(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2015년05월20일
(11) 등록번호 10-1520671
(24) 등록일자 2015년05월11일

(51) 국제특허분류(Int. Cl.)
*G06F 21/56* (2013.01) *G06F 17/00* (2006.01)
(21) 출원번호 10-2014-0050184
(22) 출원일자 2014년04월25일
심사청구일자 2014년04월25일
(56) 선행기술조사문헌
KR1020130071617 A*
KR1020090093187 A
KR1020090000216 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
**충남대학교산학협력단**
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
**조제경**
대전광역시 유성구 온천북로33번길 22-9 304호
**류재철**
대전 유성구 어은로 57, 132동 801호 (어은동, 한빛아파트)
(74) 대리인
**권혁수, 송윤호**

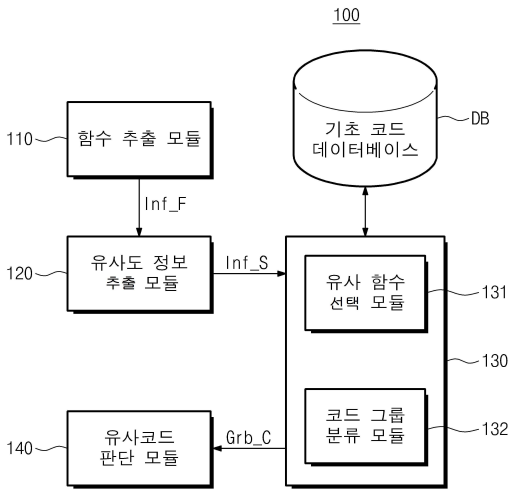
전체 청구항 수 : 총 11 항 심사관 : 구본재

(54) 발명의 명칭 **실행코드 유사도 분석 시스템 및 방법**

## (57) 요 약

본 발명의 실시 예에 따른 실행코드 유사도 분석 방법은, 분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 단계, 상기 추출된 대상 함수들 각각에 대한 유사도 정보를 획득하는 단계, 상기 획득된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수들과 유사한 기초 함수들을 선택하는 단계, 상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 단계, 그리고 상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 대상 실행코드와 유사한 기초 실행코드를 결정하는 단계를 포함한다.

*대 표 도* - 도1



이 발명을 지원한 국가연구개발사업
과제고유번호 2014-H0301-14-1010 / 1415128814
부처명 미래창조과학부
연구관리전문기관 정보통신산업진흥원
연구사업명 정보통신기술인력양성
연구과제명 클라우드 환경의 스마트 기기와 서비스 보안 기술개발 및 연구 인력양성
기 여 율 1/2
주관기관 숭실대학교산학협력단
연구기간 2014.01.01 ~ 2014.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 2010-0020726
부처명 미래창조과학부
연구관리전문기관 한국연구재단
연구사업명 차세대정보·컴퓨팅기술개발
연구과제명 개인 프라이버시 보장 원격 접근 기술
기 여 율 1/2
주관기관 고려대학교
연구기간 2013.04.01 ~ 2014.06.30

# 명 세 서

## 청구범위

### 청구항 1

정보 처리 장치가 데이터베이스에 저장된 기초 실행코드별 기초 함수 정보와 상기 기초 함수 정보에 대한 기초 유사도 정보를 이용하여 실행코드의 유사도를 분석하는 방법에 있어서,

분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 단계;

상기 추출된 대상 함수들 각각에 대한 유사도 정보를 획득하는 단계;

상기 획득된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수들과 유사한 기초 함수들을 선택하는 단계;

상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 단계; 그리고

상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 분석 대상 실행코드와 유사한 기초 실행코드를 결정하는 단계를 포함하며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 해쉬값 및 상기 추출된 각 대상 함수의 명령코드들에 대한 통계 정보를 포함하는 방법.

### 청구항 2

삭제

### 청구항 3

정보 처리 장치가 데이터베이스에 저장된 기초 실행코드별 기초 함수 정보와 상기 기초 함수 정보에 대한 기초 유사도 정보를 이용하여 실행코드의 유사도를 분석하는 방법에 있어서,

분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 단계;

상기 추출된 대상 함수들 각각에 대한 유사도 정보를 획득하는 단계;

상기 획득된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수들과 유사한 기초 함수들을 선택하는 단계;

상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 단계; 그리고

상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 분석 대상 실행코드와 유사한 기초 실행코드를 결정하는 단계를 포함하며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 해쉬값이며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 명령코드들에 대한 통계 정보, 상기 추출된 각 함수의 변수의 유형별 통계 정보, 상기 추출된 각 대상 함수의 변수 중 레지스터를 제외한 변수들을 삭제한 함수의 해쉬값, 그리고 상기 추출된 각 대상 함수에 포함된 분기문을 경계로 복수의 코드 블록들을 추출하고 추출된 각 코드 블록에 대한 해쉬값을 논리연산한 값을 더 포함하는 방법.

### 청구항 4

제 3 항에 있어서,

상기 논리연산은 AND 논리연산 또는 XOR 논리연산인 방법.

### 청구항 5

제 3 항에 있어서,

상기 기초 함수를 선택하는 단계에서, 상기 추출된 대상 함수의 해쉬값과 동일한 기초 유사도 정보가 존재하는 경우 해당 기초 유사도 정보에 대응하는 기초 함수를 유사한 기초 함수로 추출하는 방법.

**청구항 6**

제 3 항에 있어서,

상기 기초 함수를 추출하는 단계에서, 상기 유사도 정보와 상기 기초 유사도 정보의 유사도를 수치화하고 수치화된 유사도가 설정된 임계치 이상인 경우 해당하는 기초 함수를 유사한 기초 함수로 추출하는 방법.

**청구항 7**

제 1 항에 있어서,

상기 기초 실행코드를 결정하는 단계에서, 가장 많은 기초 함수들이 포함된 코드 그룹의 기초 실행코드를 상기 분석 대상 실행코드와 유사한 실행코드로 결정하는 방법.

**청구항 8**

정보 처리 장치가 데이터베이스에 저장된 기초 실행코드별 기초 함수 정보와 상기 기초 함수 정보에 대한 기초 유사도 정보를 이용하여 분석 대상 실행코드의 유사도를 분석하는 시스템에 있어서,

상기 분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 함수 추출 모듈;

상기 추출된 대상 함수들 각각에 대한 유사도 정보를 추출하는 유사도 정보 추출 모듈;

상기 추출된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수와 유사한 기초 함수를 선택하는 유사 함수 선택 모듈;

상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 코드 그룹 분류 모듈; 그리고

상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 분석 대상 실행코드와 유사한 기초 실행코드를 결정하는 유사 코드 판단 모듈을 포함하며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 해쉬값 및 상기 추출된 각 대상 함수의 명령코드들에 대한 통계 정보인 시스템.

**청구항 9**

삭제

**청구항 10**

정보 처리 장치가 데이터베이스에 저장된 기초 실행코드별 기초 함수 정보와 상기 기초 함수 정보에 대한 기초 유사도 정보를 이용하여 분석 대상 실행코드의 유사도를 분석하는 시스템에 있어서,

상기 분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 함수 추출 모듈;

상기 추출된 대상 함수들 각각에 대한 유사도 정보를 추출하는 유사도 정보 추출 모듈;

상기 추출된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수와 유사한 기초 함수를 선택하는 유사 함수 선택 모듈;

상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 코드 그룹 분류 모듈; 그리고

상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 분석 대상 실행코드와 유사한 기초 실행코드를 결정하는 유사 코드 판단 모듈을 포함하며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 해쉬값이며,

상기 유사도 정보는 상기 추출된 각 대상 함수의 명령코드들에 대한 통계 정보, 상기 추출된 각 함수의 변수의 유형별 통계 정보, 상기 추출된 각 대상 함수의 변수 중 레지스터를 제외한 변수들을 삭제한 함수의 해쉬값, 그리고 상기 추출된 각 대상 함수에 포함된 분기문을 경계로 복수의 코드 블록들을 추출하고 추출된 각 코드 블록에 대한 해쉬값을 논리연산한 값을 더 포함하는 시스템.

**청구항 11**

제 10 항에 있어서,

상기 유사 함수 선택 모듈은 상기 추출된 대상 함수의 해쉬값과 동일한 기초 유사도 정보가 존재하는 경우 해당 기초 유사도 정보에 대응하는 기초 함수를 유사한 기초 함수로 선택하는 시스템.

**청구항 12**

제 10 항에 있어서,

상기 유사 함수 선택 모듈은 상기 유사도 정보와 상기 기초 유사도 정보의 유사도를 수치화한 후 수치화된 유사도가 설정된 임계치 이상인 경우 해당 기초 함수를 유사한 기초 함수로 선택하는 시스템.

**청구항 13**

제 8 항에 있어서,

상기 유사 코드 판단 모듈은 가장 많은 기초 함수들이 포함된 코드 그룹의 기초 실행코드를 상기 대상 실행코드와 유사한 실행코드로 결정하는 시스템.

## *발명의 설명*

### *기 술 분 야*

[0001] 본 발명은 데이터 분석 시스템에 관한 것으로, 좀더 구체적으로는 실행코드를 유사도에 기반하여 분석하는 시스템 및 방법에 관한 것이다.

### *배 경 기 술*

[0002] 악성코드란 사용자가 알지 못하는 사이 컴퓨터 시스템에 침입, 설치되어 시스템이나 네트워크에 피해를 주고, 불법적으로 정보를 취득하도록 설계된 소프트웨어를 의미한다. 이러한 악성코드는 네트워크 및 컴퓨터의 발전에 따라 폭발적인 증가 추이를 보이고 있다.

[0003] 이와 같은 악성코드를 탐지하기 위해 파일 내의 문자열이나 파일의 해쉬값을 이용하여 탐지하는 방법이 주로 사용되고 있다. 하지만 이러한 방법은 동일성을 이용하여 탐지하는바, 코드의 극히 일부의 변경이 있는 경우에도 악성코드를 탐지하기 어려웠다. 현재의 많은 악성코드들이 문자열이나 실행 조건 등의 간단한 변형을 통해 생성된다. 따라서 종래의 방법을 사용하여 악성코드를 탐지하는 경우 기존 악성코드와 유사성은 높지만 새로운 악성코드로 탐지가 되거나 또는 악성코드로 탐지가 되지 않았다.

## *발명의 내용*

### *해결하려는 과제*

[0004] 상술한 문제점들을 해결하기 위한 본 발명의 목적은 기존 코드를 재사용하여 생성된 악성코드들을 정확하게 탐지하고 분류하는 실행코드 유사도 분석 시스템 및 방법을 제공하는 것이다.

### *과제의 해결 수단*

[0005] 상기 목적을 달성하기 위한 본 발명의 실시 예에 따른 실행코드 유사도 분석 방법은, 분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 단계; 상기 추출된 대상 함수들 각각에 대한 유사도 정보를 획득하는 단계; 상기 획득된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수들과 유사한 기초 함수들을 선택하는 단계; 상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로

할당하는 단계; 그리고 상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 대상 실행코드와 유사한 기초 실행코드를 결정하는 단계를 포함한다.

[0006] 상기 목적을 달성하기 위한 본 발명의 실시 예에 따른 실행코드 유사도 분석 시스템은, 상기 분석 대상 실행코드의 함수들을 대상 함수들로 추출하는 함수 추출 모듈; 상기 추출된 대상 함수들 각각에 대한 유사도 정보를 추출하는 유사도 정보 추출 모듈; 상기 추출된 유사도 정보를 상기 기초 유사도 정보와 비교하여 상기 추출된 대상 함수와 유사한 기초 함수를 선택하는 유사 함수 선택 모듈; 상기 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하는 코드 그룹 분류 모듈; 그리고 상기 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 상기 대상 실행코드와 유사한 기초 실행코드를 결정하는 유사 코드 판단 모듈을 포함한다.

## *발명의 효과*

[0007] 본 발명의 실시 예에 따르면, 실행코드에 포함된 함수의 유사도를 기반으로 악성코드를 분석하여, 기존 코드의 일부를 변경하여 생성된 악성코드를 보다 정확하게 탐지 및 분류할 수 있다.

## *도면의 간단한 설명*

[0008] 도 1은 본 발명에 따른 실행코드 유사도 분석 시스템의 실시 예를 보여주는 구성도이다.

도 2는 본 발명에 따른 유사도 정보 추출과정을 예시적으로 보여주는 도면이다.

도 3는 본 발명에 따른 실행코드 유사도 분석 과정을 예시적으로 설명하기 위한 도면이다.

도 4은 본 발명에 따른 실행코드 유사도 분석 방법의 실시 예를 보여주는 순서도이다.

## *발명을 실시하기 위한 구체적인 내용*

[0009] 앞의 일반적인 설명 및 다음의 상세한 설명들은 모두 청구된 발명의 부가적인 설명을 제공하기 위한 예시적인 것이다. 그러므로 본 발명은 여기서 설명되는 실시 예에 한정되지 않고 다른 형태로 구체화될 수도 있다. 여기서 소개되는 실시 예는 개시된 내용이 철저하고 완전해 질 수 있도록 그리고 당업자에게 본 발명의 사상이 충분히 전달될 수 있도록 하기 위해 제공되는 것이다.

[0010] 본 출원에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0011] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0012] 본 명세서에서, 어떤 부분이 어떤 구성요소를 포함한다고 언급되는 경우에, 이는 그 외의 다른 구성요소를 더 포함할 수도 있다는 것을 의미한다. 또한, 여기에서 설명되고 예시되는 각 실시 예는 그것의 상보적인 실시 예도 포함한다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0013] 도 1은 본 발명의 실시 예에 따른 실행코드 유사도 분석 시스템의 구성도이다.

[0014] 도 1을 참조하면, 본 발명의 실시 예에 따른 실행코드 유사도 분석 시스템(100)은 함수 추출 모듈(110), 유사도 정보 추출 모듈(120), 유사도 분석 모듈(130), 유사코드 판단 모듈(140), 그리고 기초 코드 데이터베이스(DB)를 포함할 수 있다. 또한, 유사도 분석 모듈(130)은 유사 함수 선택 모듈(131) 및 코드 그룹 분류 모듈(132)을 포함한다.

[0015] 본 발명에 따른 실행코드 유사도 분석 시스템의 동작을 도 1을 참조하여 설명한다. 여기에서 사용되는 어셈블리어는 Window의 x86 플랫폼인 경우를 예를 들어 설명하나 이에 한정되는 것은 아니다.

[0016] 먼저, 함수 추출 모듈(110)은 분석 대상 실행코드를 입력받아 실행코드에 포함된 함수들을 추출한다. 함수 추출

모듈(110)은 함수 구분자를 이용하여 함수들을 추출할 수 있다. 함수 구분자는 어셈블리어이며, 실행코드가 사용되는 플랫폼마다 차이가 있다. 사용되는 플랫폼이 예를 들어, x86인 경우 함수 구분자는 "push ebp", "mov edi", 그리고 "ret"일 수 있다. 이 경우, 함수는 "push ebp" 또는 "mov edi" 로 시작하여 "ret"로 끝날 수 있다. 함수 추출 모듈(110)은 이와 같은 구분자를 인식하여 분석 대상 실행코드에 존재하는 함수들을 대상 함수로 추출하여 추출된 대상 함수들에 대한 정보(Inf_F)를 유사도 정보 추출 모듈(120)로 전송한다.

[0017] 유사도 정보 추출 모듈(120)은 함수 추출 모듈(110)에서 추출된 대상 함수들을 입력받아 각각의 함수마다 기 설정된 동일한 항목의 유사도 정보를 추출한다. 유사도 정보 추출 모듈(120)은 추출된 유사도 정보(Inf_S)를 유사도 분석 모듈(130)로 전송한다.

[0018] 유사도 정보 추출 모듈(120)은 함수에 대한 해쉬값(유사도 정보 1)을 유사도 정보로 생성할 수 있다.

[0019] 또한, 유사도 정보 추출 모듈(120)은 함수에 대해 함수의 변수가 아닌 명령 코드들에 대한 통계 정보(유사도 정보 2)를 유사도 정보로 추출할 수 있다. 명령 코드들은 예를 들어, mov, jmp 등이 있을 수 있다. 통계 정보는 함수에 포함된 명령 코드들의 종류에 대한 정보와 이러한 명령 코드들이 함수 내에서 사용된 횟수에 대한 정보일 수 있다.

[0020] 또한, 유사도 정보 추출 모듈(120)은 함수의 변수 중에서 레지스터 변수를 제외한 나머지 변수들을 삭제한 함수의 해쉬값(유사도 정보 3)을 유사도 정보로 추출할 수 있다. 레지스터 변수는 예를 들어, "eax", "ecx", 그리고 "ebp" 등일 수 있다. 도 2는 함수에서 레지스터 변수를 제외한 변수들을 삭제하는 동작을 예시적으로 도시한다. 도 2를 참조하면, 레지스터 변수인 "ebx", "edi", "eax"를 제외한 나머지 변수들이 삭제되는 것을 볼 수 있다.

[0021] 또한, 유사도 정보 추출 모듈(120)은 함수에 있는 분기문을 기준으로 함수를 다수의 코드 블록으로 구분하고 각 코드 블록에 대한 해쉬값을 논리연산한 값(유사도 정보 4)을 유사도 정보로 추출할 수 있다. 유사도 정보 추출 모듈(200)은 먼저 함수에 포함된 분기문들을 확인한다. 그 후 확인된 분기문들을 기준으로 연속되는 2개의 분기문 사이를 하나의 코드 블록으로 구분하고 구분된 코드 블록들에 대한 해쉬값을 계산한다. 분기문은 예를 들어, "jmp", "jz", "jne", "jnz" 등 일 수 있다. 유사도 정보 추출 모듈(120)은 계산된 각 코드 블록들에 대한 해쉬값들을 AND 논리연산 또는 XOR 논리연산하여 결과적으로 하나의 해쉬값을 추출한다. 상술한 유사도 정보는 실행코드 사이의 연결 정도에 따른 유동성이 적어 유사도 분석의 정확성을 높일 수 있다.

[0022] 또한, 유사도 정보 추출 모듈(120)은 추출된 함수에서 사용된 변수들의 유형에 따른 통계 정보(유사도 정보 5)를 유사도 정보로 추출할 수 있다. 변수들의 유형에 따른 통계 정보는 예를 들어, 변수가 함수 주소인 경우에는 시스템 함수인지 또는 유저(user) 함수인지의 여부에 대한 정보일 수 있다. 또한, 예를 들어 변수가 값을 가지는 경우에는 그 값이 숫자인지 또는 문자열인지 여부에 대한 정보일 수 있다. 유사도 정보 추출 모듈(120)은 변수의 유형을 판단한 후 유형별 사용빈도를 통계 정보로 생성할 수 있다.

[0023] 상술한 유사도 정보는 예시적인 것이며, 이에 한정되지 않고 함수에 대한 다양한 분석값들이 유사도 정보로 이용될 수 있다.

[0024] 기초 코드 데이터베이스(DB)에는 유사도 분석이 완료된 실행코드들이 기초 실행코드로 저장되어 있다. 기초 코드 데이터베이스(DB)에는 기초 실행코드마다 추출된 기초 함수 정보가 저장되어 있다. 또한, 기초 코드 데이터베이스(DB)에는 기초 함수마다 추출된 기초 유사도 정보가 저장되어 있다. 여기에서, 기초 유사도 정보의 종류는 상술한 유사도 정보 추출 모듈(120)에서 추출한 유사도 정보와 동일한 종류이다.

[0025] 유사 함수 선택 모듈(131)은 추출된 유사도 정보를 이용하여 대상 함수와 유사한 기초 함수를 추출한다. 먼저, 유사 함수 추출 모듈(131)은 상술한 유사도 정보 중 대상 함수에 대한 해쉬값(유사도 정보 1)과 동일한 기초 유사도 정보가 기초 코드 데이터베이스(DB)에 저장되어 있는지 확인한다. 동일한 기초 유사도 정보가 존재하는 경우, 유사 함수 선택 모듈(131)은 유사도 정보 1에 해당하는 대상 함수와 확인된 기초 유사도 정보에 해당하는 기초 함수를 동일한 함수로 판단한다. 따라서, 유사 함수 선택 모듈(131) 동일한 함수로 판단된 기초 함수를 유사한 기초 함수로 추출한다. 유사도 정보 1과 동일한 기초 유사도 정보가 기초 코드 데이터베이스(DB)에 존재하지 않는 경우, 유사 함수 선택 모듈(131)은 유사도 정보 2 내지 5를 이용하여 대상 함수와 유사한 기초 함수를 추출한다. 먼저, 유사 함수 선택 모듈(131)은 저장된 기초 함수를 기본 단위로 각 기초 함수에 대한 기초 유사도 정보와 유사도 정보 2 내지 5의 유사도를 비교한다. 이때, 비교되는 유사도는 유사한지 정도에 따라 수치화될 수 있다. 수치화된 유사도가 설정된 임계치 이상인 유사도 정보는 기초 유사도 정보와 유사하다고 판단된다. 유사 함수 추출 모듈(131)은 유사하다고 판단된 유사도 정보를 종합하여 대상 함수와 유사한 기초 함수를 선택한다. 예를 들어, 각 유사도 정보마다 가중치가 주어질 수 있으며, 가중치가 반영되어 수치화된 유사도의 합이

설정된 기준값 이상인 경우 최종적으로 유사하다고 판단할 수 있다.

[0026] 코드 그룹 분류 모듈(132)은 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당한다. 즉, 코드 그룹 분류 모듈(132)은 대상 함수와 유사하다고 판단된 기초 함수들을 동일한 기초 실행 코드에 속하는 기초 함수들끼리 그룹핑하여 코드 그룹들을 생성한다. 예를 들어, 분석 대상 실행코드에서 추출된 대상 함수와 유사하다고 선택된 기초 함수들이 3개이고, 이 중에서 제1 기초 실행코드에 속하는 기초 함수가 2개이고, 제2 기초 실행코드에 속하는 기초 함수가 1개라고 가정한다. 이 경우, 코드 그룹 분류 모듈(132)은 제1 기초 실행코드속하는 기초 함수 2개를 제1 코드 블록으로 생성하고, 제2 기초 실행코드에 속하는 기초 함수 1개를 제2 코드 블록으로 생성한다. 코드 그룹 분류 모듈(132)은 생성된 코드 그룹에 대한 정보(Grb_C)를 유사 코드 판단 모듈(140)로 전송한다.

[0027] 유사 코드 판단 모듈(140)은 코드 그룹 분류 모듈(132)에서 생성된 코드 그룹들을 이용하여 분석 대상 실행코드와 유사한 기초 실행코드를 결정한다. 유사 코드 판단 모듈(140)은 가장 많은 기초 함수들을 포함하는 코드 그룹에 해당하는 기초 실행코드를 상기 대상 실행코드와 유사한 기초 실행코드로 결정할 수 있다. 상술한 예에서, 유사 코드 판단 모듈(140)은 2개의 기초 함수를 포함하는 제1 코드 블록에 속하는 제1 기초 실행코드를 분석 대상 실행코드와 유사한 실행코드로 결정한다.

[0028] 도 3는 본 발명에 따른 실행코드 유사도 분석 과정을 예시적으로 설명하기 위한 도면이다.

[0029] 도 3를 참조하여, 실행코드 유사도 분석 과정을 보다 자세히 설명한다.

[0030] 먼저, 분석 대상 실행코드는 악성코드 C이고, 기초 코드 데이터베이스에 저장된 기초 실행코드는 악성코드 A와 악성코드 B이다. 악성코드 A에서 추출한 함수테이블(tbl1)에는 복수의 함수들(A-1, A-2, A-3, A-4)이 포함되며, 악성코드 B에서 추출한 함수 테이블(tbl2)에는 복수의 함수들(B-1, B-2, B-3, B-4)이 포함된다. 기초 코드 데이터베이스에는 악성코드 A와 악성코드 B에서 추출된 각 함수들에 대한 유사도 정보(유사도 정보 A-1, 유사도 정보 A2, 유사도 정보 B1 등)가 저장되어 있다. 각 유사도 정보는 예를 들어, 상술한 5가지 유형의 유사도 정보를 포함할 수 있다.

[0031] 먼저, 함수 추출 모듈(110)은 분석 대상 실행코드인 악성코드 C에서 함수들(C-1, C-2, C-3, C-4)을 추출하여 함수 테이블(tbl3)을 생성한다.

[0032] 그 후, 유사도 정보 추출 모듈(120)은 추출된 함수들 각각에 대한 유사도 정보(유사도 정보 C-1, 유사도 정보 C-2, 유사도 정보 C-3, 유사도 정보 C-4)를 추출한다.

[0033] 유사도 분석 모듈(130)은 추출된 함수들의 유사도 정보와 기초 코드 데이터베이스(DB)에 저장된 기초 유사도 정보를 비교하여 악성코드 C에서 추출된 함수들과 유사한 기초 함수를 추출한다. 도 3의 분석 테이블(tbl4)을 참조하면, 유사도 정보 A-3과 유사도 정보 C-1, 유사도 정보 A-4와 유사도 정보 C-2가 유사하다. 또한, 유사도 정보 B-2와 유사도 정보 C-4가 유사하다. 따라서, 유사도 분석 모듈(130)은 함수 A-3과 함수 C-1이 유사하고, 함수 A-4와 함수 C-2가 유사하며, 함수 B-2와 함수 C-4가 유사하다고 판단하여 함수 A-3, A-4, 그리고 B-2를 유사한 기초 함수로 추출한다.

[0034] 유사도 분석 모듈(130)은 추출된 기초 함수의 정보를 이용하여 악성코드 A와 악성코드 C를 하나의 코드 그룹(CG1)으로 할당하고, 악성코드 B와 악성코드 C를 하나의 코드 그룹(CG2)로 할당할 수 있다.

[0035] 마지막으로, 유사 코드 판단 모듈(140)이 각 코드 그룹에 포함된 기초 함수의 수를 이용하여 악성코드 A를 악성코드 C와 유사한 악성코드로 결정한다.

[0036] 도 4는 본 발명에 따른 실행코드 유사도 분석 방법의 실시 예를 보여주는 순서도이다.

[0037] 도 4를 참조하면, 본 발명에 따른 실행코드 유사도 분석 방법의 실시 예는 분석 대상 실행코드에서 함수들을 추출한 후, 각 함수들로부터 유사도 정보를 추출한다. 다음으로, 추출된 유사도 정보와 기초 코드 데이터베이스에 저장된 기초 유사도 정보를 비교하여 추출된 함수들과 유사한 기초 함수들을 선택한다. 마지막으로, 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당하고, 할당된 코드 그룹들에 포함된 기초 함수에 기반하여 대상 실행코드와 유사한 기초 실행코드를 결정한다.

[0038] 도 4를 참조하여, 본 발명에 따른 실행코드 유사도 분석 방법의 실시 예를 설명한다.

[0039] 먼저, 함수 추출 모듈(110)은 함수 구분자를 이용하여 분석 대상 실행코드에서 함수들을 추출한다(S100). 여기에서, 함수 구분자는 함수의 시작과 끝을 정의하는 어셈블리어일 수 있으며, 이러한 구분자는 실행코드가 사용

되는 플랫폼에 따라 달라질 수 있다.

[0040] 다음으로, 유사도 정보 추출 모듈(120)은 추출된 함수들로부터 유사도 정보를 추출한다(S200). 추출되는 유사도 정보는 상술한 바와 같이 함수의 해쉬값이나, 명령코드 또는 변수들에 대한 통계 정보 등일 수 있으나, 이에 한정되지 않는다.

[0041] 유사 함수 선택 모듈(131)은 추출된 유사도 정보와 기초 코드 데이터베이스(DB)에 저장된 기초 유사도 정보의 유사도를 비교하여 추출된 대상 함수와 유사한 기초 함수를 선택한다(S300). 유사 함수 추출 모듈(131)은 각 대상 함수마다 추출된 복수의 유사도 정보와 기초 코드 데이터베이스(DB)에 저장된 기초 유사도 정보와의 유사도를 비교하여 이를 수치화한다. 그 후, 수치화된 유사도가 설정된 임계값 이상인 경우 해당 대상 함수와 기초 함수를 유사한 함수로 판단할 수 있다.

[0042] 코드 그룹 분류 모듈(132)은 선택된 기초 함수들 중 동일한 기초 실행코드에 속하는 기초 함수들을 동일 코드 그룹으로 할당한다(S400).

[0043] 마지막으로, 유사코드 판단 모듈(140)은 할당된 코드 그룹들 중 기초 함수를 가장 많이 포함하고 있는 코드 그룹에 해당하는 기초 실행코드를 대상 실행코드와 유사한 실행코드를 결정한다(S500).

[0044] 이제까지 본 발명에 대하여 그 바람직한 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

[0045] 한편, 본 발명의 범위 또는 기술적 사상을 벗어나지 않고 본 발명의 구조가 다양하게 수정되거나 변경될 수 있음은 이 분야에 숙련된 자들에게 자명하다. 상술한 내용을 고려하여 볼 때, 만약 본 발명의 수정 및 변경이 아래의 청구항들 및 동등물의 범주 내에 속한다면, 본 발명이 이 발명의 변경 및 수정을 포함하는 것으로 여겨진다.

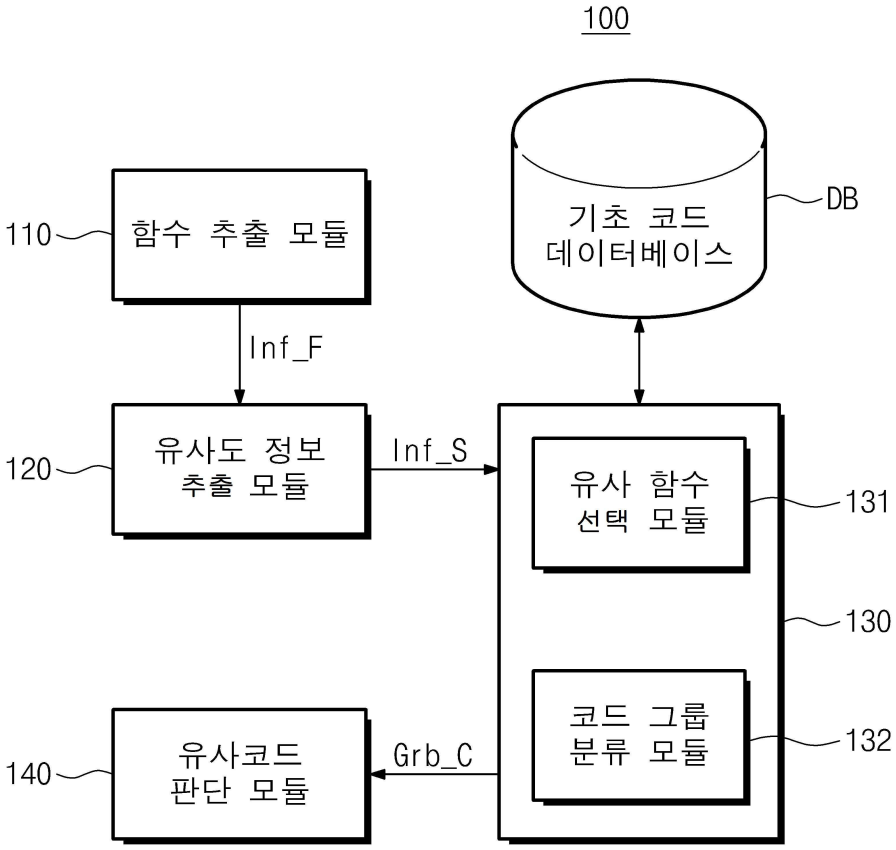
## *부호의 설명*

[0046]

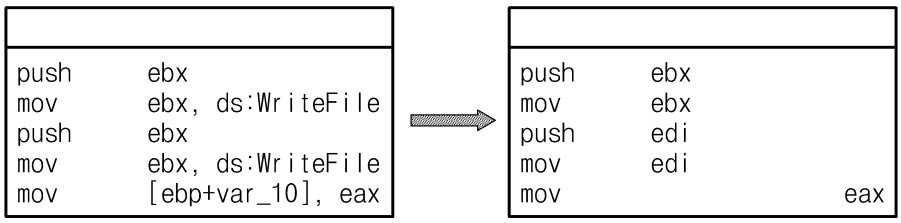
| | |
|---|---|
| 110: 함수 추출 모듈 | 120: 유사도 정보 추출 모듈 |
| 130: 유사도 분석 모듈 | 131: 유사 함수 선택 모듈 |
| 132; 코드 그룹 분류 모듈 | 140: 유사 코드 판단 모듈 |

도면

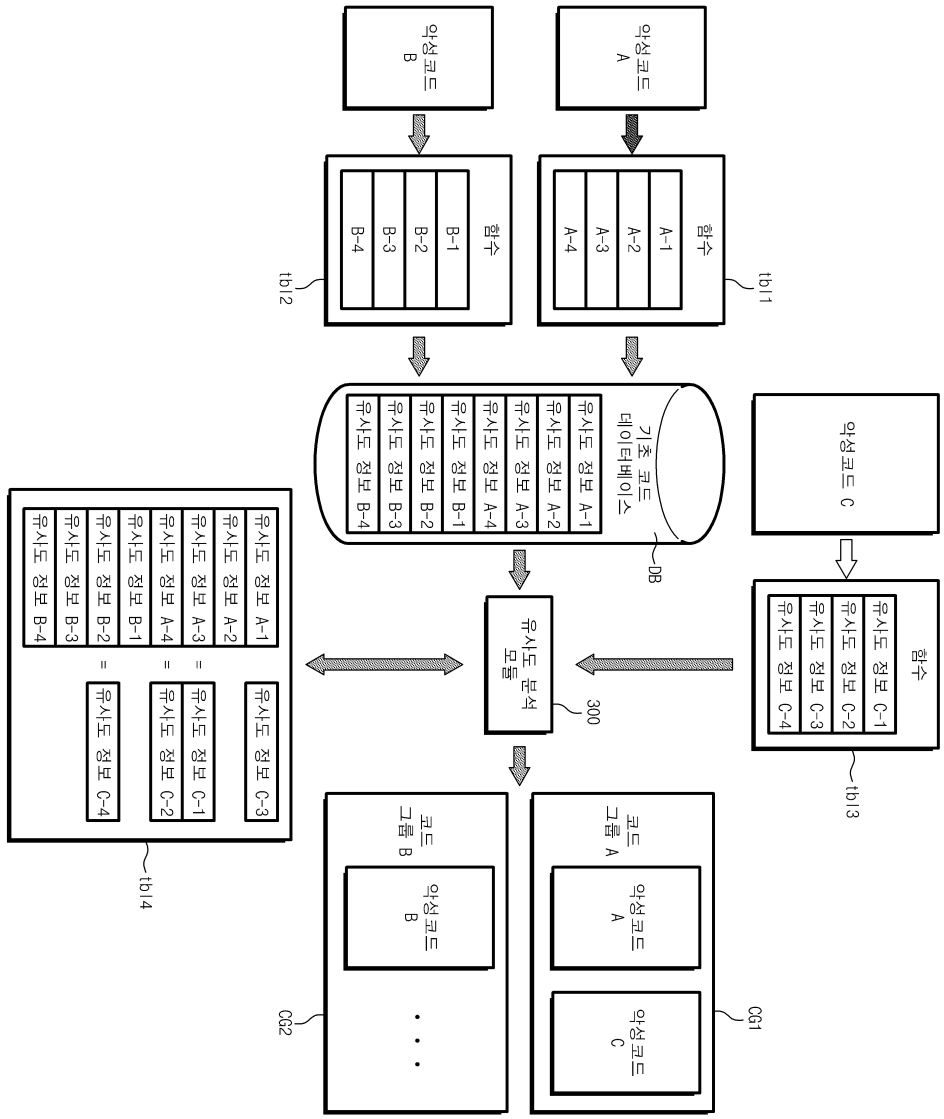
도면1


100
DB
기초 코드
데이터베이스
110
함수 추출 모듈
Inf_F
120
유사도 정보
추출 모듈
Inf_S
유사 함수
선택 모듈
131
130
코드 그룹
분류 모듈
132
140
유사코드
판단 모듈
Grb_C

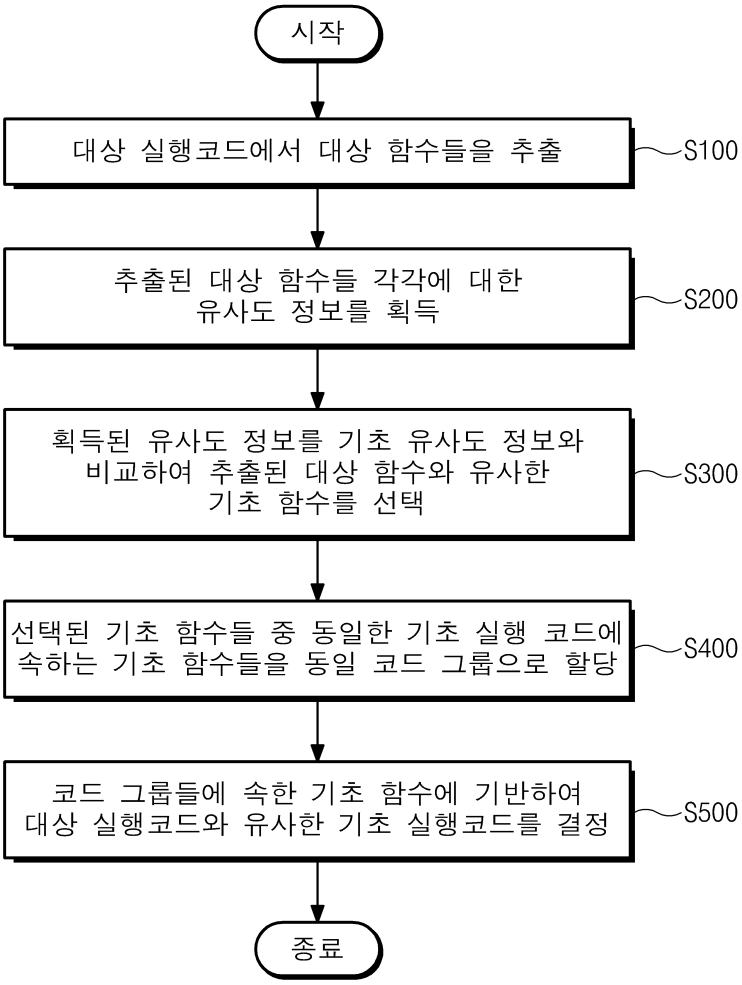
도면2


push ebx
mov ebx, ds:WriteFile
push ebx
mov ebx, ds:WriteFile
mov [ebp+var_10], eax
push ebx
mov ebx
push edi
mov edi
mov eax

도면3


악성코드 A
악성코드 B
악성코드 C
함수
A-1
A-2
A-3
A-4
B-1
B-2
B-3
B-4
tbl1
tbl2
tbl3
기초 코드 데이터베이스
유사도 정보 A-1
유사도 정보 A-2
유사도 정보 A-3
유사도 정보 A-4
유사도 정보 B-1
유사도 정보 B-2
유사도 정보 B-3
유사도 정보 B-4
유사도 정보 C-1
유사도 정보 C-2
유사도 정보 C-3
유사도 정보 C-4
DB
유사도 분석 모듈
300
tbl4
코드 그룹 A
코드 그룹 B
CG1
CG2

도면4


시작
대상 실행코드에서 대상 함수들을 추출
S100
추출된 대상 함수들 각각에 대한
유사도 정보를 획득
S200
획득된 유사도 정보를 기초 유사도 정보와
비교하여 추출된 대상 함수와 유사한
기초 함수를 선택
S300
선택된 기초 함수들 중 동일한 기초 실행 코드에
속하는 기초 함수들을 동일 코드 그룹으로 할당
S400
코드 그룹들에 속한 기초 함수에 기반하여
대상 실행코드와 유사한 기초 실행코드를 결정
S500
종료