

트랜스포머 기반 세션 클러스터 시퀀스 단위 공격 트래픽 탐지 및 분류

장윤성¹, 김지민¹, 김주성¹, 백의준¹, 최양서², 최태상³, 김명섭^{*}

¹고려대학교 네트워크 관리 연구실

²한국전자통신연구원 지능형네트워크보안연구실

³주식회사 모비젠

e-mail : brave1094@korea.ac.kr

Transformer-based Attack Traffic Detection and Classification in Session-Cluster Level

Yoon-Seong Jang¹, Ji-Min Kim¹, Ju-Sung Kim¹, Ui-Jun Baek¹,

Yang-Seo Choi², Tae-Sang Choi³, Myung-Sup Kim^{*}

¹Network Management Lab, Korea University

²Electronics and Telecommunications Research Institute (ETRI)

³MobiGen Co., Ltd.

요 약

암호화된 네트워크 환경에서 딥러닝 기반 세션 단위 공격 트래픽 탐지 및 분류 연구는 페이로드의 암호화로 충분한 분류 단서 확보의 한계를 가진다. 본 논문에서는 이를 해결하기 위해 세션 간 관계적 특성을 학습하는 세션 클러스터 시퀀스 단위 트랜스포머 기반 공격 트래픽 탐지·분류 프레임워크를 제안한다. 제안된 프레임워크는 세션 클러스터 생성기, 세션 내 패턴 특징을 학습하는 Session2Token Embedding 모듈, 그리고 세션 간 관계 특징을 학습하는 트랜스포머 기반 분류기로 구성된다. 실험 결과, 제안 모델은 암호화된 트래픽 환경에서 L4 페이로드만을 사용하여 탐지 정확도 99.99%, 분류 정확도 78.04%를 달성하며, 기존 단일 세션 단위 모델인 ET-BERT 대비 우수한 성능을 입증하였다.

1. 서론

인터넷 기술의 확산과 서비스 상용화로 암호화 프로토콜(TLS, QUIC 등)의 사용이 보편화되었고, 이와 함께 암호화된 채널을 활용한 네트워크 공격도 증가하고 있다. IoT, 클라우드, 실시간 스트리밍의 보편화는 암호화된 대용량 트래픽을 양산하여 탐지 난이도를 높였다. 전통적인 시그니처 기반·페이로드 기반 탐지 및 딥 패킷 인스펙션(DPI)은 페이로드 암호화로 인해 유효성이 저하되고 있다. 암호화로 인해 페이로드에서 미리 정의한 규칙이나 시그니처를 검출하기 어렵고, 내부 내용 없이 외형만으로는 정상 트래픽과 구별되지 않는 경우가 많아졌다. 이에 따라 CNN, RNN/LSTM, Transformer·BERT 등 다양한 ML/DL 기법을 패킷·버스트·플로우·세션 단위 입력에 적용하는 연구가 활발히 진행되고 있다.

세션 단위 접근은 세션 내부의 헤더 및 패턴을 분류 특징으로 활용해 성능을 개선해 왔으나, 최근 연구들은 암호

화된 페이로드로부터는 유의미한 분류 단서를 얻기 어렵고, 대신 헤더의 강력한 식별자(SII, Strong Identification Information)에 의존할 경우 특정 환경에 과적합되어 일반화 성능이 저하될 수 있음을 지적하였다. 본 연구는 단일 세션 내 분류 단서 확보의 한계와 분류기의 일반화 실패 문제를 해결하고자, L4 페이로드만을 사용하여 세션 간 관계를 학습하는 세션 클러스터 시퀀스 단위의 모델링을 제안한다. 본 논문에서는 세션 클러스터 시퀀스 생성 방식과 Session2Token Embedding을 통해 세션 내 패턴특징과 세션 간 관계 특징을 통합 학습하고, 암호화된 환경에서 탐지·분류 성능 및 일반화 능력 개선 연구에 도움이 되고자 한다.

2. 관련연구

딥러닝 기반의 네트워크 트래픽 분류기 연구에서는 다양한 입력 단위를 사용한 연구가 활발히 진행되고 있다. 트래픽에서 입력 단위(Traffic Granularity)는 모델이 관찰하는 최소 범위를 규정하며, 탐지 목적에 따라 적절한 단위가 달라진다. 패킷(Packet)은 헤더와 페이로드를 포함하는 가장 작은 전송 단위로 바이트·페이로드 수준의 정밀한 특징 학습에 적합하나 암호화 시 정보가 크게 제한된

이 논문은 과기정통부·정보통신기획평가원의 정보통신기술개발사업(R&D, 정보화) 사업으로 수행한 결과((No. RS-2025-02219319, 양자컴퓨터 공격에도 안전한 양자암호 기반 제로트러스트 보안 네트워크/서비스 및 제어/관리 기술 표준개발)이며, 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구(00235509, ICT융합 공공 서비스·인프라의 암호화 사이버위협에 대한 네트워크 행위 기반 보안관계 기술 개발)이며, 2024년 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임.(P0028063, 2024년 산업기술국제협력사업)

다. 버스트(Burst)는 동일 방향의 연속된 패킷 묶음으로 전송데이터의 집약성이나 순간적 패턴을 포착하는 데 유리하다. 플로우(Flow)는 5-튜플 기반의 단방향 통신 집합으로 통계적 특성 분석에 적합하며, 세션(Session)은 양방향 핸드셰이크·요청·응답을 포함하는 서비스 단위로 송·수신자 간의 데이터 송수신 및 응답을 포함하기에 가장 대표적으로 사용된다. [1]은 datagram/packet 및 burst 수준에서 BERT 계열의 사전학습을 도입하여 패킷 단위의 문맥 표현을 학습하고 이를 패킷, 세션 단위의 암호화 트래픽 분류 모델 ET-BERT를 연구하였다. [2]는 flow/session 수준에서 Transformer 기반의 사전학습 기법을 적용해 세션·플로우 전체의 통계적·구조적 패턴을 학습하는 YaTC 모델을 통한 접근을 제안하였다.

한편 최근에는 암호화된 트래픽 분석의 본질적 한계를 정리한 연구도 제기되었다. [3]은 암호화 트래픽 분류 연구에서 과적합을 유발할 수 있는 강력한 식별자들(Strong Identification Information, SII)을 정의하고 이를 데이터 누수 기반 과적합, 문맥 기반 과적합, 시간 기반 과적합으로 분류하였다. 또한 일부 연구에서 보고된 “암호화된 페이로드로부터 패턴을 학습할 수 있다”는 주장에 대해, 일반적으로 암호화된 페이로드로부터는 길이를 제외한 유의미한 분류 단서를 확보하기 어렵다고 강조하였다. 저자들은 ET-BERT와 YaTC와 같은 모델을 활용한 실험을 통해 이러한 과적합 가능성과 세션 단위 입력의 암호화된 페이로드의 분류 단서 확보 가능성에 대한 주장을 실증적으로 뒷받침하였고, 이는 암호화된 환경에서 기존 입력 단위를 사용한 딥러닝 기반 트래픽 분류 연구가 직면한 근본적 한계를 부각시켰다.

다중 세션을 입력 단위로 삼는 연구도 존재한다. [4]은 규칙·상관관계 기반으로 여러 연관 세션을 집계·분석하여 보안 이벤트를 탐지하는 다중-세션 기법을 제시하였으며,

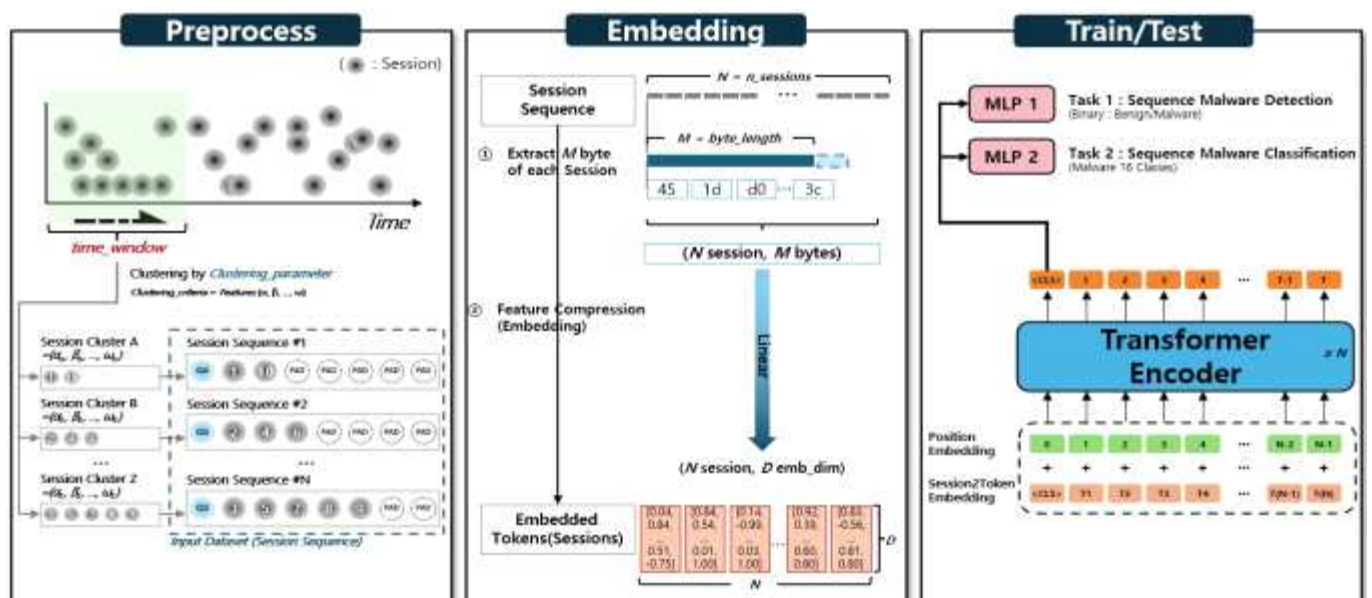
이 접근은 세션 간 상관관계 정보를 규칙·통계적 패턴으로 직접 해석 가능하도록 설계되었다. [5]는 M3F에서 다중 프로토콜에 걸친 연관 세션들을 시간 순으로 묶어 각 세션으로부터 추출된 특징(프로토콜 유형, 송신/수신 바이트 등)을 요약하여 상태열을 만들고, 패밀리별 1차 마르코프 체인을 지문(fingerprint)으로 학습·매칭하는 ML 기반 구조를 제안하였다. 본 연구에서는 이러한 규칙 기반, ML 기반 방법론과는 다르게, DL 기반 프레임워크를 사용하여 자동으로 특징을 추출하도록 한 점에서 차이가 존재한다.

본 연구는 선행 연구들을 계승하면서도 구조적으로 다음과 같이 차별화된다. 우선 여러 세션을 시간 순으로 정렬·군집화하여 세션 클러스터 시퀀스를 생성하는 전처리 파이프라인을 도입하고, 각 세션을 수동 특징이 아닌 학습 가능한 토큰 표현으로 변환하며 세션 내 패턴 특징을 학습하는 Session2Token Embedding을 적용한다. 이후 Transformer 인코더의 Self-Attention을 통해 토큰화된 세션들 간 관계 특징을 학습하여 세션 클러스터 시퀀스 단위 입력의 공격 트래픽 탐지와 기법 분류를 수행한다. 이를 통해 암호화된 네트워크 환경에서 세션 간 관계 특징이라는 추가적인 분류 단서 확보가 가능함을 보여주며, 분류 성능과 일반화 가능성의 개선을 도모하고자 한다.

3. 본론

다양한 공격 트래픽은 그 목적과 수행 방식에 따라 여러 유형으로 분류된다. 대표적으로 Port Scanning은 시스템의 접근 가능한 포트를 식별하기 위한 탐색 행위로, 이후 침입이나 악용을 위한 사전 단계로 수행된다. Flooding은 대량의 트래픽을 단시간 내 발생시켜 서버 자원을 소모시키고 서비스 거부를 유도한다. Brute-force 혹은 Credential-based 공격은 관리자 페이지 등 인증이 필요한 웹 자원에 무차별 대입이나 예상 가능한 시그니처

그림 1. 트랜스포머 기반 세션 클러스터 시퀀스 단위 공격 탐지 및 분류 도구



(Signature)를 사용해 비인가 접근을 시도한다. 이러한 공격은 세션 내 패턴 특징(Intra-session Feature)과 세션 간 관계 특징(Inter-session Feature)이 결합된 형태로 나타난다. 세션 내 패턴 특징은 개별 세션 내부의 트래픽 흐름으로, 패킷(Packet)의 헤더 필드, 바이트 수준 패턴, 특정 TCP 플래그 조합, 세션 내 L7 프로토콜의 발생 양상 등을 포함한다. 세션 간 관계 특징은 공격 수행 과정에서 연속적으로 발생하는 세션들 간의 연관성을 의미한다. 예를 들어 Flooding은 짧은 시간 내 동일 목적지 주소와 포트로 다수의 세션을 생성하는 관계 특징과, 각 세션의 패킷 수가 적고 페이로드 크기가 작은 세션 내 패턴 특징을 동시에 가진다. 대부분의 공격은 이 두 특징을 함께 가지므로, 본 연구는 암호화된 트래픽 환경에서 이를 모두 활용한 탐지·분류 방법을 제안한다.

그림 1은 제안된 트랜스포머 기반 세션 클러스터 시퀀스 단위 탐지·분류 구조를 세 부분으로 보여준다: 1) 수집된 트래픽으로부터 세션 클러스터 시퀀스를 생성하는 세션 클러스터 시퀀스 생성기, 2) 각 세션 내 패턴 특징을 추출하는 Session2Token Embedding, 3) Self-Attention으로 세션 간 관계 특징을 학습하는 트랜스포머 기반 분류기이다.

1) 세션 클러스터 시퀀스 생성기는 세션을 시작 시간 순으로 정렬하고, 지정된 time_window 내의 세션을 클러스터 후보로 묶는다. 출발지·목적지 IP, 목적지 Port, 패킷 개수, L7 프로토콜 등의 클러스터링 파라미터로 세션을 그룹화하고 이를 시간 순으로 배열해 세션 클러스터 시퀀스를 구성한다. 생성된 클러스터는 짧은 시간 내에서 유사한 특징을 공유하는 세션들의 집합이 되며, 이 세션들은 첫번째 패킷의 시작 시간에 따라 정렬되어 토큰 시퀀스의 형태로 표현된다.

2) Session2Token Embedding은 각 세션으로 고정 길이 n 바이트를 추출하고, 선형 변환 계층(Linear layer)을 통해 n 바이트를 임베딩 차원으로 압축한다. 기존 토큰나이저·임베딩 방식은 입력 어휘 크기와 임베딩 매트릭스의 비효율성 때문에 본 설정에 적합하지 않으므로 선형 변환을 선택했다. 이 Linear 레이어를 사용한 커스텀 임베딩이 [6]에서 제안되었던 반복적 Convolution-pooling 구조와의 비교 실험에서 가장 높은 성능을 보였으며, 이는 (1) 트래픽 데이터 특성상 고차원 표현이 불필요할 가능성, (2) 암호화된 페이로드가 대부분인 환경에서 저수준 통계 패턴 추출이 유효하다는 점으로 예상되나, 추가 분석이 필요하다.

표 1. 연구 활용 데이터셋 정상 6종 트래픽 및 공격 16종 트래픽

Benign	Amazon	Kakaotalk	Netflix	TencentCloud	Ucloud	Youtube
Attack	Active ScanningPort TCP	Active ScanningPort UDP	AdminPanel Discovery	BruteForce MySQL	BruteForce SSH	BruteForce Web
	BruteForce Mongo	Direct Network FloodSSH	Forbidden Bypass	GetFlood	HTTPCC	JWT
	SlowLoris	SQL Injection	SSTI	SynFlood		

다.

3) 트랜스포머 분류기는 Session2Token Embedding에서 생성된 토큰 시퀀스에 위치 임베딩(Position Embedding)을 추가해 입력한다. Self-Attention은 시간 순 토큰 간 상호작용을 학습해 세션 간 관계를 포착하며, 시퀀스 맨 앞의 [CLS] 토큰 표현을 Task1(공격 트래픽 탐지)과 Task2(공격 기법 분류)를 위한 두 개의 MLP 헤드에 연결해 멀티태스크 학습을 수행한다. 역전과 과정에서 [CLS] 토큰은 두 과제를 동시에 만족하도록 최적화되어 세션 내부·외부 특징의 통합적 활용을 가능케 한다.

제안된 모델은 패킷, 버스트, 플로우, 세션 단위를 넘어 상위 입력 단위인 세션 클러스터 시퀀스를 도입하고, Session2Token Embedding과 트랜스포머 Self-Attention을 결합해 암호화 환경에서 공격 트래픽을 효과적으로 탐지·분류하도록 설계되었다. 본 접근은 세션 내 패턴 특징과 세션 간 관계 특징을 동시에 학습한다는 점에서 기존 연구와 차별화되며, 이것은 페이로드가 암호화되는 최신의 암호화 트래픽 환경에서 세션 외부의 관계 특징을 DL 기반 공격 트래픽 탐지 및 분류에 사용하는 첫 시도이다.

4. 실험 분석

본 연구의 모델 평가는 혼동행렬의 네 값(TP, TN, FP, FN)을 기반으로 산출되는 Accuracy, Precision, Recall, F1-score를 사용하였으며, 2024년에 수집된 사실 공격 트래픽 데이터셋을 사용하였다(자동화된 공격 기법 16종, 정상 트래픽 6종). 본 논문에서의 평가 과제는 두 가지로 구성된다: (1) Detection - 정상/공격의 이진 분류, (2) Classification - 정상(통합)과 16개 공격 기법을 포함하는 총 17개 클래스 분류. 데이터셋 구성은 표 1에 정리되어 있다. 현실적으로 서로 다른 네트워크 환경에서 수집된 대규모의 공격 데이터셋을 확보하기 어려운 한계가 있어, 본 연구는 헤더 정보(환경 의존적 식별자)에 의한 과적합 가능성을 배제하고자 L3/L4 헤더를 제거한 입력 환경(L4 페이로드만)에서 추가 실험을 수행하였다. 비교 실험을 위해 ET-BERT의 공개 코드를 기반으로 세션 단위 설정을 재구성하여 동일한 데이터·전처리 조건에서 제안 모델과 비교하였다. 표 2에는 L4 페이로드만을 사용한 조건에서의 ET-BERT와 제안된 모델의 Detection 및 Classification 정확도를 정리하였다.

	L4 페이로드	
	탐지	분류
ET-BERT	98.36%	54.65%
제안된 모델	99.05%	78.04%

표 2. ET-BERT와 제안된 모델 정확도 비교

실험에서 ET-BERT와 비교하였을 때 탐지 정확도는 소폭 향상된 성능을 보였으나 미미하지만, 분류 정확도는 ET-BERT보다 약 23%p(54.65% → 78.04%) 향상된 성능을 보였다. 이를 분석하기 위해 추출된 표3은 L4페이로드만을 사용한 실험에서의 데이터셋 내 일부 암호화된 공격 기법 클래스에 대한 Precision과 Recall 비교이다. 비암호화 계열의 공격(예: BruteForce-MySQL, BruteForce-Mongo 등)은 ET-BERT와 제안 모델 모두에서 높은 Precision·Recall을 보이며 두 모델 간 차이가 거의 없다. 반면 TLS/HTTPS 계열로 페이로드가 암호화되는 공격(예: BruteForce-Web, HTTPCC, SSTI 등)에서는 ET-BERT가 L4 페이로드만을 입력으로 했을 때 거의 분류 단서를 얻지 못하거나 낮은 성능을 보이는 반면, 제안 모델은 동일한 입력 조건에서 유의미한 성능 개선을 보였다. 이러한 결과는 입력으로 주어지는 단일 세션의 L4 페이로드 자체는 동일하더라도, 제안 모델이 세션 간 관계와 세션 내부의 저수준 패턴을 함께 학습함으로써 암호화로 인해 약화된 개별 세션의 특징을 보완할 수 있었음을 시사한다. 다만 본 해석은 관찰적 근거에 기반하므로, 정확히 어떤 L7 필드나 어떤 시간적·연속적 관계가 성능 향상에 기여했는지는 추가적인 정량적 분석이 필요하다.

표 3. L4 페이로드 사용한 ET-BERT와 제안된 모델의 클래스별 Precision, Recall 비교 (BF: Brute Force)

ET-BERT	Precision	Recall	제안된 모델	Precision	Recall
BF MySQL	100%	99.7%	BF MySQL	100%	100%
BF Mongo	100%	99%	BF Mongo	100%	100%
BF Web	0%	0%	BF Web	79.2%	90.5%
HTTP CC	12.5%	46.3%	HTTP CC	22.3%	63.8%
SSTI	0%	0%	SSTI	66.4%	77.5%

5. 결론

본 연구에서는 세션 간 관계 특징을 학습하는 세션 클러스터 시퀀스 단위의 트랜스포머 기반 공격 트래픽 탐지·분류 모델을 제안한다. 제안 모델은 유사한 특성을 가진 세션들을 클러스터링하여 시간 순으로 정렬한 세션 클러스터 시퀀스를 생성하는 생성기와, 각 세션의 바이트 시퀀스를 저차원 벡터로 압축하여 세션 내 패턴 특징을 추출하는 Session2Token Embedding을 핵심 구성요소로 포함한다. 이렇게 생성된 토큰 시퀀스는 트랜스포머 인코더의 Self-Attention 연산을 통해 세션 간 관계 특징을 학습

하며, 이를 통합하여 세션 클러스터 시퀀스 단위에서 탐지 및 분류를 수행한다. 실험 결과 제안 방법은 탐지 정확도 99.99%, 분류 정확도 78.04%를 달성하여, 단일 세션 기반 접근으로는 얻기 어려웠던 암호화된 네트워크 환경에서의 분류를 세션 간 관계 특징을 통해 보완함으로써 향상된 분류 성능과 L3, L4 헤더에 의한 과적합이 완화되어 일반화 성능을 향상시킬 수 있음을 보였다.

다만, 실험에 사용한 데이터셋의 다양성 확대와, 세션 내 바이트 단위 처리로 인한 일부 필드 값 의미의 손실 문제를 해결할 필요가 있다. 향후 연구에서는 사전학습 기반 표현학습, 헤더 정보를 값으로 입력하는 대신 메타정보로 변환하여 세션 간 관계성을 반영하도록 하는 방법, 그리고 다양한 도메인에 대한 교차검증과 XAI 연구를 통해 모델의 일반화 가능성과 실운영 적용성을 검증할 것을 제안한다.

참고문헌

- [1] 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 Proc. of the ACM Web Conf. (WWW), pp. 633 - 642, Lyon, France, April 2022.
- [2] R. Zhao, M. Zhan, X. Deng, Y. Wang, Y. Wang, G. Gui and Z. Xue, "Yet Another Traffic Classifier: A Masked Autoencoder Based Traffic Transformer with Multi-Level Flow Representation," in Proc. of the AAAI Conf. on Artificial Intelligence, vol. 37, no. 4, pp. 5420 - 5427, Washington, DC, USA, February 2023.
- [3] N. Wickramasinghe, A. Shaghaghi, G. Tsudik and S. Jha, "SoK: Decoding the Enigma of Encrypted Network Traffic Classifiers," in Proc. of IEEE Symp. on Security and Privacy (SP), pp. 1825 - 1843, San Francisco, CA, USA, May 2025.
- [4] M. Mizutani, S. Shirahata, M. Minami and J. Murai, "ROOK: Multi-Session based Network Security Event Detector," in Proc. of the IEEE Int. Symp. on Applications and the Internet (SAINT), pp. 48 - 54, Turku, Finland, July 2008.
- [5] J. Liu, Q. Xiao, L. Xin, Q. Wang, Y. Yao and Z. Jiang, "M3F: A novel multi-session and multi-protocol based malware traffic fingerprinting," Computer Networks, vol. 227, 109723, pp. 1 - 16, March 2023.
- [6] Y.S. Jang, J.M. Kim, J.S. Kim, U.J. Baek, Y.S. Choi and M.S. Kim, "Multi-task detection and classification of encrypted traffic sequences via Session2Token embedding," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2025, pp. 1-4, Kaohsiung, Taiwan, September 2025.