네트워크 이상행위 탐지를 위한 암호트래픽 분석기술 동향

Trends of Encrypted Network Traffic Analysis Technologies for Network Anomaly Detection

최양서 (Y.S. Choi, yschoi92@etri.re.kr) 지능형네트워크보안연구실 책임연구원
유재학 (J.H. Yoo, dbzzang@etri.re.kr) 지능형네트워크보안연구실 선임연구원
구기종 (K.J. Koo, kjkoo@etri.re.kr) 지능형네트워크보안연구실 책임연구원
문대성 (D.S. Moon, daesung@etri.re.kr) 지능형네트워크보안연구실 책임연구원

ABSTRACT

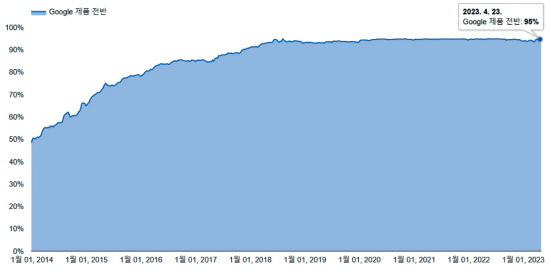
With the rapid advancement of the Internet, the use of encrypted traffic has surged in order to protect data during transmission. Simultaneously, network attacks have also begun to leverage encrypted traffic, leading to active research in the field of encrypted traffic analysis to overcome the limitations of traditional detection methods. In this paper, we provide an overview of the encrypted traffic analysis field, covering the analysis process, domains, models, evaluation methods, and research trends. Specifically, it focuses on the research trends in the field of anomaly detection in encrypted network traffic analysis. Furthermore, considerations for model development in encrypted traffic analysis are discussed, including traffic dataset composition, selection of traffic representation methods, creation of analysis models, and mitigation of AI model attacks. In the future, the volume of encrypted network traffic will continue to increase, particularly with a higher proportion of attack traffic utilizing encryption. Research on attack detection in such an environment must be consistently conducted to address these challenges.

KEYWORDS 기계학습, 딥러닝, 악성행위 탐지, 암호트래픽 분석, 트래픽 이상탐지

Ⅰ. 서론

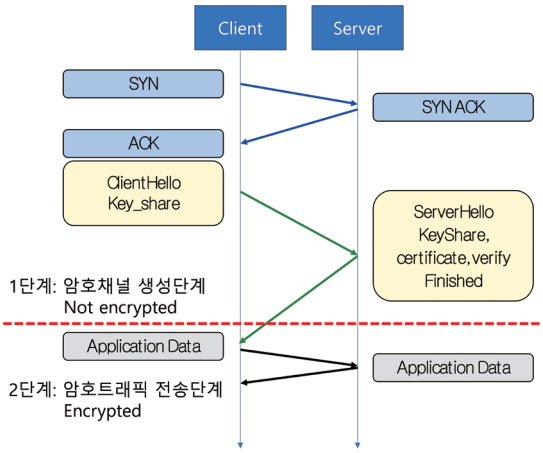
모바일 및 IoT 기기 등의 인터넷 연결로 인터넷 사용자가 폭증하면서, 매우 다양한 서비스가 인터넷을 통해 제공되고 있다. 이러한 서비스들은 은행 계좌번호 등 중요한 개인정보를 주고받기 때문에 송수신 데이터에 대한 암호화 요구가 증대되었고, 암호화된 트래픽 양은 크게 증가하였다. 실제로, 구

* DOI: https://doi.org/10.22648/ETRI.2023.J.380507
* 본 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[No. RS-2023-00235509, ICT융합 공공 서비스·인프라의 암호화 사이버위협에 대한 네트워크 행위기반 보안관제기술 개발].



출처 Reproducted from [1].

그림 1 구글 암호트래픽 규모



출처 Reproducted from [5].

그림 2 TLS 1.3 프로토콜 암호채널 생성 및 암호트래픽 전송 절차

글 서비스상의 2015년 암호화된 트래픽 비중은 약 50% 정도였으나, 2019년에는 95%로 크게 증가했음을 알 수 있다(그림 1 참고)[1].

또한, 웹 트래픽뿐만 아니라 클라우드, 이메일 등 응용프로그램이 자체적으로 암호화된 트래픽을 사용하고 있으며, 더 나아가서 VPN 서비스, 토르(ToR) 네트워크[2] 등과 같은 다양한 암호화 프로토콜이 제공되면서 일반 인터넷 서비스뿐만 아니라 사이버 공격에도 암호화된 트래픽 활용이 크게 증가되고 있다. Zscaler's 2020 Encrypted Attacks Report[3]에 따르면, 암호화된 네트워크 공격 트래픽은 2019년도에 비해 260% 증가하였다고 한다.

사이버 공격이 암호화된 트래픽을 활용하면서 기본 페이로드 데이터를 활용한 시그니처기반의 공격 탐지는 불가능하게 되었고, 이를 극복하기 위한 암호화된 트래픽 분석기술 연구가 활발히 진행되고 있다. 이에 본고에서는 암호트래픽 분석기술 동향과 해당 기술을 연구함에 있어서 고려해야 하는 사항에 대해 제시한다.

II. 암호트래픽 분석 프레임워크

일반적인 암호화 프로토콜은 그림 2와 같이, 암호화 채널 생성단계를 통해 암호화 통신을 위한 협상을 진행하고, 협상된 암호화 방식을 활용하여 실제 암호화된 네트워크 트래픽을 전송하게 된다. 이렇게 구성된 암호화된 네트워크 트래픽 분석은 1) 암호트래픽 수집, 2) 암호트래픽 특성 추출, 3) 암호트래픽 분석, 4) 암호트래픽 분석 성능 평가 등의 과정을 거치게 된다.

1. 암호트래픽 수집

암호화된 네트워크 트래픽 수집은 기존의 네트워크 트래픽 수집 방식과 동일하게 스위치, 라우터 또는 게이트웨이 등과 같은 네트워크 노드상에서 수집하거나, 물리적으로 네트워크 라인 중간에 Tap을 설치하여 네트워크 트래픽을 수집할 수 있다. 특정 서버 혹은 호스트 내에서도 트래픽을 수집할 수 있는데, 이런 경우에는 libpcap, tcpdump[4] 등과 같은 트래픽 수집 라이브러리가 활용된다. 최근에는 10Gbps 이상의 네트워크 트래픽을 실시간으로 저장할 수 있는 고성능 전용 장치도 제공되고 있다.

2. 암호트래픽 특징 추출

암호화된 네트워크 트래픽 분석을 위해 가장 중요한 것은 해당 네트워크 트래픽 특성을 적절히 표현할 수 있는 특징(Feature)을 선택하는 것인데, 이는 일반적으로 플로우(Flow)나 세션(Session) 단위로 수집된다.

플로우는 일반적으로 5-tuple(근원지 IP주소, 목적지 IP주소, 근원지 포트번호, 목적지 포트번호, 프로토콜)이 동일한 패킷들로 구성되는데, 일반적으로 플로우를 표현하는 특징으로는 평균 패킷 개수, 최대 패킷 길이, 플로우 유지시간 등이 있다.

세션은 서버와 클라이언트가 서로 주고받는 양방향 네트워크 트래픽의 집합으로, 하나의 세션에는 다수의 플로우가 속할 수 있다. 세션 단위로 네트워크 트래픽을 수집하는 경우, 특정 행위(예: 로그인 등) 동안 서버와 클라이언트가 주고받는 모든 트래픽을 수집하기 때문에 해당 행위를 표현하는 데 효율적일 수 있다.

네트워크 트래픽 특징에는 수집 대상에 따라 네트워크 패킷기반 특징, 통계적 특징, 그리고 분석 알고리즘에 맞게 네트워크 트래픽을 그래프나 이미지 등 새로운 형태로 표현하는 특징 등이 활용된다.

네트워크 패킷기반 특징은 네트워크 패킷에서 직접적으로 수집 가능한 정보를 의미하는 것으로 일반적으로 패킷의 네트워크 프로토콜 헤더에 존재하는 정보(예: IP주소, 포트번호, 프로토콜, 기타 파라미터 등)를 의미한다.

통계적 특징은 다수의 패킷을 모아 특정 특징(예: 패킷 크기, inter-arrival time, 패킷 개수, 패킷 크기의 합 등)에 대한 통계적인 값(예: 평균, 최대, 최소, 표준편차 등)을 도출하고 이를 특징으로 활용하는 것을 의미하는데, 패킷의 수집 기간이나 개수에 따라 다양한 값을 도출할 수 있다.

- TCP 윈도우 평균 크기
- 근원지 포트번호
- 수신 IP 패킷 헤더 평균 길이
- 수신 패킷 도착시간 최대 간격
- 송신 IP 패킷 헤더 평균 길이
- 송신 패킷 도착시간 최대 간격
- 송신 패킷 길이 표준편차
- 플로우 유지시간
- 송신 트래픽 지속시간
- 세션당 전체 패킷 페이로드 길이 합
- 목적지 포트번호
- 세션 내 패킷 간 도착시간 간격 표준편차
- 세션 내 패킷 간 도착시간 간격 최소값
- 송신 트래픽 도착시간 간격 표준편차
- 수신 트래픽 도착시간 간격 표준편차
- 송신 트래픽 도착시간 간격 최소값
- 수신 트래픽 도착시간 간격 최소값
- 송신 트래픽 도착시간 간격 평균값
- 수신 트래픽 도착시간 간격 평균값
- IP 패킷 길이
- TCP 페이로드 길이
- 각 세션의 패킷 간의 시간 간격
- 수신 패킷 페이로드 전체 길이
- TCP 페이로드 길이 최소값
- TCP 페이로드 길이 평균값
- TCP 페이로드 길이 중간값
- IP 패킷 길이 표준편차
- IP 패킷 비율(IP 패킷의 최대길이 / IP 패킷의 최소길이)
- goodput (세션 내 IP 패킷 전체 길이 / 플로우 유지시간)
- 세션 내 패킷 간 최대 시간 간격
- 수신 패킷 길이 표준편차
- TCP 페이로드 최대 길이
- TTL 평균값
- TTL 표준편차
- 수신 트래픽 지속시간

출처 Reproducted from [5].

그림 3 주요 네트워크 트래픽 특징(Feature)

또한 트래픽을 패킷 도착 순서[6], 그래프[7] 또는 이미지[8] 등으로 인코딩하고 이를 특징으로 활용하기도 한다. 가장 널리 사용되는 특징으로는 그림 3과 같은 것들이 있다.

네트워크 트래픽을 표현하는 특징은 전문가에 의해 직접 선택될 수도 있으나, 딥러닝 알고리즘을 이

용하여 자동으로 추출할 수도 있다.

3. 암호트래픽 분석

암호트래픽은 암호 채널 생성을 위한 협상 과정에서 암호화되지 않는 일부 데이터 외에는 페이로드상의 데이터를 이용한 분석이 불가능하다. 따라서, 페이로드 부분보다는 트래픽 자체에 이상탐지 기술을 적용하여 암호트래픽 분석을 수행한다. 암호트래픽 분석을 위한 방법으로는 전통적인 기계학습(ML: Machine Learning) 기법을 활용한 방법과 딥러닝(DL: Deep Learning) 기법을 활용하는 방법, 그리고 전문가의 지식에 기반한 방법으로 구분될 수 있다[2]. ML과 DL의 가장 큰 차이점은 ML의 경우에는 특징공학(Feature Engineering) 등의 과정을 통한 특징 선정 과정이 선행되어야 한다는 것이고, DL은 그 과정이 생략될 수 있다는 것이다.

최근까지 많은 연구는 정확한 트래픽 분류를 위해 네트워크 트래픽을 표현할 수 있는 의미 있는 정보를 추출하고, 이를 기존 기계학습 모델을 활용하여 분류하는 방법을 수행하였다. 그러나 DL은 효율적인 특징 추출 과정을 생략하고도 상당한 수준의 트래픽 분류 성능을 보이고 있기 때문에 최근에는 2가지 분석 모델을 모두 활용하고 있다.

다만, 네트워크 이상행위 탐지 분야에서는 특정 네트워크 트래픽을 이상 트래픽으로 판단하는 경우, 이에 대한 정확한 판단 근거가 제시되지 않는다면 해당 트래픽에 대한 적극적인 대응(차단 등)을 수행하는 것이 매우 어렵기 때문에 이상 트래픽 판단 근거가 명확하게 제시되는 전통적인 ML 방식이 널리 활용되고 있고, DL 방식을 활용하는 경우에도 XAI[9] 등 탐지 결과를 설명할 수 있는 모델을 활용하려는 연구가 진행되고 있다.

지식기반 방법은 사전 지식을 기반으로 이상 트래픽을 탐지하는 방식을 의미한다. 탐지의 정확도는 사전 지식의 정확도에 따라 크게 달라지고, 새로운 공격이 나타나는 경우 이에 대한 탐지 규칙을 새로 개발해야 한다는 단점이 있다.

4. 암호트래픽 분석 성능 평가

분석 모델 성능 평가 방법 중 가장 널리 활용되는 방법으로는 k-겹 교차검증(k-fold Cross Validation)이 있다. 이는 데이터셋을 k개로 분할한 뒤 k-1개를 학습용 데이터셋으로, 1개를 평가용 데이터셋으로 사용하는 방법으로, 이 방법을 k번 반복하여 k개의 성능 지표를 얻어내는 방법이다.

일반적으로 특정 탐지 모델의 성능은 탐지의 정확도와 학습 및 탐지에 소요되는 시간, 그리고 학습에 활용된 데이터셋과 다른 도메인에 대한 적용 가능성 등이 주요 지표로 활용된다.

탐지 정확도는 일반적으로 그림 4와 같이 Confusion Matrix로 알려져 있는 값들을 기반으로 TPR(True Positive Rate, Recall), FPR(False Positive Rate), Precision, Accuracy, F1 Score, Receiver Operating Characteristic(ROC) curve, Area Under ROC Curve (AUC), Precision-Recall curve(P-R curve)[10]값 등이 활용된다.

두 번째 평가 항목으로는 모델 생성과 트래픽 분석에 필요한 시간에 관한 것으로, 이는 학습과 분석

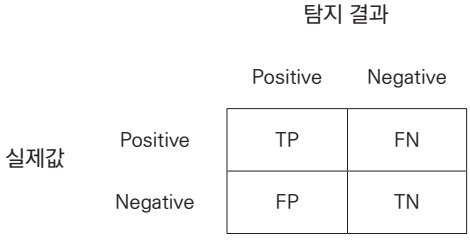


그림 4 Confusion Matrix(오차 행렬)

에 필요한 실제 시간과 이론적으로 측정하는 시간 복잡도가 포함된다.

마지막으로 일반화 가능성과 관련해서는 사전에 구축된 모델은 트래픽 특징이 변화하는 경우 분석 성능이 크게 떨어질 수 있는데, 이러한 상황에 적절히 대응하기 위해서 구축된 모델은 일반화할 수 있는 능력을 갖추고 있어야 한다. 일반적으로 재학습을 통해 모델을 변경하는데, 이때 적절한 성능이 확보될 수 있는 데이터셋의 규모로 일반화 가능성을 평가할 수 있다.

Ⅲ. 암호화된 트래픽 이상탐지

본고에서는 암호화된 네트워크 트래픽에 대한 이상탐지 분야와 암호화된 트래픽을 활용하는 악성파일 탐지 분야에 대해 논한다.

1. 네트워크 트래픽 이상탐지

암호화된 네트워크 트래픽에 대한 이상탐지는 전문가의 지식에 의존하여 특정 공격 징후를 탐지하는 방법, 네트워크 트래픽이 정상과 다른 형태를 보이는 경우 이를 탐지하는 방법 또는 악성행위를 수행하는 네트워크 트래픽을 수집하여 이를 학습함으로써 정상과 악성행위를 분류하는 방식의 탐지 방법을 활용하게 된다. 이러한 이상 상황 판단은 전통적인 ML 방식과 DL 방식 그리고 전문가의 전문 지식에 기반한 탐지 방식으로 수행할 수 있다.

가. 전통적인 기계학습기반 탐지 방식

기존의 전통적인 기계학습 방식의 네트워크 트래픽 이상탐지 기법[2]에서는 이상기반(Anomaly-based) 탐지, 분류기반(Classification-based) 탐지, 그리고 혼합(Hybrid) 탐지 방식으로 구분하고 있다.

이상기반 탐지 방식은 일반적으로 정상 트래픽만을 이용한 비지도학습 방식의 탐지 모델을 생성하고, 이를 기반으로 정상이 아닌 트래픽을 찾아내는 방식을 이용한다. 분류기반 탐지 방식은 정상 트래픽과 이상 트래픽 데이터를 이용한 지도학습 방식을 주로 활용하고, 주어진 네트워크 트래픽이 정상에 속하는지, 이상에 속하는지를 분류하는 방식을 활용한다. 혼합 탐지 방식은 앞선 2가지 방식을 혼합하여 활용하는 방식으로 준지도학습 방식을 활용하는 방법이다.

Stergiopoulos 등[11]은 사이드 채널 특징을 정의하고 이를 7가지 다른 기계학습 알고리즘에 적용하여 성능을 확인하였다. 이 논문은 악성 트래픽 탐지에 초점을 맞추었지만, 암호화된 트래픽 탐지도 포함되어 있었다. 암호화된 데이터셋은 CTU-13[12], FIRST[13], Milicenso[14]에서 추출되었으며, CART 모델에서 99.8%의 정확도를 보였다.

Qin 등[15]은 목적지 주소, 목적지 포트 및 소스 주소, 패킷 크기와 플로우 지속시간의 entropy 값을 k-means 알고리즘을 사용하여 다수의 정상 사용자 행위를 모델링하였고, 생성된 모델에서 벗어나는 정도를 이용하여 DDoS 공격이 발생했음을 판단하였다.

Zolotukhin 등[16]은 DBSCAN 알고리즘과 9개의 특징(Inter-arrival Time 등)을 이용하여 정상 사용자를 모델링하였다. 분석 대상 트래픽이 정상 모델과 임계치 값 이상 벌어지면 DoS 트래픽으로 판단한다.

나. 딥러닝기반 탐지 방식

DL은 강한 특징 추출 능력으로 원시 트래픽 데이터에서 이상탐지에 활용할 수 있는 특징을 자동으로 추출할 수 있어 알려지지 않은 공격 탐지 분야에 적용될 수 있다.

Zeng 등[17]은 DL을 활용한 End-to-End 공격 탐

지 프레임워크를 제안하였는데, 이상탐지를 위해 CNN, LSTM, SAE를 활용하였다.

Zolotukhin 등[18]은 클러스터링 알고리즘과 SAE 알고리즘을 조합하여 DDoS 공격 탐지 모델을 제안하였다. 각각 정의된 시간 간격 내에서 수집된 네트워크 트래픽으로부터 8가지 특징(서로 다른 TCP flag들의 비율 등)을 추출하여 사용하였다. 정상 사용자 행위 모델은 클러스터링 알고리즘을 이용하였고, 클러스터링 알고리즘으로는 탐지가 안 되는 정상 사용자의 웹 브라우징 행위를 모방한 공격을 탐지하는 데는 SAE가 사용되었다.

다. 전문 지식기반 탐지 방식

전문가의 지식에 기반한 네트워크 이상탐지 방식은 일반적으로 이상 여부를 판단하는 규칙에 의존하게 되는데, 이런 규칙들은 보통 네트워크 포트, 프로토콜 헤더 정보 등이 활용된다.

David 등[19]은 특정 연결(세션)에 속하는 플로우 수의 Fast entropy를 계산하여 Fingerprint로 활용하였는데, DDoS 공격 탐지를 위한 임계값을 선정하였다. 그러나 이 방법은 플로우 개수를 활용하기 때문에 Flash crowds와 DDoS를 효과적으로 분류해 내지는 못했다.

2. 악성파일 탐지

과거 악성파일은 공격에 필요한 모든 기능을 포함했기 때문에 해당 악성파일이 특정 시스템에 설치되면 공격자가 원하는 모든 기능을 수행할 수 있었다. 그러나 최근에는 침투에 필요한 기본 기능만을 탑재하여 침투를 시도하고, 침투에 성공하면 다른 기능을 추가적으로 내려받아 업데이트하는 방식을 취한다. 또한, 최근 악성파일은 공격의 최종 목표 시스템에 바로 설치되지 않고, 최종 목표 시스템에 접근할 수 있는 동일한 도메인 또는 네트워크 내의 침투 가능한 시스템에 침투 후 추가적인 공격을 시도하는 경우가 대부분이다. 이런 경우 네트워크를 통해 특정 파일을 내려받거나, C&C 서버와 통신하거나, 주변의 시스템에 대한 정보를 획득하거나, 최종 공격 목표 시스템에 침투하기 위한 공격 트래픽을 발생시키게 된다. 이러한 공격이나 통신을 시도할 때, 최근에는 공격 트래픽을 암호화하기 때문에 악성파일이 생성하는 암호화된 네트워크 트래픽을 탐지하기 위한 연구가 진행되고 있다.

가. 전통적인 기계학습기반 탐지 방식

Garg 등[20]은 시간 속성과 패킷 개수 등의 특징을 추출하여 악성파일 패밀리를 분류하는 방법을 제안하였다. 5개 분류기의 성능을 비교하였는데, k-NN과 RF가 가장 효율적이었으며, 알려지지 않은 악성파일을 탐지하기 위해 Leave-one-out 방식이 사용되었다.

Gu 등[21]은 패킷의 순서를 분석하여 소프트웨어에 의해 결정되는 특정 IoT 기기의 상태가 일정한지를 확인함으로써 악성파일의 존재를 탐지하는 방법을 제안하였다. 일반적으로 IoT 장치는 그 행위가 매우 일정한데 IoT 장치를 제어하는 악성파일은 모순되게 행동하기 때문에 이를 활용하는 방식이었다. NLP 모델을 이용하였는데 실시간으로 활용하기에는 성능이 떨어지는 방식이었다.

나. 딥러닝기반 탐지 방식

악성파일 탐지는 다양한 트래픽 특성으로 인해 기존 ML 방식으로는 해결이 어려운 문제로 인식되고 있다. 따라서 자동 특징 추출과 학습이 가능한 DL을 활용하기 시작하였다.

Feng 등[22]은 Two-layer DL 모델을 제안하였는데, 트래픽 분석은 2차원 이미지를 입력으로 사용하

는 CNN과 AutoEncoder에 기반하여 수행하였다. 이로써 악성트래픽 탐지 문제가 이미지 분류 문제로 귀결되었고, 탐지 성능은 2-label 이진 분류(악성파일이다/아니다)와 4-label 카테고리 분류(악성파일 패밀리 분류)에서 높은 성능을 보였다.

일반적으로 네트워크 트래픽은 패킷 순서의 조합이기 때문에 특정 패킷은 다른 인접한 패킷과 매우 연관이 깊을 수밖에 없다. 그러므로 이런 경우 순차적 입력과 Long-term dependency 처리가 가능한 LSTM이 트래픽 분석에 효과적일 수 있다. 이와 같은 아이디어에 맞게 Prasse 등[23]이 LSTM을 이용하여 악성파일을 탐지하는 모델을 제안하였고, RF와 비교하였을 때 더 좋은 성능을 보였다.

Banihashemi 등[24]은 UNB ISCX VPN-non-VPN 데이터셋에 대하여 모바일 트래픽이 아닌 암호화된 트래픽에 대한 분류 작업을 DNN 기반으로 수행하였다. 데이터셋은 암호화된 6가지 카테고리와 암호화되지 않은 6개의 카테고리로 구분되어 있었다. 서로 다른 응용프로그램을 분류하였는데, 분류 성능은 0.86의 정확도와 0.78의 F1-score를 보였다. 트래픽 분류를 위해 사용된 특징값들은 네트워크 패킷의 헤더 부분에서 직접적으로 얻을 수 있는 값들을 활용하였다.

Ⅳ. 암호트래픽 분석을 위한 고려사항 및 향후 연구방향

이 장에서는 암호트래픽 분석을 위해 모델을 구성할 때 고려해야 하는 사항과 향후 추가적인 연구가 필요한 분야에 대해 기술한다.

1. 트래픽 데이터셋 생성 및 구성

높은 성능의 트래픽 분석 모델 개발을 위해서는 고품질 트래픽 데이터셋을 확보하는 것이 매우 중요하다. 왜냐하면 분류 모델의 성능은 모델 개발에 사용된 데이터셋의 품질에 따라 크게 영향을 받기 때문이다. 정확하고 강건한(Robust) 분류 모델 생성을 위해서는 현실 세계가 충실히 반영된 매우 다양한 형태의 정확한 레이블링이 이루어진 고품질 데이터셋이 확보되어야만 한다.

이러한 데이터셋을 확보하는 방법으로는 다양성과 현실성이 보장된 실제 트래픽을 수집하여, 각 트래픽에 대한 정확한 레이블링을 수행하는 방법과 수집하고자 하는 행위에 대한 정확한 레이블링이 가능한 트래픽을 시뮬레이션을 통해 수집하고, 이를 확장해 나감으로써 많은 양의 트래픽을 확보하는 방식이 있다. 정확한 레이블링을 위해서는 많은 시간과 비용이 소요되기 때문에 이를 극복하기 위해 Auto-labeling 도구를 활용하기도 한다.

2. 트래픽 표현 방법

암호화된 네트워크 트래픽 이상탐지를 위한 지도·비지도 학습 ML 모델의 정확성을 향상시키기 위해서는 해당 트래픽을 표현하는 의미 있는 특징(Feature)을 선정하는 것이 매우 중요하다. 또한, 이러한 트래픽 표현 방법을 선택할 때는 다음과 같은 몇 가지 이슈에 대해 고려해야 한다.

- 정확성: 개발된 트래픽 분류 모델은 높은 정확도를 제공해야 한다. 이를 위해서는 패킷 순서, 타이밍, 방향 및 빈도 등의 표면적인 특징 외에도 서버와 클라이언트 간의 상호작용을 분석하여 보다 효율적인 트래픽 표현 방법을 개발하는 것이 필요하다. 특히, 전송되는 데이터에 대한 의미론적인 지식(Semantic Knowledge)을 접목하는 방법이 필요할 것으로 보인다.
- 적응성: 지정된 트래픽 표현 방법을 활용해 개

발된 분석 모델은 분류 대상 트래픽이 학습에 사용된 트래픽과 일정 부분 차이가 존재하더라도 최초 구성된 모델을 크게 변경하지 않고 분석할 수 있어야 한다. 트래픽의 변화는 다양한 원인(프로토콜 업그레이드, 내용 업데이트, 네트워크 상태 변화 등)으로 발생할 수 있어서 미묘한 트래픽 변화에 분류 정확성이 크게 떨어지면 해당 모델은 실제 환경에서는 사용 불가능하기 때문이다.

- 낮은 시간 복잡도: 암호화된 네트워크 트래픽 분류 모델은 실시간으로 분류 대상 네트워크 트래픽으로부터 사전에 정의된 트래픽 특징을 추출하고 이를 기반으로 네트워크 트래픽을 분류할 수 있어야 한다. 트래픽 특징을 추출할 때 직관적으로 추출 가능한 특징 외에 엔트로피 등과 같이 계산량이 많은 통계적 특징이나 이미지 또는 그래프 형태로 추상화하여 획득하는 특징 추출은 상당한 시간이 소요되기 때문에 가능한 한 복잡하지 않은 트래픽 표현 방법을 확보해야 한다.
- 해석 가능성: 분류의 효율성으로 인해 DL이 널리 적용되기 시작했으나, 분류 모델의 특징을 자동으로 선택하는 DL 특성상 분류 결과의 근거를 확인하기 어려운 경우가 많다. 그러나, 정보보안 관련 이상탐지 분야와 같은 DL의 판단 근거가 제공되어야 해당 결과에 기반한 조치가 가능한 서비스들은 DL 모델의 판단 결과 근거를 요구하고 있다. 따라서, 트래픽 표현 방법 선정 시에도 제공 서비스의 특성에 따라 분석 결과의 근거 확인이 필요한지를 고려해야 한다.

3. 분석 모델 생성

분석 모델은 크게 전통적인 ML 모델(예: k-NN, SVM, RF, DT 등)과 DL 모델(예: CNN, MLP, LSTM)로 구분되는데, 최근에는 DL을 이용하여 효과적인 특징들을 추출하고, 이를 기반으로 ML 모델을 수행하여 분류하는 혼합 모델을 활용하기도 한다. 이와 같은 분석 모델과 관련한 고려사항은 다음과 같다.

- 정확성: 분류 모델 생성에 있어 가장 중요한 지표는 정확도인데, 이는 Ⅱ장 4절에서 소개한 다양한 메트릭으로 측정할 수 있다. 대부분의 분류 모델은 학습에 활용된 데이터셋과 다른 실제 트래픽에 적용하거나, 트래픽 변조도구로 변조된 트래픽이 변조되는 경우 정확성이 크게 떨어지는 경우가 있는데, 이를 극복하기 위한 연구가 필요하다.
- 일반화 가능성: 탐지 모델은 다른 환경, 다른 네트워크에서도 그 효율성을 유지해야 하지만, 많은 연구가 특정 분류 모델의 사전 학습 과정에서 사용한 시나리오와 비슷하거나 동일한 시나리오서만 적용 가능하다고 가정하고 있다. 하지만, 현실에서는 트래픽의 형태가 매우 다양하기 때문에 이를 극복할 수 있는 연구가 이루어져야 한다.
- 전이 가능성: 전이 가능성은 한 도메인에서 사용하는 트래픽 표현 방법이 다른 도메인에 적용될 수 있는 정도를 의미하는 것으로, 특정 도메인에서 학습한 모델을 기반으로 새로운 도메인에 맞는 모델을 생성할 때 쉬운 정도를 의미한다.
  이는 일반적으로 전이학습을 통해 이루어지는데, 전이학습은 기존에 개발된 모델을 바탕으로 사용하고자 하는 학습데이터를 학습시켜 새로운 모델을 개발하는 방식으로, 도메인 간 데이터 불균형이 심한 상황 등에 유용하게 활용할 수 있다. 따라서, 다양한 도메인에 활용 가능한 모델을 개발하고자 하는 경우, 전이 가능성을 고려함으로써 모델의 일반화 능력과 유연성을 향상시키도록 해야 할 것이다.

4. AI 모델 공격 대응

트래픽 분석 모델은 이를 무력화하려는 공격에 대응 가능성을 높여야 한다. 예를 들어, 공격자가 네트워크 트래픽의 형태를 변조하여 트래픽 분석 모델에서 사용하는 특징값을 바꾸게 된다면 해당 모델에서는 해당 공격 탐지가 어려워질 수 있다.

다만, 공격자가 네트워크 트래픽 형태를 변조하는 방법은 패킷을 삭제할 수는 없고 더미 패킷을 추가하거나 패킷 전달을 지연시키는 것 정도로 제한되기 때문에 탐지 모델을 우회하는 것은 쉬운 일이 아니다. 또한, 공격자는 이상탐지 모델이 어떤 트래픽 특성을 분류에 활용하는지 모르기 때문에 적절한 공격 패킷을 생성해 내기가 쉽지 않다. 그럼에도 불구하고 모델 생성 간에는 트래픽 변조를 통해 모델을 우회할 수 있다는 점을 인지하고 이를 감안하여 모델을 생성해야 한다.

V. 결론

인터넷이 급격히 발달하면서 다양한 기기가 인터넷에 연결되고 매우 많은 서비스가 제공되기 시작하였다. 이러한 서비스 중에는 중요한 정보를 전달하는 경우가 있는데, 이때 해당 정보를 보호하기 위해 암호화된 네트워크 트래픽을 사용하기 시작하였고, 현재는 상당수의 인터넷 트래픽이 암호화되어 송수신되고 있다. 동시에, 네트워크를 통한 공격 역시 암호화된 트래픽을 활용하기 시작하면서 기존의 시그니처기반 공격 탐지 방법은 한계에 다다랐고, 이를 극복하기 위해 암호트래픽 분석 분야에서는 활발한 연구가 이루어지고 있다.

본고는 암호트래픽 분석 분야의 전반에 걸친 사항을 정리하여 암호트래픽 분석 절차, 분석 분야, 분석 모델, 모델 평가 방법 그리고 관련 연구 동향에 대해 알아보았고, 특히 암호트래픽 분석 분야 중 네트워크 트래픽 이상탐지 분야에 대한 연구 동향을 소개하였다.

추가적으로, 암호트래픽 분석을 위한 모델 개발 시 고려해야 하는 사항으로 트래픽 데이터셋 구성, 트래픽 표현 방법 선정, 분석 모델 생성 및 AI 모델 공격 대응에 대해 논하였다.

앞으로 암호화된 네트워크 트래픽의 양은 더욱더 늘어날 것이며, 특히 공격 트래픽은 더 많은 비율로 암호트래픽을 이용할 것이다. 이런 환경에서 발생하는 공격 탐지를 위한 연구는 지속적으로 이뤄져야 한다.

용어해설

ToR(The Onion Router) 미국 해군연구소에서 개발한 익명성을 제공하는 네트워크 프로토콜로서, 해당 프로토콜을 활용하기 위한 소프트웨어로 ToR 브라우저 등을 제공하고 있으며, 딥웹과 다크웹에 주로 활용되고 있음

TLS(Transport Layer Security) 네트워크 7-Layer 중 전송 계층에서 암호화를 수행하는 TLS는 인터넷상의 통신 간 개인정보와 데이터를 손쉽게 암호화하기 위해 설계된 표준 보안프로토콜임. TLS의 주요 사용처 중 하나는 웹 사이트에 접속하는 웹 브라우저와 같이 웹 응용프로그램과 서버 간의 통신을 암호화하는 것임

엔트로피(Entropy) 통계역학적으로는 주어진 거시적(Macroscopic) 상태에 대응하여 나타나는 모든 미시적(Microscopic) 상태 수의 로그값(Log)으로 생각할 수 있음. 일반적으로는 주로 무질서도(Randomness 또는 Disorder)라고 알려져 있는데 엔트로피를 완벽히 설명하는 정확한 개념은 아니지만 흔히 사용되고 있음

특징공학(Feature Engineering) 기계학습 알고리즘을 작동하기 위해 도메인 지식을 활용하여 원시 데이터를 가공하는 과정으로, 특징공학의 과정에서 새로운 특징을 만들어 내거나 불필요한 특징을 제거할 수 있음. 특징공학은 기계학습 모델의 성능에 영향을 많이 미치기 때문에 기계학습에 있어서 매우 중요한 단계임

Flash Crowds 인터넷에서 발생하는 갑작스럽고 집중적인 트래픽 증가 현상(공격 아님)

Leave-one-out 수집된 데이터셋의 일부분을 제외하고 나머지로만 탐지 모델을 학습시키고, 해당 모델을 가지고 제외했던 데이터셋에 대해 탐지 가능 여부를 확인하는 방법

Long-term Dependency 은닉층의 과거 정보가 마지막까지 전달되지 못하는 현상

약어 정리

CART	Classification And Regression Trees
CNN	Convolutional Neural Network
IoT	Internet of Things
LSTM	Long Short-Term Memory
RF	Random Forest
SAE	Stacked AutoEncoder
TLS	Transport Layer Security
ToR	The Onion Router
VPN	Virtual Private Network
XAI	Explainable Artificial Intelligence

참고문헌

[1] Google, HTTPS encryption on the web, Google Transparency Report, 2023. 4., https://transparencyreport.google.com/https/overview?hl=kr

[2] M. Shen et al., "Machine learning-powered encrypted network traffic analysis: A comprehensive survey," IEEE Commun. Surv. Tutor., vol. 25, no. 1, 2023.

[3] D. Desai, "2020: The State of Encrypted Attacks. Zscaler," Retrieved Feb. 24, 2021, https://www.zscaler.com/blogs/security-research/2020-state-encrypted-attacks

[4] https://www.tcpdump.org/

[5] IETF RFC 8446, The Transport Layer Security(TLS) Protocol Version 1.3, Aug. 2018, https://www.ietf.org/rfc/rfc8446.txt

[6] P. Sirinam et al., "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., (Toronto, Canada), Oct. 2018, pp. 1928-1943.

[7] M. Shen et al., "Accurate decentralized application identification via encrypted traffic analysis using graph neural networks," IEEE Trans. Inf. Forensics Secur., vol. 16, 2021, pp. 2367-2380.

[8] T. Shapira and Y. Shavitt, "FlowPic: A generic representation for encrypted traffic classification and applications identification," IEEE Trans. Netw. Service Manag., vol. 18, no. 2, 2021, pp. 1218-1232.

[9] 김홍비, 이태진, "정보보호 분야의 XAI 기술 동향," 정보보호학회지, 제31권 제5호, 2021.

[10] J. Lever, "Classification evaluation," Nature Methods, vol. 13, no. 8, 2016, pp. 603-604.

[11] G. Stergiopoulos et al., "Automatic detection of various malicious traffic using side channel features on TCP packets," Computer Security, Springer, Cham, Switzerland, 2018, pp. 346-362.

[12] CTU-13 dataset, CTU University, Czech Republic, 2011, https://mcfp.felk.cvut.cz/publicDatasets/CTU-Malware-Capture-Botnet-1/

[13] First.org, Hands-on Network Forensics-Training PCAP dataset from FIRST 2015, www.first.org/assets/conf2015/networkforensicsvirtualbox.zip

[14] Milicenso, Ponmocup Malware dataset, Update 2012-10-07, http://security-research.dyndns.org/pub/botnet/ponmocup/analysis2012-10-05/analysis.txt (Accessed 1 Jan. 2018)

[15] X. Qin, T. Xu, and C. Wang, "DDoS attack detection using flow entropy and clustering technique," in Proc. Int. Conf. Comput. Intell. Secur. (CIS), (Shenzhen, China), 2015, pp. 412-415.

[16] M. Zolotukhin et al., "Data mining approach for detection of DDoS attacks utilizing SSL/TLS protocol," Internet of Things, Smart Spaces, and Next Generation Networks and Systems, Springer, Cham, Switzerland, 2015, pp. 274-285.

[17] Y. Zeng et al., "Deep-full-range: A deep learning based network encrypted traffic classification and intrusion detection framework," IEEE Access, vol. 7, 2019, pp. 45182-45190.

[18] M. Zolotukhin et al., "Increasing web service availability by detecting application-layer DDoS attacks in encrypted traffic," in Proc. IEEE 23rd Int. Conf. Telecommun. (ICT), (Thessaloniki, Greece), May 2016, pp. 1-6.

[19] J. David et al., "DDoS attack detection using fast entropy approach on flow-based network traffic," Procedia Comput. Sci., vol. 50, 2015, pp. 30-36.

[20] S. Garg, S.K. Peddoju, and A.K. Sarje, "Network-based detection of Android malicious apps," Int. J. Inf. Secur. vol. 16, no. 4, 2017, pp. 385-400.

[21] T. Gu et al., "IoTGaze: IoT security enforcement via wireless context analysis," in Proc. IEEE INFOCOM Conf. Comput. Commun., (Toronto, Canada), Jul. 2020, pp. 884-893.

[22] J. Feng et al., "A two-layer deep learning method for Android malware detection using network traffic," IEEE Access, vol. 8, 2020, pp. 125786-125796.

[23] P. Prasse et al., "Malware detection by analysing network traffic with neural networks," in Proc. IEEE SPW, (San Jose, CA, USA), May 2017, pp. 205-210.

[24] S.B. Banihashem and E. Aktharkavan, "Encrypted network traffic classification using deep learning method," in Proc. Int. Conf. Web Res. (ICWR), (Tehran, Iran), May 2022.