IoT기기의 보안강화를 위한 CNN기반 정상/악성코드 분류

김수영[1], 허서영[1], 이병현[2], 이미란[1,*]
[1]대구대학교 컴퓨터정보공학부
[2]주식회사 네오스텍 기업부설연구소
kimsuyeong.hci.du@gmail.com, heosy3374@daegu.ac.kr,
windroads@naver.com, *miran@daegu.ac.kr (교신저자)

CNN-based benign/malware classification for security enhancement for IoT device

Suyeong Kim[1], Seoyoung Heo[1],
Byung-Hyun Lee[2], Miran Lee[1]
[1]Dept. of Computer and Information Engineering, Daegu University
[2]Neostack, Inc.

요 약

IoT 기기 사용량의 증가로 인해 해킹 사례도 함께 증가하며 보안의 중요성이 커지고 있다. 본 논문은 IoT 보안 취약점을 해결하기 위해 정상/악성코드의 데이터셋을 Grayscale로 변환하여 악성코드/정상코드로 분류하는 알고리즘을 개발해 IoT 기기에서 성능을 검증한다. 분류에 이용되는 딥러닝 알고리즘은 CNN(Convolutional Neural Network)으로 99.60%의 평균 정확도를 나타내며 IoT 기기(라즈베리파이)에서도 잘 작동됨을 확인할 수 있다.

1. 서론

IoT 서비스 확산과 빠른 성장에 맞물려 정보통신망 침입, CCTV, IPTV, 생활가전 제품을 통한 스팸메일 발송 등 보안 위협 사례도 다양하게 증가하고 있어 보안의 중요성도 더욱 커지고 있다. 최근, Gray-Scale 및 RGB 데이터셋을 통해 높은 정확도의 악성코드 시각화 내용을 제시하는 연구[1], 익스플로잇 페이로드 기반 IoT 공격 유형 분류[2], CNN 모델을 이용한 맬웨어 분류[3]와 같이 CNN 모델 이용, 시각화, IoT 기기에서의 분류를 개별적으로 이용한 다양한 연구가 소개되고 있다.

하지만 본 논문에서는 이를 융합해 IoT 기기에 심어진 악성코드를 학습된 CNN(Convolutional Neural Network) 모델을 통해 정상/악성코드의 여부를 판별하여 IoT 보안 취약점을 해결하고자 한다.

2. 제안하는 방법

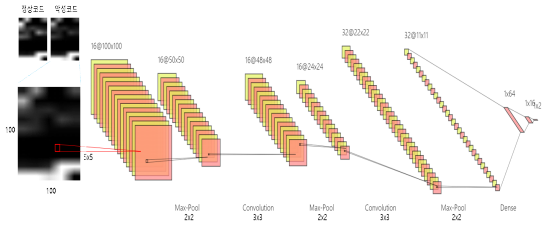
2.1. 데이터 분석

데이터셋은 University of New Brunswick의 Canadian Institute for Cybersecurity에서 제공하는 CIC-MalMem-2022 데이터를 이용하였다.[4] 총 57종의 칼럼으로 구성되어 있으나 카테고리와 클래스를 제외한 나머지 55종에 대한 칼럼에 대해 2차원 이미지를 생성하기 위해 55개의 칼럼으로 정의한다. 정상코드 29,298개, 악성코드 29,298개로 구성되어 있고 악성코드는 트로이목마(Trojan Horse), 스파이웨어(Spyware), 랜섬웨어(Ransomware)가 각 9,487개, 10,020개, 9,791개로 구성되어 있으나 본 논문에서는 멀웨어 데이터를 세분화시켜 구분하지 않고 정상/악성코드로만 분류한다.

2.2. 정상/악성 코드 분류 모델 생성

csv 파일 데이터를 5x11(55개)로 벡터화한 후, 0~255 사이의 값으로 정규화하여 흑백 이미지를 추

<그림 1. Grayscale로 정규화한 정상코드와 악성코드 이미지의 예>



<그림 2. 제안하는 정상/악성코드 분류를 위한 CNN 모델의 구조>

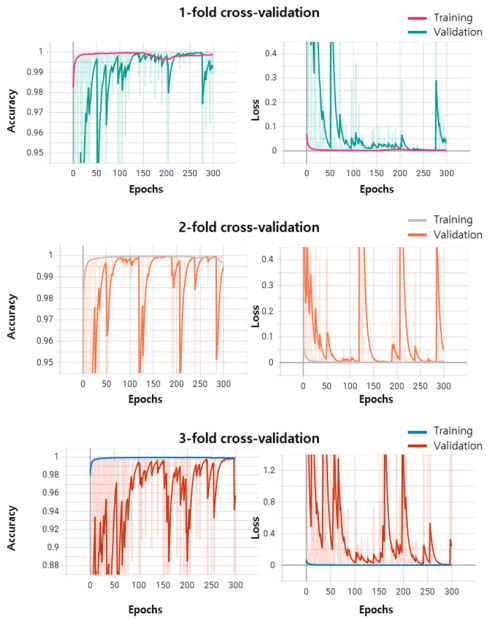
출한다 (그림 1). 해상도 향상을 위해 100x100 입력 사이즈로 업스케일링 한 후, 추출한 이미지를 CNN 모델에 입력한다.

본 논문에서 제안하는 정상/악성코드 분류를 위한 CNN 모델의 전체구조는 그림 2와 같다. Conv2d (5x5), (3x3) Kernel 필터를 통해 합성곱 연산을 진행한 후, 결과의 Overfitting을 해소하기 위해 데이터들의 표준화 및 정규화를 수행하는 Batch Normalization을 진행한다. Feature Map의 Down-scaling과 대푯값 추출을 위해 Max Pooling (2x2)를 Non-overlapping 방법으로 진행하였다. Fully Connected Layer에서는 Flatten()을 활용하여 Feature Map을 1차원으로 변환한 후, Dropout (0.5)를 각 Dense 과정에 넣어주었다. 최종적으로 악성/정상코드의 분류를 위해 Softmax 활성화 함수로 결괏값을 예측한다.

3. 결과

3.1. 모델 검증

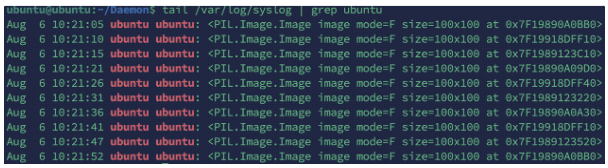
교차검증을 위해 k=3으로 고정하고 이에 대한 결과는 모두 유사한 성능을 보였다 (그림 3). 1-fold에서는 98.84%, 2-fold에서는 99.98%, 3-fold에서는 99.99%의 우수한 성능을 보이며 최종적으로 제안하는 방법은 99.60%의 평균 분류 정확도를 보였다.



<그림 3. k-fold validation 결과>

3.2. IoT 기기에서의 작동

앞에서 구축한 모델과 코드가 잘 작동되는지 실험하기 위해 IoT 기기는 라즈베리파이4 모델 B를 이용하였고, 운영체제는 Ubuntu 22.04.2 버전을 사용하였다. 3만여 개의 악성코드의 다양한 프로세스와 쓰레드를 구분하고 메모리를 덤프하여 데이터를 얻는 것은 실질적으로 오랜 시간이 걸리므로, 데이터셋 리스트에서 랜덤으로 데이터를 추출하여 악성코드를 대체하였다. 학습한 CNN 모델을 기반으로 한 Python 데몬 프로그램을 작성하여 운영체제의 상시 실행 서비스로 등록하였다. 특정 다운로드 디렉토리에 파일이 다운로드 되면 서비스가 이를 감지하여 악성코드를 판별하고 판별 유무는 운영체제 시스템 로그에 기록되며 (그림4), 악성코드로 판별된 경우 파일을 강제로 삭제한다.



<그림 4. IoT 기기에서 실행될 때마다 작성되는 로그>

4. 결론 및 기대효과

본 논문에서는 IoT 기기의 보안 사고를 유발하는 위협 요소 등의 문제를 정상/악성코드를 분류하는 방법에 대해 제안했다. 모델은 99.60%의 평균 분류 정확도를 나타냈으며, 앞서 실험한 IoT 기기에서도 경량화된 프로세스로 잘 작동됨을 확인할 수 있다. 향후, IoT 기기에서도 보안 감시가 항시 이루어질 수 있다는 점에서 편리성 또한 확보할 수 있어 우수한 사용성을 기대할 수 있다.

Acknowledgments

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] 윤채훈, “이미지로 변환한 악성코드 CNN 탐지”, 한국정보과학회 2021 한국소프트웨어종합학술대회, 2021년. pp.377-379.

[2] 김미주, 고웅, 오성택, 이재혁, 김홍근, 박순태, “IoT 기기 취약점 및 익스플로잇 수집을 통한 IoT 공격 유형 연구”, 정보보호학회지, 제 29권, 제 6호, pp.81-88, 2019년.

[3] 카방가, 김창훈, “이미지 기능을 사용하여 맬웨어를 분류하는 CNN 모델” 정보과학회 컴퓨팅의 실제 논문지, 제 24권, 제 5호, pp.256-261, 2018년.

[4] Canadian Institute for Cybersecurity, https://www.unb.ca/cic/, 2023.09.08.

[5] 김문정, 채신록, 홍은기, 황보민, 문유진, “CNN을 활용한 교통 표지판 이미지 분류 인식”, 한국컴퓨터정보학회 학술발표논문집, 제 31권, 1호, pp.317-318, 2023년.

[6] 김영민, 김태연, 한경현, 황성운, “전처리를 이용한 CNN 악성코드 탐지 기법 소개”, 2021년도 대한전자공학회 하계종합학술대회, 2021년, pp.1666-1668.