로컬 특징 기반 글로벌 이미지를 사용한 CNN 기반의 악성코드 분류 방법

장세준, 성연식*
동국대학교 멀티미디어공학과
sejun@dongguk.edu, sung@dongguk.edu

Convolutional Neural Network-based Malware Classification Method utilizing Local Feature-based Global Image

Sejun Jang, Yunsick Sung*
Dept. of Multimedia Engineering, Dongguk University-Seoul

요 약

최근 악성코드로 인한 피해가 증가하고 있다. 악성코드는 악성코드가 속한 종류에 따라서 대응하는 방법도 다르기 때문에 악성코드를 종류별로 분류하는 연구도 중요하다. 기존에는 악성코드 시각화 과정을 통해서 생성된 악성코드의 글로벌 이미지를 사용해 악성코드를 각 종류별로 분류한다. 글로벌 이미지를 악성코드로부터 추출한 바이너리 정보를 사용해서 생성한다. 하지만, 글로벌 이미지만을 사용해서 악성코드를 각 종류별로 분류하는 경우 악성코드의 종류별로 중요한 특징을 고려하기 않기 때문에 분류 정확도가 떨어진다. 본 논문에서는 악성코드의 글로벌 이미지에 악성코드의 종류별 특징을 나타내기 위한 로컬 특징 기반 글로벌 이미지를 사용한 악성코드 분류 방법을 제안한다. 첫 번째, 악성 코드로부터 바이너리를 추출하고 추출된 바이너리를 사용해서 글로벌 이미지를 생성한다. 두 번째, 악성 코드로부터 로컬 특징을 추출하고 악성코드의 종류별 핵심 로컬 특징을 단어-역문서 빈도(Term Frequency Inverse Document Frequency, TFIDF) 알고리즘을 사용해 선택한다. 세 번째, 생성된 글로벌 이미지에 악성코드의 패밀리별 핵심 특징을 픽셀화해서 적용한다. 네 번째, 생성된 로컬 특징 기반 글로벌 이미지를 사용해서 컨볼루션 모델을 학습하고, 학습된 컨볼루션 모델을 사용해서 악성코드를 각 종류별로 분류한다.

1. 서론

최근 다양한 기술들의 발달로 인해 정보 보호의 중요성이 높아지고 있다. 해커는 특정 컴퓨터에 대한 권한을 악성코드를 사용해서 탈취한다. 해커는 획득한 권한을 사용해 특정 컴퓨터에 저장되어 있는 중요 정보를 변조하거나 탈취한다. 예를 들어, 슈퍼 인텔리전스 기반의 XR SW 플랫폼으로부터 발생한 K-pop 통합 데이터를 해커로부터 보호하기 위해서 악성코드의 탐지 및 분류 과정이 필요하다.

글로벌 이미지는 악성코드의 전체 바이너리 정보를 사용해 악성코드 자체를 시각화한 이미지다. 악성코드로부터 추출한 바이너리 정보를 8bit 단위로 나누고 나뉜 8bit를 하나의 픽셀로 사용한다. 8bit는 0에서 255까지의 수를 표현할 수 있기 때문에 그레이스케일 이미지의 픽셀로 사용하기 적합하다. 악성코드의 바이너리 정보는 연산 코드, 응용 프로그램 프로그래밍 인터페이스 그리고 동적 링크 라이브러리 등의 정보를 포함하기 때문에 악성코드의 글로벌 이미지를 사용하면 악성코드의 전체적인 구조와 함께 작은 변화를 감지할 수 있다. 변종 악성코드 분류가 가능하다. 하지만, 악성코드의 글로벌 이미지만 사용해서 악성코드를 분류할 경우 패밀리별 악성코드의 행위를 고려할 수 없는 단점이 있다.

이 논문에서는 로컬 특징 기반의 글로벌 이미지를 사용한 악성코드 분류 방법을 다음과 같이 제안한다: 첫 번째, 악성코드로부터 바이너리 정보를 추출해서 악성코드의 글로벌 이미지를 생성한다. 두 번째, 악성코드로부터 로컬 특징을 추출하고 단어-역문서 빈도

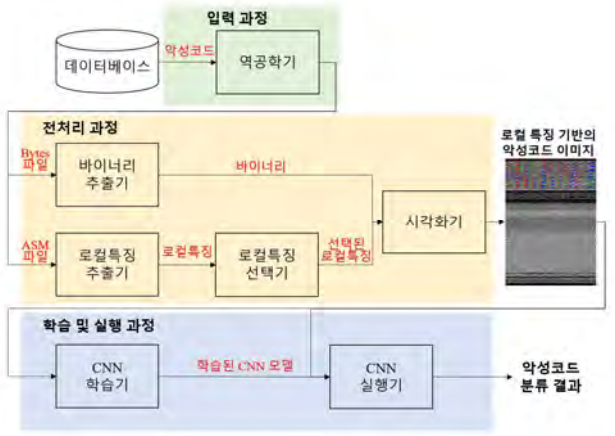
* 교신저자: 성연식 (sung@dongguk.edu)

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음(2019-0-01585)"

(Term Frequency Inverse Document Frequency, TFIDF) 알고리즘을 사용해서 악성코드의 패밀리별 중요 로컬 특징을 선택한다. 세 번째, 악성코드의 글로벌 이미지에 선택된 로컬 특징을 RGB 픽셀화해서 적용한다. 제안하는 로컬 특징 기반의 글로벌 이미지를 사용한 악성코드 분류 방법은 악성코드의 글로벌 이미지에 악성코드의 종류별 핵심 특징을 포함한다. 악성코드의 전체 구조와 로컬 특징을 고려하는 동시에 악성코드의 종류별 중요 특징을 한 장의 이미지에 포함할 수 있다.

2. 로컬 특징 기반 글로벌 이미지를 사용한 악성코드 분류 방법

(그림 1)은 제안하는 방법인 로컬 특징 기반 글로벌 이미지를 사용한 악성코드 분류 방법의 개요를 보여준다. 제안하는 방법은 입력 과정, 전처리 과정 그리고 학습 및 실행 과정으로 나뉜다.



(그림 1) 제안하는 방법 개요.

입력 과정에서는 역공학기가 데이터베이스로부터 악성코드를 입력받아서 역공학 소프트웨어를 사용해 바이트 파일과 ASM 파일을 출력한다.

전처리 과정에서는 입력 과정으로부터 입력된 바이트 파일과 ASM파일을 사용해서 로컬 특징 기반의 악성코드 이미지를 생성한다. 바이너리 추출기는 바이트 파일을 입력 받아서 바이너리를 출력한다. 로컬특징 추출기는 입력 과정으로부터 입력된 ASM 파일을 입력 받아서 로컬 특징을 추출한다. 로컬 특징은 CPU 명령어인 연산 코드다. 추출된 로컬 특징은 로컬 특징 선택기에 입력된다. 로컬 특징 선택기는 TFIDF 알고리즘을 사용해서 악성코드의 각 패밀리별 중요 특징을 선택한다[1]. 바이너리, 로컬 특징, 그리고 선택된 로컬 특징은 시각화기에 입력된다. 시각화기는 입력받은 바이너리, 로컬 특징, 그리고 선택된 로컬 특징을 사용해서 로컬 특징 기반 글로벌 이미지를 출력한다. 시각화기는 바이너리를 8bit 단위로 분할하고, 분할된 8bit 단위의 바이너리를 하나의 픽셀로 사용해서 글로벌 이미지를 생성한다[2]. 8자리의 2진수는 0부터 255까지의 값을 표현할 수 있기 때문에 그레이스케일 이미지의 픽셀로 사용하기 적합하다. 바이너리를 사용해 생성한 글로벌 이미지에 로컬 특징과 선택된 로컬 특징을 사용해서 악성코드의 각 종류별 중요한 특징을 RGB 픽셀로 적용한다.

3. 결론

이 논문에서는 로컬 특징 기반 글로벌 이미지를 사용한 악성코드 분류 방법을 다음과 같이 제안했다. 첫 번째, 데이터베이스로부터 역공학기를 사용해서 바이트 파일과 ASM 파일을 추출했다. 두 번째, 바이트 파일로부터 바이너리를 추출하고 8bit 단위로 나눴다. 8bit의 바이너리는 하나의 픽셀로 사용해서 글로벌 이미지를 생성했다. 세 번째, ASM 파일로부터 로컬 특징인 opcode를 추출하고, TF-IDF 알고리즘을 통해 악성코드의 각 종류별 중요한 특징을 선택했다. 추출된 로컬 특징과 선택된 로컬 특징을 사용해서 시각화 과정을 통해 로컬 특징 기반 글로벌 이미지를 생성했다. 네 번째, 로컬 특징 기반 글로벌 이미지를 통해서 악성코드를 각 종류별로 분류했다.

사사표기

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음(2019-0-01585)"

참고문헌

[1] H Zhang, X Xiao, F Mercaldo, S Ni, F Martinelli, A K Sangaiah, "Classification of Ransomware Families with Machine Learning based on N-gram of Opcodes", Computers & Electrical Engineering, 77, 366-375, 2019.

[2] Kesav K, Srinivas M, "Image Visualization based Malware Detection", 2013 IEEE Symposium on Computational Intelligence in Cyber Security (CICS). IEEE, Singapore, Singapore, 2013.