암호화 트래픽 분류를 위한 구조 중심 사전학습 및 미세조정

유경민, 남승우, 장윤성, 김주성, 김지민 김명섭*
고려대학교, *고려대학교
{rudals2710, nam131119, brave1094, jsung0514, illiard1209, *tmskim}@korea.ac.kr

Structure-Aware Pretraining and Fine-Tuning for Encrypted Traffic Classification

Gyeong-Min Yu, Seung-Woo Nam, Yoon-Seong Jang, Ju-Sung Kim, Ji-Min Kim, Myung-Sup Kim*
Korea Univ., *Korea Univ.

요 약

TLS, QUIC 등 암호화 프로토콜의 보편화로 기존의 패킷 내용 기반 트래픽 분류 방법이 한계에 봉착하였다. 본 논문은 암호화 환경에서도 변하지 않는 프로토콜 구조에 주목하여, 네트워크 트래픽을 필드-레이어-패킷-세션의 계층 구조로 모델링하는 구조 중심 프레임워크를 제안한다. 프레임워크는 (1) 프로토콜 경계 정렬 구조 토큰화, (2) 계층적 셔플 사전학습, (3) 계층별 중요도 기반 미세조정의 세 요소로 구성된다. IP 주소나 포트 번호 없이 응용 계층(L7) 정보만을 사용하는 제약 조건 하에서, CSTNET - TLS 1.3에서 95.0%, 300개 클래스 사설 데이터셋에서 79.2%의 정확도를 달성하며 기존 방법 대비 월등한 성능을 보였다.

Ⅰ. 서 론·

현대 네트워크 트래픽의 대부분은 TLS 1.3과 같은 강력한 암호화 프로토콜을 사용하며, 이로 인해 패킷 페이로드의 내용을 직접 분석하는 DPI(Deep Packet Inspection)[1] 기술은 사실상 무력화되었다. 이에 대응하여 최근 딥러닝 기반 연구들은 원시 바이트 시퀀스를 직접 학습하는 방향을 취해 왔다. CNN[2-3], LSTM[4], Transformer[5-8] 기반 모델들이 제안되었고 일정한 성과를 거두었으나, 암호화된 환경에서 바이트 값은 암호화 변환의 결과물로 거의 무작위 분포를 가지기 때문에 바이트 기반 표현의 신뢰성과 일반화 능력에는 근본적인 한계가 있다.

본 논문은 이러한 문제를 해결하기 위해 구조 중심(structure-centric) 관점을 제안한다. 프로토콜이 정의하는 필드 경계, 레이어 스택 구조, 패킷 순서 등의 계층적 구조 정보는 암호화 여부와 무관하게 항상 관측 가능하다. 본 연구는 이러한 구조적 속성이 트래픽 분류를 위한 견고하고 암호화 불변적인 특징을 제공한다는 가설을 검증하고, 이를 기반으로 한 표현 학습 프레임워크를 제시한다.

기존 연구는 크게 세 가지로 분류된다 :

1) 바이트 중심 접근법은 [5-6] 등이 대표적으로, 트래픽을 바이트 시퀀스로 취급하고 마스크 언어 모델(MLM)이나 마스크 오토인코더(MAE) 방식으로 사전학습한다. 암호화 환경에서 바이트 값이 무작위화되면 표현력이 급격히 저하되는 문제가 있다.

2) 부분적 구조 접근법은 대표적으로 [7] 등이 있으며, 패킷 타이밍이나 헤더 필드 정보를 일부 반영하지만, 프로토콜 계층 구조를 통합적으로 모델링하지는 않는다.

3) 구조 인식 접근법은 대표적으로 [8] 이 있으며 프로토콜 필드 구조를 표현 공간에 직접 반영하려 시도하지만, 여전히 마스크된 바이트나 필드 값 복원을 학습 목표로 삼아 암호화된 값 예측의 효용성 문제가 남는다.

본 논문은 이들 모두의 한계를 극복하기 위해 값 복원이 아닌 구조적 순서 관계 학습을 핵심 목표로 삼는다.

Ⅱ. 본론

본 논문의 전체 시스템 구조도는 Fig 1.과 같으며, 전체 프레임워크는 계층적 토큰화, 구조 중심 입력 표현, 계층적 셔플 사전학습, 계층별 중요도 기반 미세조정으로 나눌 수 있다.

2.1 계층적 토큰화

네트워크 세션은 (1)과 같이 4단계 계층 구조로 표현된다.

이 계층 구조는 프로토콜 명세에 의해 결정론적으로 정의되므로, 암호화 여부와 관계없이 유지된다. 본 논문에서는 각 계층 수준을 표현하기 위해 세 가지 특수 토큰을 도입하며, 이는 (2)에 정의된다. 입력 시퀀스는 필드 단위로 토큰화된 입력과 세 가지 특수 토큰으로 구성되며, 최대 256개의 토큰으로 제한된다. 각 필드 토큰은 최대 1,400바이트의 정보를 표현할 수 있으며, 전체 시퀀스에는 패딩(PAD) 토큰이 포함될 수 있다.

$$\text{세션 } S \rightarrow \text{패킷 } P_i \rightarrow \text{레이어 } L_{i,j} \rightarrow \text{필드 } F_{i,j,k} \quad (1)$$

CLS_S: 세션 수준 집계 토큰
CLS_P: 패킷 수준 집계 토큰 (2)
CLS_L: 레이어 수준 집계 토튼

·본 연구는 정부(중소벤처기업부)의 재원으로 중소기업기술정보진흥원(TIPA)의 창업성장기술개발사업(TIPS)의 지원을 받아 수행되었음 (RS-2025-25466990, 5G/6G 네트워크 Cross-domain Observability Engineering Orchestrator 기술 및 표준 개발). 또한, 본 연구는 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(IITP)의 지원을 받아 수행되었음 (00235509, ICT 융합 공공 서비스·인프라의 암호화 사이버 위협 대응을 위한 네트워크 행위 기반 보안 관제 기술 개발).

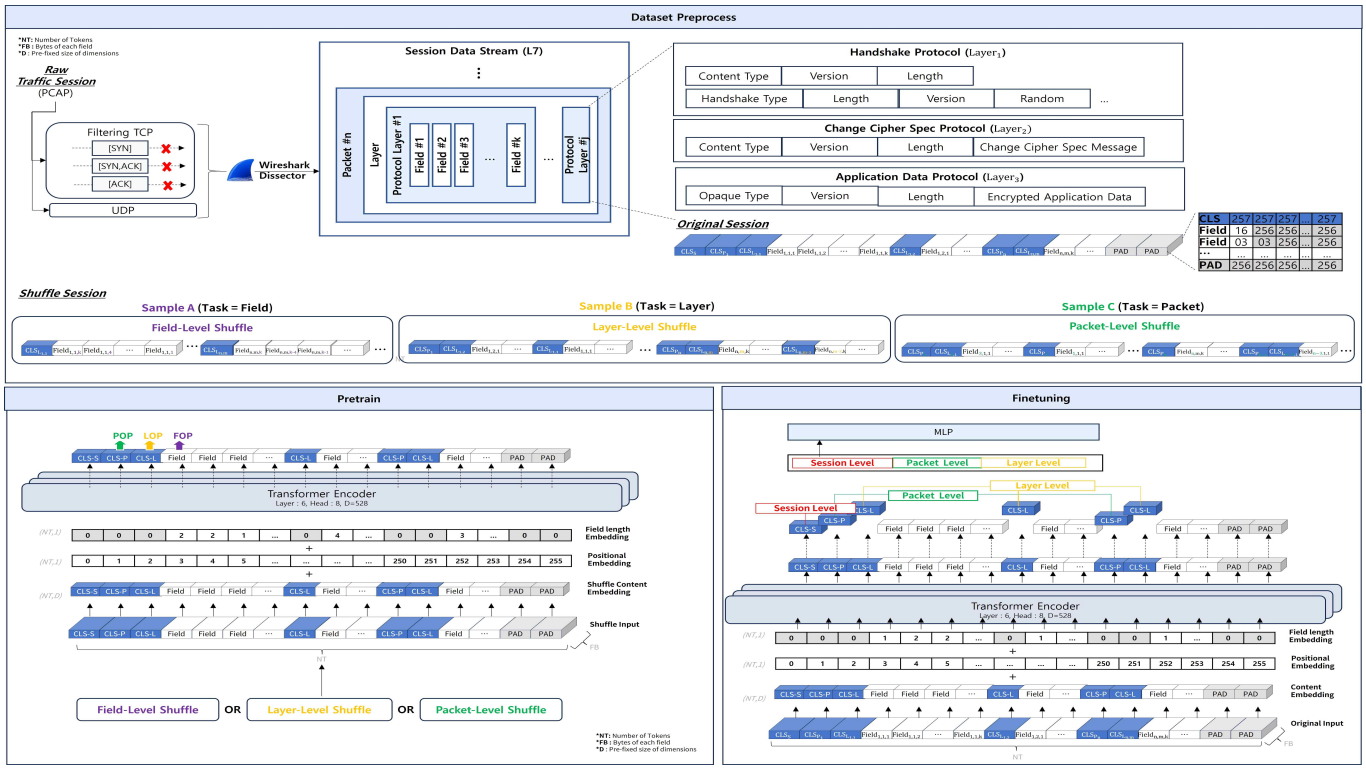

그림 1. 시스템 전체 구조

2.2 구조 중심 입력 표현

각 토큰의 임베딩은 세 가지 요소의 합으로 구성되며 (3)과 같다.

$$e = e_{content} + e_{position} + e_{\leq ngth} \quad (3)$$

-내용임베딩($e_{content}$): 필드 내 바이트를 평균/최대 풀링으로 집계하여 암호화로 인한 바이트 순서 의존성을 제거한다.

-위치임베딩($e_{position}$): 입력 시퀀스 내 전역 위치 인덱스를 학습 가능한 임베딩 테이블로 인코딩한다.

-필드길이임베딩(e_{length}): TLS 레코드 길이 등 프로토콜이 평문으로 노출하는 길이 정보를 별도 임베딩으로 인코딩한다. 길이 패턴은 암호화 여부와 무관하게 의미 있는 구조적 신호를 제공한다.

2.3 계층적 셔플 사전학습

레이블 없이 구조적 관계를 학습하기 위해 세 가지 독립 셔플 복원 태스크를 정의한다. 각 태스크는 해당 계층의 CLS 토큰이 예측을 담당하며, 전체 사전학습 손실은 세 태스크의 교차 엔트로피 손실의 합이며, (4)와 같다.

-FOP(FieldOrderPrediction): 레이어 내 필드 순서 무작위 섞기, 원래 필드 오프셋 복원

-LOP(LayerOrderPrediction): 패킷 내 레이어 순서 무작위 섞기원래 레이어 오프셋 복원

-POP(PacketOrderPrediction): 세션 내 패킷 순서 무작위 섞기원래 패킷 오프셋 복원

$$L = \lambda f \cdot L_{field} + \lambda l \cdot L_{layer} + \lambda p \cdot L_{packet} \quad (4)$$

셔플 후 위치 임베딩은 섞인 순서를 반영하여 위치 단축 학습을 방지하고, 모델이 내용 및 길이 단서로부터 구조 관계를 추론하도록 유도한다.

2.4 계층별 중요도 기반 미세조정

파인튜닝 단계에서는 단일 CLS 토큰 기반의 전역 표현 대신, 계층적 어텐션 풀링을 통해 다수준 표현을 집계한다. 각 계층 수준(레이어/패킷/세션)에 대해, 해당 수준에 속한 토큰 표현들의 어텐션 가중합을 통해 수준별 표현 벡터 v_ℓ를 생성한다. 이후, 계층 간 중요도의 차이를 반영하기 위해 학습 가능한 게이팅 계수 g_ℓ를 도입하고, 수준별 표현을 (5)와 같이 결합한다:

$$g = g_{layer} \cdot v_{layer} + g_{packet} \cdot v_{packet} + g_{session} \cdot v_{session} \quad (4)$$

여기서 v_ℓ는 계층 ℓ의 집계된 표현이며, g_ℓ는 해당 계층의 기여도를 나타내는 학습 가능한 스칼라 가중치이다. 이러한 설계를 통해 입력 트래픽의 특성에 따라 가장 유의미한 구조적 수준을 선택적으로 강조할 수 있으며, 고정된 풀링 방식 대비 보다 유연하고 표현력이 높은 특징을 학습할 수 있다. 최종 분류는 레이블 스무딩(ε=0.1)을 적용한 3층 MLP를 통해 수행된다.

Dataset	#Class	Sessions (Raw)	Sessions (Filtered)	Encryption
Private-1	48	71,841	50,166	~27%
Private-2	48	70,379	32,402	~47%
Private-3	300	399,722	351,536	~88%
ISCX VPN 2016	16	187,336	4,802	~1%
ISCX Tor 2016	14	57,605	26,165	~10%
CSTNET TLS 1.3	120	46,372	46,372	~99%

표 1. 실험에 사용된 데이터셋 요약

Method	ISCX VPN 2016		ISCX TOR 2016		CSTNET TLS 1.3		Private#3		Private#2		Private#21	
	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1
ET-BERT	82.0	84.3	86.7	79.6	69.5	68.9	32.7	37.9	36.2	36.5	48.6	46.7
YaTC	88.7	88.4	94.9	94.8	72.8	74.4	68.8	68.0	69.5	69.3	48.1	45.1
TrafficFormer	79.6	71.6	86.7	83.2	69.5	69.1	33.0	37.2	35.2	34.8	47.4	42.8
NetFound	59.3	47.2	73.6	69.0	24.8	21.3	23.2	16.8	35.4	32.1	48.2	44.9
Proposed	**90.0**	**79.8**	**95.6**	**92.6**	**95.0**	**93.8**	**79.2**	**70.8**	**80.3**	**73.3**	**82.1**	**72.3**

표 2. 비교 모델과의 실험(L7만 사용)

Method	ISCX VPN 2016		ISCX TOR 2016		CSTNET TLS 1.3		Private#3		Private#2		Private#21	
	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1
ET-BERT	76.1	64.8	85.1	80.1	81.3	78.9	71.5	74.4	61.8	61.3	63.1	59.3
YaTC	90.5	79.3	97.6	90.3	89.7	88.5	89.1	85.0	79.1	70.3	85.5	72.3
TrafficFormer	76.9	66.1	88.2	71.7	81.5	77.6	72.4	72.1	60.6	57.3	64.1	60.8
NetFound	75.2	70.9	86.7	83.0	76.9	75.7	70.1	69.4	57.2	54.1	71.2	69.2
Proposed	**93.6**	**79.8**	**98.3**	**90.8**	**94.2**	**92.9**	**92.1**	**89.9**	78.3	70.2	**86.1**	71.3

표 3. 비교 모델과의 실험(IP와 Port 마스킹 후 L3,L4 사용)

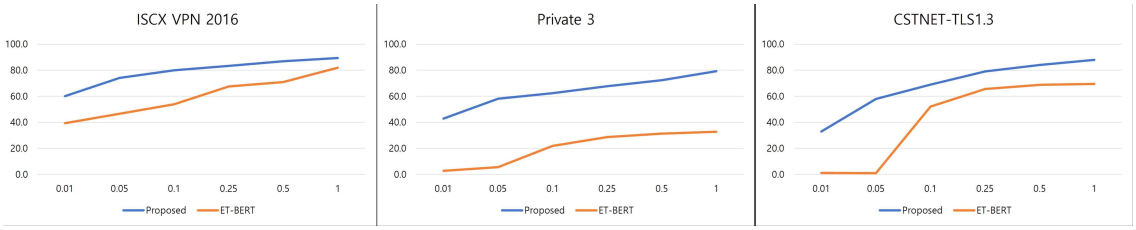

그림 2. CSTNET-TLS1.3에 대한 Few shot 성능

Config	VPN		TLS1.3		VPN			TLS		
	AC	F1	AC	F1	L	P	S	L	P	S
L	91.1	80.3	92.7	91.3	1.00	0.00	0.00	1.00	0.00	0.00
P	90.2	79.9	94.8	93.5	0.00	1.00	0.00	0.00	1.00	0.00
S	91.1	77.2	91.9	89.7	0.00	0.00	1.00	0.00	0.00	1.00
L+P	90.2	78.0	95.1	92.9	0.28	0.71	0.00	0.01	0.99	0.00
L+S	90.0	75.2	94.9	93.6	0.59	0.00	0.41	0.99	0.00	0.01
P+S	90.9	78.8	94.9	93.5	0.00	0.85	0.15	0.00	0.99	0.01
Mean	86.7	77.8	87.0	85.4	-	-	-	-	-	-
All(Proposed)	**90.0**	**79.8**	**95.0**	**93.6**	**0.24**	**0.22**	**0.54**	**0.01**	**0.98**	**0.01**

표 4. 파인튜닝 전략 별 각 계층의 웨이트 값 변화

Ⅲ. 실험

3.1 실험 설정

총 6개 데이터셋에서 평가를 수행하였으며, 자세한 정보는 표 1과 같다. 사전학습은 가장 규모가 크고 다양한 Private-3에서만 수행하며(30 에폭), 이후 각 데이터셋에 파인튜닝(50 에폭)한다. 비교 기준 모델은 ET-BERT[5], YaTC[6], TrafficFormer[7], NetFound[8]이다.

3.2 전체 성능 비교

표 2는 L7 정보만을 사용한 실험 결과이고, 표 3은 IP 주소와 Port를 마스킹한 뒤 L3, L4 정보를 함께 사용한 실험 결과이다. 표 2에서 제안 방법은 모든 데이터셋에서 일관되게 최고 정확도를 달성하였다. VPN 2016에서 정확도(90.0%)에 비해 F1(79.8%)이 다소 낮은 것은 클래스 불균형 혹은 소수 클래스에서의 분류 오류에 기인하며, F1은 정밀도와 재현율의 조화 평균으로 소수 클래스 오분류에 더 민감하기 때문이다. 특히 페이로드의 99% 이상이 암호화된 CSTNET - TLS 1.3에서 95.0% 정확도, 93.8% F1을 달성하였다는 점은 모델이 원시 바이트 내용에 의존하지 않고 구조적 특성만으로도 높은 분류 성능을 유지함을 입증한다. 또한 동일한 48개 애플리케이션을 서로 다른 환경에서 수집한 Private-1과 Private-2에서 기준 모델들은 최대 약 10%p의 정확도 차이를 보인 반면, 제안 방법은 두 데이터셋 간 비교적 일관된 성능을 유지하여 환경 변화에 대한 강인성을 보였다.

표 3에서 L3/L4 헤더 정보를 추가로 활용하면 전반적으로 성능이 향상된다. 기준 모델 중에서는 YaTC가 패킷 경계와 순차 구조를 명시적으로 모델링하는 특성 덕분에 전반적으로 강한 성능을 보였다. 제안 방법은 VPN 2016, Tor 2016, CSTNET - TLS 1.3, Private-3에서 최고 성능을 달성하였다. 다만 IP·Port를 마스킹한 후에도 L3/L4 헤더에는 프로토콜 구성, 플로우 동적 특성, 네트워크별 고유 패턴 등의 잔류 신호가 남아 있어 수집 환경에 민감하게 반응하는 경향이 있다. 이로 인해 Private-2에서는 YaTC가 소폭 우수한 결과를 보였으며, 제안 방법도 Private-1과 Private-2 간에 성능 차이가 발생하였다. 이는 L3/L4 특성의 환경 의존성에서 비롯된 것으로, 제안 모델의 구조적 표현 능력 자체의 한계라기보다는 하위 레이어 피처의 고유한 한계로 해석된다. 전반적으로 L7만 사용할 때는 제안 방법이 안정적이고 일관된 성능을 보인 반면, L3/L4를 추가하

면 일부 데이터셋에서 성능이 향상되지만 환경 변화에 대한 민감도도 함께 증가하는 트레이드오프가 존재한다.

3.3 사전학습 성능 검증

레이블 데이터를 1%만 사용하는 극소량 데이터 환경에서 제안 방법은 모든 데이터셋에서 ET-BERT를 일관되게 능가하였으며, 레이블이 적을수록 격차가 더욱 두드러졌다(그림 2). CSTNET - TLS 1.3에서 레이블 1% 조건 기준으로 제안 방법이 33.0%를 달성한 반면 ET-BERT는 1.1%에 불과하였고, Private-3에서도 각각 42.8% 대 2.8%의 큰 차이를 보였다. 이는 계층적 구조 사전학습이 강력한 귀납적 편향을 제공하여 레이블 데이터가 극히 제한된 환경에서도 효율적인 학습을 가능하게 함을 입증한다.

3.4 계층별 중요도 기반 미세조정 분석

표 4는 서로 다른 풀링 구성과 각 계층의 게이팅 가중치 변화를 보여준다. 단일 수준만 사용할 경우 데이터셋에 따라 성능이 달라진다. TLS에서는 패킷 수준(P) 표현이 정확도 94.8%, F1 93.5로 가장 높은 성능을 보인 반면, VPN에서는 세션 수준(S)이 경쟁력 있는 정확도(91.1%)를 달성하였다. 이는 어떤 단일 계층도 모든 데이터셋에서 일관되게 우수하지 않음을 보여준다. 복수 수준을 결합하면 전반적으로 더 안정적인 성능을 얻을 수 있으나, 단순 평균 풀링(Mean)은 오히려 성능을 크게 저하시킨다(VPN 86.7%, TLS 87.0%). 이는 모든 표현에 동일한 가중치를 부여하면 각 계층의 상대적 중요도를 무시하게 되어 정보 신호가 희석될 수 있음을 의미한다. 학습된 게이팅 가중치를 살펴보면, VPN에서는 세션 수준(S=0.54)이 가장 높은 가중치를 받는 반면 TLS에서는 패킷 수준(P≈0.98)이 지배적으로 선택된다. 이는 모델이 입력 특성에 따라 각 계층의 기여도를 자동으로 조절함을 보여주며, 제안 방법의 적응적 다중 수준 융합이 어떠한 단일 수준 풀링보다도 안정적으로 최고 성능(VPN 90.0% / TLS 95.0%)을 달성하는 근거가 된다.

Ⅲ. 결론

본 논문은 암호화 트래픽 분류를 위한 구조 중심 표현 학습 프레임워크를 제안하였다. 핵심 전제는 암호화가 바이트 내용을 무작위화하더라도, 프로토콜이 정의한 필드 경계, 레이어 스택, 패킷 순서 등의 계층적 구조 정보는 암호화에 무관하게 보존된다는 것이다. 본 논문은 이 구조적 속성을 학습의 중심 대상으로 삼아, 이를 기반으로 세 가지 핵심 기여를 제시하였다.

첫째, 프로토콜 경계 정렬 구조 토큰화를 통해 네트워크 트래픽을 필드-레이어-패킷-세션의 4단계 계층으로 명시적으로 표현하였다. 각 필드 토큰은 평균/최대 풀링 기반의 순서 불변 통계량으로 바이트 분포를 요약하며, 필드 길이 임베딩을 통해 암호화 환경에서도 의미 있는 구조적 신호를 제공한다.

둘째, 레이블 없이 구조적 관계를 학습하기 위해 계층적 셔플 사전학습을 제안하였다. 기존의 마스크 토큰 복원 방식과 달리, 필드(FOP)·레이어(LOP)·패킷(POP) 수준에서 무작위로 섞인 순서를 복원하는 태스크를 통해 프로토콜이 정의한 계층 간 순서 관계와 포함 관계를 명시적으로 학습한다. 소량 데이터 실험에서 제안 방법은 레이블 1% 조건에서 ET-BERT 대비 CSTNET - TLS 1.3 기준 33.0% 대 1.1%로 큰 성능 차이를 보이며, 구조적 귀납적 편향의 효과를 입증하였다.

셋째, 계층별 중요도 기반 미세조정을 통해 레이어·패킷·세션 수준의 표현을 학습 가능한 게이팅 가중치로 동적 결합하였다. 실험 결과, VPN 트래픽에서는 세션 수준(가중치 0.54)이, 암호화된 TLS 트래픽에서는 패킷 수준(가중치 0.98)이 지배적으로 선택되었다. 이는 데이터셋 특성에 따라 판별력 있는 구조 수준이 다르다는 것을 의미하며, 단일 수준 고정 풀링이나 단순 평균 풀링 대비 적응적 다중 수준 융합이 일관되게 우수함을 확인하였다.

종합적으로, 제안 방법은 6개의 다양한 데이터셋에서 기존 대표 모델들을 모두 능가하였으며, 특히 암호화 비율이 높은 환경과 소량 레이블 조건에서 강점이 두드러졌다. 이는 프로토콜이 정의한 구조적 규칙이 바이트 기반 표현에 비해 암호화 불변적이고 일반화 능력이 높은 학습 신호임을 실증적으로 뒷받침한다. 향후 연구로는 사전학습 환경과 다른 네트워크 환경에서의 교차 일반화, 그리고 새로운 TLS 버전 출현이나 신규 애플리케이션 프로토콜 등장과 같은 동적 프로토콜 변화에 대한 적응 전략을 탐구할 계획이다.

ACKNOWLEDGMENT

Put sponsor acknowledgments.

참 고 문 헌

[1] E. Papadogiannaki and S. Ioannidis, “A survey on encrypted network traffic analysis applications, techniques, and countermeasures,” ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1 - 35, 2021.

[2] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, “End-to-end encrypted traffic classification with one-dimensional convolution neural networks,” Proc. IEEE Int. Conf. Intelligence and Security Informatics (ISI), Beijing, China, 2017.

[3] Z. Wang, Q. Zeng, Y. Liu, and P. Li, “Malware traffic classification using convolutional neural network for representation learning,” Proc. Int. Conf. Information Networking (ICOIN), Da Nang, Vietnam, Jan. 11 - 13, 2017, pp. 712 - 717.

[4] Y. Yang, K. Kang, L. Jiang, Z. Gao, Y. Guo, J. Zhang, and J. Deng, “HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,” IEEE Access, vol. 5, pp. 26954 - 26964, 2017.

[5] 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,” Proceedings of the ACM Web Conference, pp. 633 - 642, 2022.

[6] R. Zhao, M. Zhan, X. Deng, Y. Wang, Y. Wang, G. Gui, and Z. Xue, “Yet another traffic classifier: A masked autoencoder based traffictransformer with multi-level flow representation,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 4, pp. 5420 - 5428, 2023.

[7] S. Guthula, R. Beltiukov, N. Battula, W. Guo and A. Gupta, “NetFound: Foundation model for network security,” arXiv preprint arXiv:2310.17025, 2023.

[8] C. Lin, W. Zhang, H. Luo, X. Meng, and Y. Zhang, “Nethira: A heterogeneity-aware hierarchical pre-trained model for network traffic classification,” arXiv preprint arXiv:2601.22494, 2026.

[9] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, “Characterization of encrypted and VPN traffic using time-related features,” Proc. 2nd Int. Conf. Information Systems Security and Privacy (ICISSP), Rome, Italy, Feb. 19 - 21, 2016, pp. 407 - 414.

[10] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, “Characterization of Tor traffic using time-based features,” Proc. 3rd Int. Conf. Information Systems Security and Privacy (ICISSP), Porto, Portugal, Feb. 19 - 21, 2017, pp. 253 - 262.