*Journal of The Korea Institute of Information Security & Cryptology* ISSN 1598-3986(Print)
VOL.25, NO.6, Dec. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.6.1485

# 복합문서 파일에 은닉된 데이터 탐지 기법에 대한 연구*

김 은 광,† 전 상 준, 한 재 혁, 이 민 욱, 이 상 진‡
고려대학교 정보보호대학원

# An effective detection method for hiding data in compound-document files*

EunKwang Kim,† SangJun Jeon, JaeHyeok Han, MinWook Lee, Sangjin Lee‡
Graduate School of Information Security, Korea University

요 약

기존 데이터 은닉은 대용량 멀티미디어 파일에 데이터를 삽입하는 방식으로 이루어졌다. 하지만 최근 Microsoft Office 2003 이하 버전 제품의 문서파일은 구조가 파일시스템과 유사하여 데이터 은닉이 비교적 용이해 커버데이터(Cover data)로 사용되고 있다. 데이터가 은닉된 문서파일을 MS Office 프로그램으로 실행할 경우 은닉 사실을 모르는 사용자는 은닉 데이터를 눈으로 쉽게 확인할 수 없다. 이에 본 논문에서는 Microsoft Office 2003 이하 버전과 한컴오피스 문서파일에서 사용되는 복합문서 파일 이진형식(Compound File Binary Format) 파일 포맷 구조를 분석하여 데이터 삽입이 가능한 공간을 살펴보고 이를 탐지하기 위한 방안을 제시하고자 한다.

ABSTRACT

Traditionally, data hiding has been done mainly in such a way that insert the data into the large-capacity multimedia files. However, the document files of the previous versions of Microsoft Office 2003 have been used as cover files as their structure are so similar to a File System that it is easy to hide data in them. If you open a compound-document file which has a secret message hidden in it with MS Office application, it is hard for users who don't know whether a secret message is hidden in the compound-document file to detect the secret message. This paper presents an analysis of Compound-File Binary Format features exploited in order to hide data and algorithms to detect the data hidden with these exploits. Studying methods used to hide data in unused area, unallocated area, reserved area and inserted streams led us to develop an algorithm to aid in the detection and examination of hidden data.

**Keywords:** Steganography, Data hiding, Compound File Binary Format, Microsoft Office

## I. 서 론

스테가노그래피는 전달하고자 하는 데이터를 은밀하게 삽입하여 원하는 대상에게 전달하기 위해 개발된 기술이다. 이는 군사 작전과 같은 기밀 통신이 필요한 상황에서부터 사회의 조직적인 범죄에까지 두루 사용되고 있다. 스테가노그래피가 조직범죄에 이용된 경우 숨겨진 데이터를 정확하게 찾아 분석에 활용할 필요가 있다.

지금까지의 스테가노그래피에서 정보은닉은 오디오나 이미지, 비디오 파일과 같은 멀티미디어 파일에 데이터를 삽입하는 방식으로 이루어졌다. 하지만 최근에는 문서파일 구조와 문서작성 응용프로그램의 기능을 이용한 데이터 은닉사례가 늘고 있다. 대표적으로 Microsoft Office 제품군은 전 세계적으로 널리 쓰이고 있는 문서작성 프로그램 중 하나로 문서파

Received(09. 25. 2015), Modified(10. 27. 2015), Accepted(10. 28. 2015)
* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단-공공복지안전사업의 지원을 받아 수행된 연구임(2012M3A2A1051106)
† 주저자, ekkim5236@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr(Corresponding author)

일구조의 특징 상 데이터 은닉이 용이하다[1].

Microsoft Office 2003 이전 버전은 복합문서 파일 (Compound-Document File) 형식을 사용하며, 2007 이후 버전은 Open Office XML(OOXML)의 구조를 사용한다. [2]에 따르면 전 세계 4000만의 MS Office 사용자들은 다양한 MS Office 버전을 사용하여 매년 400억 개 이상의 MS 문서파일을 생성하는 것으로 나타났다. 한컴오피스의 문서파일 역시 복합문서 파일 형식을 사용하기 때문에 복합문서 파일의 사용량은 적지 않다.

Microsoft Office 2003 이하 제품군에서 사용하는 파일구조와 한컴오피스에서 사용되는 복합문서 파일의 큰 구조는 동일하다. 하지만 파일을 구성하고 있는 요소들의 명칭과 각 요소에 저장되는 내용이 조금씩 상이하다. 두 형식의 파일 포맷 구조를 분석한 문서는 [3]와 [4]에서 각각 제공하고 있다.

디지털 포렌식 관점에서 데이터 은닉 여부 판단과 함께 은닉된 데이터를 확인하는 것이 매우 중요하다. 특히 문서파일에 비정상적으로 입력된 은닉 데이터는 일반인들이 쉽게 확인할 수 없다. 그리고 특정 메시지가 은닉된 문서파일이 네트워크 환경에서 전송될 경우 웹 방화벽이나 IDS, IPS에서도 탐지가 어렵다. 이와 같은 특징을 이용하여 국내외 범죄 조직들도 조직원들과 비밀 메시지를 이미지파일이나 문서파일에 담아 전달하는 스테가노그래피 기법을 이용해 사이버 교신을 했었다[5].

본 논문에서는 구조에 대해 설명하고 데이터 은닉이 가능한 방법을 조사하여 은닉된 데이터를 효과적으로 탐지하기 위한 방안을 제시한다.

## II. 관련 연구

예전에는 용량이 큰 멀티미디어 파일에 주로 데이터 은닉이 이루어졌다. 하지만 최근에는 문서파일에도 데이터 은닉이 이루어지고 있으며, 이에 따라 복합문서 파일에 데이터를 은닉하기 위한 연구와 함께 은닉된 데이터를 탐지하기 위한 기술은 계속 연구가 진행되고 있다.

MS Office 제품군은 2003 이하 버전에서 복합문서 파일형식을 사용한다[3]. 2008년 1월까지 MS Office 관련 프로그램이 복합문서 파일 형식으로 저장하는 특성을 이용하거나, 복합문서 파일의 구조를 이용하여 데이터를 은닉하는 방법들이 제시되고 있다. 우선 2003이하의 MS Office 파워포인트 응용 프로그램에서 복합문서 파일로 저장할 때 '빠르게 저장하기' 기능을 제공하는데 이 기능으로 생길 수 있는 잉여정보를 이용하여 데이터를 은닉할 수 있음이 제시되었다[6]. '빠르게 저장하기' 옵션이 선택되어 있으면 파일을 저장할 때, 삭제되거나 수정되기 이전의 슬라이드를 삭제하는 것이 아니라 해당 스트림의 오프셋만 변경하는 식으로 파일을 저장한다. 따라서 파워포인트 응용 프로그램으로 확인할 수 없는 영역이 파일 내에 존재한다.

또 다른 은닉 방법으로는, 복합문서 파일이 파일 시스템의 구조와 유사성을 가지는 특징을 이용하여 파일 시스템처럼 복합문서 파일 역시 슬랙(Slack) 영역을 가지며 이러한 슬랙 영역을 이용하여 데이터를 은닉할 수 있는 방법이 제시되었다[7]. 또한 MS Office 응용 프로그램에서 인식할 수 없는 스트림을 생성하여 데이터를 은닉하는 것이 가능함을 보인 논문도 발표되었다[8]. 한편, MS Office 2007 응용프로그램은 2003 버전에서 사용했던 복합문서 파일이 아닌 Open XML 형식을 이용하여 데이터를 저장한다. MS Office 응용 프로그램에서는 복합문서 파일들에 대한 호환성을 유지하고 있지만, 문서 형식이 완전히 달라졌기 때문에 복합문서 파일에 데이터를 은닉하는 방법들은 Open XML 형식을 사용하는 파일에는 사용할 수 없다.

속성정보가 저장되는 스트림 내 은닉된 데이터를 탐지하고 삭제해주는 도구도 존재하지만(Payne Group(2005)[9], Soft Experience (2005)[10], Workshare(2005)[11] 등) 도구에서 적용된 방법으로 은닉된 데이터만 삭제할 수 있으며 새로운 기법으로 은닉된 데이터는 탐지할 수 없다.

문서파일 안에 텍스트 데이터를 효과적으로 은닉하기 위한 몇 가지 연구가 진행되었다. 텍스트 기반의 데이터 은닉 방법은 크게 어의문자를 이용한 방법과 문서 내 공백공간을 이용한 방법으로 나뉜다[12]. 어의문자(語義文字, semagram)란 그림이나 상징으로 표현되는 문자로 어의문자를 이용한 방법은 문서 내 여분의 공백 및 탭을 추가하여 문서 내 텍스트의 위치를 변경해 비밀 메시지를 표시한다. 문서 내 공백공간을 이용한 방법은 문서 내 존재하는 줄 간격의 끝 공백 공간과 문자 간 공백을 조작하여 인코딩된 은닉 데이터를 삽입하는 방법이다[13]. 하지만 위 2가지 방법은 문서 파일 내에 은닉할 수 있는 데이터 크기가 작다는 단점이 있다.

Abdul Monem S. Rahma는 복합문서 파일의 미사용 영역 내에 데이터를 은닉하는 방법을 제시하였다. 데이터가 은닉된 복합문서 파일은 문서응용 프로그램에서 실행했을 때 확인할 수 없고 데이터가 은닉된 복합문

서 파일을 복사하거나 전자 우편으로 전송해도 은닉 데이터가 여전히 존재한다[14].

Wesam Bhaya는 워드문서파일에서 사용하는 글꼴 중 육안으로 봤을 때 유사해 보이는 글꼴 15쌍을 선정하여, 유사한 글꼴이지만 실제 문서를 작성했을 때 발생하는 글꼴의 높이 차를 이용한 데이터 은닉 방법을 제시하였다[15].

## III. 복합문서 파일 포맷

복합문서 파일은 하나의 통합된 지각 환경을 이루는 사용자 인터페이스의 조직화된 모음으로 텍스트, 오디오, 동영상 등 서로 다른 데이터 형식을 포함할 수 있는 구조이다. 복합문서 파일 형식을 사용하는 파일들의 구조를 살펴보면 Table 1.과 같다.

MS PowerPoint나 MS Excel을 구동시키지 않고 삽입된 문서를 편집할 수 있게 해준다. 이러한 특성을 OLE(Object Linking Embedding)라고 하며, 복합문서 파일을 OLE 복합문서라고 부르기도 한다. 복합문서 파일은 운영체제에서 사용하는 파일 시스템(예:FAT 등)과 유사하다.

복합문서 파일은 스토리지(storage)와 스트림(stream)의 계층 구조로 구성되며, 이들을 관리하기 위한 메타데이터가 존재한다. 이 포맷은 여러 응용프로그램에서 생성된 데이터를 한 문서 파일에서 편집할 수 있는 환경을 제공한다. 메타데이터란 사용자가 입력한 데이터를 설명하거나 관리하기 위한 데이터로 응용 프로그램에서 자동적으로 생성된다.

복합문서 파일의 최상위 저장소는 루트이며, 루트는 스트림 또는 스토리지를 가질 수 있다. 스트림은 파일시스템에서의 파일에 해당하며, 스토리지는 파일 시스템에서의 폴더와 유사하다. 즉, 스토리지는 하위

Table 1. Stream and Storage of Compound-Document files (: Stream, : Storage)

| MS Word | MS Excel | MS PowerPoint | Hancom Office |
|---|---|---|---|
| WordDocument<br>1Table/0Table<br>Data<br>\0x005SummaryInformation<br>\0x005DocumentSummaryInformation<br>\0x001CompObj<br>Macros Storage<br>VBA<br>_VBA PROJECT<br>dir<br>Module <Name><br>_SRP_ <Number><br>PROJECTlk<br>PROJECTwm<br>PROJECT<br>ObjectPool<br>\0x003ObjInfo<br>\0x003Print<br>\0x003EPRINT<br>MsoDataStore<br>_Encryption <N><br>Item<br>Properties<br>xmlsignatures<br>_signatures<br>encryption | 0x005SummaryInformation<br>0x005DocumentSummaryInformation<br>LNK + 8 Hex<br>MBD + 8 Hex<br>0x0001Compobj<br>VBA<br>VBA Storage<br>_VBA PROJECT<br>dir<br>Module<N><br>_SRP_<N><br>PROJECTlk<br>PROJECTwm<br>PROJECT<br>Workbook<br>_SX_DB_CUR<br>000<N><br>XML<br>_xmlsignatures<br>5 Decimal<br>Revision Log<br>User Names<br>Ctls | Picture<br>Current User<br>0x005SummaryInformation<br>PowerPoint Document<br>0x005DocumentSummaryInformation<br>VBA<br>VBA Storage<br>_VBA PROJECT<br>dir<br>Module<N><br>_SRP_<N><br>PROJECTlk<br>PROJECTwm<br>PROJECT<br>Encrypted Summary Information<br>Digital Signature<br>Custom XMLData<br>Signatures | FileHeader<br>PrvText<br>PrvImage<br>DocOptions<br>_LinkDoc<br>Script<br>DefaultJScript<br>JScriptVersion<br>0x0005HwpSummaryInformation<br>DocInfo<br>BodyText<br>Section0 |

에 또 다른 스토리지 또는 스트림을 가질 수 있다.

## IV. 복합문서 파일을 대상으로 한 데이터 은닉 방법

복합문서 파일에 데이터를 은닉할 수 있는 방법은 크게 4가지로 분류할 수 있다. 4가지 방법 중 하나로 본 논문에서 자세히 다루지 않지만 다음과 같은 방법으로 응용프로그램 조작을 통해 복합문서 파일 내 데이터를 은닉할 수 있다.

- 문서 속성정보에 데이터 은닉
- 매우 작은 글자 크기로 작성하기
- 배경색과 같은 글자색으로 작성하기
- 이미지 뒤에 텍스트 숨기기
- 문서에서 보이는 URL과 다른 링크로 하이퍼링크를 설정하기

### 4.1 복합문서 파일의 미사용 영역에 데이터 은닉방법

복합문서 파일은 일반 파일 시스템과 매우 유사하기 때문에 파일 시스템에서의 특성이 그대로 존재한다. 이러한 구조적 특징을 이용한 데이터 은닉 방법을 크게 3가지로 분류할 수 있다.

#### 4.1.1 미할당 영역

복합문서 파일에서는 섹터(또는 소형 섹터)단위로 데이터를 저장하기 때문에 일반 파일 시스템처럼 미할당 영역이 존재할 수 있다. 미할당 영역은 복합문서 파일의 파일 할당 표에서 값이 0xFFFFFFFF(-1)인 인덱스에 해당하는 부분인 영역이다.

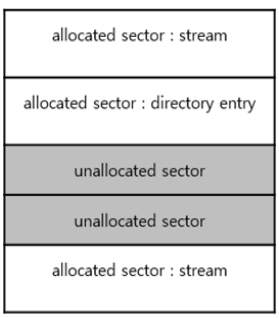


Fig. 1. Unallocated area

#### 4.1.2 슬랙 영역

할당된 섹터 내에서도 슬랙 영역이 존재할 수 있는데 디렉터리 엔트리, 스트림의 마지막 부분이 슬랙 영역으로 남을 수 있다. Fig. 2.의 디렉터리 엔트리와 스트림 영역에서 존재할 수 있는 슬랙 영역을 보여준다. 예를 들어, Fig. 2.에서 첫 번째 스트림의 크기는 2,762 바이트이다. 복합문서 파일은 섹터 단위(512 바이트)로 데이터가 저장되기 때문에 2,762 바이트의 데이터를 저장하기 위해서는 6개의 섹터(512*6=3072바이트)가 할당된다. 이 때 데이터는 2762 바이트까지 저장되므로, 3072 - 2762=310 바이트의 슬랙 영역이 생기게 된다. 310 바이트 크기의 슬랙 영역에 데이터를 은닉할 수 있다.

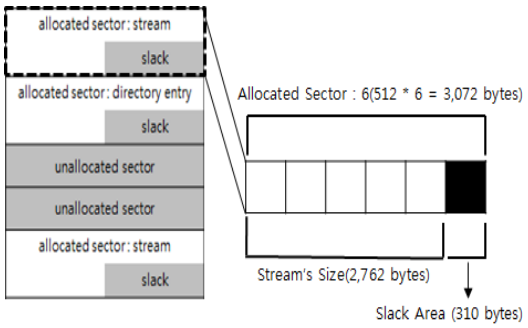


Fig. 2. Slack area and Unallocated sector

#### 4.1.3 예약 영역

복합문서 파일 내부에 정상적으로 존재하는 스트림 내부의 미사용 영역이 존재할 수 있다. Fig. 3.은 정상 스트림 내에 존재하는 미사용 영역을 보여준다. 정상스트림 내 미사용 영역에 특정 데이터를 은닉하게 될 경우 응용프로그램 상에서 은닉된 데이터를 확인할 수 없다. 위에서 살펴본 미사용 영역과 할당된 디렉터리 엔트리 및 스트림에서의 슬랙 영역, 정상스트림 내 미사용 영역은 문서 파일 내에서 사용되지 않는 영역으로 이 위치 역시 은닉데이터가 존재하더라도 응용프로그램에서 인식하지 않는다. 바로 이러한 사용되지 않는 영역에 데이터를 은닉시킬 수 있으며 복합문서 파일 형식에 대한 이해 없이는 이러한 데이터를 탐지하기 어렵다. 예를 들어, 한컴오피스 한글파일에 FileHeader 스트림이 있는데 이 스트림은 한글 파일을 인식하기 위한 정보를 저장하고 있다. 이 스트림은 256바이트의 고정 길이를 가지고 있

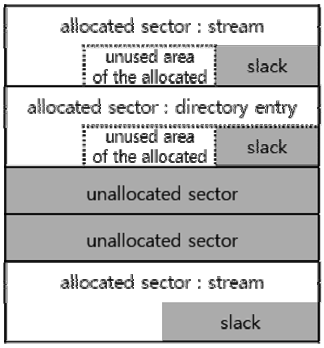


Fig. 3. Reserved area in Allocated area

으며, 한글 HwpScan2(signature), 문서 버전, 문서 속성 정보 등이 저장된다. FileHeader 스트림에 기록된 데이터가 256바이트 이하일 경우 스트림 내 미사용 영역이 발생하게 된다. Fig. 4.는 FileHeader 스트림의 구조를 나타낸다.

예시로 작성된 스트림 내에는 한글 문서의 압축여부와 버전정보, 암호설정 여부, 배포용 문서 설정 여부 정보만 기록되어 FileHeader 스트림의 총 크기 256바이트 중 48바이트 크기만 사용하고 있어 208바이트만큼의 미사용 영역이 발생하게 된다. Fig 4. 예시처럼 할당된 스트림 내 예약 영역에 데이터가 은닉되어도 응용프로그램 실행에 아무런 문제가 없으며, 응용프로그램 상에서 확인이 불가능하다.

### 4.2 복합문서 파일 형식 조작을 통한 데이터 은닉방법

복합문서 파일은 데이터의 추가 및 삭제가 자유롭다. 데이터를 추가하여 임의로 스트림, 스토리지를 생성할 수 있는데 그 방법은 크게 2가지로 나눌 수 있다. 첫 번째 방법으로는 Windows에서 제공하는 API를 직접 호출하는 프로그램을 구현하는 방법이 있고, 두 번째 방법으로는 SS Viewer[16], CFX[17] 등의 복합문서 파일 관리 도구를 사용하여 스트림, 스토리지를 추가할 수 있다. Fig. 5.은 복합문서 파일 관리 도구를 사용하여 임의로 스트림, 스토리지를 추가한 것이며, 이와 같이 임의로 추가된 스트림, 스토리지는 응용프로그램에서 인식하지 않으므로 Fig. 6.과 같이 숨겨진 영역으로 활용하여 데이터를 은닉시킬 수 있다.

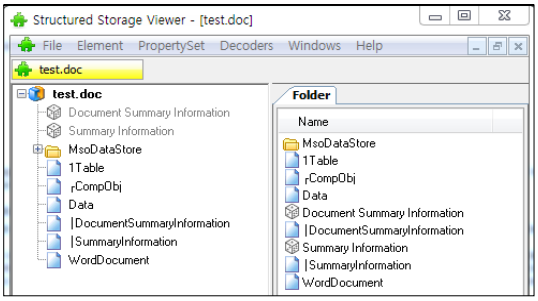


Fig. 5. Additional streams and storage using SS viewer

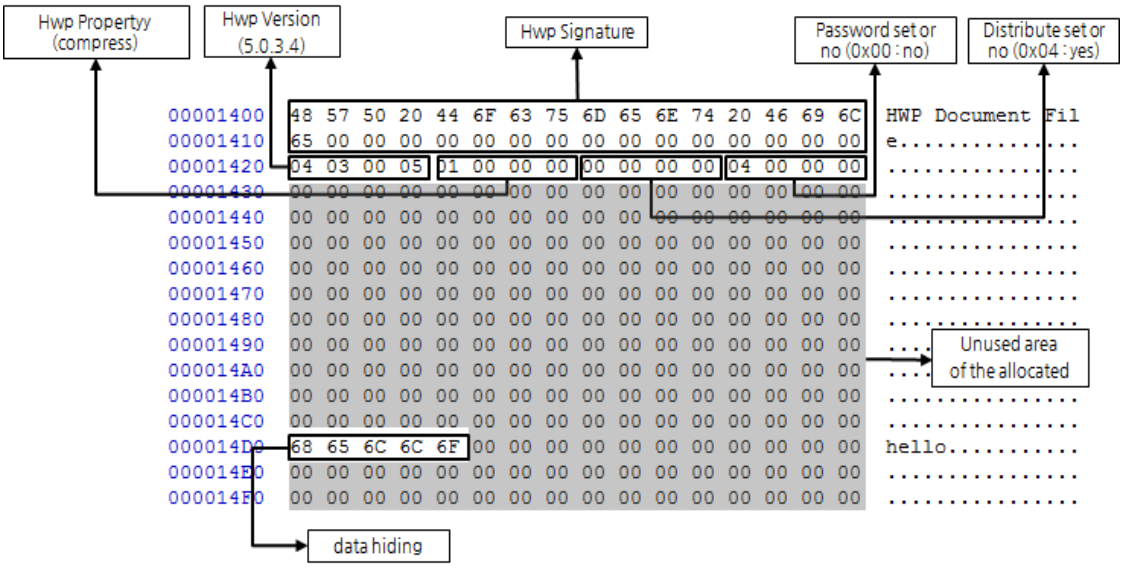


Fig. 4. FileHeader Streams Structure

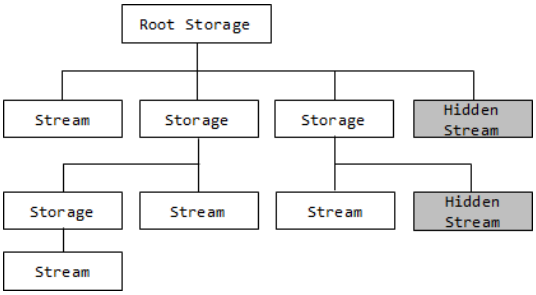


Fig. 6. Recognition of streams and storage added randomly in the application

### 4.3 기타

기존에 연구된 데이터 은닉 방법을 이용하여 복합문서 파일에 데이터 은닉이 가능하다. 은닉 방법은 다음과 같다.

- 스테가노그래피 도구를 이용하여 데이터가 은닉된 멀티미디어 파일을 문서에 삽입하기
- 은닉 메시지를 인코딩하여 입력된 공백이나 탭의 개수, 입력된 구두점의 개수로 표현하기

제시된 방법들은 복합문서 파일의 특징을 이용한 것이 아닌 기존에 알려진 방법들을 활용하여 복합문서 파일에 은닉하는 방법들이다.

## V. 데이터 은닉 탐지도구 평가

4장에서 제시된 11가지 데이터 은닉 방법을 사용하여 복합문서 파일에 데이터를 은닉해보고 복합문서 파일 내 데이터 은닉 여부를 탐지하는 도구를 사용하여 은닉 방법 별 탐지여부를 확인해보았다. 데이터 은닉 실험을 위한 커버 파일로 MS Office 제품군과 한컴오피스 파일을 사용했다.

Table 2.의 실험결과를 보면 문서작성 프로그램의 기능을 이용하여 문서 속성정보 항목에 데이터를 은닉한 경우 3개 도구 모두 탐지하였지만 속성정보가 저장되는 스트림 내 미사용 영역에 은닉한 데이터는 탐지되지 않았다. 이 외 다른 방법으로 은닉된 데이터는 대부분의 탐지도구에서 탐지되지 않았다.

## VI. 은닉된 데이터의 탐지 방법

복합문서 파일 형식의 파일에 임의로 추가된 스트림, 스토리지에 대한 탐지 방법은 워드, 엑셀, 파워포인트, 한글 등의 복합문서 파일에서 유효한 스트림, 스토리지 목록을 체크하는 것이다. 체크 방법은 Windows에서 제공하는 API(StgOpenStorage, StgOpenStorageEx)를 호출하는 프로그램을 구현하거나 SS Viewer와 같은 복합문서 파일 관리 도구를 통해 임의로 추가되거나 수정된 스트림, 스토리지 확인이 가능하다.

또한, 복합문서 파일은 구조 상 존재하는 3가지 종류의 미사용 영역을 상세 분석하여 은닉된 데이터를 찾는 방법 역시 가능하다. 예를 들어, 한글파일에서 Hwpsummaryinformation 스트림 영역을 살펴보도록 하자. 은닉된 데이터를 탐지하기 위해서는 해당 스트림을 분석하여 저장된 속성 정보의 개수와 마지막 속성정보가 기록된 위치 정보를 확인한 후 스트림 내 미사

Table 2. Experiment results of the hidden data detection tool (P:PPT, W:WORD, E:EXCEL, H:HWP)

| Method \ Tool | | Catalogue | | | | Workshare Professional | | | | Metadata Assistant 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P | W | E | H | P | W | E | H | P | W | E | H |
| application control | small size font | X | X | X | X | X | X | X | X | X | X | X | X |
| | same font color and background color | X | X | X | X | X | X | X | X | X | X | X | X |
| | hididen text behind a image file | X | X | X | X | X | X | X | X | X | X | X | X |
| | insert a data in property information area | O | O | O | X | O | O | O | X | O | O | O | X |
| | other hyperlink | X | X | X | X | O | O | O | X | O | O | O | X |
| unused area | unallocated area | X | X | X | X | X | X | X | X | X | X | X | X |
| | slack area | X | X | X | X | X | X | X | X | X | X | X | X |
| | reserved area | X | X | X | X | X | X | X | X | X | X | X | X |
| compound file control | add stream | X | X | X | X | X | X | X | X | X | X | X | X |
| etc. | insert a data-hidden file | X | X | X | X | X | X | X | X | X | X | X | X |
| | present space or dot to encoded data | X | X | X | X | X | X | X | X | X | X | X | X |

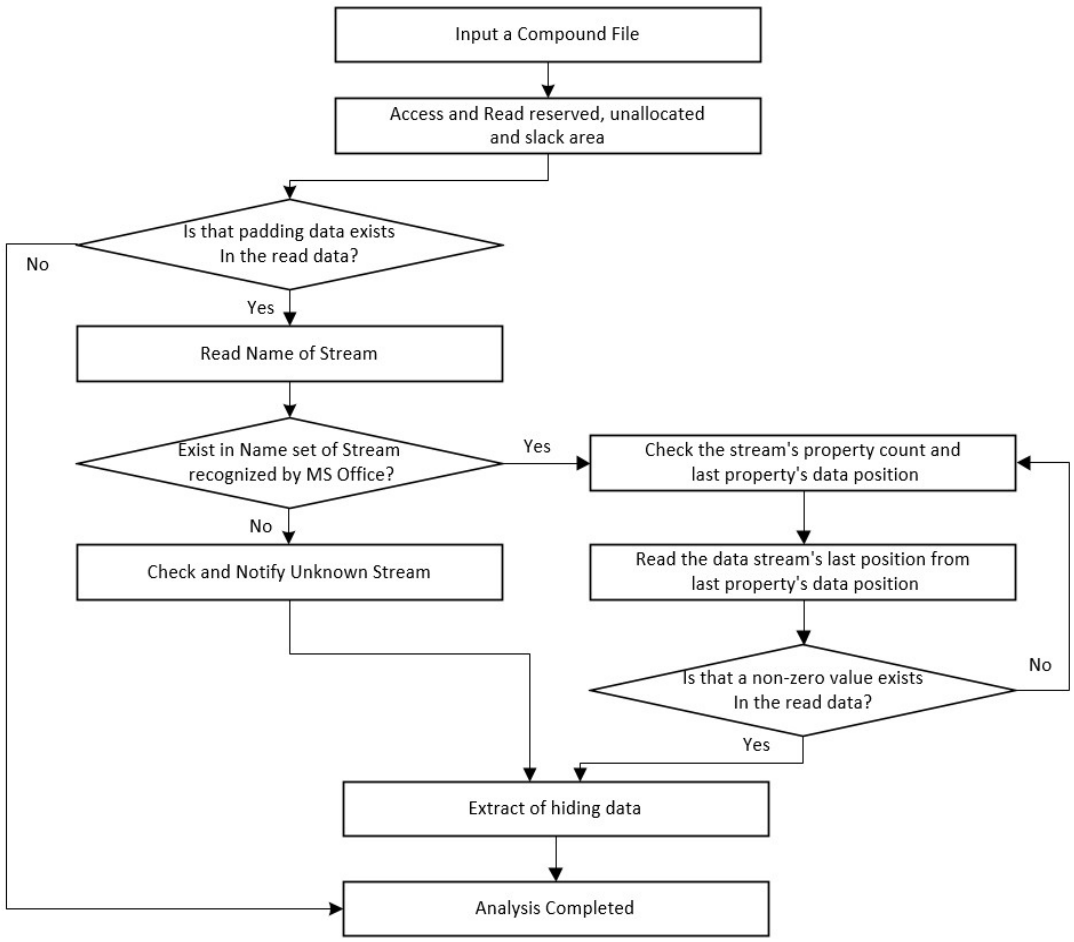


Fig. 7. Hidden data in Compound-file detection algorithm

용 영역에 0x00이 아닌 다른 데이터가 기록되어 있는지 확인하면 된다. 만약 0x00이 아닌 다른 데이터가 존재한다면 은닉된 데이터가 있다고 판단할 수 있다. Hwpsummaryinformation 스트림 외에도 다른 복합문서 파일 내 생성되는 다른 스트림 역시 스트림 내 미사용 영역을 파악하여 은닉 데이터의 존재 여부확인이 가능하다. Fig. 7.은 복합문서 파일 내에 은닉된 데이터를 탐지하기 위한 알고리즘을 나타낸다.

탐지 순서는 확인하려는 복합문서 파일의 미사용 영역을 검사하고 패딩 데이터가 아닌 다른 데이터가 존재한다면 데이터 추출 후 분석을 수행한다. 미사용 영역에 패딩 데이터 존재 시 복합문서 파일 내 존재하는 모든 스트림의 이름을 확인한 후 MS Office에서 지원하지 않는 스트림 이름이 존재할 경우 해당 스트림 내에 존재하는 데이터를 추출 후 분석을 수행한다. 마지막으로 스트림 내 미사용 영역을 확인하여 0x00이 아닌 다른 값이 존재한다면 추출 후 분석을 수행한다.

## VII. 탐지 알고리즘 평가

위에서 제안한 복합문서 파일 내 은닉된 데이터 탐지 알고리즘을 평가하기 위해 복합문서 파일구조를 사용하고 있는 4가지 파일형식에 본 논문에 제시된 데이터 은닉 11가지 방법을 적용하여 탐지 여부를 확인하였다. 실험 결과는 Table 3.과 같다.

용프로그램 조작을 통해 복합문서 내 눈으로 식별하기 어렵게 데이터를 은닉한 경우 문서 내 입력된 데이터들이 저장되는 스트림을 확인하여 은닉 여부

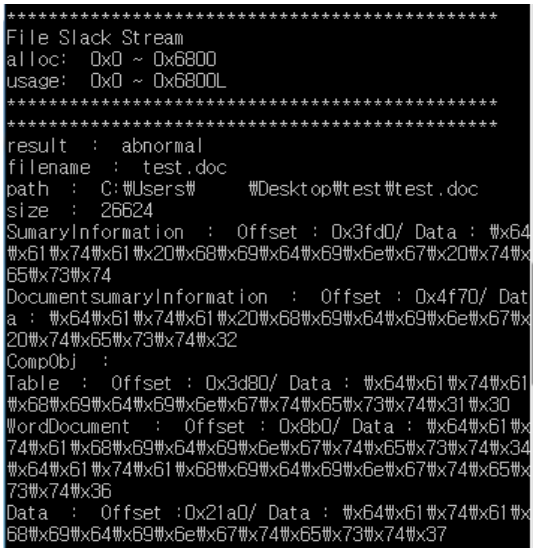


Fig 8. detect to hidden data

확인이 가능하다. Table 4.는 복합문서 파일 내 입력된 데이터가 저장되는 스트림과 스토리지 목록이다.

복합문서의 구조적 특징을 이용하여 데이터를 은닉할 경우 구조에 대한 이해가 없이는 쉽게 확인 할 수 없다. 본 논문에서 제안한 알고리즘으로 은닉 여부를 확인 결과 복합문서 파일의 구조적 특징을 이용하여 은닉된 데이터를 정상적으로 탐지한다.

Table 4. Check Stream / Storage List

| extension | stream / storage |
|---|---|
| PPT | PowerPointDocument |
| EXCEL | WorkBook |
| WORD | WordDocument |
| HWP | BodyText |

## VIII. 결 론

기존의 정보 은닉은 멀티미디어 파일이나 파일시스템의 슬랙 영역에 데이터를 삽입하는 방법이 대부분이었다. 하지만 최근에는 정보 은닉기술이 고도화되면서 특정 파일의 구조나 특징을 이용하여 데이터를 은닉할 수 있는 방법이 이용되고 있다. 이러한 방법으로 데이터가 은닉될 경우 데이터 은닉 여부를 모르는 일반인은 쉽게 파악할 수 없다.

본 논문에서는 복합문서 파일의 구조 상 발생하는 미사용 영역을 분류하고 해당 영역에 데이터를 은닉하고 문서작성 프로그램의 기능을 이용하여 데이터를 은닉할 수 있는 방법과 함께 미사용 영역에 은닉된 데이터를 탐지할 수 있는 방법을 살펴보았다.

향후에는 Microsoft Office 2007 이상 버전에서 사용되는 Open Office XML 파일에서 데이터를 은닉할 수 있는 방법과 은닉된 데이터를 탐지하기

Table 3. Experiment results of the hidden data detection algorithm

| method | | result | | | |
|---|---|---|---|---|---|
| | | PPT | WORD | EXCEL | HWP |
| application control | small size font | X | X | X | X |
| | same font color and background color | X | X | X | X |
| | hididen text behind a image file | X | X | X | X |
| | insert a data in property information area | O | O | O | O |
| | other hyperlink | O | O | O | O |
| unused area | unallocated area | O | O | O | O |
| | slack area | O | O | O | O |
| | reserved area | O | O | O | O |
| compound file control | add stream | O | O | O | O |
| etc. | insert a data-hidden file | X | X | X | X |
| | present space or dot to encoded data | X | X | X | X |

위한 방법을 연구할 예정이다.

## References


[1] Byers, S., "Information leakage caused by hidden data in published documents," IEEE Security Privacy vol. 2, no. 2, pp. 23-27, Mar. 2004.

[2] T. Ngo, "Office Open XML Overview," ECMA TC45 white paper, online at http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf, last accessed April. 2011.

[3] Daniel Rentz, "Microsoft Compound Document- File Format", http://www.openoffice.org/sc/compdocfileformat.pdf

[4] http://www.hancom.com/forMatQna.boardIntro.do

[5] http://www.yonhapnews.co.kr/bulletin/2015/03/02/0200000000AKR20150302142100009.HTML

[6] Jung Heum Park, Bora Park, Sangjin Lee, Seokhie Hong, and Jong Hyuk Park, "Extraction of Residual Information in the Microsoft PowerPoint file from the Viewpoint of Digital Forensics considering PerCom Environment," The 2nd International Workshop on Web and Pervasive Security, IEEE, pp.584-589, Mar. 2008.

[7] A. Castiglione, De Santis, and C. Soriente, "Taking advantages of a disadvantage : Digital forensics and steganography using document metadata," The Journal of Systems and software, vol 80, Issue 5, pp.750-764, May. 2007.

[8] Hyukdon Kwon, Yeog Kim, Sangin 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 ICISS, p.141-146, Jan. 2008.

[9] http://www.payneconsulting.com/products.

[10] http://peccatte.karefil.com/software/Catalogue.

[11] http://www.workshare.com

[12] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for datahiding?," IBM Syst. J., vol. 35, no. 3.4, pp. 313-336, Apr. 1996.

[13] G. Sui and H. Luo, "A new steganography method based on hypertext?," in Proc. Radio Science Conf. pp. 181-184. Aug. 2004.

[14] A. M. S. Rahma, B. AbdulWahab, and A. Y. Al-Noori, "Proposed steganographic method for data hiding in Microsoft word documents structure," Al-Mansour Journal, no. 15, pp. 1-29, 2011.

[15] W. Bhaya, A. Rahma, and D. Al-Nasrawi, "Text steganography based on font type in ms-word documents," Journal of Computer Science, vol. 9, no. 7, pp. 898-904, 2013.

[16] http://www.mitec.cz/ssv.html

[17] http://compound-file-explorer.software.informer.com/1.8/

## 〈저자소개〉


김 은 광 (Eun-kwang Kim) 학생회원
2013년 2월: 가천대학교 수학정보학과 졸업
2015년 2월: 서울과학종합대학원 MBA 졸업
2014년 3월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 디지털 포렌식, 정보은닉

전 상 준 (Sang-jun Jeon) 학생회원
2010년 2월: 고려대학교 정보경영공학과 공학사
2010년 2월~현재: 고려대학교 디지털 포렌식 연구센터 석박사통합과정
〈관심분야〉 디지털 포렌식

한 재 혁 (Jae-hyeok Han) 학생회원
2011년 2월: 서울시립대학교 수학과 졸업
2014년 3월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 디지털 포렌식, 파일시스템

이 민 욱 (Min-wook Lee) 학생회원
2014년 2월: 경기대학교 전자공학과 졸업
2014년 9월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 디지털 포렌식, 역공학

이 상 진 (Sang-jin Lee) 종신회원
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