Zhangjie Fu 외 3명은 저작권 분쟁 발생 시 문서의 출처를 조사하는데 유효한 증거를 추출할 수 있는 방법을 제시 하고 RI와 문서 속성의 조사를 통해 원본과 파생본 여부를 파악할 수 있으며, OOXML포맷 내부의 생성자 정보와 시간정보를 통해 실제 저작권 소유자를 파악할 수 있다는 것을 발견하였다[7].

위 연구들을 통해 2007버전 이상의 Microsoft 파워포인트에 대해 적용하여, 포렌식 속성을 갖는 식별자가 존재할 수 있으며, 새로운 OOXML 형식에도 데이터를 은닉할 수 있음을 확인할 수 있다. 본 논문은 Microsoft 파워포인트 파일을 대상으로 포렌식 관점에서 중요한 데이터를 의미할 수 있는 속성들과 응용프로그램에 의해 데이터가 생성될 수 있음을 설명한다.

3. Microsoft 파워포인트의 파일 형식

이 장에서는 Microsoft 파워포인트 파일의 이해를 위한 OOXML형식을 설명한다. 또한 파워포인트 파일에 대하여 패키지의 전반적인 형태와 주요 구성 요소를 설명한다.

3.1 OOXML 파일 형식

각각의 OOXML파일은 Fig. 1과 같이 패키지(Package)라 불리는 컨테이너(Container) 안에 압축된 여러 파트(part)들과 그와 관련된 것들이 저장되어 있다. 하나의 패키지는 패키지의 콘텐츠-유형(content-type), 관계(relationship), 파트 항목들을 포함하는 일반적인 ZIP 압축 파일이다. 패키지-관계(package-relationship) ZIP 항목에 파트와 패키지 사이의 관계 정보가 저장되며, 파트-관계(part-relationship) ZIP 항목에는 여러 파트 사이의 관계 정보가 저장된다[6].

새로운 OOXML 파일은 이전 파일 유형과 구별하기 위해서 새로운 파일 확장자로 변경하였다[7]. 이전 파일 유형의 확장자는 워드는 .doc, 엑셀은 .xls, 파워포인트는 .ppt였지만 새로운 파일 유형의 확장자는 기존의 확장자 명에 'x'를 붙

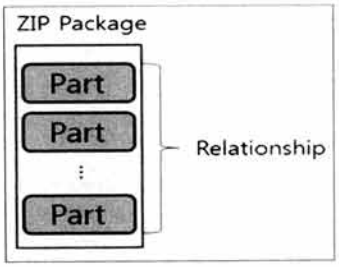

그림 1. OOXML 파일 유형의 구조
Fig. 1. The structure of OOXML file types

여 .docx, .xlsx, .pptx 등과 같이 표현한다[5][8]. 매크로를 사용하는 경우에는 'x'대신 'm'을 확장자 명에 붙인다[8].

3.2 Microsoft 파워포인트 파일 형식

Microsoft 파워포인트 문서 파일은 기능들을 단일 파트에 모두 저장하지 않고, 특정 기능들을 그룹화 하여 각 요소별로 나눠진 파트에 저장된다. 일반적인 Microsoft 파워포인트 파일의 그룹화 된 구조는 다음 Fig. 2와 같다. Fig. 2를 보면 컨테이너 루트에 _rels, docProps, ppt로 각 기능을 그룹화 하여 구성됨을 확인할 수 있다.

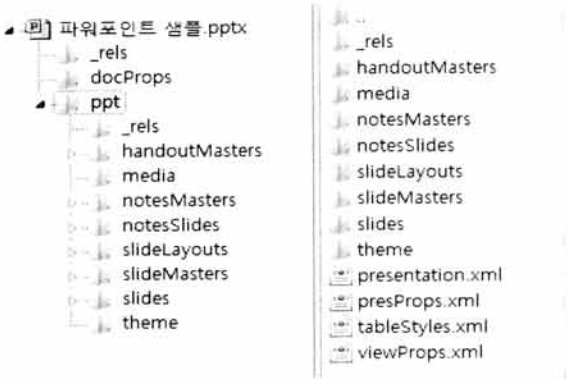

그림 2. Microsoft 파워포인트 파일 형식
Fig. 2. The internal structure sample of the Microsoft PowerPoint

"_rels" 구성 요소에는 ".rels" 확장자를 갖는 관계 파트가 저장되며, 이 파트는 "ppt" 구성 요소 아래의 기능별로 나뉘어져 있는 그룹에서도 저장된다. 이러한 관계 파트는 문서의 다양한 파트들을 연결시켜 주는 역할을 한다. "docProps" 구성 요소에는 응용프로그램에 의해 규정된 파일등록정보, 핵심 파일 등록 정보 파트 그리고 파워포인트의 최초 슬라이드 미리보기 이미지가 저장된다. 그리고 루트에 있는 "[Content_Types].xml"파트에는 총 사용된 슬라이드, 레이아웃, 노트 등 패키지 안에 있는 모든 다른 파트에 대한 콘텐츠의 유형 정보가 저장된다. 또한 파워포인트 문서 파일에는 이미지, 애니메이션, 오디오, 비디오 기능 등이 포함될 수 있다. 이러한 기능이 포함되면 "media"그룹을 만들어 내부에 실제 파일을 저장한다.

대표적인 파트 구성요소들로는 Table 1과 같이 프레젠테이션, 슬라이드, 슬라이드 마스터, 노트 마스터, 유인물 마스터(handoutMaster) 항목 등이 있다. 프레젠테이션 파트에는 사용된 각 파트들에 대한 ID정보를 가지고 있다. 이러한 ID 정보는 서로 다른 문서 간에 작성 순서 및 연관성을 파악하는데 결정적인 정보가 된다. 슬라이드 항목은 프레젠테이션에 있는 모든 슬라이드를 말하며, 슬라이드 마스터 항목은 프레젠테이션에 사용되는 모든 슬라이드 마스터를 말한다. 노트 마스터는 페이지 형식에 대한 정보가 포함되며, 노트 슬라이드에는 작성된 노트에 대한 정보를 담고 있다. 노트 슬라이드 내부에는 노트가 작성된 슬라이드를 찾을 수 있는 정보를 포함하고 있기 때문에 응용프로그램에서 확인하지 못하는 슬라이드가 있더라도 연관되어있는 내용을 추측 가능하게 한다. 유인물 마스터는 유인물을 보는 방법을 나타내며, 유인물을 나누어주기 위한 인쇄 설정을 나타내므로 유인물을 위한 인쇄여부를 파악하고자 할 때 참조할 수 있다.

표 1. 주요 파트 설명
Table 1. Describes the main parts

파트	설명
Handout Master	유인물을 보는 방법
Notes Master	페이지 형식 정보
Notes Slide	노트 정보
Presentation	사용되어진 파트 ID
Slide	사용되는 모든 슬라이드
Slide Master	사용되는 모든 슬라이드 마스터

4. OOXML 형식을 사용하는 Microsoft 파워포인트 파일의 포렌식 속성

이 장에서는 Microsoft 파워포인트 응용프로그램에 의해 생성될 수 있는 포렌식 속성을 소개한다. 이에 관하여 서로 다른 파워포인트 파일의 연관성 수사 시 도움이 될 수 있는 생성과 수정에 해당된 정보들의 소개와 Microsoft 파워포인트 파일 내의 각 파트에 부여되는 식별자를 설명한다.

4.1 자동 복구 정보 저장

Microsoft 파워포인트 응용프로그램은 사용자가 파워포인트 파일을 편집할 때 '저장'을 하지 않더라도 비정상적으로 종료되거나 사용자의 실수로 인해 작업 내용을 저장하지 못하였을 때를 대비하기 위해 '자동 복구 정보 저장' 기능을 가지고 있다. 이 기능은 Microsoft 파워포인트 2010의 경우 기본적으로 10분마다 복구를 위한 정보를 자동으로 저장하도록 설정되어 있으며, 저장되는 임시파일은 C:\Users\[사용자계정]\AppData\Roaming\Microsoft\PowerPoint\ 아래에 위치하도록 설정되어 있다. 이 위치에 생성된 파일은 사용자가 문서 작업을 할 때 최초 저장은 하였으나, 저장 이후 문서의 수정 및 추가 작업을 할 때 저장되지 않았던 파워포인트 문서가 저장된다. 임시파일이 저장되는 위치와 시간 간격의 설정 및 확인은 Fig. 3과 같이 파워포인트 옵션에서 할 수 있다.

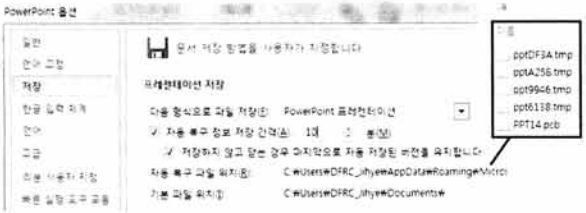

그림 3. 파워포인트 2010 버전 응용프로그램의 자동 복구 정보 저장 설정을 위한 옵션 페이지
Fig. 3. The option page for Save Auto Recover information (PowerPoint 2010 version of the application)

한편 최초 저장이 없는 경우 응용프로그램은 '저장하지 않은 프레젠테이션 복구' 기능을 따로 수행한다. 이 기능으로 임시 저장되는 파일은 C:\Users\[사용자계정]\AppData\Local\Microsoft\Office\UnsavedFiles 하위에 위치한다.

4.2 문서 속성

문서 속성은 app.xml과 core.xml에 기록되어있다. 다양한 요소들이 있지만 그 중 중요하다고 판단되는 요소를 간추려 Table 2에 정리하였다. app.xml이 가진 요소 중 서로 다른 문서간의 연관성 조사 관점에서 중요하게 보이는 속성은 TotalTime과 HiddenSlides 요소이다.

TotalTime은 총 편집시간을 의미하며, 시간 단위는 분이다. TotalTime을 확인하여 각 문서의 편집 소요 시간을 비교하면 더 큰 편집 시간을 갖는 문서를 파악할 수 있게 된다. 만약 파워포인트 문서를 복사한 후 원본은 접근하지 않고 복사본만을 가지고 편집 작업을 했다면 TotalTime이 더 큰 값을 가지고 있을 것이다.

또한 HiddenSlides는 슬라이드 중 숨겨진 슬라이드의 수를 표시한다. 파워포인트 쇼의 경우 응용프 로그램의 뷰어에서는 숨겨진 슬라이드를 볼 수 없지만, 파일 내부에 HiddenSlides 값을 가지고 있다면 전체슬라이드 수의 값이 뷰어에서 확인할 수 있는 슬라이드 수보다 클 것이다. 숨겨진 슬라이드를 알아낼 수 있다는 점은 편집자의 의도를 파악하는데 있어 매우 중요하다.

이러한 HiddenSlides 속성을 파악할 수 있는 또 다른 파트로는 \ppt\slides\slide*.xml이 있다. 숨김 설정을 한 슬라이드는 Fig. 4와 같이 show 속성 값을 '0'으로 설정한다. 파워포인트 파일의 편집 상태를 열람 및 수정할 수 없는 경우라 하더라도 각 슬라이드별 숨김 속성을 파악하여 숨겨진 슬라이드를 찾아낼 수 있다.

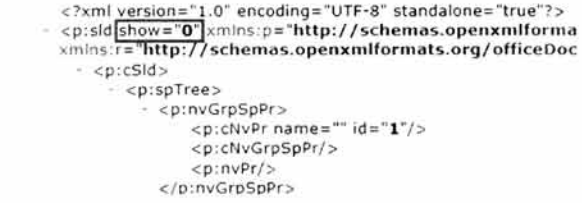

그림 4. 숨긴슬라이드 설정을 알 수 있는 HiddenSlide 속성
Fig. 4. HiddenSide properties can see hidden slide settings

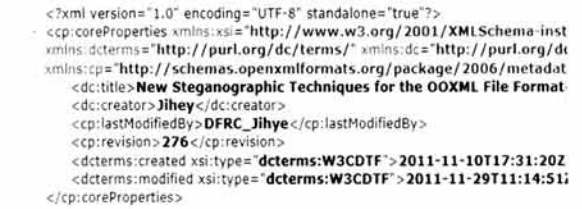

그림 5. core.xml 페이지 샘플
Fig. 5. core.xml page sample

core.xml 내부에는 파워포인트 파일의 생성자와 생성 시간, 최종 수정자와 수정 시간, 몇 번의 수정을 하였는지에 대한 정보를 가지고 있다. Fig. 5를 보면 처음에는 Jihey라는 사용자가 2011년 11월 10일 07시 31분 20초에 본 파일을 최초 생성하였으나 2011년 11월 29일 11시 4분 51초에 DFRC_Jihye로부터 최종 수정되었음을 확인할 수 있다. 그리고 이 파일은 전체 276번의 수정을 거쳤음을 알 수 있다. 즉, core.xml을 보면 파워포인트 파일의 최초 생성자를 파악할 수 있고 타인이 파일을 수정했다면 그에 대한 증명도 가능하다.

4.3 식별자

식별자에 대한 포렌식 속성을 소개하기 전, 식별자에 관하여 설명하겠다. Microsoft 파워포인트 파일을 구성하는 모든 파트(슬라이드 마스터(slideMasters), 핸드아웃 마스터(handoutMasters), 슬라이드, 노트 마스터(notesMasters) 등)는 개별 식별자와 저장된 데이터의 내용을 포함한다. 이러한 식별자는 파트별로 유일한 값을 가지며, r:id와 id로 표현된다. r:id는 Relationship ID를 의미하고 id는 파트 자체에 해당하는 고유 값을 갖는다. id 값은 유일한 값을 가지며 rId1은 슬라이드 마스터, rId2는 슬라이드1로 절대적인 값으로 고정된 이후 rId3부터는 고정된 값이 아닌 상대적인 값을 갖는다. 다음 소단원에서 주요 파트의 식별자를 분석하여 포렌식 속성이 있음을 설명한다.

표 2. app.xml와 core.xml에 포함된 주요 요소
Table 2. The main elements included in the app.xml and core.xml

app.xml 주요 요소	설 명	core.xml 주요 요소	설 명
TotalTime	총 편집시간(단위=분)	creator	생성자
Slides	슬라이드의 수	lastModifiedBy	마지막 수정자
Notes	노트를 포함하는 슬라이드 수	revision	수정 횟수
HiddenSlides	숨겨진 슬라이드 수	created time	파일 생성 시간
AppVersion	Application 버전	modified time	마지막으로 수정된 시간

4.3.1 슬라이드 마스터 식별자

Microsoft 파워포인트 문서 작성 시, 사용자들은 일반적으로 디자인 서식을 설정한다. 각각의 디자인 서식은 슬라이드 마스터에서 상세히 설정할 수 있으며 정해진 디자인 서식마다 슬라이드 마스터 아이디(sldMasterId) 값이 주어진다. 슬라이드 마스터 식별자인 슬라이드 마스터 아이디의 특징은 다음과 같다.

① 슬라이드 마스터에만 부여되며, 슬라이드 마스터 아이디는 (r:id=" ")와 (id=" ") 쌍으로 구성된다.
② 슬라이드 마스터 r:id는 "rId1"이고, id는 "xxxxxxxxxx" 열자리의 숫자로 표현된다.
③ 슬라이드 마스터 id는 어떠한 서식도 선택하지 않는다면 최초 값은 '2147483648'값을 갖는다.
④ 파워포인트 파일을 복사나 다른 이름으로 저장하여도 슬라이드 마스터 r:id와 id는 값이 변하지 않는다.
⑤ 슬라이드 마스터 내부 서식을 일부 수정하여도 슬라이드 마스터 id는 변하지 않는다.

사용자들은 파워포인트 파일의 디자인 서식을 설정할 때, 서식을 전부 수정하는 것이 아니라 일부 수정 혹은 기존의 파일을 복사하여 그대로 사용한다. 이러한 일반 사용자들의 파워포인트 사용 특성은 같은 서식을 갖는 서로 다른 파일이 존재할 경우 복사 여부를 가리는데 중요한 역할을 한다. 서식을 조금 수정한다고 하여도 슬라이드 마스터 아이디 값은 동일하기 때문에 복사된 파일이 있다면 원본과 같은 슬라이드 마스터 아이디 값을 가질 것이다. 각각 생성된 파일의 슬라이드 마스터 아이디가 같을 경우는 희박하기 때문에 두 문서가 동일 파일로부터 파생되었음을 입증하는데 활용할 수 있다.

4.3.2 슬라이드 식별자(sldID)

OOXML형식을 사용하는 Microsoft 파워포인트 파일의 각 슬라이드에는 '슬라이드 식별자'로써 슬라이드 아이디(sldID)가 부여된다. 이 번호는 슬라이드에만 부여(슬라이드 마스터, 노트, 그림 등에는 부여되지 않음)된다[4]. 슬라이드 식별자는 Fig. 6에서와 같이 presentation.xml 페이지 내 슬라이드 리스트(sldIdLst)요소에서 확인할 수 있다. 슬라이드 식별자는 다음과 같은 특징이 있다.

① 슬라이드에만 부여되며, (r:id=" ")와 (id=" ") 쌍으로 구성된다.
② 첫 번째 슬라이드의 r:id는"rId2"이고, id는 "256"이다.
③ 기존 슬라이드를 모두 삭제한 후에 생성한 슬라이드의 식별자 id는 "256"이다.
④ 슬라이드 식별자 r:id는 응용프로그램의 뷰어에서 확인할 수 있는 슬라이드의 순서와 동일하다.
⑤ 슬라이드 식별자 id는 응용프로그램의 뷰어에서 보이는 슬라이드 순서와 관계없이 슬라이드가 생성된 순서대로 '가장 높은 슬라이드 id + 1'로 순차적으로 증가한다.
⑥ 파워포인트 파일을 복사나 다른 이름으로 저장하여도 슬라이드 식별자 r:id와 id는 값은 변하지 않는다.
⑦ 슬라이드가 중간에 삭제되거나 이동되면, 슬라이드 식별자 id는 불연속성을 가지게 된다.
⑧ 숨긴 슬라이드도 슬라이드 식별자 특징이 그대로 적용된다.

이와 같은 특징들로 슬라이드 식별자의 비교를 통해 슬라이드 각각의 생성, 삭제, 이동 여부를 파악할 수 있다. 예를 들어 Fig. 6을 분석하면, 처음 세 슬라이드가 순차적으로 생성되었음을 알 수 있다. 그리고 4번째 페이지 rId5에 슬라이드 id 260이 있다는 것은 슬라이드 id 259가 어디론가 이동되었거나 삭제되었음을 알 수 있다. 그림에서는 슬라이드 id 259가 rId6에 위치하며 이는 5번째 슬라이드가 4번째 슬라이드 앞에 삽입 생성되었거나 생성된 후 이동되었음을 알 수 있다. 이와 같은 방식으로 분석을 하면 슬라이드 id 261, 267을 가지는 슬라이드도 삽입 생성되었거나 생성된 이후 이동되었음을 알 수 있다. 결국 모든 슬라이드 id는 순차적이지 않은 모습을 띄게 되며, 슬라이드 id 값의 분석을 통해 슬라이드의 편집 이력을 파악할 수 있게 된다. Fig. 6에는 나와 있지 않지만 슬라이드가 삭제되었을 경우 슬라이드 id가 중간에 빠져 있을 수도 있다.

```
<p:sldIdLst>
    <p:sldId r:id="rId2" id="256"/>
    <p:sldId r:id="rId3" id="257"/>
    <p:sldId r:id="rId4" id="258"/>
    <p:sldId r:id="rId5" id="260"/>
    <p:sldId r:id="rId6" id="259"/>
    <p:sldId r:id="rId7" id="262"/>
    <p:sldId r:id="rId8" id="263"/>
    <p:sldId r:id="rId9" id="267"/>
    <p:sldId r:id="rId10" id="264"/>
    <p:sldId r:id="rId11" id="265"/>
    <p:sldId r:id="rId12" id="266"/>
    <p:sldId r:id="rId13" id="261"/>
</p:sldIdLst>
```

그림 6. 슬라이드 리스트 요소 내 슬라이드 식별자
Fig. 6. The slide identifier in the slide list element

일반 사용자들은 파워포인트 파일 문서를 작성할 때, 매번 새로 작성하는 것이 아니라 기존의 문서를 복사하여 불필요한 슬라이드를 삭제하고 상황에 맞게 슬라이드 순서를 조정하고 필요한 내용이 있을 시 슬라이드를 새로 추가하는 등의 수정 작업을 한다. 이러한 사용자들의 파워포인트 작성 특성과 슬라이드 식별자의 포렌식 특성은 하나의 파일로부터 파생된 서로 다른 파일들 간의 편집 이력을 추적하는데 중요한 역할을 할 것이다.

5. OOXML형식의 Microsoft파워포인트 파일의 연관성 및 편집 이력 조사 방법

파워포인트 파일 생성 시 함께 부여되는 식별자는 유일한 값을 가지는 특성이 있으므로 파일의 파생여부를 알 수 있음을 이전 장의 설명을 통해서 알 수 있다. 이 장에서는 임의로 생성한 파일의 편집 이력 조사 방법을 앞 장에서 설명한 Microsoft 파워포인트 파일의 포렌식 속성을 이용하여 설명한다. 이러한 조사 방법은 파워포인트 문서의 편집이력을 파악하는 디지털 포렌식 수사 시 유용하게 쓰일 수 있다.

표 3. 슬라이드 식별자와 편집 이력 조사 내용(사선으로 채워진 칸은 삭제된 슬라이드이다.)
Table 3. Slide identifier and Edit History investigation(Filled with diagonal cells are deleted-slide)

편집 이력 분석			X파일	
슬라이드 마스터 아이디	2147483684	이동	슬라이드 마스터 아이디	2147483684
순차적으로 생성된 슬라이드 식별자			확장자	.ppsx
256			281	
257			279(HiddenSlide)	
258~263			280	
264			282	
265			283	
266~270			264	
271			265	
272			271	
273~276			272	
277			277	
278			257(HiddenSlide)	
279			285	
280				
281				
282				
283				
284				
285				

5.1 연관성 및 편집 이력 조사 내용

Microsoft 파워포인트로 작성된 임의의 T파일을 샘플로 설명한다. T파일에 대하여 4장에서 설명한 포렌식 속성을 조사하여 어떠한 연관성과 편집 이력을 갖는지 추적한다.

'.ppsx'확장자를 가진 Microsoft 파워포인트 쇼 파일은 편집이 불가능하다. T파일의 슬라이드 마스터 아이디가 아무것도 설정하지 않았을 때 갖는 값이 아닌 '2147483684'를 갖는다는 것은 고유의 서식을 가지고 있다는 의미이다. 고유의 서식을 갖는 파워포인트 파일은 어떠한 다른 파워포인트 파일과 비교 시 연관성을 파악하는데 도움이 된다. 전체적인 문장 구조와 사용단어가 유사성을 보인다거나 슬라이드 마스터 아이디 값이 동일하면 같은 파일로부터 파생되었다는 가능성이 높아지면서 연관성이 깊어진다.

다음으로 슬라이드 식별자를 조사하여 Table 3에 나타내었고 T파일의 슬라이드 편집 이력을 분석하였다. 또한 T파일에는 숨겨진 슬라이드(HiddenSlides)가 발견되었다. Table 3을 보면 순차적으로 생성된 슬라이드 식별자와는 다르게 T파일은 슬라이드가 화살표처럼 이동되었음을 확인할 수 있다. 또한 일부 슬라이드 식별자(279, 257)는 의도적으로 숨겨졌고, T파일에는 존재하지 않는 식별자 값 256, 258~263, 266~270, 273~276, 278, 284는 편집 중 슬라이드가 삭제되었음을 알 수 있다.

5.2 조사 결과

고유한 슬라이드 마스터 아이디를 갖는다는 것은 파일이 특정 서식을 갖추고 있다는 의미이다. 독립적으로 파일을 생성할 경우 아무 서식을 만들지 않는 한 동일 서식을 만들어내기는 흔치 않기 때문에 서로 다른 파일이 있을 시 연관성 조사에 도움이 될 수 있다. 연관성을 파악하는데 도움이 되는 또 다른 정보로는 생성자와 생성시간이 있다. 생성자와 생성시간의 분석은 동일 생성자로부터 파일이 만들어졌는지 파악할 수 있게 한다.

T파일에서는 슬라이드 식별자 값 256 등 일부 슬라이드가 보이지 않음으로 보아 삭제 작업이 이뤄졌음을 알 수 있다. 게다가 T파일의 슬라이드 식별자는 순차적인 흐름이 아닌 불연속성을 보이므로 편집 중 슬라이드들이 삽입 생성되거나 이동되었음을 파악할 수 있다. 또한 숨겨진 슬라이드를 가지고 있는 T파일은 파일 작성 시 의도적으로 보이지 않게 하기 위해 숨겨졌다는 확증이 된다.

이러한 내용들을 확인할 수 있다는 것을 일반 사용자는 쉽게 알 수 없고, Microsoft 파워포인트의 내부 형식에 대해 전문적으로 알지 못하면 수정 또한 어렵다. 따라서 이러한 분석은 파워포인트 파일들의 편집이력을 조사하는데 상당한 도움이 될 수 있다.

6. OOXML형식의 Microsoft파워포인트 파일의 포렌식 조사 방법

지금까지 설명한 OOXML의 포렌식 속성과 파워포인트 파일의 편집이력 조사 방법을 통하여 이 장에서는 Fig. 7과 같이 원본과 파생본의 파일 추적성을 고려하여 발생 가능한 가상의 사건을 소개한다. 사건의 조사 과정 및 방법은 앞장에서 설명한 Microsoft 파워포인트 파일의 포렌식 속성을 이용하여 설명한다. 이러한 디지털 포렌식 조사의 결과는 사건의 정황 증거로 활용 될 수 있다.

6.1 사건 배경

기술개발업체인 X사에 근무하는 A씨는 연구 기획 팀장으로 근무하며 다양한 프로젝트를 직접 통제하기 위해 팀원

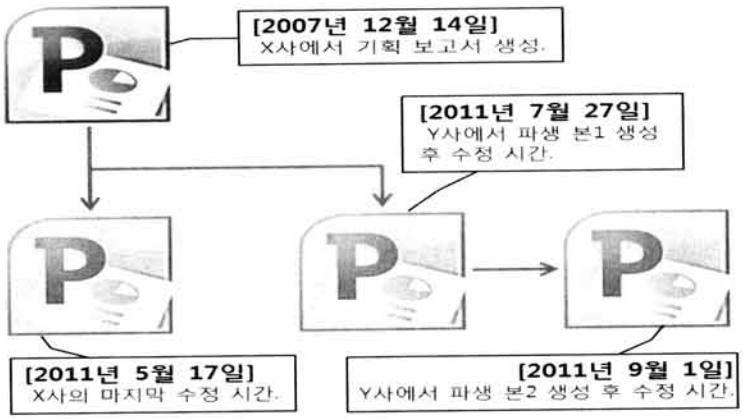

그림 7. 파워포인트 원본과 파생 파일의 추적
Fig. 7. Tracking of the source and derived PowerPoint files

들의 연구 개발 문서 관리 등 전반적인 연구소 관리 업무를 수행 했다. A씨는 일정기간 근무 후 퇴사를 하게 되었고 경쟁업체인 Y사에 취업했다. 몇 년 후인 2011년, 유명 제조업체인 Z사는 제품을 생산하는데 있어 Y사로부터 받은 기술 제안서를 검토 후 Y사의 기술을 적용하여 제품을 개발하기로 하였다.

수개월 후 Y사의 기술이 적용된 부품이 X사가 개발한 기술과 유사성을 보이면서 외부로 반출될 수 없는 기획 문서 파일이 유출되었다는 수상한 정황을 발견하였고, 이에 대한 조사가 시작 되었다. 오랜 시간이 지난 후 남아있는 증거는 Z사에 건네진 기술 제안서와 X사의 연구개발문서인 기획 보고서 파일이다. 사건해결에 실마리를 줄 수 있는 중요한 기획 보고서 문서들은 Microsoft 파워포인트로 작성된 것이다.

증거에 대한 초기 조사 결과 Z사가 가지고 있는 Y사의 기술 제안서는 편집이 불가능한 '.ppsx'확장자를 가진 Microsoft 파워포인트 쇼 파일임에도 불구하고 X사의 기획 보고서와 매우 유사한 단어를 사용하고 있었다. 또한 수상하게도 전체적인 문장의 구조와 서식도 유사성을 보임을 발견하였다.

6.2 조사 목적

X사에서 작성된 기획 보고서가 Y사로 유출되어 이를 이용하여 기술 제안서를 작성하였는지를 판단한다.

6.3 조사 대상

Table 4와 같이 X사에서 작성된 기획 보고서와 Y사가 Z사에게 건넨 기술 제안서와 제안 요약서를 조사한다.

6.4 조사 방법

조사 대상인 Table 4의 Microsoft 파워포인트 파일에 대하여 4장에서 설명한 포렌식 속성을 가지는 저장된 자동 복구 정보, 문서 속성, 슬라이드 식별자, 슬라이드 마스터 식별자(서식) 등을 조사한다. 이를 통해 X사의 기획 보고서와 Y사의 기술 제안서가 서로 연관성이 있는지, 각각 독립적으로 생성된 것인지를 판단한다.

표 4. 조사 대상 분류
Table 4. The classification of investigation files

	분석 대상	파일 종류
X사	기획보고서	Microsoft 파워포인트 (.pptx)
Y사	기술 제안서 (제안 요약, 제안서)	Microsoft 파워포인트 쇼 (.ppsx)

6.5 조사 내용

6.5.1 문서 속성

① 생성 시간

X사가 작성한 기획 보고서 Microsoft 파워포인트 파일의 생성 시간과 Y사가 제안서 작성에 사용된 파일들의 생성시간이 동일하였다. 이 속성 정보를 확인함으로써 X사와 Y사의 Microsoft 파워포인트 문서의 원본 파일이 동일할 가능성이 높아진다.

② 마지막 수정자(마지막으로 수정하여 저장한 사람)

X사가 가지고 있는 Microsoft 파워포인트의 마지막으로 저장한 사람은 X사 내부 직원의 ID로 되어 있었으나 Y사의 기술 제안을 목적으로 사용된 Microsoft 파워포인트 문서의 마지막 수정자는 과거 X사에 있었으나 현재는 Y사에서 근무하는 A씨의 ID로 확인 되었다.

③ 수정시간(마지막으로 수정하여 저장한 시간)과 Total Time

Y사의 기술 제안서에 적용된 Microsoft 파워포인트 파일의 마지막 저장 시간이 X사의 기획 보고서의 수정시간보다 더 나중이다. 또한 Y사의 TotalTime이 X사의 TotalTime 보다 큰 값을 가지고 있었다. 즉, 동일 시간에 생성된 Microsoft 파워포인트 문서지만 X사의 기획 보고서보다 Y

사의 기술 제안서가 더 오랜 시간동안 편집 작업이 있었음을 알 수 있다.

④ HiddenSlides(숨긴 슬라이드)

X사의 기획보고서 파일에는 숨겨진 슬라이드가 발견되지 않았으나, Y사의 제안서에 삽입된 Microsoft 파워포인트 파일에는 숨겨진 슬라이드가 있음이 확인되었다. 위에서 조사한 속성 결과를 Table 5에 정리하였다.

6.5.2 식별자

유사한 서식을 가지고 있는 각 파일들의 슬라이드 마스터 아이디를 조사하였다. 조사 대상으로 있는 세 파일은 모두 같은 슬라이드 마스터 아이디 값을 가지고 있음이 확인되었다. 같은 서식을 가지고 있는 각 파일들의 연관성이 깊어졌다.

다음으로 슬라이드 식별자를 조사하였다. 조사 대상 파일 내의 모든 슬라이드 식별자를 Table 6에 나타내었고 X사의 기획보고서와 Y사의 기술 제안서 간의 슬라이드 편집 관계를 파악하였다.

6.6 조사 결과

Microsoft 파워포인트의 생성시간, 생성자, 슬라이드 마스터 아이디를 조사한 결과 X사의 기획 보고서가 복사되어 Y사의 기술제안서가 수정되어 작성된 것으로 판단된다. 그 이유는 다음과 같다.

슬라이드 마스터 아이디가 모두 동일한 값을 가진다는 것을 확인하였고, 동일 값을 갖는다는 것은 세 파일이 전부 같은 서식을 사용한다는 것이다. 독립적으로 파일을 생성할 경우 아무 서식을 만들지 않는 한 동일 서식을 만들어내기는 흔치 않다. 때문에 같은 슬라이드 마스터 아이디를 갖는다는 것은 세 파일중 하나가 원본이거나 세파일 모두 같은 파일로부터 복사되어진 파일이라는 것을 추정할 수 있다.

X사와 Y사의 파일 생성자와 생성시간이 같다 것 또한 모두 같은 생성자로부터 파일이 만들어졌다는 것을 알 수 있다. Y사의 기술제안서에서는 슬라이드 식별자 값 256 등 일부 슬라이드가 보이지 않음으로 보아 삭제작업이 이뤄졌음을 알 수 있다. Y사의 슬라이드 아이디에는 X사에서 가지고 있는 슬라이드 아이디보다 더 큰 값이 있다는 점은 기

표 5. 조사 대상 파일의 문서 속성 내용 요약
Table 5. Summary of the document properties of the investigation file

	X사	Y사	
분류	Microsoft파워포인트 (기획보고서)	Microsoft파워포인트 쇼 (기술제안서)	Microsoft 파워포인트 쇼 (제안 요약)
생성시간	2007년 12월 14일 화요일, 오전 9:33:51	좌동	좌동
수정자	X사 직원	A씨	좌동
수정시간	2011년 5월 17일 화요일, 오후 6:37:14	2011년 7월 27일 수요일, 오후 3:31:35	2011년 9월 1일 목요일, 오전 8:58:50
숨긴 슬라이드	없음	2개	5개

표 6. 슬라이드 식별자 조사 내용과 슬라이드 편집 관계 파악(방사형 배경 칸은 추가 생성된 슬라이드이다.)
Table 6. Investigate slide identifier and analyze Edit the relationship between the slide(Filled with croos striped cells are created-slide)

	X사	관계 연결	Y사	
분류	기획보고서 .pptx		기술 제안서 .ppsx	제안 요약서 .ppsx
슬라이드 식별자	256		281	281
	257		279(HiddenSlide)	279(HiddenSlide)
	258		280	280
	260		282	282
	261		283	283(HiddenSlide)
	264		260	260
	...		261	261(HiddenSlide)
	273		264~265	264~265
	275		267~268	267~268
	276		270~273	270~273
	277		275	275
	279		276	276(HiddenSlide)
	280		277	277
			257(HiddenSlide)	257(HiddenSlide)
			258	258
			285	285

획 보고서를 복사해서 슬라이드를 새로 추가하여 기술 제안서를 편집한 사실을 뒷받침하는 결정적 증거이다. 또한 숨겨진 슬라이드를 많이 가지고 있는 제안 요약서 파일은 기술제안서 작성 시 의도적으로 보이지 않게 하기 위해 숨겨졌다는 확증이 된다.

이러한 내용들을 확인 할 수 있다는 것을 일반 사용자는 쉽게 알 수 없고, Microsoft 파워포인트의 내부 형식에 대해 전문적으로 알지 못하면 수정 또한 어려우므로 조사 대상의 파일들은 변경되지 않은 무결성을 갖추고 있다고 할 수 있다.

7. 결 론

기업과 개인의 정보를 담고 있는 문서 파일이 자산으로 취급되면서 문서 관리가 중요해진 오늘날 문서파일의 유출로 디지털 포렌식 조사가 필요한 경우가 많다. 현재까지는 Microsoft office 2007 하위 버전인 복합문서 형식 파일에 대한 디지털 포렌식 조사가 주를 이루었겠지만 OOXML형식이 사용되기 시작한 시점에서는 문서파일의 증거 확보가 미비하다. 문서파일의 증거 확보에 관련하여 워드에 대한 연구는 있었으나 파워포인트에 대한 자세한 연구가 없었다. 따라서 본 논문은 파워포인트 파일의 편집 이력 추적과 서로 다른 문서간의 연관성 조사에 활용될 수 있어 증거 확보의 증대를 이끌어 줄 것으로 기대된다.

대부분의 중요한 개인 및 기업의 발표, 보고서 등의 수많은 문서들은 Microsoft office의 파워포인트로 제작되며, 보통 제작자는 이전의 서식을 유지하기 위해 파일을 복사 및 수정하여 내용을 추가하거나 삭제하는 작업을 하는 특성을 가진다. 이러한 사용자의 특성을 이용하여 파워포인트 파일을 분석하면 문서 간의 연관성을 파악해야 하거나 편집이력을 추적해야 할 필요가 있을 경우 정황 파악에도 큰 도움을 줄 수 있을 것이다.

이 논문에서는 OOXML을 사용하는 Microsoft 파워포인트 응용프로그램에 의해 생성될 수 있는 포렌식 속성을 소개하였다. 이러한 속성들은 응용프로그램에 의해 자동으로 저장되는 데이터로 일반적인 사용자는 그 내용을 쉽게 알 수 없다. 이렇게 쉽게 알지 못하는 파워포인트의 포렌식 속성과 일반 사용자들의 파워포인트 제작 특성을 이용한 파일 편집 이력 조사 과정 및 방법을 설명하였다. 이러한 편집 이력 조사 결과는 사건의 정황 증거로 활용될 수 있다.

참 고 문 헌

[1] Simon Byers, "Information leakage caused by hidden data in published documents", *IEEE Security & Privacy*, Vol.2, No.2, pp.23-27, 2004.

[2] A. Castiglionea, A. De Santisa and C. Sorienteb, "Taking advantages of a disadvantage: Digital forensics and steganography using document metadata", *The Journal of Systems and Software*, Volu.80, Issue 5, pp.750-764, May, 2007.

[3] Simson L. Garfinkel and James J.Migletz, "New XML-Based Files Implications for Forensics", *IEEE Security & Privacy*, Vol.7, Issue 2, pp.38-44, 2009.

[4] Park Jungheum and Lee Sangjin, "Forensic investigation of Microsoft PowerPoint files", *Digital Investigation 6(1-2)*, pp.16-24, 2009.

[5] Aniello Castiglione, Bonaventura D'Alessio, Alfredo De Santis and Francesco Palmieri, "New Steganographic Techniques for the OOXML File Format", *Availability, Reliability and Security for Business, Enterprise and Health Information Systems (C)*, LNCS Vol.6908, pp.344-358, 2011.

[6] Microsoft Corporation, Standard ECMA-376, Office Open XML file formats. 3rd edition, 2011.

[7] Zhangjie Fu, Xingming Sun, Yuling Liu and Bo Li, "Forensic investigation of OOXML format documents", *Digital Investigation*, Vol.8, Issue 1, pp.48-55, July, 2011.

[8] Microsoft Corporation, "XML file name extensions in Office 2010", URL : http://technet.microsoft.com/en-us/library/cc179191.aspx, June, 2010.

윤 지 혜

e-mail : jihey_2000@nate.com

2007년 2월 성결대학교 정보통신공학과 (공학사)

2009년 9월~현 재 정보경영공학전문대학원 석사과정

관심분야 : 모바일 포렌식, 스마트기기 보안, 네트워크보안, 네트워크 포렌식 등

박 정 흠

e-mail : junghmi@korea.ac.kr

2007년 2월 한양대학교 정보통신대학 컴퓨터전공(공학사)

2009년 2월 고려대학교 정보경영공학전문대학원(공학석사)

2009년 3월~현 재 고려대학교 정보경영공학전문대학원 박사과정

관심분야 : 디지털 포렌식, 안티-안티 포렌식 등

이 상 진

e-mail : sangjin@korea.ac.kr

1987년 2월 고려대학교 수학과

1989년 2월 고려대학교 수학과(이학석사)

1994년 8월 고려대학교 수학과(이학박사)

1989년 10월~1999년 2월 ETRI 선임연구원

1999년 3월~2001년 8월 고려대학교 자연과학대학 조교수

2001년 9월~현 재 고려대학교 정보경영공학전문대학원 교수

관심분야 : 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수 등