Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.24, NO.2, Apr. 2014 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2014.24.2.319

MS 엑셀 파일의 텍스트 셀 입력 순서에 관한 연구*

이 윤 미,† 정 현 지, 이 상 진‡
고려대학교 정보보호대학원

A Study on Edit Order of Text Cells on the MS Excel Files*

Yoonmi Lee,† Hyunji Chung, Sangjin Lee‡
Center for Information Security Technologies, Korea University

요 약

스마트폰이나 태블릿 PC 사용이 보급화 되면서 장소에 구애받지 않고 실시간으로 문서의 생성과 편집이 일어나고 있다. 이처럼 학교나 회사에서 업무처리 방법의 한 부분을 차지하고 있는 문서 파일들을 분석하여 데이터가 입력되거나 편집된 흐름을 추적할 수 있다면 디지털 포렌식 수사에서 증거 자료로 활용될 수 있을 것이다. 대표적인 문서 프로그램으로 Microsoft 사의 Office 시리즈를 꼽을 수 있다. MS Office 프로그램은 복합 문서 파일 형식(Compound Document File Format)을 사용하는 97-2003 버전, OOXML 파일 형식(Office Open XML File Format)을 사용하는 2007-현재 버전까지 두 가지 파일 형식으로 구성된다. 지금까지 연구된 MS 파일에 대한 디지털 포렌식 분석 방법은 파일에 은닉된 정보를 탐지하거나 문서의 속성 정보를 통해 위변조 여부를 판단하는 것이었다. 본 논문에서는 디지털 포렌식 관점에서 MS 엑셀 파일에 텍스트 셀이 입력된 순서를 분석하여 문서의 입력 순서와 마지막으로 수정한 셀을 파악하는 방법을 연구하였다.

ABSTRACT

Since smart phones or tablet PCs have been widely used recently, the users can create and edit documents anywhere in real time. If the input and edit flows of documents can be traced, it can be used as evidence in digital forensic investigation. The typical document application is the MS(Microsoft) Office. As the MS Office applications consist of two file formats that Compound Document File Format which had been used from version 97 to 2003 and OOXML(Office Open XML) File Format which has been used from version 2007 to now. The studies on MS Office files were for making a decision whether the file has been tampered or not through detection of concealed items or analysis of documents properties so far. This paper analyzed the input order of text cells on MS Excel files and shows how to figure out what cell is the last edited in digital forensic perspective.

Keywords: Digital Forensics, MS Excel, Document File Forensics, OOXML File Format, Compound Document File Format

I. 서 론

Microsoft(이하 MS) Office는 문서 작업에 사용하는 대표적인 프로그램이다. PC나 모바일 등 다양한 기기에서 지원하고, 다른 종류의 문서와 호환성도 뛰어나 전 세계적으로 문서 관련 소프트웨어 시장 점유율이 93% 이상이다[1].

MS Office 97부터 2003 버전까지는 복합 문서

접수일(2013년 10월 21일), 수정일(2014년 2월 14일), 게재확정일(2014년 3월 6일)
* 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단-공공복지안전사업의 지원을 받아 수행된 연구임(2012M3A2A1051106)
† 주저자, 2yoonmi@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr(Corresponding author)

파일 형식(Compound Document File Format)을 사용하였고, 2007 버전부터 현재까지 OOXML 파일 형식(Office Open XML File Format)을 사용한다[3].

MS 엑셀은 스프레드시트(spreadsheet) 프로그램의 한 종류로, 표 형식으로 나타낸다. 이 표는 셀(cell) 단위로 나뉘고, 하나의 셀은 행(column; x)과 열(row; y)로 이름이 매겨지는 매트릭스 구조의 저장방식이다[6]. 셀에는 숫자, 텍스트, 수식 등의 값을 입력할 수 있다. 그 중 텍스트 셀은 MS 엑셀 파일 형식의 특성에 의해 텍스트가 입력된 순서를 파악할 수 있다.

엑셀 시트에 데이터가 입력되거나 편집된 흐름을 추적할 수 있다면 사용자의 행위 분석과 데이터 변조 측면에서 디지털 포렌식 수사에 증거 자료로 활용될 가능성이 있다.

본 논문에서는 MS Office 엑셀 파일의 텍스트 셀 입력 순서를 디지털 포렌식 관점에서 분석하고 분석한 내용을 실제 조사에 활용할 수 있는 방법을 제시한다.

II. 관련 연구

디지털 포렌식 관점에서 MS Office 문서 파일 분석에 관한 연구가 다수 이루어졌다.

Zhangjie Fu 등[3]은 의심스러운 MS 워드 문서의 출처 파악을 위해 문서로부터 추출한 고유한 값인 수정 식별자(Revision Identifier, RI) 값을 비교하여 해당 문서의 복사본과 원본 구분 방법을 제시하였다. 수정 식별자 값은 문서를 수정할 때마다 새로 생성되기 때문에 원본의 수정 식별자 값과 의심되는 파일의 수정 식별자 값을 비교하면 위변조 여부를 판단할 수 있다. 또한 OOXML 형식 내부의 *'core.xml'* 파트로부터 생성자 및 시간 정보를 통해 실제 문서 소유자를 파악할 수 있다는 것을 밝혔다.

Bora Park 등[5]은 MS Office 2007에서 사용자 정의 파트와 파트 간의 관계(relationship)를 이용하여 문서 파일 안에 정보를 숨기는 방법과 은닉된 정보를 탐지하는 방법에 대해 제시하였다. 사용자 정의 파트란 관계 파일에서 유효한 타겟으로 지정되지 않은 모든 파트를 의미하는데, MS Office 프로그램 자체에서 이를 확인하지 않기 때문에 해당 부분에 임의의 파일을 삽입할 수 있다.

지금까지 연구된 MS Office 파일에 대한 디지털 포렌식 분석 방법은 파일 안에 은닉된 정보를 탐지하거나 문서의 속성 및 식별자 값 등을 통해 위변조 여부를 판단하는 것들이었다. 본 논문은 MS 엑셀 파일 분석을 통해 사용자의 텍스트 입력 순서를 파악할 수 있고, 이러한 흔적은 사용자 행위 분석과 데이터 변조 증거로 활용할 수 있음을 설명한다.

III. MS 엑셀 파일 형식

3.1 복합 문서 파일 형식

복합 문서 파일 형식은 MS Office 97부터 2003 버전에서 제공하는 문서 파일 형식이며, *.xls*, *.ppt*, *.doc* 등의 확장자로 저장된다.

복합 문서 파일 형식은 구조가 FAT 또는 NTFS 등의 파일 시스템과 유사하다. 스토리지(storage)와 스트림(stream)의 계층구조로 구성되며, 스토리지와 스트림을 관리하기 위한 메타데이터가 존재한다.

스토리지는 파일 시스템의 폴더와 동일한 기능을 하며, 실제 데이터는 존재하지 않는다. 스트림은 실제 데이터 파일을 의미한다. 메타데이터에는 스토리지와 스트림에 접근하기 위해 필요한 정보를 담고 있다[7].

3.2 OOXML 파일 형식

OOXML 파일 형식은 MS Office 2007 버전부터 현재까지 사용하고 있는 문서 파일 형식으로 *.xlsx*, *.pptx*, *.docx* 등의 확장자로 저장된다.

OOXML 파일은 여러 개의 파트(part)들과 파트 간의 관계를 나타내는 관계(relationship) 등이 패키지(package)라는 컨테이너에 ZIP 압축 형식으로 저장되어 있다[3].

파트는 패키지 내에 존재하는 파일들을 의미하며, 각 파트는 서로 다른 콘텐츠의 유형(content type)을 갖는다. 예를 들어, *'sheet#.xml'* 파트는 엑셀의 본문 내용을 저장하고, *'app.xml'*, *'core.xml'* 파트는 문서 속성에 대한 정보를 담고 있다[2].

IV. MS 엑셀 파일 분석

이 절에서는 MS 엑셀 파일을 분석하여 사용자가 입력한 숫자, 텍스트 등의 값들이 파일 내부에 어떤 방식으로 저장되는지 설명한다.

4.1 데이터 저장 방식

MS 엑셀은 데이터를 표 형식으로 처리하는 문서 프로그램으로 스프레드시트(spreadsheet)라고 한다. 표는 여러 개의 셀로 구성되어 있고, 하나의 셀은 행과 열로 이름이 매겨지는 매트릭스 구조이다.

행과 열은 (x, y) 형태로 각 셀에 저장된 값(value)과 값이 존재하는 위치를 'XY' 형식으로 저장한다.

4.2 숫자 입력

복합 문서 파일 형식의 엑셀 파일에서 숫자 값은 본문의 내용이 저장되는 *'Workbook 스트림'*의 *'Worksheets 서브스트림(Substream)'* 영역에 저장된다. *'Worksheet# 서브스트림'*은 각 *worksheet* 별로 나뉘어져 있으며 *worksheet*의 개수만큼 생성된다.

엑셀 파일의 내부 구조에서는 숫자를 입력한 순서와 관계없이 각 셀에 입력된 값을 행→열 순으로 순차적으로 저장하기 때문에 숫자 값이 어떤 순서로 입력되었는지 알 수 없다. 또한, 이 값은 'IEEE 754' 부동소수점 형식(64-bit binary floating-point number)으로 저장[8]되기 때문에 별도의 변환 과정을 거치지 않으면 파일의 내부 구조에서 눈으로 숫자 값을 읽기는 어렵다.

OOXML 파일 형식의 엑셀 파일에서 숫자 값은 본문의 내용이 저장되는 *'sheet#.xml'* 파트에 저장된다. Fig.1.은 숫자를 입력했을 때, *'sheet#.xml'* 파트에 저장되는 정보를 보여준다. MS 엑셀의 셀에 숫자를 입력하면 '값(value)'을 뜻하는 *'〈v〉'* 요소 사이에 입력한 숫자 값이 그대로 저장된다. 이 때, *'〈c r〉'* 요소로 해당 숫자 값이 입력된 셀의 위치를 나타낸다.

OOXML 파일 형식의 엑셀 파일에서 숫자 값은 복합 문서 파일 형식과 마찬가지로 셀에 해당하는 값이 저장될 뿐 입력된 순서를 나타내는 요소는 존재하지 않는다.

음수, 소수, 백분율, 그리고 셀 서식 메뉴를 이용하여 날짜 혹은 시간 등의 형식을 따르는 숫자 표시 또한 *'sheet#.xml'* 파트를 통해 어떤 값이 입력되었는지는 알 수 있으나 작성 순서는 알 수 없다.

```
<sheetData>
  - <row r="2" x14ac:dyDescent="0.3" spans="2:3">
      - <c r="B2">
            <v>2222</v>
        </c>
      - <c r="C2">
            <v>4444</v>
        </c>
    </row>
  - <row r="3" x14ac:dyDescent="0.3" spans="2:3">
      - <c r="B3">
            <v>3333</v>
        </c>
      - <c r="C3">
            <v>1111</v>
        </c>
    </row>
</sheetData>
```

Fig.1. The numeral cell values in the 'sheet#.xml' part

```
<row r="3" x14ac:dyDescent="0.3" spans="2:4">
  - <c r="B3" t="s">
        <v>5</v>
    </c>
  - <c r="C3" t="s">
        <v>2</v>
    </c>
  - <c r="D3" t="s">
        <v>0</v>
    </c>
</row>
<row r="4" x14ac:dyDescent="0.3" spans="2:4">
  - <c r="B4" t="s">
        <v>3</v>
    </c>
  - <c r="D4" t="s">
        <v>6</v>
    </c>
</row>
```

Fig.2. The index values of text cell in the 'sheet#.xml' part

4.3 텍스트 입력

MS 엑셀 파일은 숫자와 달리 텍스트 입력에 대해 입력한 순서를 별도로 관리한다.

복합 문서 파일 형식의 엑셀 파일에서 텍스트를 입력하면 *'Workbook 스트림'*의 *'Globals 서브스트림'* 영역에 저장된다. 그 중 *'SST(string constants) 레코드'* 영역을 살펴보면 엑셀 파일에 텍스트를 입력한 순서대로 데이터를 저장하고 있다. 'XLUnicodeRichExtendedString' 형식으로 표현하고 있으며, MS에서 제공하는 'OffVis(Office Visualization Tool)[9]'를 통해 텍스트가 입력된 순서대로 저장되었다는 것을 쉽게 확인할 수 있다.

```
<si>
    <t>Friday</t>
    <phoneticPr type="noConversion" fontId="1"/>
</si>
<si>
    <t>Saturday</t>
    <phoneticPr type="noConversion" fontId="1"/>
</si>
<si>
    <t>Sunday</t>
    <phoneticPr type="noConversion" fontId="1"/>
</si>
<si>
    <t xml:space="preserve">the end of text </t>
    <phoneticPr type="noConversion" fontId="1"/>
</si>
```

Fig.3. The text values of text cell in the 'sharedStrings.xml' part

OOXML 파일 형식의 엑셀 파일에서 텍스트를 입력하면 *'sheet#.xml'* 파트에 각 셀 별로 텍스트가 입력된 순서를 나타내는 인덱스 값이 생성된다. 인덱스 값은 '0'부터 시작하며, 인덱스 '0' 값을 갖는 셀은 가장 처음에 입력된 텍스트 셀을 뜻한다. Fig.2.와 같이 *'〈v〉'* 요소 사이에 텍스트 입력 순서를 나타내는 숫자 값이 생성된 것을 확인할 수 있다[2].

인덱스 값과 매치되는 텍스트 값은 *'sharedStrings.xml'*이라는 별도의 파트에 존재한다. *'sharedStrings.xml'* 파트에서 가장 첫 줄에 저장된 텍스트 값은 *'sheet#.xml'* 파트에서 인덱스 '0'을 갖는 셀과 일치하고, 마지막 줄에 저장된 텍스트 값은 마지막 숫자의 인덱스 값을 갖는 셀과 일치한다. Fig.3.은 *'sharedStrings.xml'* 파트에 저장된 텍스트 값을 보여준다. Fig.3.에서 *'〈t xml:space="preserve"〉'* 요소는 텍스트 앞/뒤로 공백이 입력되었거나 셀에 공백만 입력된 경우를 나타낸다. 따라서 셀에 공백만 입력이 되어도 기록에 남는다는 것을 알 수 있다.

4.4 텍스트 셀 특성

일반적으로 텍스트 값은 작성한 순서대로 저장되나 셀을 복사하여 붙여넣기 하거나 자동 완성 기능을 사용하여 동일한 데이터를 채웠을 경우에는 최초 입력 값만 저장한다.

복합 문서 파일 형식의 MS 엑셀 파일은 이런 경우 어떤 셀이 복사/붙여넣기 하거나 자동 완성 기능을 사용하여 값이 채워졌는지 알 수 없으나 OOXML 파일 형식의 MS 엑셀 파일은 *'sheet#.xml'* 파트와 *'sharedStrings.xml'* 파트에서 확인할 수 있다.

'sharedStrings.xml' 파트에서 *'〈count〉'* 요소는 파일에 포함된 전체 텍스트 셀의 개수를 나타내며, *'〈uniqueCount〉'* 요소는 붙여넣기 하거나 자동 완성 기능을 사용하여 동일한 데이터가 입력된 셀을 제외한 셀의 개수를 나타낸다. *'sheet#.xml'* 파트에서 동일한 인덱스 값을 갖는 텍스트 셀은 붙여넣기 하거나 자동 완성 기능을 사용하여 데이터를 채운 셀이라는 것을 의미한다. 이러한 특성은 드래그하여 텍스트 값을 채운 경우에도 동일하게 발생한다.

이처럼 복합 문서 파일 형식의 *'SST 레코드'*, *'sheet#.xml'* 파트, *'sharedStrings.xml'* 파트를 통해 텍스트 셀의 입력 순서, 자동으로 데이터가 채워진 셀, 그리고 그 셀의 값을 파악할 수 있다.

V. MS 엑셀 파일 포렌식 조사 방법

5.1 사건 배경

아파트 건설 건에 대한 수급사업자인 A 건설은 원사업자인 B 건설과 계약을 체결한 상태이다. 어느 날 계약서를 검토하던 중 계약 내용이 기존에 합의된 사항과 다른 내용으로 작성된 것을 확인하였다.

'6. 대금의 지급' 항목에 기존 'contract.xlsx' 파일에는 존재하지 않았던 '부대조건' 항목이 나중에 확인한 'contract.xlsx' 파일에 추가되어 있었다. B 건설은 '부대조건'이 포함되어 있는 'contract.xlsx' 파일이 원본이라고 주장하고 있으며, 두 파일의 모든 시간 정보는 동일했다.

A 건설은 두 개의 'contract.xlsx' 파일 간의 위변조 여부 판별을 의뢰했다.

5.2 조사 목적

MS 엑셀 파일 분석을 통해 문서의 위변조 여부를 판별한다.

5.3 조사 대상

Fig.4.와 같이 B 건설에서 작성한 동일한 파일명의 'contract.xlsx' 파일 2개를 조사한다.

5.4 조사 방법

조사 대상인 두 개의 'contract.xlsx' 파일의 텍스트 입력 순서를 확인할 수 있는 '*sharedStrings.xml*', '*sheet1.xml*' 파트 분석을 통해 각 파일 간 텍스트가 입력된 순서와 계약서 내용상의 상이한 부분에 대해 분석한다.

5.5 조사 내용

'*sharedStrings.xml*' 파트에 입력된 텍스트 값들과 '*sheet1.xml*' 파트에 저장된 텍스트의 인덱스 값을 비교하기 위해 문서의 첫 행('건설공사 표준계약서(기본, 변경)')부터 마지막 행('성명:대표이사 OOO(인)')까지 모든 텍스트 값과 인덱스 값을 출력하고 비교했다.

5.6 조사 결과

두 개의 'contract.xlsx' 파일을 조사한 결과 첫 번째 파일은 첫 행부터 마지막 행까지 순차적으로 텍스트가 입력된 것을 확인했다. 그러나 두 번째 파일의 경우 '6. 대금의 지급 ※ 부대조건'에 해당하는 3개의 행이 가장 마지막에 입력된 것을 확인했다.

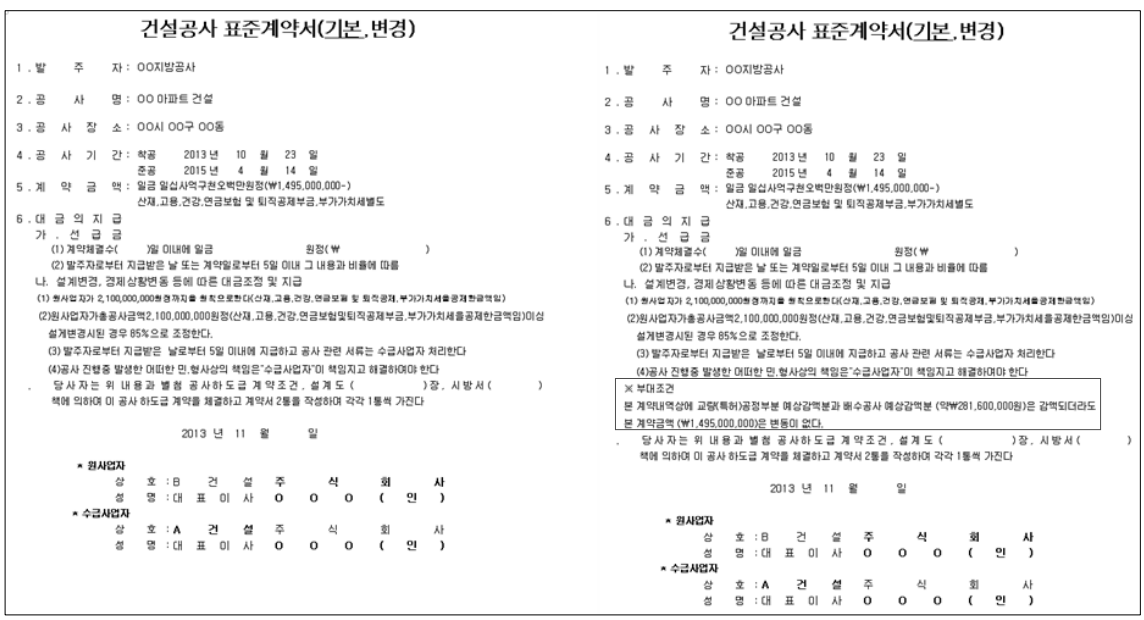
결과적으로 두 번째 'contract.xlsx' 파일은 B 건설이 A 건설에게 계약내용을 제시한 이후 '※ 부대조건' 항목에 대한 텍스트 셀을 의도적으로 추가 삽입한 것으로 판단되기 때문에 두 번째 'contract.xlsx' 파일은 위변조 되었다.

VI. 결 론

본 논문에서는 MS Office 프로그램 중 MS 엑셀 파일 분석을 통하여 사용자가 텍스트를 입력한 순서를 추정하는 방법에 대해 연구했다.

MS 엑셀 파일은 복합 문서 파일 형식, OOXML 파일 형식 등 두 가지 파일 형식을 사용하며, 숫자 값을 입력하는 것과 달리 텍스트 입력에 대해 데이터를 입력한 순서를 별도로 관리한다.

복합 문서 파일 형식의 MS 엑셀 파일은 'SST 레코드' 영역에 텍스트를 입력한 순서대로 데이터를 저장한다. OOXML 파일 형식의 MS 엑셀 파일에서는 'sheet#.xml', 'sharedStrings.xml' 등 두 파트를 매치하여 텍스트가 입력된 셀의 순서를 알아낼 수 있다. 'sheet#.xml' 파트에 각 셀 별로 텍스트가 입력된 순서를 나타내는 인덱스 값이 저장되고, 'sharedStrings.xml' 파트에는 'sheet#.xml' 파트의 인덱스 값과 매치되는 실제 텍스트 값이 저장되어 텍스트 입력 순서 및 마지막으로 수정한 셀이 무엇인지 파악 가능하다.


건설공사 표준계약서(기본,변경)

1. 발 주 자 : OO지방공사
2. 공 사 명 : OO 아파트 건설
3. 공 사 장 소 : OO시 OO구 OO동
4. 공 사 기 간 : 착공 2013년 10 월 23 일
준공 2015년 4 월 14 일
5. 계 약 금 액 : 일금 일십사억구천오백만원정(₩1,495,000,000-)
산재,고용,건강,연금보험 및 퇴직공제부금,부가가치세별도
6. 대 금 의 지 급
가. 선 급 금
(1) 계약체결수( )일 이내에 일금 원정(₩)
(2) 발주자로부터 지급받은 날 또는 계약일로부터 5일 이내 그 내용과 비율에 따름
나. 설계변경, 경제상황변동 등에 따른 대금조정 및 지급
(1) 동사업자가 2,100,000,000원증액지출 청직으로한다(산재,고용,건강,연금보험 및 퇴직공제,부가가치세출공제한금액임)
(2)원사업자가출공사금액2,100,000,000원정(산재,고용,건강,연금보험및퇴직공제부금,부가가치세출공제한금액임)이상
설계변경시된 경우 85%으로 조정한다.
(3) 발주자로부터 지급받은 날로부터 5일 이내에 지급하고 공사 관련 서류는 수급사업자 처리한다
(4)공사 진행중 발생한 어떠한 민,형사상의 책임은"수급사업자"이 책임지고 해결하여야 한다
. 당사자는 위 내용과 별첨 공사하도급 계약조건, 설계도( )장, 시방서( )
책에 의하여 이 공사 하도급 계약을 체결하고 계약서 2통을 작성하여 각각 1통씩 가진다

2013 년 11 월 일

* 원사업자
상 호 : B 건 설 주 식 회 사
성 명 : 대 표 이 사 O O O (인)
* 수급사업자
상 호 : A 건 설 주 식 회 사
성 명 : 대 표 이 사 O O O (인)

건설공사 표준계약서(기본,변경)

1. 발 주 자 : OO지방공사
2. 공 사 명 : OO 아파트 건설
3. 공 사 장 소 : OO시 OO구 OO동
4. 공 사 기 간 : 착공 2013년 10 월 23 일
준공 2015년 4 월 14 일
5. 계 약 금 액 : 일금 일십사억구천오백만원정(₩1,495,000,000-)
산재,고용,건강,연금보험 및 퇴직공제부금,부가가치세별도
6. 대 금 의 지 급
가. 선 급 금
(1) 계약체결수( )일 이내에 일금 원정(₩)
(2) 발주자로부터 지급받은 날 또는 계약일로부터 5일 이내 그 내용과 비율에 따름
나. 설계변경, 경제상황변동 등에 따른 대금조정 및 지급
(1) 동사업자가 2,100,000,000원증액지출 청직으로한다(산재,고용,건강,연금보험 및 퇴직공제,부가가치세출공제한금액임)
(2)원사업자가출공사금액2,100,000,000원정(산재,고용,건강,연금보험및퇴직공제부금,부가가치세출공제한금액임)이상
설계변경시된 경우 85%으로 조정한다.
(3) 발주자로부터 지급받은 날로부터 5일 이내에 지급하고 공사 관련 서류는 수급사업자 처리한다
(4)공사 진행중 발생한 어떠한 민,형사상의 책임은"수급사업자"이 책임지고 해결하여야 한다
※ 부대조건
본 계약내역상에 교량(특허)공정부분 예상감액분과 폐수공사 예상감액분 (약₩281,600,000원)은 감액되더라도
본 계약금액 (₩1,495,000,000)은 변동이 없다.
. 당사자는 위 내용과 별첨 공사하도급 계약조건, 설계도( )장, 시방서( )
책에 의하여 이 공사 하도급 계약을 체결하고 계약서 2통을 작성하여 각각 1통씩 가진다

2013 년 11 월 일

* 원사업자
상 호 : B 건 설 주 식 회 사
성 명 : 대 표 이 사 O O O (인)
* 수급사업자
상 호 : A 건 설 주 식 회 사
성 명 : 대 표 이 사 O O O (인)

Fig.4. Two 'contract.xlsx' files; 1st file(left), 2nd file(right)

특히, 문자 셀에 자동 완성 기능을 사용하여 데이터를 채운 경우, 드래그하여 데이터를 채운 경우 등 우리가 흔히 이용하는 엑셀의 편리한 기능에 대해서도 사용 흔적이 존재한다.

'5절. MS 엑셀 파일의 포렌식 조사 방법'에서 제시했듯이 엑셀 파일에서 작성된 표, 특히 계약이나 회계와 관련된 문서는 텍스트가 일반적으로 위에서 아래로 순차적으로 입력된다. 이러한 특성으로 계약서나 회계 장부 조작을 위해 중간에 셀을 삽입하거나 텍스트를 수정한 경우 문서의 텍스트 입력순서를 파악하여 추가적으로 작업된 셀인지 판단할 수 있다.

이 논문에서는 텍스트 셀이 입력된 순서를 파악함으로써 사용자의 행위를 분석했다. 이러한 분석 방법을 활용하여 원본 파일에서 추가적으로 입력된 데이터가 무엇인지, 마지막으로 입력하거나 수정한 셀이 무엇인지 파악하고 문서의 조작여부를 판단하여 MS 엑셀 파일이 실제 사건에서도 정황 증거로 사용될 수 있다.

References

〔1〕 Office 365 Could Boost Microsoft's Market Share In Shift To Cloud, http://seekingalpha.com/article/1437661-office-365-could-boost-microsofts-market-share-in-shift-to-cloud

〔2〕 Microsoft Corporation, "Office Open XML File Formats," Standard ECMA-376, 4th Edition, Dec. 2012.

〔3〕 Zhangjie Fu, Xingming Sun, Yuling Liu and Bo Li, "Forensic investigation of OOXML format documents," Digital Investigation, vol 8, issue 1, pp.48-55, Jul. 2011.

〔4〕 JiHye Youn, JungHeum Park and Sangjin Lee, "Methods for Investigating of Edit History about MS PowerPoint Files That Using the OOXML Formats," Journal of The Korea Institute of information Security & Cryptology, C19(4), pp.215-224, Aug. 2012.

〔5〕 Bora Park, Jungheum Park and Sangjin Lee, "Data concealment and detection in Microsoft Office 2007 files," Digital Investigation, vol 5, issues 3-4, pp.104-114, Mar. 2009.

〔6〕 D. J. Power, A Brief History of Spreadsheets, http://dssresources.com/history/sshistory.html

〔7〕 Hyukdon Kwon, Yeog Kim, Sangjin Lee and Jongin Lim, "A Tool for the Detection of Hidden Data in Microsoft Compound Document File Format," International Conference on Information Science and Security, pp.141-146, Jan. 2008

〔8〕 Microsoft Corporation, "〔MS-XLS〕 : Excel Binary File Format (.xls) Structure," Feb. 2013

〔9〕 Microsoft Corporation, The Microsoft Office Visualization Tool (OffVis) Fact Sheet, http://www.microsoft.com/en-us/download/details.aspx?id=2096

〈저자소개〉

이 윤 미 (Yoonmi Lee) 학생회원
2011년 2월: 순천향대학교 정보보호학과 공학사
2013년 3월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 디지털 포렌식, 정보보호

정 현 지 (Hyunji Chung) 학생회원
2010년 2월: 고려대학교 컴퓨터공학, 산업시스템공학 공학사
2010년 3월~2012년 2월: 고려대학교 정보보호대학원 공학석사
2012년 3월~현재: 고려대학교 정보보호대학원 박사과정
〈관심분야〉 디지털 포렌식, 데이터 마이닝

이 상 진 (Sangjin Lee) 종신회원
1987년 2월: 고려대학교 수학과 학사
1989년 2월: 고려대학교 수학과 석사
1994년 8월: 고려대학교 수학과 박사
1989년 10월~1999년 2월: ETRI 선임 연구원
1999년 3월~2001년 8월: 고려대학교 자연과학대학 조교수
2001년 9월~현재: 고려대학교 정보보호대학원 교수
2008년 3월~현재: 고려대학교 디지털포렌식연구센터 센터장
〈관심분야〉 디지털 포렌식, 심층 암호, 해쉬 함수