

Charted-Input CNN을 활용한 응용 레벨 트래픽 분류

2022270607 이윤서 X 2022270639 김수안

BACKGROUND

네트워크 트래픽?

컴퓨터 네트워크를 통해 송수신되는 데이터의 흐름
사용자 간 통신, 서버 요청 및 응답, 애플리케이션 실행 등
다양한 네트워크 활동으로 발생

트래픽 분류?

패킷의 특징이나 흐름(flow) 정보를 기반으로 트래픽
유형이나 소스를 식별하는 기술

* 유형 : (예: 웹 브라우징, 스트리밍, 파일 전송)

* 소스 : (예: 특정 애플리케이션, 프로토콜)

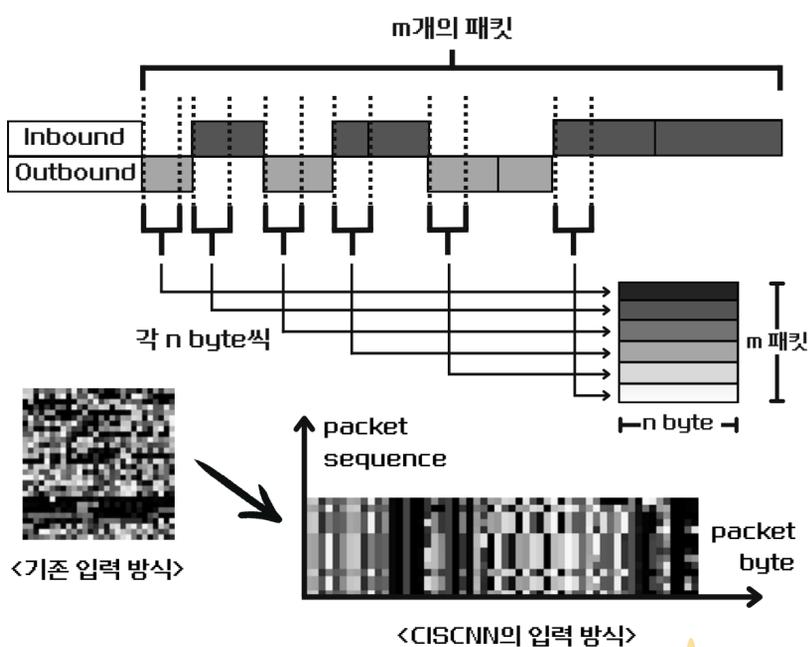
트래픽 분류의 필요성?

네트워크 자원 관리, QoS(Quality of Service),
보안 위협 탐지 및 대응, 대역폭 최적화, 비용 절감

CONCEPT & BUILD

Charted-input Shape CNN (CISCNN)?

차트 형태의 입력 형식을 가진 CNN 모델
Session 별 초기 m개 패킷에서 각 n byte씩 사용



기대 효과 및 구현 방식

트래픽을 2D 형태로 변환하는 새로운 관점 제시
값 자체가 가지는 정적 특징 뿐 아니라, 값이 변화하는
동적 패턴을 학습에 포함 >> 성능 향상 및 경량화 가능성

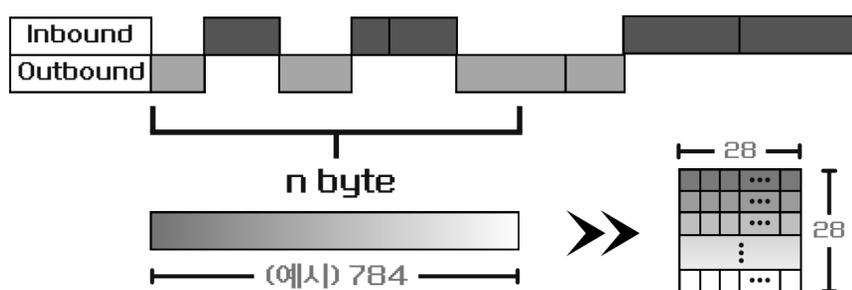
PyTorch Pytorch로 구현,
컨테이너로 구동

* 이식성 확보를 위해 데이터셋/모델을 라이브러리로 관리

CURRENT LIMITS

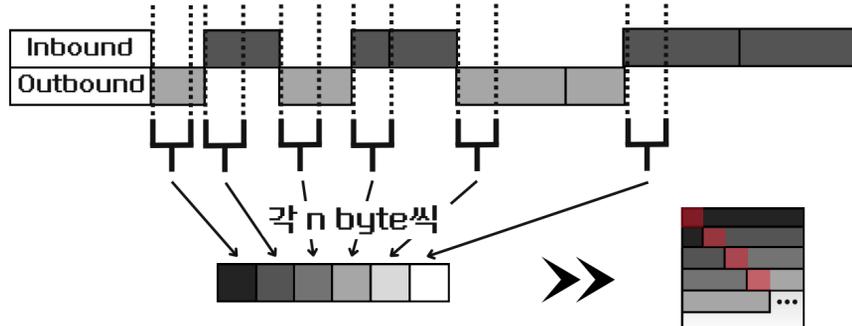
CNN을 활용한 트래픽 분류의 기존 방식

1 Session 별 초기 n byte 사용
(예시) n = 600, 784, 1500, ...



한계점? (1) 세션 내의 많은 패킷들 중 협상과정만 포함
그 이후 패킷들의 정보, 특징, 패턴이 무시
(2) 세션의 시간적 흐름/맥락의 미포함

2 Session 별 초기 m개 패킷에서 각 n byte씩 사용
(예시) (m,n) = (6, 100), (5, 100), (10,50), ...



한계점? 각 패킷의 경계 구분이 모호 * : 각 패킷의 시작점
특정 패킷 필드의 변화 패턴을 학습하기 어려운 입력 형태

패킷의 각 필드나 데이터의 변화 등, 시계열적인

“동적 패턴” 을 학습에 포함한다면?

PROGRESS

실험 결과

기존 입력 방식과 비교했을 때의 Accuracy 변화?

사실 데이터셋 1 : 14*56에서 7.99% 향상

사실 데이터셋 2 : 8*98에서 4.95% 향상

공개 데이터셋 (ISCX/VPN 2016) : 4*196에서 2.99% 향상

추후 계획

다양한 입력 크기 적용(MISCNN), 임베딩을 통한 순서 개념
추가 적용 등 향후 연구 필요