https://doi.org/10.33778/kcsa.2024.24.3.013

단일 클래스 모델을 활용한 네트워크 침입 탐지

민 병 준*, 박 대 경**

요 약

4차 산업혁명의 발전으로 네트워크가 급속히 확산되면서 사이버 보안 위협이 더욱 증가하고 있다. 기존의 시그니처 기반 네트워크 침입 탐지 시스템(NIDS)은 알려진 공격을 탐지하는 데 효과적이지만, APT와 같은 새로운 공격에는 한계가 있다. 또한, 지도 학습 기반 딥러닝 모델은 불균형 데이터 문제로 인해 정상 데이터에 편향된 결과를 낳을 위험이 있다. 이러한 문제를 해결하기 위해 본 논문은 정상 데이터만을 학습하여 비정상 데이터를 탐지하는 단일 클래스 모델 기반의 네트워크 침입 탐지 방법을 제안한다. DeepSVDD와 MemAE 모델을 활용해 NSL-KDD 데이터 셋에서 제안하는 방법의 효율성을 검증하며, 지도 학습 모델과의 비교를 통해 제안된 방법이 실제 네트워크 침입 탐지 문제에서 더욱 효과적임을 확인한다.

Network Intrusion Detection Using One-Class Models

Byeongjun Min*, Daekyeong Park**

ABSTRACT

Recently, with the rapid expansion of networks driven by the advancements of the Fourth Industrial Revolution, cybers ecurity threats are becoming increasingly severe. Traditional signature-based Network Intrusion Detection Systems (NIDS) are effective in detecting known attacks but show limitations when faced with new threats such as Advanced Persistent Threats (APT). Additionally, deep learning models based on supervised learning can lead to biased decision boundaries due to the imbalanced nature of network traffic data, where normal traffic vastly outnumbers malicious traffic. To address the se challenges, this paper proposes a network intrusion detection method based on one-class models that learn only from n ormal data to identify abnormal traffic. The effectiveness of this approach is validated through experiments using the Deep SVDD and MemAE models on the NSL-KDD dataset. Comparative analysis with supervised learning models demonstrates that the proposed method offers superior adaptability and performance in real-world scenarios.

Key words : Network Intrusion Detection, Cybersecurity, One-Class Anomaly Detection, Deep Learning

접수일(2024년 08월 19일), 게재확정일(2024년 09월 02일)

* 한화시스템 사이버전장팀 (주저자/교신저자)

** 한화시스템 사이버전장팀 (공동저자)

1. 서 론

최근 4차 산업혁명의 발전에 따라 네트워크의 급속한 확산이 이루어지면서, 사이버 보안 위협이 점차 심각해지고 있다. 특히, 사이버 공격의 수법이 점점 더 정교화되고 다양화됨에 따라 기존의 보안 체계는 새로운 위협을 효과적으로 방어하는 데 한계를 드러내고 있다. 전통적으로 네트워크 침입 탐지 시스템(Network-based Intrusion Detection System, NIDS)은 시그니처 기반 탐지 방법(Signature-based Detection)에 의존해 왔다. 이 방법은 과거의 침입 흔적을 분석하여 이미 알려진 공격 패턴을 정의하는 방식으로, 기존의 사이버 위협을 탐지하는 데는 효과적이었다. 그러나 APT(Advanced Persistent Threat)와 같은 새로운 공격 기법에 대해서는 탐지의 한계가 명확히 드러나고 있다.

딥러닝 모델을 활용한 지도 학습 기반 연구도 많이 수행되고 있으나[1-3], 실제 네트워크 환경에서는 정상 트래픽에 비해 위협 트래픽의 발생 빈도가 매우 낮아 이를 식별하고 라벨링 하는 데 막대한 비용이 소모된다. 또한, 소수의 공격 데이터를 포함한 학습 데이터 셋은 불균형 데이터(Imbalanced Data) 문제를 야기하여, 지도 학습 모델이 정상 데이터에 편향된 결정 경계(decision boundary)를 형성할 위험이 있다.

이러한 문제를 해결하기 위해, 본 논문에서는 단일 클래스(one-class) 모델 기반의 네트워크 침입 탐지 방법을 제안한다. 단일 클래스 모델은 정상 데이터만을 학습하여 비정상적인 데이터를 탐지하는 방식으로, 정상 패턴에서 벗어나는 모든 트래픽을 잠재적인 위협으로 간주한다. 이 접근 방식은 학습되지 않은 새로운 유형의 위협에도 효과적으로 대응할 수 있으며, 위협 데이터를 충분히 수집하기 어려운 실제 환경에서도 적합하다. 단일 클래스 기반 이상 탐지 모델인 DeepSVDD 모델[4]과 MemAE 모델[5]을 활용하여 제안하는 방법이 네트워크 침입 탐지 도메인에서 유효함을 NSL-KDD 데이터 셋[6]을 통해 검증한다. 실험을 통한 검증 과정에서는 신규 위협이 존재한다고 가정하여, 일부 위협 클래스에 해당하는 데이터를 제거한 채 학습된 지도 학습 모델들의 분류 성능과 비교함으로써, 제안된 네트워크 침입 탐지 방법이 다양한 실세계 위협에 대해 보다 우수한 대응력을 제공할 수 있음을 검증한다.

본 연구의 결과는 향후 네트워크 보안 체계에 있어 실질적이고 효과적인 방향성을 제시하는 데 기여할 것으로 기대된다.

2. 관련연구

2.1 단일 클래스 기반 이상탐지

단일 클래스 기반 이상 탐지(One-Class Anomaly Detection)[7]는 주로 정상 데이터만을 사용하여 학습하고, 비정상 데이터를 탐지하는 방법론으로, 비정상 데이터가 드물거나 불규칙한 패턴을 보이는 상황에서 특히 유용하다. 이 접근법의 핵심은 정상 데이터만을 학습하여 정상적인 행동 또는 패턴을 정의하고, 이와 다른 데이터를 비정상으로 탐지하는 것이기 때문에 새로운 비정상 데이터를 수집하지 않아도 효과적인 탐지가 가능하다. 주로 네트워크 보안, 금융 사기 탐지, 의료 진단 등의 분야에서 주로 사용되며, 비정상 데이터의 수집과 라벨링이 어려운 문제를 해결하는 데 큰 장점을 지닌다. 또한, 비정상 데이터가 알려지지 않은 상황에서도 정상 패턴에 벗어나는 데이터를 자동으로 비정상으로 분류할 수 있어, 새롭게 등장하는 위협에도 잘 대응할 수 있다.

단일 클래스 기반 이상 탐지의 연구 갈래로는 GAN(Generative Adversarial Network) 기반 방법론[8]과 오토인코더(autoencoder) 기반 방법론[4-5]이 두드러진다. GAN은 생성자(generator)와 판별자(discriminator)라는 두 개의 신경망으로 구성되는 모델이다. 생성자는 정상 데이터와 유사한 데이터의 분포를 학습하고, 판별자는 이 데이터가 정상인지 가짜인지 판단하는 역할을 한다. 이러한 경쟁적 학습을 통해 정상 데이터의 특성을 학습한 후, 새로운 데이터가 이 정상 분포에 얼마나 적합한지를 평가하여 비정상 여부를 결정한다. 이 방

법은 고차원 데이터에서 정상 데이터의 복잡한 분포를 잘 학습할 수 있는 장점이 있지만, 학습 과정이 불안정할 수 있으며, 판별자의 성능에 따라 탐지 능력이 달라질 수 있다.

오토인코더 기반 방법론은 입력 데이터를 압축하여 저차원 잠재 공간에 표현한 후, 이를 다시 원래 데이터로 복원하는 과정을 학습한다. 이 과정에서 정상 데이터는 낮은 복원 오류를, 비정상 데이터는 높은 복원 오류를 보이게 된다. 따라서 새로운 데이터가 들어오면, 오토인코더는 이를 인코딩하고 복원하여 그 오류를 계산하며, 이 오류가 일정 임계값보다 크다면 비정상 데이터로 간주하게 된다.

2.2 메모리 증강 오토인코더

메모리 증강 오토인코더(Memory-Augmented Autoencoder, MemAE)는 오토인코더에 메모리 네트워크를 추가하여 이상 탐지 성능을 향상시킨 모델이다. 메모리 모듈은 인코더(encoder)와 디코더(decoder) 사이에서 위치하여, 학습과정에서는 정상 데이터의 특정 패턴을 대표하는 잠재 벡터(latent vector)를 학습한다. 추론 과정에서는 인코더의 출력으로 만들어진 잠재 벡터 z를 기존에 저장된 정상 패턴들 m_i의 가중 합으로 변환하는 역할을 수행한다. 이는 식 1과 같이 표현된다.

$$z_{mem} = \sum_{i=1}^{K} w_i m_i \qquad (1)$$

여기서 w_i는 각 메모리 슬롯의 가중치를, K는 메모리 슬롯의 사이즈를 의미하며, z_{mem}은 메모리 모듈을 통해 새롭게 계산된 잠재 표현을 의미한다. 메모리 슬롯의 가중치 벡터 w_i는 입력 데이터 잠재 벡터 z와 메모리 슬롯 m_i간의 유사도를 기반으로 계산되며, 식 2와 같이 표현된다.

$$w_i = \frac{\exp(d(z-m_i))}{\sum_{j=1}^{K}\exp(d(z-m_i))} \qquad (2)$$

여기서 d는 유사도 측정함수로 코사인 함수가 일반적으로 사용된다.

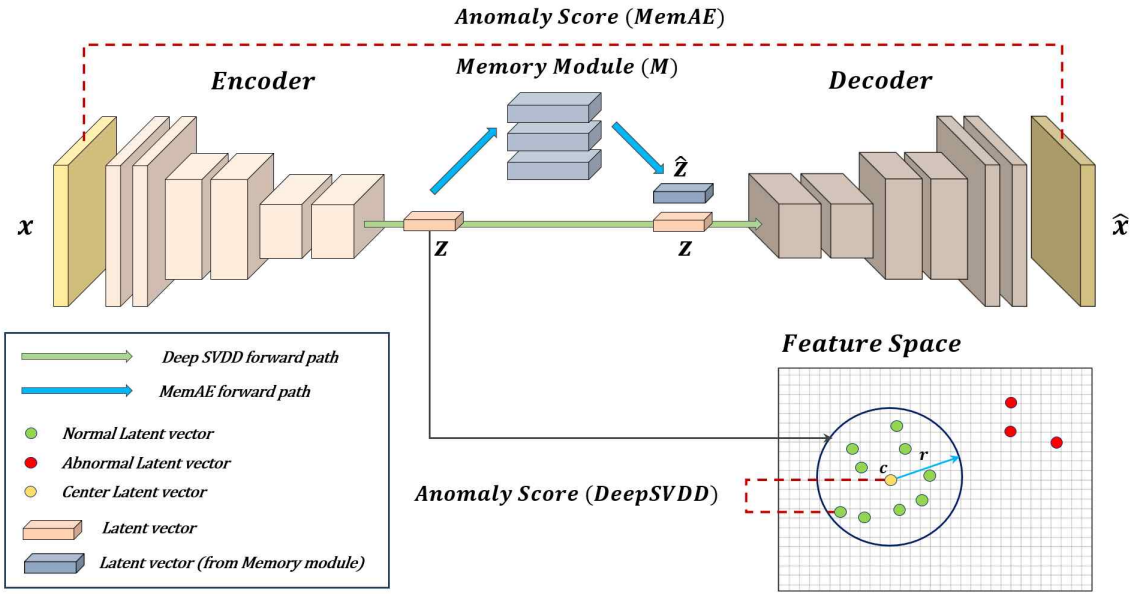
정상 데이터만을 학습한 메모리 모듈은 비정상적인 입력 데이터의 잠재 벡터를 정상 데이터의 패턴들로 강제 변환하여, 디코더로부터 복원된 결과가 높은 재구성 오류(reconstruction error)가 발생하는 것을 유도하도록 동작한다. 이는 일반적인 오토인코더 모델에서 발생할 수 있는 정상 데이터의 과적합 문제를 완화하며, 다양한 이상 패턴에 대해 높은 탐지 성능을 보이도록 만든다.

2.3 DeepSVDD

Deep Support Vector Data Description (Deep SVDD)는 심층 신경망(Deep Neural Network)을 기반으로 한 비지도 학습 방식의 이상 탐지 모델이다. DeepSVDD는 정상 데이터의 잠재 표현을 대표하는 중심 벡터로부터 정상 데이터의 경계를 표현하는 초구(hypersphere)를 학습하는 것을 목표로 한다. 중심 벡터는 정상 데이터의 잠재 표현들로부터 결정되며, 모델은 이러한 정상 데이터들의 잠재 벡터들이 구의 중심 벡터에 가까이 위치하도록 학습한다. 이러한 학습 과정은 식 3과 같이 표현된다.

$$\min_{W,c} \frac{1}{N}\|\phi(x_i;W)-c\|^2 + \lambda\|W\|^2 \qquad (3)$$

여기서 x_i는 입력 데이터를, W는 신경망의 가중치 파라미터를, $\phi(x_i;W)$ 는 인코더의 출력으로 생성된 잠재벡터, c는 잠재공간에서의 데이터 중심 벡터, λ는 정규화 파라미터로 과적합을 방지하기 위해 사용하는 파라미터를 의미한다. 이러한 목적 함수는 입력 데이터의 잠재벡터 $\phi(x_i;W)$ 와 중심 벡터 c의 차이를 최소화하는 것으로 더 작은 결정경계 초구를 형성하기 위함이다.



(그림 1). Network Intrusion Anomaly Detection 모델

따라서 입력 데이터의 잠재 표현벡터가 초구의 바깥에 위치할 경우 이상 데이터로 간주한다. 이 방법 또한 학습 데이터가 편향되어 있을 때, 또는 비정상적인 데이터가 극히 드문 경우에 유용하다.

2.4 NSL-KDD 데이터 셋

NSL-KDD 데이터 셋은 네트워크 이상 탐지 연구에서 가장 널리 사용되는 데이터 셋 중 하나로, 1999년 DARPA KDD 컵 대회에서 제공된 KDD'99 데이터 셋의 개선된 버전이다. KDD'99 데이터 셋은 네트워크 침입 탐지를 위한 표준 데이터 셋으로 오랜 기간 사용되었으나, 데이터 중복성과 불균형성으로 인해 연구자들 사이에서 다양한 비판을 받았다. 이러한 문제를 해결하기 위해 NSL-KDD 데이터 셋이 개발되었다.

3. 제안하는 방법

본 논문에서는 비지도 학습 기반 접근 방식 중 관련 연구 2.2-2.3에서 소개한 MemAE와 DeepSVDD 모델을 활용하여 네트워크 이상 탐지를 수행하고, 새로운 사이버 위협에 대한 대응력을 향상시키는 방법을 제안한다. 이 두 모델은 단일 클래스 이상 탐지 방법을 기반으로 정상적인 행동이나 패턴을 학습한 후, 이와 다른 데이터를 비정상으로 탐지한다. 이를 통해 새로운 비정상 데이터를 수집하지 않고도 효과적인 탐지가 가능하며, 사이버 위협에 대한 대응력을 크게 향상시킬 수 있다. 또한, 추가적인 공격 데이터를 수집하고 라벨링 하는 데 따르는 시간 및 비용 문제를 해결하여, 실제 네트워크 이상 탐지 환경에 매우 적합한 방법을 제공한다.

3.1 네트워크 트래픽 이상점수

제안된 모델들은 오토인코더를 기반으로 발전한 모델들이만, 정상 데이터와 비정상 데이터를 구분하는 이상 점수를 획득하는 방식에서, (그림 1)과 같이 중요한 차이점을 보인다. DeepSVDD는 잠재 공간(latent space)에서 이상 점수를 계산한다. 모델은 정상 데이터를 잠재 공간으로 매핑한 후, 이 공간에서 정상 데이터가 특정 중심 주위에 모이도록 학습한다. 이상 점수는 새로운 입력 데

이터가 이 잠재 공간에서 중심으로부터 얼마나 멀리 떨어져 있는지를 기반으로 계산되며, 식 4와 같이 표현된다.

$$S_{svdd} = \|\phi(x_i; W) - c\|^2 \quad (4)$$

여기서 x는 입력데이터를, ϕ는 학습된 인코더 네트워크를, W는 인코더 네트워크의 가중치를, c는 학습된 인코더 네트워크로부터 획득된 정상 데이터들의 중심벡터를 의미한다. 이상 점수 S_{svdd}는 입력 특징벡터 $\phi(x; W)$와 정상의 중심점 c와의 거리를 의미한다. 즉, 입력 데이터가 정상 클래스일 경우 중심 벡터에 가까운 위치에 있게 되며, 비정상 데이터일 경우 중심에서 멀리 떨어지게 되어 높은 이상 점수를 부여받는다.

반면, MemAE는 입력 공간(input space)에서 이상 점수를 계산한다. 모델은 정상 데이터를 기반으로 학습된 메모리 모듈을 사용하여 입력 데이터를 재구성한다. 재구성된 데이터와 원본 데이터 간의 차이를 재구성 오류로 정의하고, 이 오류가 클수록 해당 데이터는 비정상으로 간주된다. 이를 나타낸 이상 점수는 식 5와 같이 표현된다.

$$S_{MemAE} = \|x - \hat{x}\|^2 \quad (5)$$

여기서 x는 입력 데이터를 $\hat{x}$는 모델의 출력 즉 복원된 입력 데이터를 의미한다. 즉 MemAE 모델의 이상 점수는 입력 데이터가 잠재 공간에서 어떻게 표현되는지와 무관하게, 원본 데이터의 복원 가능성을 통해 이상성을 평가한다는 점에서 차이가 있다.

3.2 임계값 설정 및 위협탐지

이상 점수를 기반으로 정상 및 비정상 네트워크 트래픽을 구분하기 위해서는 임계값이 필요하다. 이는 검증 데이터 셋을 통해 획득되며 이보다 작을 경우에는 정상으로 간주되고, 클 경우에는 위협으로 판단한다. 검증 데이터 셋의 정상 샘플들의 이상 점수들을 모두 획득한 뒤, 백분위 수를 통해 검증 데이터 셋에서 가장 좋은 F1 점수를 보여준 임계값을 사용한다. 이때 검증 셋에서도 두 클래스 간 데이터의 편차가 매우 커서 편향된 임계값이 나오는 것을 고려하여, 균형 샘플링을 통해서 F1 점수를 계산하였다.

3.3 네트워크 특징 전처리

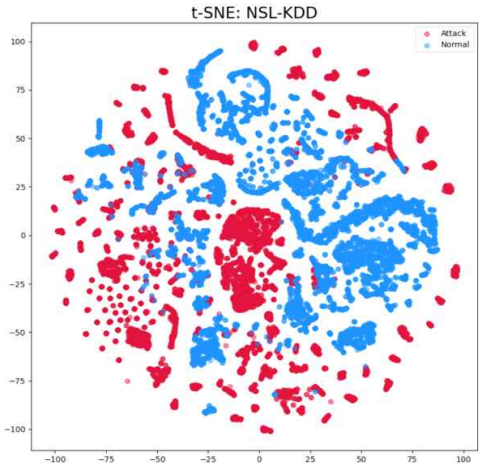
데이터의 전처리는 딥러닝 모델의 성능과 안정성을 확보하는 데 중요한 역할을 한다. 정규화와 같은 전처리 과정은 데이터의 각 속성을 동일한 스케일로 조정함으로써, 최적화 과정의 효율성을 높이고 학습을 보다 안정적으로 수행할 수 있게 한다. 또한, 속성 간 비교 가능성을 증대시켜 모델이 데이터의 중요한 패턴을 효과적으로 학습할 수 있도록 지원하며, 이상치의 영향을 줄여 모델의 학습 성능을 향상시킨다.

3.3.1 불필요 특징 제거

NSL-KDD 데이터는 네트워크로부터 추출한 442개의 속성 값으로 구성되어 있으나, difficulty 속성은 학습과 무관하기에 사전에 제거하였다. 또한, 모든 데이터에서 값이 동일하여 표준 편차가 0으로 확인된 num_outbound_cmds 속성도 제거하였다. 이를 통해 40개의 입력 특징을 가진 데이터로 변경하였으며, 결측값이 포함된 데이터는 모두 제거하였다.

3.3.2 데이터 정규화

3.3.1절에서 불필요한 특징이 제거된 40개의 특징 중 연속형 속성 데이터는, 속성 값들의 범위 차이를 왜곡하지 않고 공통된 스케일로 변경하기 위해 최소-최대 정규화(Min-max Normalization)를 진행하였다. 이후, 범주형 문자 데이터는 신경망의 입력으로 사용할 수 없는 형태이므로, 원핫(one-hot) 인코딩을 수행하였다. 이를 통해 40개의 특징으로부터 121 입력 특징으로 최종 변환되었다.



(그림 2) NSL-KDD 데이터 셋 t-SNE 시각화

(그림 2)는 네트워크 t-분산 확률적 이웃 임베딩(t-SNE) 기법을 통해 전처리된 121개의 속성으로 구성된 네트워크 트래픽을 2차원의 공간으로 임베딩하여 시각화 한 결과이다. 정상 샘플과 공격 샘플이 일부 동일한 특징 공간을 공유하여 선형적으로 분리하기 어려운 상황을 보여준다. 이는 네트워크 트래픽에서 이상 탐지 문제가 얼마나 어려운지를 잘 나타낸다.

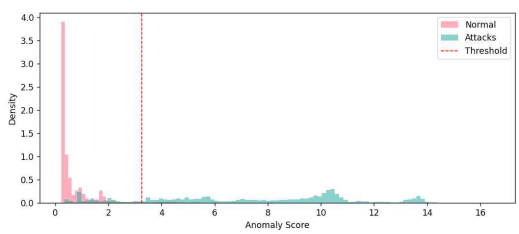
4. 실험

본 실험에서는 제안된 단일 클래스 기반 네트워크 이상 탐지 방법의 효율성을 검증하고, 지도 학습 모델이 네트워크 침입 탐지 문제에 적합하지 않음을 확인하기 위해 일부 위협 클래스 데이터를 누락한 상태에서 학습된 지도 학습 모델의 분류 성능을 측정하여 비교하였다. 실험에는 NSL-KDD 데이터 셋을 활용하였으며, 성능 평가를 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 점수(F1 Score) 지표를 사용하였다. 또한, NSL-KDD 데이터 셋의 38개 세부 공격 기법은 <표 1>에서 제시된 4개의 공격 유형으로 일반화하여 사용하였다.

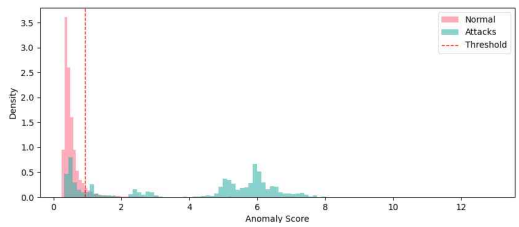
<표 1> NSL-KDD 데이터 셋

Type	Description	Train	Test
Normal	normal traffic	67343	9711
DoS	Denial of Service	45927	7458
Probe	Pre-operation for vulnerability analysis before intrusion	11656	2754
U2R	Unauthorized access to take over root authority	995	2421
R2L	Attempting unauthorized access from remote	52	200
Total		125973	22544

(그림 3-4)는 NSL-KDD 데이터 셋의 정상 데이터를 학습한 MemAE와 DeepSVDD 모델에서 테스트 셋의 이상 점수 분포를 히스토그램으로 시각화한 결과를 보여준다. x축은 이상 점수를, y축은 밀도를 나타내며, 붉은색으로 표시된 임계값은 검증 데이터 셋을 통해 도출된 값이다. 좌측에 밀집된 분포는 정상 데이터의 이상 점수 분포를 나타내며, 두 모델 모두 정상 데이터의 이상 점수가 매우 낮음을 확인할 수 있다. 또한, DeepSVDD 모델은 정상 데이터를 중심 벡터에 가깝게 학습하기 때문에 더 타이트한 임계값을 가지는 것을 알 수 있다. 이를 통해 두 모델 모두 정상 데이터만으로도 위협 데이터를 효과적으로 구분할 수 있음을 확인할 수 있다.

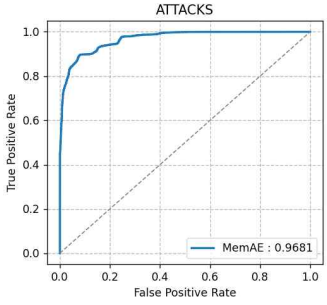


(그림 3) MemAE 모델의 이상점수 분포

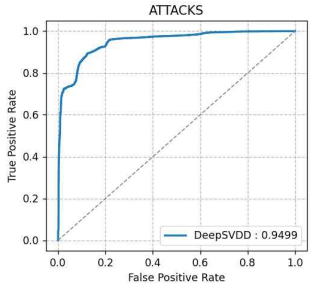


(그림 4) DeepSVDD 모델의 이상점수 분포

(그림 5-6)은 정상 트래픽만을 학습한 제안된 모델들의 ROC(receiver operating characteristic) 커브를 나타낸다. x축은 거짓 양성 비율 (False Positive Rate), y축은 참 양성 비율(True Positive Rate)를 나타내며, ROC 커브는 다양한 임계값에서 모델의 분류 성능을 평가한다. ROC 커브는 좌상단에 가깝게 위치할수록 모델의 분류 성능이 우수하다는 것을 나타내며, AUC (Area Under the Curve) 값이 1에 가까울수록 모델이 이상 탐지를 정확하게 수행할 수 있음을 의미한다.



(그림 5) MemAE 모델의 ROC 커브



(그림 6) DeepSVDD 모델의 ROC 커브

실험 결과, AUC 값은 MemAE 모델에서 0.9681, DeepSVDD 모델에서 0.9499로 계산되었으며, 이는 두 모델이 네트워크 침입 탐지 문제에서 정상 데이터와 공격 데이터를 매우 효과적으로 구분함을 나타낸다.

<표 2> 제안된 방법의 위협탐지 결과

Model	Accuracy	Precision	Recall	F1 Score
DeepSVDD	0.82	0.86	0.82	0.82
MemAE	0.89	0.90	0.89	0.89

<표 2>는 본 논문에서 제안한 네트워크 위협 탐지 방법의 성능 결과를 보여준다. 실험 결과, MemAE 모델의 분류 성능이 약 7% 더 높은 것으로 확인되었으며, 이는 (그림 4)에서 일부 비정상 샘플들의 이상 점수가 정상 데이터의 이상 점수 분포와 일부 겹치는 현상으로 인해 DeepSVDD 모델이 일반화 문제를 겪고 있음을 시사한다. 이러한 결과는 비정상 데이터와 정상 데이터 간의 경계가 명확하지 않을 때 발생할 수 있는 문제로, 모델의 성능 향상을 위해 추가적인 연구가 필요할 수 있음을 보여준다.

<표 3> DoS 공격을 학습하지 못한 지도학습 모델들과 제안된 방법의 위협탐지 결과 비교

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.67	0.79	0.67	0.72
AdaBoostClassifier	0.67	0.79	0.67	0.72
LogisticRegression	0.70	0.78	0.70	0.74
DeepSVDD	0.82	0.86	0.82	0.82
MemAE	0.89	0.90	0.89	0.89

<표 3>은 NSL-KDD 데이터 셋에서 DoS(서비스 거부) 공격을 학습하지 않고, 이를 테스트 셋에서 분류한 실험 결과를 나타낸다. 이 표는 지도 학습 모델들과 제안된 단일 클래스 모델인 DeepSVDD 및 MemAE 모델의 성능을 비교한 것이다.

지도 학습 모델인 Random Forest, AdaBoostClassifier, Logistic Regression은 각각 정확도(Accuracy) 0.67에서 0.70 사이의 성능을 보였으며, 정밀도(Precision)와 재현율(Recall)에서도 비슷한 성능을 나타냈다. 그러나 F1 점수는 0.72에서 0.74로, DoS 공격을 학습하지 않은 상태에서의 탐지 성능이 제한적임을 보여준다. 반면, 제안된 단일 클래스 모델인 DeepSVDD와 MemAE는 각각 정확도 0.82와 0.89, F1 점수 0.82와 0.89로 매우 높은 성능을 기록하였다. 특히, MemAE 모델은 Precision과 Recall 모두에서 0.90 이상의 높은 점수를 달성하여, 학습되지 않은 DoS 공격을 효과적으로 탐지할 수 있음을 보여준다. 이 결과는 제안된 단일 클래스 기반 네트워크 침입 탐지 방법이 기존의 지도 학습 모델에 비해 새로운 유형의 공격에 대한 탐지 성능이 우수하다는 것을 강력히 시사한다.

5. 결 론

본 연구에서는 기존의 네트워크 침입 탐지 시스템(NIDS)과 딥러닝 기반 지도 학습 모델들이 직면한 한계, 특히 새로운 사이버 위협과 데이터 불균형 문제를 해결하기 위해 단일 클래스 모델 기반의 네트워크 침입 탐지 방법을 제안하였다. MemAE와 DeepSVDD 모델을 중심으로 실험을 수행한 결과, 제안된 방법은 학습되지 않은 공격 유형에 대해서도 높은 탐지 성능을 보이며, 특히 기존의 지도 학습 모델들이 놓치는 부분을 효과적으로 보완할 수 있음을 확인하였다.

NSL-KDD 데이터 셋을 활용한 실험에서, DoS 공격을 학습하지 않은 상태에서도 제안된 단일 클래스 모델들이 높은 정확도와 F1 점수를 기록하였으며, 이는 새로운 위협에 대한 강력한 대응 능력을 보여준다. 특히, MemAE 모델은 모든 성능 지표에서 가장 뛰어난 결과를 보였다.

결론적으로, 본 연구의 결과는 단일 클래스 기반 이상 탐지 모델이 새로운 사이버 위협에 효과적으로 대응할 수 있는 유망한 방법임을 시사한다. 이는 미래의 네트워크 보안 체계에서 더욱 실질적이고 효율적인 솔루션을 제공할 수 있으며, 불확실성이 높은 사이버 환경에서 신뢰할 수 있는 탐지 모델로 자리매김할 수 있을 것이다. 향후 연구에서는 다양한 실제 네트워크 환경에 제안된 방법을 적용하여 추가적인 검증을 진행하고, 모델의 성능을 더욱 향상시키기 위한 방법론적 개선이 필요할 것이다.

참고문헌

[1] RAJAGOPAL, Smitha, et al. “Towards effective network intrusion detection: from concept to creation on Azure cloud”, IEEE Access, vol. 9, pp. 19723-19742, 2021.

[2] SU, Tongtong, et al. “BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset”, IEEE Access, vol. 8, pp. 29575-29585, 2020.

[3] GAO, Xianwei, et al. “An adaptive ensemble machine learning model for intrusion detection”, Ieee Access, vo. 7, pp. 82512-82521, 2019.

[4] GONG, Dong, et al. “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection”, In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1705-1714, 2019.

[5] RUFF, Lukas, et al. “Deep one-class classification. In: International conference on machine learning”, PMLR, pp. 4393-4402. 2018.

[6] TAVALLAEE, Mahbod, et al, “A detailed analysis of the KDD CUP 99 data set”, IEEE symposium on computational intelligence for security and defense applications, pp.1-6, 2009.

[7] PANG, Guansong, et al. “Deep learning for anomaly detection”, ACM Computing Surveys(CSUR), vol. 54, no.2, pp. 1-38, 2021.

[8] XIA, Xuan, et al. “GAN-based anomaly detection: A review”, Neurocomputing, vol. 493, pp. 497-535. 2022.

〔저자소개〕

민 병 준 (Byeong-jun Min)
2019년 2월 세종대학교 일반대학원 컴퓨터공학(공학석사)
2023년 2월 세종대학교 일반대학원 컴퓨터공학(공학박사)
2023년~현재 한화시스템(주) 기반기술연구소 연구원
email: alsqod1015@hanwha.com

박 대 경 (Dae-kyeong Park)
2020년 2월 숭실대학교 컴퓨터공학(공학사)
2022년 2월 세종대학교 일반대학원 컴퓨터공학(공학석사)
2022년~현재 한화시스템(주) 기반기술연구소 연구원
관심분야: 사이버 상황인식, 이상탐지, 정보보호, 시스템 침입분석 etc.
email: daekyeong.park@hanwha.com