악성코드 이미지화와 전이학습을 이용한 악성코드 분류 기법

이종관[1*] · 이민우[2]

[1]육군사관학교 · [2]아주대학교

Malware Classification Method using Malware Visualization and Transfer Learning

Jong-Kwan Lee[1,*] · Minwoo Lee[2]

[1]Korea Military Academy · [2]Ajou University

E-mail : jklee64@kma.ac.kr / iminu@ajou.ac.kr

요 약

본 논문은 악성코드의 이미지화와 전이학습을 이용한 악성코드 분류 방안을 제안한다. 공개된 악성코드는 쉽게 재사용 또는 변형이 가능하다. 그런데 전통적인 악성코드 탐지 기법은 변형된 악성코드를 탐지하는데 취약하다. 동일한 부류에 속하는 악성코드들은 서로 유사한 이미지로 변환된다. 따라서 제안하는 기법은 악성코드를 이미지화하고 이미지 분류 분야에서 검증된 딥러닝 모델을 사용하여 악성코드의 부류를 분류한다. Malimg 데이터셋에 대해 VGG-16 모델을 이용하여 실험한 결과 98% 이상의 분류 정확도를 나타냈다.

ABSTRACT

In this paper, we propose a malware family classification scheme using malware visualization and transfer learning. The malware can be easily reused or modified. However, traditional malware detection techniques are vulnerable to detecting variants of malware. Malware belonging to the same class are converted into images that are similar to each other. Therefore, the proposed method can classify malware with a deep learning model that has been verified in the field of image classification. As a result of an experiment using the VGG-16 model on the Malimg dataset, the classification accuracy was over 98%.

키워드

Malware Classification, Transfer Learning, Deep Learning, Visualization

Ⅰ. 서 론

악성코드는 ‘악의적인 목적을 위해 작성된 실행 가능한 코드’로서 프로그램, 매크로, 스크립트 뿐 아니라 취약점을 악용하는 데이터 형태들도 포함한다. FireEye의 2020년 사이버 위협 보고서에 의하면 식별된 악성코드 중 41%는 이전에 알려지지 않은 새로운 것으로 악성코드 탐지 시스템에 의한 탐지를 회피하기 위해 지속적으로 진화하고 있다. 그런데 새롭게 등장하는 악성코드는 기존의 악성코드와 전혀 다른 새로운 유형이 아니라, 기존 악성코드를 일부 재사용하거나 변형한 형태가 대부분이다. 이는 오랜 시간과 노력을 들여 악성코드를 정교하게 제작하는 것보다 단시간 내에 다수의 악성코드를 생산하는 것이 더욱 경제적이기 때문이다. 그런데 이렇게 제작된 악성코드는 시그니쳐(signature) 기반의 전통적인 탐지 시스템을 무력화하는데 유용하다. 악성코드의 일부만 변형되더라도 해당 악성코드를 식별하는 시그니쳐가 변경되기 때문이다[1].

한편, 동일한 부류에 속하는 악성코드들은 서로 유사한 형태의 이미지로 변환된다[2, 3]. 따라서 이미지 분류에 우수한 성능을 나타내는 딥러닝 모델을 이용하면 변형된 악성코드도 쉽게 탐지할 수 있다. 본 논문은 이러한 아이디어에 기초하여 악성코드 분류 기법을 제안하고 그 성능을 실험을 통해 확인한다.

* corresponding author

II. 제안하는 분류 기법

가. 악성코드 이미지화

Nataraj 등은 악성코드를 이미지화하는 방법을 최초로 제안하였다. 악성코드를 그레이 스케일의 이미지로 변환하면, 다른 부류에 속하는 악성코드들은 패턴이 다른 이미지들로 각각 나타나지만, 같은 부류에 속하는 악성코드들은 서로 매우 유사한 이미지들로 변환된다는 것을 밝혔다. 그림 1은 악성코드를 그레이스케일의 이미지로 변환하는 과정을 나타낸다. 악성코드를 8비트 단위로 구분하여 0~255 값으로 매핑함으로써 이미지를 생성한다.

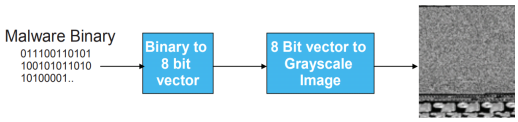

Fig. 1 Process of malware visualization[2]

나. 전이학습

전이학습은 특정 분야에서 학습된 모델의 일부를 타 분야에서 재사용하는 학습기법이다. 딥러닝 모델을 학습하기 위해서는 양질의 풍부한 데이터와 많은 연산장치가 필요하다. 그런데 특정 분야에 최적화된 학습된 모델이 있다면 데이터가 부족한 유사 분야에서 이를 활용하여 쉽게 딥러닝 모델을 구현할 수 있다.

다. 악성코드 분류 절차

악성코드 분류 절차는 악성코드를 이미지로 변환하는 ① 전처리 단계, 이미지로부터 특징을 추출하는 ② 특징추출 단계, 추출된 특징을 통해 악성코드의 부류를 분류하는 ③ 분류 단계로 구성된다. 이때 전처리 단계에서 이미지의 크기는 특징추출 모델의 입력계층에 따라 결정된다.

III. 성능 분석

가. 데이터셋

실험을 위해 Malimg 데이터셋을 사용한다. Malimg 데이터셋은 25종의 악성코드 부류로 구성된 9,340개의 악성코드 샘플을 포함하고 있다.

나. 딥러닝 모델

실험을 위해 사용된 딥러닝 모델은 특징추출 모델과 분류 모델로 구성된다. 특징추출 모델은 VGG-16 모델을 사용하였으며, 분류 모델은 1개의 은닉층을 갖는 fully connected network를 사용하였다. 분류 모델의 은닉계층과 출력계층에는 활성화 함수로 각각 relu, sigmoid를 사용하였으며, 출력계층의 노드 수는 Malimg 데이터셋의 클래스가 25개이므로 25로 하였다.

다. 실험 결과

제안한 기법에 대한 분류 정확도, 정밀도, 재현율, F1-Score는 표 1과 같다.

Table 1. Performance of the proposed method

Accuracy	Precision	Recall	F1 Score
0.9853	0.9851	0.9853	0.9843

IV. 결론 및 향후연구

본 논문은 악성코드 이미지화와 전이학습을 이용한 악성코드 분류 방안을 제안하였다. 제안하는 방안은 악성코드를 이미지로 변환하고 VGG-16를 통해 이미지의 특징을 추출하여 악성코드의 부류를 최종 분류한다. 제안하는 기법을 Malimg 데이터셋에 적용하였을 때 98% 이상의 분류 정확도를 나타냈다. 대량의 악성코드 샘플을 수집하는 것이 쉽지 않다는 점과 빠르게 진화하는 악성코드의 변이 속도를 고려할 때 제안하는 기법이 악성코드에 대응하는데 효과적이라 판단된다.

현재 실험결과는 VGG-16 모델의 가중치를 학습에 사용하지 않고 그대로 사용한 것이다. 향후, 악성코드 샘플로 fine tuning을 하여 VGG-16 모델의 가중치를 보다 최적화함으로써 분류 정확도를 향상시키고 VGG-16 이외의 다양한 모델을 사용했을 때의 성능을 확인할 예정이다.

Acknowledgement

이 논문은 2019년 한국연구재단의 지원을 받아 수행하였음. (No. 2019R1G1A100303013)

References

[1] Prayudi, Yudi, and Imam Riadi. "Implementation of malware analysis using static and dynamic analysis method." *International Journal of Computer Applications*, Vol. 117, No. 6 , 2015, pp. 11-15.

[2] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath. “Malware images: visualization and automatic classification.” *Proceeding of the 8th international symposium on visualization for cyber security*, pp. 1-7, 2011.

[3] M. Song, J. Lee, "Analysis of the impact of the visualization methods on deep learning-based malware classification performance", *Korea Journal of Military Art and Science*, Vol. 77, No. 1, pp. 511-530, 2021.