딥 러닝을 이용한 시드 기반 악성 트래픽 탐지

백의준, 박지태, Huru Hasanova, 김명섭

고려대학교 컴퓨터정보학과

{pb1069, pjj5846, hhuru, tmskim}@korea.ac.kr,

요 약

오늘날의 네트워크가 급속하게 성장하고 네트워크 기능이 다양한 응용 및 서비스 개발에서 중요한 요소로 자리잡음과 동시에 네트워크 환경을 위협하는 다양한 악성 트래픽이 발생하고 있다. 이러한 악성 트래픽은 네트워크 환경에 막대한 피해를 입히므로 악성 트래픽 탐지 연구는 네트워크 관리 분야에서 필수 불가결하다. 이에 본 논문에서는 딥 러닝을 이용한 시드 기반 악성 트래픽 탐지 방법을 제안한다. 또한 다양한 악성 트래픽 탐지 실험 결과를 통해 기존 악성 트래픽 탐지 모델과 비교하고 평가한다.

1. 서 론

오늘날의 네트워크가 급속하게 성장하고 네트워크 기능이 다양한 응용 및 서비스 개발에서 중요한 요소로 자리잡고 있다. 동시에 네트워크 환경을 위협하는 다양한 악성 트래픽이 발생하고 있다. 이러한 악성 트래픽은 네트워크 환경에 막대한 피해를 입히므로 악성 트래픽 탐지 연구는 필수적이다. 기존에 악성 트래픽 탐지에 대한 연구가 진행되었으나 정확한 탐지에 한계점을 지닌다. [1]에서는 악성 트래픽을 플로우 형태로 변환하여 이 중 한 플로우에서 5-tuple 를 추출하고 이를 Seed 라고 정의한다. 이 Seed 와 다른 플로우 간 유사성, 연결성을 계산하여 악성 트래픽 플로우를 탐지하고 이 과정을 연속적으로 반복하여 탐지를 진행한다. 하지만 가중치의 조절을 통한 연결성 계산이 정교하지 못하여 악성 트래픽 플로우가 가진 다양한 특성을 모두 반영하기 힘들고 가중치를 조절하는 과정이 Brute-Force 로 진행하기 때문에 시간-복잡도 측면에서 효율적이지 못하다. 이에 본 논문에서는 다수의 Layer 구성을 통해 다양한 패턴 인식에 강점을 가지고 Back-Propagation 을 통해 효율적으로 가중치를 조절할 수 있는 딥 러닝을 이용한 시드 기반 악성 트래픽 탐지 방법을 제안한다.

본 논문은 1 장 서론에 이어, 2 장 본론에서 수집한 데이터의 구조와 데이터 추출 및 전처리 과정, 탐지 모델의 구조, 탐지 모델을 학습하고 테스트하는 과정에 대해 서술한다. 3 장 실험결과에서 임의로 구성된 실험 트레이스 별 탐지 정확도를 표로 보이고 실험결과를 분석 및 설명한다. 마지막 4 장에서 결론 및 향후 연구에 대해 서술하고 본 논문을 마친다.

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원(No.2015R1D1A3A01018057) 및 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(No. 2017-0-00513, Security Analytics 기반의 이기종 보안솔루션 위협 분석 및 대응 기술 개발)을 받아 수행된 연구임.

2. 본 론

본 장에선 수집한 트래픽으로부터 데이터를 추출하고 학습을 위해 변환하는 과정, 딥 러닝 기반 탐지 모델의 구조 그리고 탐지 모델을 학습하고 테스트하는 과정에 대해 서술한다.

실험 데이터는 악성 트래픽과 정상 트래픽으로 구성되며 악성 트래픽은 [2]로부터 수집하고 정상 트래픽은 Wireshark 를 이용하여 다양한 응용을 통해 수집한다. 수집한 트래픽을 플로우의 형태로 변환하고 IP(Destination), Port(Source, Destination), Protocol 정보를 추출한다. 추출한 데이터는 정상 및 악성 여부에 따라 0, 1 로 라벨링되며 구조는 그림 1 과 같다.

Data Structure

Destination IP [0:4]				Src_Port	Dst_Port	Protocol	Label
Example of Data							
239	255	255	250	2844	443	6	0

그림 1. 추출한 데이터의 구조 및 예시

추출한 플로우 데이터는 Seed 를 포함하는 Source 플로우와 Target 플로우로 구성되며 수식 1 과 같다.

(1) $$SourceFlow = \{x | x \in Malicious\ Traffc\}$$
$$TargetFlow = \{x, y | x \in Malicious\ Traffic, y \in Normal\ Traffic\}$$

Source 플로우와 Target 플로우 간 연결성 계산을 위해 Source 플로우와 Target 플로우를 그림 2 와 같은 구조로 병합한다.

Data Structure

Source Flow	Target Flow	Label

그림 2. 병합된 플로우 데이터의 구조

Source 플로우와 Target 플로우가 병합된 플로우로 변환될 때 Source 플로우, Target 플로우의 정상, 악성 여부에 따라 새롭게 라벨링되며 그림 3 을 통해 간략히 설명한다. 플로우 데이터 병합에서 [악성,정상] 꼴의 형태는 0 으로 라벨링되며 이는 악성 플로우로부터 정상 플로우는 검출되지 않도록 학습시키는 것이다.

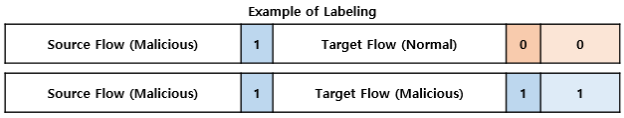

그림 3. 플로우 데이터 병합 예시

또한 [악성, 악성]꼴의 형태는 1 로 라벨링되며 이는 악성 플로우로부터 악성 플로우를 검출하도록 학습시키는 것이다. 학습은 Multi-Layer Logistic Regression [3]을 사용하였으며 Input Layer, 3 개의 Hidden Layer, Output Layer 로 구성되고 가설 함수는 Sigmoid, Cost 최소화 알고리즘으로 경사하강법을 사용한다. 모델의 간략한 구조는 그림 4 와 같다.

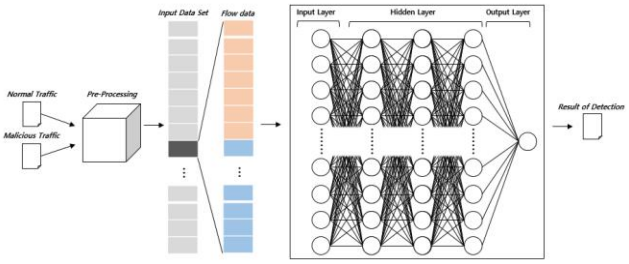

그림 4. 학습 모델 구조

학습 데이터 셋과 정해진 학습 횟수만큼 반복적으로 학습하고 악성 트래픽 탐지에 적합하게 가중치를 조절한다. 모델의 학습 과정은 그림 5 와 같다.

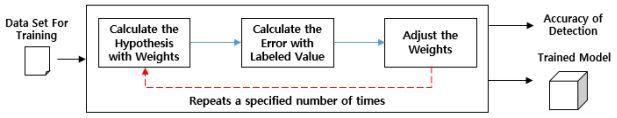

그림 5. 모델의 학습 과정

모델을 학습하는 과정을 마친 후 테스트 데이터 셋과 학습된 모델을 통해 악성 트래픽을 탐지한다. 학습된 모델의 테스트 과정은 그림 6 과 같다.

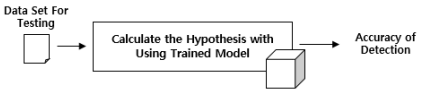

그림 4. 학습된 모델의 테스트 과정

계산된 가설 함수 값을 설정한 기준값(Threshold)에 따라 0,1 로 나누고 이를 데이터의 라벨값과 대조하여 탐지 정확도를 계산한다. 탐지 정확도는 Recall, Precision 으로 표현되며 계산은 수식 2 와 같다.

$$(2)\quad Recall = \frac{detected\ TRUE}{total\ number\ of\ existing\ TRUE}$$
$$Precision = \frac{TRUE\ detections}{whole\ detections\ of\ an\ algorithm}$$

3. 실험 결과

본 장에서는 앞서 생성한 데이터 셋과 학습된 모델을 통해 기존 모델과의 탐지 정확도 비교 결과를 표로 보이고 분석하고 이에 대해 설명한다. 실험 결과는 표 1 의 기존 모델 및 딥 러닝 모델의 학습 과정의 정확도, 표 2 의 기존 모델 및 딥러닝 모델의 테스트 과정의 정확도로 구성되고 Recall 및 Precision 은 트레이스 별 모든 시드 각각에 대한 탐지 결과의 평균으로 계산하였다.

실험한 결과, 딥 러닝을 이용한 모델이 기존 모델보다 학습 과정에서 높은 Recall 및 Precision 을 나타내는 것을 확인하였고 테스트 과정에서도 높은 탐지 성능을 내는 것을 확인하였다. 또한 딥 러닝 기법의 효율적인 가중치 계산으로 기존의 Brute-Force 로 조절하는 것보다 시간-복잡도 측면에서 높은 성능을 내는 것을 확인하였다.

	기존 모델		딥 러닝	
	Recall	Precision	Recall	Precision
T1	0.3535	1.00	1.00	1.00
T2	0.2193	1.00	1.00	1.00
T3	0.4866	1.00	1.00	1.00
T4	0.538	1.00	1.00	1.00

표 1. 기존 및 딥 러닝 모델 간 학습 정확도 비교

	기존 모델		딥 러닝	
	Recall	Precision	Recall	Precision
T1	0.1202	1.00	0.4819	1.00
T2	0.1097	1.00	0.3765	1.00
T3	0.3406	1.00	0.5672	1.00
T4	0.2152	1.00	0.4349	1.00

표 2. 기존 및 딥 러닝 모델 간 테스트 정확도 비교

4. 결 론

본 논문은 딥 러닝을 이용하여 시드 기반으로 악성 트래픽 탐지하는 방법을 제안하고 실험하였다. 실험 결과를 통해 기존 모델의 탐지 방법보다 탐지 정확도에서 좋은 결과를 나타내는 것을 확인하였고 이를 통해 시드 기반 악성 트래픽 탐지에 딥 러닝 기법을 적용할 수 있는 가능성을 제시하였다. 그러나 실제 환경에서 발생하는 다양한 악성 트래픽들을 수용할 수 있는 Coverage 가 작고 실제적인 악성 트래픽들의 행위 분석보다 화이트/블랙 리스트 방법에 가깝게 학습이 되는 한계점을 지닌다. 이에 향후 연구로 명확한 실험 및 결과 검증을 위해 악성 트래픽이 발생하는 실제 환경을 면밀히 분석하고 실제 네트워크에서 발생하는 것과 유사하게 데이터를 수집하고 기존의 다양한 딥 러닝 기법을 적용하여 탐지 알고리즘을 개선하고 고도화할 예정이다.

5. 참고문헌

[1] 박지태, 이성호, 구영훈, Huru Hasanova, “플로우의 연관성 모델과 시드 기반의 연속적인 그룹핑을 이용한 악성 트래픽 탐지”, 한국통신학회 학술대회논문집, pp. 608-609, 2017 년 11 월

[2] “Malware-traffic-analysis.net.”. Malware-Traffic-Analysis.net. last modified Mar 28. 2018, accessed Mar 15. 2018, https://www.malware-traffic-analysis.net/index.html

[3] “GitHub.”. hunkim/DeepLearningZeroToAll. Last modified May 13. 2017, accessed Jan 11. 2018 https://github.com/hunkim/DeepLearningZeroToAll/blob/master/lab-09-2-xor-nn.py