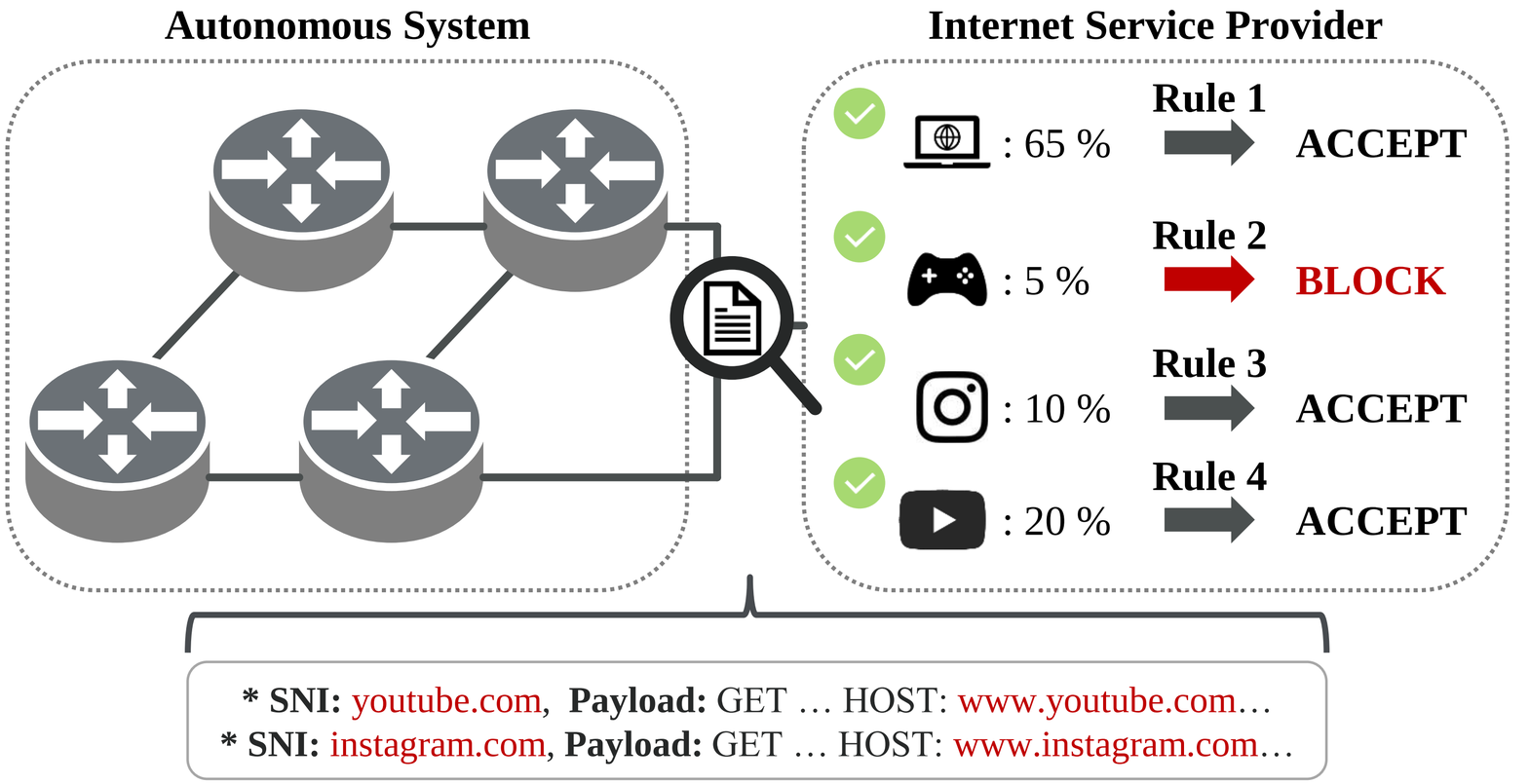
프로젝트 소개

2024.10.30.(목)

국민대학교 정보보호연구실

배경 (전통적 앱 식별 기술과 그 활용)

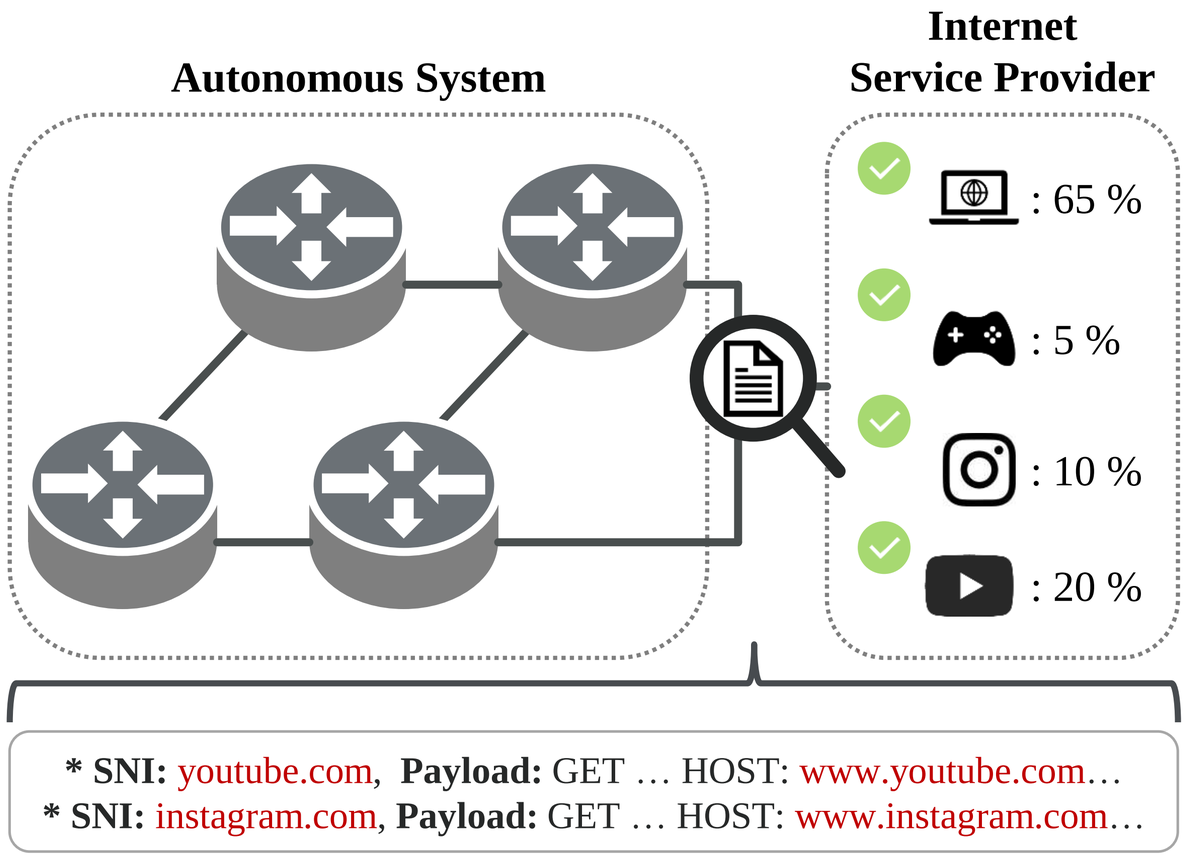
ISP(Internet Service Provider) 등은 전통적으로 **패킷 페이로드 분석 및 SNI 기반 앱 식별 기술**을 통해 QoS 관리 및 특정 앱 차단 등을 수행.



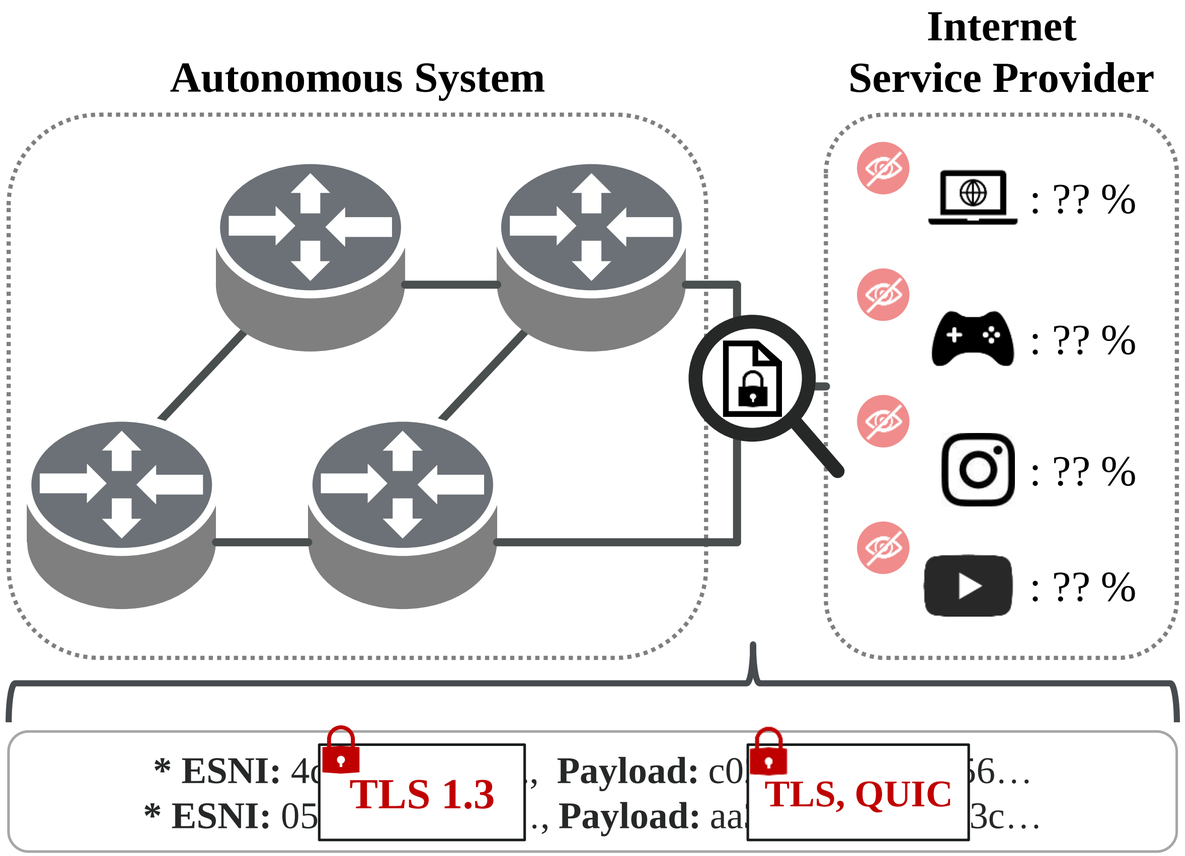
Application Classification on Unencrypted Traffic

배경 (기존 앱 식별 기술의 무력화)

암호화 통신(TLS, QUIC, VNP 등) 발달에 의해 기존 방식의 **앱 식별 기술 무력화.**



(a) Application Classification on Unencrypted Traffic



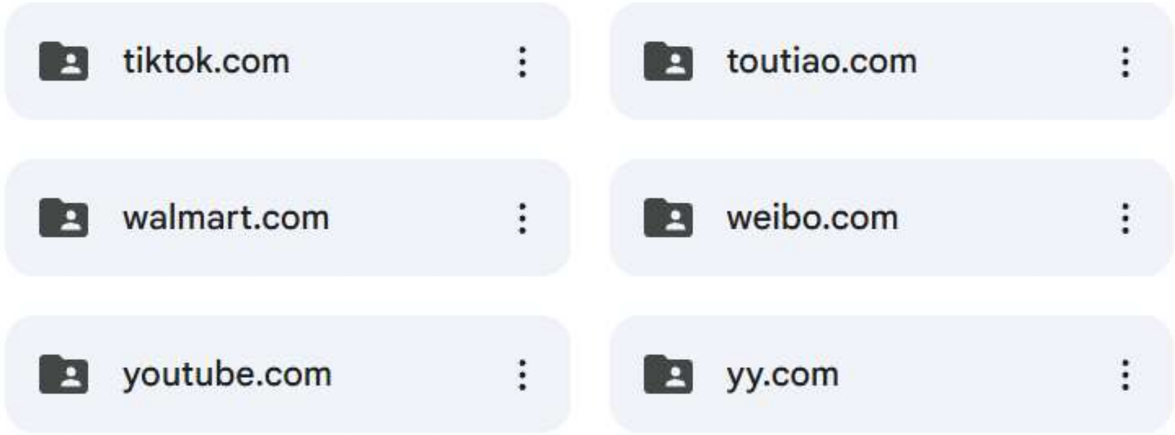
(b) Application Classification on Encrypted Traffic

개요

- 목표: 암호화된 네트워크 트래픽에서 앱을 식별하는 모델 개발
- 데이터: CSTNET-TLS 1.3
- 인원
  - 최대 3인
- 구성
  - 과제 1: 데이터 전처리 및 feature 추출
  - 과제 2: 모델 학습 및 성능 평가

데이터셋

- 공개 데이터셋인 CSTNET
- .pcap 형식
- 특징
  - 암호화된 트래픽 (TLS)
  - 라벨 총 120개



과제 1

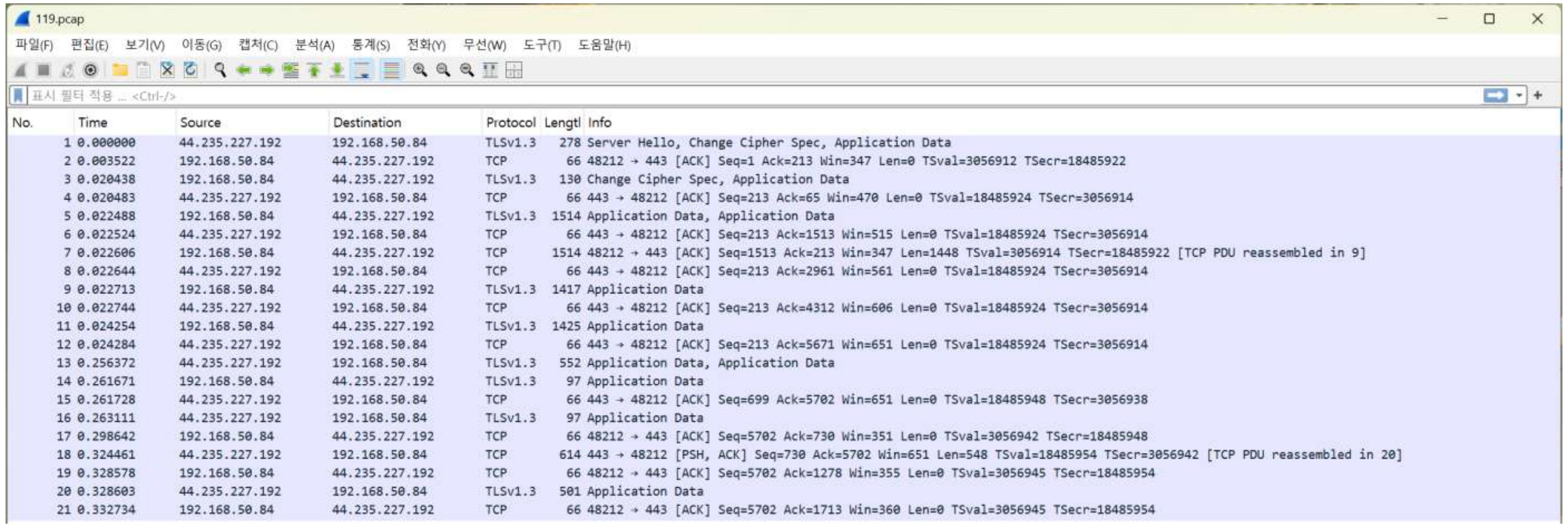
- 목표: Raw pcap 파일에서 flow별 feature 추출

- 기준

  1. Flow 식별 (5-tuple 기준)
  2. 플로우 별 **패킷 길이 시퀀스** 추출
  3. 플로우 개수 400개 미만 라벨 삭제
  4. 결과 저장: CSV (flow-level features)

Src IP	Src Port	Dst IP	Dst Port	Protocol	Label	Pkt Length Sequence
8.6.0.1	0	8.0.6.4	0	0	benign	0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0
192.168.1	123	192.168.1	123	17	benign	48:48:-48:-48:48:48:-48:-48:48:48:-48:-48
192.168.1	5353	224.0.0.25	5353	17	benign	95:95:95:95:95:95:95:95:95:95:95:95:95:95:95:9
192.168.1	123	152.2.133.	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-
192.168.1	123	45.79.11.2	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-
192.168.1	123	91.189.89.	123	17	benign	48:-48:48:-48:48:-48
192.168.1	123	199.199.2	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-
192.168.1	123	204.11.20	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-
192.168.1	123	192.168.1	123	17	benign	48:48:-48:-48:48:48:-48:-48:48:48:-48:-48
192.168.1	123	108.61.56.	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-
192.168.1	123	96.126.10	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-

과제 1



119.pcap

파일(F) 편집(E) 보기(V) 이동(G) 캡처(C) 분석(A) 통계(S) 전화(Y) 무선(W) 도구(T) 도움말(H)

표시 필터 적용 ... <Ctrl-/>

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	44.235.227.192	192.168.50.84	TLSv1.3	278	Server Hello, Change Cipher Spec, Application Data
2	0.003522	192.168.50.84	44.235.227.192	TCP	66	48212 → 443 [ACK] Seq=1 Ack=213 Win=347 Len=0 TSval=3056912 TSecr=18485922
3	0.020438	192.168.50.84	44.235.227.192	TLSv1.3	130	Change Cipher Spec, Application Data
4	0.020483	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=213 Ack=65 Win=470 Len=0 TSval=18485924 TSecr=3056914
5	0.022488	192.168.50.84	44.235.227.192	TLSv1.3	1514	Application Data, Application Data
6	0.022524	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=213 Ack=1513 Win=515 Len=0 TSval=18485924 TSecr=3056914
7	0.022606	192.168.50.84	44.235.227.192	TCP	1514	48212 → 443 [ACK] Seq=1513 Ack=213 Win=347 Len=1448 TSval=3056914 TSecr=18485922 [TCP PDU reassembled in 9]
8	0.022644	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=213 Ack=2961 Win=561 Len=0 TSval=18485924 TSecr=3056914
9	0.022713	192.168.50.84	44.235.227.192	TLSv1.3	1417	Application Data
10	0.022744	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=213 Ack=4312 Win=606 Len=0 TSval=18485924 TSecr=3056914
11	0.024254	192.168.50.84	44.235.227.192	TLSv1.3	1425	Application Data
12	0.024284	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=213 Ack=5671 Win=651 Len=0 TSval=18485924 TSecr=3056914
13	0.256372	44.235.227.192	192.168.50.84	TLSv1.3	552	Application Data, Application Data
14	0.261671	192.168.50.84	44.235.227.192	TLSv1.3	97	Application Data
15	0.261728	44.235.227.192	192.168.50.84	TCP	66	443 → 48212 [ACK] Seq=699 Ack=5702 Win=651 Len=0 TSval=18485948 TSecr=3056938
16	0.263111	44.235.227.192	192.168.50.84	TLSv1.3	97	Application Data
17	0.298642	192.168.50.84	44.235.227.192	TCP	66	48212 → 443 [ACK] Seq=5702 Ack=730 Win=351 Len=0 TSval=3056942 TSecr=18485948
18	0.324461	44.235.227.192	192.168.50.84	TCP	614	443 → 48212 [PSH, ACK] Seq=730 Ack=5702 Win=651 Len=548 TSval=18485954 TSecr=3056942 [TCP PDU reassembled in 20]
19	0.328578	192.168.50.84	44.235.227.192	TCP	66	48212 → 443 [ACK] Seq=5702 Ack=1278 Win=355 Len=0 TSval=3056945 TSecr=18485954
20	0.328603	44.235.227.192	192.168.50.84	TLSv1.3	501	Application Data
21	0.332734	192.168.50.84	44.235.227.192	TCP	66	48212 → 443 [ACK] Seq=5702 Ack=1713 Win=360 Len=0 TSval=3056945 TSecr=18485954

tshark, pyshark, scapy 등

Src IP	Src Port	Dst IP	Dst Port	Protocol	Label	Pkt Length Sequence
8.6.0.1	0	8.0.6.4	0	0	benign	0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0:0
192.168.10	123	192.168.10	123	17	benign	48:48:-48:-48:48:48:-48:-48:48:48:-48:-48
192.168.10	5353	224.0.0.25	5353	17	benign	95:95:95:95:95:95:95:95:95:95:95:95:95:95:95:9
192.168.10	123	152.2.133.	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-4
192.168.10	123	45.79.11.2	123	17	benign	48:-48:48:-48:48:-48:48:-48:48:-48:48:-48:48:-4
192.168.10	123	91.189.89.	123	17	benign	48:-48:48:-48:48:-48

과제 1

- 결과물
  - flow_features.csv
  - 전처리 코드 (.ipynb or .py)

과제 2

- 목표: Flow-level feature를 이용한 앱 분류 모델 학습
- 데이터: 과제 1에서 생성한 flow 데이터
- 추가 feature 생성 가능
  - 패킷 수, 평균/최대 패킷 크기, 평균 inter-arrival time 등
- 출력: 모델 F1-score

과제 2

- 추천 모델
  - 트리 계열, Transformer 계열 등등
- 평가 지표
  - Macro F1-score

과제 2

- 결과물
  - 학습 코드 (train_model.py / .ipynb)
  - 보고서 (모델 설명 + 성능표)

참고 자료

- Lin, Xinjie, et al. "ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification." *Proceedings of the ACM Web Conference 2022*. ACM, 2022.
- Liu, Chang, et al. "Fs-net: A flow sequence network for encrypted traffic classification." *IEEE INFOCOM 2019-IEEE Conference On Computer Communications*. IEEE, 2019.
- Xie, Renjie, et al. "Rosetta: Enabling robust tls encrypted traffic classification in diverse network environments with tcp-aware traffic augmentation." *Proceedings of the ACM turing award celebration conference-China 2023*. 2023.

추가 문의 사항

- 2025 정보보호와 시스템보안 슬랙 채널에서 질의

감사합니다!

https://infosec.kookmin.ac.kr