

외부주의와 잔차연결에 기반한 효과적인 네트워크 침입탐지 모델

신 창 의*, 권 태 웅*, 이 준°

An Effective Network Intrusion Detection Model based on External Attention and Residual Connections

Chang-Yui Shin*, Tae-Woong Kwon*, Jun Lee°

요 약

컴퓨터 네트워크 및 네트워크 기술에 대한 수요가 증가함에 따라 악성공격에 노출된 표면이 자연스럽게 확장되어 악성공격은 점차 증가하고 있습니다. 따라서 네트워크에서 침입탐지 기술은 보안분야로서의 역할이므로 점점 더 중요해지고 있습니다. 하지만 정보보호 측면에서 트래픽의 페이로드가 암호화되면서 이러한 악성 트래픽에 대한 분석이 네트워크 장비에서 제한적이게 됨에 따라 침입탐지는 더욱 어려워졌습니다. 데이터를 기반으로 하는 네트워크 침입탐지에 딥러닝을 적용하면서부터 딥러닝 모델의 계산 복잡성 또한 지속적으로 제기되었습니다. 본 연구에서는 딥러닝 모델 중 이차적 계산 복잡도를 가진 트랜스포머 모델의 셀프어텐션을 대상으로 그 메커니즘적 발전을 이룬 외부주의 구조와 잔차연결을 순차 활용한 효과적인 침입탐지 모델을 설계하여 제안합니다. 제안하는 모델을 최신의 UKM-IDS20 데이터셋을 활용하여 성능을 실험한 결과로써, 오탐지율 0.24%라는 매우 우수한 결과를 보였습니다.

Key Words : Encrypted Network Traffic, Network Intrusion Detection, External Attention, Residual Connections

ABSTRACT

With the increasing demand for computer networks and networking technologies, the attack surface exposed to malicious threats has naturally expanded, leading to a steady rise in such attacks. Consequently, intrusion detection technologies within networks have become increasingly important as a core component of cybersecurity. However, from the perspective of information security, as the payload of traffic becomes encrypted, the analysis of malicious traffic by network devices has become limited, thereby making intrusion detection even more challenging. Since the application of deep learning to data-driven network intrusion detection, the computational complexity of deep learning models has been continuously raised as a concern. In this study, we propose an effective intrusion detection model that sequentially leverages external attention structures and residual connections —mechanistic advancements targeting the self-attention mechanism of transformer models, which inherently possess quadratic computational complexity. Using the latest UKM-IDS20 dataset to evaluate the performance of the proposed model, we achieved an exceptionally strong result, with a false detection rate of only 0.24%.

※ 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (RS-2023-00236032, 도로교통 공공 서비스 인프라의 암호화 사이버위협에 대한 네트워크 행위기반 보안관제 기술 개발)

• First Author : KISTI(Korea Institute of Science and Technology Information), Science and Technology Security R&D Center, superego99@kisti.re.kr

° Corresponding Author : KISTI(Korea Institute of Science and Technology Information), Science and Technology Security R&D Center, jun.lee@kisti.re.kr

* KISTI(Korea Institute of Science and Technology Information), Science and Technology Security R&D Center, taewoong.kwon@kisti.re.kr

논문번호 : KNOM2025-02-005, Received November 6, 2025; Revised November 18, 2025; Accepted December 06, 2025

I. 서 론

오늘날 컴퓨터 네트워크와 기술에 대한 수요가 급증하면서, 새로운 유형의 공격 발생 위험 또한 높아지고 있습니다. 사이버 위협 탐지는 현대 사이버 보안에서 중요한 방어 기술이 되었습니다. 네트워크 기술이 발전하고 공격 전략이 정교해지면서, 매일 발생하는 공격의 빈도가 증가하고 있습니다 [1]. 전통적인 네트워크 침입탐지 시스템(Network Intrusion Detection Systems, IDS)은 새롭고 더욱 정교한 공격 기술에 적응하는 데 어려움을 겪으며, 이를 업데이트하는 것은 계산 비용이 많이 들고 복잡할 수 있습니다. 최근에는 머신러닝을 활용하게 됨에 따라 네트워크 침입탐지 시스템에 의해 지원되는 네트워크 방어 또한 데이터 기반 접근 방식을 점점 더 많이 채택하여 데이터에 대한 의존도가 높아지고 있습니다.

게다가 기술의 발전과 함께, 사용자들은 최신 기술이 탑재된 네트워크 지원 장치를 점점 더 많이 활용하게 되면서 네트워크 트래픽 소비가 크게 증가하고 있습니다. 더 나아가, 엣지 컴퓨팅이 사물인터넷(IoT)에 점점 더 통합됨에 따라, 네트워크 전반에 걸친 데이터 교환과 증가된 연산 요구로 인해 네트워크 트래픽의 양은 빠르게 확장되고 있습니다 [2]. 이렇게 급속히 증가하는 데이터들 중에 악성 트래픽을 신속정확하게 분류하는 것이 네트워크 침입탐지의 핵심이 될 수 있습니다.

그런데 대부분의 네트워크 장비들은 보안을 강화하기 위해 TLS 및 HTTPS와 같은 암호화 통신 프로토콜을 사용합니다. 그러나 이는 네트워크 트래픽의 내용을 직접 분석하거나 분류하는 것을 점점 더 어렵게 만들 수 있습니다. 다양한 대규모의 네트워크 트래픽으로부터 악성트래픽을 탐지하거나 애플리케이션을 정확히 분류하는 태스크는 현대 통신 시스템에서 보안적 측면, 효과적인 자원 관리 측면, 그리고 서비스 품질(QoS) 향상 측면에서 필수적입니다 [2], [3].

한편, 분류학적인 측면에서 네트워크 침입탐지(Network Intrusion Detection, NID)나 네트워크 트래픽 분류(Network Traffic Classification, NTC)는 네트워크 트래픽 분석(Network Traffic Analysis, NTA)과 분석적 메커니즘의 맥락이 유사하다는 점을 염두해두고 바라볼 필요가 있습니다.

앞서 언급했듯이, 암호화 기술의 광범위한 채택으

로 인해 트래픽관련 태스크는 점점 더 어려워지고 있습니다. 페이로드 기반 분석은 더 이상 가능하지 않으며, 포트 번호나 IP 주소에만 의존하는 방식은 애플리케이션을 정확히 식별하는 데 자주 실패하는 경우가 많습니다. 포트 번호, 통계적 특징, 전문가가 정의한 특성에 기반한 기존의 분류 방법은 낮은 정확도를 보이며 실시간 트래픽 분석에 한계를 드러냅니다. 이에 따라 암호화된 트래픽에서도 침입을 탐지하거나 효과적인 애플리케이션 분류를 위해 의미 있는 특징을 추출하려는 새로운 다양한 접근법들이 등장하였습니다 [5], [6], [7], [8]. 특히, 합성곱 신경망(Convolutional Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN), 트랜스포머(Transformer) 기반 모델은 원시 바이트 수준에서 특징을 자동으로 추출할 수 있으며, 암호화된 트래픽에서도 의미 있는 패턴을 학습할 수 있는 잠재력을 보여주었습니다.

트랜스포머 기반 모델은 합성곱 신경망 및 순환 신경망 구조에 비해 우수한 성능을 입증했으나, 모델 복잡성과 계산 비용 측면에서 효율성에 대한 우려가 제기되었습니다 [7]. 트랜스포머 기반 트래픽 분류 모델은 주로 자기주의(self-attention) 메커니즘을 활용하는데, 이는 시퀀스 내 종속성을 포착하기 위해 이차적 계산 복잡도를 가지는 것이 특징입니다. 그 결과, 긴 시퀀스를 처리할 때 상당한 계산 및 메모리 오버헤드가 발생하게 됩니다. 이러한 이유로 자기주의 메커니즘을 사용하는 트랜스포머 모델은 실시간 온라인 트래픽 분류에는 적합하지 않습니다. 따라서 제한된 계산 자원을 가진 네트워크 장치에 배포하기에는 더욱 부적절할 수 있습니다.

이러한 문제를 해결하기 위해 선형 시간 상태 공간 모델(linear-time state space model)이 제안되어 더 빠른 추론과 감소된 메모리 소비를 달성하고자 하였습니다. 그러나 해당 모델 [7] 역시 한계가 존재하며, 초당 분류할 수 있는 샘플 수가 종종 열개를 넘지 못합니다. 같은 맥락에서, 자기주의의 이차적 계산 복잡도를 완화하기 위해 선형 복잡도를 제공하는 외부주의(External Attention) 메커니즘이 도입되었습니다. 이에 따라, 네트워크 트래픽 분류에 외부주의 메커니즘을 적용하는 가능성을 연구가 시도되었습니다. Yahui Hu 등 [8]은 1차원 합성곱 신경망과 외부주의 메커니즘을 계층적으로 통합하였으나, 그들의 구조에서 합성곱 신경망의 역할은 포괄적인 특징을 추출하기에는 충분하지 않았습니다.

본 연구에서는 외부주의 메커니즘 이후에 보다

포괄적인 특징을 포착하는 데 적합한 발전된 구조를 설계하고, 이에 대한 실험 결과를 제시하였습니다. 따라서 본 논문에서는 외부주의 메커니즘에 1차원 레즈넷 구조를 결합한 모델을 제안합니다. 제안하는 구조는 암호화, 익명화, 구조적 불균형 조건 하에서 네트워크 트래픽 분류를 위한 강건하고 일반화된 해결책입니다. 제안하는 구조는 실제 환경에서 흔히 발생하는 잡음이 많고 이질적인 바이트 수준 입력 시퀀스로부터 전역적 의미적 연관성과 세밀한 1차원적 패턴을 효과적으로 포착하도록 설계되었습니다.

제안하는 구조의 아키텍처는 자연어 처리(Natural Language Processing)와 시계열 분석의 강점을 결합하여 바이트 기반 입력을 효율적으로 처리할 수 있도록 합니다. 구체적으로, 외부주의 메커니즘은 전역적 맥락 인식을 담당하며, 1차원 레즈넷은 세부 특징을 추출하고 이를 더욱 표현력 있는 특징 공간으로 재구성합니다. 외부주의 메커니즘은 입력 시퀀스 전반에서 중요한 영역을 선택적으로 강조하고, 1차원 레즈넷은 이러한 영역을 정제하여 효과적으로 재표현합니다.

본 논문의 구성은 다음과 같습니다. 제 II장은 네트워크 침입탐지에 대한 기존 연구를 개관하고, 본 모델 개발의 동기를 논의합니다. 제 III장은 제안된 프레임워크의 아키텍처 구성 요소와 구현을 상세히 설명합니다. 제 IV장은 포괄적인 실험 설정과 결과를 제시하며, 제안하는 모델의 성능을 기존 다른 방법들과 비교합니다. 마지막으로 제 V장은 본 논문의 결론과 향후 연구방향을 제시합니다.

II. 관련연구

2.1. 네트워크 트래픽 분류 목적 딥러닝 모델

딥러닝은 이미지 처리, 자연어 이해, 시계열 분석 등 다양한 분야에서 탁월한 성능을 입증해왔습니다. 이러한 기술적 발전은 네트워크 트래픽 분석 분야에도 적극적으로 적용되고 있습니다. 초기 연구에서는 2차원 합성곱 신경망을 활용하여 패킷 데이터를 이미지와 유사한 형식으로 변환하여 구조적 특징을 추출하고, 이어서 장단기 메모리(Long Short-Term Memory, LSTM) 네트워크를 통해 흐름 수준의 시계열 종속성을 포착하였습니다. 그 결과로 제안된 2차원 합성곱 신경망-장단기 메모리 아키텍처 [9]는 공간적 및 시계열적 특징을 통합하는 것을 목표로

하였으나, 단일 패킷 데이터를 1차원적 관점에서 다루지는 못하였습니다.

한편, 시계열 데이터 처리에서 매우 효과적인 것으로 입증된 트랜스포머 모델 [10]의 자기주의 메커니즘은 최근 자연어 처리 분야에서 암호화된 네트워크 트래픽 분석 작업에 채택되었습니다 [11]. 그리고 BERT 아키텍처를 기반으로 한 ET-BERT [12] 모델은 패킷 및 흐름 수준에서 분류 성능을 향상시키기 위해 제안되기도 하였습니다. Masked BURST 및 Same-origin BURST Prediction 기법을 도입함으로써, ET-BERT는 ISCXP2016 데이터셋 [13]에서 각각 98.90%와 99.37%의 F1-score를 달성하였다. 그러나 강력한 성능에도 불구하고, 이 모델은 암호화된 트래픽의 바이트 수준 특징을 포착하지 못하며, 대규모 사전학습에 의존하기 때문에 막대한 계산 및 자원 비용을 초래하는 실질적 한계를 가지는 단점이 있습니다.

트랜스포머를 활용한 네트워크 트래픽 분석 분야의 또 다른 연구[14]에서는 DistilBERT 기반의 경량 모델이 제안되었으며, 최소 63바이트의 입력 데이터만으로도 높은 성능을 달성하였다. 이 연구에서는 다중 속성 패킷 간의 중복 정보를 제거하고 표현 능력을 강화하기 위해 비교학습(Comparative Learning)을 적용하였다. 그 결과, 해당 모델은 실시간 처리 및 모델 압축 측면에서 ET-BERT보다 더 실용적인 대안을 제공하였습니다. 해당 XENTC 모델은 불필요한 특징을 제거함으로써 네 개의 다중 속성 패킷을 동시에 효과적으로 처리하며, 97.0%에서 98.1% 범위의 F1-score를 달성하였습니다. 학습된 모델을 파인튜닝하여 단순화된 63바이트 표현된 패킷을 데이터로 입력받아 분류하는 데 소요된 시간은 패킷당 약 0.0093초였습니다.

네트워크 트래픽 분석에 특화된 딥러닝 모델들은 성능 향상을 위해 서로 다른 설계 철학에 기반한 독창적인 아키텍처와 데이터 처리 전략을 보여줍니다. Self-Attentive Method(SAM) [15]는 패킷 내 요소들 간의 관계를 학습하기 위해 자기주의 메커니즘을 활용하고, 그 출력을 1차원 합성곱 신경망과 통합했습니다. 저자들은 50 바이트 패킷 입력을 SAM 모델에 적용하여 94.48%의 F1-score를 달성하였습니다. 게다가 해당 연구에서는 패킷당 추론 시간이 2밀리초 미만임을 보고하였으며, 이는 온라인 환경에서 실시간 분류에 적합함을 나타내었습니다. 게다가 각 피처의 중요도를 주의(attention) 관계를 통해 분석한 결과도 제시하였습니다.

2.2 외부주의 모델

자기주의 메커니즘은 트랜스포머 모델의 기본적인 구조로써, 단일 샘플 내에서 장기 종속성을 포착할 수 있는 장점을 제공합니다. 이에 따라 자기주의 메커니즘은 컴퓨터 비전과 자연어 처리 등 다양한 분야에서 널리 채택되어 왔습니다. 해당 메커니즘은 샘플 내 모든 위치 간의 특징 관계를 집계함으로써 각 위치의 표현을 정교화합니다. 그러나 이 과정은 입력 위치의 수에 대해 이차적 계산 복잡도를 초래하기도 합니다. 또한 자기주의 메커니즘은 본질적으로 샘플 내부 관계에 집중하기 때문에, 서로 다른 샘플 간의 잠재적 상관관계는 간과되기도 합니다. 즉, 단일 샘플 내 관계적 표현을 강화하는 데에는 효과적이지만, 샘플 간 관계 모델링에는 기여하지 못합니다.

앞서의 자기주의 메커니즘의 계산 오버헤드와 샘플 간 상관관계 모델링의 부재를 해결하기 위해 외부주의 메커니즘 [16]이 제안되었습니다. 성능비교를 통한 우수성을 제시했습니다. 자기주의 메커니즘은 쿼리(query)와 키(key) 벡터 간의 유사도를 측정하여 어텐션 맵을 계산한 뒤, 이를 값(value) 벡터의 가중 집계와 결합하여 새로운 특징 맵을 생성합니다. 반면, 외부주의 메커니즘은 쿼리 벡터와 외부 학습 가능한 키 메모리 간의 유사도를 계산하고, 그 결과 어텐션 맵을 학습 가능한 값을 가지는 외부 메모리에 곱하여 새로운 특징을 생성합니다. 이 두 외부 메모리는 선형 계층을 통해 구현되며 전체 데이터셋에 걸쳐 공유됨으로써 외부주의 메커니즘의 일반화 능력을 강화합니다. 그 결과, 외부주의 메커니즘은 여러 샘플을 효과적으로 구별할 수 있는 고도로 판별적인 특징을 학습합니다. 또한 입력 크기에 대해 선형적인 계산 복잡도를 가지므로, 실행 시간 측면에서 향상된 효율성을 제공합니다.

Yahui Hu 등 [8]은 네트워크 트래픽 분석 분야에서 패킷 수준의 중요한 바이트 관계를 포착하기 위해 외부주의 메커니즘과 1차원 합성곱 신경망을 통합한 연구를 제시하였습니다. 그들의 실험에서는 패킷 헤더로부터 12바이트 특징을 추출하여 여러 데이터셋에서 96% 이상의 분류 정확도를 달성하였으며, 패킷당 평균 약 0.36밀리초의 추론 시간을 기록하였습니다. 그러나 해당 모델은 외부주의 메커니즘 출력으로부터 파생된 패킷 내 바이트 단위 종속성을 효율적으로 처리하는 데 한계가 있었습니다.

이러한 한계를 해결하기 위해, 본 연구에서는 단

일 샘플 내 중요한 바이트 수준 상호작용을 모델링할 때, 1차원 합성곱 신경망으로 인해 발생하는 관계 정보 손실을 고려하였습니다. 이에 따라, 표준화된 표 형식 입력에 적합하도록 2차원 레즈넷 아키텍처를 최적화하여 1차원 입력에 맞게 재설계해 잔차연결을 1차원 합성곱 신경망에 반영했습니다. 이러한 아키텍처적 개선이 패킷 수준에서 관계적 특성을 포착하는 데 미치는 영향은 제4장에서 제시되는 실험 결과와 분석을 통해 검증하였습니다.

III. 데이터 및 모델 구성

3.1. 데이터셋 선정

UKM-IDS20 [17]은 실제 네트워크 환경 (Universiti Kebangsaan Malaysia)에서 수집된 데이터를 기반으로 생성된 새로운 벤치마킹 데이터셋입니다. UKM-IDS20 데이터셋이 생성된 환경은 다음과 같습니다.

표 1. UKM-IDS20내 정상/공격 데이터 분포

Table 1. Distribution of Normal and Attack Data in the UKM-IDS20 Dataset

Connection Record Type			Training set / Testing set
Sum			10,308 / 2,579
Normal		-	7,140 / 1,769
Abnormal	ARP poisoning	-	476 / 116
	DoS	Mass HTTP requests	461 / 140
		TCP flood	467 / 121
		UDP data flood	445 / 108
	Scans	-	474 / 123
	Exploits	BeEF HTTP exploits	404 / 96
		Metasploits exploits	441 / 106

Hyper-V를 이용한 가상화 환경에서 Office A(서버, IDS, 스위치, 4개 호스트)와 Office B(호스트, 스위치, 웹 서버)로 구성된 네트워크를 모의했습니다. TShark, Kali Linux, Metasploit, Nmap 등의 도구를 사용하여 정상 트래픽과 악성 공격 트래픽을 생성하고 캡처되었습니다.

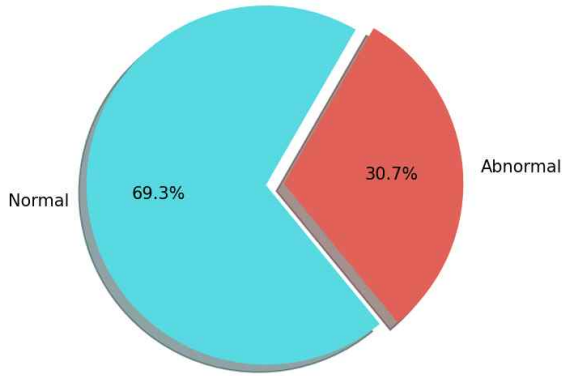


그림 1. 이진분류 데이터 분포

Fig. 1. Binary Class Data Distribution

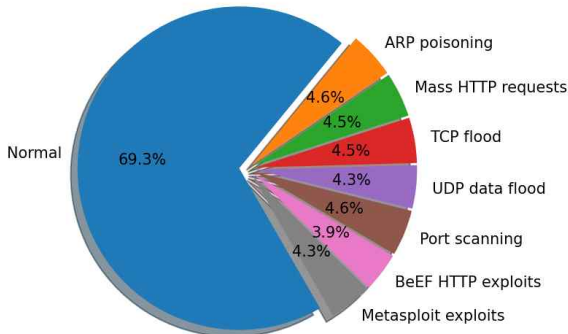


그림 2. 멀티클래스 데이터 분포

Fig. 2. Multi-class Data Distribution

데이터셋의 특징 갯수와 공격유형을 살펴보면, 총 46개의 특징을 포함하며, ARP poisoning, DoS (HTTP 요청, TCP/UDP 플러드), Scans (포트 스캔), Exploits (Metasploit MS17-010, BeEF)의 네 가지 주요 공격 유형과 정상(Normal) 트래픽을 포함합니다. 46개의 특징이 전처리되는 과정은 데이터셋 공개논문[17]에 세부적으로 설명되어 있습니다. 데이터셋의 구성측면에서 총 12,887개의 연결 레코드로 구성되며, 훈련 세트(10,308개)와 테스트 세트(2,579개)로 나뉩니다. 표 1은 데이터셋의 클래스별

세부 현황을 나타냅니다. 정상 레코드 수가 공격 레코드보다 많도록 구성하여 기계 학습 모델의 분류 성능 저하 문제를 완화하였습니다.

한편, Rough-set 이론을 이용한 특징 의존도 (Dependency Ratio, DR) 분석 결과, UKM-IDS20 데이터셋의 특징들이 KDD99 및 UNSW-NB15 데이터셋보다 높은 관련성을 가지는 것이 제시되기도 하였습니다. 특히 기본 연결정보 외에 시간 기반 특징과 ARP 관련 특징이 공격 탐지에 중요한 역할을 하는 것으로 분석되었습니다. 이에 따라 이러한 특징들이 피처에 반영되었다는 점이 해당 데이터셋이 가지는 특장점입니다.

UKM-IDS20 데이터셋은 12,887개의 데이터로 구성되어 있습니다. 다음 그림 1, 2는 그에 대한 수량적 비율 분포도를 파이차트로 나타낸 것인데, 그림 1은 정상 트래픽과 공격트래픽을 기준으로 2진 분류한 것이며, 그림 2는 공격트래픽을 각 7개의 클래스로 세분하고 정상트래픽을 1개 클래스로 구분하여 총 8개 클래스로 분류한 것입니다. 이러한 데이터의 클래스 라벨을 원한 인코딩하고, 숫자 데이터를 일정 구간으로 나눈 뒤, 결측치를 처리해서 PyTorch 텐서 데이터셋으로 변환하는 과정을 거쳐 모델에 입력되도록 했습니다. 이러한 입력 시퀀스에 대해 포지션 임베딩과 바이트 임베딩을 적용하여 각 위치와 바이트 단위 토큰을 128차원 벡터로 매핑하였습니다..

3.2. 모델 설계

제안하는 모델은 입력 데이터를 순차적 아키텍처를 통해 처리하며, 전처리와 임베딩 과정을 거친 후 외부주의 메커니즘과 1차원 레즈넷 모듈을 적용했습니다. 그림 3은 제안하는 아키텍처적 계층 구조를 나타냅니다. 외부주의 메커니즘과 1차원 레즈넷 모두 입력 시퀀스를 처리하도록 설계되었지만, 그 기능적 역할과 초점은 근본적으로 상이합니다. 설계 관점에서 볼 때, 두 모듈은 모델 전체 아키텍처 내에서 상호 보완적인 작업을 수행하도록 순차적으로 설계되었습니다.

외부주의 메커니즘은 전체 입력 시퀀스에서 전역적으로 중요한 영역을 식별하고 선택하는 역할을 주로 담당합니다. 이는 각 요소의 상대적 중요성을 평가하는 메커니즘으로 작동하며, 일반적으로 소프트맥스(soft-max) 기반 가중치 부여 방식을 통해 구현됩니다. 이를 통해 모델은 데이터 내에서 가장 정보의 의미성이 높은 부분에 계산 자원을 집중할 수

있으며, 이러한 선택적 강조는 모델이 고수준 의미적 연관성을 포착하는 능력이 향상되도록 합니다.

반면, 1차원 레즈넷은 외부주의 메커니즘에 의해 강조된 영역으로부터 세밀한 특징을 추출하고 재표현하는 역할을 합니다. 잔차 연결(residual connections) [18], [19], [20]과 1차원 합성곱 연산을 활용하여 선택된 정보를 정제하고, 이를 보다 표현력 있고 판별적인 특징 공간으로 변환합니다. 본질적으로, 외부주의 메커니즘이 어디에 집중할지를 결정한다면, 1차원 레즈넷은 무엇을 추출하고 어떻게 더 정밀하게 표현할지를 결정합니다.

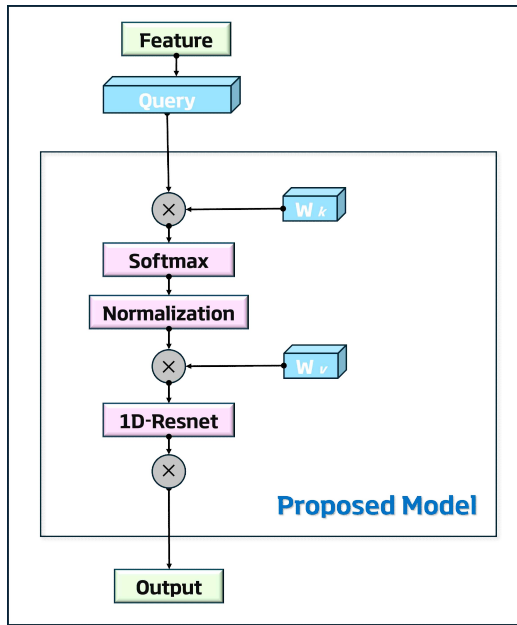


그림 3. 제안하는 모델

Fig. 3. Proposed Model

3.2.1 외부주의 레이어

수식 (1)를 순차적으로 설명하면, $K = QW_k$ 로 쿼리 행렬 Q 와 학습 가능한 가중치 W_k 를 곱해 K 를 얻습니다. 이어서 $A = \text{Softmax}(K/\tau)$ 로 스케일링된 소프트맥스를 적용합니다. 마지막으로 각 행을 정규화하여 $\hat{A} = \frac{A}{\sum_{j=1}^S A_{ij}}$ 를 얻습니다. 즉 쿼리행렬을 선형적으로 변환하고, 소프트맥스를 적용해 각 행단위 정규화 과정을 모두 포함한 하나의 식입니다.

$$\hat{A}_i = \frac{\text{Softmax}(\frac{QW_k}{\tau})_i}{\sum_{j=1}^S \text{Softmax}(\frac{QW_k}{\tau})_{ij}} \quad (1)$$

외부주의 메커니즘은 본 신경망 아키텍처에서 핵심적인 역할을 수행하며, 입력 데이터 내 전역적으로 중요한 위치를 식별하고 선택적으로 강조할 수 있도록 모델을 지원합니다. 이러한 메커니즘은 분류, 생성, 이해와 같은 다운스트림 작업에 필수적인 두드러진 표현을 추출하는 데 기여합니다. 기존의 자기주의 메커니즘이 내부 토큰 간 상호작용에만 의존하는 것과 달리, 외부주의 메커니즘은 추가적인 키 공간을 도입하여 보다 유연하고 문맥 인식적인 주의 점수(attention score) 계산을 가능하게 합니다.

외부주의 메커니즘의 내부 동작 논리는 다음과 같이 설명될 수 있습니다. 먼저, 입력 시퀀스는 사전에 정의된 어텐션 키 공간으로 투영되며, 이는 각 입력 요소의 상대적 중요성을 평가하기 위한 참조 프레임워크로써 기능을 합니다. 이러한 투영은 일반적으로 학습 가능한 선형 변환을 통해 수행되며, 입력 특징을 일련의 어텐션 키로 매핑됩니다. 이후 이 키들은 소프트맥스(soft-max) 함수를 통해 처리되어 값이 정규화되고, 입력 시퀀스 전반에 걸친 중요도 점수 분포가 산출됩니다. 이러한 확률적 가중 메커니즘은 모델이 가장 정보성이 높은 구성 요소에 집중하고, 덜 관련된 요소는 억제할 수 있도록 합니다.

마지막 단계에서는 어텐션으로 강화된 표현을 재구성합니다. 전역적 문맥 강화(global contextual enhancement) 이후, 가중된 입력 특징은 단순하더라도 효율적인 투영 계층을 통해 다시 초기 특징 공간으로 선형 매핑됩니다. 이러한 설계는 입력 데이터의 구조적 무결성을 유지하는 동시에 전역적으로 인식된 의미 정보를 풍부하게 하여, 모델이 보다 정교한 이해와 의사결정을 수행할 수 있는 능력을 강화합니다.

3.2.2 1차원 레즈넷 레이어

수식 (2)는 잔차학습 구조(Residual Learning Framework)를 일반화하여 (1)에서 얻은 결과를 처리하는 과정입니다. 따라서 수식 (2)를 위에서와 마찬가지로 설명하면, $T(Data)$ 는 입력텐서($Data$)의 축을 교환하는 전치연산을 의미합니다. $F_{Conv1D_{max(2)}}$ 는 동일 패딩을 갖는 1차원 합성곱 연산을 2회 적용하는 변환 연산을 적용하는 것을 의미합니다. 이어서 배치 정규화 연산자인 N_{BN} 을 적용하고, σ 는 비선형 활성화 함수인 ReLU 연산이 적용됨을 의미합니다. 마지막으로 이러한 과정에서 얻은 비선형 변환 경로에 $S(T(EA_{out}))$ 이라는 잔차경로를 더함으로써 최

중 출력값을 정의하는 형태입니다.

$$y = F_{Conv1D_{mac(2)}}(\sigma(N_{BN}(F_{Conv1D_{same(2)}}(S(T(EA_{out})))) + S(T(EA_{out}))) \quad (2)$$

본 모델은 다양한 길이의 입력 시퀀스를 유연하게 수용할 수 있도록 설계된 점이 중요합니다. 서로 다른 입력 크기에서도 구조적 일관성을 보장하기 위해 모든 합성곱 연산은 same padding [18]을 사용하며, 이를 통해 입력과 출력 간의 시퀀스 길이가 유지됩니다. 이러한 설계 선택은 잔차 학습을 안정화시키고, 수동적인 크기 조정(adjustment)이나 절단(truncation)의 필요성을 제거합니다. 그리고 어텐션 가중 입력(attention-weighted input)을 수신함으로써, 1차원 레즈넷 모듈은 목표 지향적 특징 정제를 수행하고, 국소적으로 두드러진 정보를 보다 판별력 있고 표현력이 풍부한 특징 공간으로 변환됩니다. 이는 모델의 전체적인 표현 능력을 향상시키며, 실제 데이터에 내재된 변동성에 대해 강건하게 적응할 수 있도록 합니다.

IV. 성능실험 및 분석

4.1. 실험 환경 및 모델 주요 파라미터

모델은 CUDA 12.9 기반의 Python 3.10.18 및 Pytorch 2.7.1를 기반으로 구현되었습니다. 하드웨어는 20Core Intel(TM) Ultra 7 265K CPU @ 3.90GHz의 CPU와 NVIDIA GeForce RTX(12GB 메모리)의 GPU가 장착된 서버로 구성되었습니다.

앞서 기술된 바와 같이 외부주의모델은 128차원의 벡터를 입력으로 받음에 따라 1차원 레즈넷 모델도 동일한 벡터를 받도록 처리하도록 했습니다. 필터와 커널 사이즈는 각각 256, 16으로 설정하였습니다.

4.2. 실험결과 및 분석

모델을 평가하기 위한 요소로 사용된 지표들은 아래 수식 (3) ~ (8)과 같습니다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

$$Detection Rate(DR) = \frac{TP}{TP + FN} \quad (7)$$

$$False Alarm Rate(FAR) = \frac{FP}{FP + TN} \quad (8)$$

위에서 True Positive(TP)는 실제 True인 정답을 True라고 예측(정답)하는 것이고, False Positive(FP)는 실제 False인 정답을 True라고 예측(오답)하는 것입니다. False Negative(FN)는 실제 True인 정답을 False라고 예측(오답)하는 것이고, True Negative(TN)는 실제 False인 정답을 False라고 예측(정답)하는 경우입니다. 수식(7)은 네트워크 침입 탐지 모델의 성능을 측정할 때 주로 쓰이는 Detection Rate(DR)이며, 분류모델의 성능을 측정할 때 사용되는 수식 (5)의 Recall과 같은 값을 나타냅니다.

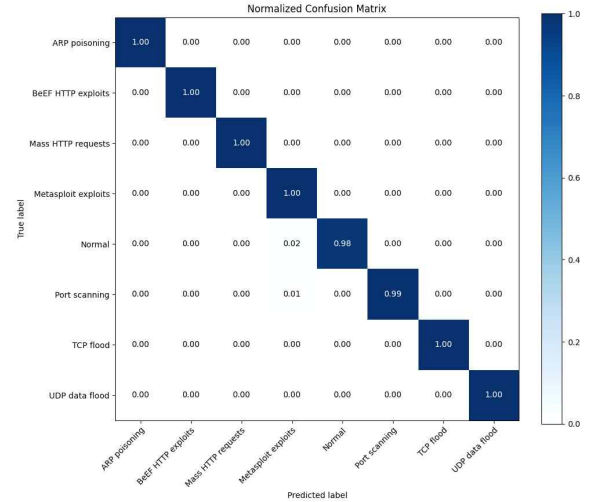


그림 4. 컨퓨전 매트릭스

Fig. 4. Confusion Matrix

그림 4는 제안하는 모델의 학습과정 이후에 예측 성능을 나타낸 컨퓨전 매트릭스입니다. 그림 4를 통해서 모델의 성능을 살펴보자면, 컨퓨전 매트릭스의 좌측 대각선 값은 모델의 정답을 맞춘 결과를 나타냅니다. 따라서 대각선 값들이 0.98~1.00 사이의 값을 가지므로 라벨별로 정답률이 매우 높은 수준임을 확인할 수 있습니다. 그리고 분류결과에 대해 모

델이 예측한 것 중 잘못된 것을 기준으로 모델이 잘못 판단할 수 있는 부분으로, Normal 트래픽이나 Port scanning 트래픽을 일부 Metasploits으로 잘못된 예측을 할 수 있음을 나타냅니다.

표 2. 다른 연구들과의 성능비교

Table 2. Performance comparison with other research

	DANN	HOE-DANN	Proposed
Detection Rate (=Recall)	95.18%	96.91%	98.23%
False Alarm Rate	7.97%	7.57%	0.24%
Accuracy	93.60%	94.66%	99.30%
Precision	-	-	99.51%
F1-score	-	-	98.83%

표 2에서는 제안하는 모델과 다른 모델들이 동일 한UKM-IDS20 데이터셋을 대상으로 성능을 비교한 결과를 확인 할 수 있습니다. 세부적으로 살펴보면 탐지율(Detection Rate)이 98.23%을 보임으로써 제안하는 모델이 HOE-DANN보다 1.32% 높은 성능을 가집니다. 그리고 오경보율(False Alarm Rate)은 0.24%를 보임으로써 HOE-DANN보다 7.33% 낮아, 허위경보비율이 매우 낮아 뛰어난 탐지 성능을 가지고 있는 것을 의미합니다.

표 3. 제안 모델의 총 파라미터 갯수

Table 3. Total Number of Parameters in Our Proposed Model

	Number of parameters	Magnitude (Size)
BERT	109.0 M	440.0 MB
DistilBERT	66.0 M	267.0 MB
ALBERT	11.0 M	49.0 MB
Proposed model	2.7 M	10.9 MB

표 3에서는 모델의 총 파라미터 개수와 모델의 크기를 나타내는데, 학습가능한 파라미터의 수가 3 백만개 이하이고 모델의 사이즈도 10MB정도입니다. 따라서 다른 딥러닝 모델보다 매우 간결한 정도이면서도, 앞서의 우수한 성능을 보여주는데서 효과적인 네트워크 침입탐지 모델임을 알 수 있습니다. 46 개의 피쳐로된 입력을 처리하는데 걸리는 시간은 초당 4,638개로 매우 빠른 처리 성능을 보여줍니다.

V. 결론 및 향후 연구방향

5.1. 결론

본 논문에서는 효과적인 외부주의 메커니즘과 1차원 레즈넷 모델을 연계하는 레이어 구조를 설계하여 효과적인 침입탐지모델을 설계한 결과를 보여주었습니다. 제안된 모델에서 외부주의 메커니즘은 전체적인 전역적 맥락 인식을 담당하며, 1차원 레즈넷 레이어는 세부 특징을 추출하고 이를 더욱 표현력 있는 특징 공간으로 재구성하는 역할을 담당하도록 설계되었습니다.

제안하는 모델의 성능을 실험하는데 사용한 UKM-IDS20데이터셋은 기존의 다른 데이터셋 보다 최신 공격 양상을 반영하여 더 높은 특징 복잡성을 가진 데이터셋이기에 본 연구의 성능실험에 의미있는 기여를 한 것으로 판단할 수 있습니다. 제안하는 모델의 구성적 특징을 바탕으로 UKM-IDS20 데이터셋을 대상으로 실험한 결과 오경보율이 0.24% 정도의 매우 낮은 수치를 보여 매우 우수한 IDS 탐지 모델임을 드러냈습니다.

5.2. 향후 연구방향

향후 연구 진전을 통해 발전시키고 싶은 방향은 다음과 같습니다. 첫째, 앞서 제시한 UKM-IDS20 데이터셋은 수집되는 환경에서 정상과 공격 트래픽의 비율을 불균형한 특성을 자연스럽게 반영합니다. 이러한 불균형한 데이터셋에 대한 여러 접근 방법에 대해 살펴보고, 모델의 성능을 보다 향상시킬 수 있는 방법론에 대한 새로운 접근을 시도하려고 합니다. 둘째, 보다 확장된 메커니즘 연구와 함께 다양한 데이터와 모델을 다루어, 추가 실험결과를 통해 성능 일반화 검증을 하려고 합니다. 셋째, 클래스간 오탐 발생원인을 분석하는 것에 XAI를 적용하여 인지가능한 해석을 통해 모델의 실용적 측면에 대한 접근으로 보다 확장하고자 합니다.

References

- [1] E. Kabir, J. Hu, H. Wang, and G. Zhuo, "A novel statistical technique for intrusion detection systems," *Future Generation Computer Systems*, vol. 79, pp. 303–318, Feb. 2018, doi: 10.1016/j.future.2017.01.029.
- [2] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017, doi: 10.1109/JIOT.2017.2683200.
- [3] R. H. Serag, M. S. Abdalzaher, H. A. E. A. Elsayed, and M. Sobh, "Software Defined Network Traffic Classification for QoS Optimization Using Machine Learning," *J. Netw. Syst. Manag.*, vol. 33, no. 2, p. 41, Apr. 2025, doi: 10.1007/s10922-025-09911-6.
- [4] J. H. Kalwar and S. Bhatti, "Deep Learning Approaches for Network Traffic Classification in the Internet of Things (IoT): A Survey," Feb. 01, 2024, arXiv: arXiv:2402.00920. doi: 10.48550/arXiv.2402.00920.
- [5] P. Zhu, G. Wang, J. He, Y. Dong, and Y. Chang, "An encrypted traffic identification method based on multi-scale feature fusion," *Array*, vol. 21, p. 100338, Mar. 2024, doi: 10.1016/j.array.2024.100338.
- [6] S. Xu, J. Han, J. Wang, and Y. Bai, "An encrypted traffic classification method based on autoencoders and convolutional neural networks," *PLOS One*, vol. 20, no. 9, p. e0333276, Sept. 2025, doi: 10.1371/journal.pone.0333276.
- [7] T. Wang, X. Xie, W. Wang, C. Wang, Y. Zhao, and Y. Cui, "NetMamba: Efficient Network Traffic Classification via Pre-training Unidirectional Mamba," Oct. 20, 2024, arXiv: arXiv:2405.11449. doi: 10.48550/arXiv.2405.11449.
- [8] J. Yu, Y. Choi, K. Koo, and D. Moon, "A novel approach for application classification with encrypted traffic using BERT and packet headers," *Comput. Netw.*, vol. 254, p. 110747, Dec. 2024, doi: 10.1016/j.comnet.2024.110747.
- [9] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network traffic classifier with convolutional and recurrent neural networks for Internet of Things," *IEEE Access*, vol. 5, pp. 18042–18050, 2017, doi: 10.1109/ACCESS.2017.2747560.
- [10] A. Vaswani et al., "Attention Is All You Need," Aug. 02, 2023, arXiv: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
- [11] H. Y. He, Z. Guo Yang, and X. N. Chen, "PERT: Payload Encoding Representation from Transformer for Encrypted Traffic Classification," in *2020 ITU Kaleidoscope: Industry-Driven Digital Transformation (ITU K)*, Feb. 2020, pp. 1–8. doi: 10.23919/ITUK50268.2020.9303204.
- [12] X. Lin, G. Xiong, G. Gou, Z. Li, J. Shi, and J. Yu, "ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification," in *Proceedings of the ACM Web Conference 2022*, Apr. 2022, pp. 633–642. doi: 10.1145/3485447.3512217.
- [13] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of Encrypted and VPN Traffic using Time-related Features," in *Proceedings of the 2nd International Conference on Information Systems Security and Privacy*, Rome, Italy: SCITEPRESS - Science and Technology Publications, 2016, pp. 407–414. doi: 10.5220/0005740704070414.
- [14] C.-Y. Shin, J.-T. Park, U.-J. Baek, and M.-S. Kim, "A Feasible and Explainable Network Traffic Classifier Utilizing DistilBERT," *IEEE Access*, vol. 11, pp. 70216–70237, 2023, doi: 10.1109/ACCESS.2023.3293105.
- [15] G. Xie, Q. Li, and Y. Jiang, "Self-attentive deep learning method for online traffic classification and its interpretability," *Comput. Netw.*, vol. 196, p. 108267, Sept. 2021, doi: 10.1016/j.comnet.2021.108267.
- [16] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond Self-Attention: External Attention

Using Two Linear Layers for Visual Tasks,”
IEEE Trans. Pattern Anal. Mach. Intell., vol. 45,
no. 5, pp. 5436–5447, May 2023, doi:
10.1109/TPAMI.2022.3211006.

- [17] M. S. Al-Daweri, S. Abdullah, and K. A. Zainol Ariffin, “An adaptive method and a new dataset, UKM-IDS20, for the network intrusion detection system,” Computer Communications, vol. 180, pp. 57–76, Dec. 2021, doi: 10.1016/j.comcom.2021.09.007.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 10, 2015, arXiv: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.
- [19] P. Tchatchoua, G. Graton, M. Ouladsine, J. Muller, A. Traore, and M. Juge, “1D ResNet for Fault Detection and Classification on Sensor Data in Semiconductor Manufacturing,” in 2022 International Conference on Control, Automation and Diagnosis (ICCAD), Lisbon, Portugal: IEEE, July 2022, pp. 1–6. doi: 10.1109/ICCAD55197.2022.9853997.
- [20] C. Szegedy et al., “Going Deeper with Convolutions,” Sept. 17, 2014, arXiv: arXiv:1409.4842. doi: 10.48550/arXiv.1409.4842.

신 창 의 (Chang-Yui Shin)



2003년 2월 : 육군사관학교 운영분석학과 졸업
2007년 2월 : 고려대학교 전자컴퓨터공학과 석사
2025년 2월 : 고려대학교 컴퓨터정보학과 박사
2025년 7월~현재 : 한국과학기술정보연구원 과학기술보안연구센터 박사후연구원

<관심분야> 네트워크관리, 네트워크 트래픽 분석, Ad-hoc 라우팅, 전송데이터링크, AI, 정보보호

권 태 웅 (Taewoong Kwon)



2012년 2월 : 숭실대학교 컴퓨터학부 졸업
2014년 8월 : 고려대학교 정보보호대학원 정보보호학과 석사
2014년 12월~현재 : 한국과학기술정보연구원 과학기술보안연구센터 선임연구원

<관심분야> 정보보호, 보안관계, 네트워크 보안, 네트워크 가시화

이 준 (Jun Lee)



2017년 8월 : 한국항공대학교 정보공학 박사
2016년 10월~2017년 11월 : 일본산업기술종합연구소(AIST) 인공지능연구센터(AIRC) Research Assistant

2017년 12월~2019년 12월 : 일본산업기술종합연구소(AIST) 인공지능연구센터(AIRC) 박사후연구원

2020년 1월~2024년 12월 : 한국과학기술정보연구원 과학기술보안연구센터 선임연구원

2025년 1월~현재 : 한국과학기술정보연구원 과학기술보안연구센터 책임연구원

<관심분야> 인공지능, 보안관계, 지식추출, 자연언어처리, 네트워크 보안