소개영상 : https://www.youtube.com/watch?v=wjlrIJas8TQ

O24Sec

(Object Oriented Clustering for Security Monitoring)

부제 : 객체 중심 보안관제 오탐제거

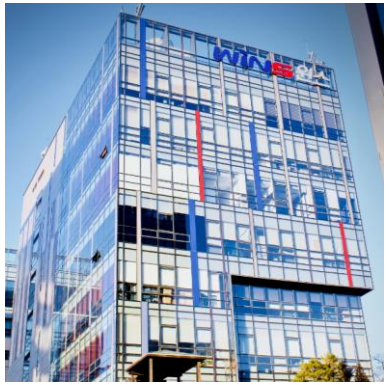
팀 (18조) : 맬러리를 찾아서

팀 소개 (맬러리를 찾아서)

김민송

장우혁

산학 담당
윤명근 교수님



㈜윈스
최병환 팀장님

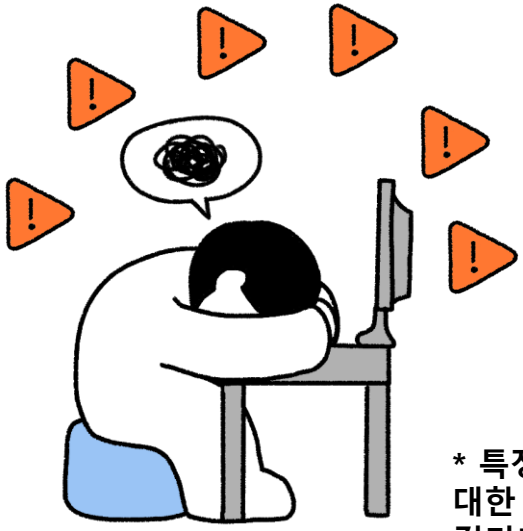
Contents

산학 기관 요구 사항

자사 정보보안 솔루션 : “**SNIPER BD1 오탐(False Positive)** 제거”

〈Sniper IPS〉

〈Sniper BD1〉



〈Alert fatigue〉

*** 특정 경우에는 알람 1개에 대한 분석에 10분~40분까지 걸리기도 함**

제안 기술 (배경)

데이터 전처리의 중요성

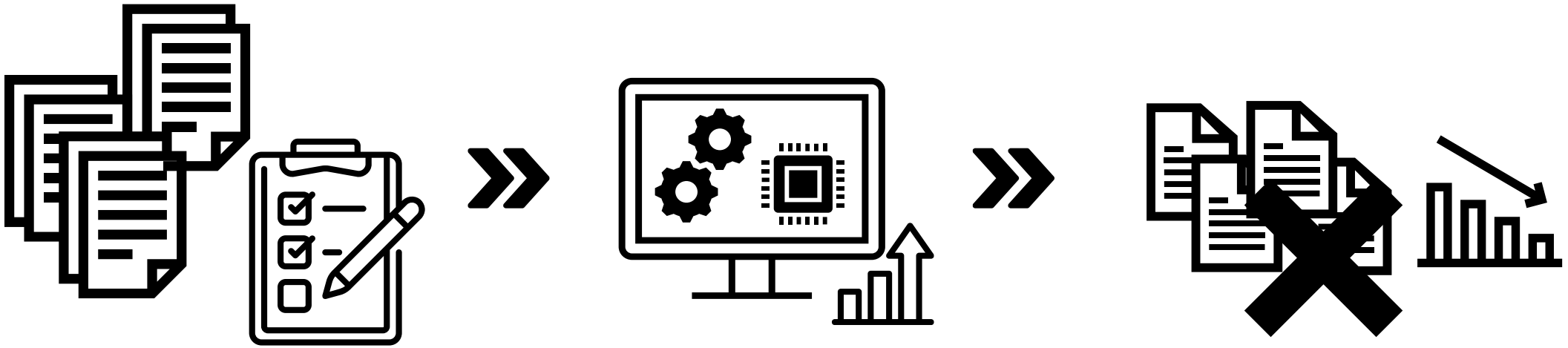
〈“Garbage in Garbage out”〉

모델	정확도	정밀도	재현율	F1-score
정합 전	98.02%	93.76%	59.6%	72.8%
정합 후	98.23%	96%	95.42%	94.3%

〈“데이터 정합 후 성능 향상” , NetSec2020 송중석 박사〉

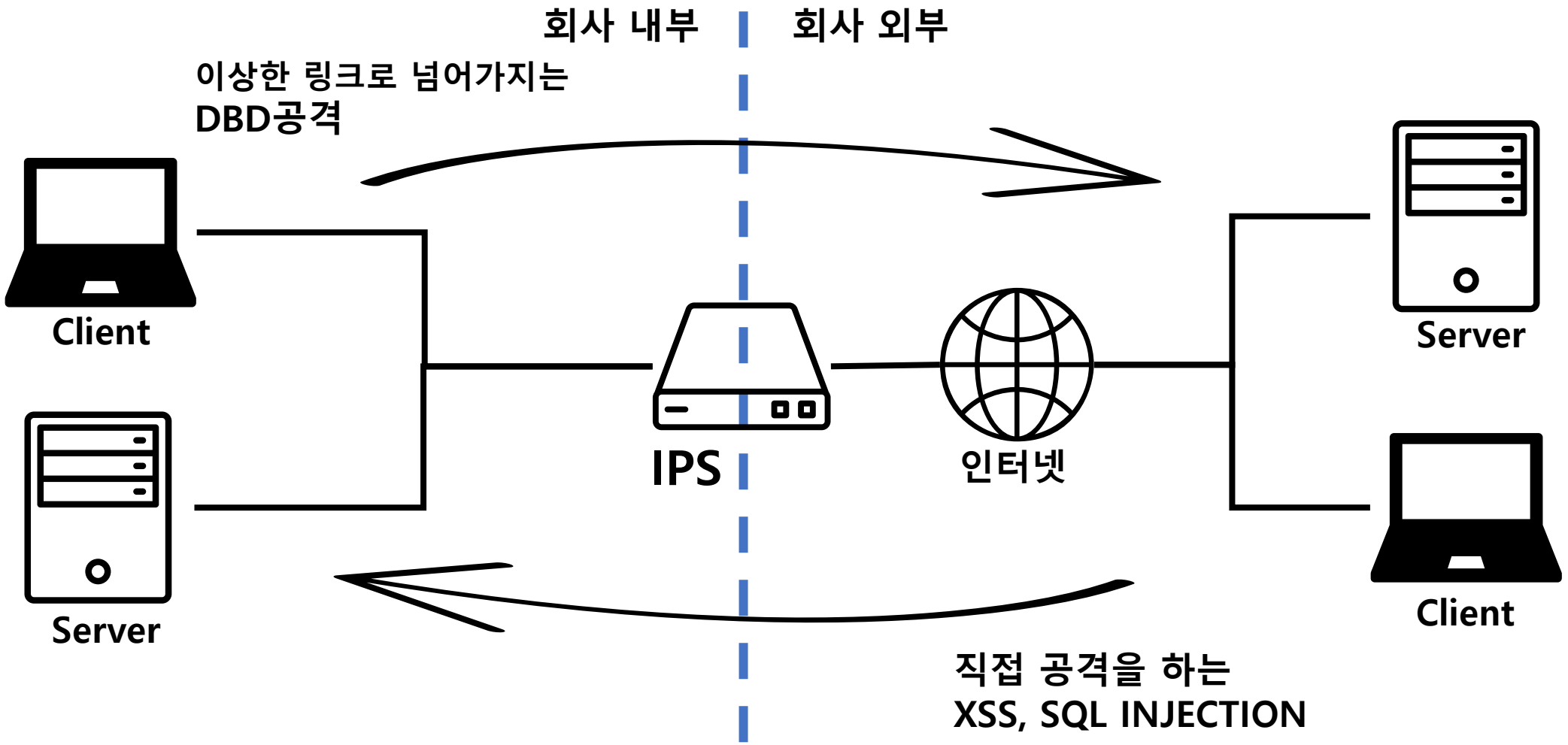
제안 기술 (배경)

"**인공지능 학습데이터 셋에 대한 라벨의 정확도**를 높여 모델의 오탐률을 낮추는 것"

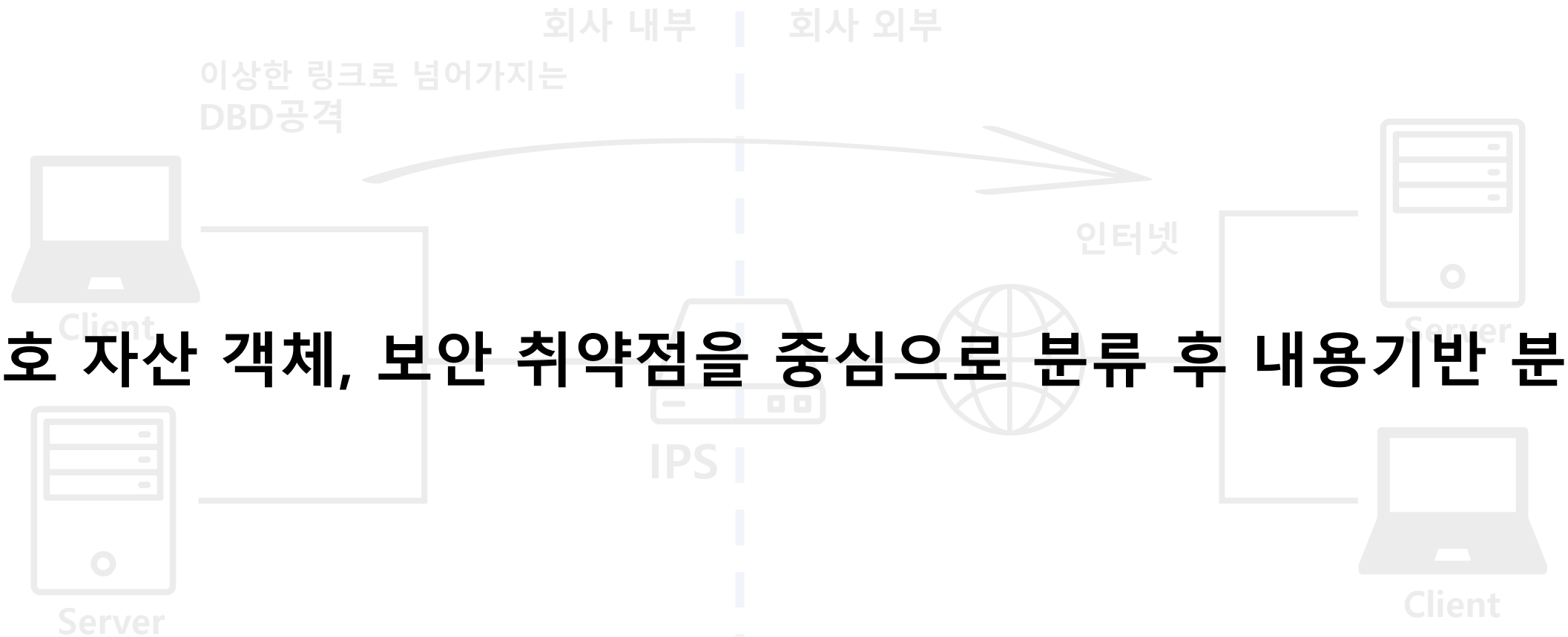


재라벨링 **모델 성능 향상** **오탐 제거**

제안 기술 (배경 : 네트워크 흐름)

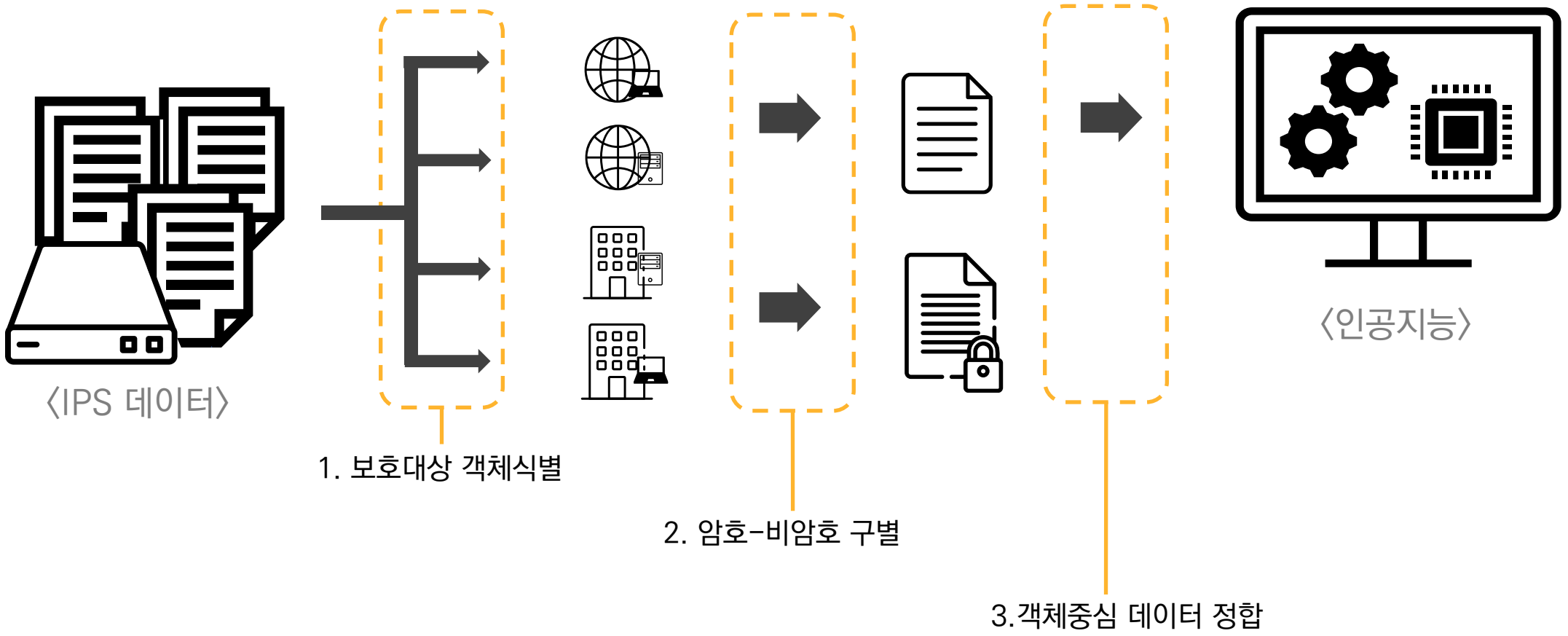


제안 기술 (배경 : 네트워크 흐름)



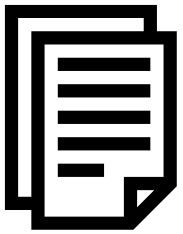
보호 자산 객체, 보안 취약점을 중심으로 분류 후 내용기반 분석

제안 기술 (Object Oriented Clustering for Security Monitoring)



프로젝트 개발 기술 (1. 보호대상 객체식별)

- 사용데이터: IPS 이벤트 약 2000만건(1년)



- **Source IP**
- **Destination IP**
- **Source Port**
- **Destination Port**
- **Protocol**

〈5-Tuple〉

From 외부 클라이언트 To 내부 서버

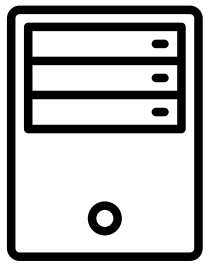
From 외부 서버 To 내부 클라이언트

From 내부 서버 To 외부 클라이언트

From 내부 클라이언트 To 외부 서버

프로젝트 개발 기술 (1. 보호대상 객체식별)

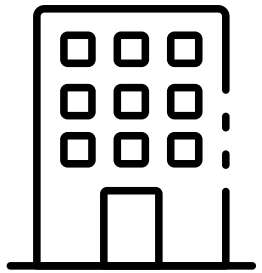
Step1



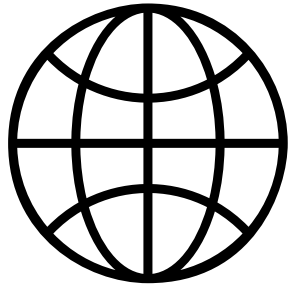
Server

Client

Step2



회사 내부



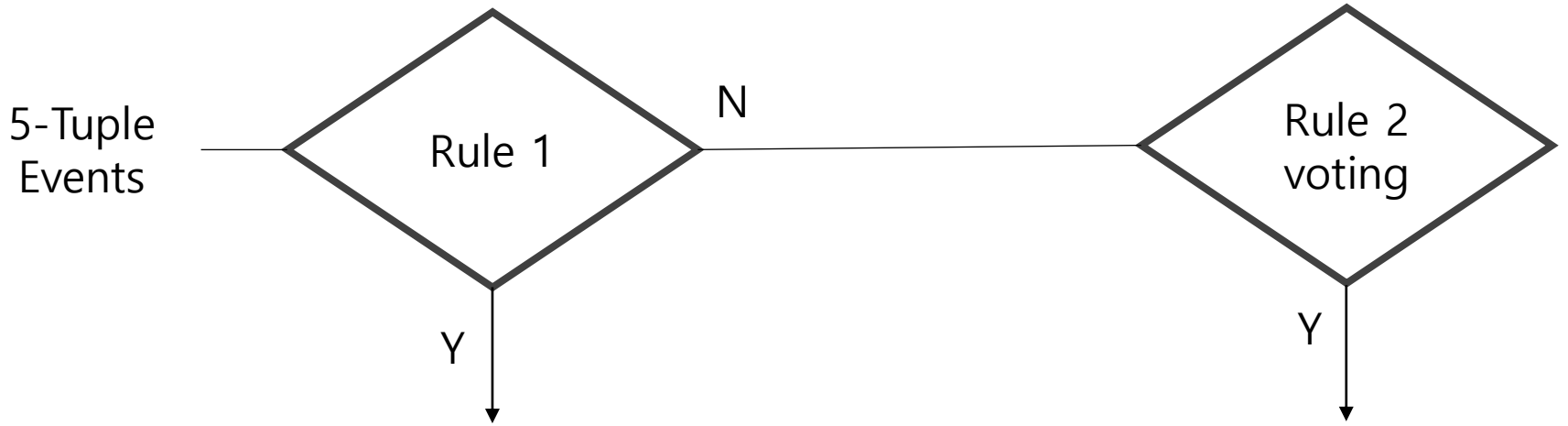
회사 외부

프로젝트 개발 기술 (1. 보호대상 객체식별)

Step1: 서버-클라이언트 분류

1. IP:PORT 카운트가 높은 쪽
2. IP의 PORT 종류가 적은 쪽
3. IP에서 해당 PORT 점유율이 높은 쪽

한 쪽만 1024 미만의 포트를 가짐



1024 미만: 서버
1024 이상: 클라이언트

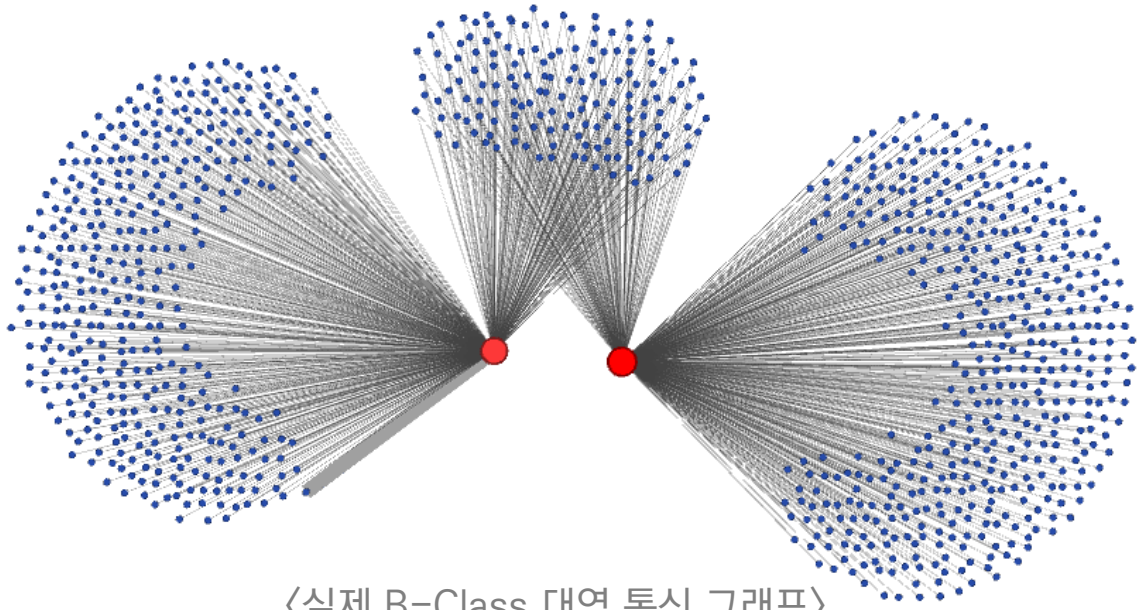
과반수쪽: 서버
반대쪽 : 클라이언트

프로젝트 개발 기술 (1. 보호대상 객체식별)

Step2: 내부-외부 분류

- IP를 B-Class대역으로 묶었을 때 통신한 상대 B-Class대역이 많으면 내부로 판별

B-class 대역 ex) 203.116.X.X, 168.102.X.X



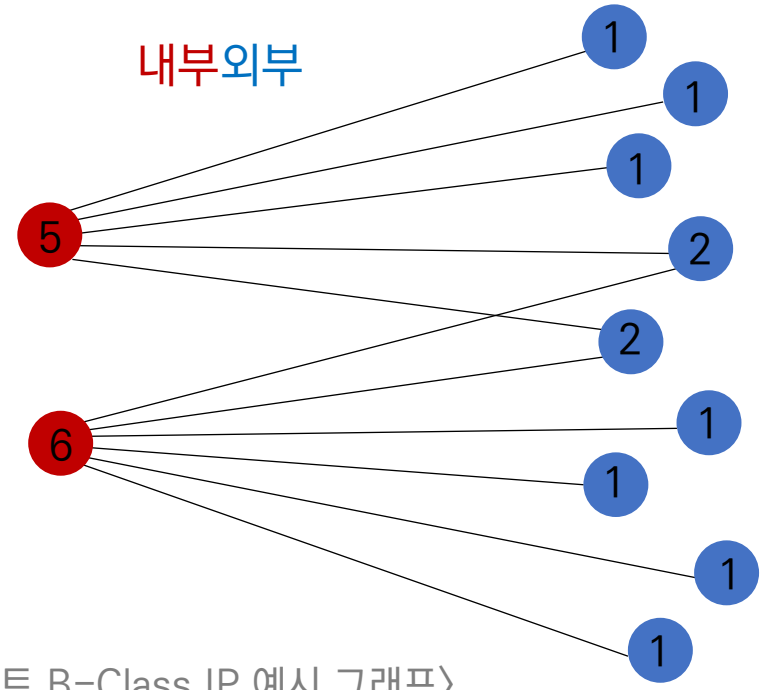
〈실제 B-Class 대역 통신 그래프〉

빨간 노드: 내부
파란 노드: 외부
엣지 : 통신

프로젝트 개발 기술 (1. 보호대상 객체식별)

Step2: 내부-외부 분류

- 통신한 상대 B-Class대역이 많은 순으로 내림차순 정렬 후 임계점까지 내부로 분류



〈IPS이벤트 B-Class IP 예시 그래프〉

내림차순 정렬

B-Class	상대방 종류	내부 개수(가정)
내부1	6	1
내부2	5	2
외부1	2	3
외부2	2	
외부3	1	
외부4	1	
외부5	1	
외부6	1	
외부7	1	
외부8	1	
외부9	1	

외부

내부 개수를 1씩 추가,

상대방 종류가 내부 개수보다 작은 B-Class 부터 외부

프로젝트 개발 기술 (2. 암호화 분리)

Entropy(복잡도)

Payload에 Byte가 얼마나 여러 종류(최대 256가지)로 들어가 있는지

일반적으로 암호화 되었을 경우 Entropy값이 높음
난수화로 인해 여러가지 Byte 값이 쓰이기 때문
Ex) SSH, TLS

<Byte Stream>

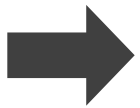
ADDEB79218960CA907EE9907EABCFGDD4
8CE1WGN5T4RGHER1D2FQ5VRNY4NUM

일반적으로 암호화 되지 않았을 경우 Entropy값이 낮음
비슷한 범위의 Byte 값이 계속 쓰이기 때문
Ex) HTTP, FTP

474554202F434F4E54454E542E7068702048
5454502F312E31
("GET /CONTENT.php HTTP/1.1")

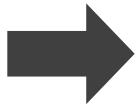
프로젝트 개발 기술 (2. 암호화 분리)

1. Entropy가 높지만 식별 가능한 정보가 담겨져 있는 경우



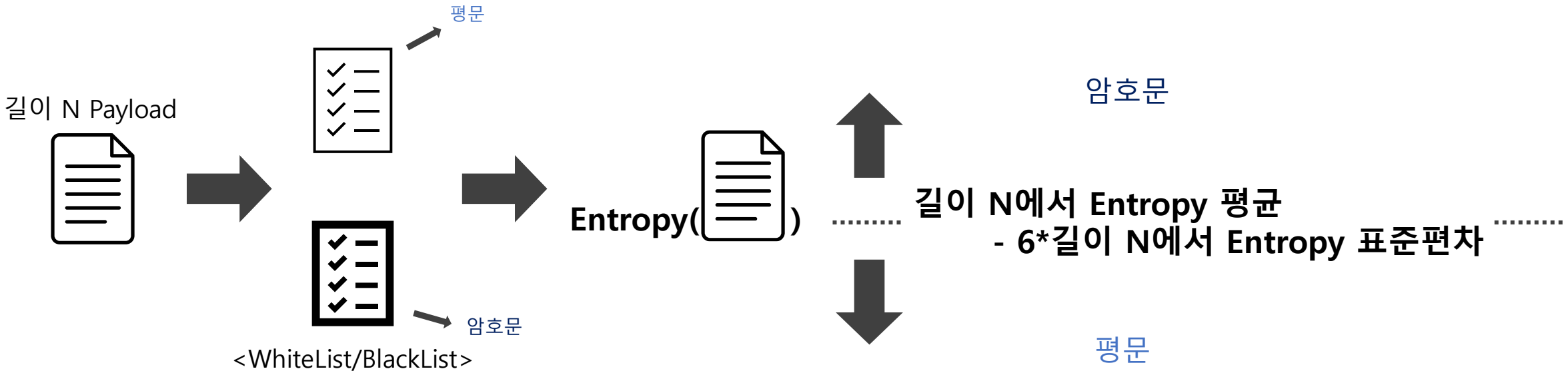
White List

2. Entropy가 낮지만 식별 가능한 정보가 없는 경우

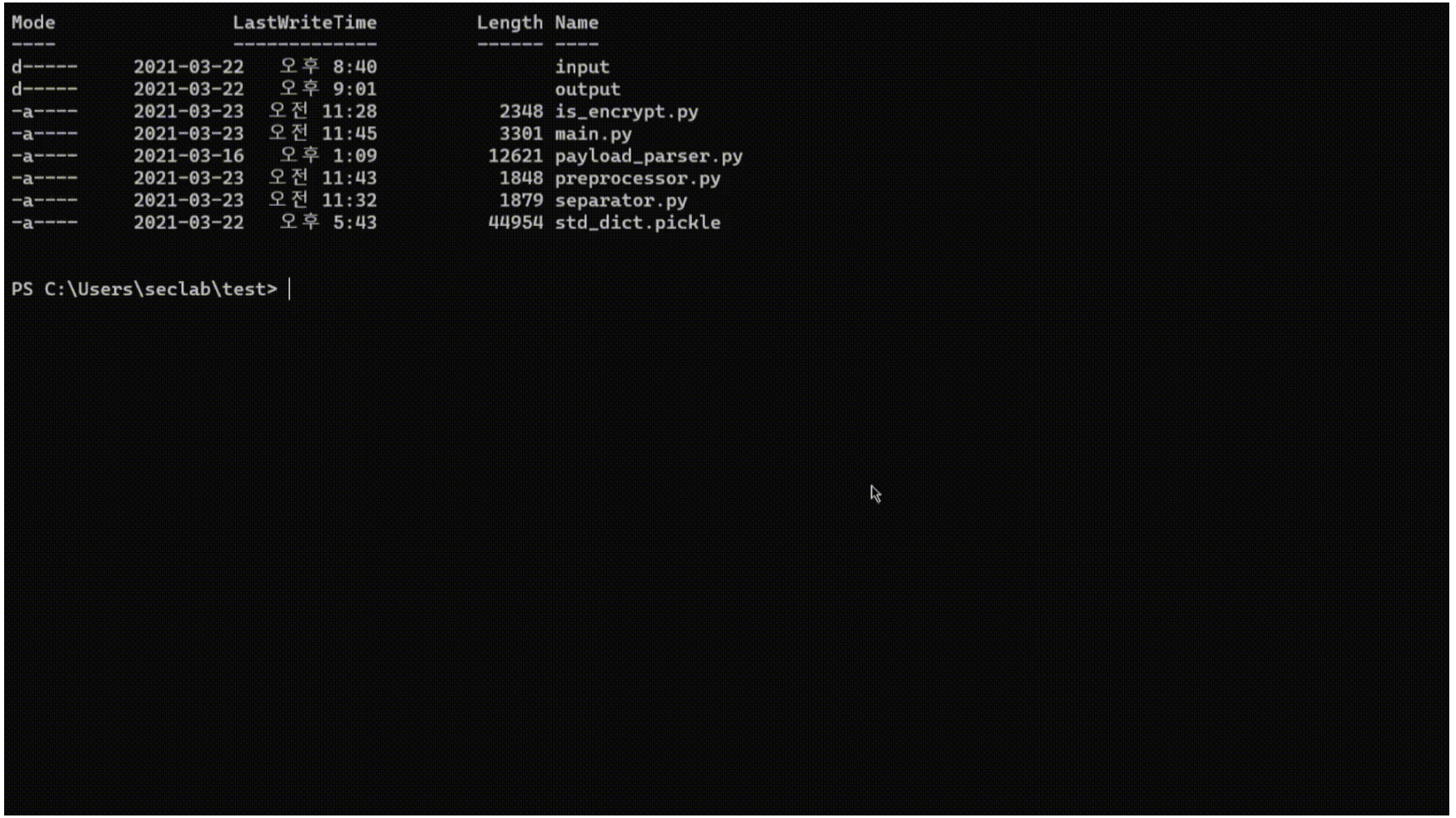


Black List

프로젝트 개발 기술 (2. 암호화 분리)



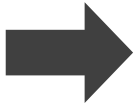
프로젝트 개발 기술 (목표 1, 2 실행 화면)



프로젝트 개발 기술 (3. 유사도 기반 클러스터링)

특수문자와 공백제거를 통한 단어화

POST /sops/setMallData.sops HTTP/1.1₩r₩n
Content-Type: multipart/form-data;boundary=***b*o*u*n*d*a*r*y*****₩r₩n**
Connection: close₩r₩n
User-Agent: Dalvik/2.1.0 (Linux; U; Android 8.0.0; LM-G710N Build/OPR1.170623.032)₩r₩n
Host: parcelapi.enuri.com₩r₩n
Accept-Encoding: gzip₩r₩n
Content-Length: 16228₩r₩n₩r₩n



[POST, sops, setMallData, HTTP, 1, Content, Type, multipart, form, data, boundary, b, o, u, n, d, a, r, y, Connection, close, User, Agent, Dalvik, 2, 0, Linux, U, Android, 8, LM, G710N, Build, OPR1, 170623, 032, Host, parcelapi, enuri, com, Accept, Encoding, gzip, 16228]

프로젝트 개발 기술 (3. 유사도 기반 클러스터링)

자카드 유사도를 통한 군집화 $$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

[POST, sops, setMallData, HTTP, 1, Content, Type, multipart, form, data, boundary, b, o, u, n, d, a, r, y, Connection, close, User, Agent, Dalvik, 2, 0, Linux, U, Android, 8, LM, G710N, Build, OPR1, 170623, 032, Host, parcelapi, enuri, com, Accept, Encoding, gzip, 16228]

[POST, sops, setMallData, HTTP, 1, … , 14422]

[POST, sops, setMallData, HTTP, 1, … , 25673]

[POST, sops, setMallData, HTTP, 1, … , 64323]

[GET, favicon, ico, HTTP, 1, Host, pigbooktv, net, Accept, Language, ko, kr, Connection, keep, alive, Encoding, gzip, deflate, User, Agent, MobileSafari, 604, CFNetwork, 901, Darwin, 17, 6, 0]

[GET, favicon, … , 17, 6, 0]

[GET, favicon, … , 17, 6, 0]

[GET, favicon, … , 17, 6, 0]

프로젝트 개발 기술 (3. 유사도 기반 클러스터링)

차이점을 통해 데이터 정제(사람)

클러스터 내 대표 벡터
[POST, sops, setMallData, HTTP, 1, Content, Type, multipart, form, data, boundary, b, o, u, n, d, a, r, y, Connection, close, User, Agent, Dalvik, 2, 0, Linux, U, Android, 8, LM, G710N, Build, OPR1, 170623, 032, Host, parcelapi, enuri, com, Accept, Encoding, gzip, 16228]

EVENT	기존 LABEL	차이점	교정 된 LABEL
1	0	16228 -> 14422	0
2	1	16228 -> 25673	0
3	0	16228 -> 64323	0
4	0	16228 -> 49859	0

공격적으로 의미 없는 차이

프로젝트 결과 (실제 데이터에서 정합 성공)

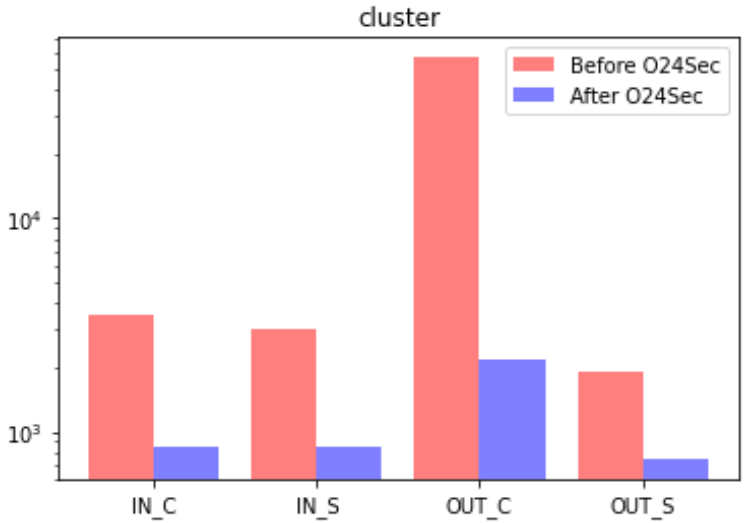
<기술 적용 후 LABEL 상황>

	A	B	C	D	E	F	G	H	I	J
1	Cluster	DetectN	Result	_id	Payload	label_P	DetectN	diffrent		
5233	1547		0	e06b2ecb-				["b'953'"]		
5234	1547		0	f9808adb-				["b'3980'"]		
5235	1547		1	f0fdc411-4				["b'64498'"]		
5236	1547		0	67aa372b-				["b'SM-G965N'", "b'59745'"]		
5237	1547		1	11b4e606-				["b'64498'"]		
5238	1547		1	2cf7501b-				["b'21942'"]		
5239	1547		0	69bbfb24-				["b'35701'"]		
5240	1547		0	13c614cd-				["b'953'"]		
5241	1547		0	a9f30000-				["b'4030'"]		
5242	1547		1	17b1a019-				["b'64498'"]		
5243	1547		1	a3330350-				["b'21942'"]		
5244	1547		0	8b53ba8a-				["b'953'"]		
5245	1547		0	9e705bda-				["b'4005'"]		
5246	1547		1	4d2727b3-				["b'64498'"]		
5247	1547		1	4d9478d9-				["b'21942'"]		
5248	1547		0	37cb5b65-				["b'35488'"]		

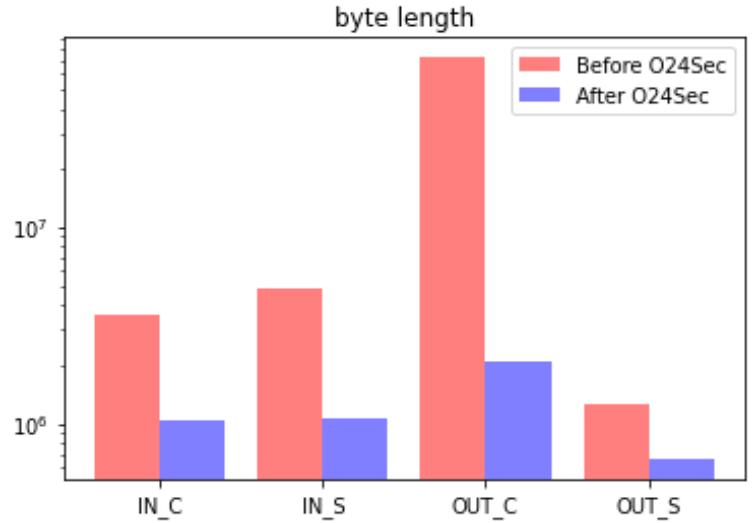
<재라벨링 의뢰 후 LABEL 상황>

	A	B	C	D	E	F	G	H	I	J
1	Cluster	DetectN	Result	_id	Payload	label_P	DetectN	diffrent		
5233	1547		0	e06b2ecb-				["b'953'"]		
5234	1547		0	f9808adb-				["b'3980'"]		
5235	1547		0	f0fdc411-4				["b'64498'"]		
5236	1547		0	67aa372b-				["b'SM-G965N'", "b'59745'"]		
5237	1547		0	11b4e606-				["b'64498'"]		
5238	1547		0	2cf7501b-				["b'21942'"]		
5239	1547		0	69bbfb24-				["b'35701'"]		
5240	1547		0	13c614cd-				["b'953'"]		
5241	1547		0	a9f30000-				["b'4030'"]		
5242	1547		0	17b1a019-				["b'64498'"]		
5243	1547		0	a3330350-				["b'21942'"]		
5244	1547		0	8b53ba8a-				["b'953'"]		
5245	1547		0	9e705bda-				["b'4005'"]		
5246	1547		0	4d2727b3-				["b'64498'"]		
5247	1547		0	4d9478d9-				["b'21942'"]		
5248	1547		0	37cb5b65-				["b'35488'"]		

프로젝트 결과 (시간 절약)



분석 유형 평균 80% 감소

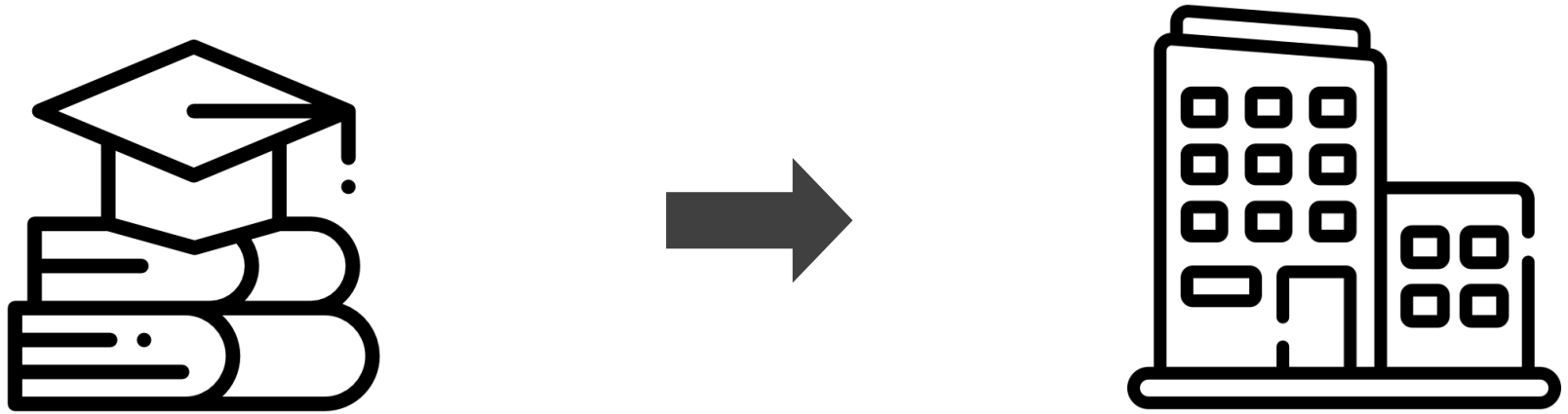


분석 byte 수 평균 70% 감소

멘토링

지도교수님이신 윤명근 교수님과, 실리콘벨리 현업자 이신 문성익 멘토님의 도움을 받음

학교에서 배운 이론들을 어떻게 회사 현업에 적용시킬지

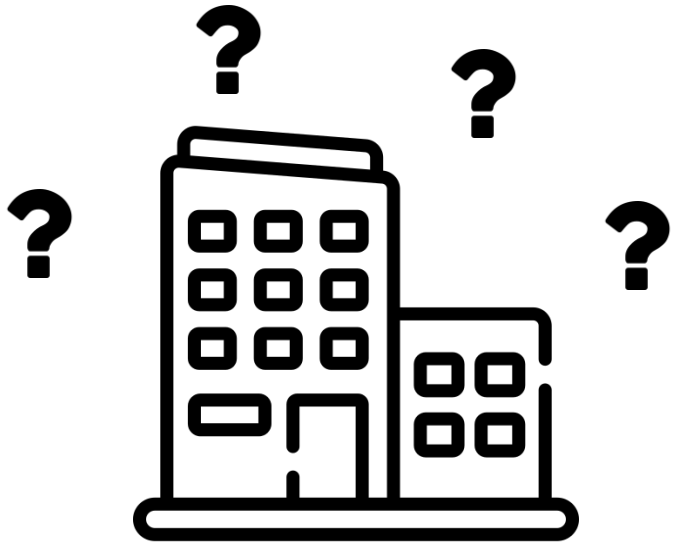


EX) Clustering, 유사도, Graph, NLP, 네트워크, 임계값 설정

멘토링

지도교수님이신 윤명근 교수님과, 실리콘벨리 현업자 이신 문성익 멘토님의 도움을 받음

회사에서 사용하는 최신 기술은 어떠한 것들인지



EX) TF-IDF, 코사인 유사도, Clustering 기법

멘토링

지도교수님이신 윤명근 교수님과, 실리콘벨리 현업자 이신 문성익 멘토님의 도움을 받음

주제에 대해 토론해보며 또 다른 방법은 없는지

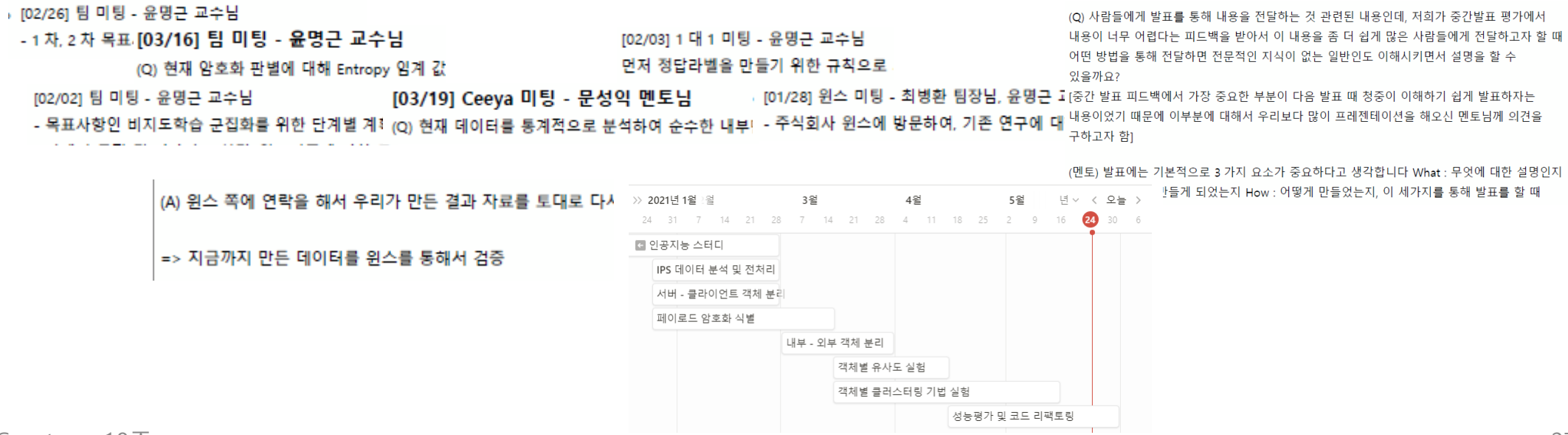
멘티의 질문 및 토론

- 클러스터링 feature / distance measure 를 어떻게 정하면 좋을까?
- 네트웍 구성에 대한 정보가 없는 상태에서 내부/외부 객체를 어떻게 구분 할 수 있을까?
- 위 질문에서 NAT 이 포함된 경우에는?
- 세션을 재구성 할 수 있나? 할 필요가 있나?
- 패킷 암호화 여부를 판별할 방법은?
- 학습된 클러스터로 Classification 을 어떻게 할 수 있을까?

멘토링

지도교수님이신 윤명근 교수님과, 실리콘벨리 현업자 이신 문성익 멘토님의 도움을 받음

그 외에도 시간 관리, 발표기술, 실험 방법 등 프로젝트 전체적 흐름 잡는데에 도움을 많이 받음



감사합니다.