(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년11월10일
(11) 등록번호 10-2466167
(24) 등록일자 2022년11월08일

(51) 국제특허분류(Int. Cl.)
G06F 21/56 (2013.01) *G06N 3/08* (2006.01)
(52) CPC특허분류
G06F 21/56 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2021-0013287
(22) 출원일자 2021년01월29일
심사청구일자 2021년01월29일
(65) 공개번호 10-2022-0109814
(43) 공개일자 2022년08월05일
(56) 선행기술조사문헌
Daniel Gibert et al., "Classification of Malware by Using Structural Entropy on Convolutional Neural Networks"(2018.04.)*
Prudkovskiy Nikolay, "Static analysis of executable files by machine learning methods"(2020.07.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
충남대학교 산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
김정우
대전광역시 유성구 대학로76번안길 45 204호
조은선
대전광역시 유성구 어은로 57 한빛아파트, 110동 1504호
백준영
대전광역시 중구 문화로87번길 72
(74) 대리인
이은철, 김재문

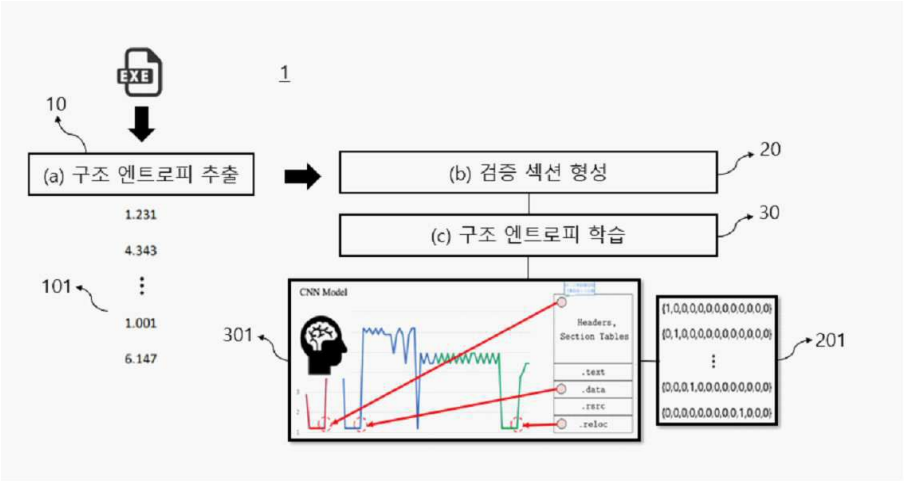
전체 청구항 수 : 총 3 항 심사관 : 정성훈

(54) 발명의 명칭 **컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 악성코드 탐지 프로그램 및 방법**

(57) 요 약

본 발명은, 컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 악성코드 탐지 프로그램에 있어서, 상기 파일의 세그먼트를 분류하여 구조 엔트로피 값을 추출하는 제1 기능; 및 상기 제1 기능으로 추출된 상기 구조 엔트로피 값에 위치 정보인 희소 행렬 벡터(one-hot vector)를 부여하여 검증 섹션을 형성하는 제2 기능;을 실행시키기 위하여 매체에 저장되며, 상기 검증 섹션의 상기 희소 행렬 벡터의 피처가 반영된 구조 엔트로피를 분석하여 상기 파일의 악성코드 감염 여부를 판단하는 것을 특징으로 한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	1711103337
과제번호	2019-0-01343-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	융합보안핵심인재양성사업
기 여 율	1/1
과제수행기관명	한국인터넷진흥원
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램에 있어서,

상기 파일의 세그먼트를 분류하여 구조 엔트로피 값을 추출하는 제1 기능; 및

상기 제1 기능으로 추출된 상기 구조 엔트로피 값에 위치 정보인 희소 행렬벡터(one-hot vector)를 부여하여 검증 섹션을 형성하는 제2 기능;을 실행시키기 위하여 매체에 저장되며,

상기 검증 섹션의 상기 희소 행렬 벡터의 피처(feature)가 반영된 구조 엔트로피를 분석하여 상기 파일의 악성코드 감염 여부를 판단하고,

상기 제1 기능은,

상기 파일의 섹션을 일정한 바이트(byte) 사이즈의 청크(chunk)로 나누고,

상기 제2 기능은,

상기 청크의 각각에 상기 희소 행렬 벡터를 부여하고,

상기 제2 기능은,

상기 희소 행렬 벡터의 위치 정보에 해당하는 상기 구조 엔트로피 값을 매칭하여 2차원 벡터로 구성된 상기 검증 섹션을 형성하는 것을 특징으로 하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램.

청구항 2

제 1 항에 있어서,

상기 제2 기능이 실행되어 형성된 상기 검증 섹션의 상기 희소 행렬 벡터와 상기 구조 엔트로피 값을 인공신경망에 연결하여 구조 엔트로피를 학습하는 제3 기능을 더 실행시키는 것을 특징으로 하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램.

청구항 3

제 2 항에 있어서,

상기 제3 기능은,

딥러닝 알고리즘인 CNN(Convolutional neural network) 모델로 상기 검증 섹션을 학습하는 것을 특징으로 하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램.

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 악성코드 탐지 프로그램 및 방법에 관한 것으로서, 특히 컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 악성코드 탐지 프로그램 및 방법에 관한 것이다.

배 경 기 술

[0003] 악성코드는 사용자 컴퓨터에 위협을 가하며, 시간이 지날수록 나날이 지능화되고 있다. 악성코드를 탐지하기 위해서는 악성코드의 시그니쳐를 확인하여 판단해야 한다. 그러나 시그니쳐를 확인하고 악성코드로 판별하기까지의 시간은 매우 오래 걸린다. 시간을 단축하기 위해 다양한 인공지능 기반의 악성코드 탐지 시스템이 개발되고 있으며, 그중에서 여러 가지 악성코드의 피쳐들을 분석해서 악성코드를 자동으로 탐지하기 위한 기술의 수요가 증가하고 있다.

[0004] 악성코드를 탐지하는 대표적인 피쳐로는 구조 엔트로피가 있는데, 구조 엔트로피는 파일을 일정한 세그먼트로 쪼개 각각의 세그먼트에서 엔트로피 수치를 추출한 일련의 숫자 나열이다. 이러한 구조 엔트로피를 분석을 통해서 높은 성능을 가지는 악성코드 탐지 시스템을 구현할 수 있다. 그러나 악성코드와 같은 바이너리 파일은 시그니쳐 패턴이 파일에서 존재하는 위치에 따라 다른 의미를 가진다. 따라서, 기존의 구조 엔트로피를 기반으로 한 악성코드 탐지 시스템은 시그니쳐 패턴이 존재하는 위치의 정보를 반영하지 못하였다.

[0005] 종래의 특허문헌으로, 한국등록특허 제10-1526500호는 정보 엔트로피를 이용한 악성 의심 웹사이트 탐지 방법을 개시한다. 상기 선행문헌은 엔트로피 값을 기 설정된 임계값과 비교하여 악성코드가 삽입되어 있는지 여부를 판단한다. 이는 엔트로피 값을 이용하여 악성코드의 존재 여부를 판단하지만, 시그니쳐 패턴의 위치 정보를 반영하는 기술은 개시되지 않고 있다.

[0006] 이에, 본 출원인은 악성코드를 효과적으로 분석할 수 있도록 구조 엔트로피에 파일의 구조 정보를 분석하고 파일 위치 정보를 추가하여 인공지능의 학습을 통해 악성코드를 탐지하는 새로운 모델을 고안하게 되었다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국등록특허 제10-1526500호

발명의 내용

해결하려는 과제

[0009] 본 발명은 악성코드를 효과적으로 분석할 수 있도록 구조 엔트로피에 파일의 구조 정보를 분석하여 위치 정보를 추가하는 새로운 피쳐엔지니어링을 개발하는데 목적이 있다.

과제의 해결 수단

[0011] 상기 목적을 달성하기 위하여 본 발명은, 컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 악성코드 탐지 프로그램에 있어서, 상기 파일의 세그먼트를 분류하여 구조 엔트로피 값을 추출하는 제1 기능; 및 상기 제1 기능으로 추출된 상기 구조 엔트로피 값에 위치 정보인 희소 행렬 벡터(one-hot vector)를 부여하여 검증 섹션을 형성하는 제2 기능;을 실행시키기 위하여 매체에 저장되며, 상기 검증 섹션의 상기 희소 행렬 벡터의 피쳐가 반영된 구조 엔트로피를 분석하여 상기 파일의 악성코드 감염 여부를 판단하는 것을 일 특징으로 한다.

[0012] 바람직하게, 상기 제2 기능이 실행되어 형성된 상기 검증 섹션의 상기 희소 행렬 벡터와 상기 구조 엔트로피 값을 인공신경망에 연결하여 구조 엔트로피를 학습하는 제3 기능을 더 실행시키는 것을 특징으로 한다.

[0013] 바람직하게, 상기 제3 기능은, 딥러닝 알고리즘인 CNN(Convolutional neural network) 모델로 상기 검증 섹션을 학습하는 것을 특징으로 한다.

[0014] 바람직하게, 상기 제1 기능은, 상기 파일의 섹션을 일정한 바이트(byte) 사이즈의 청크(chunk)로 나누고, 상기 제2 기능은, 상기 청크의 각각에 상기 희소 행렬 벡터를 부여하는 것을 특징으로 한다.

[0015] 바람직하게, 상기 제2 기능은, 상기 희소 행렬 벡터의 위치 정보에 해당하는 상기 구조 엔트로피 값을 매칭하여 2차원 벡터로 구성된 상기 검증 섹션을 형성하는 것을 특징으로 한다.

[0016] 또한, 본 발명은 컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 악성코드 탐지 방법에 있어서, 상기 파일의 세그먼트를 분류하여 구조 엔트로피 값을 추출하는 (a)단계; 및 상기 (a)단계에서 추출된 상기 구조 엔트로피 값에 위치 정보인 희소 행렬 벡터(one-hot vector)를 부여하여 검증 섹션을 형성하는 (b)단계;를 포함하고, 상기 검증 섹션의 상기 희소 행렬 벡터의 피처가 반영된 구조 엔트로피를 분석하여 상기 파일의 악성코드 감염 여부를 판단하는 것을 다른 특징으로 한다.

[0017] 바람직하게, 상기 (b)단계가 수행되어 형성된 상기 검증 섹션의 상기 희소 행렬 벡터와 상기 구조 엔트로피 값을 인공신경망에 연결하여 구조 엔트로피를 학습하는 (c)단계를 더 포함하는 것을 특징으로 한다.

발명의 효과

[0019] 본 발명에 따르면, 컴퓨터에서 실행되는 파일의 구조를 분석하여, 섹션의 희소 행렬 벡터의 피처가 반영된 구조 엔트로피를 통해 악성코드를 탐지하는 효과가 있다. 보다 상세하게, 본 발명에 따르면 종래의 바이너리 파일이 다른 문맥마다 다른 의미를 갖기 때문에 다른 의미에 따른 유사한 구조 엔트로피 패턴을 찾는 것이 어려워 구조 엔트로피를 통한 악성코드 탐지에 한계가 있었다. 본 발명은 구조 엔트로피에 섹션 정보인 검증 섹션을 새롭게 추가하여 검증 섹션의 구조 엔트로피 분석으로 악성코드를 탐지할 수 있는 이점이 있다. 이 경우, 구조 엔트로피 분석에는 딥러닝 알고리즘인 CNN(Convolutional neural network) 모델을 적용하여 정확도 높은 악성코드의 탐지가 가능하다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 실시예에 따른 엔트로피를 기반으로 한 악성코드 탐지 프로그램의 구성도이다.

도 2는 도 1의 실시예에 따른 악성코드 탐지 프로그램의 특징 개요도이다.

도 3은 본 발명의 실시예에 따른 구조 엔트로피를 나타낸다. 도 3a는 악성코드 탐지를 위한 파일의 구조 엔트로피의 패턴 모습을 나타낸다. 도 3b는 본 발명의 실시예에 따른 검증 섹션이 추가된 파일의 구조 엔트로피를 나타낸다.

도 4는 본 발명의 실시예에 따른 악성코드 탐지 프로그램의 성능 결과를 나타낸다. 도 4a는 악성코드(malware)의 파일과 정상 파일(benign)의 테스트 셋 설정 모습을 나타낸다. 도 4b는 도 4a의 테스트 셋을 인공신경망에 연결하여 구조 엔트로피를 학습한 뒤 악성코드 검출 성능을 비교한 결과를 나타낸다.

발명을 실시하기 위한 구체적인 내용

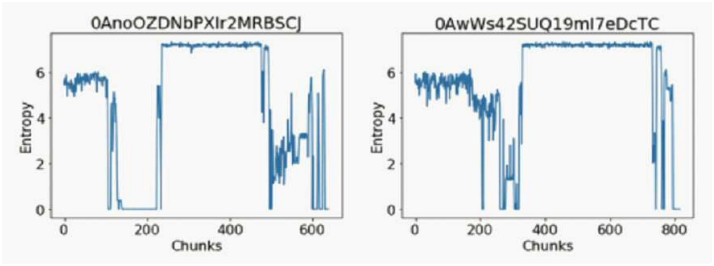
[0022] 하, 첨부된 도면들에 기재된 내용들을 참조하여 본 발명을 상세히 설명한다. 다만, 본 발명이 예시적 실시 예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일 참조부호는 실질적으로 동일한 기능을 수행하는 부재를 나타낸다.

[0023] 본 발명의 목적 및 효과는 하기의 설명에 의해서 자연스럽게 이해되거나 보다 분명해질 수 있으며, 하기의 기재만으로 본 발명의 목적 및 효과가 제한되는 것은 아니다. 또한, 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다.

[0024] 도 1은 본 발명의 실시예에 따른 엔트로피를 기반으로 한 악성코드 탐지 프로그램(1)의 구성도이다. 도 2는 도 1의 실시예에 따른 악성코드 탐지 프로그램(1)의 특징 개요도이다.

[0025] 도 1 및 도 2를 참조하면, 엔트로피를 기반으로 한 악성코드 탐지 프로그램은 컴퓨터에서 실행되기 위한 제1 기능(10), 제2 기능(20) 및 제3 기능(30)을 포함하여 매체에 저장될 수 있다.

[0026] 배경기술에 전술한 바와 같이, 악성코드를 탐지하는 대표적인 피처인 구조 엔트로피는 각각의 세그먼트에서 엔트로피 수치를 추출한 일련의 숫자 나열로 고유의 패턴을 갖는다. 청크(chunk) 파일의 같은 패밀리(family) 내에서는 구조 엔트로피가 같은 패턴을 형성하기 때문에 악성코드의 탐지를 위한 피쳐로 활용되기에 적합하다.



[0027]

[0028] <구조 엔트로피의 참고도>

[0030] 상기 구조 엔트로피의 참고도는 고유의 구조 엔트로피 패턴을 예시한 것이다. 그러나, 바이너리(binary) 파일은 다른 문맥(context)에서 다른 의미를 가지므로 유사 패턴의 구분이 어려운 문제가 있다. 이하에서 설명하게 될 본 실시예에 따른 악성코드 탐지 프로그램은 제1 기능(10), 제2 기능(20) 및 제3 기능(30)의 실행으로, 구조 엔트로피에 분석 가능한 섹션을 형성하여 악성코드의 검증 가능한 피쳐 엔지니어링을 수행한다. 보다 상세하게, 본 발명의 실시예에 따른 악성코드 탐지 프로그램(1)은 검증 섹션(201)의 희소 행렬 벡터의 피쳐가 반영된 구조 엔트로피를 분석하여 파일의 악성코드 감염 여부를 판단할 수 있다.

[0031] 제1 기능(10)은 파일의 세그먼트를 분류하여 구조 엔트로피 값(101)을 추출할 수 있다. 파일의 섹션을 일정한 바이트(Byte) 사이즈의 청크(chunk)로 나눌 수 있다. 제1 기능(10)이 적용되는 파일은 컴퓨터에서 실행되는 파일일 수 있다. 본 실시예로, 본 명세서에서 지칭하는 파일은 실행 파일인 PE 파일 및 윈도우에서 실행 가능한 파일을 총칭할 수 있다.

[0032] 구조 엔트로피 값(10)은 파일을 일정한 세그먼트로 분류하여 각각의 세그먼트에서 엔트로피 수치를 추출한 일련의 숫자 나열을 의미한다. 하나의 세그먼트에는 여러개의 섹션이 존재한다.

[0033] 제1 기능(10)은 구조 엔트로피의 바이너리를 청크(chunk)로 분류하고 각 청크에서 구조 엔트로피 값(101)을 추출한다.

[0034] 제2 기능(20)은 제1 기능(10)으로 추출된 구조 엔트로피 값(101)에 위치 정보인 희소 행렬 벡터(one-hot vector)를 부여하여 검증 섹션(201)을 형성할 수 있다.

[0035] 제2 기능(20)은 청크(chunk)의 각각에 희소 행렬 벡터를 부여하여 각각의 구조 엔트로피 값(101)에 대응되는 희소 행렬 벡터를 생성한다.

[0036] 제2 기능(20)은 희소 행렬 벡터의 위치 정보에 해당하는 구조 엔트로피 값(101)을 매칭하여 2차원 벡터로 구성되는 검증 섹션(201)을 형성할 수 있다.

[0037] 제3 기능(30)은 제2 기능(20)이 실행되어 형성된 검증 섹션(201)의 희소 행렬 벡터와 구조 엔트로피 값(101)을 인공신경망에 연결하여 구조 엔트로피를 학습할 수 있다. 인공신경망은 딥러닝 알고리즘인 CNN(Convolutional neural network) 모델(301)로 검증 섹션(201)을 학습할 수 있다.

[0038] CNN 모델(301)은 이미지에서 객체, 얼굴, 장면 및 텍스트를 인식하기 위해 패턴을 찾는 데 유용하다. 특히, 데이터에서 직접 학습하며, 패턴을 사용하여 이미지를 분류하고 특징을 수동으로 추출할 필요가 없다.

[0039] PE(Portable Excutable)는 윈도우에서 사용되는 실행 가능한 파일의 형식을 의미한다. PE의 구조는 크게 헤더와 섹션으로 구분되는데, 헤더는 섹션 테이블이라고도 불린다. 섹션 테이블은 파일을 실행할 때 처음 시작해야 할 코드의 시작 부분에 대한 정보와 섹션의 구조 정보를 관리하는 구조체라 할 수 있다.

[0040] PE 파일의 섹션은 사양에 따라 13개의 섹션으로 분류된다. 13개의 섹션은 헤더, data, .edata, .idata, .pdata, .rdata, .rsrc, .reloc, .text, .tls, .sdata, .xdata, Undefined를 포함한다. 12개의 섹션으로 분류

되지 않는 나머지 섹션을 Undefined로 정의한다. 분류된 섹션은 섹션별로 바이너리에서 바이트를 추출하고, 추출된 바이트는 동일한 크기의 청크로 나뉜다. 본 실시예에 따른 엔트로피를 기반으로 한 악성코드 탐지 시스템은 7번의 과정을 통해 진행된다.

[0041] 도 2에 도시된 악성코드 탐지 프로그램(1)의 피쳐 엔지니어링 과정은 하기의 7가지 스텝으로 진행될 수 있다.

[0042] 첫 번째로, 13개의 섹션을 기반으로 바이너리 프로그램에서 섹션을 식별한다. 두 번째로, 식별된 13개의 섹션별로 바이너리 프로그램에서 바이트를 추출한다. 세 번째로, 섹션에서 추출된 바이트를 동일한 크기의 청크로 분류한다. 이를 $Chunk_i^S$ 로 표기하며, 여기서 i는 청크의 인덱스를 의미하고, S는 섹션을 의미한다. S에는 헤더, data, .edata, .idata, .pdata, .rdata, .rsrc, .reloc, .text, .tls, .sdata, .xdata, Undefined가 포함한다. 또한, 인덱스 i는 0보다 크거나 같고 m보다는 작은 범위를 가질 수 있다. 네 번째로, 바이너리 프로그램에서 추출한 바이트의 모든 청크에 대해 구조 엔트로피 값(201)을 추출한다. 이를 $Entropy_i^S$ 로 표기하며, S는 섹션을 의미하고, i는 청크의 엔트로피 값을 의미한다. 다섯 번째로, 모든 청크를 대상으로 희소 행렬 백터를 생성한다. 이는 ohV_i^S 로 표기되며, 섹션 S 내에 인덱스 i의 청크에 대한 희소 행렬 벡터를 의미한다. 본 실시예에 따른 ohV_i^S 는 .edata 섹션의 10번째 청크는 $ahV_{10}^{.cdatc}$ 으로 표기되며, 희소 행렬 벡터의 3번째 요소가 .edata 섹션에 해당하는 경우는 <0,0,1,0,0,0,0,0,0,0,0,0,0,0> 값으로 표기된다. 여섯 번째로, $Entropy_i^S$ 와 ohV_i^S 는 결합이 가능하며, 결과 벡터는 $chunkVec_i^S$ 로 표기한다. 본 실시예에 따른 $Entropy_{10}^{.cdata}$ 는 $Entropy_{10}^{.cdata}$ 가 2.3이면 $chunkVec_{10}^{.cdata}$ 는 <2.3, 0,0,1,0,0,0,0,0,0,0,0,0,0> 로 표기되며, 2.3은 $ahV_{10}^{.cdatc}$ 보다 앞에 표기된다. 일곱 번째로, 각 청크에 대해 희소 행렬 벡터와 구조 엔트로피 값(101)이 연결되며, 각 섹션은 희소 행렬 벡터로 표기되어, <m, 14> 형태의 2차원 벡터를 형성된다. 희소 행렬 백터의 단일 요소는 1, 그 외 다른 요소는 0으로 표기된다. 이후에 CNN 모델(301)에 제공된다.

[0043] 도 3은 본 발명의 실시예에 따른 구조 엔트로피를 나타낸다. 도 3a는 악성코드 탐지를 위한 파일의 구조 엔트로피의 패턴 모습을 나타낸다. 도 3b는 본 발명의 실시예에 따른 검증 섹션이 추가된 파일의 구조 엔트로피를 나타낸다.

[0044] 도 3a을 참조하면, 구조 엔트로피는 동일한 계열 내에서 유사한 패턴을 보이는데, 각각의 패턴은 서로 다른 의미를 갖는다. 패턴이 유사하기 때문에 다른 의미를 가져도 구별하기가 어렵다. 도 3b는 도 3a의 구조 엔트로피에 도 2의 실시예에 따른 섹션의 정보를 구조 엔트로피의 패턴에 추가한 결과를 나타낸다.

[0045] 도 4는 본 발명의 실시예에 따른 악성코드 탐지 프로그램의 성능 결과를 나타낸다. 도 4a는 악성코드(malware)의 파일과 정상 파일(benign)의 테스트 셋 설정 모습을 나타낸다. 도 4b는 도 4a의 테스트 셋을 인공신경망에 연결하여 구조 엔트로피를 학습한 뒤 악성코드 검출 성능을 비교한 결과를 나타낸다.

[0046] 도 4a를 참조하면, 본 실시예에서 사용되는 PE 파일은 컴퓨터 윈도우의 System32 폴더에서 정상으로 판별된 파일로 선정하였다. 악성코드를 내장한 PE 파일은 Roberts J.-M. Virusshare(https://virusshare.com/)의 VirusShare_0로 선정하였다.

[0047] 도 4a의 Datasets used는 Dataset이 Benign 및 Malware에 대한 세부 정보를 나타낸다. Benign는 Train Set이 3,409, Test Set은 1,461로, Malware는 Train Set이 11,695, Test Set이 5,017로 70/30의 비율로 분할된다.

[0048] 청크를 4,096 바이트로 설정하는 동안 모든 Datasets에 대해 feature engineering을 수행하고 CNN 모델(301)에 결과가 제공된다. CNN 모델(301)에서 variable-length 입력을 처리하기 희소 행렬 벡터의 범위를 3,600으로 제한하였다. 따라서 파일은 <3600, 14>의 희소 행렬 벡터로 표현된다.

[0049] 도 4b를 참조하면, PE 파일의 섹션 정보가 악성코드 탐지에 대한 중요성을 확인하기 위해 종래의 구조 엔트로피와 본 실시예에 따른 섹션 정보가 추가된 구조 엔트로피를 비교하였다. 일반 구조 엔트로피는 Entropy Streams(w/o information on sections)으로, 본 실시예에 따른 구조 엔트로피는 The Proposed one(W/information on sections)으로 표기하였다. 각각의 구조 엔트로피는 CNN 모델(301)에 제공되어 정확성(Accuracy)과 효과성(F1-Score)의 테스트를 진행하였다. 도 4b의 Performance comparison with Benign and

Malware를 참고하면, The Proposed one(W/information on sections)은 Entropy Streams(w/o information on sections)보다 정확성(Accuracy)이 3.3% 및 효과성(F1-Score)이 0.05 향상되었다.

[0050] 본 발명의 다른 실시예로, 컴퓨터에서 실행되는 파일의 구조 정보를 이용하여 엔트로피 기반으로 악성코드를 탐지하는 방법은 악성코드를 탐지 프로그램(1)에서 수행되는 수행 단계가 될 수 있다. 본 실시예에 따른 악성코드를 탐지하는 방법은 파일의 세그먼트를 분류하여 구조 엔트로피 값(101)을 추출하는 (a) 단계; (a)단계에서 추출된 구조 엔트로피 값(101)에 위치 정보인 희소 행렬 벡터(one-hot vector)를 부여하여 검증 섹션을 형성하는 (b)단계; (b)단계가 수행되어 형성된 검증 섹션(201)의 희소 행렬 벡터와 구조 엔트로피 값(101)을 인공신경망에 연결하여 구조 엔트로피를 학습하는 (c)단계를 포함할 수 있다. (a)단계는 전술한 제1 기능(10), (b)단계는 전술한 제2 기능(20), (c)단계는 전술한 제3 기능(30)의 실시예가 원용될 수 있다.

[0051] 이상에서 대표적인 실시예를 통하여 본 발명을 상세하게 설명하였으나, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리 범위는 설명한 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 특허청구범위와 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태에 의하여 정해져야 한다.

부호의 설명

[0053] 1: 악성코드 탐지 프로그램

10: 제1 기능

101: 구조 엔트로피 값

20: 제2 기능

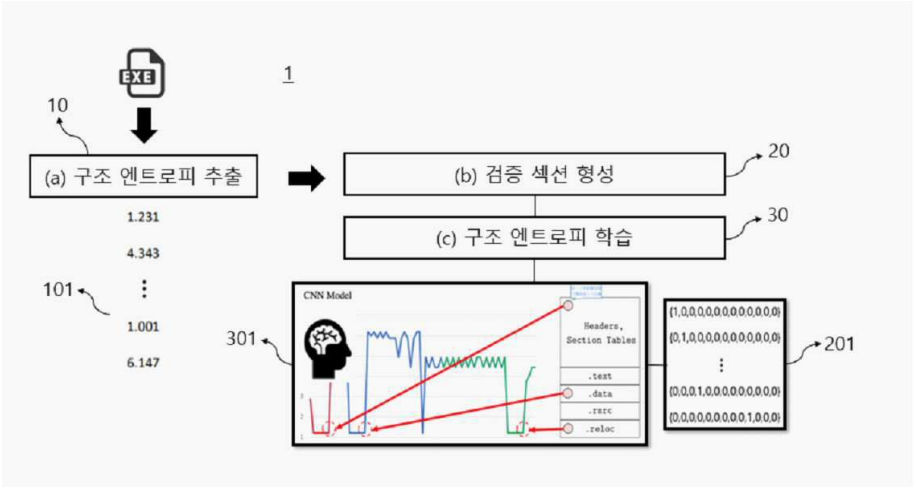
201: 검증 섹션

30: 제3 기능

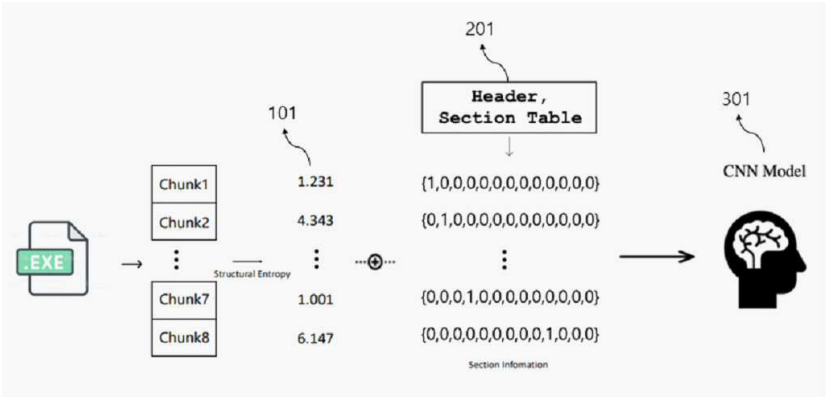
301: CNN 모델

도면

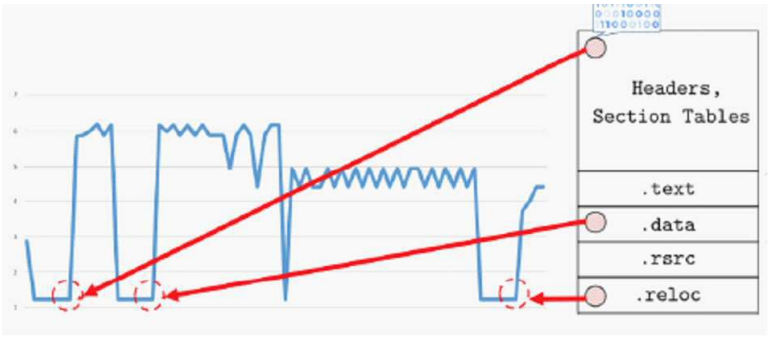
도면1



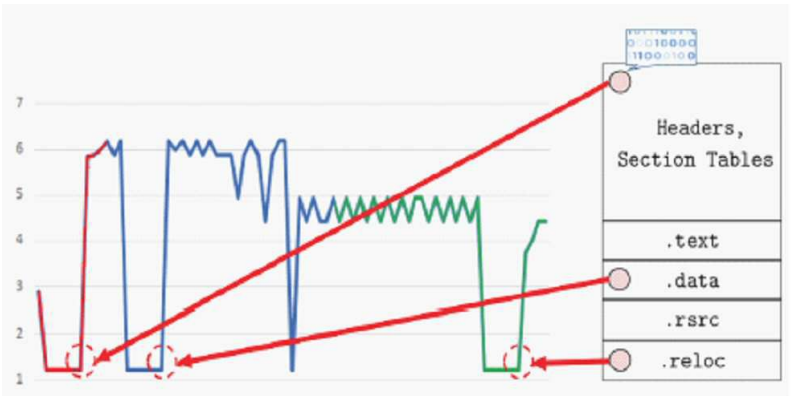
도면2



도면3a



도면3b



도면4a

Dataset	The number of Train Set	The number of Test Set	Total
Benign	3,409	1,461	4,870
Malware	11,695	5,017	16,712

도면4b

Feature	Accuracy	F1-score
Entropy streams (w/o information on sections)	95.8 %	0.94
The proposed one (w/ information on sections	99.1 %	0.99

【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 2

【변경전】

제 1 항에 있어서,

상기 제2 기능이 실행되어 형성된 상기 검증 섹션의 상기 희소 행렬 벡터와 상기 구조 엔트로피 값을 인공신경망에 연결하여 구조 엔트로피를 학습하는 제3 기능을 더 실행시키는 것을 특징으로 하는 매체에 저장된 악성코드 탐지 프로그램.

【변경후】

제 1 항에 있어서,

상기 제2 기능이 실행되어 형성된 상기 검증 섹션의 상기 희소 행렬 벡터와 상기 구조 엔트로피 값을 인공신경망에 연결하여 구조 엔트로피를 학습하는 제3 기능을 더 실행시키는 것을 특징으로 하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램.

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 3

【변경전】

제 2 항에 있어서,

상기 제3 기능은,

딥러닝 알고리즘인 CNN(Convolutional neural network) 모델로 상기 검증 섹션을 학습하는 것을 특징으로 하는 매체에 저장된 악성코드 탐지 프로그램.

【변경후】

제 2 항에 있어서,

상기 제3 기능은,

딥러닝 알고리즘인 CNN(Convolutional neural network) 모델로 상기 검증 섹션을 학습하는 것을 특징으로 하는 컴퓨터로 읽을 수 있는 기록 매체에 저장된 악성코드 탐지 프로그램.