악성트래픽의 실시간 탐지를 위한 CNN기반 분류 모델 비교 및 분석

이민성, 박지태, 최정우, 최동근*, 김명섭

고려대학교

{min0764, pjj5846, choigoya97, tmskim}@korea.ac.kr *dkchoi@doctorsoft.co.kr

Comparison and analysis of CNN-based classification models for real-time detection of malicious traffic

Lee Min-Seong, Park Jee-tae, Choi jeong-woo, Dongkeun Choi*, Kim Myung-Sup

Korea Univ., *Doctorsoft

요 약

네트워크 환경이 성장함에 따라 네트워크 환경을 위협하는 악성 트래픽도 성장하고 있다. 정보 보호의 중요성이 증가하면서 악성 트래픽 탐지 및 분류가 보안 분야에서 지속적으로 연구가 이루어지고 있다. 악성 트래픽을 분석하는 방법으로 포트 기반, 페이로드 기반, 통계 기반의 시그니처를 정의하여 악성 트래픽을 탐지하고 분석하는 연구들이 진행되었지만, 최근 암호화된 데이터의 전송으로 암호화된 트래픽이 늘어남에 따라 악성 트래픽도 암호화된 상태로 침입하여 공격하는 경향을 보이고 있다. 이에 따라 시그니처 기반의 악성 트래픽 탐지보다 머신 러닝 및 딥 러닝을 이용한 악성 트래픽 탐지 및 분류에 대한 연구가 활발하게 이루어지고 있다. 머신 러닝 및 딥 러닝 기반의 악성 트래픽 탐지는 탐지 정확도에 중점을 두고 있으며, 마찬가지로 트래픽 분류에서도 분류 정확도에 중점을 두고 있다. 트래픽을 분류한다는 측면에서 분류 정확도도 중요하지만 네트워크 트래픽의 실시간성을 고려하면 분류 정확도뿐만 아니라 빠르게 탐지할 수 있는 방법도 중요하다. 본 논문에서는 CNN 및 LSTM기반의 악성 트래픽 분류 방법을 사용하여 분류 정확도 및 분류 시간을 비교하였다.

Ⅰ. 서 론

네트워크 환경이 성장함에 따라 네트워크 환경을 위협하는 악성 트래픽도 성장하고 있다 . COVID-19로 인하여 재택근무 환경이 늘어남에 따라 인터넷 트래픽의 사용량이 증가하고 있다. 이러한 환경에서 정보 보호의 중요성이 나날이 증가하고 있고 악성 트래픽의 탐지 및 분류 연구가 보안 분야에서 지속적으로 이루어지고 있다.

악성 트래픽 분석 방법으로는 포트 기반 방법, 시그니처 기반 방법, 통계 기반의 분석 방법을 이용하여 시그니처를 정의하고 정의된 시그니처로 악성 트래픽을 탐지하고 분류하는 연구 방법이 많이 사용되었다. 안전한 데이터 전송의 중요성으로 인해 대부분의 트래픽들이 SSL/TLS 프로토콜을 사용하여 암호화된 데이터 전송 방식을 사용하고 있다. 발전하는 네트워크 환경에 맞춰 악성 트래픽들도 암호화된 트래픽이 늘어남에 따라 악성 트래픽도 암호화된 상태로 침입하여 공격하는 경향을 보이고 있다. 악성 트래픽이 암호화되어 발생하기 때문에 기존에 사용되던 포트 기반, 페이로드 시그니처 기반의 트래픽 탐지 및 분류 방법들이 사용하기 어려워지고 있다. 443번 포트로 고정되어 있어 포트 기반의 분류를 하기 어려우며, 페이로드 자체도 암호화되어 있기 때문에 페이로드 시그니처를 정의하여 악성 트래픽을 탐지하기 어렵다. 이러한 이유로 통계 기반의 방법을 적용하면서 머신 러닝 및 딥 러닝을 이용한 악성 트래픽 탐지 및 분류 연구가 활발하게 진행되고 있다.

현재 진행되고 있는 머신 러닝 및 딥 러닝을 이용한 악성 트래픽 탐지 및 분류 연구들은 알고리즘의 탐지 및 분류 정확도에 중점을 두고 있다. 여러 다양한 분야에서 머신 러닝 및 딥 러닝을 이용한 연구가 활발하게 진행되고 있고, 각 분야에서 높은 탐지 및 분류 정확도를 보이고 있다. 하지만 악성 트래픽의 경우 가능한 빠르게 탐지하고 대처를 하는 것이 중요하다. 공격 트래픽이 들어왔을 때 대비를 하기 위해서는 악성 트래픽의 종류가 어떤 종류인지 파악하는 것이 필수적이다. 기존에 발생한 악성 트래픽의 경우 공격이 발생한 트래픽을 분석하여 공격의 패턴을 정의하고 같은 공격이 발생하였을 때 대비를 할 수 있다, 하지만 실시간 네트워크 상황에서 발생하는 사이버 공격들은 기존 방식의 공격을 변환하여 새로운 공격 방식으로 공격하기 때문에 공격이 발생하였을 때 탐지나 분류 알고리즘이 새롭게 트래픽을 받아들여 학습하고 다음 공격에 대비할 수 있어야 한다. 다음 공격에 대비하기 위해서는 알고리즘이 탐지나 분류를 할 수 있도록 빠른 시간 안에 패턴을 정의할 수 있어야 한다. 본 논문에서는 악성 트래픽 분석 분야에서 실시간으로 탐지 및 분류 할 수 있는 머신 러닝 모델 및 딥 러닝 모델에 대하여 알고리즘을 정의하고 정의된 알고리즘이 트래픽을 학습하고 분류를 하는 전체적인 시간에 대한 연구 결과를 언급한다.

서론에 이어 본론에서는 연구에 사용된 악성 트래픽 데이터 셋과 사용된 알고리즘에 대해 언급한다. 3장에서는 탐지 정확도와 탐지 시간을 비교한 실험 결과에 대해서 언급하고, 4장에서는 결론 및 향후 연구에 대하여 언급한 뒤 논문을 마친다.

Ⅱ. 본 론

본 장에서는 IoT 환경에서 발생한 악성 트래픽 및 정상 트래픽을 포함한 공공 데이터 셋에 대하여 설명한다. 해당 데이터 셋에 대한 설명 이후 데이터 셋을 사용하여 악성 트래픽을 탐지하고 결과를 비교 분석하기 위한 탐지 알고리즘 대하여 설명한다.

2.1 데이터셋

본 연구에서 사용된 데이터 셋은 IoT 환경에서 발생한 악성 트래픽을

※ 이 논문은 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발)

모아놓은 공공 데이터 셋을 사용하였다. IoT-23 Dataset이라고 명명된 이 데이터 셋은 13가지의 악성 트래픽과 정상 트래픽을 포함한 데이터 들이 Pcap파일로 구성되어 있다. 각각의 악성 트래픽 및 정상 트래픽들이 Pcap 파일로 구성된 데이터 셋에서 4가지 악성 트래픽인 Harai, Mirai, Torii, Trojan과 1가지 정상 트래픽을 분류 할 수 있도록 데이터 셋을 구성하였다.

표 1. IoT-23 데이터 셋

Name	Duration(hours)	Flows
Hakai	24	10,404
Mirai	24	11,162
Torii	24	6,497
Trojan	8	4,427
Normal	24	1,983

분류 알고리즘에서 CNN알고리즘을 기반으로 사용하기 때문에 전처리 과정을 통해 악성 트래픽의 플로우들을 이미지 파일로 변환하여야 한다. 공공 데이터 셋의 Pcap파일을 JSON 파일로 변환하여 플로우 데이터를 저장한다. 플로우를 구성하는 패킷 데이터들은 각자 다른 길이로 구성되어 있다. 악성 트래픽의 특징을 포함한 데이터를 구성하기 위하여 플로우의 패킷 데이터 중 앞부분에 해당하는 데이터를 추출하여 저장한다. 저장한 데이터를 20*20 사이즈의 이미지 데이터로 구성하여 전처리과정을 마친다. 3D CNN은 한 플로우에서 발생하는 패킷 중 첫 10가지 패킷을 사용한다. 각 패킷의 400바이트를 20x20 사이즈의 이미지로 구성하고 10가지 패킷을 적용하여 20x20x10의 3D 데이터로 구성한다.

2.2 분류 알고리즘

머신 러닝 및 딥 러닝 분야에서 탐지 및 분류 정확도와 탐지 및 분류 시간은 트레이드 오프 관계라고 볼 수 있다. 탐지 및 분류 정확도를 높이기 위해 학습 모델을 추가하고 구체화 한다면 그만큼 탐지 및 분류하는 시간이 늘어나게 되고, 시간적인 부분이 중요시되는 분야에서는 탐지 모델을 최대한 가볍게 하고 정확도를 타협하여 낮추는 방법을 적용하고 있다. 악성 트래픽은 탐지 정확도가 높아야 공격에 대처할 수 있지만 공격이 발생하였을 때 얼마나 빠르게 대처할 수 있는가에 대한 문제도 고려해야 한다. 가장 좋은 모델은 탐지 정확도가 높으면서 탐지 시간이 낮은 모델이라고 할 수 있다.

본 연구에서 사용한 분류 알고리즘은 이미지 처리에서 많이 사용되고 있는 CNN 알고리즘을 기반으로 모델을 구성하였다. 이 중 RNN 알고리즘을 추가하여 모델의 구성이 무거워졌을 때 탐지 및 분류 정확도 및 시간의 변화가 어느 정도 일어나는 지 파악한다.. CNN 기반의 3가지 알고리즘과 RNN 알고리즘을 조합한 총 4가지 분류 알고리즘을 사용하여 모델을 구성하였다. 트래픽을 분석하는 방법들을 다양한 비교를 위하여 CNN알고리즘에서는 기본적인 2D CNN알고리즘을 사용하였고, 추가적으로 3D CNN알고리즘을 사용하였다. 또한, 분류 시간이 짧으면서 좋은 분류 정확도를 얻기 위하여 다중 2D CNN알고리즘을 사용하였고, 다중 2D CNN에 RNN 알고리즘 중 LSTM 알고리즘을 추가한 모델을 사용하였다.

Ⅲ. 실험 결과

전처리된 데이터의 이미지를 이용하여 4가지 알고리즘 모델을 학습하고 분류 정확도와 분류 시간을 측정하였다. 모델들의 분류 시간을 측정하기 위하여 데이터를 학습하는데 10000번의 epoch를 고정으로 두었다.

가장 기본적인 2D CNN의 경우 %의 분류 정확도로 분류가 가능하였다. 3D CNN은 71.58%의 분류 정확도를 나타냈다. Multiple CNN 모델은 67.23%의 정확도로 3D CNN보다는 낮은 정확도를 나타내었다. 마지막으로 Multiple CNN에 LSTM을 추가한 모델은 85.47%로 가장 높은 정확도를 보였다. 모델을 구체적으로 구성하는 경우 악성 트래픽을 분류하는데 높은 정확도를 보이는 것을 알 수 있다.

실험 결과 가장 빠른 분류 시간을 보이는 모델이 Multiple CNN 모델이다. Multiple CNN 모델을 기준으로 나머지 3가지 모델의 분류 시간을 비교한 결과 3D CNN 모델은 분류하는 데 4%의 시간이 더 걸렸으며, 2D CNN은 11% 더 증가된 시간이 측정되었다. 분류 정확도가 높은 Multiple CNN과 LSTM으로 구성된 모델은 2.5% 더 증가되었다.

표 2. 실험 결과

Model	Accuracy	Time
2D CNN	83.97%	11%
3D CNN	71.58%	4%
Multiple 2D CNN	67.23%	-
Multiple 2D CNN + LSTM	85.86%	2.5%

실험 결과 모델의 분류 정확도는 모델을 구성함에 따라서 좋은 정확도를 나타낼 수 있다. 하지만 측정 시간은 모델이 추가됨에 따라서 더 증가된 분류 시간을 가지게 된다. 악성 트래픽을 분류하기 위한 상황에서 높은 정확도가 필요하다면 Multiple CNN에 LSTM 모델을 사용하는 것이 좋지만 빠른 탐지를 위해서는 Multiple CNN 모델을 사용하는 것이 좋다. 실시간 네트워크 상황에서는 트래픽의 양이 많고, 트래픽의 종류가 다양하기 때문에 분류 시간이 더 차이가 날 수 있다.

Ⅳ. 결론 및 향후 연구

본 논문은 네트워크의 실시간성을 고려하여 공격이 발생하였을 때 빠르게 공격을 식별하고 대처할 수 있는 분류 알고리즘에 대한 연구이다. 정상 트래픽과 악성 트래픽을 분류하는 탐지 알고리즘들의 탐지 정확도와 탐지 시간을 비교한다. 탐지 및 분류 모델을 구성할 때 정확도를 높이기 위하여 모델을 추가하여 탐지 시간이 증가하는 경우와 탐지 시간을 빠르게 하기 위하여 모델을 가볍게 하는 경우를 모두 고려하여 모델을 구성할 수 있다. 네트워크 상황에 따라 비교한 알고리즘들 통해 탐지 정확도를 타협하면서 탐지 시간이 조금 더 빠른 알고리즘을 선택하여 모델을 구성할 수 있도록 한다.

향후 연구로는 악성 트래픽을 포함한 다양한 공공 트래픽에 대한 분석과 함께 탐지 및 분류 정확도를 높이면서 탐지 시간을 빠르게 할 수 있는 모델에 대한 연구를 진행한다.

참 고 문 헌

[1] Garcia S, Parmisano A, Erquiaga MJ. In: Zenodo. Iot-23: a labeled dataset with malicious and legitimate iot network traffic (version 1.0.0) [data set]; 2020. doi: 10.5281/zenodo.4743746

[2] 지세현, 박지태, 백의준, 김명섭,(2018). 안전한 네트워크 구축을 위한 컨볼루션 신경망 기반 악성트래픽 탐지. 한국통신학회 학술대회논문집, pp861-862.