MaxDAPH 기반 비지도 표현 학습을 통한 네트워크 이상 탐지 방법

조영준, 박혜원, 박요한*
계명대학교 컴퓨터공학
{5907022, wldnjsfuf, yhpark}@kmu.ac.kr

A Study on Network Intrusion Detection Systems Using Unsupervised Representation Learning Based on MaxDAPH

Youngjun Cho, Hyewon Park, Yohan Park*

Department of Computer Engineering, Keimyung University

요 약

본 논문에서는 MaxDAPH(Maximum Distance from Average Point in Hyperspace) 기반 비지도 학습을 활용한 네트워크 이상 탐지 방법을 제안하였다. 제안된 방법은 MLP 인코더를 통해 데이터를 잠재공간으로 매핑하고, 중심점(Centroid)과의 거리 최소화를 목표로 하는 MaxDAPH 손실 함수를 적용하여 정상과 공격 데이터를 구분한다. 또한 Early Stopping 기법을 도입하여 학습의 안정성과 효율성을 확보하였다. UNSW-NB15와 CICDarknet2020 데이터셋 실험 결과, Accuracy 0.86, Precision 0.92, Recall 0.89, F1 Score 0.91을 달성하여 기존 비지도 학습 기반 방법보다 우수한 성능을 보였다. 따라서 제안된 방법은 라벨 정보가 없는 환경에서도 효과적인 이상 탐지를 수행할 수 있으며, 암호화 트래픽 및 Zero-day 공격 탐지에도 높은 확장 가능성을 지닌다.

Ⅰ. 서 론

컴퓨터 네트워크의 확장은 시스템의 가용성과 기밀성을 위협하는 사이버 공격의 증가로 이어지고 있다[1]. 기존 IDS(Intrusion Detection System)는 알려진 공격에서는 효과적이지만, 제로데이터 공격이나 변종 탐지에는 한계가 있으며 암호화 트래픽 환경에서는 규칙 기반 탐지가 어렵다[2][3]. 이러한 한계를 극복하기 위해 비지도 학습 기반 이상 탐지 연구가 활발히 진행되고 있으며, ET-SSL 모델[4]은 높은 성능을 보였으나 완전한 비지도 학습으로는 보기 어렵다.

본 연구는 MaxDAPH(Maximum Distance from Average Point in Hyperspace) 기반의 완전 비지도 이상 탐지 방법을 제안한다. 제안한 모델은 MLP 인코더를 통해 입력 데이터를 잠재 공간으로 매핑하고, 중심점(centroid)으로부터의 거리를 최소화하는 MaxDAPH 손실 함수를 사용하여 정상 데이터는 중심 주변에 이상 데이터는 외곽으로 분리되도록 학습한다. 또한 Early Stopping 기법을 적용해 학습 안정성을 확보하였다. UNSW-NB15와 CIC-Darknet2020 데이터셋 실험 결과, 제안된 MaxDAPH는 라벨 정보가 없는 환경에서도 F1 Score 0.91을 달성하여, 라벨 의존적인 ET-SSL보다 구조적으로 단순하면서도 안정적인 탐지 성능을 보여주었다. 따라서 MaxDAPH는 완전 비지도 기반 네트워크 이상 탐지의 효율적 대안으로서, 암호화 트래픽 및 제로데이 고역 탐지 분야에 기여할 것으로 기대된다.

Ⅱ. MaxDAPH 기반의 완전 비지도 이상 탐지 방법

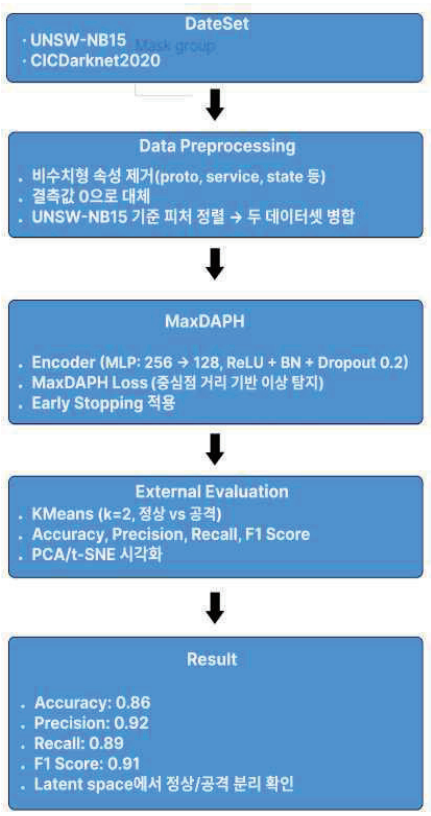

그림 1 MaxDAPH 기반 탐지 과정

그림1은 제안한 탐지 기법의 전체 흐름도를 나타낸다. 성능 평가는 공용 데이터셋인 UNSW-NB15와 CICDarknet2020을 이용하여 수행하였다.

2-1. UNSW-NB15:

Moustafa 와 Slay(2015) 가MiCIS 컨퍼런스에서 제안한 데이터셋으로 정상 트래픽과 함께 Dos, Exploits, Fuzzers, Generic, Reconnaissance 등 다양한 공격유형을 포함한다.

2-2. CICDarknet2020:

Canadian Institute for Cybersecurity(CIC)에서 구축한 데이터셋으로, VPN 및 Tor 환경의 암호화된 네트워크 트래픽을 포함한다. 정상(Benign)과 Botnet, DDoS, Infiltration 등 다양한 공격 데이터를 제공한다. 비수치형 속성을 제거하고 결측값을 0으로 대체한 후, UNSW-NB15의 피처 구성을 기준으로 재정렬하여 병합한 결과 총 223,862개의 샘플과 44개의 특징으로 구성되었다.

2-3. 제안기법:MaxDAPH기반비지도학습

본 연구에서 MaxDAP 기반 비지도 학습 모델을 제안한다.

인코더(Encoder): 입력 데이터를 잠재공간(latent space)으로 매핑하기 위해 다층 퍼셉트론(MLP) 구조를 사용하였다. 은닉층은 256 및 128 뉴런으로 구성하였으며, ReLU활성함수, Batch Normalization, Droput(0.2)을 적용하였다.

MaxDAPH 손실 함수(MaxDAPH Loss)

잠재공간에서 샘플 임베딩 벡터 Zi와 중심점(centroid) C 간의 거리를 최소화하는 손실 함수를 적용하였다. 이를 수식으로 표현한다면 다음과 같다.

$$L_{MAXDAPH} = \frac{1}{N}\sum_{i=1}^{N}\|z_i - c\|_2 \qquad (1)$$

여기서 N은 샘플의 개수를 의미하며, 각 데이터 임베딩 Zi가 중심점 C로부터 얼마나 떨어져 있는지를 L2 거리로 계산한다.

정상 데이터는 중심점 주변에 밀집하게 되고, 중심점으로부터 멀리 떨어진 데이터는 잠재적으로 이상 공격으로 탐지된다.

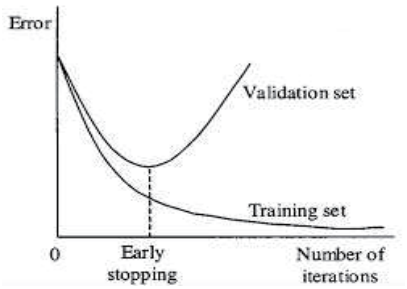

그림2 조기중단(Early Stopping)

Early Stopping : 그림 2와 같이 손실 값이 개선되지 않는 경우 학습을 조기에 종료하여 과적합을 방지하였다.

2-4. 클러스터링 및 평가 방법(Clustering and Evaluation)

잠재공간으로 매핑된 임베딩 벡터는 K-Means 알고리즘을 이용해 두 개의 클러스터로 분류하였다. 클러스터 수는 정상 및 공격 클래스에 대응하도록 2으로 설정하였으며, 성능 평가는 Accuracy, Precision, Recall, F1 Score로 수행하였다. 또한 PCA와 t-SNE를 통해 잠재공간을 시각화하여 정상과 공격 데이터의 분리 양상을 확인하였다.

2-5. 실험 환경(Experimental Setup)

실험은 Python 3.10, TensorFlow 2.15, scikit-learn 환경에서 수행하였다.

데이터는 MinMaxScaler를 사용해 0~1 구간으로 정규화하였다. 하이퍼파라미터는 Latent Dimension 64, Learning Rate 1e-6, Batch Size 128, Epoch 200으로 설정하였다. 모든 학습은 동일한 조건에서 수행되었다.

Ⅲ. 실험 결과 및 분석

학습 과정에서 MaxDAPH 손실 값은 epoch이 증가함에 따라 안정적으로 감소하며 수렴하였다. 이는 모델이 중심점을 기준으로 정상 데이터를 효과적으로 학습했음을 의미한다. t-SNE 시각화 결과 정상 데이터는 잠재공간 내 특정 영역에 밀집된 형태로 분포하였으며, 공격데이터는 MaxDAPH 손실 함수가 중심점(Centroid)을 기준으로 정상 데이터를 수렴시키고, 이상 데이터를 외곽으로 이동시키는 구조적 특성을 잘 반영하고 있음을 확인 할 수 있었다.

표 1. ET-SLL,MaxDAPH 비교분석

Metric	ET-SSL	MaxDAPH
Accuracy	0.90	0.86
Precision	0.91	0.92
Recall	0.90	0.89
F1 Score	0.90	0.91

K-Means 기반 클러스터링 성능 평가 결과 표 1을 보면 Accuracy 0.86, Precision 0.92, Recall 0.89, F1 Score 0.91을 달성하였다. 이는 라벨 정보가 없는 완전 비지도 환경에서도 정상 공격을 효과적으로 구분했음을 보여준다. 동일 데이터셋에서 재현한 ET-SSL 모델은 Accuracy 0.90, F1 Score 0.90으로 원 논문에서 보고된 0.95 보다 낮게 측정되었다. 반면 제안한 MaxDAPH 모델은 단순한 MLP 구조와 완전 비지도 학습에도 F1 Score 0.91으로 거의 동일하거나, 일부 지표에서는 더 우수한 성능을 보였다. 결과적으로, MaxDAPH는 구조적 단순성과 학습 효율성을 동시에 확보하면서도, 복잡한 자기지도 모델(ET-SSL)에 근접한 이상 탐지 성능을 보여준다.

참 고 문 헌

[1] Liao, H. J., Lin, C. H. R., Lin, Y. C., and Tung, K. Y.,Intrusion Detection System: A Comprehensive Review," Journal of Network and Computer Applications, Vol. 36, No. 1, pp. 16 - 24, 2013.

[2] Moustafa, N., and Slay, J.,UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," Proceedings of the Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 2015.

[3] Panigrahi, R., and Borah, S.,A Comprehensive Survey on Machine Learning for Intrusion Detection," Journal of Network and Computer Applications, Vol. 182, pp. 103 - 107, 2021.

[4] Brissaud, E., et al.,Encrypted Traffic Self-Supervised Learning for Network Anomaly Detection (ET-SSL)," Scientific Reports, 2025.