

MS-Word 문서 접근 제어시스템 설계

장승주*

Design of the Access Control System for MS-WORD Document System

Jang, Seung-Ju*

Department of Computer Engineering, Dongeui University, Busan 47340, Korea

요 약

본 논문은 MS-워드 문서 시스템에 대한 접근 제어 시스템을 설계한다. 본 논문에서 설계하는 시스템은 MS-워드 문서 구조를 분석하여 문서 관련 정보를 활용한다. MS-워드문서 정보를 일부 변형하여 변형된 정보에 접근할 수 없는 사용자는 접근을 차단하도록 설계하는 것이다. 이렇게 함으로써 MS-워드문서에 대해서 접근 권한을 가진 사용자 외에는 문서를 읽을 수 없도록 한다. 즉, MS-워드문서에 대한 접근 권한을 통제할 수 있도록 한다. MS-워드문서에 대한 접근 권한을 가진 사용자는 변형된 정보를 원래 정보로 복구할 수 있도록 하여 정상적으로 문서를 읽을 수 있도록 한다. 본 논문에서 설계하는 내용을 실제 구현하고 실험을 수행하였다. 실험에서는 MS-워드문서 정보를 변형하였을 경우 문서 접근이 되는지를 수행하였다. 실험을 수행한 결과 MS-word 접근제어시스템이 정상적으로 잘 동작됨을 확인할 수 있었다.

ABSTRACT

This paper designs access control system for MS-word(Microsoft-word) document system. The system designed in this paper uses the document-related information by analyzing the MS-word document structure. It is designed to block access to users who can not access the modified information by partially modifying MS-word document information. This makes it impossible to read documents other than those who have access to the MS-word document. This allows you to control access to the MS-word document. A user with access to the MS-word document will be able to retrieve the modified information back to the original information so that the document can be read normally. In this paper, we design and implement experiments. In the experiment, we performed document access if MS-word document information was modified. Experimental results show that the MS-word access control system operates normally.

키워드 : MS-워드, 접근 제어, 접근 권한, 데이터 필드 변경, 실험

Key word : MS-word, Access Control, Access Permission, Modified Data Field, Experiment

Received 3 July 2018, Revised 19 July 2018, Accepted 27 September 2018

* Corresponding Author Jang, Seung-Ju(E-mail:sjjang@deu.ac.kr, Tel: +82-51-890-1710)

Department of Computer Engineering, Dongeui University, Busan 47340, Korea

Open Access <http://doi.org/10.6109/jkiice.2018.22.10.1405>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

인터넷의 급속한 발전으로 인하여 각종 인터넷 사고가 급증하면서 개인정보 및 중요자산에 대한 보안 대책이 시급한 실정이다. 최근에 컴퓨터 보안과 관련한 문제들이 끊임없이 발생되고 있다. 전자상거래 뿐만 아니라 인터넷 포털 사이트 홈페이지 이용 시 다양한 개인 정보 입력을 요구하고 있다. 이러한 개인 정보는 개인이 사용하는 PC환경을 중심으로 저장 및 관리되고 있다. 개인 PC에 대한 보안서비스는 약간의 차이가 있지만 열악한 수준을 벗어나지 못하고 있다. 그림 1은 인터넷을 이용할 경우에 발생할 수 있는 여러 가지 문제들을 보여주고 있다. 특히 보안에 문제가 발생할 소지가 있다.

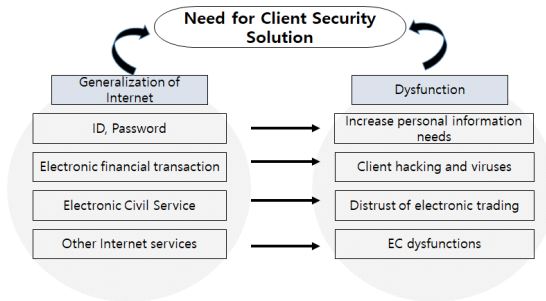


Fig. 1 Security Technology for Dysfunction of the Internet

따라서 중요 정보 유출에 따른 손실은 가능해지기 어려울 정도다. 정보기술의 순기능과 함께 정보기술의 발전은 업무 조직에게 다양한 혜택을 주고 있으나, 치명적인 역기능도 발생하고 있다.

또한 최근 국내 각종 문서 또는 서류의 유통 보관에 상당한 비용이 소요되고 있는 것으로 추정되고 있다. 전자 문서에 대한 보안 문제 해결을 위해서는 이보다 훨씬 더 많은 비용을 지불해야 한다. 개인이나 기업의 경우 중요한 전자 문서의 관리를 위해서 암호 복호화 기술을 활용하는 경우가 많다 [1, 2].

특히 우리가 매일 사용하는 문서 파일과 관련한 보안 사고의 빈도가 증가되고 있는 추세이다. 문서보안에 대한 국내시장은 DRM이 주도하고 있었다. 최근에는 DRM, DLP, 암호화, 문서 중앙화등 수많은 솔루션들이 소개되고 있다.

또 다른 보안 방법으로 디스크 암호화 혹은 파일 암호 등과 같은 다양한 기능을 사용할 수 있다. 디스크 암호

화는 암호로 전체 디스크 또는 부분 파일을 보호 할 수 있다. 파일의 이름을 숨길 수도 있다. 디스크(드라이브) 암호화는 보안을 위해 HDD, SSD, USB메모리, SD카드 등의 저장장치를 암호화하거나 파일 시스템을 암호화 하는 것이다 [1, 2]. 문서 중앙화의 경우는 조직 내에서 생산되는 모든 문서를 중앙 서버에서 관리하는 방식이다. 문서를 관리 및 통제하기에는 좋은 방법이라고 할 수 있다.

본 논문은 전 세계적으로 가장 많이 사용하는 MS-워드 문서 사용시 접근 권한 통제를 통한 권한 관리 시스템을 설계한다. 2장은 관련연구, 3장은 MS-워드 문서 접근 제어 시스템 설계, 4장은 설계된 시스템에 대한 실험 및 결과, 5장 결론의 순으로 설명한다.

II. 관련 연구

Microsoft(이하 MS) 워드는 문서 작성 및 편집 등에 있어서 전 세계적으로 사용되는 대표적인 문서 편집 프로그램이다. MS-워드는 PC나 모바일 단말기 등 다양한 기기에서 사용할 수 있도록 기능을 지원하고, 다른 종류의 문서와 호환성도 뛰어나 전 세계적으로 문서 관련 소프트웨어 시장 점유율이 높다.

MS-워드 파일의 구조는 OLE(Object Linking and Embedded) 포맷 구조로 되어 있다. OLE는 Microsoft에서 개발한 독점 기술로, 문서 및 기타 개체에 포함 및 연결될 수 있다. 개발자에게는 사용자 정의 인터페이스 요소를 개발하고 사용하는 방법인 OCX(OLE Control Extension)가 도입되었다.

OLE 파일 포맷은 내부가 하나의 작은 하드 디스크와 같은 파일 시스템 구조를 가진다. OLE 파일 내부에는 폴더 및 파일의 개념도 있다. OLE 파일 내부가 하나의 작은 파일 시스템이므로 MS 오피스는 하위 호환성 및 상위 호환성도 뛰어나다.

MS 오피스 프로그램은 복합 문서 파일 형식(CDFS, Compound Document File Format)을 사용하는 97-2003 버전, OOXML 파일 형식(오피스 Open XML File Format)을 사용하는 2007-현재 버전까지 두 가지 파일 형식으로 구성된다.

OOXML 파일 형식은 MS 오피스 2007 버전부터 현재까지 사용하고 있는 문서 파일 형식으로 .xlsx, .pptx,

.docx 등의 확장자로 저장된다. OOXML 파일은 여러 개의 파트(part)들과 파트간의 관계를 나타내는 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 [2, 4]. OpenXML은 처음부터 Microsoft Corporation에서 정의한 이진 형식으로 인코딩된 기존의 워드 프로세서 문서, 프레젠테이션 및 스프레드시트 자료를 정확하게 표현할 수 있도록 하기 위해 설계되었다. 이러한 파일의 이진 형식은 공간을 중시하고 구문 분석 시간이 사용자 경험에 상당한 영향을 미치던 시기에 만들어졌다. 각 문서 유형은 WordprocessingML, PresentationML 또는 SpreadsheetML과 같은 기본 표시 언어를 통해 지정된다 [5, 6].

OPC(Open Packaging Convention)는 문서를 완전히 표시하기 위해 여러 유형의 콘텐츠를 ZIP 보관 파일과 같은 컨테이너에 저장하는 방식을 제공한다. OPC의 권장 구현 방법은 ZIP 보관 파일 형식을 사용하는 것이다.

본 논문과 관련한 문서 접근 제어 방식은 대부분의 경우 압축화 방식을 사용하고 있다. 이러한 방식은 구현하기가 간단한 반면에 암호, 복호화하는데 시간이 많이 소요되는 단점이 있다.

III. MS-워드 문서 접근 제어 시스템 설계

본 논문에서는 MS-워드문서에 대한 접근 제어 시스템을 설계한다. 우선 접근 제어 시스템을 설계하기 위하여 MS-워드의 문서 구조를 분석한다. MS-워드의 문서 구조는 다음 그림 2와 같다 [7].

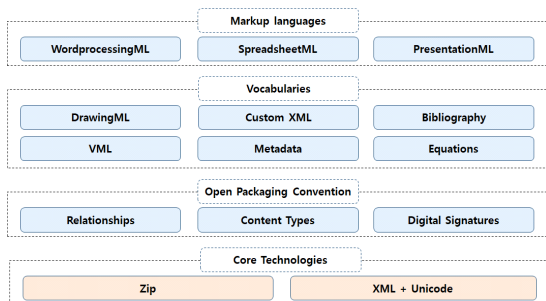


Fig. 2 Structure of MS-word Document

MS-word 문서는 XML 기반 형식으로 표준화되어 있으며 많은 소프트웨어 플랫폼 및 운영 체제에서 솔루션을 사용할 수 있는 개방형 구조로 되어 있다. 문서에는 WordprocessingML, 스프레드시트에는 SpreadsheetML, 프레젠테이션에는 PresentationML이 있다. Open XML 문서는 많은 부분을 포함하는 컨테이너로 저장된다. 현재 컨테이너는 ZIP 파일 형식이며 ZIP 내의 파트로 볼 수 있다.

Microsoft 오피스 2007버전이 등장하면서 파일 포맷이 바뀌었는데, Microsoft 오피스의 2003 이하 버전 파일들은 복합문서 파일 형태로 저장하였으나 2007 이상의 버전에서는 오피스 Open XML 파일 형태로 저장하기 시작하였다. 기업 및 개인은 문서 작성 시 Microsoft 오피스를 많이 사용하므로 OOXML형태의 파일 또한 비중이 크게 늘어났다

새롭게 바뀐 OOXML파일 포맷에 대해서 데이터가 숨겨져 있다. 각각의 OOXML파일은 패키지라 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 파일 형식으로 되어 있다. 패키지-관계 ZIP 항목에 파트와 패키지 사이의 관계 정보가 저장되며, 파트-관계 ZIP 항목에는 여러 파트 사이의 관계 정보가 저장된다 [8, 9].

새로운 OOXML 파일은 이전 파일 유형과 구별하기 위해서 새로운 파일 확장자로 변경하였다. 변경되기 전 MS-워드의 확장자는 .doc 였다. 하지만 새로운 MS-워드 파일 유형의 확장자는 기존 확장자 명에 'x'를 붙여 .docx로 표현한다.

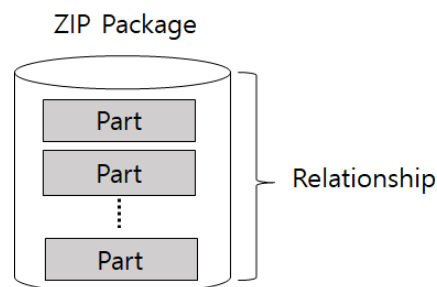


Fig. 3 Internal Structure of ZIP File

그림 3은 MS-워드 파일의 ZIP 구조내의 실제적인 모습이다. ZIP 파일 형식의 MS-워드 문서는 패키지라 불리는 컨테이너 안에 압축된 파트(part)들과 그와 관련된 것들이 저장되어 있다.

MS-워드 파일 구조를 살펴볼 수 있는 도구를 이용할 수 있다. Structured Storage View(이하 SSView)과 같은 도구를 이용할 수 있다. 그림 4는 MS-워드 파일에 대해서 SSView를 실행하여 구조를 본 화면이다. MS-워드 파일을 SSView를 통해서 열기하면 해당 문서의 내부 구조와 정보를 알 수 있다. 기본적으로 Microsoft 오피스 문서에 대한 뷰 기능을 지원한다.

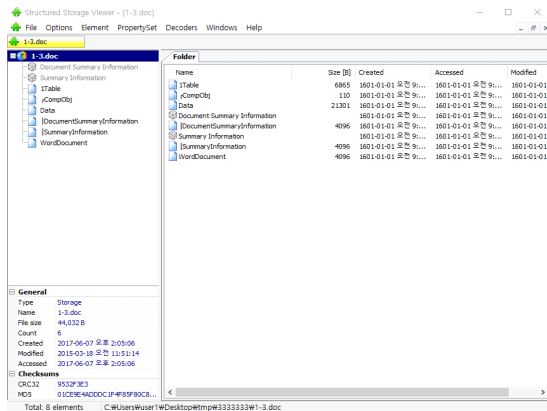


Fig. 4 Internal Structure of MS-word File

그림 4에서 PropertySet에는 ID, Name, Type, Value 정보가 있다. 여기에 있는 정보를 삭제, 추가, 수정 등의 편집이 가능하다. Decoders 메뉴에는 ZLib codec을 선택할지 안할지를 설정하는 기능이 있다. ZLib codec를 선택하고 싶으면 별다른 설정 없이 ZLib codec 메뉴를 클릭하면 된다.

3.1. 접근 제어 시스템

본 논문은 MS-워드 파일에 대해서 접근을 제어하기 위하여 특정 필드 정보를 변형하게 된다. 본 논문에서 제안하는 MS-워드 파일에 대한 접근 제어 시스템 구조는 다음그림 5와 같다.

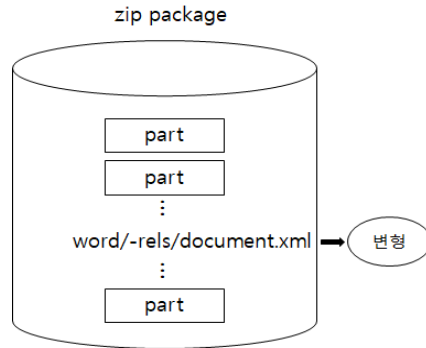


Fig. 5 Structure for Proposed System

기존의 문서 접근 방식이 암호화, 복호화를 통해서 하고 있다. 하지만 본 논문은 특정 필드 변형을 통해서 수행함으로써 비용적인 측면, 성능적인 측면에서 유리하다고 할 수 있다. 본 논문에서 제안하는 방식이 이러한 장점을 가지고 있는 반면에 변형된 필드 정보의 유출로 인한 단점이 발생할 수 있다 [10, 11].

그림 5와 같이 특정 필드 정보를 변형하기 전에 MS-워드 파일의 구조를 분석하게 된다. 분석된 정보를 기초로 해서 MS-워드의 특정 필드를 변형한다. 특정 필드를 변형하기 위해서 MS-워드 파일이 가지고 있는 구조에서 압축된 파일이 압축 해제가 되지 않도록 필드를 변형하게 된다. 특정 필드를 변형할 경우에 압축해제가 되지 않도록 하는 것이 목표이다.

그리고, 접근 권한을 가진 사용자는 변형된 정보를 환원할 수 있도록 하는 프로그램을 통해서 쉽게 사용이 가능하도록 한다. 이러한 목표를 만족시키기 위하여 변경할 특정 필드를 찾게 된다. 변경할 특정 필드는 가능한 파일의 앞부분을 선택할 수 있도록 한다. 파일의 앞부분에 파일과 관련한 정보들이 있기 때문이다.

MS-워드 파일, 즉, docx 파일은 zip 파일 형식으로 되어 있다. 이 zip 파일은 여러 개의 “parts”로 구성되어 있다. 이 zip 파일은 이미지나 비디오와 같은 미디어 파일도 포함하고 있다.

OOXML 파일 패키지 구조는 파일들 간의 관계를 _rels 폴더 내의 파일에 정의하도록 되어 있다. 예를 들면 그림 6과 같은 경우는 파일 내용이 어느 파일에 저장되어 있는지를 나타낸다.

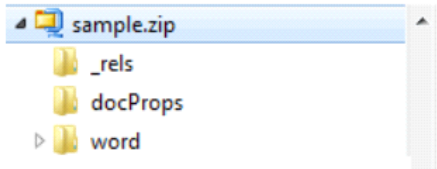


Fig. 6 Structure of MS-word Compressed File

각 파일들의 관계는 각 폴더의 하위 폴더 _rels에 정의되어 있다. 그림 7은 _rels에 정의되어 있는 모양을 보여준다.

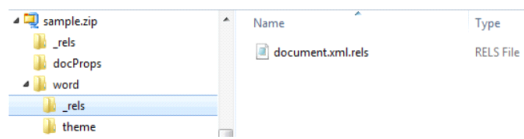


Fig. 7 Structure of Folder in MS-word Compressed File

MS-워드 파일의 구조에서 압축 해제하여 document.xml을 추출한다. document.xml 파일의 특정 필드 값의 변경을 통한 MS-워드 파일 접근 제어를 하도록 설계한다.

MS-워드 압축 파일에서 압축해제 한 파일 중에 document.xml 파일을 이용한다. 이 파일에서 0xA15 주소 위치 값을 특정값 '00' '00'으로 변경한다. 이 값의 강제 변경을 위해서 프로그램을 이용하여 수행한다.

이렇게 특정 필드 값의 변경후 MS-워드 파일 열기를 하게 되면 정상적인 열기 수행이 되지 않는다. 이렇게 함으로서 지정된 사용자 이외에 해당 파일에 대한 접근을 원천적으로 차단하도록 한다. 이 파일에 대한 접근을 하기 위해서는 변경한 필드 내의 값을 원래 값으로 환원시켜줘야 한다. 특정 필드 값을 원래 값으로 환원 시켜주게 되면 MS-워드 파일을 정상적으로 사용이 가능하다.

IV. 실험

본 논문에서 제안하는 MS-워드 문서 접근 제어 시스템 설계 내용에 대한 실험을 수행하였다. 본 논문에서 제안하는 시스템에 대한 실험은 윈도우 10 운영체제 환경에서 수행되었다. MS-워드는 2013 버전을 가지고 수행되었다. 프로그램 환경은 비주얼 스튜디오 2012 버전에서 이루어졌다.

먼저 MS-워드 파일을 압축 해제 시킨다. MS-워드 파일을 압축 해제 시키게 되면 앞에서 그림 7과 같은 형식으로 해제가 되게 된다. 이 압축 파일 내에서 document.xml 파일의 내용을 확인해보면 다음 그림 8과 같다.

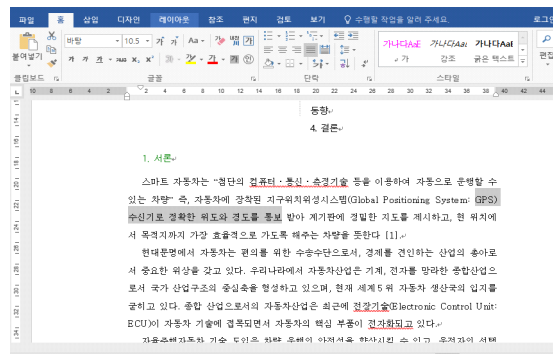


Fig. 8 Original File of MS-word

위 그림 8과 같이 document.xml 파일을 통해서 원본 파일 내용을 확인할 수 있다. MS-워드 파일에 대한 접근을 제어하기 위하여 많은 방법들을 이용한다. 일반적으로 많이 사용하는 것이 암호화 방법이다. 본 논문에서는 암호화 방식보다 효율적인 필드 정보 관리를 통한 접근 제어를 수행한다. 아래 그림 9는 MS-워드 원본 파일에 대한 폴더 내용이다.

필드 정보 관리는 MS-워드 파일의 특정 필드를 특정한 값으로 변형함으로써 파일을 읽을 수 없도록 한다. 이 방법은 간단하게 프로그램을 이용하여 수행이 가능하다. 프로그램으로 특정 필드 값을 변형하고, MS-워드 파일을 정상적으로 읽으려고 하면, 변형된 필드 값을 정상적으로 복원해주면 된다.

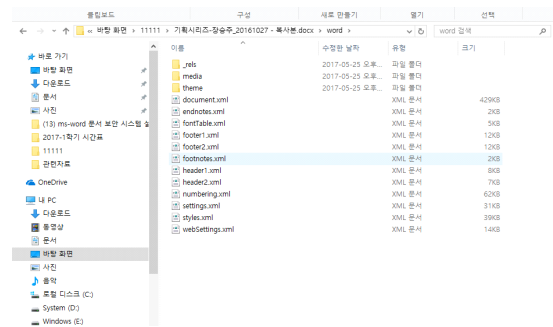


Fig. 9 Folder Shape of MS-word File after Releasing Compression



Fig. 10 Content of document.xml File

그림 10은 document.xml파일의 실제적인 내용을 보여준다. MS-워드 문서에서 0xA15 위치 값을 '00 00 00'으로 변경한다. 이 필드 값을 '00 00 00'로 변경할 경우 MS-워드 문서 읽기가 가능하지 등에 대한 실험을 수행하였다. 다음은 접근 제어 기능을 설정한 후의 실제 MS-워드 파일에 대한 압축 해제 후 xml 파일을 읽은 화면은 그림 11과 같다.

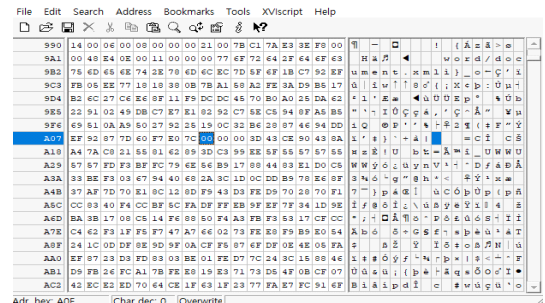


Fig. 11 Change of 0xA15 Position Value

그림 11은 0xA15 위치 값을 00 00 00으로 변경하는 과정이다. 이 값을 변경하기 위해서 바이너리 편집기를 사용하였다. 이 값을 변경하고 난후 MS-워드 파일을 열기했을 경우 결과 화면은 그림 12와 같다.

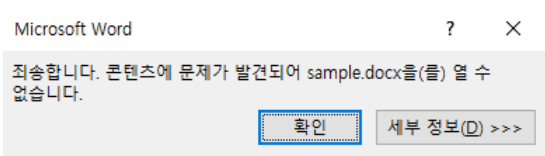


Fig. 12 Reading MS-word File after Changing 0xA15 Position Value

그림 12에서 보는바와 같이 0xA15 위치 값의 변경으로 정상적인 MS-워드 파일 읽기가 되지 않는 것을 확인할 수 있다.

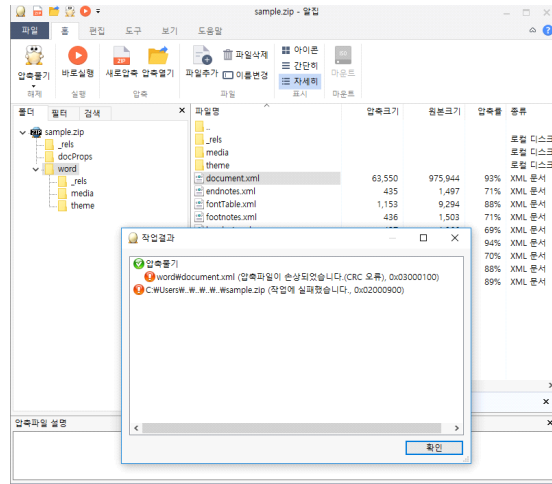


Fig. 13 Decompression Process of MS-word File

그림 13은 접근 제어 설정 후 MS-워드 압축 해제하고 나서 document.xml 파일이 열리지 않음을 확인할 수 있다. 이와는 반대로 정상적인 사용자가 MS-워드파일을 사용하고자 하는 경우 MS-워드 파일에서 특정 필드 값을 정상적으로 복원해 주게 된다. 변경된 특정 필드 값을 복원시켜 주게 되면 정상적인 파일 읽기가 가능함을 확인할 수 있다.

V. 결론

본 논문은 세계적으로 가장 많이 사용하는 MS-워드 문서에 대한 접근 제어 시스템을 설계한다. 본 논문에서 설계하는 MS-워드 파일에 대한 접근 권한 제어 시스템은 특정 필드 정보의 은닉을 통한 접근 제어를 수행하도록 한다. 이렇게 함으로써 기존의 문서 보안 시스템보다 간략하면서 보다 안정적인 보안을 할 수 있도록 한다.

본 논문에서 제안하는 MS-워드 문서 보안 설계 내용은 특정한 필드 정보를 은닉하여 문서 읽기를 할 수 없도록 한다. 또한, 정상적으로 접근 가능한 사용자는 은닉된 특정 필드 정보를 복구하여 MS-워드 문서를 읽을

수 있도록 한다. 이렇게 함으로써 암호화 등의 방법을 사용할 경우에 발생하는 부담을 줄일 수 있다.

본 논문에서 제안하는 MS-워드 문서 보안 시스템은 특정 필드값을 변경한다. 이렇게 변경할 경우에 MS-워드 문서를 열어서 읽어보지 못하도록 한다. 물론 정상적인 사용자를 위해서 변경된 필드 정보를 원래 값으로 복원시킬 수 있도록 함으로써 정상적으로 MS-워드 파일을 읽을 수 있도록 한다.

본 논문에서 제안하는 MS-워드 문서 접근 제어 시스템 설계 내용을 중심으로 실험을 수행하였다. 실험에서 MS-워드의 특정 필드 정보를 은닉한 결과 MS-워드 문서를 읽을 수 없음을 확인할 수 있었다. 특정 필드 변경으로 인한 단점은 추후 보안을 위한 연구를 진행할 예정이다.

References

- [1] J. H. Youn, "Methods for investigating of edit history about MS PowerPoint files that using the OOXML formats," M.S thesis, Department of Information Management, Korea University, 2013.
- [2] J. H. Youn, J. H. Park, and S. J. Lee, "Methods for Investigating of Edit History about MS," *Journal C of Korea Information Processing System*, vol. 19, no. C, pp. 215-224, Aug. 2012.
- [3] W. Vugt, "Open XML The markup explained," Microsoft, Oct. 2007.
- [4] Y. M. Lee, H. J. Chung, and S. J. Lee, "A Study on Edit Order of Text Cells on the MS Excel Files," *Journal of The Korea Institute of Information Security & Cryptology*, vol. 24, no. 2, pp. 319-325, Apr. 2014.
- [5] H. B. Chang, S. S. Yeo, G. C. Park, and C. H. Lee, "The Study on Development of Document Security Components," *Journal of Security Engineering*, vol. 5. no 2, pp. 123-132, Apr. 2008.
- [6] Digital Forensics Wikipedia [Internet]. Available: http://forensic.korea.ac.kr/DFWIKI/index.php/%EB%AC%B8%EC%84%9C%ED%8C%8C%EC%9D%BC_%EB%82%B4%EC%9A%A9_%EB%B0%8F_%EB%82%B4%EB%B6%80_%EA%B5%AC%EC%A1%B0_%ED%99%95%EC%9D%B8_%EC%9C%A0%ED%8B%B8%EB%A6%AC%ED%8B%B0.
- [7] Object Linking Embedding [Internet]. Available: https://en.wikipedia.org/wiki/Object_Linking_and_Embedding.
- [8] Y. M. Lee, "A Study for Forensic Methods of MS Excel Files," M.S thesis, Department of Information Management, Korea University, 2014.
- [9] A. Ziad, A. H. Bagci, T. Zaitoun, and A. Asad, "Towards the Memory Forensics of MS Word Documents," *Advances in Intelligent Systems and Computing*, vol. 558, pp 179-185, July 2017.
- [10] S. J. Jang, "Implementation of the Access Control System for Hangul Document System," *Journal of the Korea Institute of Information and Communication Engineering*, vol.22, no.2, pp. 323-329, Feb. 2018.
- [11] S. Park, and K. Jang, "Cases Analysis and Implications of Personal Information Leaking in focusing on precedent cases," *Journal of Security Engineering*, vol.13, no.3, pp. 249-260, Jun. 2016.



장승주(Jang, Seung Ju)

1985년 부산대학교 계산통계학(전산학) 학사
 1991년 부산대학교 계산통계학(전산학) 석사
 1996년 부산대학교 컴퓨터공학 박사
 1987년~1996년 한국전자통신연구원(ETRI) 시스템 SW연구실
 1993년~1996년 부산대학교 시간강사
 2001년~2002년 Univ. of Missouri at Kansas City, visiting professor
 1996년 ~ 현재 동의대학교 컴퓨터공학과 교수
 ※관심분야 : 운영체제, 임베디드 운영체제, 분산시스템, 시스템 보안, 스마트 폰 시스템 운영체제, 자율주행차 운영체제