



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0091839  
(43) 공개일자 2019년08월07일

(51) 국제특허분류(Int. Cl.)  
G06F 21/56 (2013.01) G06F 16/00 (2019.01)  
(52) CPC특허분류  
G06F 21/56 (2013.01)  
G06F 16/35 (2019.01)  
(21) 출원번호 10-2018-0010905  
(22) 출원일자 2018년01월29일  
심사청구일자 없음

(71) 출원인  
삼성전자주식회사  
경기도 수원시 영통구 삼성로 129 (매탄동)  
고려대학교 산학협력단  
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)  
(72) 발명자  
서재우  
경기도 수원시 권선구 장다리로 10 에버그린 202호  
강수인  
서울특별시 광진구 자양로7길 30 준빌라트 302호  
(뒷면에 계속)  
(74) 대리인  
이건주, 김정훈

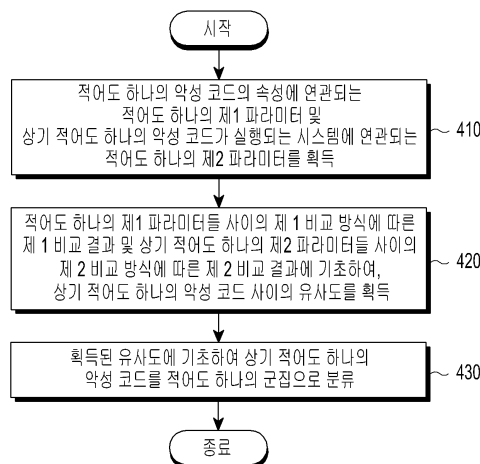
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 악성 코드를 분류하는 전자 장치 및 그 동작 방법

### (57) 요약

전자 장치가 개시된다. 다양한 실시예에 따른 전자 장치는, 프로세서; 및 상기 프로세서에 전기적으로 연결되는 메모리를 포함하고, 상기 프로세서는, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하고, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하고, 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하도록 설정될 수 있다. 그 밖에 다양한 실시예가 제공될 수 있다.

대표도 - 도4



(72) 발명자

**김민철**

서울특별시 중랑구 양원역로 18 금호어울림아파트  
105동 1004호

**김혜민**

서울특별시 동대문구 고산자로 534 제기한신아파트  
101동 101호

**김휘강**

서울특별시 서초구 태봉로2길 65 네이처힐4단지아  
파트 404동 302호

**도기석**

경기도 수원시 영통구 태장로 45 망포마을현대2차  
아이파크아파트 205동 801호

**문주연**

서울특별시 마포구 월드컵북로 235 성산시영아파트  
4동 903호

**송현민**

서울특별시 성북구 종암로19나길 20-3, 603호

**오세준**

서울특별시 동대문구 전농로38길 82-4, 204호

**이수연**

서울시 성북구 안암로 9가길 71, 203호

## 명세서

### 청구범위

#### 청구항 1

전자 장치에 있어서,

프로세서; 및

상기 프로세서에 전기적으로 연결되는 메모리를 포함하고,

상기 프로세서는,

적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하고,

상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하고,

상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하도록 설정된 전자 장치.

#### 청구항 2

제 1항에 있어서,

상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 사이즈(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정되는 전자 장치.

#### 청구항 3

제1항에 있어서, 상기 프로세서는,

상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하도록 설정되는 전자 장치.

#### 청구항 4

제1항에 있어서,

상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함하는 전자 장치.

#### 청구항 5

제4항에 있어서,

상기 프로세서는, 상기 적어도 하나의 악성 코드 중 제1 악성 코드의 상기 복수의 제1 파라미터의 집합과 상기 적어도 하나의 악성 코드 중 제2 악성 코드의 상기 복수의 제2 파라미터의 집합의 교집합의 크기와 합집합의 크기의 비율을 Jaccard 유사도로 획득하도록 설정된 전자 장치.

## 청구항 6

제1항에 있어서,

상기 프로세서는 상기 제1 비교 결과 및 상기 제2 비교 결과의 평균 값에 기초하여 상기 유사도를 획득하도록 설정된 전자 장치.

## 청구항 7

제1항에 있어서,

상기 프로세서는, 상기 적어도 하나의 악성 코드 중 제1 악성 코드 및 제2 악성 코드의 상기 유사도가 임계치 이상인 경우 상기 제1 악성 코드 및 상기 제2 악성 코드를 같은 군집으로 분류하는 전자 장치.

## 청구항 8

제1항에 있어서,

상기 프로세서는 상기 적어도 하나의 악성 코드 사이의 유사도 및 상기 적어도 하나의 군집에 기초하여 상기 적어도 하나의 악성 코드의 네트워크 그래프를 생성하도록 설정된 전자 장치.

## 청구항 9

적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작,

상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및

상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 포함하는 악성 코드를 분류하는 방법.

## 청구항 10

제9항에 있어서,

상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 사이즈(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정되는 악성 코드를 분류하는 방법.

## 청구항 11

제9항에 있어서,

상기 복수의 제2 파라미터를 획득하는 동작은,

상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하는 동작을 포함하는 악성 코드를 분류하는 방법.

## 청구항 12

제9항에 있어서,

상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함하는 악성 코드를 분류하는 방법.

### 청구항 13

제12항에 있어서,

상기 유사도를 획득하는 동작은,

상기 적어도 하나의 악성 코드 중 제1 악성 코드의 상기 복수의 제1 파라미터의 집합과 상기 적어도 하나의 악성 코드 중 제2 악성 코드의 상기 복수의 제2 파라미터의 집합의 교집합의 크기와 합집합의 크기의 비율을 Jaccard 유사도로 획득하는 동작을 포함하는 악성 코드를 분류하는 방법.

### 청구항 14

제9항에 있어서,

상기 유사도를 획득하는 동작은,

상기 제1 비교 결과 및 상기 제2 비교 결과의 평균 값에 기초하여 상기 유사도를 획득하는 동작을 더 포함하는 악성 코드를 분류하는 방법.

### 청구항 15

제9항에 있어서,

상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작은, 상기 적어도 하나의 악성 코드 중 제1 악성 코드 및 제2 악성 코드의 상기 유사도가 임계치 이상인 경우 상기 제1 악성 코드 및 상기 제2 악성 코드를 같은 군집으로 분류하는 동작을 포함하는 악성 코드를 분류하는 방법.

### 청구항 16

제9항에 있어서,

상기 적어도 하나의 악성 코드 사이의 유사도 및 상기 적어도 하나의 군집에 기초하여 상기 적어도 하나의 악성 코드의 네트워크 그래프를 생성하는 동작을 더 포함하는 악성 코드를 분류하는 방법.

### 청구항 17

컴퓨터 상에서 수행하기 위한 프로그램을 기록한 비일시적 컴퓨터 판독 가능한 기록 매체에 있어서,

상기 프로그램은, 프로세서에 의한 실행 시, 상기 프로세서가,

적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작,

상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및

상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 수행하는, 기록 매체.

## 청구항 18

제17항에 있어서,

상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 크기(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정되는 기록 매체.

## 청구항 19

제17항에 있어서,

상기 복수의 제2 파라미터를 획득하는 동작은,

상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하는 동작을 수행하는 기록 매체.

## 청구항 20

제17항에 있어서,

상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함하는 기록 매체.

## 발명의 설명

### 기술 분야

[0001] 다양한 실시예는, 악성 코드에 연관되는 다양한 파라미터에 기초하여 악성 코드를 분류하는 전자 장치 및 그 동작 방법에 관한 것이다.

### 배경 기술

[0003] 악성 코드는 멀웨어(malware) 라고도 하며 컴퓨터에 악영향을 끼칠 수 있는 모든 소프트웨어를 통칭한다. 유사한 특성을 갖는 악성 코드들은 컴퓨터 상에서 감염 방법이나 증상들이 유사하게 발현되기 때문에, 유사한 특성을 갖는 악성 코드들을 분류하여 군집 별로 구분 지어 놓으면, 새로운 악성 코드가 발견되었을 때 새로운 악성 코드와 유사한 악성 코드들에 대한 대응 방식과 유사한 방식으로 대응할 수 있다.

[0004] 다양한 실시예는, 악성 코드의 분석 방법에는 악성 코드를 디스어셈블(disassemble)하는 코드 정적 분석 방법 및 런타임 디버깅(runtime debugging) 기법을 이용하여 악성 코드를 시스템 내에서 직접 실행시키며 발생하는 변화들을 분석하는 코드 동적 분석 방법이 있다.

### 발명의 내용

#### 해결하려는 과제

[0006] 악성 코드에 대하여 코드 정적 분석 방법 또는 코드 동적 분석 방법 중 어느 하나만을 선택적으로 수행하여 악성 코드를 분석하는 경우 분석 결과에 대한 신뢰도가 낮아질 수 있다.

[0007] 다양한 실시예에서는, 악성 코드에 대해 코드 정적 분석 방법과 코드 동적 분석 방법을 모두 이용하여 악성 코드를 분석하고, 분석 결과에 따라 악성 코드 간 유사도 획득 및 유사 군집을 분류할 수 있다.

## 과제의 해결 수단

- [0009] 다양한 실시예에 따른 전자 장치(예: 도 1의 전자 장치(101))에 있어서, 프로세서, 및 상기 프로세서에 전기적으로 연결되는 메모리를 포함하고, 상기 프로세서는, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하고, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하고, 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하도록 설정될 수 있다.
- [0010] 다양한 실시예에 따른 악성 코드를 분류하는 방법은, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 포함할 수 있다.
- [0011] 다양한 실시예에 따른 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 비일시적 컴퓨터 판독 가능한 기록 매체에 있어서, 상기 프로그램은, 프로세서에 의한 실행 시, 상기 프로세서가, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 수행할 수 있다.

## 발명의 효과

- [0013] 다양한 실시예에 따라, 악성 코드에 대한 코드 정적 분석과 코드 동적 분석 결과에 기초하여 유사도를 획득하고 악성 코드들을 군집 별로 분류할 수 있다. 또한, 획득된 유사도 및 분류된 군집에 기초하여 악성 코드들의 네트워크 그래프를 생성함으로써 악성 코드들의 관계를 시각화할 수 있다.

## 도면의 간단한 설명

- [0015] 도 1은 다양한 실시예에 따른 전자 장치를 포함하는 네트워크 환경을 도시한다.
- 도 2는 다양한 실시예들에 따른, 악성 코드를 분류하는 방법에 대한 예시적인 도면이다.
- 도 3은 다양한 실시예에 따른 운영 체제에 따라 악성 코드를 분석하는 방법에 대한 예시적인 도면이다.
- 도 4는 다양한 실시예에 따른 악성 코드를 분류하는 방법에 대한 예시적인 흐름도이다.
- 도 5는 다양한 실시예에 따른 악성 코드의 속성에 연관되는 복수의 제1 파라미터들에 대한 예시이다.
- 도 6은 다양한 실시예에 따른 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터들에 대한 예시이다.
- 도 7a 및 7b는 다양한 종류의 제1 파라미터에 대한 Jaccard 유사도를 획득하는 구성에 대한 예시적인 도면이다.
- 도 8은 다양한 실시예에 따른 Jaccard 유사도를 획득하는 방법을 설명하기 위한 도면이다.
- 도 9 및 도 10은 다양한 실시예에 따른 복수의 제2 파라미터를 비교하여 유사도를 획득하는 예시적인 도면이다.
- 도 11은 다양한 실시예에 따른 파라미터들의 비교 결과에 따라 악성 코드 사이의 유사도를 획득하는 구성에 대한 예시적인 도면이다.
- 도 12 및 도 13은 다양한 실시예에 따른 악성 코드 사이의 유사도에 기초하여 복수의 악성 코드를 시각화하는

예시적인 도면이다.

### 발명을 실시하기 위한 구체적인 내용

- [0016] 도 1은, 다양한 실시예들에 따른, 네트워크 환경(100) 내의 전자 장치(101)의 블록도이다. 도 1을 참조하면, 네트워크 환경(100)에서 전자 장치(101)는 제 1 네트워크(198)(예: 근거리 무선 통신 네트워크)를 통하여 전자 장치(102)와 통신하거나, 또는 제 2 네트워크(199)(예: 원거리 무선 통신 네트워크)를 통하여 전자 장치(104) 또는 서버(108)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(101)는 서버(108)를 통하여 전자 장치(104)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(101)는 프로세서(120), 메모리(130), 입력 장치(150), 음향 출력 장치(155), 표시 장치(160), 오디오 모듈(170), 센서 모듈(176), 인터페이스(177), 햅틱 모듈(179), 카메라 모듈(180), 전력 관리 모듈(188), 배터리(189), 통신 모듈(190), 가입자 식별 모듈(196), 또는 안테나 모듈(197)을 포함할 수 있다. 어떤 실시예에서는, 전자 장치(101)에는, 이 구성요소들 중 적어도 하나(예: 표시 장치(160) 또는 카메라 모듈(180))가 생략되거나, 하나 이상의 다른 구성 요소가 추가될 수 있다. 어떤 실시예에서는, 이 구성요소들 중 일부들은 하나의 통합된 회로로 구현될 수 있다. 예를 들면, 센서 모듈(176)(예: 지문 센서, 홍채 센서, 또는 조도 센서)은 표시 장치(160)(예: 디스플레이)에 임베디드된 채 구현될 수 있다.
- [0017] 프로세서(120)는, 예를 들면, 소프트웨어(예: 프로그램(140))를 실행하여 프로세서(120)에 연결된 전자 장치(101)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 일실시예에 따르면, 데이터 처리 또는 연산의 적어도 일부로서, 프로세서(120)는 다른 구성요소(예: 센서 모듈(176) 또는 통신 모듈(190))로부터 수신된 명령 또는 데이터를 휘발성 메모리(132)에 로드하고, 휘발성 메모리(132)에 저장된 명령 또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리(134)에 저장할 수 있다. 일실시예에 따르면, 프로세서(120)는 메인 프로세서(121)(예: 중앙 처리 장치 또는 어플리케이션 프로세서), 및 이와는 독립적으로 또는 함께 운영 가능한 보조 프로세서(123)(예: 그래픽 처리 장치, 이미지 시그널 프로세서, 센서 허브 프로세서, 또는 커뮤니케이션 프로세서)를 포함할 수 있다. 추가적으로 또는 대체적으로, 보조 프로세서(123)은 메인 프로세서(121)보다 저전력을 사용하거나, 또는 지정된 기능에 특화되도록 설정될 수 있다. 보조 프로세서(123)는 메인 프로세서(121)와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0018] 보조 프로세서(123)는, 예를 들면, 메인 프로세서(121)가 인액티브(예: 슬립) 상태에 있는 동안 메인 프로세서(121)를 대신하여, 또는 메인 프로세서(121)가 액티브(예: 어플리케이션 실행) 상태에 있는 동안 메인 프로세서(121)와 함께, 전자 장치(101)의 구성요소들 중 적어도 하나의 구성요소(예: 표시 장치(160), 센서 모듈(176), 또는 통신 모듈(190))와 관련된 기능 또는 상태들의 적어도 일부를 제어할 수 있다. 일실시예에 따르면, 보조 프로세서(123)(예: 이미지 시그널 프로세서 또는 커뮤니케이션 프로세서)는 기능적으로 관련 있는 다른 구성 요소(예: 카메라 모듈(180) 또는 통신 모듈(190))의 일부로서 구현될 수 있다.
- [0019] 메모리(130)는, 전자 장치(101)의 적어도 하나의 구성요소(예: 프로세서(120) 또는 센서모듈(176))에 의해 사용되는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(예: 프로그램(140)) 및, 이와 관련된 명령에 대한 입력 데이터 또는 출력 데이터를 포함할 수 있다. 메모리(130)는, 휘발성 메모리(132) 또는 비휘발성 메모리(134)를 포함할 수 있다.
- [0020] 프로그램(140)은 메모리(130)에 소프트웨어로서 저장될 수 있으며, 예를 들면, 운영 체제(142), 미들 웨어(144) 또는 어플리케이션(146)을 포함할 수 있다.
- [0021] 입력 장치(150)는, 전자 장치(101)의 구성요소(예: 프로세서(120))에 사용될 명령 또는 데이터를 전자 장치(101)의 외부(예: 사용자)로부터 수신할 수 있다. 입력 장치(150)은, 예를 들면, 마이크, 마우스, 또는 키보드를 포함할 수 있다.
- [0022] 음향 출력 장치(155)는 음향 신호를 전자 장치(101)의 외부로 출력할 수 있다. 음향 출력 장치(155)는, 예를 들면, 스피커 또는 리시버를 포함할 수 있다. 스피커는 멀티미디어 재생 또는 녹음 재생과 같이 일반적인 용도로 사용될 수 있고, 리시버는 착신 전화를 수신하기 위해 사용될 수 있다. 일실시예에 따르면, 리시버는 스피커와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0023] 표시 장치(160)는 전자 장치(101)의 외부(예: 사용자)로 정보를 시각적으로 제공할 수 있다. 표시 장치(160)은, 예를 들면, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 일실시예에 따르면, 표시 장치(160)는 터치를 감지하도록 설정된 터치 회로(touch circuitry), 또는 상기

터치에 의해 발생하는 힘의 세기를 측정하도록 설정된 센서 회로(예: 압력 센서)를 포함할 수 있다.

- [0024] 오디오 모듈(170)은 소리를 전기 신호로 변환시키거나, 반대로 전기 신호를 소리로 변환시킬 수 있다. 일실시예에 따르면, 오디오 모듈(170)은, 입력 장치(150)를 통해 소리를 획득하거나, 음향 출력 장치(155), 또는 전자 장치(101)와 직접 또는 무선으로 연결된 외부 전자 장치(예: 전자 장치(102)) (예: 스피커 또는 헤드폰))를 통해 소리를 출력할 수 있다.
- [0025] 센서 모듈(176)은 전자 장치(101)의 작동 상태(예: 전력 또는 온도), 또는 외부의 환경 상태(예: 사용자 상태)를 감지하고, 감지된 상태에 대응하는 전기 신호 또는 데이터 값을 생성할 수 있다. 일실시예에 따르면, 센서 모듈(176)은, 예를 들면, 제스처 센서, 자이로 센서, 기압 센서, 마그네틱 센서, 가속도 센서, 그립 센서, 근접 센서, 컬러 센서, IR(infrared) 센서, 생체 센서, 온도 센서, 습도 센서, 또는 조도 센서를 포함할 수 있다.
- [0026] 인터페이스(177)는 전자 장치(101)이 외부 전자 장치(예: 전자 장치(102))와 직접 또는 무선으로 연결되기 위해 사용될 수 있는 하나 이상의 지정된 프로토콜들을 지원할 수 있다. 일실시예에 따르면, 인터페이스(177)는, 예를 들면, HDMI(high definition multimedia interface), USB(universal serial bus) 인터페이스, SD카드 인터페이스, 또는 오디오 인터페이스를 포함할 수 있다.
- [0027] 연결 단자(178)는, 그를 통해서 전자 장치(101)가 외부 전자 장치(예: 전자 장치(102))와 물리적으로 연결될 수 있는 커넥터를 포함할 수 있다. 일실시예에 따르면, 연결 단자(178)는, 예를 들면, HDMI 커넥터, USB 커넥터, SD 카드 커넥터, 또는 오디오 커넥터(예: 헤드폰 커넥터)를 포함할 수 있다.
- [0028] 햅틱 모듈(179)은 전기적 신호를 사용자가 촉각 또는 운동 감각을 통해서 인지할 수 있는 기계적인 자극(예: 진동 또는 움직임) 또는 전기적인 자극으로 변환할 수 있다. 일실시예에 따르면, 햅틱 모듈(179)은, 예를 들면, 모터, 압전 소자, 또는 전기 자극 장치를 포함할 수 있다.
- [0029] 카메라 모듈(180)은 정지 영상 및 동영상을 촬영할 수 있다. 일실시예에 따르면, 카메라 모듈(180)은 하나 이상의 렌즈들, 이미지 센서들, 이미지 시그널 프로세서들, 또는 플래시들을 포함할 수 있다.
- [0030] 전력 관리 모듈(188)은 전자 장치(101)에 공급되는 전력을 관리할 수 있다. 일실시예에 따르면, 전력 관리 모듈(388)은, 예를 들면, PMIC(power management integrated circuit)의 적어도 일부로서 구현될 수 있다.
- [0031] 배터리(189)는 전자 장치(101)의 적어도 하나의 구성 요소에 전력을 공급할 수 있다. 일실시예에 따르면, 배터리(189)는, 예를 들면, 재충전 불가능한 1차 전지, 재충전 가능한 2차 전지 또는 연료 전지를 포함할 수 있다.
- [0032] 통신 모듈(190)은 전자 장치(101)와 외부 전자 장치(예: 전자 장치(102), 전자 장치(104), 또는 서버(108))간의 직접(예: 유선) 통신 채널 또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신 모듈(190)은 프로세서(120)(예: 어플리케이션 프로세서)와 독립적으로 운영되고, 직접(예: 유선) 통신 또는 무선 통신을 지원하는 하나 이상의 커뮤니케이션 프로세서를 포함할 수 있다. 일실시예에 따르면, 통신 모듈(190)은 무선 통신 모듈(192)(예: 셀룰러 통신 모듈, 근거리 무선 통신 모듈, 또는 GNSS(global navigation satellite system) 통신 모듈) 또는 유선 통신 모듈(194)(예: LAN(local area network) 통신 모듈, 또는 전력선 통신 모듈)을 포함할 수 있다. 이들 통신 모듈 중 해당하는 통신 모듈은 제 1 네트워크(198)(예: 블루투스, WiFi direct 또는 IrDA(infrared data association) 같은 근거리 통신 네트워크) 또는 제 2 네트워크(199)(예: 셀룰러 네트워크, 인터넷, 또는 컴퓨터 네트워크(예: LAN 또는 WAN)와 같은 원거리 통신 네트워크)를 통하여 외부 전자 장치와 통신할 수 있다. 이런 여러 종류의 통신 모듈들은 하나의 구성 요소(예: 단일 칩)으로 통합되거나, 또는 서로 별도의 복수의 구성 요소들(예: 복수 칩들)로 구현될 수 있다. 무선 통신 모듈(192)은 가입자 식별 모듈(196)에 저장된 가입자 정보(예: 국제 모바일 가입자 식별자(IMSI))를 이용하여 제 1 네트워크(198) 또는 제 2 네트워크(199)와 같은 통신 네트워크 내에서 전자 장치(101)를 확인 및 인증할 수 있다.
- [0033] 안테나 모듈(197)은 신호 또는 전력을 외부(예: 외부 전자 장치)로 송신하거나 외부로부터 수신할 수 있다. 일실시예에 따르면, 안테나 모듈(197)은 하나 이상의 안테나들을 포함할 수 있고, 이로부터, 제 1 네트워크 198 또는 제 2 네트워크 199와 같은 통신 네트워크에서 사용되는 통신 방식에 적합한 적어도 하나의 안테나가, 예를 들면, 통신 모듈(190)에 의하여 선택될 수 있다. 신호 또는 전력은 상기 선택된 적어도 하나의 안테나를 통하여 통신 모듈(190)과 외부 전자 장치 간에 송신되거나 수신될 수 있다.
- [0034] 상기 구성요소들 중 적어도 일부는 주변 기기들간 통신 방식(예: 버스, GPIO(general purpose input and output), SPI(serial peripheral interface), 또는 MIPI(mobile industry processor interface))를 통해 서로 연결되고 신호(예: 명령 또는 데이터)를 상호간에 교환할 수 있다.

[0035] 일실시예에 따르면, 명령 또는 데이터는 제 2 네트워크(199)에 연결된 서버(108)를 통해서 전자 장치(101)와 외부의 전자 장치(104)간에 송신 또는 수신될 수 있다. 전자 장치(102, 104) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 일실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 외부 전자 장치들(102, 104, or 108) 중 하나 이상의 외부 장치들에서 실행될 수 있다. 예를 들면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로, 또는 사용자 또는 다른 장치로부터의 요청에 반응하여 수행해야 할 경우에, 전자 장치(101)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 하나 이상의 외부 전자 장치들에게 그 기능 또는 그 서비스의 적어도 일부를 수행하라고 요청할 수 있다. 상기 요청을 수신한 하나 이상의 외부 전자 장치들은 요청된 기능 또는 서비스의 적어도 일부, 또는 상기 요청과 관련된 추가 기능 또는 서비스를 실행하고, 그 실행의 결과를 전자 장치(101)로 전달할 수 있다. 전자 장치(101)는 상기 결과를, 그대로 또는 추가적으로 처리하여, 상기 요청에 대한 응답의 적어도 일부로서 제공할 수 있다.. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.

[0037] 도 2는 다양한 실시예들에 따른, 악성 코드를 분류하는 방법에 대한 예시적인 도면이다.

[0038] 본 발명의 다양한 실시예에 따른 전자 장치(예: 도 1의 전자 장치(101))의 프로세서(예: 도 1의 프로세서(120))는, 적어도 하나의 악성 코드를 분석하여 악성 코드 사이의 유사도를 획득하여 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류할 수 있다. 전자 장치(101)는 메모리(예: 도 1의 메모리(130))는 적어도 하나의 군집으로 분류된 적어도 하나의 악성 코드에 대한 정보를 저장할 수 있다.

[0039] 다양한 실시예에 따라, 프로세서(120)는 악성 코드(210)를 분석하여 다양한 피처들(220)을 생성할 수 있다. 프로세서(120)는 악성 코드(210)에 코드 동적 분석 방법 및 코드 정적 분석 방법을 이용하여 피처들(220)을 생성할 수 있다. 피처들(220)은 사용자 또는 프로세서(120)에 의해 지정된 개수로 획득되도록 설정될 수 있으며, 피처들(220) 각각은 악성 코드(210)에 연관되는 특정 파라미터들의 집합일 수 있다. 예를 들어, 프로세서(120)는 악성 코드(210)의 속성에 연관되는 복수의 제1 파라미터들의 집합인 피처 1을 생성하거나 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터들의 집합인 피처 2를 획득할 수 있다. 프로세서(120)는 코드 정적 분석 방법을 이용하여 악성 코드(210) 자체의 속성에 기초하여 복수의 제1 파라미터를 획득할 수 있다. 예를 들어, 제1 파라미터는 악성 코드(210)의 파일 이름, 파일 사이즈, 섹션 마다 할당되는 정보의 속성 중 적어도 하나에 기초하여 획득될 수 있으나, 제1 파라미터를 결정하는 속성은 위 예시들에 한정되는 것은 아니며, 악성 코드(210)의 다양한 속성에 기초하여 결정될 수 있다. 프로세서(120)는 코드 동적 분석 방법을 이용하여 악성 코드(210)가 시스템에서 행동하는 동작의 유사성에 기초하여 악성 코드(210)의 시스템에 연관되는 제2 파라미터를 생성할 수 있다. 예를 들어, 프로세서(120)는 악성 코드(210)가 실행되는 운영 체제에서 호출되는 콜 시퀀스를 이용하여 제2 파라미터를 생성할 수 있다. 메모리(130)는 획득된 제1 파라미터 및 제2 파라미터를 저장할 수 있다.

[0040] 도 3은 다양한 실시예에 따른 운영 체제에 따라 악성 코드를 분석하는 방법에 대한 예시적인 도면이다.

[0041] 다양한 실시예에 따라, 전자 장치(예: 도 1의 전자 장치(101))의 프로세서(예: 도 1의 프로세서(120))는 분석을 수행할 악성 코드(310)가 이미 메모리(130)에 저장된 악성 코드(310)인지 확인할 수 있다. 예를 들어, 프로세서(120)는 악성 코드(310)의 형식을 확인한 후 해시(hash)를 이용하여 이미 메모리(130)에 저장되어 있던 악성 코드(310)인지 확인할 수 있다. 악성 코드(310)가 메모리(130)에 저장된 악성 코드(310)가 아닌 것으로 확인된 경우, 프로세서(120)는 악성 코드(310)가 어떤 시스템, 예를 들어 어떤 운영 체제에서 실행되는 지 확인할 수 있다. 예를 들어, 프로세서(120)는 악성 코드(310)가 윈도우에서 실행되는 것으로 확인된 경우 윈도우 태스크 큐(Windows task queue, 321)에 악성 코드(310)를 할당할 수 있다. 동일하게, 프로세서(120)는 악성 코드(310)가 리눅스(Linux)에서 실행되는 것으로 확인된 경우 악성 코드(310)를 리눅스 태스크 큐(Linux task queue, 322)에 할당할 수 있고, 안드로이드(Android)에서 실행되는 것으로 확인된 경우 악성 코드(310)를 안드로이드 태스크 큐(Android task queue, 323)에 할당할 수 있다. 윈도우 태스크 큐(321), 리눅스 태스크 큐(322), 안드로이드 태스크 큐(323) 중 어느 하나에 할당된 악성 코드(310)는 각각 윈도우 악성 코드 분석기(331), 리눅스 악성 코드 분석기(332), 안드로이드 악성 코드 분석기(333) 중 어느 하나에 할당될 수 있다. 윈도우 악성 코드 분석기(331), 리눅스 악성 코드 분석기(332), 안드로이드 악성 코드 분석기(333)는 도 2에서 설명한 프로세서(120)와 같은 동작을 수행하여 악성 코드(310)에 연관되는 복수의 제1 파라미터 및 복수의 제2 파라미터를 생성할 수 있다. 프로세서(120)는 악성 코드(310)의 제1 파라미터 및 제2 파라미터를 메모리(340, 예: 도 1의 메모리(130))에 저장하거나, 제1 파라미터 및 제2 파라미터를 이용하여 다른 악성 코드와 유사도를 획득할 수 있다.

- [0042] 도 2를 참고하면, 프로세서(120)는 악성 코드(210)의 분석 결과에 따라 파라미터들의 집합인 피처들(220)을 획득하는 경우, 피처들(220)의 특성에 맞는 유사도 알고리즘에 피처들(220)을 입력할 수 있다. 예를 들어, 프로세서(120)는  $n$  개의 피처들(220) 중 피처 1 및 피처 6 (231)은 유사도 알고리즘 A(234)에 입력하고, 피처 2 및 피처 4 (232)는 유사도 알고리즘 B(235)에 입력하고, 피처 3 및 피처 5 (233)는 유사도 알고리즘 C(236)에 할당하여 각각의 피처에 기초한 악성 코드(210)와 다른 악성 코드와의 유사도를 확인 또는 획득할 수 있다. 프로세서(120)는 유사도 알고리즘 A(234), 유사도 알고리즘 B(235), 유사도 알고리즘 C(236)에서 획득한 유사도를 통해 악성 코드(210)와 다른 악성 코드와의 최종 유사도를 획득할 수 있다. 프로세서(120)는 최종 유사도 획득 시에 유사도 알고리즘 별로 가중치를 적용하여 최종 유사도를 획득할 수 있다. 예를 들어, 프로세서는  $n$  개의 피처들(220)에 대해 피처 별로 각각의 가중치를  $1/n$ 로 할당하여  $n$  개의 피처들에 대한 평균 값을 획득하고, 획득된 평균 값을 최종 유사도(237)로서 획득할 수 있다. 하지만  $1/n$ 의 가중치는 예시적인 것일 뿐 본원의 내용을 한정하는 것은 아니며, 특정 피처의 중요도에 따라서 피처 별로 가중치를 상이하게 할당할 수 있다. 다양한 실시예에 따라, 프로세서(120)는 획득된 최종 유사도에 기초하여 시각화(238)를 수행할 수 있다. 예를 들어, 프로세서(120)는 2차원 또는 3차원 공간 상에서 복수의 악성 코드들 간의 유사도를 그래프를 이용하여 시각화할 수 있다.
- [0044] 도 4는 다양한 실시예에 따른 악성 코드를 분류하는 방법에 대한 예시적인 흐름도이다. 도 5는 다양한 실시예에 따른 악성 코드의 속성에 연관되는 복수의 제1 파라미터들에 대한 예시이다. 도 6은 다양한 실시예에 따른 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터들에 대한 예시이다.
- [0045] 동작 410에서, 프로세서(예: 도 1의 프로세서(120))는 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득할 수 있다. 프로세서(120)는 도 2에서 설명한 바와 같이 악성 코드의 분석을 통해 복수의 제1 파라미터들 및 복수의 제2 파라미터들을 획득할 수 있다. 도 5를 참고하면, 복수의 제1 파라미터들은 아이디 1 내지 6을 가지는 악성 코드들의 섹션 정보일 수 있다. 프로세서(120)는, 악성 코드 사이의 유사도를 획득하기 위해 악성 코드들 각각에 대응하는 복수의 제1 파라미터들에 대하여 집합 유사도 알고리즘을 이용할 수 있다. 예를 들어, 프로세서(120)는 Jaccard 유사도 측정 방법을 이용하여 악성 코드들 사이의 유사도를 획득할 수 있다.
- [0046] 도 6을 참고하면, 복수의 제2 파라미터들은 아이디 1 내지 6을 갖는 악성 코드들의 콜 시퀀스(call sequence)에 의해 생성될 수 있다. 예를 들어, 복수의 제2 파라미터들은 악성 코드가 시스템 내에서 실행될 때 호출되는 시스템 내의 함수들을 이용하여 획득될 수 있다. 프로세서(120)는 악성 코드들의 콜 시퀀스에 기초하여 악성 코드 간 유사도를 측정하기 위해 시퀀스 유사도 알고리즘을 이용할 수 있다. 예를 들어, 프로세서(120)는 Nilsimsa 알고리즘을 이용하여 악성 코드 간 유사도를 측정할 수 있다.
- [0047] 도 4의 동작 420에서, 프로세서(120)는 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득할 수 있다. 앞서 설명한 바와 같이, 프로세서(120)는 악성 코드의 속성에 연관되는 복수의 제1 파라미터들에는 집합 유사도 알고리즘을 적용하고, 악성코드가 실행되는 시스템에 연관되는 제2 파라미터인 콜 시퀀스에는 시퀀스 유사도 알고리즘을 적용할 수 있다.
- [0048] 도 7a 및 7b는 다양한 종류의 제1 파라미터에 대한 Jaccard 유사도를 획득하는 구성에 대한 예시적인 도면이다.
- [0049] 도 7a를 참고하면, 다양한 실시예에 따른 프로세서(예: 도 1의 프로세서(120))는 악성 코드 A(701)와 악성 코드 B(702) 각각의 복수의 제1 파라미터들을 악성 코드 A 및 B의 ELF(executable and linkable format) 섹션으로 설정할 수 있다. ELF 섹션은 링커가 오브젝트 파일을 해석하기 위한 대부분의 정보를 의미할 수 있다. 프로세서(120)는 악성 코드 A(701)와 악성 코드 B(702)의 복수의 제1 파라미터들 사이에 Jaccard 유사도(703) 알고리즘을 적용하여 섹션 유사도(704)를 결정할 수 있다.
- [0050] 동일하게 도 7b를 참고하면, 다양한 실시예에 따른 프로세서(120)는 악성 코드 A(701)와 악성 코드 B(702) 각각의 복수의 제1 파라미터들을 악성 코드 A(701) 및 B(702)의 심볼 리스트(symbol list)로 설정할 수 있다. 심볼 리스트는, 오브젝트 파일, 라이브러리와 같은 바이너리 파일에 대한 심볼을 포함할 수 있다. 프로세서(120)는 악성 코드 A(701)와 악성 코드 B(702)의 복수의 제1 파라미터들 사이에 Jaccard 유사도(703) 알고리즘을 적용하여 심볼 유사도(714)를 결정할 수 있다.
- [0051] 도 8은 다양한 실시예에 따른 Jaccard 유사도를 획득하는 방법을 설명하기 위한 도면이다.

- [0052] 다양한 실시예에 따른 프로세서(예: 도 1의 프로세서(120))는, 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820) 간의 포함 관계를 이용하여 유사도를 획득할 수 있다. 예를 들어, 도 8을 참고하면 악성 코드 A의 복수의 제1 파라미터들의 집합(810)은 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 교집합(830) 영역에 총 14개의 파라미터들이 포함될 수 있다. 또한, 악성 코드 A의 복수의 제1 파라미터들의 집합(810)에서 교집합(830)을 제외한 영역에는 하나의 파라미터가 포함될 수 있고 악성 코드 B의 복수의 제1 파라미터들의 집합(820)에서 교집합(830)을 제외한 영역에는 파라미터가 포함되지 않을 수 있다. 예를 들어, 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 합집합의 원소의 개수는 15개일 수 있다. Jaccard 유사도 측정 방식에 의하면, 프로세서(120)는 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 교집합의 원소 개수 14를 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 합집합의 원소 개수 15로 나누어 0.93을 악성 코드 A와 악성 코드 B 사이의 Jaccard 유사도로 획득할 수 있다. 동일하게, 프로세서(120)는 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 교집합의 크기와 악성 코드 A의 복수의 제1 파라미터들의 집합(810)과 악성 코드 B의 복수의 제1 파라미터들의 집합(820)의 합집합의 크기의 비율을 악성 코드 A와 악성 코드 B 사이의 Jaccard 유사도로 획득할 수 있다.
- [0053] 도 9 및 도 10은 다양한 실시예에 따른 복수의 제2 파라미터를 비교하여 Nilsimsa 유사도를 획득하는 예시적인 도면이다.
- [0054] 다양한 실시예에 따라, 프로세서(예: 도 1의 프로세서(120))는 문자열 간의 유사도를 나타내는 LSH(locality sensitive hashing) 기반의 Nilsimsa 알고리즘을 이용하여 악성 코드들 사이의 복수의 제2 파라미터들의 유사도를 획득할 수 있다. 프로세서(120)는 Nilsimsa 알고리즘을 통해 악성 코드에 대한 Nilsimsa 해시를 생성할 수 있다. 유사한 악성 코드들은 유사한 Nilsimsa 해시를 가지므로, 프로세서(120)는 서로 다른 악성 코드에 대응하는 Nilsimsa 해시들에 포함된 문자열 간의 유사도에 기초하여 악성 코드 사이의 Nilsimsa 유사도를 획득할 수 있다.
- [0055] 도 9를 참고하면, 프로세서(120)는 악성 코드 A(910), 악성 코드 B(920), 및 악성 코드 C(930) 각각에 대해 n 개의 시스템 콜을 포함하는 콜 시퀀스들(911, 921, 931)을 획득할 수 있다. 프로세서(120)는 콜 시퀀스들(911, 921, 931) 각각을 Nilsimsa 알고리즘(940)에 입력시켜 Nilsimsa 해시 1(951), Nilsimsa 해시 2(952), Nilsimsa 해시 3(953)을 각각 생성할 수 있다. 프로세서(120)는 Nilsimsa 해시 1(951), Nilsimsa 해시 2(952), Nilsimsa 해시 3(953)을 각각 비교하여 악성 코드 A(910), 악성 코드 B(920), 및 악성 코드 C(930) 간의 Nilsimsa 유사도를 획득할 수 있다. 예를 들어, 프로세서(120)는 Nilsimsa 해시 1(951)과 Nilsimsa 해시 2(952)를 비교하여 1-2 API call 유사도(961)를 획득하고, Nilsimsa 해시 1(951)과 Nilsimsa 해시 3(953)을 비교하여 1-3 API call 유사도(962)를 획득하고, Nilsimsa 해시 2(952)와 Nilsimsa 해시 3(953)을 비교하여 2-3 API call 유사도(963)를 획득할 수 있다.
- [0056] 도 10을 참고하면, 프로세서(예: 도 1의 프로세서(120))는 악성 코드 1 내지 4의 Nilsimsa 유사도를 악성코드1을 기준으로 획득할 수 있다. 프로세서(120)는 악성 코드 1 내지 4의 콜 시퀀스 각각에 대해 Nilsimsa 해시를 생성할 수 있다. 콜 시퀀스에는 시스템 콜 호출 시 사용되는 경로, 옵션, 변수 등의 정보가 포함될 수 있다. 프로세서(120)는, 악성 코드 1의 Nilsimsa 해시와 악성 코드 2 내지 4의 Nilsimsa 해시의 문자열을 비교하여 Nilsimsa 유사도를 획득할 수 있다.
- [0057] 도 11은 다양한 실시예에 따른 파라미터들의 비교 결과에 따라 악성 코드 사이의 유사도를 획득하는 구성에 대한 예시적인 도면이다.
- [0058] 다양한 실시예에 따라, 프로세서(120)는 도 7a, 도 7b, 도 9에서 획득한 Nilsimsa 유사도와 Jaccard 유사도를 이용하여 악성 코드들 사이의 유사도를 획득할 수 있다. 프로세서(120)는, 악성 코드들 사이의 유사도를 획득된 Nilsimsa 유사도와 Jaccard 유사도들의 평균 값으로 결정할 수 있다. 예를 들어, Nilsimsa 유사도와 Jaccard 유사도들에 1/3의 가중치를 곱하여 합산함으로써 악성 코드 사이의 유사도를 결정할 수 있다. 하지만, 1/3의 가중치 값은 예시적인 것일 뿐, 각각의 파라미터에 따른 유사도 별로 다른 가중치를 곱하여 합산하여 악성 코드 사이의 유사도를 결정할 수 있다. 예를 들어, ID 4를 갖는 악성 코드와 ID 2를 갖는 악성 코드의 API call 유사도가 0.726이고, 심볼 유사도가 0.013, 및 섹션 유사도가 0.307인 경우, ID 4의 악성 코드와 ID 2의 악성 코드의 유사도는 평균 값인 0.349로 획득될 수 있다.

- [0060] 도 4의 동작 430에서, 프로세서(120)는 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류할 수 있다. 프로세서(120)는 비교 결과에 따라 악성 코드들 사이의 유사도가 임계치 이상이라고 판단되는 경우, 악성 코드들이 유사하다고 판단할 수 있다. 프로세서(120)는 악성 코드들 사이의 유사도를 시각적으로 확인하기 위해 적어도 하나의 악성 코드들을 이용하여 네트워크 그래프를 생성할 수 있다.
- [0061] 도 12 및 도 13은 다양한 실시예에 따른 악성 코드 사이의 유사도에 기초하여 복수의 악성 코드를 시각화하는 예시적인 도면이다.
- [0062] 도 12를 참고하면, 프로세서(예: 도 1의 프로세서(120))는 악성 코드 A(1210), 악성 코드 B(1220), 및 악성 코드 C(1230)을 이용하여 네트워크 그래프를 생성할 수 있다. 예를 들어, 악성 코드 A(1210), 악성 코드 B(1220)의 비교 결과 유사도가 임계치 이상인 것으로 판단된 경우, 노드(node)를 구성하고 있는 악성 코드 A(1210)와 악성 코드 B(1220) 사이를 엣지(edge)로 연결하여 악성 코드 A(1210)와 악성 코드 B(1220)가 유사함을 시각적으로 표현할 수 있다. 프로세서(120)는 악성 코드 C(1230)가 악성 코드 A(1210) 및 악성 코드 B(1220)와 유사하지 않은 것으로 판단한 경우, 악성 코드 C(1230)와 악성 코드 A(1210) 및 악성 코드 B(1220)를 엣지로 연결하지 않음으로써 시각적으로 표현할 수 있다.
- [0063] 도 13을 참고하면, 프로세서(예: 도 1의 프로세서(120))는 도 12의 2차원 네트워크 그래프를 3차원으로 확장하여 악성 코드들 사이의 유사도를 시각적으로 표현할 수 있다. 프로세서(120)는 비교 결과 임계치 이상의 유사도를 갖는 악성 코드들을 엣지로 연결하여 복수의 군집들(1310, 1320, 1330, 1340, 1350)로 시각화할 수 있다. 프로세서(120)는 새로운 악성 코드가 발견되는 경우, 복수의 군집들(1310, 1320, 1330, 1340, 1350)에 포함된 악성 코드들과 비교를 통해 새로운 악성 코드와 유사한 군집을 확인할 수 있다.
- [0065] 다양한 실시예에 따른 전자 장치(예: 도 1의 전자 장치(101))에 있어서, 프로세서(예: 도 1의 프로세서(120)), 및 상기 프로세서(120)에 전기적으로 연결되는 메모리(예: 도 1의 메모리(130))를 포함하고, 상기 프로세서(120)는, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하고, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하고, 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하도록 설정될 수 있다.
- [0066] 일 실시예에 따라, 상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 사이즈(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정될 수 있다. 일 실시예에 따라, 상기 프로세서(120)는, 상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하도록 설정될 수 있다. 일 실시예에 따라, 상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함할 수 있다. 일 실시예에 따라, 상기 프로세서(120)는, 상기 적어도 하나의 악성 코드 중 제1 악성 코드의 상기 복수의 제1 파라미터의 집합과 상기 적어도 하나의 악성 코드 중 제2 악성 코드의 상기 복수의 제2 파라미터의 집합의 교집합의 크기와 합집합의 크기의 비율을 Jaccard 유사도로 획득하도록 설정될 수 있다. 일 실시예에 따라, 상기 프로세서(120)는 상기 제1 비교 결과 및 상기 제2 비교 결과의 평균 값에 기초하여 상기 유사도를 획득하도록 설정될 수 있다. 일 실시예에 따라, 상기 프로세서(120)는, 상기 적어도 하나의 악성 코드 중 제1 악성 코드 및 제 2 악성 코드의 상기 유사도가 임계치 이상인 경우 상기 제1 악성 코드 및 상기 제2 악성 코드를 같은 군집으로 분류할 수 있다. 일 실시예에 따라, 상기 프로세서(120)는 상기 적어도 하나의 악성 코드 사이의 유사도 및 상기 적어도 하나의 군집에 기초하여 상기 적어도 하나의 악성 코드의 네트워크 그래프를 생성하도록 설정될 수 있다.
- [0067] 다양한 실시예에 따른 악성 코드를 분류하는 방법은, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 포함할 수 있다.

- [0068] 일실시예에 따라, 상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 사이즈(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정될 수 있다. 일실시예에 따른 상기 복수의 제2 파라미터를 획득하는 동작은, 상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하는 동작을 포함할 수 있다. 일실시예에 따라, 상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함할 수 있다. 일실시예에 따른 상기 유사도를 획득하는 동작은, 상기 적어도 하나의 악성 코드 중 제1 악성 코드의 상기 복수의 제1 파라미터의 집합과 상기 적어도 하나의 악성 코드 중 제2 악성 코드의 상기 복수의 제2 파라미터의 집합의 교집합의 크기와 합집합의 크기의 비율을 Jaccard 유사도로 획득하는 동작을 포함할 수 있다. 일실시예에 따른 방법은, 상기 제1 비교 결과 및 상기 제2 비교 결과의 평균 값에 기초하여 상기 유사도를 획득하는 동작을 더 포함할 수 있다. 일실시예에 따른 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작은, 상기 적어도 하나의 악성 코드 중 제1 악성 코드 및 제2 악성 코드의 상기 유사도가 임계치 이상인 경우 상기 제1 악성 코드 및 상기 제2 악성 코드를 같은 군집으로 분류하는 동작을 포함할 수 있다. 일실시예에 따른 방법은, 상기 적어도 하나의 악성 코드 사이의 유사도 및 상기 적어도 하나의 군집에 기초하여 상기 적어도 하나의 악성 코드의 네트워크 그래프를 생성하는 동작을 더 포함할 수 있다.
- [0069] 다양한 실시예에 따른 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 비일시적 컴퓨터 판독 가능한 기록 매체에 있어서, 상기 프로그램은, 프로세서(120)에 의한 실행 시, 상기 프로세서(120)가, 적어도 하나의 악성 코드의 속성에 연관되는 복수의 제1 파라미터 및 상기 적어도 하나의 악성 코드가 실행되는 시스템에 연관되는 복수의 제2 파라미터를 획득하는 동작, 상기 복수의 제1 파라미터들 사이의 제 1 비교 방식에 따른 제 1 비교 결과 및 상기 복수의 제2 파라미터들 사이의 제 2 비교 방식에 따른 제 2 비교 결과에 기초하여, 상기 적어도 하나의 악성 코드 사이의 유사도를 획득하는 동작, 및 상기 획득된 유사도에 기초하여 상기 적어도 하나의 악성 코드를 적어도 하나의 군집으로 분류하는 동작을 수행할 수 있다.
- [0070] 일실시예에 따라, 상기 적어도 하나의 제1 파라미터는 상기 적어도 하나의 악성 코드의 명칭(name), 파일 타입(file type), 파일 사이즈(file size), 헤더(header) 정보 중 적어도 하나에 기초하여 결정될 수 있다. 일실시예에 따라, 상기 프로세서(120)는 상기 악성 코드가 시스템 상에서 실행될 경우에 생성되는 콜 시퀀스의 문자열을 비교하여 상기 복수의 제2 파라미터를 획득하는 동작을 수행할 수 있다. 일실시예에 따라, 상기 제1 비교 방식은 Jaccard 유사도 측정 방식을 포함하며, 상기 제2 비교 방식은 Nilsimsa 유사도 측정 방식을 포함할 수 있다.
- [0072] 본 문서에 개시된 다양한 실시예들에 따른 전자 장치는 다양한 형태의 장치가 될 수 있다. 전자 장치는, 예를 들면, 휴대용 통신 장치 (예: 스마트폰), 컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 또는 가전 장치를 포함할 수 있다. 본 문서의 실시예에 따른 전자 장치는 전술한 기기들에 한정되지 않는다.
- [0073] 본 문서의 다양한 실시예들 및 이에 사용된 용어들은 본 문서에 기재된 기술적 특징들을 특정한 실시예들로 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 또는 관련된 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 아이템에 대응하는 명사의 단수 형은 관련된 문맥상 명백하게 다르게 지시하지 않는 한, 상기 아이템 한 개 또는 복수개를 포함할 수 있다. 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C," "A, B 및 C 중 적어도 하나," 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제 1", "제 2", 또는 "첫째" 또는 "둘째"와 같은 용어들은 단순히 해당 구성요소를 다른 해당 구성요소와 구분하기 위해 사용될 수 있으며, 해당 구성요소들을 다른 측면(예: 중요성 또는 순서)에서 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에, "기능적으로" 또는 "통신적으로"라는 용어와 함께 또는 이런 용어 없이, "커플드" 또는 "커넥티드"라고 언급된 경우, 그것은 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로(예: 유선으로), 무선으로, 또는 제 3 구성요소를 통하여 연결될 수 있다는 것을 의미한다.
- [0074] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구현된 유닛을 포함할 수 있으며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는, 상기 부품의 최소 단위 또는 그 일부가 될 수 있다. 예를 들

면, 일실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.

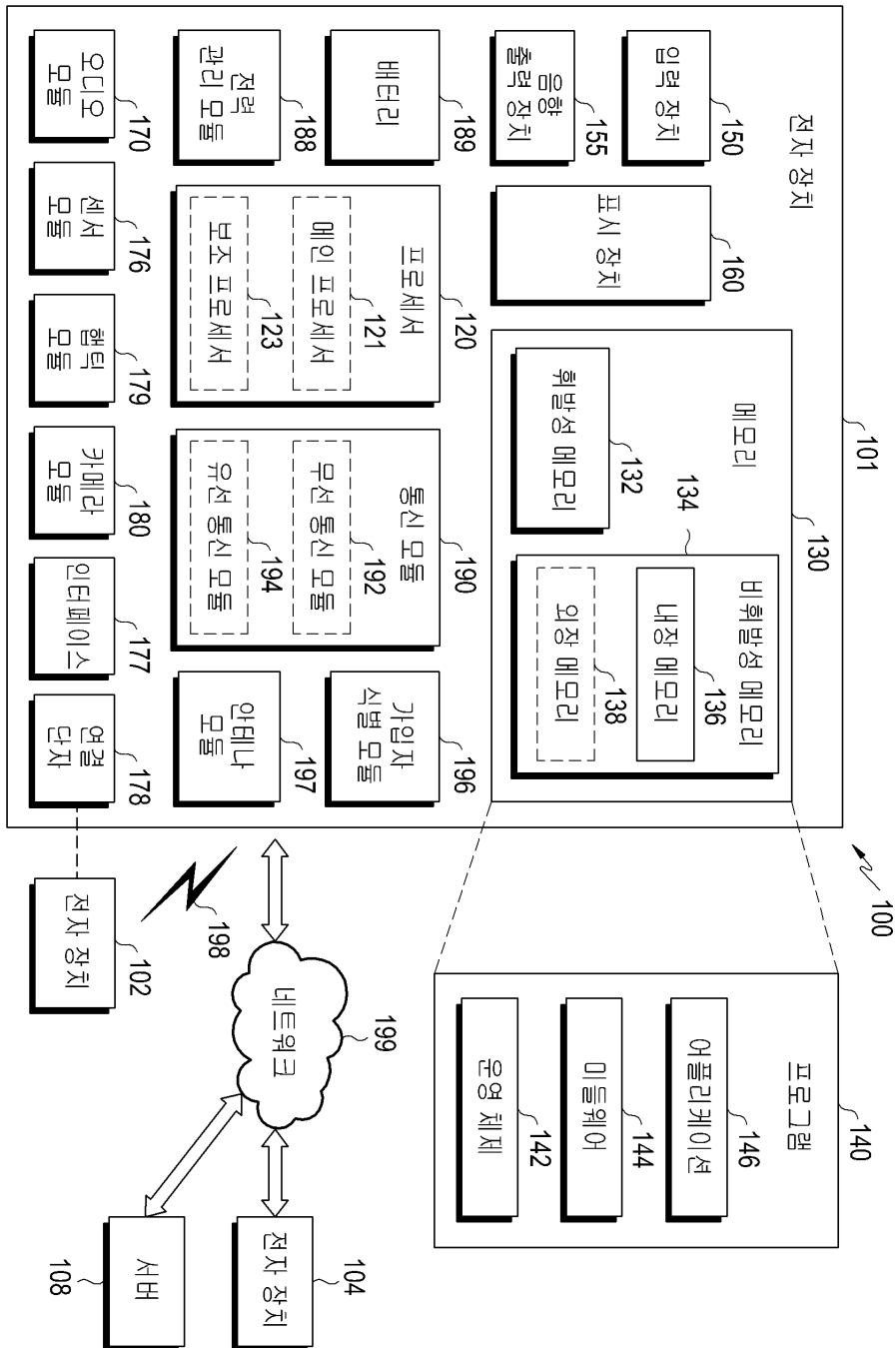
[0075] 본 문서의 다양한 실시예들은 기기(machine)(예: 전자 장치(101)) 의해 읽을 수 있는 저장 매체(storage medium)(예: 내장 메모리(136) 또는 외장 메모리(138))에 저장된 하나 이상의 명령어들을 포함하는 소프트웨어(예: 프로그램(140))로서 구현될 수 있다. 예를 들면, 기기(예: 전자 장치(101))의 프로세서(예: 프로세서(120))는, 저장 매체로부터 저장된 하나 이상의 명령어들 중 적어도 하나의 명령을 호출하고, 그것을 실행할 수 있다. 이것은 기기가 상기 호출된 적어도 하나의 명령어에 따라 적어도 하나의 기능을 수행하도록 운영되는 것을 가능하게 한다. 상기 하나 이상의 명령어들은 컴파일러에 의해 생성된 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, ‘비일시적’은 저장매체가 실재(tangible)하는 장치이고, 신호(signal)(예: 전자 기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장매체에 반영구적으로 저장되는 경우와 임시적으로 저장되는 경우를 구분하지 않는다.

[0076] 일실시예에 따르면, 본 문서에 개시된 다양한 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로 배포되거나, 또는 어플리케이션 스토어(예: 플레이 스토어™)를 통해 또는 두개의 사용자 장치들(예: 스마트폰들) 간에 직접, 온라인으로 배포(예: 다운로드 또는 업로드)될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 기기로 읽을 수 있는 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

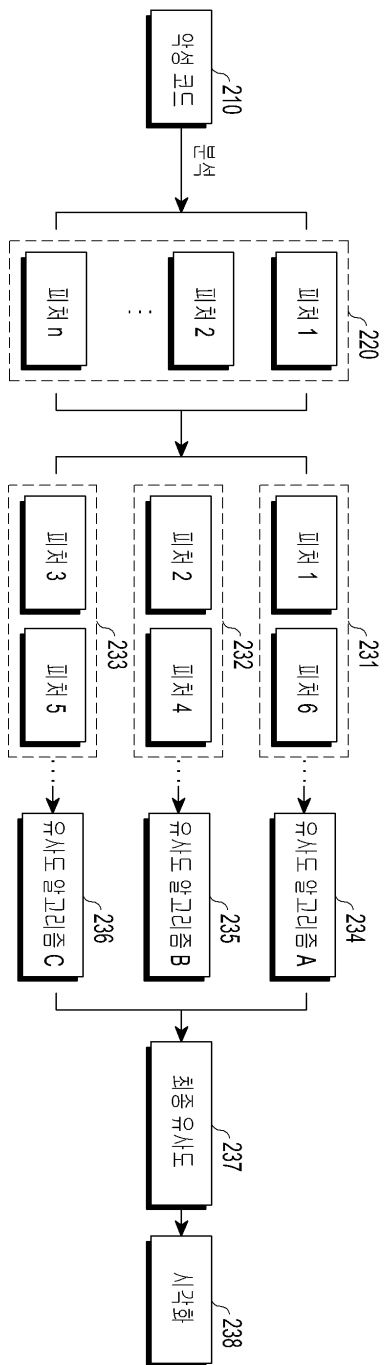
[0077] 다양한 실시예들에 따르면, 상기 기술한 구성요소들의 각각의 구성요소(예: 모듈 또는 프로그램)는 단수 또는 복수의 개체를 포함할 수 있다. 다양한 실시예들에 따르면, 전술한 해당 구성요소들 중 하나 이상의 구성요소들 또는 동작들이 생략되거나, 또는 하나 이상의 다른 구성요소들 또는 동작들이 추가될 수 있다. 대체적으로 또는 추가적으로, 복수의 구성요소들(예: 모듈 또는 프로그램)은 하나의 구성요소로 통합될 수 있다. 이런 경우, 통합된 구성요소는 상기 복수의 구성요소들 각각의 구성요소의 하나 이상의 기능들을 상기 통합 이전에 상기 복수의 구성요소들 중 해당 구성요소에 의해 수행되는 것과 동일 또는 유사하게 수행할 수 있다. 다양한 실시예들에 따르면, 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적으로, 병렬적으로, 반복적으로, 또는 휴리스틱하게 실행되거나, 상기 동작들 중 하나 이상이 다른 순서로 실행되거나, 생략되거나, 또는 하나 이상의 다른 동작들이 추가될 수 있다.

도면

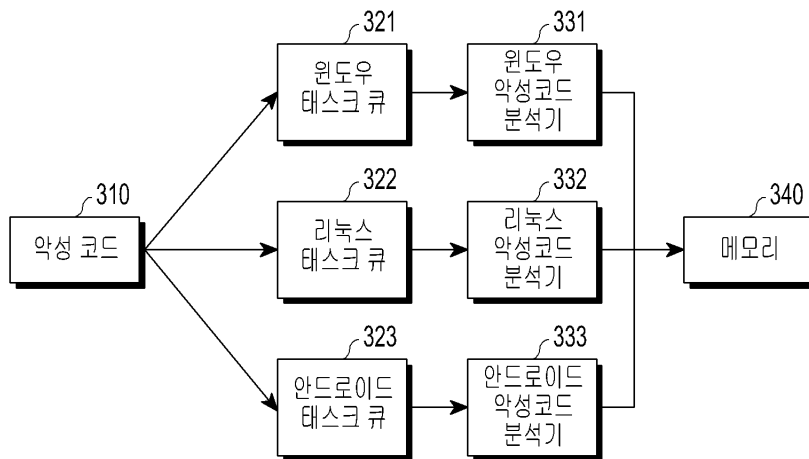
도면1



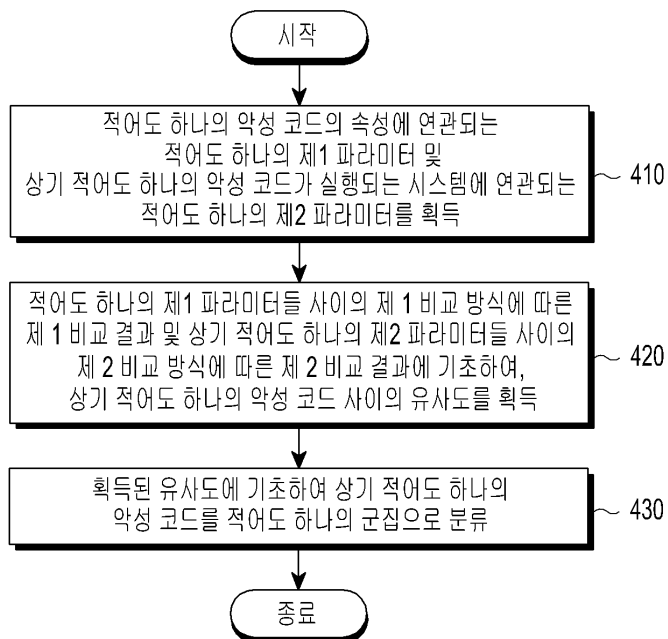
도면2



도면3



도면4



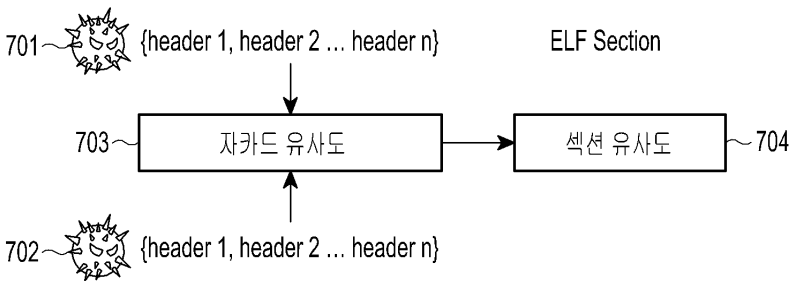
| ⌘ ID | Section                                                                                            |
|------|----------------------------------------------------------------------------------------------------|
| 1    | [".bss";comment,"ctors","data","ctors","eh_frame","fini","gcc_except_table","gnu.attributes","g... |
| 2    | [".bss";comment,"ctors","data","data.rel.ro","debug_abbrev","debug_aranges","debug_info","de...    |
| 3    | [".bss";comment,"ctors","data","ctors","eh_frame","fini","got","jcr","mdebug.abi32","pdr...        |
| 4    | [".bss";comment,"ctors","data","dynamic","dynsym","eh_frame","eh_frame_hdr","fini","fini_arr...    |
| 5    | [".bss";comment,"ctors","data","data.rel.ro","ctors","eh_frame","fini","gcc_except_table","got...  |
| 6    | [".bss";comment,"ctors","data","data.rel.ro","debug_abbrev","debug_aranges","debug_frame","d...    |

도면5

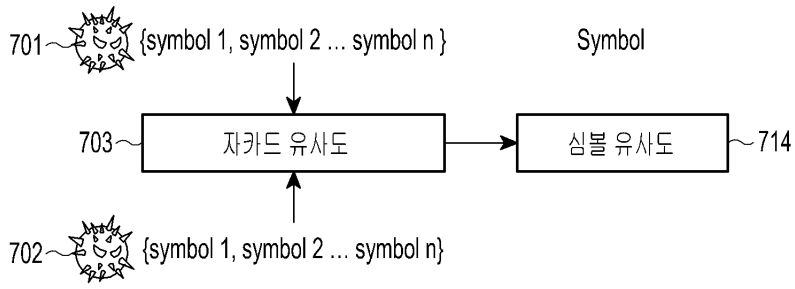
도면6

| ⌘ ID | CALLS                                                                                                                 |
|------|-----------------------------------------------------------------------------------------------------------------------|
| 1    | [\"GetSystemTimeAsFileTime\", \"SetUnhandledExceptionFilter\", \"FindFirstFileExW\", \"NtClose\", \"FindFirstFileE... |
| 2    | [\"FindFirstExW\", \"NtDelayExecution\", \"GetTempPathW\", \"FindFirstExW\", \"NtDelayExecution\", \"LdrG...          |
| 3    | [\"GetSystemTimeAsFileTime\", \"SetUnhandledExceptionFilter\", \"NtOpenThread\", \"LdrGetDllHandle\", \"LdrGet...     |
| 4    | [\"GetSystemTimeAsFileTime\", \"SetUnhandledExceptionFilter\", \"NtClose\", \"NtOpenKey\", \"NtOpenKey\", \"NtC...    |
| 5    | [\"GetSystemTimeAsFileTime\", \"NtClose\", \"NtOpenKey\", \"NtQueryValueKey\", \"NtClose\", \"NtOpenKey\", \"NtQ...   |
| 6    | [\"InternetOpenA\", \"LdrGetProcedureAddress\", \"ReadProcessMemory\", \"ReadProcessMemory\", \"ReadProc...           |

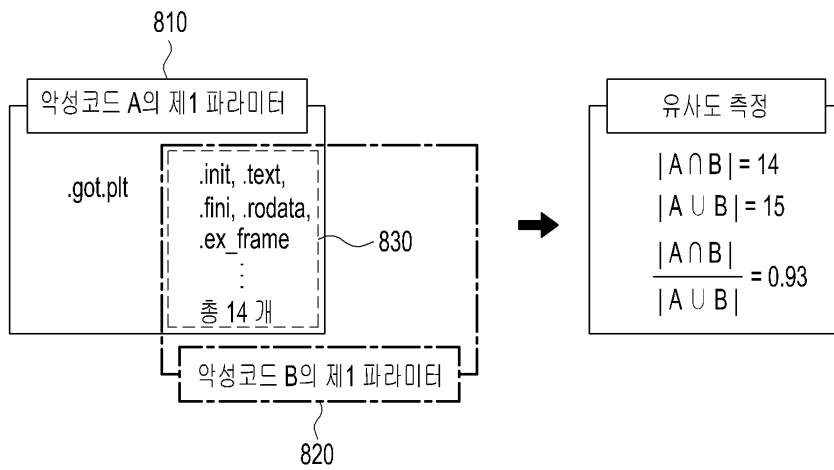
도면7a



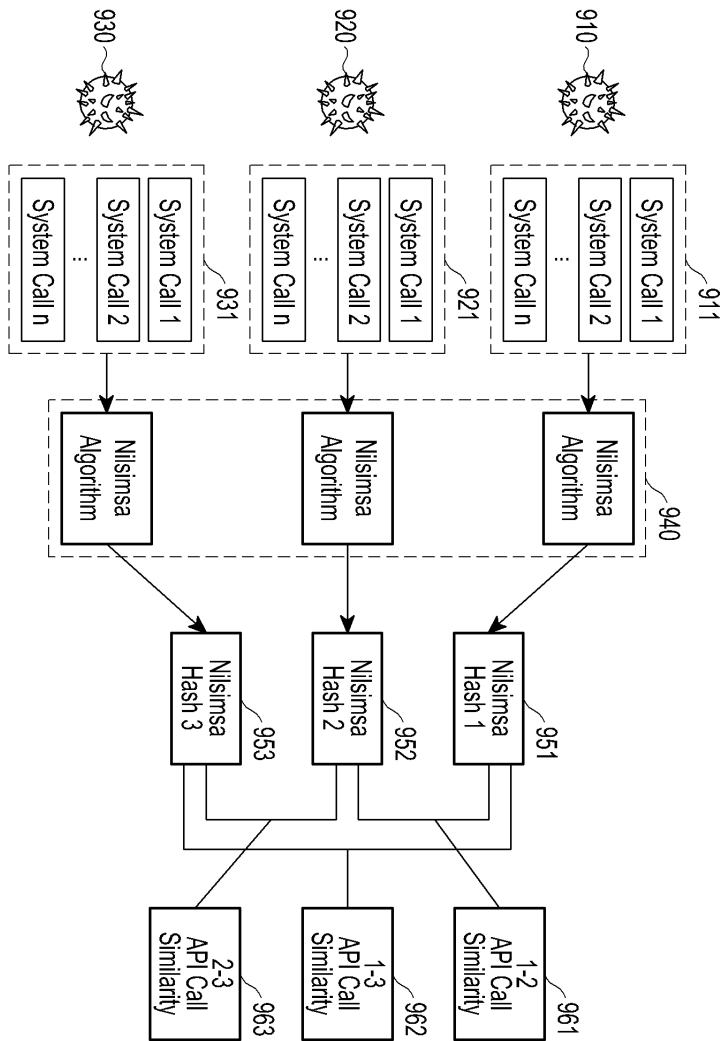
도면7b



도면8



도면9



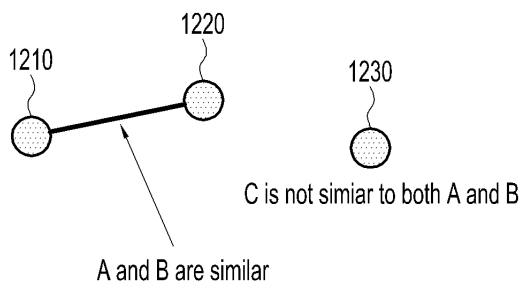
도면10

| 악성코드 ID | 진단명 (Kaspersky)             | Nilsinsa Hash (Hex)                                              | 유사도    |
|---------|-----------------------------|------------------------------------------------------------------|--------|
| 악성코드 1  | Trojan-Ransom.Win32.Wanna.m | 3e2be5319bc0fa3d35911dfd7011c01fa82c06080c40eb0e431cd5033bd7173e | 기준     |
| 악성코드 2  | rojan-Random.Win32.Wanna.m  | 3e2be5319bc0fa3d35911dfd7811c01ea82c06080c40eb0e431cd5033bd7163e | 0.9883 |
| 악성코드 3  | HEUR:Trojan.Win32.Generic   | 287a69741b4176373f8054f4520288192831040445a189e34a9dc129fb03e28  | 0.6758 |
| 악성코드 4  | Backdoor.Win32.GF.j         | 49524a00c223293c8e0e22aea6181e8d1be2807f28c44280523660508b231640 | 0.4805 |

도면11

| ID 1 | ID 2 | APIcall_Similarity | symbol_Similarity | section_Similarity | all_Similarity |
|------|------|--------------------|-------------------|--------------------|----------------|
| 4    | 2    | 0.726562           | 0.013082          | 0.307692           | 0.349112       |
| 5    | 2    | 0.84375            | 0.322629          | 0.710526           | 0.625635       |
| 5    | 4    | 0.678688           | 0.007248          | 0.357143           | 0.348026       |
| 6    | 2    | 0.71875            | 0.32206           | 0.775              | 0.60527        |
| 6    | 4    | 0.648438           | 0.015663          | 0.3125             | 0.325534       |
| 6    | 5    | 0.671875           | 0.223036          | 0.666667           | 0.520526       |
| 8    | 2    | 0.761719           | 0.087811          | 0.394737           | 0.414756       |
| 8    | 4    | 0.691406           | 0.028986          | 0.40625            | 0.375547       |

도면12



도면13

