



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0090247
(43) 공개일자 2023년06월21일

- | | |
|---|--|
| <p>(51) 국제특허분류(Int. Cl.)
G06F 21/56 (2013.01)</p> <p>(52) CPC특허분류
G06F 21/562 (2013.01)</p> <p>(21) 출원번호 10-2022-0170106</p> <p>(22) 출원일자 2022년12월07일
심사청구일자 2022년12월07일</p> <p>(30) 우선권주장
1020210178386 2021년12월14일 대한민국(KR)</p> | <p>(71) 출원인
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)</p> <p>(72) 발명자
김형식
대전광역시 유성구 대학로 99, 공대 5호관 618호 (궁동)</p> <p>조우진
대전광역시 유성구 대학로 99, 공대 5호관 604호 (궁동)</p> <p>(74) 대리인
제일특허법인(유)</p> |
|---|--|

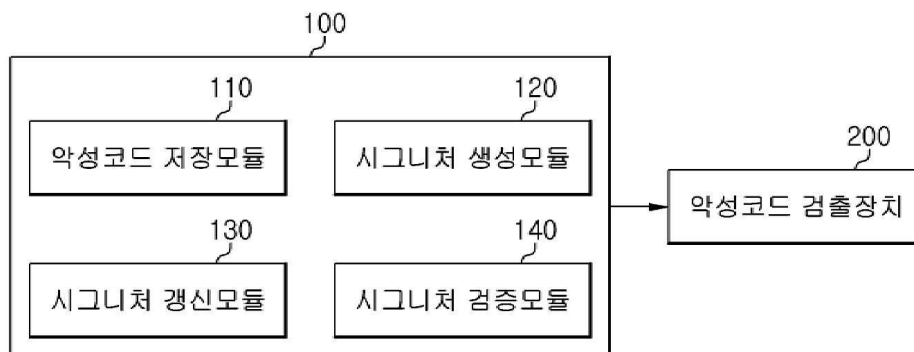
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 특정 악성코드에 최적화된 시그니처 생성방법 및 장치

(57) 요약

입력된 악성코드에 대한 최적화된 시그니처(signature)를 생성하는 방법 및 장치가 제공된다. 시그니처 생성방법은, 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성하고, 악성코드에 대응되는 패밀리 악성코드에 기초하여 시그니처를 갱신하고, 시그니처의 기준 특징값에 기초하여 갱신된 시그니처의 재갱신 여부를 판단한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	1711125781
과제번호	2019-0-01343-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	융합보안핵심인재양성사업
기 여 율	1/1
과제수행기관명	한국인터넷진흥원
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

기 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성하는 단계;

상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코드에 기초하여 시그니처를 갱신하는 단계; 및

갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함하는 시그니처 생성방법.

청구항 2

제1항에 있어서,

상기 시그니처를 생성하는 단계는,

상기 악성코드에서 다수의 초기 특징값을 추출하는 단계;

상기 다수의 기존악성코드 중에서 상기 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출하여 상기 다수의 초기 특징값 각각에 매칭하는 단계;

기 설정된 제1임계값에 기초하여 상기 다수의 초기 특징값 중 일부를 선별하는 단계; 및

선별된 초기 특징값을 포함하는 상기 시그니처를 생성하는 단계를 포함하는 시그니처 생성방법.

청구항 3

제2항에 있어서,

상기 다수의 초기 특징값 중 일부를 선별하는 단계는,

상기 다수의 초기 특징값 각각에 매칭된 상기 하나 이상의 기존악성코드의 개수가 큰 순서로 상기 다수의 초기 특징값 각각을 정렬하는 단계; 및

상기 다수의 초기 특징값 중에서 상기 하나 이상의 기존악성코드의 개수가 상기 제1임계값 미만인 하나 이상의 초기 특징값을 선별하는 단계를 포함하는 시그니처 생성방법.

청구항 4

제1항에 있어서,

상기 시그니처를 갱신하는 단계는,

상기 악성코드에서 추출한 패밀리 라벨에 기초하여 상기 하나 이상의 패밀리 악성코드를 추출하는 단계;

추출된 상기 하나 이상의 패밀리 악성코드 각각과 상기 시그니처 간 일치율을 판단하는 단계;

판단 결과에 따라 상기 하나 이상의 패밀리 악성코드 중에서 상기 시그니처와 최상의 일치율을 갖는 하나의 패밀리 악성코드를 추출하는 단계;

추출된 상기 하나의 패밀리 악성코드와 상기 시그니처 간 공통 특징값을 추출하는 단계; 및

상기 시그니처와 상기 공통 특징값 각각을 갱신 전 시그니처와 갱신된 시그니처로 출력하는 단계를 포함하는 시

그니처 생성방법.

청구항 5

제4항에 있어서,

상기 하나의 패밀리 악성코드를 추출하는 단계는,

상기 시그니처의 특징값과 전체 일치하는 특징값을 갖는 패밀리 악성코드를 제외한 나머지 패밀리 악성코드 중에서 상기 최상의 일치율을 갖는 상기 하나의 패밀리 악성코드를 추출하는 단계인 시그니처 생성방법.

청구항 6

제1항에 있어서,

상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계는,

기 설정된 제2임계값에 기초하여 갱신 전 시그니처의 다수의 특징값 중 일부를 상기 하나 이상의 기준 특징값으로 추출하는 단계;

상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 비율을 산출하는 단계; 및

상기 비율을 기 설정된 제3임계값과 비교하고, 비교 결과에 따라 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함하는 시그니처 생성방법.

청구항 7

제6항에 있어서,

상기 기준 특징값으로 추출하는 단계는,

상기 갱신 전 시그니처의 상기 다수의 특징값 중에서 상기 제2임계값 이하의 개수를 상기 기준 특징값으로 추출하는 단계인 시그니처 생성방법.

청구항 8

제6항에 있어서,

상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계는,

상기 갱신된 시그니처의 상기 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 상기 제3임계값 이상이면, 상기 갱신된 시그니처의 재 갱신을 판단하고,

상기 갱신된 시그니처의 상기 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 상기 제3임계값 미만이면, 상기 갱신 전 시그니처를 상기 악성코드의 시그니처로 출력하는 단계인 시그니처 생성방법.

청구항 9

제1항에 있어서,

상기 악성코드에서 패밀리 라벨 및 다수의 초기 특징값을 추출하는 단계를 더 포함하는 시그니처 생성방법.

청구항 10

다수의 기존악성코드가 저장된 악성코드 저장모듈;

외부에서 악성코드가 입력되면, 상기 다수의 기존악성코드에 기초하여 상기 악성코드에 대한 시그니처를 생성하여 출력하는 시그니처 생성모듈;

상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코드를 추출하고, 추출된 상기 패밀리 악성코드에 기초하여 상기 시그니처를 갱신하여 갱신 전 시그니처와 갱신된 시그니처를 출력하는 시그니처 갱신모듈; 및

상기 갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 시그니처 검증모듈을 포함하는 시그니처 생성장치.

청구항 11

제10항에 있어서,

상기 시그니처 생성모듈은,

상기 악성코드를 분석하여 패밀리 라벨 및 다수의 초기 특징값을 추출하는 악성코드 분석부;

기 설정된 제1임계값에 기초하여 상기 다수의 초기 특징값 중 일부를 선별하여 출력하는 특징값 선별부; 및

선별된 제1특징값을 포함하는 상기 시그니처를 생성하는 시그니처 생성부를 포함하는 시그니처 생성장치.

청구항 12

제11항에 있어서,

상기 특징값 선별부는,

상기 다수의 기존악성코드 중에서 상기 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출하여 상기 다수의 초기 특징값 각각에 매칭하고, 상기 다수의 초기 특징값 각각에 매칭된 상기 하나 이상의 기존악성코드의 개수가 큰 순서로 상기 다수의 초기 특징값 각각을 정렬하며, 정렬된 상기 다수의 초기 특징값 중에서 상기 하나 이상의 기존악성코드의 개수가 상기 제1임계값 미만인 하나 이상의 특징값을 선별하는 시그니처 생성장치.

청구항 13

제10항에 있어서,

상기 시그니처 갱신모듈은,

상기 악성코드에서 추출한 패밀리 라벨에 기초하여 상기 다수의 기존악성코드 중에서 상기 하나 이상의 패밀리 악성코드를 추출하는 패밀리 악성코드 추출부;

추출된 상기 하나 이상의 패밀리 악성코드 각각과 시그니처 간 일치율을 판단하고, 판단 결과에 따라 상기 하나 이상의 패밀리 악성코드 중에서 상기 시그니처와 최상의 일치율을 갖는 하나의 패밀리 악성코드와 상기 시그니처 간 공통 특징값을 추출하는 공통 특징값 추출부; 및

상기 공통 특징값에 기초하여 상기 시그니처를 갱신하는 시그니처 갱신부를 포함하는 시그니처 생성장치.

청구항 14

제13항에 있어서,

상기 공통 특징값 추출부는,

상기 시그니처의 특징값과 전체 일치하는 특징값을 갖는 패밀리 악성코드를 제외한 나머지 패밀리 악성코드 중에서 상기 최상의 일치율을 갖는 상기 하나의 패밀리 악성코드를 추출하는 시그니처 생성장치.

청구항 15

제10항에 있어서,

상기 시그니처 검증모듈은,

갱신 전 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값을 추출하고, 상기 갱신된 시그니처의 다수의 특징값 중 상기 하나 이상의 기준 특징값이 포함된 비율을 산출하는 특징값 비율산출부; 및

산출된 상기 비율에 기초하여 상기 갱신된 시그니처의 재 갱신 여부를 판단하고, 판단 결과에 따라 상기 시그니처 갱신모듈의 동작을 제어하는 시그니처 갱신제어부를 포함하는 시그니처 생성장치.

청구항 16

제15항에 있어서,

상기 특징값 비율산출부는,

기 설정된 제2임계값에 기초하여 상기 갱신 전 시그니처의 다수의 특징값 중 일부를 상기 하나 이상의 기준 특징값으로 추출하고, 상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율을 산출하는 시그니처 생성장치.

청구항 17

제16항에 있어서,

상기 특징값 비율산출부는,

상기 갱신 전 시그니처의 상기 다수의 특징값 중에서 대응되는 기존악성코드의 개수가 적은 순서대로 상기 제2 임계값 이하의 개수를 상기 기준 특징값으로 추출하는 시그니처 생성장치.

청구항 18

제15항에 있어서,

상기 시그니처 갱신제어부는,

상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 기 설정된 제3임계값 이상이면 상기 갱신된 시그니처의 재 갱신을 판단하여 상기 시그니처 갱신모듈의 동작을 제어하고,

상기 비율이 상기 제3임계값 미만이면 상기 갱신 전 시그니처를 상기 악성코드의 시그니처로 출력하는 시그니처 생성장치.

청구항 19

컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 기록매체로서,

상기 컴퓨터 프로그램은,

기 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성하는 단계;

상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코드에 기초하여 상기 시그니처를 갱신하는 단계; 및

상기 갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함하는 시그니처 생성방법을 수행하기 위한 명령어를 포함하는 컴퓨터 판독 가능한 기록매체.

청구항 20

컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은,

기 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성하는 단계;

상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코드에 기초하여 상기 시그니처를 갱신하는 단계; 및

상기 갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함하는 시그니처 생성방법을 수행하기 위한 명령어를 포함하는 기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 외부에서 입력된 특정 악성코드에 대한 최적화된 시그니처(signature)를 생성하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 기업 또는 연구소 등의 네트워크에 악성코드가 침입된 경우, 해당 기업 또는 연구소의 보안담당자는 침입한 악성코드에 대한 시그니처를 생성하여 악성코드로 인한 피해를 방지한다.

[0003] 그러나, 악성코드의 시그니처를 생성하기 위해서는 해당 악성코드에 포함된 다수의 정보, 예컨대 악성코드의 특징값 중에서 어떠한 정보가 유의미한 정보인지를 파악하기 어려워 악성코드에 대해 부정확한 시그니처가 생성되고, 이로 인해 악성코드에 의한 피해 방지의 효율이 저하된다.

[0004] 기존에는 침입한 악성코드 외에 다수의 악성코드를 수집하고, 이들에 대한 유사도 기반의 클러스터링(clustering)을 통해 그룹을 구성하고, 각 그룹별로 공통의 정보를 추출하여 그룹별 시그니처를 생성하는 방식이 사용되고 있다.

[0005] 그러나, 기존의 시그니처 생성방식에서는 악성코드에 대한 최적의 시그니처를 생성하기 위해 많은 수의 악성코드들을 클러스터링하며, 이로 인해 클러스터링에 많은 시간이 소요되는 문제가 있었다.

[0006] 또한, 침입한 악성코드에 임의의 기능이 추가되거나 삭제된 경우에는, 기존의 유사도 기반 클러스터링은 악성코드와 유사한 변종 악성코드를 정확하게 식별하지 못하는 문제가 있었다.

[0007] 더욱이, 기존에는 클러스터링을 통해 하나의 그룹으로 묶인 악성코드들에서 공통의 특징을 시그니처로 생성하는 방식이며, 이로 인해 그룹으로 묶인 개별 악성코드의 경우에는 시그니처 생성 시 악성코드에 대한 유의미한 특징이 제거되는 문제가 있다.

[0008] 이러한 문제로 인해, 기존의 시그니처 생성방식에서는 침입한 악성코드에 대한 빠른 대처가 어렵고, 악성코드에 대한 부정확한 시그니처가 생성되어 악성코드에 의한 피해가 증가되고 있는 현실이다.

선행기술문헌

특허문헌

[0009] (특허문헌 0001) 한국등록특허 제10-1337216호(2013.12.05.)

발명의 내용

해결하려는 과제

[0010] 본 발명은 외부에서 입력된 특정 악성코드에 대한 최적화된 시그니처를 생성하는 방법 및 장치를 제공하고자 하는 데 있다.

과제의 해결 수단

[0011] 본 발명의 일 실시예에 따른 시그니처 생성방법은, 기 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성하는 단계; 상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코드에 기초하여 시그니처를 갱신하는 단계; 및 갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함한다.

[0012] 상기 시그니처를 생성하는 단계는, 상기 악성코드에서 다수의 초기 특징값을 추출하는 단계; 상기 다수의 기존악성코드 중에서 상기 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출하여 상기 다수의 초기 특징값 각각에 매칭하는 단계; 기 설정된 제1임계값에 기초하여 상기 다수의 초기 특징값 중 일부를 선별하는 단계; 및 선별된 초기 특징값을 포함하는 상기 시그니처를 생성하는 단계를 포함한다.

[0013] 상기 다수의 초기 특징값 중 일부를 선별하는 단계는, 상기 다수의 초기 특징값 각각에 매칭된 상기 하나 이상의 기존악성코드의 개수가 큰 순서로 상기 다수의 초기 특징값 각각을 정렬하는 단계; 및 상기 다수의 초기 특징값 중에서 상기 하나 이상의 기존악성코드의 개수가 상기 제1임계값 미만인 하나 이상의 초기 특징값을 선별하는 단계를 포함한다.

[0014] 상기 시그니처를 갱신하는 단계는, 상기 악성코드에서 추출한 패밀리 라벨에 기초하여 상기 하나 이상의 패밀리 악성코드를 추출하는 단계; 추출된 상기 하나 이상의 패밀리 악성코드 각각과 시그니처 간 일치율을 판단하는 단계; 판단 결과에 따라 상기 하나 이상의 패밀리 악성코드 중에서 시그니처와 최상의 일치율을 갖는 하나의 패밀리 악성코드를 추출하는 단계; 추출된 상기 하나의 패밀리 악성코드와 시그니처 간 공통 특징값을 추출하는 단계; 및 상기 시그니처와 상기 공통 특징값 각각을 갱신 전 시그니처와 갱신된 시그니처로 생성하는 단계를 포함한다.

[0015] 상기 하나의 패밀리 악성코드를 추출하는 단계는, 시그니처의 특징값과 전체 일치하는 특징값을 갖는 패밀리 악성코드를 제외한 나머지 패밀리 악성코드 중에서 상기 최상의 일치율을 갖는 상기 하나의 패밀리 악성코드를 추출하는 단계이다.

[0016] 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계는, 기 설정된 제2임계값에 기초하여 상기 갱신 전 시그니처의 다수의 특징값 중 일부를 상기 하나 이상의 기준 특징값으로 추출하는 단계; 상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 비율을 산출하는 단계; 및 상기 비율을 기 설정된 제3임계값과 비교하고, 비교 결과에 따라 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계를 포함한다.

[0017] 상기 기준 특징값으로 추출하는 단계는, 상기 갱신 전 시그니처의 상기 다수의 특징값 중에서 대응되는 기존악성코드의 개수가 적은 순서대로 상기 제2임계값만큼의 개수를 상기 기준 특징값으로 추출하는 단계이다.

[0018] 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 단계는, 상기 갱신된 시그니처의 상기 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 상기 제3임계값 이상이면, 상기 갱신된 시그니처의 재 갱신을 판단하고, 상기 갱신된 시그니처의 상기 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 상기 제3임계값 미만이면, 상기 갱신 전 시그니처를 상기 악성코드의 시그니처로 출력하는 단계이다.

[0019] 본 실시예의 시그니처 생성방법은 상기 악성코드에서 패밀리 라벨 및 다수의 초기 특징값을 추출하는 단계를 더 포함한다.

[0020] 본 발명의 일 실시예에 따른 시그니처 생성장치는, 다수의 기존악성코드가 저장된 악성코드 저장모듈; 외부에서 악성코드가 입력되면, 상기 다수의 기존악성코드에 기초하여 상기 악성코드에 대한 시그니처를 생성하여 출력하는 시그니처 생성모듈; 상기 다수의 기존악성코드 중에서 상기 악성코드에 대응되는 하나 이상의 패밀리 악성코

드를 추출하고, 추출된 상기 패밀리 악성코드에 기초하여 시그니처를 갱신하여 갱신 전 시그니처와 갱신된 시그니처를 출력하는 시그니처 갱신모듈; 및 상기 갱신 전 시그니처의 하나 이상의 기준 특징값에 기초하여 상기 갱신된 시그니처의 재 갱신 여부를 판단하는 시그니처 검증모듈을 포함한다.

[0021] 상기 시그니처 생성모듈은, 상기 악성코드를 분석하여 패밀리 라벨 및 다수의 초기 특징값을 추출하는 악성코드 분석부; 기 설정된 제1임계값에 기초하여 상기 다수의 초기 특징값 중 일부를 선별하여 출력하는 특징값 선별부; 및 선별된 특징값을 포함하는 시그니처를 생성하는 시그니처 생성부를 포함한다.

[0022] 상기 특징값 선별부는, 상기 다수의 기존악성코드 중에서 상기 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출하여 상기 다수의 초기 특징값 각각에 매칭하고, 상기 다수의 초기 특징값 각각에 매칭된 상기 하나 이상의 기존악성코드의 개수가 큰 순서로 상기 다수의 초기 특징값 각각을 정렬하며, 정렬된 상기 다수의 초기 특징값 중에서 상기 하나 이상의 기존악성코드의 개수가 상기 제1임계값 미만인 하나 이상의 특징값을 선별한다.

[0023] 상기 시그니처 갱신모듈은, 상기 악성코드에서 추출한 패밀리 라벨에 기초하여 상기 다수의 기존악성코드 중에서 상기 하나 이상의 패밀리 악성코드를 추출하는 패밀리 악성코드 추출부; 추출된 상기 하나 이상의 패밀리 악성코드 각각과 시그니처 간 일치율을 판단하고, 판단 결과에 따라 상기 하나 이상의 패밀리 악성코드 중에서 시그니처와 최상의 일치율을 갖는 하나의 패밀리 악성코드와 시그니처 간 공통 특징값을 추출하는 공통 특징값 추출부; 및 상기 공통 특징값에 기초하여 시그니처를 갱신하는 시그니처 갱신부를 포함한다.

[0024] 상기 공통 특징값 추출부는, 시그니처의 특징값과 전체 일치하는 특징값을 갖는 패밀리 악성코드를 제외한 나머지 패밀리 악성코드 중에서 상기 최상의 일치율을 갖는 상기 하나의 패밀리 악성코드를 추출한다.

[0025] 상기 시그니처 갱신부는, 갱신하기 전의 시그니처와 갱신한 후의 시그니처 각각을 갱신 전 시그니처와 갱신된 시그니처로 생성한다.

[0026] 상기 시그니처 검증모듈은, 상기 갱신 전 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값을 추출하고, 상기 갱신된 시그니처의 다수의 특징값 중 상기 하나 이상의 기준 특징값이 포함된 비율을 산출하는 특징값 비율산출부; 및 산출된 상기 비율에 기초하여 상기 갱신된 시그니처의 재 갱신 여부를 판단하고, 판단 결과에 따라 상기 시그니처 갱신모듈의 동작을 제어하는 시그니처 갱신제어부를 포함한다.

[0027] 상기 특징값 비율산출부는, 기 설정된 제2임계값에 기초하여 상기 갱신 전 시그니처의 다수의 특징값 중 일부를 상기 하나 이상의 기준 특징값으로 추출하고, 상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율을 산출한다.

[0028] 상기 특징값 비율산출부는, 상기 갱신 전 시그니처의 상기 다수의 특징값 중에서 대응되는 기존악성코드의 개수가 적은 순서대로 상기 제2임계값 이하의 개수를 상기 기준 특징값으로 추출한다.

[0029] 상기 시그니처 갱신제어부는, 상기 갱신된 시그니처의 다수의 특징값 중에서 상기 하나 이상의 기준 특징값이 포함된 상기 비율이 기 설정된 제3임계값 이상이면 상기 갱신된 시그니처의 재 갱신을 판단하여 상기 시그니처 갱신모듈의 동작을 제어하고, 상기 비율이 상기 제3임계값 미만이면 상기 갱신 전 시그니처를 상기 악성코드의 시그니처로 출력한다.

발명의 효과

[0030] 본 발명은 역색인 저장구조로 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 빠르게 생성할 수 있다.

[0031] 또한, 본 발명은 갱신 전 시그니처와 갱신 후 시그니처 간 특징값 동일비율에 따라 갱신 후 시그니처의 재 갱신을 제어함으로써, 입력된 악성코드에 대한 최적의 시그니처를 생성할 수 있다.

도면의 간단한 설명

[0032] 도 1은 본 발명의 실시예에 따른 시그니처 생성장치를 나타내는 블록도이다.

도 2는 도 1의 시그니처 생성모듈의 블록도이다.

도 3은 도 1의 시그니처 갱신모듈의 블록도이다.

도 4는 도 1의 시그니처 검증모듈의 블록도이다.

도 5는 본 발명의 실시예에 따른 시그니처 생성방법을 나타내는 순서도이다.

도 6 내지 도 8은 본 발명의 실시예에 따른 시그니처 생성방법을 구체적으로 나타내는 순서도들이다.

발명을 실시하기 위한 구체적인 내용

- [0033] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0034] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 그리고 후술되는 용어들은 본 발명의 실시예에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0035] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시 예에 대하여 상세하게 설명한다.
- [0036] 도 1은 본 발명의 실시예에 따른 시그니처 생성장치를 나타내는 블록도이다.
- [0037] 도면을 참조하면, 본 실시예의 시그니처 생성장치(100)는 외부에서 임의의 악성코드가 입력되면, 상기 악성코드에 대한 시그니처(signature)를 생성할 수 있다. 시그니처 생성장치(100)는 생성된 시그니처를 외부장치, 예컨대 악성코드검출장치(200)로 전송할 수 있다.
- [0038] 이에, 악성코드 검출장치(200)는 수신된 악성코드의 시그니처를 내부의 악성코드 데이터베이스(미도시)에 저장하고, 이러한 데이터베이스에 저장된 다수의 시그니처를 이용하여 네트워크를 통해 유입되는 다수의 코드들로부터 악성코드의 포함여부를 검출할 수 있다.
- [0039] 여기서, 시그니처 생성장치(100)로 입력되는 임의의 악성코드는 기존에 검출되었던 악성코드가 아닌 신규 악성코드이거나 또는 기존악성코드에서 특정 기능이 추가되거나 또는 삭제된 변종 악성코드일 수 있다.
- [0040] 또한, 시그니처 생성장치(100)가 생성하는 시그니처는 입력된 악성코드에서 추출한 데이터, 예컨대 하나 이상의 특징값의 집합으로 표현되는 악성코드 패턴일 수 있다.
- [0041] 또한, 본 실시예는 시그니처 생성장치(100)가 악성코드 검출장치(200)와 별개의 구성으로 설명하나, 본 발명은 이에 제한되지 않는다. 예컨대, 시그니처 생성장치(100)는 악성코드 검출장치(200) 내에 하나의 모듈로 포함될 수도 있다.
- [0042] 본 실시예의 시그니처 생성장치(100)는 악성코드 저장모듈(110), 시그니처 생성모듈(120), 시그니처 갱신모듈(130) 및 시그니처 검증모듈(140)을 포함할 수 있다.
- [0043] 악성코드 저장모듈(110)은 다수의 기존악성코드를 저장할 수 있다. 다수의 기존악성코드 각각은 하나 이상의 특징값의 집합을 포함할 수 있다. 또한, 다수의 기존악성코드 각각은 해당 악성코드에 대한 패밀리 라벨을 더 포함할 수 있다.
- [0044] 여기서, 패밀리 라벨은 악성코드가 해당하는 패밀리의 명칭이며, 악성코드 패밀리는 중복되는 특성을 공유하는 하나 이상의 악성코드의 집합일 수 있다. 이러한 악성코드의 패밀리 라벨은 백신의 진단명을 통해 수집할 수 있다. 하나의 예로 바이러스토탈(Virustotal) 서비스를 이용하면 입력된 악성코드에 대하여 약 70개 이상의 백신으로부터 진단명을 수집할 수 있다.
- [0045] 또한, 악성코드 저장모듈(110)은 아래 [표 1]과 같이, 다수의 기존악성코드를 역색인(inverted indexing) 구조로 저장할 수 있다.

표 1

ID	특징값	악성코드 hash
1	CreateDirectoryW+"dirpath_r": "C:\\\\Users\\W\\Administrator" "dirpath": "C:\\\\Users\\W\\Administrator"	b643d5ae9a8aecc5b092e7bd8ea6f43,d5159b90b87669c6fadd15ed3fc13ba1,f1264a968aa937344a9591be33409e69, ...

2	DrawTextExW+"string": "Get MD5 signature from:"	1c5549ed4f02f206f961c4e8e4bb63b6,9a00a22edc7aed4bfc ccca10b3f16c31
.	.	.
.	.	.
.	.	.

- [0047] [표 1]에 도시된 바와 같이, 본 실시예의 악성코드 저장모듈(110)은 다수의 특징값 각각에 대하여 하나 이상의 악성코드, 예컨대 악성코드 해시(hash)값이 매칭되도록 역색인 구조로 다수의 기존악성코드를 저장할 수 있다.
- [0048] 여기서, [표 1]의 특징값에서 플러스(+) 기호 좌측은 API명칭이고, 플러스기호 우측은 API에 정의된 매개변수의 명칭 및 전달인자를 차례로 나타낸다.
- [0049] 이러한 다수의 기존악성코드에 대한 역색인 저장 구조로 인해, 추후 설명될 시그니처 생성모듈(120)이 악성코드 저장모듈(110)에서 특징값 및 대응되는 악성코드를 추출하는 속도가 향상될 수 있다.
- [0050] 시그니처 생성모듈(120)은 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성할 수 있다.
- [0051] 도 2는 도 1의 시그니처 생성모듈의 블록도이다.
- [0052] 도 1 및 도 2를 참조하면, 본 실시예의 시그니처 생성모듈(120)은 악성코드 분석부(121), 특징값 선별부(122) 및 시그니처 생성부(123)를 포함할 수 있다.
- [0053] 악성코드 분석부(121)는 입력된 악성코드를 분석하여 상기 악성코드에 대한 패밀리 라벨 및 다수의 특징값을 추출할 수 있다. 이하에서는, 설명의 편의를 위하여, 입력된 악성코드에서 추출되는 특징값은 초기 특징값으로 명명하기로 한다.
- [0054] 악성코드에서 추출되는 다수의 초기 특징값 각각은 정적 정보와 동적 정보를 포함할 수 있다. 여기서, 정적 정보는 악성코드 자체에서 추출되는 특징값이고, 동적 정보는 악성코드가 실제 실행되었을 때 추출되는 특징값일 수 있다.
- [0055] 이에, 본 실시예의 악성코드 분석부(121)는 입력된 악성코드를 가상 머신에서 실제 실행시킬 수 있는 소정의 프로그램, 예컨대 쿠쿠 샌드박스(cuckoo sandbox)를 통해 상기 악성코드로부터 다수의 초기 특징값을 추출할 수 있다.
- [0056] 특징값 선별부(122)는 악성코드로부터 추출된 다수의 초기 특징값 중 일부 개수의 초기 특징값을 선별하여 출력할 수 있다.
- [0057] 특징값 선별부(122)는 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드 중에서 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출할 수 있다. 특징값 선별부(122)는 다수의 초기 특징값 각각에 추출된 기존악성코드의 개수를 매칭하고, 기 설정된 제1임계값에 기초하여 다수의 초기 특징값 중 일부를 선별하여 출력할 수 있다.
- [0058] 여기서, 제1임계값은 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드 전체 개수에 대한 소정의 퍼센트(%) 비율일 수 있으며, 대략 50%일 수 있다. 예컨대, 다수의 기존악성코드의 총 개수가 3,054이면, 제1임계값은 대략 1,527일 수 있다.
- [0059] 이에, 특징값 선별부(122)는 각각에 하나 이상의 기존악성코드가 매칭된 다수의 특징값 중에서 제1임계값 이상의 악성코드가 매칭된 특징값을 제외한 나머지 특징값을 선별하여 출력할 수 있다.
- [0060] 시그니처 생성부(123)는 특징값 선별부(122)에서 출력된 하나 이상의 초기 특징값에 기초하여 이를 포함하는 악성코드에 대한 시그니처를 생성할 수 있다.
- [0061] 다시 도 1을 참조하면, 시그니처 갱신모듈(130)은 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드에 기초하여 시그니처 생성모듈(120)에서 생성된 악성코드에 대한 시그니처를 갱신할 수 있다.
- [0062] 도 3은 도 1의 시그니처 갱신모듈의 블록도이다.
- [0063] 도 1 및 도 3을 참조하면, 시그니처 갱신모듈(130)은 패밀리 악성코드 추출부(131), 공통 특징값 추출부(132)

및 시그니처 갱신부(133)를 포함할 수 있다.

- [0064] 패밀리 악성코드 추출부(131)는 앞서 시그니처 생성모듈(120)의 악성코드 분석부(121)에서 추출된 악성코드의 패밀리 라벨에 기초하여 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드 중에서 동일한 패밀리 라벨을 갖는 하나 이상의 기존악성코드를 패밀리 악성코드로 추출할 수 있다.
- [0065] 공통 특징값 추출부(132)는 시그니처의 특징값과 하나 이상의 패밀리 악성코드 각각의 특징값 간 일치율을 판단하고, 판단 결과에 따라 하나의 패밀리 악성코드를 추출할 수 있다.
- [0066] 여기서, 공통 특징값 추출부(132)는 패밀리 악성코드의 특징값과 시그니처의 특징값 간 일치 여부에 따른 일치율을 판단할 수 있다. 이때, 공통 특징값 추출부(132)는 하나 이상의 패밀리 악성코드 중에서 시그니처의 특징값과 완전하게 동일한, 예컨대 100% 일치하는 특징값을 갖는 패밀리 악성코드를 제외할 수 있다. 그리고, 공통 특징값 추출부(132)는 나머지 패밀리 악성코드 중에서 시그니처의 특징값과 최상의 일치율을 갖는 특징값을 포함하는 하나의 패밀리 악성코드를 추출할 수 있다.
- [0067] 이에, 공통 특징값 추출부(132)는 추출된 하나의 패밀리 악성코드에 포함된 다수의 특징값과 시그니처의 다수의 특징값에서 공통으로 나타나는 하나 이상의 특징값을 추출하고, 이를 공통 특징값으로 출력할 수 있다.
- [0068] 시그니처 갱신부(133)는 상기 시그니처와 상기 하나 이상의 공통 특징값 각각을 갱신 전 시그니처와 갱신된 시그니처로 생성할 수 있다.
- [0069] 시그니처 갱신부(133)는 상기 갱신 전 시그니처와 상기 갱신된 시그니처를 함께 출력할 수 있다.
- [0070] 다시 도 1을 참조하면, 시그니처 검증모듈(140)은 갱신 전 시그니처에서 추출되는 기준 특징값에 기초하여 갱신된 시그니처의 재 갱신 여부를 판단하고, 그에 따라 시그니처 갱신모듈(130)의 동작을 제어할 수 있다.
- [0071] 도 4는 도 1의 시그니처 검증모듈의 블록도이다.
- [0072] 도 1 및 도 4를 참조하면, 시그니처 검증모듈(140)은 특징값 비율산출부(141) 및 시그니처 갱신제어부(142)를 포함할 수 있다.
- [0073] 특징값 비율산출부(141)는 갱신 전 시그니처의 다수의 특징값 중에서 하나 이상의 기준 특징값을 추출할 수 있다.
- [0074] 여기서, 특징값 비율산출부(141)는 갱신 전 시그니처의 다수의 특징값 중에서 대응되는 기존악성코드의 개수가 적은 순서대로 제2임계값만큼의 일부의 특징값을 기준 특징값으로 추출할 수 있다.
- [0075] 이때, 제2임계값은 갱신 전 시그니처의 다수의 특징값 전체 개수에 대한 소정의 퍼센트(%) 비율일 수 있으며, 대략 20%일 수 있다.
- [0076] 예컨대, 갱신 전 시그니처의 다수의 특징값의 총 개수가 707이면, 제2임계값은 대략 141일 수 있고, 특징값 비율산출부(141)는 다수의 특징값 중에서 대응되는 기존악성코드의 개수가 적은 순서대로 141개의 특징값을 갱신 전 시그니처에 대한 기준 특징값으로 추출할 수 있다.
- [0077] 또한, 특징값 비율산출부(141)는 갱신된 시그니처의 다수의 특징값 중에서 기 추출된 기준 특징값이 포함된 비율을 산출할 수 있다.
- [0078] 시그니처 갱신제어부(142)는 특징값 비율산출부(141)에서 산출된 기준 특징값의 포함 비율에 기초하여 갱신된 시그니처의 재 갱신 여부를 판단할 수 있다.
- [0079] 이때, 시그니처 갱신제어부(142)는 기 설정된 제3임계값에 기초하여 기준 특징값의 포함 비율에 따라 갱신된 시그니처의 재 갱신 여부를 판단할 수 있다. 여기서, 제3임계값은 기준 특징값의 총 개수에 대한 소정의 퍼센트 비율일 수 있으며, 대략 80%일 수 있다.
- [0080] 여기서, 기 산출된 기준 특징값의 포함 비율이 제3임계값 이상이면, 다시 말해 갱신된 시그니처의 다수의 특징값에 기준 특징값이 80% 이상 포함되어 있으면, 시그니처 갱신제어부(142)는 갱신된 시그니처의 재 갱신을 판단할 수 있다. 이에, 시그니처 갱신부(133)는 시그니처 갱신모듈(130)에서 갱신된 시그니처를 재 갱신하도록 그 동작을 제어할 수 있다.
- [0081] 또한, 기 산출된 기준 특징값의 포함 비율이 제3임계값 미만이면, 시그니처 갱신제어부(142)는 갱신 전 시그니처를 시그니처 생성장치(100)로 입력된 악성코드에 대한 최종 시그니처, 예컨대 최적의 시그니처로 출력할 수

있다.

- [0082] 이와 같이, 본 실시예의 시그니처 생성장치(100)는 역색인 저장구조로 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 빠르게 생성할 수 있다.
- [0083] 또한, 본 실시예의 시그니처 생성장치(100)는 2개의 시그니처 간, 즉 갱신 전 시그니처와 갱신된 시그니처 간 특징값 동일비율에 따라 시그니처 갱신의 반복을 제어함으로써, 입력된 악성코드에 대한 최적의 시그니처를 생성할 수 있다.
- [0084] 도 5는 본 발명의 실시예에 따른 시그니처 생성방법을 나타내는 순서도이고, 도 6 내지 도 8은 본 발명의 실시예에 따른 시그니처 생성방법을 구체적으로 나타내는 순서도들이다.
- [0085] 도 5 및 도 6을 참조하면, 시그니처 생성모듈(120)은 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 생성할 수 있다(S10).
- [0086] 먼저, 시그니처 생성모듈(120)의 악성코드 분석부(121)는 입력된 악성코드를 분석하여 상기 악성코드의 패밀리 라벨 및 다수의 특징값, 예컨대 다수의 초기 특징값을 추출할 수 있다(S110).
- [0087] 다음으로, 특징값 선별부(122)는 기 저장된 다수의 기존악성코드 중에서 다수의 초기 특징값 각각을 포함하는 하나 이상의 기존악성코드를 추출할 수 있다(S120).
- [0088] 이어, 특징값 선별부(122)는 다수의 초기 특징값 각각에 추출된 기존악성코드의 개수를 매칭하고, 기 설정된 제 1임계값에 기초하여 다수의 초기 특징값 중 일부를 선별하여 출력할 수 있다(S130).
- [0089] 여기서, 특징값 선별부(122)는 다수의 초기 특징값 각각을 이에 매칭된 기존악성코드의 개수가 큰 순서로 정렬할 수 있다. 그리고, 정렬된 다수의 초기 특징값 중에서 매칭된 기존악성코드의 개수가 제1임계값 미만의 개수를 갖는 하나 이상의 초기 특징값을 선별하여 추출할 수 있다.
- [0090] 이때, 제1임계값은 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드의 총 개수에 대한 소정의 퍼센트 비율, 예컨대 대략 50%의 비율에 해당하는 값일 수 있다.
- [0091] 예컨대, 아래 [표 2]와 같이, 특징값 선별부(122)는 다수의 초기 특징값 각각에 이들을 포함하는 기존악성코드의 개수를 매칭할 수 있다. 그리고, 매칭된 기존악성코드의 개수가 큰 순서로 다수의 초기 특징값을 정렬할 수 있다.

표 2

[0092]

ID	초기 특징값	기존악성코드 개수
1	NtClose+"handle": masked	2925
2	RegCloseKey+"key_handle": masked	2562
..	.	.
.	.	.
.	.	.
25	GetFileType+"file_handle": masked	1559
26	NtProtectVirtualMemory+"process_identifier":masked "stack_dep_bypass": masked "stack_pivoted": masked "heap_dep_bypass": masked "length": masked "protection": masked "process_handle": masked "base_address": masked	1430
27	LdrGetDllHandle_"module_name": masked:"kernel32.dll" "stack_pivoted": masked "module_address":	1384
.	.	.
.	.	.
.	.	.
731	UuidCreate+"uuid": "{fc7ae18d-a45d-46ca-a289-864d213cd522}"	0
732	NtWriteFile+"file_handle": masked "filepath": "C:WWWUsersWWAdministratorWWAppDataaWWWLocalWWWTempWWWserverr.exe"	0

- [0093] 이때, 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드의 총 개수가 3,054개이면, 제1임계값에 따른 기존 악성코드의 개수는 대략 1,527개일 수 있다. 이에, 특징값 선별부(122)는 정렬된 다수의 초기 특징값 중에서 제1임계값 미만의 기존악성코드를 갖는 초기 특징값, 예컨대 [표 2]에서 26~732번까지의 707개의 초기 특징값을 선별하여 출력할 수 있다.
- [0094] 이어, 시그니처 생성부(123)는 선별된 하나 이상의 초기 특징값에 기초하여 이를 포함하는 악성코드의 시그니처를 생성할 수 있다(S140).
- [0095] 도 5 및 도 7을 참조하면, 시그니처 갱신모듈(130)은 다수의 기존악성코드 중에서 기 입력된 악성코드에 대응되는 하나 이상의 패밀리 악성코드에 기초하여 기 생성된 시그니처를 갱신하여 갱신된 시그니처를 출력할 수 있다(S20).
- [0096] 시그니처 갱신모듈(130)의 패밀리 악성코드 추출부(131)는 앞서 시그니처 생성모듈(120)의 악성코드 분석부(121)에서 추출된 악성코드의 패밀리 라벨에 기초하여 악성코드 저장모듈(110)에 저장된 다수의 기존악성코드 중에서 패밀리 라벨에 대응되는 하나 이상의 기존악성코드를 패밀리 악성코드로 추출할 수 있다(S210).
- [0097] 다음으로, 시그니처 갱신모듈(130)의 공통 특징값 추출부(132)는 시그니처와 하나 이상의 패밀리 악성코드 간 일치율을 판단하고, 판단된 일치율에 기초하여 하나의 패밀리 악성코드를 추출할 수 있다(S220).
- [0098] 구체적으로, 공통 특징값 추출부(132)는 패밀리 악성코드의 특징값과 시그니처의 특징값, 예컨대 선별된 초기 특징값 간 일치 여부에 따른 일치율을 판단할 수 있다.
- [0099] 이때, 공통 특징값 추출부(132)는 하나 이상의 패밀리 악성코드 중에서 시그니처의 특징값과 완전하게 동일한, 예컨대 100% 일치하는 특징값을 갖는 시그니처를 포함하는 패밀리 악성코드를 제외할 수 있다.
- [0100] 이에, 공통 특징값 추출부(132)는 나머지 패밀리 악성코드 중에서 시그니처의 특징값과 최상의 일치율을 갖는 특징값을 포함하는 시그니처를 추출하고, 추출된 시그니처에 대응되는 하나의 패밀리 악성코드를 추출할 수 있다.
- [0101] 계속해서, 공통 특징값 추출부(132)는 추출된 하나의 패밀리 악성코드의 특징값과 시그니처 간 공통 특징값을 추출할 수 있다(S230).
- [0102] 여기서, 공통 특징값 추출부(132)는 하나의 패밀리 악성코드에 포함된 다수의 특징값과 시그니처의 다수의 초기 특징값 중 공통으로 나타나는 하나 이상의 특징값을 추출할 수 있다. 공통 특징값 추출부(132)는 추출된 하나 이상의 특징값을 공통 특징값으로 출력할 수 있다.
- [0103] 다음으로, 시그니처 갱신부(133)는 추출된 공통 특징값에 기초하여 시그니처를 갱신하고, 갱신된 시그니처를 출력할 수 있다(S240).
- [0104] 이때, 시그니처 갱신부(133)는 갱신 전 시그니처와 갱신된 시그니처를 함께 출력할 수 있다.
- [0105] 도 5 및 도 8을 참조하면, 시그니처 검증모듈(140)은 갱신된 시그니처를 검증하여 재 갱신 여부를 판단할 수 있다. 시그니처 검증모듈(140)은 판단 결과에 따라 시그니처 갱신모듈(130)의 동작을 제어할 수 있다(S30).
- [0106] 먼저, 시그니처 검증모듈(140)의 특징값 비율산출부(141)는 기 설정된 제2임계값에 기초하여 갱신 전 시그니처의 정렬된 다수의 특징값 중에서 일부의 특징값을 기준 특징값으로 추출할 수 있다(S310).
- [0107] 이때, 제2임계값은 갱신 전 시그니처의 다수의 특징값의 총 개수에 대한 소정의 퍼센트 비율, 예컨대 대략 20%의 비율에 해당하는 값일 수 있다.
- [0108] 아래 [표 3]을 참조하면, 갱신 전 시그니처의 다수의 특징값, 예컨대 707개의 특징값은 각각에 매칭된 악성코드의 개수가 작은 순서로 정렬되어 있다.
- [0109] 또한, 제2임계값은 갱신 전 시그니처의 다수의 특징값의 총 개수의 대략 20%인 141일 수 있다.
- [0110] 이에, 특징값 비율산출부(141)는 정렬된 707개의 특징값 중에서 기 설정된 제2임계값 이하의 특징값, 예컨대 1~141개의 특징값을 갱신 전 시그니처의 기준 특징값으로 추출할 수 있다.

표 3

[0111]

ID	초기 특징값	기준악성코드 개수
1	NtWriteFile_"file_handle": masked "filepath": "C:WWUsersWWAdministratorWWAppDataWWLocalWWTempWW serverr.exe"	0
2	UuidCreate_"uuid": "{fc7ae18d-a45d-46ca-a289-864d213cd522}"	0
..	.	.
.	.	.
.	.	.
141	RegQueryValueExW_"key_handle": masked "value": "" "regkey_r": "Modules" "reg_type": masked "regkey": "HKEY_LOCAL_MACHINE\\SOFTWARE\\Microsoft\\Fusion\\NativeImagesIndex \\v2.0.50727_32\\IL\\f6e8397\\46ad0879\\6f\\Modules"	139
142	RegQueryValueExW_"key_handle": masked "value": "System.Runtime.Remoting,2.0.0.0,,b77a5c561934e089" "regkey_r": "DisplayName" "reg_type": masked "regkey": "HKEY_LOCAL_MACHINE\\SOFTWARE\\Microsoft\\Fusion\\NativeImagesIndex \\v2.0.50727_32\\IL\\f6e8397\\46ad0879\\6f\\DisplayName"	139
.	.	.
.	.	.
.	.	.
706	LdrGetDllHandle_"module_name": "kernel32.dll" "stack_pivoted": masked "module_address": masked	1384
707	NtProtectVirtualMemory+"process_identifier":masked "stack_dep_bypas s": masked "stack_pivoted": masked "heap_dep_bypass": masked "length": masked "protection": masked "process_handle": masked "base_address": masked	1430

[0112]

다음으로, 특징값 비율산출부(141)는 갱신된 시그니처의 다수의 특징값을 추출하고, 추출된 다수의 특징값에 기
추출된 하나 이상의 기준 특징값의 포함 비율을 산출할 수 있다(S320).

[0113]

이어, 시그니처 검증모듈(140)의 시그니처 갱신제어부(142)는 기 설정된 제3임계값에 기초하여 갱신된 시그니처
의 다수의 특징값 중 기준 특징값의 포함 비율로부터 갱신된 시그니처의 재 갱신 여부를 판단할 수 있다(S330).

[0114]

예컨대, 시그니처 갱신제어부(142)는 기 산출된 기준 특징값의 포함 비율이 제3임계값 이상이면, 갱신된 시그니
처의 재 갱신이 필요함을 판단할 수 있다. 이에, 시그니처 갱신제어부(142)는 시그니처 갱신모듈(130)의 동작을
제어하여 상기 시그니처 갱신모듈(130)에서 갱신된 시그니처를 재 갱신하여 새로운 시그니처가 출력되도록 할
수 있다.

[0115]

반면에, 시그니처 갱신제어부(142)는 기 산출된 기준 특징값의 포함 비율이 제3임계값 미만이면, 유의미한 특징
값이 충분하지 않다고 판단할 수 있다. 이에, 시그니처 검증모듈(140)은 갱신 전 시그니처를 기 입력된 악성코
드에 대한 최적의 시그니처로 출력할 수 있다.

[0116]

여기서, 제3임계값은 기준 특징값의 총 개수에 대한 소정의 퍼센트 비율, 예컨대 대략 80%의 비율에 해당하는
값일 수 있다.

[0117]

따라서, 갱신된 시그니처의 다수의 특징값에 기준 특징값, 즉 갱신 전 시그니처에서 추출된 특징값이 제3임계값
이상 포함된 경우에, 갱신된 시그니처에는 갱신 전 시그니처의 유의미한 시그니처가 유효함을 알 수 있다.
이에, 시그니처 갱신제어부(142)는 갱신된 시그니처의 재 갱신을 판단하여 시그니처 갱신모듈(130)의 동작을 제
어할 수 있다.

[0118]

또한, 갱신된 시그니처의 다수의 특징값에 기준 특징값이 제3임계값 미만 포함된 경우에, 갱신 전 시그니처의
유의미한 시그니처가 유효하지 않음을 알 수 있다. 이에, 시그니처 갱신제어부(142)는 갱신된 시그니처를 제거

하고 갱신 전 시그니처를 기 입력된 악성코드의 최적 시그니처로 출력할 수 있다.

[0119] 이와 같이, 본 실시예의 시그니처 생성장치(100)는 역색인 저장구조로 저장된 다수의 기존악성코드에 기초하여 외부에서 입력된 악성코드에 대한 시그니처를 빠르게 생성할 수 있다.

[0120] 또한, 본 실시예의 시그니처 생성장치(100)는 2개의 시그니처 간, 즉 갱신 전 시그니처와 갱신 후 시그니처 간 특징값 동일비율에 따라 시그니처 갱신의 반복을 제어함으로써, 입력된 악성코드에 대한 최적의 시그니처를 생성할 수 있다.

[0121] 이상에서 설명된 본 발명의 블록도의 각 블록과 순서도의 각 단계의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수도 있다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 인코딩 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 인코딩 프로세서를 통해 수행되는 그 인스트럭션들이 블록도의 각 블록 또는 순서도의 각 단계에서 설명된 기능들을 수행하는 수단을 생성하게 된다. 이들 컴퓨터 프로그램 인스트럭션들은 특정 방법으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 메모리에 저장되는 것도 가능하므로, 그 컴퓨터 이용가능 또는 컴퓨터 판독 가능 메모리에 저장된 인스트럭션들은 블록도의 각 블록 또는 순서도 각 단계에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다. 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 블록도의 각 블록 및 순서도의 각 단계에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

[0122] 또한, 각 블록 또는 각 단계는 특정된 논리적 기능(들)을 실행하기 위한 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실시예들에서는 블록들 또는 단계들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들 또는 단계들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들 또는 단계들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

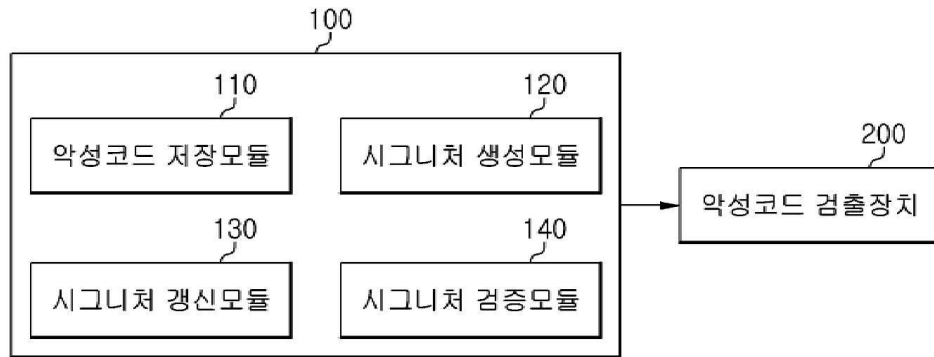
[0123] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 품질에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명에 개시된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 균등한 범위 내에 있는 모든 기술사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

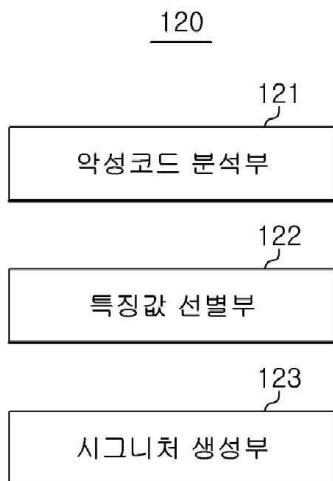
[0124] 100: 시그니처 생성장치 110: 악성코드 저장모듈
120: 시그니처 생성모듈 121: 악성코드 분석부
122: 특징값 선별부 123: 시그니처 생성부
130: 시그니처 갱신모듈 131: 패밀리 악성코드 추출부
132: 공통 특징값 추출부 133: 시그니처 갱신부
140: 시그니처 검증모듈 141: 특징값 비율산출부
142: 시그니처 갱신제어부

도면

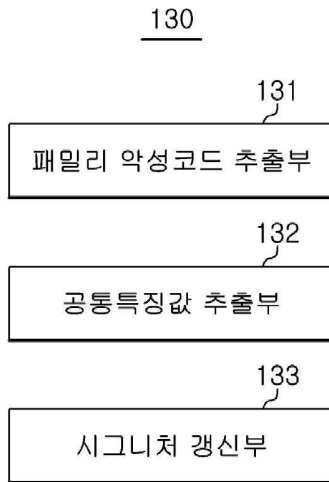
도면1



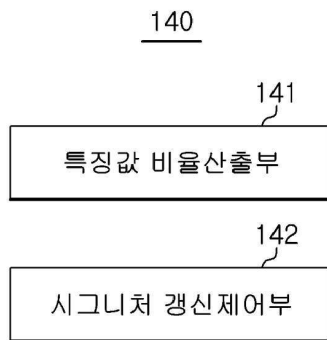
도면2



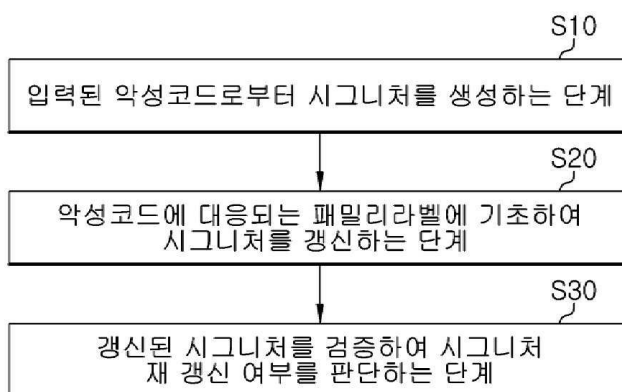
도면3



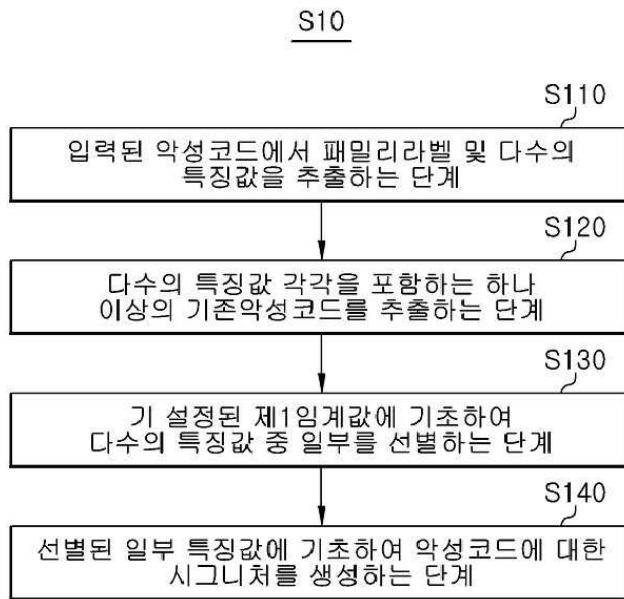
도면4



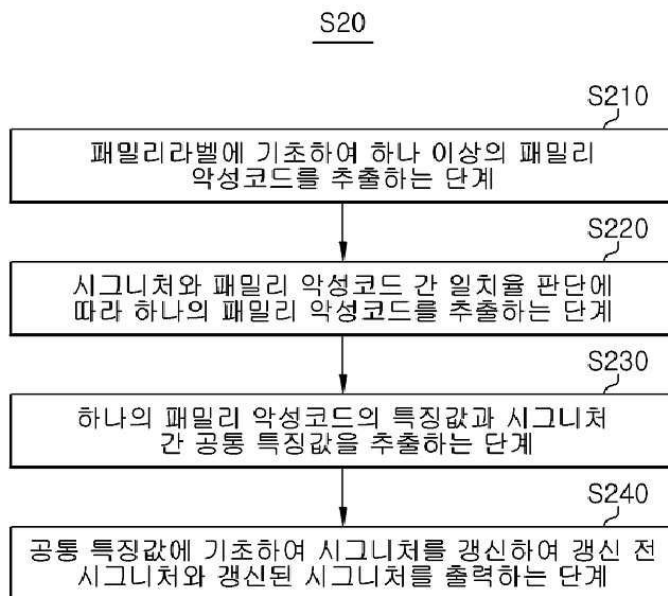
도면5



도면6



도면7



도면8

