(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년06월27일
(11) 등록번호 10-1745873
(24) 등록일자 2017년06월05일

(51) 국제특허분류(Int. Cl.)
G06F 21/56 (2013.01) *G06F 21/50* (2013.01)
G06F 21/60 (2013.01)
(52) CPC특허분류
G06F 21/56 (2013.01)
G06F 21/50 (2013.01)
(21) 출원번호 10-2015-0182057
(22) 출원일자 2015년12월18일
심사청구일자 2015년12월18일
(56) 선행기술조사문헌
KR1020090055669 A
KR1020080043201 A
KR100628869 B1
KR100620313 B1

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이상진
서울특별시 종로구 사직로8길 20, 101동 1503호 (내수동, 파크팰리스)
박정흠
경기도 부천시 소사구 심곡로10번길 17, 1동 904호 (심곡본동, 부천아름다운나날아파트)
(뒷면에 계속)
(74) 대리인
특허법인충현

전체 청구항 수 : 총 11 항 심사관 : 윤혜숙

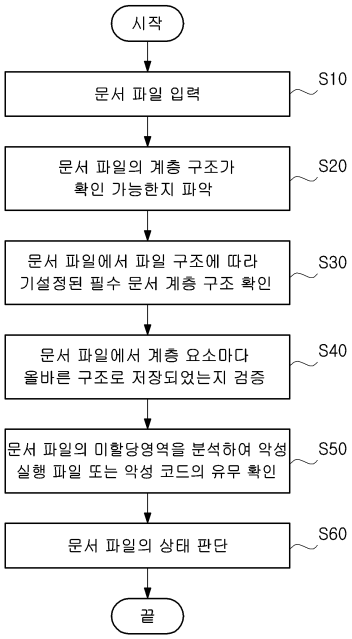
(54) 발명의 명칭 **악성 문서 파일 식별 시스템 및 방법**

(57) 요 약

본 발명은 악성 문서 파일 식별 시스템 및 방법에 관한 것으로, 악성 여부 확인이 필요한 문서 파일을 입력 받는 단계와, 상기 문서 파일의 계층 구조가 확인 가능한지 파악하는 제 1 분석 단계와, 상기 문서 파일에서 파일 구조에 따라 기설정된 필수 문서 계층 구조를 확인하는 제 2 분석 단계와, 상기 문서 파일에서 계층 요소마다 올바

(뒷면에 계속)

대 표 도 - 도3



른 구조로 저장되었는지 검증하는 제 3 분석 단계와, 상기 문서 파일의 미할당 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 제 4 분석 단계 및 상기 제 1 내지 제 4 분석 단계의 결과를 통합하여 상기 문서 파일의 상태를 판단하는 단계를 포함하여 구성된다.

상기와 같이 구성된 본 발명에 따른 악성 문서 파일 식별 시스템 및 방법에 의하면, 문서 파일을 직접 실행하지 않고 새롭게 나타난 악성 문서 파일을 탐지할 수 있으며, 문서 파일의 악성 여부와 함께 손상된 파일에 대해서도 탐지할 수 있고, 매크로 기능을 사용하는 악성 문서 파일에 대한 탐지율도 향상시키는 효과가 있다.

(52) CPC특허분류
G06F 21/563 (2013.01)
G06F 21/565 (2013.01)
G06F 21/566 (2013.01)
G06F 21/60 (2013.01)

(72) 발명자
연규철
서울특별시 서대문구 증가로20나길 28, 101호 (북가좌동, 길훈맨션)
최종현
서울특별시 동대문구 천장산로6길 35-4 (이문동)

최호용
서울특별시 성북구 안암로9가길 34 (안암동5가)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711026813
부처명 미래창조과학부
연구관리전문기관 정보통신기술진흥센터
연구사업명 정보통신기술인력양성사업
연구과제명 제어 시스템 보안 연구(Control System Security)
기 여 율 1/1
주관기관 고려대학교 산학협력단
연구기간 2015.01.01 ~ 2015.12.31

명 세 서

청구범위

청구항 1

입력부가 악성 여부 확인이 필요한 문서 파일을 입력 받는 단계;

제1 분석부가 상기 문서 파일마다 띄는 고유의 계층 구조가 확인 가능한지 파악하는 제 1 분석 단계;

제2 분석부가 상기 계층 구조 중 파일 포맷 별로 필수적으로 가지고 있어야 할 기설정된 필수 문서 계층 구조의 존재여부를 확인하는 제 2 분석 단계;

제3 분석부가 상기 계층 구조에 저장된 계층 요소마다 올바른 구조로 저장되었는지 검증하는 제 3 분석 단계;

제4 분석부가 상기 문서 파일 내의 미사용 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 제 4 분석 단계; 및

판단부가 상기 제 1 내지 제 4 분석 단계의 결과를 통합하여 상기 문서 파일의 상태를 판단하는 단계;를 포함하는 악성 문서 파일 식별 방법.

청구항 2

제 1 항에 있어서,

상기 제 1 분석 단계는,

상기 문서 파일의 계층 구조 확인이 불가능한 경우 상기 문서 파일을 손상된 파일로 판단하는 것을 특징으로 하는 악성 문서 파일 식별 방법.

청구항 3

제 1 항에 있어서,

상기 제 2 분석 단계는,

상기 필수 문서 계층 구조가 올바르게 저장되지 않은 경우 손상된 파일로 판단하는 것을 특징으로 하는 악성 문서 파일 식별 방법.

청구항 4

제 1 항에 있어서,

상기 제 3 분석 단계는,

각 계층 요소의 구조를 확인하여 이상 유무를 확인하는 단계;

상기 문서 파일에서 각 계층 요소의 크기 또는 사이즈를 검사하는 단계;

상기 각 계층 요소가 공개 문서에서 정의된 데이터인지 확인하는 단계; 및

상기 계층 요소 중 매크로 기능이 포함된 경우 매크로 코드를 검사하는 단계;를 포함하는 것을 특징으로 하는 악성 문서 파일 식별 방법.

청구항 5

제 1 항에 있어서,

상기 제 4 분석 단계는,

상기 계층 요소 중 파일 유효 범위 외의 미할당 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계;

상기 계층 요소에서 발생하는 슬랙(Slack) 부분에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계; 및

상기 계층 요소의 삭제로 파일 중간에 생기는 상기 미사용 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계;를 포함하는 것을 특징으로 하는 악성 문서 파일 식별 방법.

청구항 6

제 1 항 내지 제 5 항 중 어느 한 항에 따른 방법을 컴퓨터로 실행하기 위한 프로그램이 기록된 컴퓨터 판독가능 기록매체.

청구항 7

악성 여부 확인이 필요한 문서 파일을 입력 받는 입력부;

상기 문서 파일마다 띄는 고유의 계층 구조가 확인 가능한지 파악하는 제 1 분석부;

상기 계층 구조 중 파일 포맷 별로 필수적으로 가지고 있어야 할 기설정된 필수 문서 계층 구조의 존재 여부를 확인하는 제 2 분석부;

상기 문서 파일에서 상기 계층 구조에 저장된 계층 요소마다 올바른 구조로 저장되었는지 검증하는 제 3 분석부;

상기 문서 파일 내의 미사용 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 제 4 분석부; 및

상기 제 1 내지 제 4 분석부의 결과를 통합하여 상기 문서 파일의 상태를 판단하는 판단부;를 포함하는 악성 문서 파일 식별 시스템.

청구항 8

제 7 항에 있어서,

상기 제 1 분석부는,

상기 문서 파일의 계층 구조 확인이 불가능한 경우 상기 문서 파일을 손상된 파일로 판단하는 것을 특징으로 하는 악성 문서 파일 식별 시스템.

청구항 9

제 7 항에 있어서,

상기 제 2 분석부는,

상기 필수 문서 계층 구조가 올바르게 저장되지 않은 경우 손상된 파일로 판단하는 것을 특징으로 하는 악성 문서 파일 식별 시스템.

청구항 10

제 7 항에 있어서,

상기 제 3 분석부는,

각 계층 요소의 구조를 확인하여 이상 유무를 확인하고, 상기 문서 파일에서 각 계층 요소의 크기 또는 사이즈를 검사하고, 상기 각 계층 요소가 공개 문서에서 정의된 데이터인지 확인하며, 상기 계층 요소 중 매크로 기능이 포함된 경우 매크로 코드를 검사하는 것을 특징으로 하는 악성 문서 파일 식별 시스템.

청구항 11

제 7 항에 있어서,

상기 제 4 분석부는,

상기 계층 요소 중 파일 유효 범위 외의 미할당 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유

무를 확인하고, 상기 계층 요소에서 발생하는 슬랙(Slack) 부분에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하며, 상기 계층 요소의 삭제로 파일 중간에 생기는 상기 미사용 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 것을 특징으로 하는 악성 문서 파일 식별 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 악성 문서 파일 식별 시스템 및 방법에 관한 것으로, 더욱 상세하게는 문서 파일의 구조를 파악하고 파악한 문서 파일의 구조와 맞지 않는 부분에 대해서 검증하여 악성 문서 파일인지 여부를 식별하는 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 워드프로세서는 문서를 작성, 편집, 저장 및 인쇄할 때 사용하는 소프트웨어를 말한다. 대표적인 워드프로세서로는 한글과 컴퓨터에서 출시한 한컴오피스, Microsoft에서 출시한 Microsoft Office, Apache재단에서 개발하고 있는 Apache OpenOffice, Microsoft에서 출시한 RTF(Rich Text Format), Adobe에서 출시한 PDF(Portable Document Format) 등이 있다.

[0003] 한컴오피스는 한글, 한셀, 한쇼 등이 있다. 한글은 hwp, 한셀은 cell, 한쇼는 hpt 확장자를 사용한다. 한컴오피스는 복합 파일 이진구조(Compound File Binary Format)로 파일을 저장한다. 복합 파일 이진 구조는 텍스트, 오디오, 동영상 등 서로 다른 데이터 형식을 포함할 수 있는 구조이다.

[0004] Microsoft Office는 Word, Excel, PowerPoint 등이 있다. Microsoft Office는 97~2003까지 버전은 복합 파일 이진 구조로 파일을 저장하고 있다. Word는 doc 확장자를 사용하며, Excel은 xls, PowerPoint는 ppt 확장자를 사용한다. 그 이후 버전에서는 OOXML(Office Open XML)로 저장하며, Word는 docx, Excel은 xlsx, PowerPoint는 pptx 확장자를 사용한다.

[0005] Apache OpenOffice는 XML 형식을 기반으로 한 파일 구조(Open Document Format)를 사용하고 있다. Apache OpenOffice는 odt, odp, ods 확장자를 사용한다.

[0006] 그 밖에 RTF, PDF는 독자적인 파일 구조를 가지고 있으며, 확장자는 rtf와 pdf를 사용하고 있다.

[0007] 각 워드프로세서는 각각 다른 파일 구조를 가지고 있지만, 문서를 작성, 편집, 저장 및 인쇄하는 것에 초점이 맞춰져 있는 프로그램이다. 그러므로 기능들은 비슷하고, 저장하는 데이터도 비슷하다.

[0008] 각 워드프로세서의 문서 파일 구조에서 지칭하는 용어는 다르나, 크게 보면 폴더와 파일과 같은 개념으로 저장된다. 폴더는 파일들을 저장하고 있으며, 파일들은 해당 문서파일에서 실질적으로 저장하고자 하는 데이터를 저장하고 있다.

선행기술문헌

특허문헌

[0009] (특허문헌 0001) 대한민국 등록특허 제10-1060596호 (2011년 08월 24일)

발명의 내용

해결하려는 과제

[0010] 본 발명은 상술한 각각의 워드프로세서에 대해서 악성 문서 파일을 판별하는 새로운 방법 및 시스템을 제안하기 위한 것으로, 문서 파일 구조에 기반을 두어 악성 문서 파일을 탐지함으로써, 문서 파일을 직접 실행하지 않고 새롭게 나타난 악성 문서 파일을 탐지할 수 있는 악성 문서 파일을 판별하는 방법 및 시스템을 제공하는 것에 그 목적이 있다.

[0011] 또한 본 발명은 문서 파일 구조 및 미할당 영역을 검사함으로써, 문서 파일의 악성 여부와 함께 손상된 파일에 대해서도 탐지할 수 있으며, 매크로 기능을 사용하는 악성 문서 파일에 탐지율을 향상시킬 수 있는 악성 문서

파일을 판별하는 방법 및 시스템을 제공하는 것에 그 목적이 있다.

과제의 해결 수단

[0012] 상기 목적을 달성하기 위해 본 발명은 입력부가 악성 여부 확인이 필요한 문서 파일을 입력받는 단계와, 제1 분석부가 상기 문서 파일마다 띄는 고유의 계층 구조가 확인 가능한지 파악하는 제 1 분석 단계와, 제2 분석부가 상기 계층 구조 중 파일 포맷 별로 필수적으로 가지고 있어야 할 기설정된 필수 문서 계층 구조의 존재여부를 확인하는 제 2 분석 단계와, 제3 분석부가 상기 계층 구조에 저장된 계층 요소마다 올바른 구조로 저장되었는지 검증하는 제 3 분석 단계와, 제4 분석부가 상기 문서 파일 내의 미사용 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 제 4 분석 단계 및 판단부가 상기 제 1 내지 제 4 분석 단계의 결과를 통합하여 상기 문서 파일의 상태를 판단하는 단계를 포함하여 구성된다.

[0013] 본 발명에 따른 악성 문서 파일 식별 방법에 있어서, 상기 제 1 분석 단계는 상기 문서 파일의 계층 구조 확인이 불가능한 경우 상기 문서 파일을 손상된 파일로 판단하는 것을 특징으로 한다.

[0014] 본 발명에 따른 악성 문서 파일 식별 방법에 있어서, 상기 제 2 분석 단계는 상기 필수 문서 계층 구조가 올바르게 저장되지 않은 경우 손상된 파일로 판단하는 것을 특징으로 한다.

[0015] 본 발명에 따른 악성 문서 파일 식별 방법에 있어서, 상기 제 3 분석 단계는 상기 각 계층 요소의 구조를 확인하여 이상 유무를 확인하는 단계와, 상기 문서 파일에서 각 계층 요소의 파일 크기 또는 사이즈를 검사하는 단계와, 상기 각각의 계층 요소가 공개 문서에서 정의된 데이터인지 확인하는 단계 및 상기 계층 요소 중 매크로 기능이 포함된 경우 매크로 코드를 검사하는 단계를 포함하는 것을 특징으로 한다.

[0016] 본 발명에 따른 악성 문서 파일 식별 방법에 있어서, 상기 제 4 분석 단계는 상기 계층 요소 중 파일 당유효 범위 외의 미할당 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계와, 상기 계층 요소에서 발생하는 슬랙(Slack) 부분에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계 및 상기 계층 요소의 삭제로 파일 중간에 생기는 상기 미사용 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 단계를 포함하는 것을 특징으로 한다.

[0017] 또한 본 발명은 악성 여부 확인이 필요한 문서 파일을 입력 받는 입력부와, 상기 문서 파일마다 띄는 고유의 계층 구조가 확인 가능한지 파악하는 제 1 분석부와, 상기 계층 구조 중 파일 포맷 별로 필수적으로 가지고 있어야 할 기설정된 필수 문서 계층 구조의 존재 여부를 확인하는 제 2 분석부와, 상기 계층 구조에 저장된 계층 요소마다 올바른 구조로 저장되었는지 검증하는 제 3 분석부와, 상기 문서 파일 내의 미사용 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 제 4 분석부 및 상기 제 1 내지 제 4 분석부의 결과를 통합하여 상기 문서 파일의 상태를 판단하는 판단부를 포함하여 구성된다.

[0018] 본 발명에 따른 악성 문서 파일 식별 시스템에 있어서, 상기 제 1 분석부는 상기 문서 파일의 계층 구조 확인이 불가능한 경우 상기 문서 파일을 손상된 파일로 판단하는 것을 특징으로 한다.

[0019] 본 발명에 따른 악성 문서 파일 식별 시스템에 있어서, 상기 제 2 분석부는 상기 필수 문서 계층 구조가 올바르게 저장되지 않은 경우 손상된 파일로 판단하는 것을 특징으로 한다.

[0020] 본 발명에 따른 악성 문서 파일 식별 시스템에 있어서, 상기 제 3 분석부는 각 계층 요소의 구조를 확인하여 이상 유무를 확인하고, 상기 문서 파일에서 각 계층 요소의 파일 크기 또는 사이즈를 검사하고, 상기 각각의 계층 요소가 공개 문서에서 정의된 데이터인지 확인하며, 상기 계층 요소 중 매크로 기능이 포함된 경우 매크로 코드를 검사하는 것을 특징으로 한다.

[0021] 본 발명에 따른 악성 문서 파일 식별 시스템에 있어서, 상기 제 4 분석부는 상기 계층 요소 중 파일 유효 범위 외의 미할당 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하고, 상기 계층 요소에서 발생하는 슬랙 부분에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하며, 상기 계층 요소의 삭제로 파일 중간에 생기는 상기 미사용 영역에 대해서 상기 악성 실행 파일 또는 상기 악성 코드의 유무를 확인하는 것을 특징으로 한다.

발명의 효과

[0022] 상기와 같이 구성된 본 발명에 따른 악성 문서 파일 식별 시스템 및 방법에 의하면, 문서 파일 구조에 기반을 두어 악성 문서 파일을 탐지함으로써, 문서 파일을 직접 실행하지 않고 새롭게 나타난 악성 문서 파일을 탐지할

수 있다.

[0023] 또한 본 발명에 따른 악성 문서 파일 식별 시스템 및 방법에 의하면, 문서 파일 구조 및 미할당 영역을 검사함으로써, 문서 파일의 악성 여부와 함께 손상된 파일에 대해서도 탐지할 수 있으며, 매크로 기능을 사용하는 악성 문서 파일에 대한 탐지율도 향상시키는 효과가 있다.

도면의 간단한 설명

[0024] 도 1은 본 발명의 일 실시 예에 따른 악성 문서 파일 식별 시스템을 나타내는 구성도이다.

도 2는 본 발명의 일 실시 예에 따른 미할당 영역을 나타내는 도면이다.

도 3은 본 발명의 바람직한 실시 예에 따른 악성 문서 파일 식별 방법을 나타내는 순서도이다.

도 4는 본 발명의 일 실시 예에 따른 문서 파일 구조에 따른 필수 문서 계층 구조를 나타내는 도면이다.

도 5는 본 발명의 일 실시 예에 따른 문서 파일에서 계층 요소마다 올바른 구조로 저장되었는지 검증하는 방법을 나타내는 순서도이다.

도 6은 본 발명의 일 실시 예에 따른 문서 파일의 미할당 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 방법을 나타내는 순서도이다.

도 7은 본 발명의 일 실시 예에 따른 악성 문서 파일 식별 결과를 나타내는 도면이다.

도 8은 본 발명의 일 실시 예에 따른 JSON 파일로 저장된 악성 문서 파일 식별 결과를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0025] 본 발명은 다양한 변형 및 여러 가지 실시 예를 가질 수 있는바, 특정 실시 예들을 도면에 예시하고 상세한 설명에 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0026] 이하, 본 발명의 바람직한 실시 예를 첨부된 도면과 참조하여 상세히 설명하기로 한다.

[0027] 도 1은 본 발명의 일 실시 예에 따른 악성 문서 파일 식별 시스템을 나타내는 구성도이다. 도 1을 참조하면, 악성 문서 파일 식별 시스템(100)은 입력부(110), 분석부(120), 판단부(130)를 포함한다.

[0028] 입력부(110)는 분석하고자 하는 문서 파일을 입력 받고 분석부(120)는 제 1 분석부(121), 제 2 분석부 (122), 제 3 분석부(123), 제 4 분석부(124)를 포함하며, 입력부(110)로부터 입력 받은 문서 파일을 분석한다.

[0029] 여기서 제 1 분석부(121)는 문서 파일의 계층 구조가 확인 가능한지 파악하는 역할을 수행하고, 제 2 분석부(122)는 문서 파일에서 파일 구조에 따라 기설정된 필수 문서 계층 구조를 확인한다.

[0030] 또한 제 3 분석부(123)는 문서 파일에서 계층 요소마다 올바른 구조로 저장되었는지 검증하며, 제 4 분석부(124)는 문서 파일의 미할당 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인한다.

[0031] 특히 제 4 분석부(124)에서 확인하는 미할당 영역은 크게 세 종류로 구분할 수 있으며, 세 종류의 미할당 영역은 도 2를 참조하여 설명하겠다.

[0032] 도 2는 본 발명의 일 실시 예에 따른 미할당 영역을 나타내는 도면이다. 문서 파일에서 미할당 영역은 도 2와 같이 크게 세 종류로 구분할 수 있다.

[0033] 먼저, 첫 번째 미할당 영역은 도 2의 (a)와 같이 파일 유효 범위 외의 미할당 영역으로 정상적인 파일 뒤에 악의적인 행위를 하는 코드나 실행 파일을 덧붙인 파일이 대표적인 예이다. 이 경우 파일 구조를 해석하면, 할당 영역을 계산할 수 있다.

[0034] 다음으로, 두 번째 미할당 영역은 도 2의 (b)와 같이 계층 요소에서 슬랙(Slack) 부분이 발생한 경우를 의미하며, 이 경우에는 파일이 할당한 영역보다 데이터가 짧은 경우 생성될 수 있다.

[0035] 마지막으로, 세 번째 미할당 영역은 도 2의 (b)와 같이 계층 요소가 삭제되어 파일의 중간에 사용하지 않은 영역을 의미한다.

[0036] 이와 같은 세 종류의 미할당 영역에는 악의적인 사용자가 악의적인 프로그램이나 악의적인 악성 코드를 저장하

는데 사용된다. 따라서, 미할당 영역에서 악성 실행 코드와 악성 실행 파일의 존재 여부를 확인해야 한다.

[0037] 다음으로는 다시 도 1로 돌아와 악성 문서 파일 식별 시스템에 대해서 설명하겠다. 판단부(130)는 제 1 분석부(121), 제 2 분석부(122), 제 3 분석부(123), 제 4 분석부(124) 각각에서 분석한 결과를 통합하여 입력부(110)에서 입력 받은 문서 파일의 상태를 판단한다.

[0038] 여기서 판단부(130)는 문서 파일의 상태를 판단 시 악성 여부에 대한 판단 뿐만 아니라 문서 파일의 손상에 대해서도 확인할 수 있다.

[0039] 이상으로 악성 문서 파일 식별 시스템에 대해서 설명하였다. 다음으로는 도 3을 참조하여 악성 문서 파일 식별 방법에 대해서 설명하겠다.

[0040] 도 3은 본 발명의 바람직한 실시 예에 따른 악성 문서 파일 식별 방법을 나타내는 순서도이다.

[0041] 도 3을 참조하면, 악성 문서 파일 식별 시스템은 악성 여부 확인이 필요한 문서 파일을 입력 받고(S10), 입력 받은 문서 파일에 대해서 악성 여부를 분석한다.

[0042] 여기서 악성 문서 파일 식별 시스템은 악성 여부 확인을 위해 문서 파일을 직접 실행하거나 가상의 환경에서 실행하여 악성 여부를 판단하지 않고, 문서 파일 구조에 기반을 두어 악성 문서 파일을 탐지한다.

[0043] 보다 구체적으로 악성 여부를 분석하는 방법은 크게 4종류의 분석을 수행하여 입력 받은 문서 파일의 악성 여부를 식별한다.

[0044] 먼저, 문서 파일의 계층 구조가 확인 가능한지 파악한다(S20). 각각의 문서 파일들은 고유의 계층 구조를 띄고 있으므로, 계층 구조를 파악하여 파싱한다. 여기서 문서 파일의 계층 구조를 파악할 수 없는 경우 해당 파일은 손상된 파일을 의미한다. 이와 같이 문서 파일의 계층 구조를 파악하고 나면 단계적으로 분석이 가능하다.

[0045] 다음으로 문서 파일에서 파일 구조에 따라 기설정된 필수 문서 계층 구조를 확인한다(S30). 즉, 문서 파일의 계층 구조 중에서 문서 파일의 구조에 따라 기설정된 요소에 대해서 검사를 한다. 문서 파일 구조를 보면 필수적으로 가져야 할 요소가 있다. 이에 대해서 도 4를 참조하여 설명하겠다.

[0046] 도 4는 본 발명의 일 실시 예에 따른 문서 파일 구조에 따른 필수 문서 계층 구조를 나타내는 도면이다. 문서 파일은 파일 구조에 따라 도 4와 같이 문서 파일 정보를 저장하고 있는 부분, 사용자가 입력한 텍스트 데이터를 저장하는 부분, 사용자가 설정한 텍스트 크기와 텍스트 정렬 등을 저장하는 부분이 저장되어야 한다.

[0047] 해당 파일이 올바르게 저장되지 않은 경우에는 필수 요소들이 손상됐을 가능성이 있다. 또한, 파일의 계층 구조를 손상시키면서까지 악성 코드를 문서 파일에 삽입하는 경우도 있다. 이러한 경우에는 악성 파일일 가능성이 있다.

[0048] 이와 같은 필수 문서 계층 구조를 확인함으로써, 해당 파일의 손상 여부를 판단할 수 있다.

[0049] 다음으로 다시 도 3으로 돌아와 문서 파일에서 계층 요소마다 올바른 구조로 저장되었는지 검증한다(S40). 여기서는 계층 요소를 검증하는 단계로 파일의 계층 구조에 대해서 파악을 완료하였으므로, 계층 구조의 요소마다 올바른 구조로 저장되었는지 검증을 하며, 구체적인 검증 방법은 도 5를 참조하여 설명하겠다.

[0050] 도 5는 본 발명의 일 실시 예에 따른 문서 파일에서 계층 요소마다 올바른 구조로 저장되었는지 검증하는 방법을 나타내는 순서도이다. 도 5를 참조하면, 악성 문서 파일 식별 시스템은 문서 파일의 구성 요소를 확인하여 이상 유무를 확인한다(S41).

[0051] 여기서 파일 설정에는 각 파일에서 사용하는 부분에 대해서 명시하고 있다. 예를 들면 해당 파일의 데이터 압축 여부, 암호화 여부, 읽기 전용 여부 등을 저장하게 된다. 만약 파일 설정에 암호화 여부에 체크가 되어 있지만, 암호화했을 경우 저장되는 폴더나 파일들이 없을 경우에는 손상된 파일이나 악성 파일로 의심할 수 있다.

[0052] 다음으로, 문서 파일에서 각 계층 요소의 크기 또는 사이즈를 검사한다(S42). 여기서 문서 파일은 파일 구조 별로 해당 요소들이 특정 바이트나 크기로 정의되어 있는 경우가 있다. 예를 들어 256 바이트의 고정 길이를 갖는 요소인데, 그 이상의 크기를 갖고 있는 경우 의심요소로 판단할 수 있다.

[0053] 계층 요소의 크기 또는 사이즈 검사 이후에는 각 계층 요소를 방문하면서 공개 문서에 정의되어 있는 데이터인지 아닌지 확인한다(S43). 악의적인 사용자가 문서 파일과 상관없는 악성 파일이나 악성 코드를 삽입할 경우 해당 데이터는 문서 파일에 정의되어 있지 않은 데이터이므로 이에 대해서 악성 의심 요소로 판별할 수 있다.

[0054] 여기서 공개 문서에 정의된 데이터인지 식별하는 방법으로는 파일 구조의 구문 검사를 이용하여 수행할 수 있다.

[0055] 마지막으로 악성 문서 파일 식별 시스템은 매크로 기능이 들어있는 요소는 매크로 코드를 검사한다(S44).

[0056] 여기서 매크로란 반복적인 작업을 빠르고 효율적으로 수행하기 위해 만든 기능으로, 프로그램에 따라 다르지만 매크로에 Visual Basic 언어, Javascript 등을 삽입할 수 있다. 이 기능은 사용자의 편리성을 제공하지만, 프로그램 언어를 삽입할 수 있으므로 문서 프로그램에 악의적인 행위를 하는 프로그램을 삽입할 수 있다.

[0057] 예를 들어 매크로 기능을 사용하여 인터넷에서 특정 프로그램을 다운로드 받아 실행시킬 수 있다. 따라서, 매크로 코드를 검사하여 악의적인 행위로 의심되는 부분을 검출한다.

[0058] 다시 도 3으로 돌아와 계층 요소 검증 후 문서 파일의 미할당 영역을 분석하여 악성 파일 또는 악성 코드의 유무를 확인한다(S50). 여기서 미할당 영역은 크게 세 종류로 분류를 하며, 구체적으로 세 종류의 미할당 영역에 대한 분석은 도 6을 참조하여 설명하겠다.

[0059] 도 6은 본 발명의 일 실시 예에 따른 문서 파일의 미할당 영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무를 확인하는 방법을 나타내는 순서도이다.

[0060] 먼저, 악성 문서 파일 식별 시스템은 파일 유효 범위 외의 미할당 영역에 대해서 악성 실행 파일 또는 악성 코드의 유를 확인한다(S51). 이는 정상적인 파일 뒤에 악의적인 행위를 하는 코드나 실행 파일을 덧붙인 파일이 대표적인 예로, 파일 구조를 해석하면 할당영역을 계산할 수 있어, 미할당 영역의 위치를 쉽게 파악할 수 있다.

[0061] 다음으로, 계층 요소에서 발생하는 슬랙 부분에 대해서 악성 실행 파일 또는 악성 코드의 유무를 확인한다(S52). 이는 파일이 할당한 영역보다 데이터가 짧은 경우 생기는 슬랙에 악의적인 행위를 하는 코드나 실행 파일을 덧붙이는 경우로, 해당 부분에 대해서 확인하여 악성 여부를 확인할 수 있다.

[0062] 마지막으로, 계층 요소의 삭제로 파일 중간에 생기는 미할당 영역에 대해서 악성 실행 파일 또는 악성 코드의 유무를 확인한다(S53).

[0063] 특히, 요즘 악의적인 실행 코드는 쉽게 발견되어 실행 코드를 난독화시키는 악성 문서 파일들이 많아지고 있다. 따라서, 미할당 영역에서 난독화를 해제하고, 해제된 미할당 영역에서 악성 실행 파일이나 악성 코드를 검사하여 악성 의심 요소가 있는지 확인한다.

[0064] 이와 같이 미할당 영역에 대해서 악성 실행 파일 또는 악성 코드의 유무 확인과 함께 할당 영역에 대해서도 악성 코드의 유무를 확인할 수 있다.

[0065] 구체적으로 할당 영역에서 악성행위를 수행하는 악성 코드와, 할당 영역에는 미할당 영역으로 이동할 수 있는 데이터를 삽입하고 악성 코드를 미할당 영역에 저장하여 악성행위를 수행하는 경우에 대해서 악성 코드의 유무를 확인한다.

[0066] 다시 도 3으로 돌아와 미할당 영역에 대한 분석 완료 후 악성 문서 파일 식별 시스템은 계층 구조, 필수 문서 계층 구조, 계층 요소, 미할당 영역 각각에서 분석한 결과를 통합하여 입력 받은 문서 파일의 상태를 판단한다(S60).

[0067] 이를 통해 악성 문서 파일 식별 시스템은 입력 받은 문서가 정상 파일인지 아니면 손상 또는 악성 파일인지 판단하며, 판단 결과에 대해서는 도 7을 참조하여 설명하겠다.

[0068] 도 7은 본 발명의 일 실시 예에 따른 악성 문서 파일 식별 결과를 나타내는 도면이다. 도 7을 참조하면, 악성 문서 파일 식별 시스템으로 입력 받은 문서 파일에 대한 악성 여부를 판단한 결과 파일을 시스템의 메모리에 저장할 수 있다.

[0069] 여기서 판단 결과에는 입력 받은 문서 파일의 식별 성공 여부와 파일의 상태를 확인할 수 있다. 또한, 악성 의심 요소와 악성 의심 요소로 판단되는 위치에 대해서도 확인할 수 있다.

[0070] 더불어 악성 문서 파일 식별 시스템은 판단 결과에 대해서 JSON 파일로 저장할 수 있으며, 이에 대해서는 도 8을 참조하여 설명하겠다.

[0071] 도 8은 본 발명의 일 실시 예에 따른 JSON 파일로 저장된 악성 문서 파일 식별 결과를 나타내는 도면이다. 도 8을 참조하면, 실제로 JSON 파일로 저장된 문서 파일을 식별한 결과인 Result에서 의심으로 나와 있음을 확인할

수 있다.

[0072] 이에 따라서 Vuln_Info에서 각 항목들은 어떠한 취약점이 식별되었는지를 나타내고 있다. 즉, Vuln 1은 매크로 코드 중 VBA코드를 삽입한 것이고, Offset은 해당 파일의 주소 값 297에서 해당 취약점이 발견되어 있음을 나타내고 있으며, Description은 해당 취약점에 대한 설명을 나타내고 있다.

[0073] 이와 같이 각 항목들에서 나타내고 있는 값들을 이용하여 악성파일임을 판단할 수 있다.

[0074] 상기와 같이 구성된 본 발명에 따른 악성 문서 파일 식별 시스템 및 방법에 의하면, 문서 파일 구조에 기반을 두어 악성 문서 파일을 탐지함으로써, 문서 파일을 직접 실행하지 않고 새롭게 나타난 악성 문서 파일을 탐지할 수 있다.

[0075] 또한 본 발명에 따른 악성 문서 파일 식별 시스템 및 방법에 의하면, 문서 파일 구조 및 미할당 영역을 검사함으로써, 문서 파일의 악성 여부와 함께 손상된 파일에 대해서도 탐지할 수 있으며, 매크로 기능을 사용하는 악성 문서 파일에 대한 탐지율도 향상시키는 효과가 있다.

[0076] 본 명세서에 기재된 본 발명의 실시 예와 도면에 도시된 구성은 본 발명의 가장 바람직한 실시 예에 관한 것이고, 발명의 기술적 사상을 모두 포괄하는 것은 아니므로, 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형 예들이 있을 수 있음을 이해하여야 한다. 따라서 본 발명은 상술한 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형실시가 가능한 것은 물론이고, 그와 같은 변경은 청구범위 기재의 권리범위 내에 있게 된다.

부호의 설명

[0077] 100: 악성 문서 파일 식별 시스템

110: 입력부

120: 분석부

121: 제 1 분석부

122: 제 2 분석부

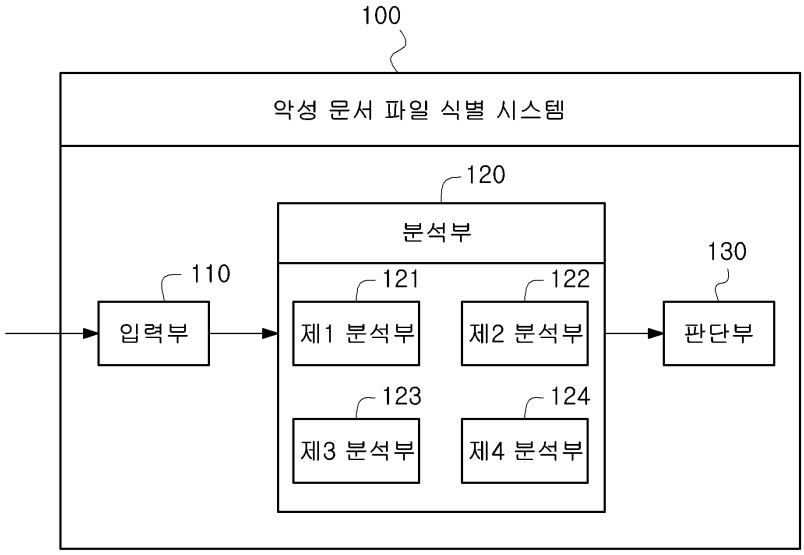
123: 제 3 분석부

124: 제 4 분석부

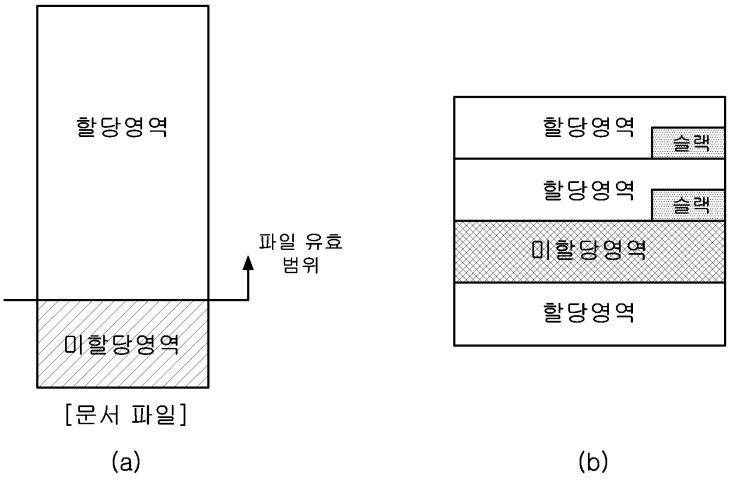
130: 판단부

도면

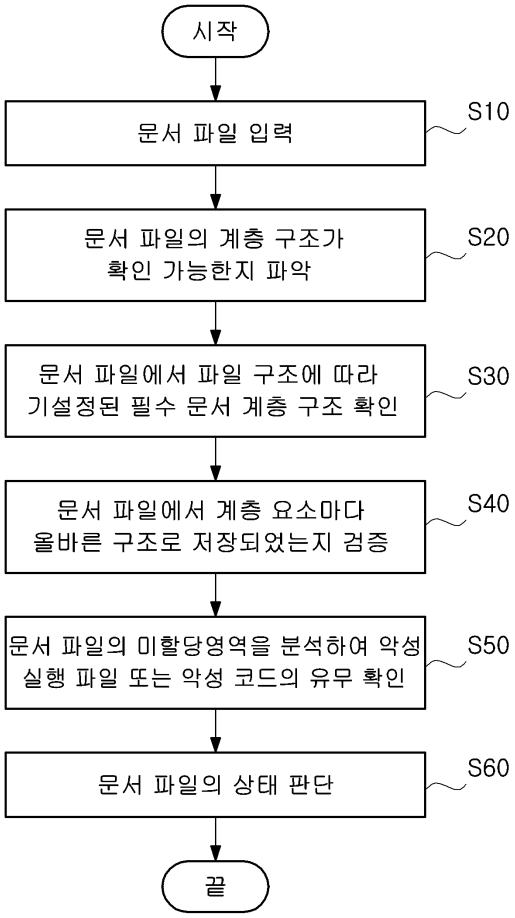
도면1


100
악성 문서 파일 식별 시스템
120
분석부
110
입력부
121
제1 분석부
122
제2 분석부
123
제3 분석부
124
제4 분석부
130
판단부

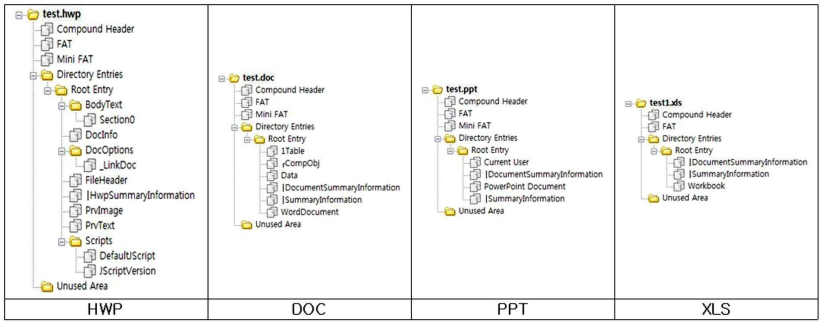
도면2

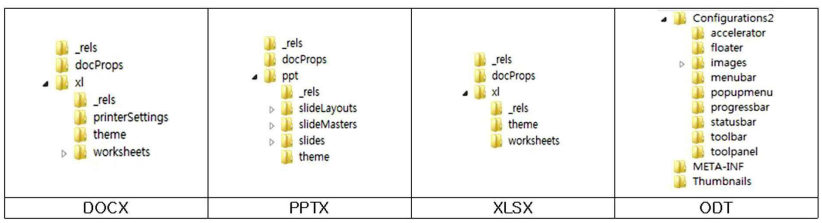

할당영역
파일 유효
범위
미할당영역
[문서 파일]
(a)
할당영역
슬랙
할당영역
슬랙
미할당영역
할당영역
(b)

도면3

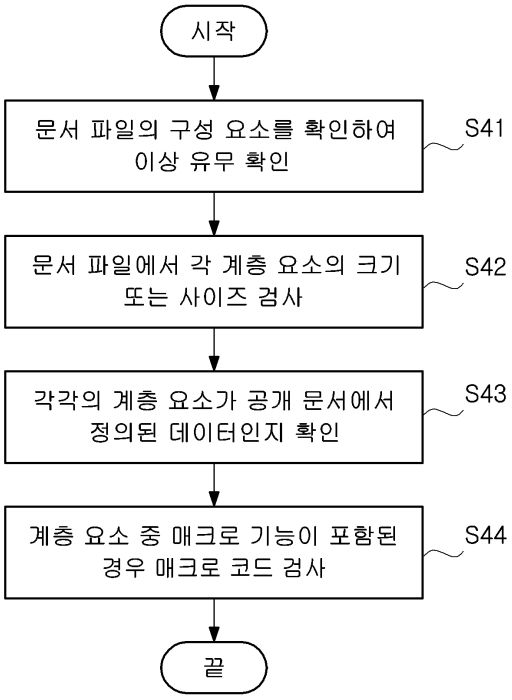

시작
문서 파일 입력
S10
문서 파일의 계층 구조가 확인 가능한지 파악
S20
문서 파일에서 파일 구조에 따라 기설정된 필수 문서 계층 구조 확인
S30
문서 파일에서 계층 요소마다 올바른 구조로 저장되었는지 검증
S40
문서 파일의 미할당영역을 분석하여 악성 실행 파일 또는 악성 코드의 유무 확인
S50
문서 파일의 상태 판단
S60
끝

도면4

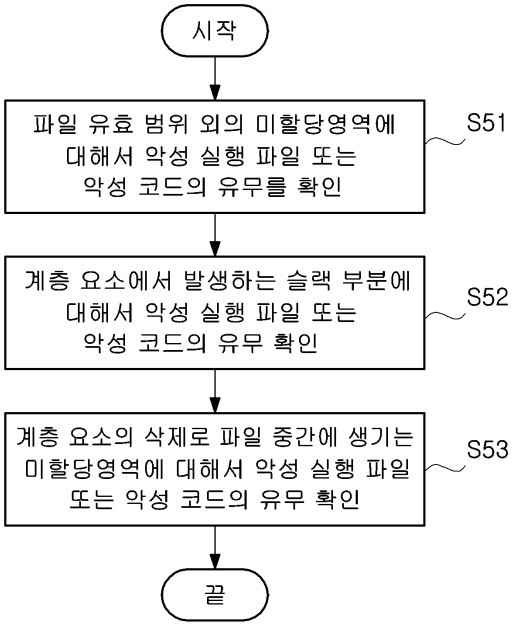

HWP
DOC
PPT
XLS


DOCX
PPTX
XLSX
ODT

도면5


시작
문서 파일의 구성 요소를 확인하여 이상 유무 확인
S41
문서 파일에서 각 계층 요소의 크기 또는 사이즈 검사
S42
각각의 계층 요소가 공개 문서에서 정의된 데이터인지 확인
S43
계층 요소 중 매크로 기능이 포함된 경우 매크로 코드 검사
S44
끝

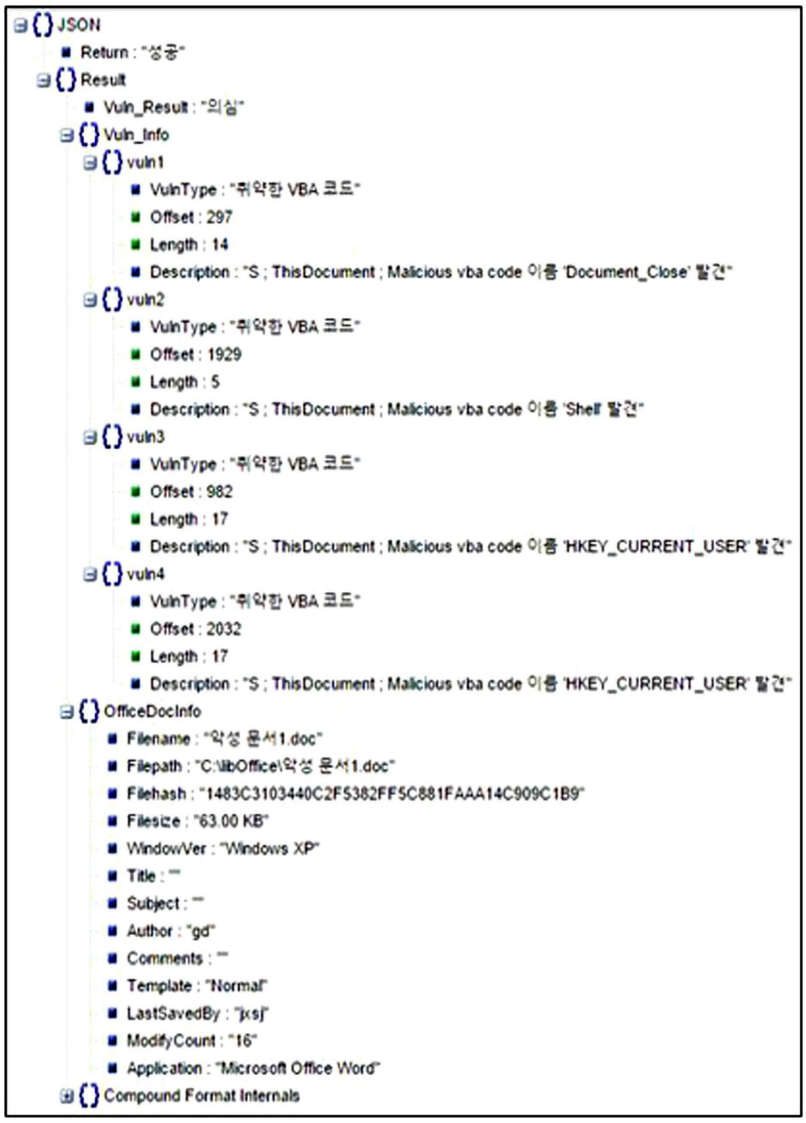
도면6


시작
파일 유효 범위 외의 미할당영역에 대해서 악성 실행 파일 또는 악성 코드의 유무를 확인
S51
계층 요소에서 발생하는 슬랙 부분에 대해서 악성 실행 파일 또는 악성 코드의 유무 확인
S52
계층 요소의 삭제로 파일 중간에 생기는 미할당영역에 대해서 악성 실행 파일 또는 악성 코드의 유무 확인
S53
끝

도면7

```
{
      “Return” : (성공/실패),
      “Result” : {
            “Vuln_Result”: (의심/정상/알수없음),
            “Vuln_Info”  : (결과 판단 요소),
            “OfficeDocInfo”: (문서관련 부가정보),
            “Compound Format Internals” : {
                  “Compound Header”   : (복합 파일 헤더 정보),
                  “Directory Entries” : (디렉터리 엔트리 정보)
            }
      }
}
```

도면8



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 제4항

【변경전】

상기 각 계층 요소

【변경후】

각 계층 요소

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 제10항

【변경전】

상기 각각의 계층 요소

【변경후】

상기 각 계층 요소

【직권보정 3】

【보정항목】 청구범위

【보정세부항목】 제10항

【변경전】

상기 각 계층 요소

【변경후】

각 계층 요소

【직권보정 4】

【보정항목】 청구범위

【보정세부항목】 제4항

【변경전】

상기 각각의 계층 요소

【변경후】

상기 각 계층 요소