머신러닝 기반 네트워크 침입탐지 및 상황 분석 솔루션 프레임워크 구현

강도희*, 김유진*, 연예림*, 곽병일**

*,** 한림대학교 소프트웨어학부 (학부생, 교수)

Implementation of Framework for Machine Learning-Based Network Intrusion Detection and Situational Analysis Solutions

DO-Hee Kang* Yu-Jin Kim* Ye-Rim Yeon* and Byung-Il Kwak**

*, ** Hallym University Software (Undergraduate student, Professor)

요 약

본 논문에서는 네트워크 보안 상황 파악이 가능한 네트워크 트래픽 분석 솔루션을 구현하였다. 웹 프레임워크를 통해 제공하는 대시보드는 데이터 시각화, 데이터 분석, 보안 솔루션 및 백신 추천으로 이루어져 있다. 본 프레임워크는 입력한 네트워크 트래픽을 머신러닝 기반 분류 알고리즘을 통해 네트워크 침입을 12가지 공격 유형으로 식별 및 분류할 수 있으며, Circle Network 및 Parallel Coordinates Attack Visualization(PCAV) 등 11가지 시각화 그래프를 통해 공격 패턴에 대한 가시성을 높였다. 또한, 공격 분석표를 통해 구체적인 정보를 제공하고, 분석 데이터를 기반으로 보안 솔루션 및 백신을 추천하여 컴퓨터 네트워크를 보호하도록 구성하였다.

I. 서론

네트워크 규모의 지속적인 확장은 정보의 양과 복잡성을 증가시켜 네트워크 보안 상황 인지를 어렵게 만든다.[1] 이에 따라, 네트워크 보안 상황을 분석하기 위한 방법으로 보안 이벤트 시각화 기술이 연구되고 있다.[2] 보안 이벤트 시각화란 네트워크상에서 발생하는 방대한 양의 이벤트를 시각화하는 기술로 보안과 관련된 많은 정보를 신속하고 정확하게 전달한다.[3]

본 논문에서는 네트워크 보안 상황을 파악할 수 있는 웹 기반 네트워크 트래픽 분석 솔루션을 구현하였다. 본 프레임워크는 입력한 네트워크 트래픽을 Random Forest 알고리즘을 통해 네트워크 침입을 12가지 공격 유형으로 식별 및 분류할 수 있으며, 11가지 시각화 그래프를 통해 공격 패턴에 대한 가시성을 높였다. 또한, 공격 분석표를 통해 구체적인 정보를 제공하고, 분석 데이터를 기반으로 보안 솔루션을 추천하여 컴퓨터 네트워크를 보호하도록 구성하였다.

II. 침입 탐지 시스템

2.1 학습 데이터셋

본 논문에서는 여러 공격 데이터가 포함된 CICIDS2017[4] 데이터셋 및 직접 수집한 네트워크 데이터를 사용하였다. 학습에 사용된 데이터는 Normal, DoS Hulk, DoS slowhttptest, DoS slowloeis, DoS GoldenEye, Heartbleed, ICMP Flooding, Infiltration, PortScan, Botnet, FTP-Patator, SSH-Patator, DDoS로 총 13개의 유형으로 분류된다.

2.2 피처 엔지니어링

침입 탐지 시스템을 서버에서 운영하기 위해 경량화 하였다. 이를 위해 모델 학습에 사용되는 피처의 수를 최소화하고, 데이터 시각화 및 모델 학습에 사용되는 피처를 동일하게 구성했다. 또한, 데이터 시각화에서 네트워크 트래픽을 쉽게 파악할 수 있도록, 송수신 IP 주소와

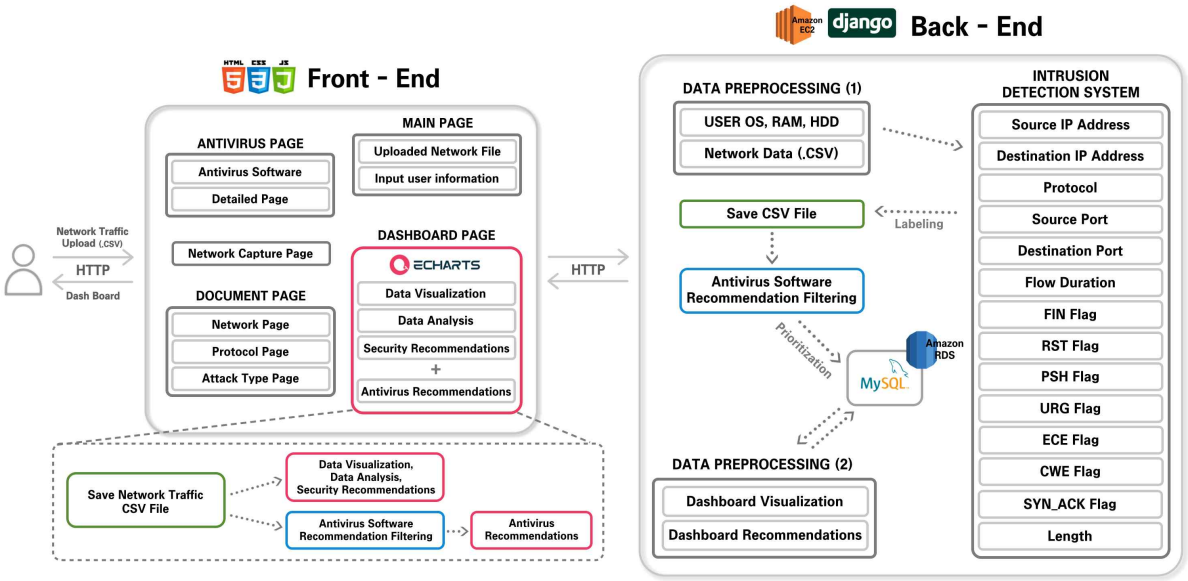


Fig 1. Overview

같은 기본적인 정보로만 피처 엔지니어링을 수행하였다. CICIDS2017 데이터셋에서 13개의 피처를 선택하였으며, 수신 상태를 확인하기 위해 SYN Flag와 ACK Flag를 동시에 카운트하는 SYN_ACK Flag를 재구성하여 추가했다. 학습에 사용된 피처는 Fig 1의 침입 탐지 시스템(Intrusion Detection System)에 나타내었다.

2.3 모델 성능

본 논문에서는 데이터셋을 8:2 비율로 랜덤하게 분할하여 Random Forest 알고리즘으로 학습 및 테스트를 수행하였다. 성능 지표로는 혼동행렬을 기반으로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score를 계산하여 평가하였다. 공격을 공격으로 식별했을 때, True Positive(TP)로 설정하였으며, 정상을 정상으로 식별했을 때, True Negative(TN)로 설정했다. 실험 결과, Accuracy 99.99%, Precision 99.82%, Recall 98.27%, F1-Score 98.95% 성능을 보였다.

III. 데이터 시각화

데이터 시각화의 상위 영역(Line Chart, Circle Network, 정상/이상 비율, 패킷 길이 통계 및 빈도, Protocol & Port)은 네트워크 트래픽의 전체적인 통계와 흐름을 파악할 수 있으며, 하위 영역(Flag Count, Flow Duration, Parallel Coordinates Attack Visualization, Conversation)은 데이터를 세분화하여 공격 유형을 식별할 수 있다. 이를 통해 네트워크 보안 상황을 종합적으로 분석할 수 있다.

3.1 Line Chart

네트워크 트래픽 흐름을 파악하기 위해 시간에 따른 BPS(Bytes Per Second), PPS(Packets Per Second), QPS(Queries Per Second), RPS(Requests Per Second)를 Fig 2와 같이 Line Chart로 나타내었다. 정상 및 공격 데이터는 투명도 차이를 활용해 구분하였다. BPS는 초당 바이트 수, PPS는 초당 패킷 수를 나타내며, DoS 및 포트 스캔과 같은 공격을 확인할 수 있다. QPS 및 RPS는 초당 DNS 쿼리, HTTP(S) 요청 수를 의미하며, 응용 계층 공격 및 서비스 부하 상태를 파악할 수 있다. 이를 통해 사용자는 보안 이벤트가 언제 발생했는지 직관적으로 검출할 수 있다.

3.2 Circle Network

전체적인 IP 주소 간의 관계를 Fig 3-(a)과 같이 Circle Network 형태로 나타내었다. IP 주소는 노드로, IP 주소 간의 교류량은 노드와 노드 사이를 연결하는 선으로 표현하였다. 또한, 교류량이 일정 임계치를 초과하면 상위 노드는 노란색으로, 연결선은 파랑, 노랑, 흰색 순으로

시각화하였다. Fig 3-(b)과 같이 특정 노드를 선택할 경우 해당 IP 주소와 트래픽 교류량을 구체적으로 확인할 수 있다.

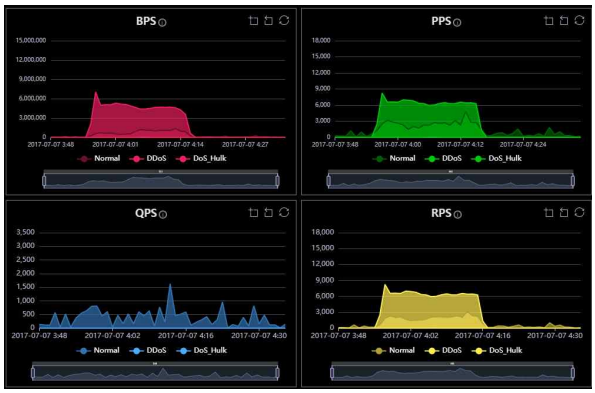
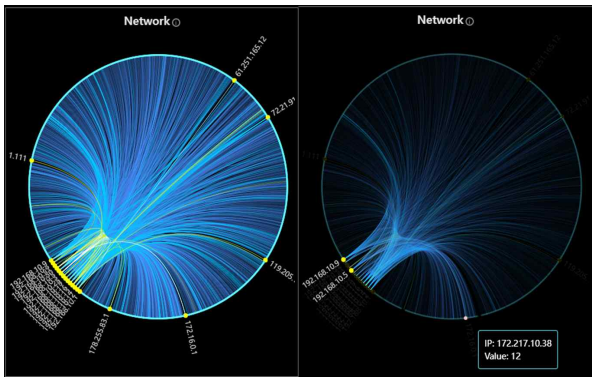


Fig 2. Line Chart



(a) 전체 네트워크 (b) 특정 노드 선택

Fig 3. Circle Network Graph

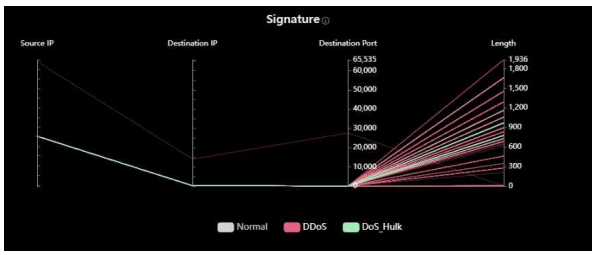


Fig 4. Parallel Coordinates Attack Graph

3.3 Parallel Coordinates Attack Visualization

Parallel Coordinates Attack Visualization (PCAV)[5]을 활용하여 복잡한 네트워크 데이터를 간단하게 시각화하였다. PCAV는 Source IP, Destination IP, Destination Port, Packet Length를 x축에 두며, 공격 데이터의 경우 특정 패턴을 형성한다. Fig 4는 DDoS 및 DoS Hulk의 공격 데이터가 특정 패턴으로 시각화된다는 것을 보여준다. 이 외에도 대시보드에 시각화한 그래프는 Table 1에 설명하였다.

그래프	설명
정상/이상 비율	정상과 공격 데이터의 전체 비율을 도넛 그래프로 표현
패킷 길이 통계	트래픽 길이의 평균, 표준편차, 최댓값, 최솟값을 표로 표현
패킷 길이 빈도	트래픽 길이의 빈도를 히스토그램으로 표현
Flow Duration	초당 평균 Flow Duration을 선 그래프로 표현
Protocol & Port	프로토콜별 포트 빈도 수를 도넛 그래프로 표현
Flag Count	초당 Flag Count를 누적 막대 그래프로 표현
Conversation	트래픽 교류량이 가장 높은 IP들의 특징을 표로 표현

Table 1. 대시보드 그래프 설명

IV. 구현

본 프레임워크는 입력한 네트워크 트래픽을 Pandas로 전처리하고, 오픈소스 EChart를 통해 대시보드를 구성한다. Fig 1은 구현한 웹 프레임워크의 오버뷰를 나타낸다.

4.1 Dashboard

대시보드는 데이터 시각화, 데이터 분석, 보안 솔루션 및 백신 추천으로 이루어져 있다. 해당 내용은 보고서 형태로 PDF 다운로드가 가능하며, Fig 5를 통해 확인할 수 있다.

4.1.1 데이터 분석

Fig 5-(a)와 같이 전체 데이터를 통계 분석한 후, 공격 유형이 식별되면 Fig 5-(b)와 같이 해당 공격에 대한 분석표를 제공한다. 공격 분석표의 구성은 Table 2와 같다.

4.1.2 보안 솔루션 및 백신 추천

공격이 식별되면 해당 공격에 대한 보안 솔루션을 Fig 5-(c)와 같이 제시한다. 공격 데이터가 없는 경우, 백신 업데이트 등 가장 기본적인 보안 솔루션을 제공한다. 백신은 분석 데이터를 기반으로 연관성을 계산하여 Fig 5-(d)와 같이 추천한다. 분석 데이터는 네트워크 트래픽 정보, 사용 유형(개인용, 가정용 등), 사용 환경(OS, RAM, HDD), 가격 등으로 구성된다.

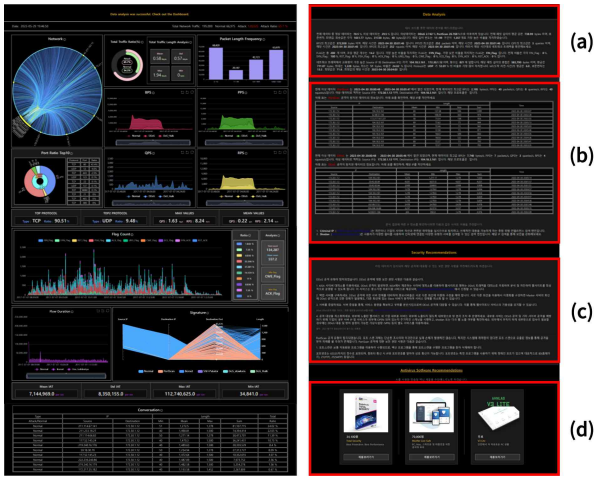


Fig 5. Dashboard

종류		설명
IP 주소	Source	송신 IP 주소
	Destination	수신 IP 주소
Length	Mean	트래픽 길이의 평균값
	Min	트래픽 길이의 최솟값
	Max	트래픽 길이의 최댓값
	Sum	트래픽 길이의 총합
Time		트래픽이 들어온 시간

Table 2. 공격 분석표

V. 결론

본 논문은 네트워크 보안 상황을 인지할 수 있는 웹 기반 네트워크 트래픽 분석 솔루션을 구현하였다. 본 프레임워크는 머신러닝 분류 알고리즘인 Random Forest를 활용하여 네트워크 침입을 12가지 공격 유형으로 식별 및 분류하고, 다양한 보안 이벤트 시각화 기술을 사용하여 네트워크상의 여러 공격을 직관적으로 파악할 수 있다. 또한, 공격 분석표 및 보안 솔루션을 제공하여 네트워크 보안 상황을 인지하고 대응할 수 있도록 구성하였다.

[참고문헌]

[1] 박재범, 김휘강, 김은진.(2014).대규모 네트워크의 효과적 보안상황 인지를 위한 벌집 구조 시각화 시스템의 설계 및 구현.정보보호학회논문지,24(6),1197-1213.

[2] H. Shiravi, A. Shiravi and A. A. Ghorbani, "A Survey of Visualization Systems for Network Security," in IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 8, pp. 1313-1329, Aug. 2012, doi: 10.1109/TVCG.2011.144.

[3] 정치윤, 손선경, 장범환 and 나중찬. (2009). 시각화 기반의 효율적인 네트워크 보안 상황 분석 방법. 정보보호학회논문지, 19(3), 107-117.

[4] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

[5] Hyunsang Choi, Heejo Lee, Hyogon Kim, Fast detection and visualization of network attacks on parallel coordinates, Computers & Security, Volume 28, Issue 5, 2009, Pages 276-288, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2008.12.003

[6] K. hun Han and H. K. Kim, "FDANT-PCSV: Fast Detection of Abnormal Network Traffic Using Parallel Coordinates and Sankey Visualization," Journal of the Korea Institute of Information Security & Cryptology, vol. 30, no. 4, pp. 693 - 704, Aug. 2020.