(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년07월20일
(11) 등록번호 10-1641295
(24) 등록일자 2016년07월14일

(51) 국제특허분류(Int. Cl.)
G06F 21/56 (2013.01) *G06F 21/60* (2013.01)
(21) 출원번호 10-2014-0182578
(22) 출원일자 2014년12월17일
심사청구일자 2014년12월17일
(65) 공개번호 10-2016-0073801
(43) 공개일자 2016년06월27일
(56) 선행기술조사문헌
KR1020060067117 A*
KR1020130071617 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
주식회사 한글과컴퓨터
주식회사 한컴시큐어

(72) 발명자
이희조
권종훈
김종민

(74) 대리인
특허법인엠에이피에스

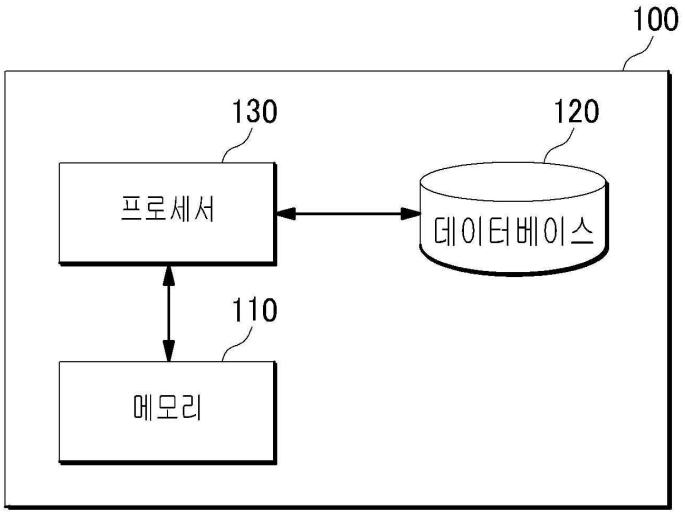
전체 청구항 수 : 총 7 항 심사관 : 구본재

(54) 발명의 명칭 **문서에 포함된 악성 공격 코드 탐지 시스템 및 방법**

(57) 요 약

본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드를 탐지하는 시스템은 악성 공격 코드를 탐지하는 프로그램이 저장된 메모리, 기 수집된 악성 공격 코드에 기초하여 분류된 행위 시그너처가 저장된 데이터베이스 및 프로그램을 실행시키는 프로세서를 포함하되, 프로세서는 상기 악성 공격 코드를 탐지하는 프로그램이 실행됨에 따라, 문서를 파싱하고, 파싱된 문서의 구조적 특징에 기초하여 악성 공격 코드를 추출하며, 추출된 악성 공격 코드에 기초하여 행위 시그너처를 추출하고, 데이터베이스에 저장된 행위 시그너처와 문서에서 추출된 행위 시그너처가 매칭되는 경우, 악성 공격 코드가 공격 행위를 하는 것으로 탐지한다.

대 표 도 - 도1



명 세 서

청구범위

청구항 1

문서에 포함된 악성 공격 코드를 탐지하는 시스템에 있어서,

상기 악성 공격 코드를 탐지하는 프로그램이 저장된 메모리,

기 수집된 악성 공격 코드에 기초하여 분류된 행위 시그너처가 저장된 데이터베이스 및

상기 프로그램을 실행시키는 프로세서를 포함하되,

상기 프로세서는 상기 악성 공격 코드를 탐지하는 프로그램이 실행됨에 따라,

상기 문서를 파싱하고, 상기 파싱된 문서의 구조적 특징에 기초하여 악성 공격 코드를 추출하며, 상기 추출된 악성 공격 코드에 기초하여 행위 시그너처를 추출하고,

상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처가 매칭되는 경우, 상기 악성 공격 코드가 공격 행위를 하는 것으로 탐지하며,

상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처가 매칭되지 않는 경우, 상기 데이터베이스에 상기 문서에서 추출된 행위 시그너처를 업데이트하는 것인 악성 공격 코드 탐지 시스템.

청구항 2

제 1 항에 있어서,

상기 구조적 특징은 상기 파싱된 문서에서 악성 공격 코드가 저장될 수 있는 구조체 및 상기 악성 공격 코드의 크기를 포함하는 악성 공격 코드 탐지 시스템.

청구항 3

제 1 항에 있어서,

상기 프로세서는 상기 파싱된 문서의 구조적 특징을 이용하여 탐색 범위를 줄이고, 상기 탐색 범위에서 악성 공격 코드로 의심되는 부분을 찾아 바이너리 코드화를 수행한 이후에 상기 악성 공격 코드를 추출하는 악성 공격 코드 탐지 시스템.

청구항 4

제 1 항에 있어서,

상기 악성 공격 코드는 상기 악성 공격 코드의 공격 방법을 정의하는 익스플로잇 코드 및 상기 악성 공격 코드의 공격 유형을 정의하는 쉘 코드를 포함하되,

상기 프로세서는 상기 익스플로잇 코드 및 상기 쉘 코드에 기초하여 상기 악성 공격 코드의 행위 시그너처를 분류하는 악성 공격 코드 탐지 시스템.

청구항 5

삭제

청구항 6

문서에 대한 악성 공격 코드 탐지 방법에 있어서,

데이터베이스에 기 수집된 악성 공격 코드에 기초하여 추출된 행위 정보를 저장하는 단계;

상기 문서에서 악성 공격 코드를 추출하는 단계;

상기 악성 공격 코드에 기초하여 행위 시그너처를 추출하는 단계;

상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처를 매칭하는 단계;

상기 매칭하는 단계를 통하여, 상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처가 매칭되면, 상기 악성 공격 코드가 공격 행위를 하는 것으로 탐지하는 단계; 및

상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처가 매칭되지 않으면, 상기 데이터베이스에 상기 문서에서 추출된 상기 행위 시그너처를 업데이트하는 단계를 포함하는 악성 공격 코드 탐지 방법.

청구항 7

제 6 항에 있어서,

상기 악성 공격 코드를 추출하는 단계는,

상기 문서를 파싱하는 단계; 및

상기 파싱된 문서의 구조적 특징에 기초하여 악성 공격 코드를 추출하는 단계를 포함하되,

상기 구조적 특징은 상기 파싱된 문서에서 악성 공격 코드가 저장될 수 있는 구조체 및 상기 악성 공격 코드의 크기를 포함하는 악성 공격 코드 탐지 방법.

청구항 8

제 6 항에 있어서,

상기 악성 공격 코드는 상기 악성 공격 코드의 위치 및 공격 방법을 정의하는 익스플로잇 코드 및

상기 악성 공격 코드의 공격 유형을 정의하는 쉘 코드를 포함하되,

상기 악성 공격 코드에 포함된 상기 익스플로잇 코드 및 상기 쉘 코드에 기초하여 상기 악성 공격 코드의 행위 시그너처를 분류하는 악성 공격 코드 탐지 방법.

청구항 9

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 문서에 포함된 악성 공격 코드 탐지 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 오피스 응용프로그램(office application)은 워드 프로세서(word processor), 스프레드 시트(spread sheet), 데이터베이스(database), 프리젠테이션(presentation) 등의 응용 소프트웨어의 모음이다. 예를 들어, "한글과 컴퓨터" 사의 "한컴오피스", "마이크로소프트(Microsoft)" 사의 "마이크로소프트 오피스(Microsoft office)", "Apple" 사의 "Works" 및 "Apache" 사의 "아파치 오픈 오피스(Apache Open Office)" 등이 있다.

[0003] 최근에는 이러한 오피스 응용프로그램에 의해 생성된 문서 파일에 악성 공격 코드를 삽입하고 정상 문서처럼 위장하여 실행을 유도하는 해킹 공격 방법이 전 세계적으로 늘어나고 있다. 특히 오피스 응용프로그램 파일은 개인뿐 아니라 기업, 관공서 및 군부대 등에서 널리 사용하고 있어 그 위험성이 커지고 있다.

[0004] 그러므로 오피스 응용프로그램을 사용하는 사용자들에게 안전한 오피스 환경을 보장하기 위한 보안 대응이 필요하다.

[0005] 오피스 응용프로그램에 삽입된 악성 공격 코드를 탐지하는 기술에 대한 종래 발명은 다음과 같다. 한국 공개특허공보 제10-2012-0130626호(발명의 명칭: "발명의 명칭 문서기반 악성코드 탐지 장치 및 방법")는 가상 머신에서 행위 분석 프로그램을 이용하여 문서의 행위 분석 결과를 검출하고, 검출된 결과를 이용하여 악성 공격 코드

를 판단하는 문서기반 악성 탐지 장치 및 방법을 개시하고 있다.

[0006] 또한, 한국 공개특허공보 제10-2013-0078960호(발명의 명칭: "오피스 프로그램의 취약점을 이용한 악성 공격 코드의 행위기반 진단 및 차단방법")는 오피스 프로그램의 취약점을 이용하여 문서에 삽입된 악성 공격 코드의 행위 기반의 진단 및 차단을 하는 방법을 개시하고 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 본 실시예는 문서에 포함된 악성 공격 코드를 추출하고 공격 행위를 탐지하는 시스템 및 방법을 제공하는데 그 목적이 있다.

[0008] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0009] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드를 탐지하는 시스템은 악성 공격 코드를 탐지하는 프로그램이 저장된 메모리, 기 수집된 악성 공격 코드에 기초하여 분류된 행위 시그너처가 저장된 데이터베이스 및 프로그램을 실행시키는 프로세서를 포함하되, 프로세서는 상기 악성 공격 코드를 탐지하는 프로그램이 실행됨에 따라, 문서를 파싱하고, 파싱된 문서의 구조적 특징에 기초하여 악성 공격 코드를 추출하며, 추출된 악성 공격 코드에 기초하여 행위 시그너처를 추출하고, 데이터베이스에 저장된 행위 시그너처와 문서에서 추출된 행위 시그너처가 매칭되는 경우, 악성 공격 코드가 공격 행위를 하는 것으로 탐지한다.

[0010] 또한, 본 발명의 일 실시예에 따른 문서에 대한 악성 공격 코드 탐지 방법은 데이터베이스에 기 수집된 악성 공격 코드에 기초하여 추출된 행위 정보를 저장하는 단계; 문서에서 악성 공격 코드를 추출하는 단계; 악성 공격 코드에 기초하여 행위 시그너처를 추출하는 단계; 및 상기 데이터베이스에 저장된 행위 시그너처와 상기 문서에서 추출된 행위 시그너처를 매칭하여 상기 악성 공격 코드가 공격 행위를 하는 것으로 탐지하는 단계를 포함한다.

발명의 효과

[0011] 전술한 과제 해결 수단 중 어느 하나에 의하면, 본 발명의 일 실시예는 다수의 악성 공격 코드에 기초하여 생성된 적은 수의 행위 시그너처 만을 이용하여 악성 공격 코드의 공격 행위를 탐지할 수 있으므로 악성 공격 코드를 탐지하는데 소요되는 시간 및 자원을 대폭 감소시킬 수 있다. 또한, 행위 시그너처를 이용하므로 향후 발생할 수 있는 유사 행위를 수행하는 변종 악성 공격 코드를 사전에 탐지하고 공격을 예방할 수 있다.

[0012] 즉, 적은 시간과 자원을 이용하여 문서에 포함된 다양한 변종의 악성 공격 코드를 사전에 탐지하므로 사용자의 개인 컴퓨터의 악성 공격 코드의 감염을 예방할 수 있고 문서에 대한 안정성 및 신뢰성을 증대시킬 수 있다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드 탐지 시스템을 개략적으로 도시한 구성도이다.

도 2는 본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드를 개략적으로 도시한 블록도이다.

도 3은 본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드 탐지 시스템의 악성 공격 코드를 탐지하는 프로그램을 개략적으로 도시한 구성도이다.

도 4는 본 발명의 일 실시예에 따른 문서에 대한 악성 공격 코드 탐지 방법을 설명하기 위한 순서도이다.

도 5는 본 발명의 일 실시예에 따른 문서에 대한 악성 공격 코드 탐지 방법에서 악성 공격 코드 추출 방법을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할

수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였다.

[0015] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0016] 이하에서는 도 1 내지 도 3을 참조하여 본 발명의 일 실시예에 따른 문서(200)에 포함된 악성 공격 코드 탐지 시스템(100)을 설명하도록 한다.

[0017] 도 1은 본 발명의 일 실시예에 따른 문서(200)에 포함된 악성 공격 코드 탐지 시스템(100)을 개략적으로 도시한 구성도이다.

[0018] 악성 공격 코드 탐지 시스템(100)은 문서(200) 내에서 추출한 행위 시그너처를 이용하여 악성 공격 코드를 탐지하고 악성 공격 코드의 행위 정보를 제공한다. 이때, 악성 공격 코드 탐지 시스템(100)은 기존의 오피스 프로그램과 연동하거나, 독립적인 형태(stand only)로 실행될 수 있다.

[0019] 예를 들어, 기존의 오피스 프로그램과 연동한 악성 공격 코드 탐지 시스템(100)은 사용자가 오피스 프로그램을 이용하여 문서(200)를 편집하기 위하여 오피스 프로그램을 실행하고 문서(200)를 로드할 때, 악성 공격 코드 탐지를 수행할 수 있다. 이때, 문서(200)에 악성 공격 코드가 포함된 경우, 오피스 프로그램의 실행을 중단하고 사용자에게 악성 공격 코드의 행위 정보 및 위험성을 통보할 수 있다.

[0020] 그리고 악성 공격 코드 탐지 시스템(100)은 메일 서버(mail server) 및 사용자의 메일 시스템(mail system)과 연동되어 메일에 첨부된 문서(200) 파일을 대상으로 악성 공격 코드의 유무를 탐지할 수 있다. 만약 메일에 첨부된 문서(200) 파일에서 악성 공격 코드가 탐지되는 경우, 사용자이나 메일 서버 관리자에게 악성 공격 코드를 포함하는 문서(200) 파일에 대한 정보 및 해당 문서(200) 파일에 포함된 악성 공격 코드의 행위 정보를 통보하고, 사전에 차단할 수 있도록 할 수 있다.

[0021] 또한, 독립적인 형태로 실행되는 악성 공격 코드 탐지 시스템(100)은 사용자의 컴퓨터 내에 포함된 복수개의 문서(200) 파일을 대상으로 악성 공격 코드의 유무를 탐지하고 사용자에게 악성 공격 코드를 포함하는 문서(200) 파일에 대한 정보 및 해당 문서(200) 파일에 포함된 악성 공격 코드의 행위 정보를 통보할 수 있으나 이에 한정된 것은 아니다.

[0022] 본 발명의 일 실시예에 따른 악성 공격 코드 탐지 시스템(100)은 메모리(110), 데이터베이스(120) 및 프로세서(130)를 포함한다.

[0023] 이때, 메모리(110)에는 악성 공격 코드를 탐지하는 프로그램(150)이 저장된다. 여기에서 메모리(110)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장 장치를 통칭하는 것이다. 예를 들어, 메모리(110)는 콤팩트 플래시(compact flash; CF) 카드, SD(secure digital) 카드, 메모리 스틱(memory stick), 솔리드 스테이트 드라이브(solid-state drive; SSD) 및 마이크로(micro) SD 카드 등과 같은 낸드 플래시 메모리(NAND flash memory), 하드 디스크 드라이브(hard disk drive; HDD) 등과 같은 마그네틱 컴퓨터 기억 장치 및 CD-ROM, DVD-ROM 등과 같은 광학 디스크 드라이브(optical disc drive) 등을 포함할 수 있다.

[0024] 또한, 메모리(110)에 저장된 악성 공격 코드를 탐지하는 프로그램은 소프트웨어 또는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)와 같은 하드웨어 형태로 구현될 수 있으며, 소정의 역할들을 수행할 수 있다.

[0025] 그렇지만 '구성 요소들'은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, 각 구성 요소는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다.

[0026] 따라서, 일 예로서 구성 요소는 소프트웨어 구성 요소들, 객체지향 소프트웨어 구성 요소들, 클래스 구성 요소들 및 태스크 구성 요소들과 같은 구성 요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다.

[0027] 구성 요소들과 해당 구성 요소들 안에서 제공되는 기능은 더 작은 수의 구성 요소들로 결합되거나 추가적인 구

성 요소들로 더 분리될 수 있다.

[0028] 데이터베이스(120)에는 기 수집된 악성 공격 코드에 기초하여 분류된 행위 시그너처(semantic signature)가 저장된다. 이때, 데이터베이스(120)는 악성 공격 코드 탐지 시스템(100)에 서버 프로그램(server program) 형태로 실행되거나, 악성 공격 코드 탐지 시스템(100)과 네트워크로 연결된 독립적인 데이터베이스 서버(database server)일 수 있다.

[0029] 그리고 데이터베이스(120)는 캐쉬, ROM(Read Only Memory), PROM(Programmable ROM), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 및 플래쉬 메모리(Flash memory)와 같은 비휘발성 메모리 소자 또는 RAM(Random Access Memory)과 같은 휘발성 메모리 소자 또는 하드디스크 드라이브(HDD, Hard Disk Drive), CD-ROM과 같은 저장 매체 중 적어도 하나로 구현될 수 있으나 이에 한정되지는 않는다.

[0030] 또한, 데이터베이스(120)에 기 수집된 악성 공격 코드는 알려진 악성 공격 코드 샘플이거나, 악성 공격 코드 탐지 시스템(100)을 수행하며 수집된 악성 공격 코드일 수 있으나, 이에 한정된 것은 아니다.

[0031] 본 발명의 일 실시예에 따른 프로세서(130)는 메모리(110)에 저장된 악성 공격 코드를 탐지하는 프로그램(150)이 실행됨에 따라, 문서(200)를 파싱하고, 파싱된 문서(200)의 구조적 특징에 기초하여 악성 공격 코드를 추출한다.

[0032] 이때, 프로세서(130)는 문서(200)를 파싱하기 위하여 문서(200)의 종류에 따라 복호화하고 악성 공격 코드가 의심되는 파일 내의 오프셋(offset)을 파악할 수 있다. 복호화 방법은 각각 문서(200)의 종류에 따라 상이할 수 있다. 예를 들어, 입력되는 문서(200)가 "한글과컴퓨터"의 "한글워드프로세서"를 이용하여 작성된 "한글 문서 파일"(파일 확장자: "hwp")의 경우에는 공개된 한글 워드프로세서 포맷에 기반하여 복호화를 수행할 수 있다. 또한, 입력되는 문서(200)가 "마이크로소프트 오피스"를 이용하여 작성된 "워드 파일"(파일 확장자: "doc" 또는 "docx") 및 "엑셀 파일"(파일 확장자: "xls" 또는 "xlsx") 등의 "마이크로 소프트 오피스 문서 파일"인 경우에는 "오피스 API(application programming interface)"를 이용하여 복호화를 수행할 수 있으나, 이에 한정된 것은 아니다.

[0033] 그리고 프로세서(130)는 악성 공격 코드가 의심되는 파일 오프셋에 기반하여 악성 공격 코드가 시작되는 앵커포인터(anchor point)를 찾아 바이너리 코드화를 수행한다. 이때, 바이너리 코드화는 프로그램이나 파일의 구조적인 부분을 분석하여 바이너리 코드나 데이터를 추출하는 역공학(reverse engineering)을 이용할 수 있다.

[0034] 이때, 앵커포인터는 악성 공격 코드에 따라 다양하게 정의될 수 있다. 예를 들어, 앵커포인터는 "0x0c", "0x0d"와 같은 힙 오버플로어(heap overflow)가 가능한 1바이트(byte) 기계어로 구성될 수 있으며, "PUSH"나 "NOP 코드"와 같은 명령어 세트로 구성될 수 있다.

[0035] 이렇게 문서(200)를 파싱하고 파싱된 문서의 구조적 특징을 이용하여 탐색 범위를 줄인 후에 악성 공격 코드로 의심되는 부분을 찾아 바이너리 코드화를 거친 후, 악성 공격 코드에 기초하여 행위 시그너처를 추출한다.

[0036] 이때, 행위 시그너처는 악성 공격 코드의 위치 정보, 구조 정보 및 행위 정보 등을 포함할 수 있다. 위치 정보는 문서(200) 내의 악성 공격 코드가 존재하는 위치에 대한 정보이며, 구조 정보는 악성 공격 코드의 구조적 특징에 대한 정보일 수 있다. 그리고 행위 정보는 문서(200) 내의 악성 공격 코드가 포함하고 있는 악성 행위 정보이며, 이 악성 행위 정보는 악성 공격 코드가 실행되었을 때 발생하는 사용자 컴퓨터의 이벤트(event) 일 수 있다.

[0037] 예를 들어, 위치 정보는 "헤더 영역", "본문 영역", "이미지" 및 "스크립트" 등이 될 수 있고, 구조 정보는 "일반적인 구조", "암호화 루틴이 포함된 구조" 및 "체인 형태로 실행되는 구조" 등이 될 수 있다. 그리고 행위 정보는 "파일 실행", "라이브러리 호출" 및 "레지스트리 접근" 등이 될 수 있다.

[0038] 행위 시그너처는 위치 정보, 구조 정보 및 행위 정보를 코드화하여 표현(representation)할 수 있다. 예를 들어, 위치 정보는 "헤더 영역"의 코드를 1, "본문 영역"의 코드를 2로 설정될 수 있다. 또한, 구조 정보는 "알 수 없음"의 코드를 0, "일반적인 구조"의 코드를 1, "암호화 루틴이 포함된 구조"의 코드를 2로 설정할 수 있다. 그리고 행위 정보는 "알 수 없음"의 코드를 0, "파일 실행"의 코드를 1로 설정할 수 있다.

[0039] 이때, 행위 시그너처는 위치 정보, 구조 정보 및 행위 정보 코드에 기초하여 3자리의 코드를 생성할 수 있다. 그러므로 위치 정보가 "본문 영역", 구조 정보가 "일반적인 구조"이며, 행위 정보가 "파일 실행"인 악성 공격 코드의 행위 시그너처는 "211"과 같이 코드화 되어 표현될 수 있다. 또한, 위치 정보가 "헤더 영역", 구조 정

보가 "암호화 루틴이 포함된 구조"이며, 행위 정보가 "알 수 없음"인 악성 공격 코드의 행위 시그너처는 "120"과 같이 코드화되어 표현될 수 있다.

[0040] 한편, 프로세서(130)는 행위 시그너처를 추출한 다음, 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처의 매칭 여부를 판단한다. 이때, 프로세서(130)은 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처가 매칭되는 경우, 문서(200)에 포함된 악성 공격 코드가 공격 행위를 하는 것으로 탐지한다.

[0041] 한편, 파싱된 문서(200)에서 악성 공격 코드를 추출하기 위하여 사용되는 구조적 특징은 파싱된 문서(200)에서 악성 공격 코드가 저장될 수 있는 구조체 및 악성 공격 코드의 크기를 포함할 수 있다. 이때, 악성 공격 코드가 저장될 수 있는 구조체는 문서의 포맷(format)에 따른 스트림(stream) 구조체를 의미하며, 태그(tag) 등을 포함할 수 있다. 예를 들어, 구조체는 문서의 내용, 문서의 형식 및 문장의 형태 등을 포함할 수 있다.

[0042] 그러므로 프로세서(130)는 파싱된 문서(200)에서 특정한 구조체의 길이가 악성 공격 코드를 구성하는 최소 크기 이상인지 판별한 후, 악성 공격 코드의 최소 길이 이상인 경우, 의심 구조체로 분류할 수 있다. 의심 구조체로 분류한 다음 프로세서(130)은 특정한 구조체를 악성 공격 코드 탐색 영역에 포함되는 것으로 탐색할 수 있다.

[0043] 또한, 프로세서(130)는 역어셈블(disassemble)에 기반하여 파싱된 문서(200)에서 악성 공격 코드가 시작되는 것으로 의심되는 앵커 포인터를 찾고 어셈블리어로 변환하여 바이너리 코드화를 수행할 수 있다. 역어셈블은 역공학 방법의 하나로 실행 파일이나 문서 파일을 어셈블리어(assembly language)로 변환하는 것이다.

[0044] 그리고 프로세서(130)는 어셈블리어로 변환된 문서(200)에 포함된 명령어(instruction) 및 인자(operand) 등의 구조적 특징에 기초하여 악성 공격 코드를 추출할 수 있다.

[0045] 도 2는 본 발명의 일 실시예에 따른 악성 공격 코드를 개략적으로 도시한 블록도이다.

[0046] 도 2를 참조하면, 문서(200)에서 추출되는 악성 공격 코드는 악성 공격 코드의 공격 방법을 정의하는 익스플로잇 코드(exploit code) 및 악성 공격 코드의 공격 유형을 정의하는 쉘 코드(shell code)를 포함할 수 있다.

[0047] 그러므로 프로세서(130)는 문서(200)에서 추출되는 악성 공격 코드의 익스플로잇 코드 및 쉘 코드에 기초하여 악성 공격 코드의 행위 시그너처를 분류할 수 있다. 예를 들어, 프로세서(130)는 악성 공격 코드의 익스플로잇 코드 및 쉘 코드를 분석하여 공격 행위를 추출하여 행위 시그너처를 분류할 수 있다.

[0048] 한편, 문서(200)에서 추출된 악성 공격 코드의 행위 시그너처가 데이터베이스(120)에 저장된 행위 시그너처와 매칭되지 않을 경우, 프로세서(130)은 문서(200)에서 추출된 행위 시그너처를 데이터베이스(120)에 업데이트할 수 있다. 이렇게 데이터베이스(120)에 업데이트된 행위 시그너처는 앞으로 다른 문서(200)의 악성 공격 코드 탐지 시 활용될 수 있다.

[0049] 다음은 도 3을 참조하여 본 발명의 일 실시예에 따른 문서에 포함된 악성 공격 코드 탐지 시스템(100)을 상세하게 설명한다.

[0050] 도 3은 본 발명의 일 실시예에 따른 문서(200)에 포함된 악성 공격 코드 탐지 시스템(100)의 악성 공격 코드를 탐지하는 프로그램(150)을 개략적으로 도시한 구성도이다.

[0051] 도 3을 참조하면 악성 공격 코드 탐지 시스템(100)의 악성 공격 코드 탐지 프로그램(150)은 문서 파싱 모듈(151) 및 악성 공격 코드 탐지 모듈(152)을 포함하며, 문서(200), 문서 파일 뷰어(300) 및 데이터베이스(120)와 연결될 수 있다.

[0052] 악성 공격 코드 탐지 프로그램(150)은 하나 이상의 문서(200)가 입력되면 문서(200)의 악성 공격 코드를 추출하기 위하여 문서(200)를 문서 파싱 모듈(151)에 전달할 수 있다. 그리고 문서(200)를 수신한 문서 파싱 모듈(151)은 문서(200)의 파싱을 수행할 수 있다. 이때, 악성 공격 코드 탐지 시스템(100)으로 입력되는 하나 이상의 문서(200)는 모두 동일한 종류이거나, 각각 상이한 종류일 수 있다.

[0053] 또한, 문서 파싱 모듈(151)은 문서(200)의 종류에 따라 악성 공격 코드를 탐지하기 위하여 문서(200)의 복호화 라이브러리를 포함할 수 있다. 이때, 복호화 라이브러리는 각각 문서(200)의 종류에 따라 상이할 수도 있다. 예를 들어, 입력되는 문서(200)가 "한글과컴퓨터"의 "한글워드프로세서"를 이용하여 작성된 "한글 문서 파일(확장자: "hwp")"의 경우에는 복호화 라이브러리로 공개된 한글 워드프로세서 포맷에 기반하여 복호화를 수행할 수 있다. 또한, 입력되는 문서(200)가 "마이크로소프트 오피스"를 이용하여 작성된 "워드 파일(확장자: "doc" 또는 "docx")" 및 "엑셀 파일(확장자: "xls" 또는 "xlsx")" 등의 "마이크로 소프트 오피스 문서 파일"인 경우에는

복호화 라이브러리로 "오피스 API(application programming interface)"를 이용하여 복호화를 수행할 수 있으나, 이에 한정된 것은 아니다.

[0054] 그리고 문서 파싱 모듈(151)은 문서의 파싱을 수행한 다음 파싱된 문서를 악성 공격 코드 탐지 모듈(152)에 전달할 수 있다. 악성 공격 코드 탐지 모듈(152)은 파싱된 문서(200)에서 악성 공격 코드가 시작될 것으로 의심되는 파일 내의 오프셋에 기반하여 악성 공격 코드가 시작되는 앵커포인터를 찾아 바이너리 코드화를 수행할 수 있다.

[0055] 바이너리 코드화를 수행한 다음 악성 공격 코드 탐지 모듈(152)은 문서(200)의 구조적 특징을 분석하여, 악성 공격 코드를 추출할 수 있다. 이때, 구조적 특징은 파싱된 문서(200)에서 악성 공격 코드가 저장될 수 있는 구조체 및 악성 공격 코드의 크기를 포함할 수 있다.

[0056] 또한, 악성 공격 코드 탐지 모듈(152)은 추출된 악성 공격 코드에 기초하여 행위 시그너처를 추출할 수 있다. 이때, 행위 시그너처는 악성 공격 코드의 위치 정보 및 행위 정보 등을 포함할 수 있다.

[0057] 악성 공격 코드 탐지 모듈(152)은 행위 시그너처를 추출하고 문서(200)에서 추출된 행위 시그너처와 데이터베이스(120)에 저장된 행위 시그너처의 매칭 여부를 판단할 수 있다. 이때, 데이터베이스(120)에 저장된 행위 시그너처는 악성 공격 코드 탐지 시스템(100)에 의하여 기 수집된 악성 공격 코드에서 추출한 행위 시그너처이거나 악성 공격 코드 탐지 시스템(100) 구축을 위하여 수집된 악성 공격 코드 샘플에서 추출된 행위 시그너처일 수 있다.

[0058] 프로세서(130)는 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처가 매칭되는 경우, 악성 공격 코드 탐지 모듈(152)은 문서(200)에 포함된 악성 공격 코드가 공격 행위를 하는 것으로 탐지하고 문서 파일 뷰어(300)에 탐지 결과 및 문서 내용을 전달할 수 있다. 이때, 탐지 결과는 악성 공격 코드의 행위 시그너처에 포함된 위치 정보 및 행위 정보 등이 포함될 수 있다.

[0059] 이와 반대로 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처가 매칭되지 않는 경우, 악성 공격 코드 탐지 모듈(152)은 데이터베이스(120)에 행위 시그너처를 업데이트하고, 문서 파일 뷰어(300)에 알려지지 않은 악성 공격 코드가 탐지되었음을 통보하는 탐지결과 및 문서의 내용을 전달할 수 있다.

[0060] 다음은 도 4 및 도 5를 이용하여 본 발명의 일 실시예에 따른 문서(200)에 대한 악성 공격 코드 탐지 방법을 설명한다.

[0061] 도 4는 본 발명의 일 실시예에 따른 문서(200)에 대한 악성 공격 코드 탐지 방법을 설명하기 위한 순서도이다.

[0062] 도 4에 도시된 것처럼, 문서(200)에 대한 악성 공격 코드 탐지 방법은 기 수집된 악성 공격 코드에 기초하여 추출된 행위 정보를 데이터베이스(120)에 저장한다(S400). 이때, 기 수집된 악성 공격 코드는 악성 공격 코드 탐지 시스템(100)에 의하여 기 수집된 악성 공격 코드이거나 악성 공격 코드 탐지 시스템(100)을 구축하기 위하여 수집된 악성 공격 코드 샘플일 수 있다.

[0063] 악성 공격 코드 탐지 시스템(100)에 문서(200)가 입력되면, 악성 공격 코드 탐지 방법은 문서(200)에서 악성 공격 코드를 추출하고(S410), 악성 공격 코드에 기초하여 행위 시그너처를 추출한다(S420). 이때, 행위 시그너처는 악성 공격 코드의 위치 정보 및 행위 정보 등을 포함할 수 있다. 위치 정보는 문서(200) 내의 악성 공격 코드가 존재하는 위치에 대한 정보이며, 행위 정보는 문서(200) 내의 악성 공격 코드가 포함하고 있는 악성 행위 정보가 될 수 있다.

[0064] 또한, 악성 공격 코드 탐지 방법은 문서(200)에서 행위 시그너처를 추출한 다음, 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처를 매칭한다(S440). 매칭 여부를 판단하여(S440) 문서(200)의 행위 시그너처와 데이터베이스(120)에 저장된 행위 시그너처가 매칭되면 악성 공격 코드가 공격 행위를 하는 것으로 탐지한다(S450).

[0065] 한편, 악성 공격 코드 탐지 방법은 데이터베이스(120)에 저장된 행위 시그너처와 문서(200)에서 추출된 행위 시그너처를 매칭하고(S440) 매칭되지 않으면, 데이터베이스(120)에 문서(200)에서 추출된 행위 시그너처를 업데이트할 수 있다(S460). 이렇게 업데이트된 행위 시그너처는 다른 문서(200)의 악성 공격 코드 탐지를 위하여 활용될 수 있다.

[0066] 도 5는 본 발명의 일 실시예에 따른 문서(200)에 대한 악성 공격 코드 탐지 방법에서 악성 공격 코드 추출 방법을 설명하기 위한 순서도이다.

[0067] 악성 공격 코드 탐지 방법은 악성 공격 코드를 추출하기 위하여(S410) 문서(200)를 파싱하고(S411), 파싱된 문서(200)의 구조적 특징에 기초하여 악성 공격 코드를 추출할 수 있다(S412). 이때, 구조적 특징은 파싱된 문서(200)에서 악성 공격 코드가 저장될 수 있는 구조체 및 악성 공격 코드의 크기를 포함할 수 있다. 특히, 악성 공격 코드가 저장될 수 있는 구조체는 문서(200)의 포맷에 따른 스트림 구조체를 의미하며, 태그(tag) 등을 포함할 수 있다. 예를 들어, 구조체는 문서의 내용, 문서의 형식 및 문장의 형태 등을 포함할 수 있다.

[0068] 그러므로 악성 공격 코드 탐지 방법은 파싱된 문서(200)에서 특정한 구조체의 길이가 악성 공격 코드를 구성하는 최소 크기 이상인지 판별한 후, 악성 공격 코드의 최소 길이 이상인 경우, 의심 구조체로 분류할 수 있다. 의심 구조체로 분류한 다음 악성 공격 코드 탐지 방법은 특정한 구조체를 악성 공격 코드 탐색 영역에 포함되는 것으로 탐색할 수 있다. 또한, 악성 공격 코드는 악성 공격 코드의 위치 및 공격 방법을 정의하는 익스플로잇 코드 및 악성 공격 코드의 공격 유형을 정의하는 쉘 코드를 포함할 수 있다. 이때, 악성 공격 코드에 포함된 익스플로잇 코드 및 쉘 코드에 기초하여 공격 코드의 행위 시그너처를 추출할 수 있다.

[0069] 본 발명의 일 실시예에 따른 악성 공격 코드 탐지 시스템(100) 및 방법은 다수의 악성 공격 코드에 기초하여 생성된 적은 수의 행위 시그너처 만을 이용하여 악성 공격 코드의 공격 행위를 탐지할 수 있으므로 악성 공격 코드를 탐지하는데 소요되는 시간 및 자원을 대폭 감소시킬 수 있다. 또한, 행위 시그너처를 이용하므로 향후 발생할 수 있는 유사 행위를 수행하는 변종 악성 공격 코드를 사전에 탐지하고 공격을 예방할 수 있다.

[0070] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 메커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.

[0071] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.

[0072] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0073] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

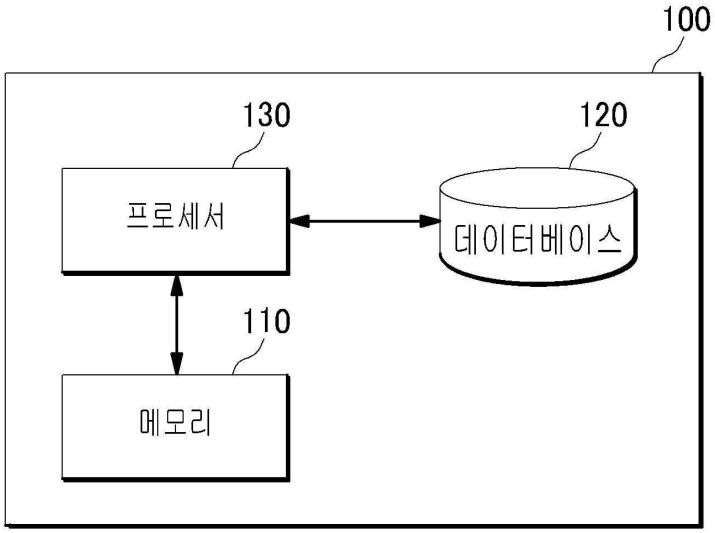
부호의 설명

[0074] 100: 악성 공격 탐지 시스템 110: 스토리지 장치

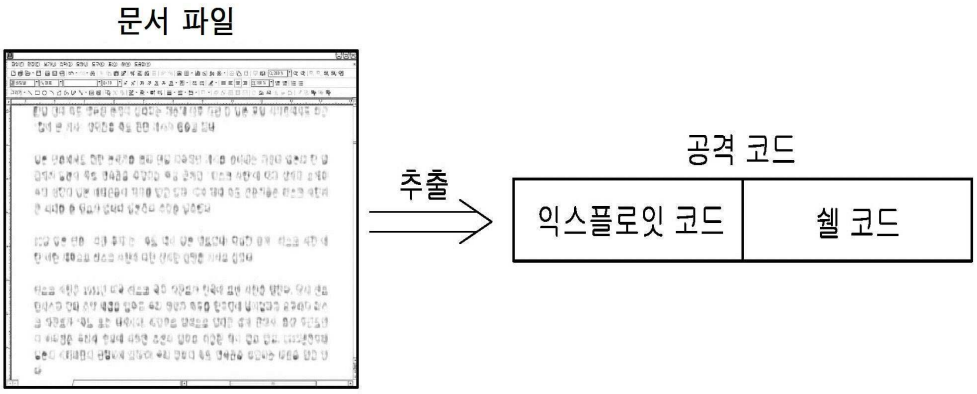
120: 데이터베이스 130: 프로세서

도면

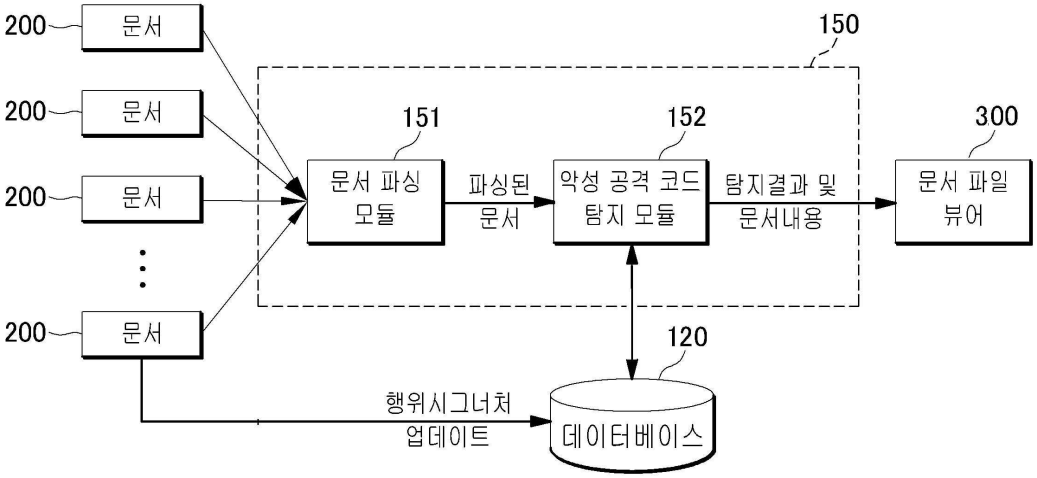
도면1



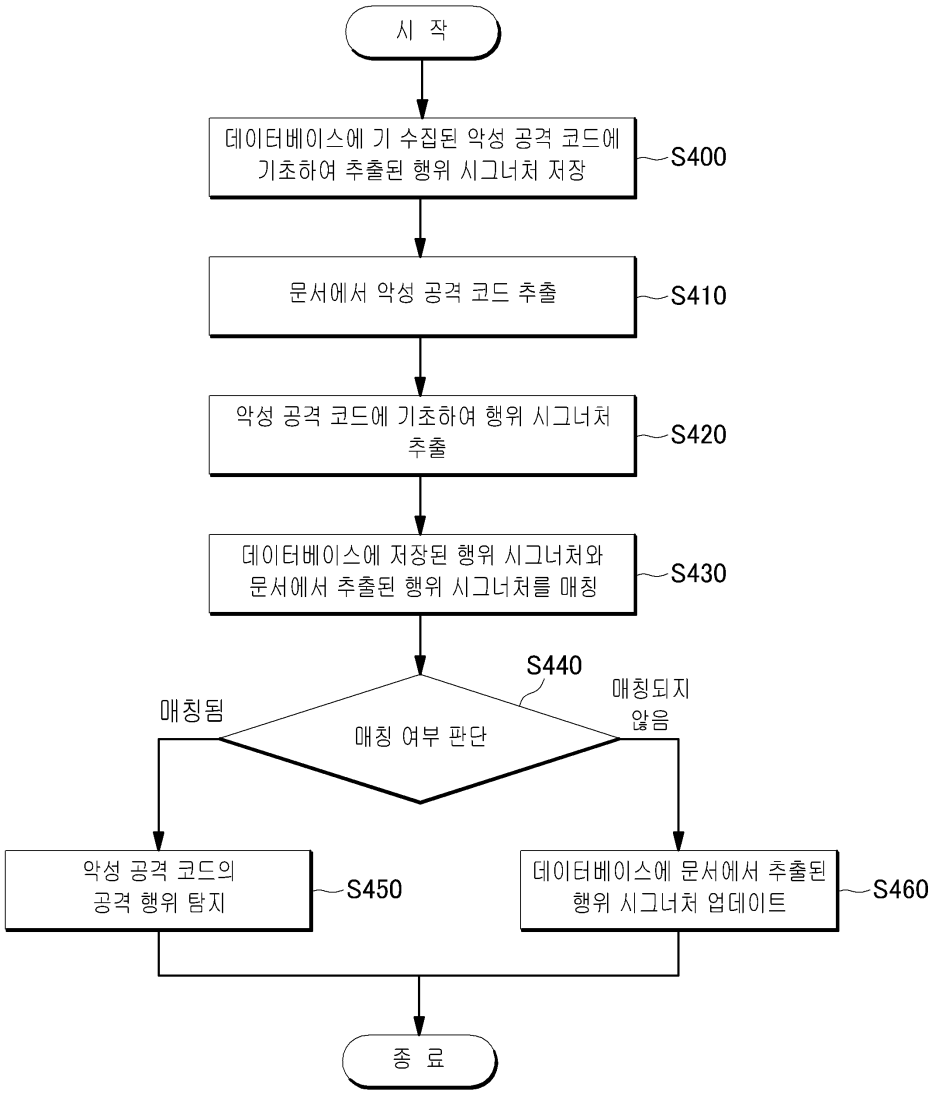
도면2



도면3


200
문서
200
문서
200
문서
200
문서
150
151
문서 파싱
모듈
파싱된
문서
152
악성 공격 코드
탐지 모듈
탐지결과 및
문서내용
300
문서 파일
뷰어
120
행위시그너처
업데이트
데이터베이스

도면4


시 작
데이터베이스에 기 수집된 악성 공격 코드에
기초하여 추출된 행위 시그너처 저장
S400
문서에서 악성 공격 코드 추출
S410
악성 공격 코드에 기초하여 행위 시그너처
추출
S420
데이터베이스에 저장된 행위 시그너처와
문서에서 추출된 행위 시그너처를 매칭
S430
S440
매칭 여부 판단
매칭됨
매칭되지
않음
악성 공격 코드의
공격 행위 탐지
S450
데이터베이스에 문서에서 추출된
행위 시그너처 업데이트
S460
종 료

도면5