CNN기반 가상화 난독화된 악성코드 탐지 기법

유동운, 황선진, 손기수, 이혜주, 조은혜, 최윤호

부산대학교

dorong842@gmail.com, unlockable7@gmail.com, sonkisoo97@pusan.ac.kr, gpwn0797@gmail.com, sun00098@gmail.com, yhchoi@pusan.ac.kr

CNN based Virtualized Obfuscated malware detection technique

Dongwoon Yoo, Seon-Jin Hwang, Ki Soo Son, Hye Ju Lee, Jo Eunhye, Yoon-Ho Choi

Pusan National University

요 약

본 논문에서는 가상화 난독화된 악성 코드에 등장에 따른 가상화 난독화된 악성코드 탐지의 필요성을 보이고 가상화 난독화된 악성코드 탐지를 위해 CNN기반의 탐지 기법을 제시한다. 해당 기법은 이미지화한 가상화 난독화 데이터를 이용해 CNN 분류 모델을 학습시키고 학습된 CNN 분류 모델을 통해서 가상화 난독화된 악성코드를 탐지하는 모델이다. 해당 기법에서 사용하는 CNN 분류 모델은 customCNN, ResNet과 InceptionNet을 softvoting을 통해서 만든 Ensemble 분류 모델이며, Ensemble 분류 모델을 사용하여 가상화 난독화된 악성코드 탐지를 수행하고 결과를 분석하여 가상화 난독화된 악성코드 탐지에 CNN 분류 모델을 사용하는 것의 효과를 검증한다.

Ⅰ. 서 론

한국인터넷진흥원의 2023년 상반기 사이버 위협 동향 보고서에 의하면 악성코드는 2022년 하반기 전체 침해사고에서 33.2%를 차지했으며, 2023 상반기 전체 침해사고에서 23.5%를 차지했다[1]. 악성코드가 전체 침해사고에서 차지하는 비율은 2023년에 약 10% 감소하였으나, 악성코드는 여전히 주요 침해사고 유형이며, AhnLab[2]에서 발표한 바에 따르면 최신 악성코드의 유포는 주로 이메일을 통해서 이뤄지기 때문에 신종 악성코드를 탐지하기 위해 발견된 악성코드에 대한 분석이 꾸준히 이뤄지고 있으며, 이를 통해 악성코드 탐지를 위한 다양한 방법이 제시되고 있다. 하지만 악성코드 개발자들도 악성코드 탐지를 회피하기 위해서 가상화 난독화와 같은 회피기법을 사용하고 있다. 가상화 난독화는 프로그램의 바이너리를 보호하기 위한 정보보호기술로, 바이너리의 일부를 다른 아키텍처의 명령어로 치환하고 가상머신을 통해서 해당 명령어를 실행한다. 바이너리를 치환하기 때문에 기존 동작을 유지하고 원본 바이너리와 다른 형태를 가지게 된다. 이 때문에, 악성코드에 가상화 난독화 기술을 이용하게 되면 원본 바이너리와 다른 형태를 가지기 때문에 기존의 악성코드 탐지방법을 회피할 수 있다. 현재 가상화 난독화된 악성코드에 관한 연구는 가상화 난독화된 프로그램에 대한 복원에 대해서 중점적으로 이뤄지고 있어, 가상화 난독화된 악성코드를 탐지하는 방법에 대한 연구는 미흡하다.

따라서 본 논문에서는 가상화 난독화된 악성코드 탐지를 위해서 기존의 악성코드 탐지[3][4]에서 효과적이었던 CNN모델을 사용하는 가상화 난독화된 악성코드의 탐지 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 배경지식을 서술하고, 3장에서는 제안하는 탐지 기법에 관해서 서술한다, 4장에서 제안하는 탐지 기법의 실험 결과를 분석하고, 5장에서는 본 논문의 결론을 서술한다.

Ⅱ. 배경지식

Ⅱ-1. 가상화 난독화

가상화 난독화는 프로그램에 대한 리버싱을 통한 정보유출을 막기 위한 소프트웨어 보호 기술이다. 가상화 난독화 적용시 사용자가 지정한 범위의 바이너리를 다른 아키텍처의 바이트코드로 치환하며, 바이트코드를 실행하기 위한 가상머신을 소프트웨어로 구현하여 코드에 삽입한다. 복잡한 연산을 지원하는 바이너리를 바이트코드로 치환하기 때문에 코드의 전체 길이가 늘어나며, 가상머신을 추가하는 것, 역시 코드의 길이를 늘이며, 프로그램 실행의 부하가 늘어나게 된다. 이처럼 가상화 난독화 기술은 프로그램의 동작을 느리게 하는 단점이 존재하지만, 프로그램 바이너리를 보호하는 데 효과적이라는 커다란 장점이 있다.

Ⅱ-2. ResNet & InceptionNet

ResNet[5]은 마이크로소프트에서 개발한 알고리즘으로 해당 알고리즘은 CNN의 layer를 너무 깊게 쌓았을 때 발생하는 degradation 문제를 해결하여 152층의 layer를 쌓아 2015년 ILSVRC에서 우승을 차지하였다. ResNet의 가장 중요한 특징은 Residual block으로 기존 방식에 skip connection을 추가한 Residual block으로, H(x)=y가 되도록 하는 기존 신경망을 H(x)-x=F(x)로 만들어 F(x)=0으로 학습시켜 미지의 y가 아닌 0을 목표로 주어 학습을 쉽게 만들었다. skip connection은 x를 사용하기 위한 것으로 일종의 short cut이 되어 optimal한 x가 존재한다면 이후의 layer를 건너뛰는 것으로 다중 layer의 문제를 해결하였다.

InceptionNet[6]은 구글이 발표한 모델 구조로 CNN구조의 모델이 커짐에 따라 들어가는 연산량을 줄여 더 깊고 큰 모델을 적은 연산량으로 만들 수 있게 만든 모델이다. InceptionNet의 Inception module은 각각에 모듈에서 1*1 convolution으로 채널을 축소하고 각각의 convolution을 병렬적으로 수행한다. convolution을 여러개로 분해하는 것으로 파라미터 수를 감소시켜 연산량을 줄이도록 하였다.

Ⅲ. CNN기반 가상화된 악성코드 탐지 기법

본 논문에서 제안하는 가상화된 악성코드 탐지 기법의 개요는 그림 1과

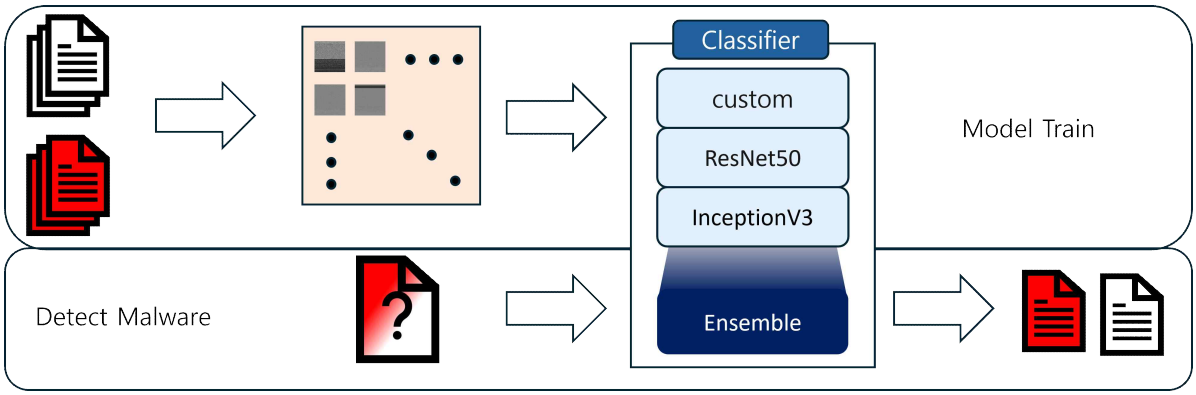

그림 1 제안하는 탐지 기법 개요

같다. 먼저 데이터셋에 VMProtect 3.8버전을 이용해 가상화 난독화를 적용하여 가상화된 데이터셋을 만든다. 이후 CNN모델에 입력으로 사용하기 위해서 가상화 난독화된 데이터셋을 이미지화 한다. 데이터의 이미지화는 각 데이터의 바이너리를 참고하여 바이너리를 10진수 벡터를 만들고 만들어진 10진수 벡터를 기반으로 각 픽셀별로 0~255사이의 값을 가지는 흑백 이미지를 생성하는 방식이다. CNN 분류 모델은 custom, ResNet50, InceptionV3모델을 사용한다. Custom 모델은 간단한 CNN모델로 conv 레이어와 maxpooling 레이어 3개를 사용하여 구성하였다. ResNet50과 InceptionV3는 imagenet에서 미리 학습한 모델을 통해서 구성하였으며. 3가지 CNN 모델 모두 Adam Optimizer와 활성함수로 softmax를 사용하였다. 마지막으로 학습이 모두 끝난 모델들을 softvoting을 통해서 하나의 Ensemble 분류 모델을 만들었고 만들어진 Ensemble 분류 모델을 이용하여 악성코드 탐지를 진행한다.

Ⅳ. 실험결과

제안 기법은 Windos 11 pro 64bit, AMD Ryzen 5 3600, NVIDIA GeForce GTX 1600, 16GB에서 실험을 진행하였다. 실험 데이터는 정상코드 167개와 악성코드 176개를 이용하였으며 각각의 데이터는 4:1비율로 학습과 검증데이터로 나누어 실험을 진행하였다. 표 1는 실험 결과이다.

표 1 가상화 난독화된 악성코드 탐지 결과

CNN 분류 모델	정확도(%)
custom	50
ResNet	61.76
InceptionNet	63.24
Ensemble	66.18

실험 결과 Ensemble 모델에서 66.18%로 가장 높은 정확도를 보이는 것을 볼 수 있으며, InceptionNet이 63.24%로 2번째로 높은 정확도를 보인다.

Ensemble 모델에서 가장 높은 정확도를 보였지만, 66.18%의 정확도는 실제 악성코드 탐지에 사용하기에는 너무 낮은 정확도이므로 제안하는 탐지 기법은 실제 가상화 난독화된 악성코드의 탐지에 사용하기에는 부적절할 것으로 보인다. Ensemble 모델에서도 절대적인 성능이 낮게 측정된 이유는 가상화 난독화를 적용함에 따라서 악성코드와 정상코드 사이에서 볼 수 있는 특징 간의 차이점이 가상화 요소에 의해 대부분 가려지게 되고, 이에 따라 이미지화한 데이터에서 악성과 정상 사이의 차이를 경계가 희미해지기 때문으로, 이에 따라 Ensemble 모델에서도 탐지 정확도가 낮게 측정된다.

Ⅴ. 결론

본 논문에서는 가상화된 악성코드의 탐지를 위한 CNN기반의 탐지 기법을 제시하였다. 해당 탐지 기법은 기존의 악성코드 탐지에서 높은 성능을 보인 탐지 기법이었으나, 가상화된 악성코드의 탐지에서는 좋은 성능을 보이지 못함을 알 수 있었다. 이에 따라 향후 연구에서는 일반적인 악성코드에서 사용된 탐지방법이 아닌 난독화된 악성코드의 탐지에서 사용되는 방법들[7]을 이용하여 가상화된 악성코드를 탐지하는 연구가 수행되어야 할 것이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(No. RS-2023-00217689)이며, 본 연구는 한국전력공사의 사외공모 기초연구 (개별)에 의해 지원되었음 (과제번호: R22XO01-3)

참 고 문 헌

[1] 한국인터넷진흥원. “2023년 상반기 사이버 위협 동향 보고서” (https://www.kisa.or.kr/skin/doc.html?fn=20230822_165603_185.pdf&rs=/result/2023-08/)

[2] AhnLab. “최신 악성코드 동향, 해커들은 이것을 주로 활용했다”, 2023 (https://m.ahnlab.com/ko/contents/content-center/34146)

[3] Yeo, Minsoo, et al. "Flow-based malware detection using convolutional neural network." 2018 International Conference on Information Networking (ICOIN). IEEE, 2018.

[4] Catak, Ferhat Ozgur, et al. "Data augmentation based malware detection using convolutional neural networks." Peerj computer science 7 (2021): e346.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 2016, Deep Residual Learning for Image Recognition, CVPR 2016

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet et al. 2015, Going Deeper With Convolutions, CVPR 2015

[7]Kim, Jin-Young, and Sung-Bae Cho. "Obfuscated malware detection using deep generative model based on global/local features." Computers & Security 112 (2022): 102501.