# AI 보안관제 2.0

2023.05.11.(목)

국민대학교 소프트웨어융합대학 교수
국민대학교 인공지능연구소 소장
윤명근

# 목차

# 발표자 & 연구그룹

- 발표자: 윤명근
  - 1998~2010 금융결제원 금융ISAC/**보안관제센터**
  - 2010~2022 국민대학교 소프트웨어융합대학 인공지능학부
- 연구그룹: 국민대학교 정보보호연구실
  - https://infosec.kookmin.ac.kr
- 연구분야
  - 2010~2015: 네트워크 보안, 침입탐지, 핀테크
  - **2016~2022: 보안빅데이터, 보안관제, 인공지능 보안**
- 주요실적
  - 논문 SCIE 30편, 특허 30편, 수상 및 기술이전 다수
- 산학협력/자문기관
  - KISTI 과학기술사이버안전센터, 윈스, 지란지교시큐리티, 금융보안원, 금융결제원 등

# 보안관제기술동향

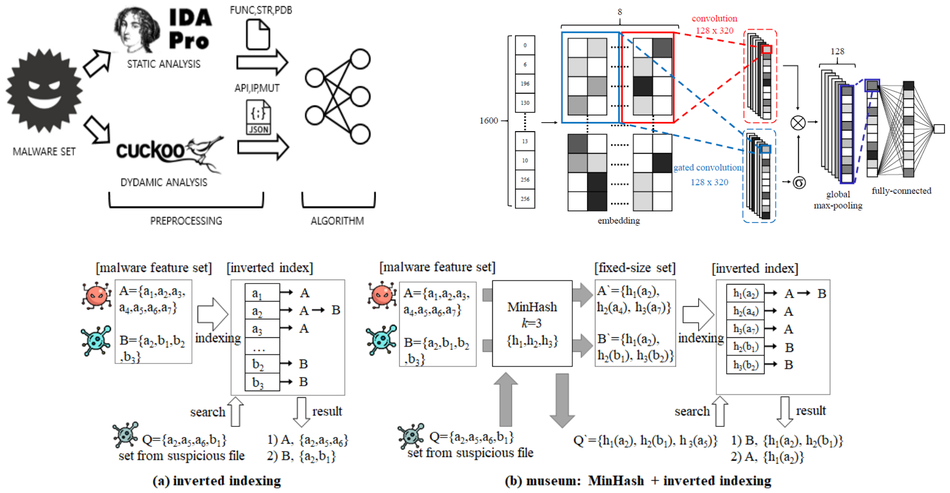
- AI & Security
  - AI for security, **AI for SOC**
    - ✓SOC: Security Operations Center (보안관제센터)
  - Security for AI

## [사람들의 생각]

영화 터미네이터의 한 장면

## [AI for security]

- AI 기반 보안빅데이터 분석 & (반)자동화 연구
- 보안산업 현장 + 대학 연구실

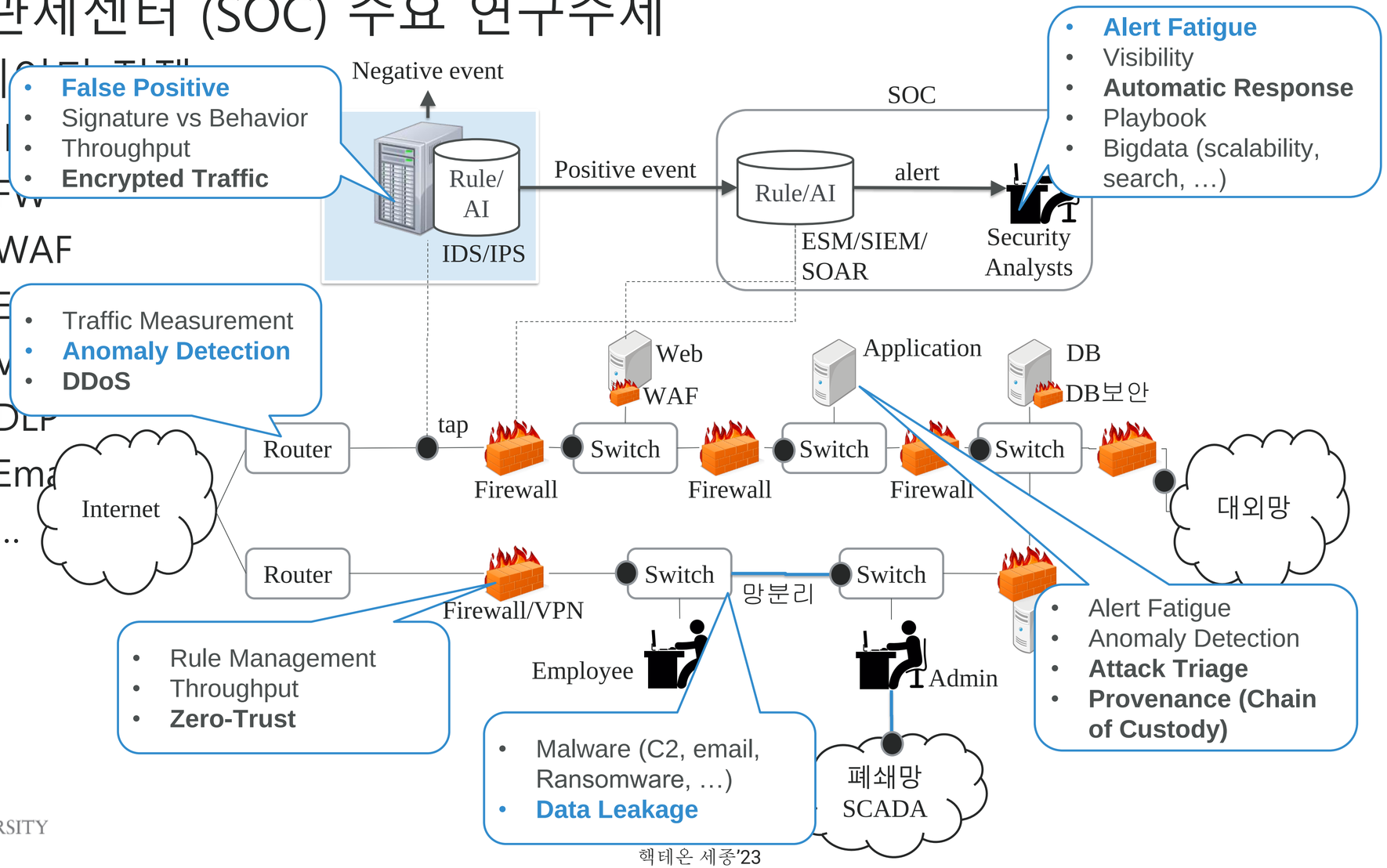



## [Security for AI]


https://medium.com/self-driving-cars/adversarial-traffic-signs-fd16b7171906

# 보안관제기술동향

- 보안관제센터 (SOC) 주요 연구주제
  - 빅데이터 적용
    - ✓IPS
    - ✓FW
    - ✓WAF
    - ✓E
    - ✓V
    - ✓DLP
    - ✓Ema
    - ✓…

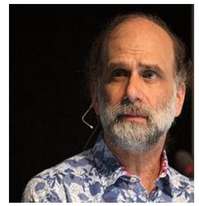


# 보안관제기술동향

- MSS (Managed Security Service)
  - N-IDS (Network-based Intrusion Detection System)
- ESM (Enterprise Security Management)
  - RDBMS
- SIEM (Security Information and Event Management)
  - 빅데이터 구축 및 검색
- SOAR (Security Orchestration, Automation and Response)
  - AI, TI (Threat Intelligence), RPA (Robotic Process Automation)

- SNORT (1998)
- Bro/Zeek (1999)

**Managed Security Monitoring:Network Security for the 21st Century**



Bruce Schneier Computers & Security 20 (2001) 491-503

- 국내 ISAC 출범 (2002)
- 데이터마이닝으로 보안관제 로그 분석 자동화 (상관관계 분석, 자산 목록 및 취약점점검 결과 연계)
- 알파고 (2016)

**1세대 단위보안관제**

Perimeter Security

- 운영 중심, IPS 자동 차단 의존적

**2세대 통합보안관제**

Data Security

- ESM (Enterprise Security Management)
- 이벤트(Alert) 중심 관제 센터

**3세대 빅데이터보안관제**

Threat Intelligence

- SIEM(Security Information and Event Management
- 취약점 예방 중심의 관제 센터

**4세대 AI & SOAR**

Zero Trust & Threat Hunting

- SOAR (Security Orchestration, Automation, and Response)
- 침해대응 자동화

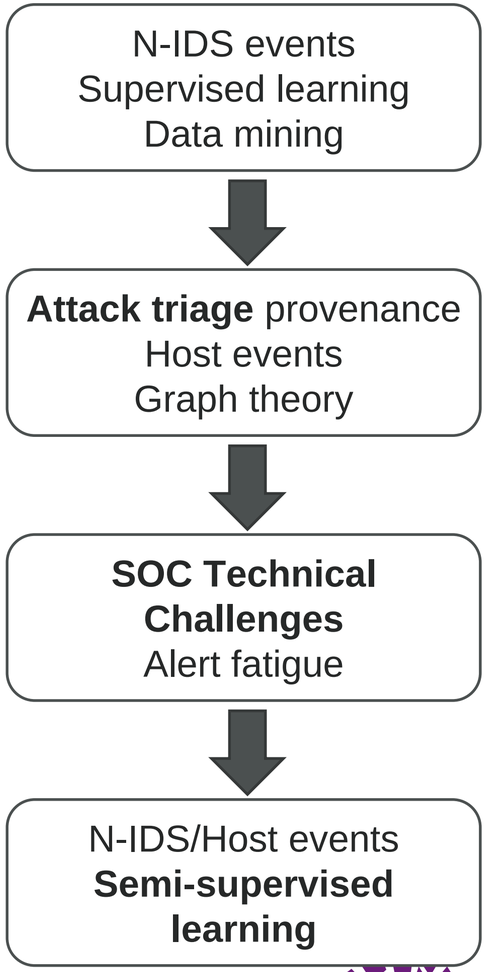
"보안 오케스트레이션을 이용한 관제 적용 사례", 이글루시큐리티, https://www.igloo.co.kr/security-information/보안-오케스트레이션을-이용한-관제-적용-사례/

# 보안관제기술동향

- 사이버보안 핵심 역할 수행
- 보안관제센터 (SOC) 국내 현황
  - 국내 42개 부문보안관제센터 운영
    - ✓금융-금융보안원, 통신·과학-KISTI과학기술사이버안전센터, 교육-교육부사이버안전센터, …
  - 국내 17개 보안관제 전문기업 지정
    - ✓윈스, 이글루시큐리티, 삼성에스디에스, 한전KDN, …
  - ISAC (Information Sharing and Analysis Center)
    - ✓정보통신(2002), 금융(2002), 행정(2015), 의료(2018)
  - 기타 사내 보안관제센터

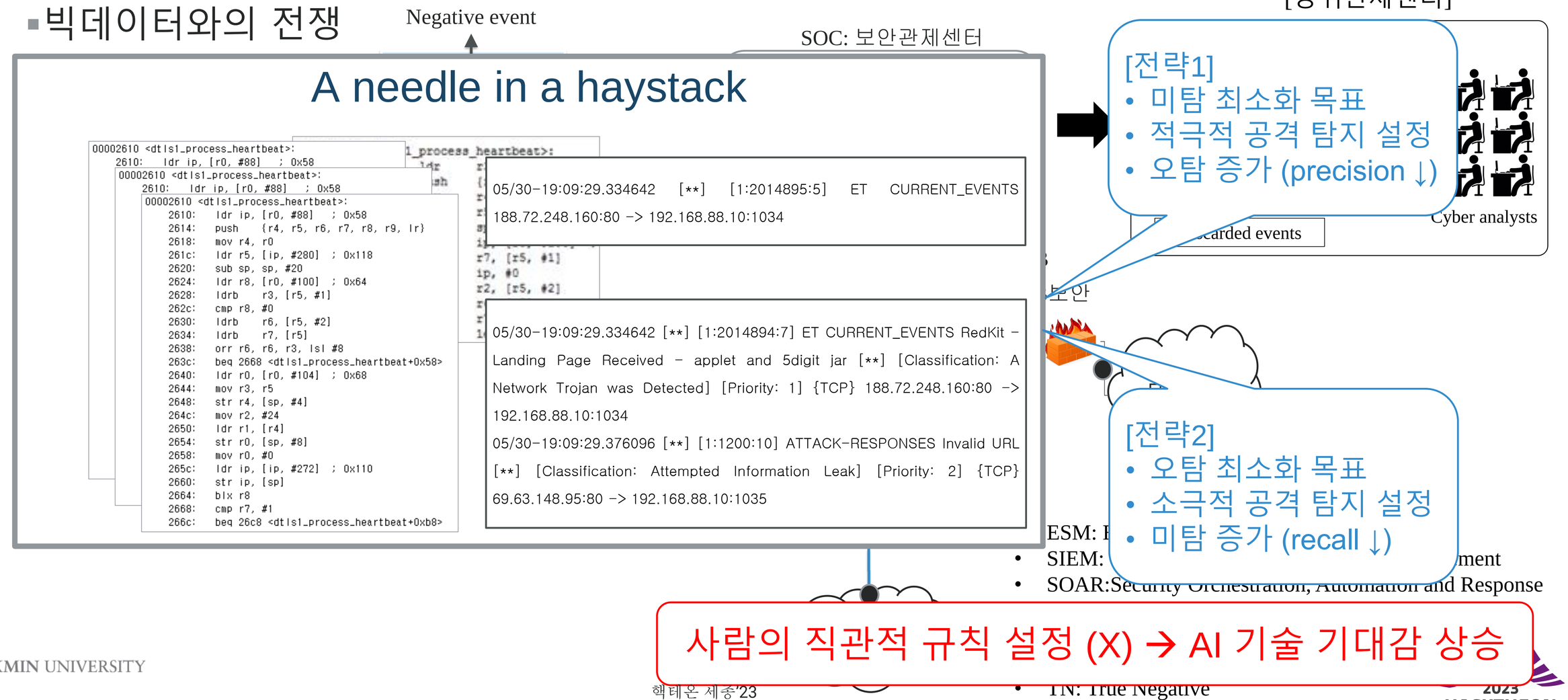
# 보안관제기술동향

- 2020s: AI보안관제
  - Y. Shen, et al., “Tiresias: Predicting security events through deep learning,” ACM CCS'18
  - Y. Shen, et. al, “Attack2vec: Leveraging temporal word embeddings to understand the evolution of cyberattacks,” USENIX Security'19
  - **F. Kokulu, et al., ‘‘Matched and mismatched SOCs,’’ ACM CCS’19**
  - W., et al., “Nodoze: Combatting threat alert fatigue with automated provenance triage,” NDSS'19
  - F. Liu, et al., “Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise,” ACM CCS'19
  - R. Tang, et al., “Zerowall: Detecting zero-day web attacks through encoder-decoder recurrent neural networks,” INFOCOM'20
  - **B. Alahmadi, et al., “99% False Positives: A Qualitative Study of SOC Analysts’ Perspectives on Security Alarms,” USENIX Security'22**
  - **T. Ede, et al., “DEEPCASE: Semi-Supervised Contextual Analysis of Security Events,” IEEE Security & Privacy'22**
  - J. Zeng, et al,. “SHADEWATCHER: Recommendation-guided Cyber Threat Analysis using System Audit Records,” IEEE Security & Privacy'22

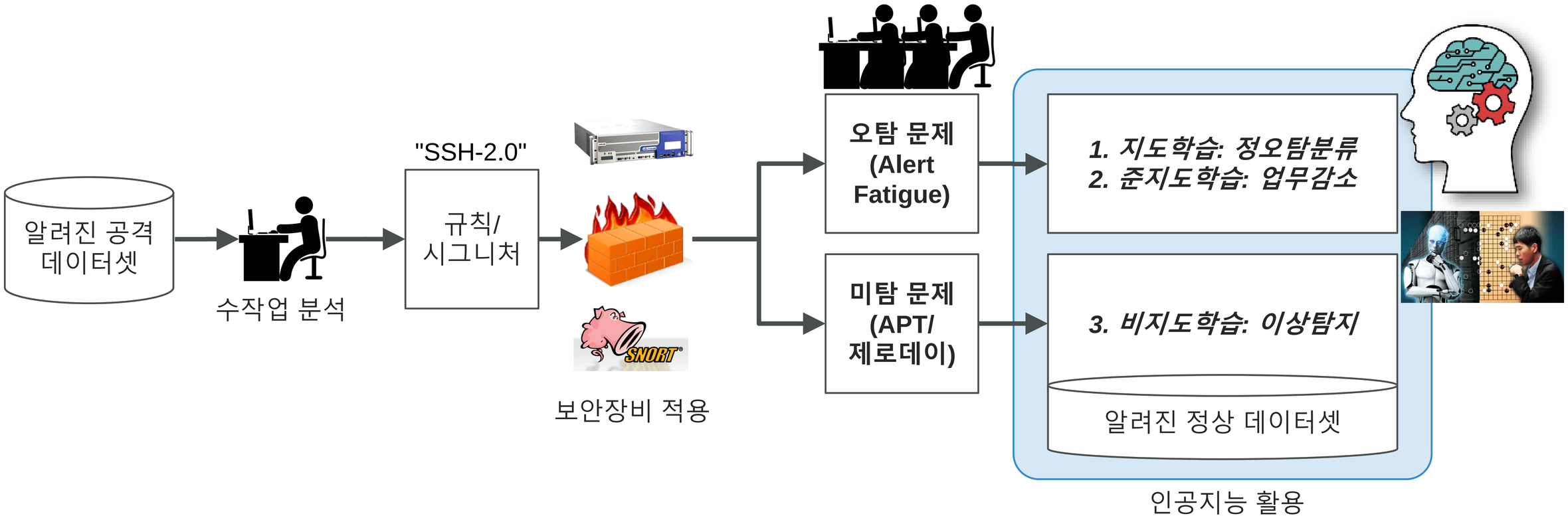


# 문제 정의

- 보안관제센터 (SOC) since 2000
  - 빅데이터와의 전쟁



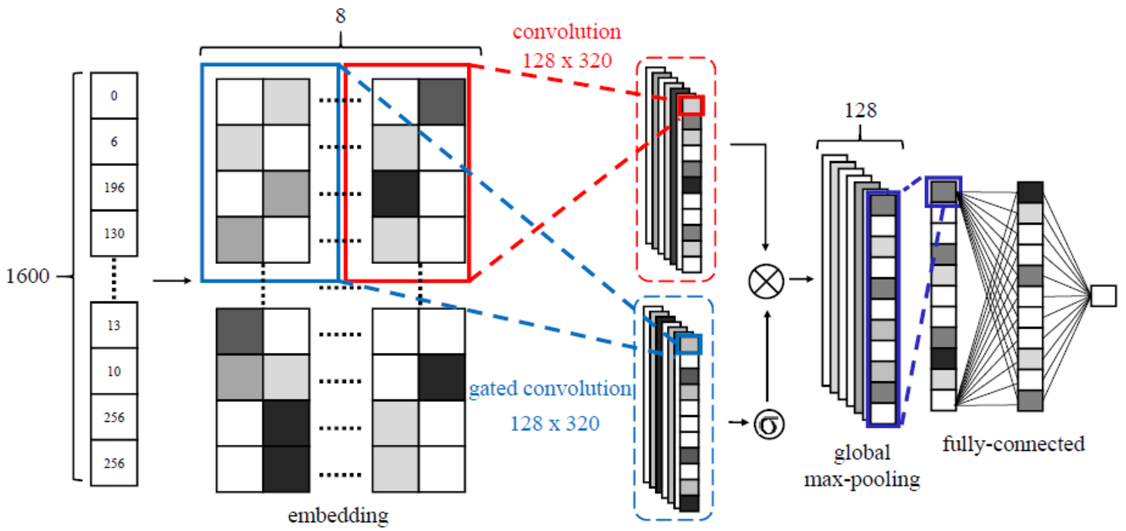
# 문제 정의


알려진 공격
데이터셋
수작업 분석
"SSH-2.0"
규칙/
시그니처
SNORT
보안장비 적용
오탐 문제
(Alert
Fatigue)
미탐 문제
(APT/
제로데이)
1. 지도학습: 정오탐분류
2. 준지도학습: 업무감소
3. 비지도학습: 이상탐지
알려진 정상 데이터셋
인공지능 활용

# AI 보안관제 1.0

- 지도학습: 정오탐분류 연구사례
  - N-IPS 패킷 정보로부터 피처 생성

    -GET /web_resources/js/jquery_ui.js?20191127&vEhF=6587 AND 1=1 UNION ALL SELECT 1
  - 오탐 이벤트 경우 IPS 설치 사이트 특성 때문에 발생
    - ✓비슷한 패킷 패턴 자주 목격
  - 모델 개발 및 검증
    - ✓1dConv + DNN
    - ✓NLP parser + Bi-LSTM
    - ✓TF-IDF + XGBoost

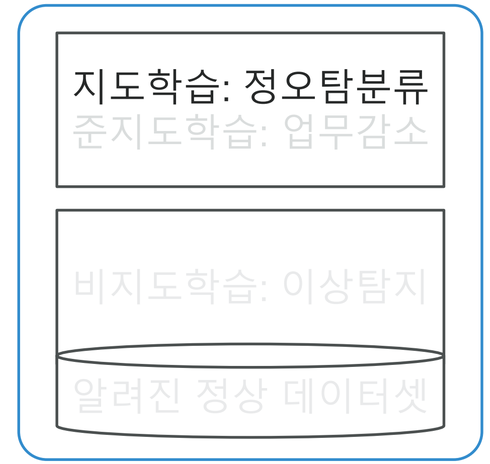




사이버안보 논문 공모전 최우수 논문, “멀티모달 딥러닝을 이용한 보안관제 이벤트 자동 분류 연구”, 2020.9

# AI 보안관제 1.0

- 지도학습: 정오탐분류 연구사례
  - 장점
    - ✓주어진 학습데이터 내에서 잘 동작 (F1스코어 95% 이상)
  - 단점
    - ✓OOD(Out-Of-Distribution) 문제, Concept-drift, Data-drift,…
      - 수집 사이트별로 데이터 양상 차이 (공간적)
      - 신규 이벤트 등장, 신규 서버 등장 (시간적)

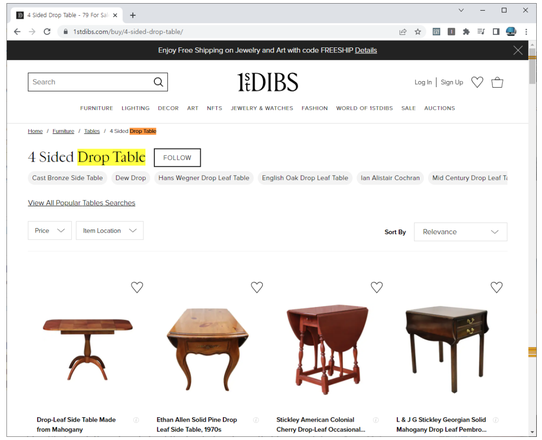



```
var statement = "SELECT * FROM users
WHERE name = '" + userName + "'";
```

```
a';DROP TABLE users; SELECT * FROM
userinfo WHERE 't' = 't
```

```
SELECT * FROM users WHERE name =
'a';DROP TABLE users; SELECT * FROM
userinfo WHERE 't' = 't';
```

오탐!

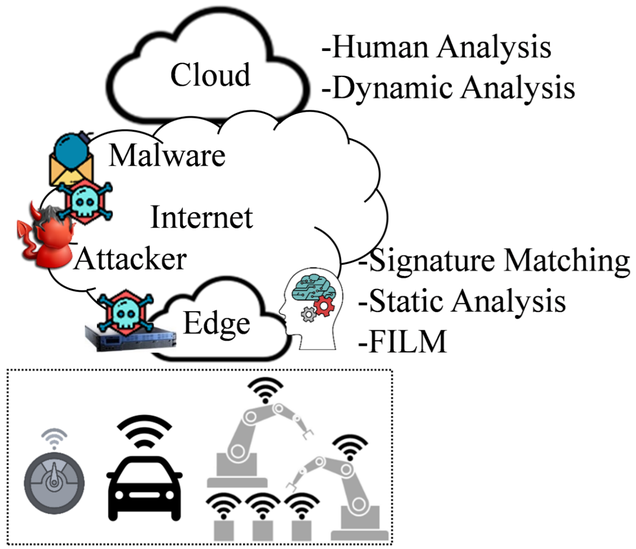



Server IP: 100.100.100.100

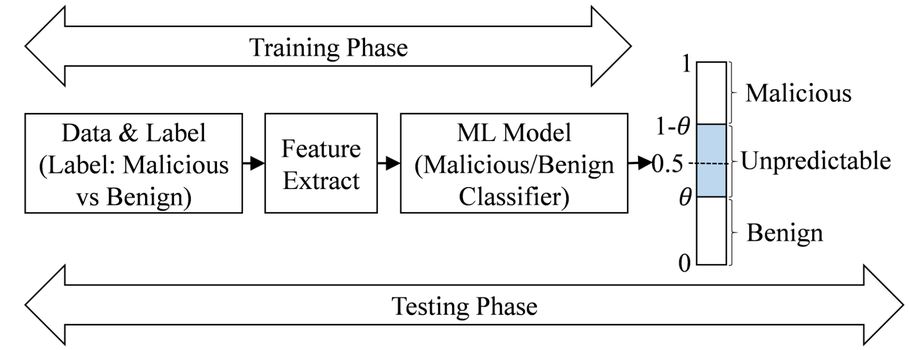
: detect signature

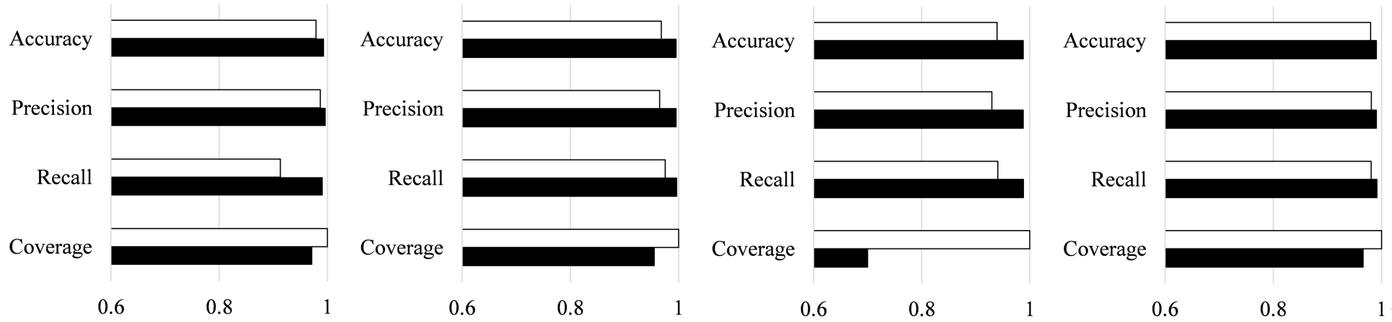
# AI 보안관제 1.0

- 지도학습: 정오탐분류 연구사례
  - OOB (Out-of-Bound) 고려
  - 틀리는 답을 하는 AI < 모른다고 답하는 AI

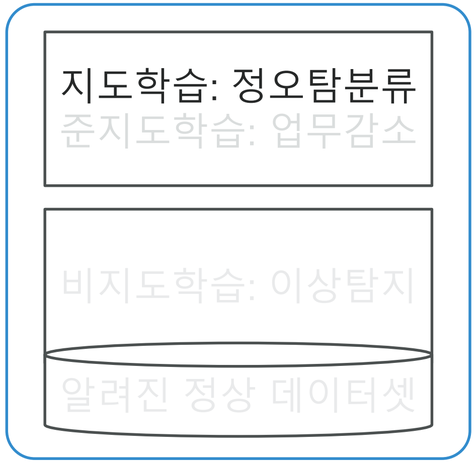



(b) Malware Detection for Edge Computing



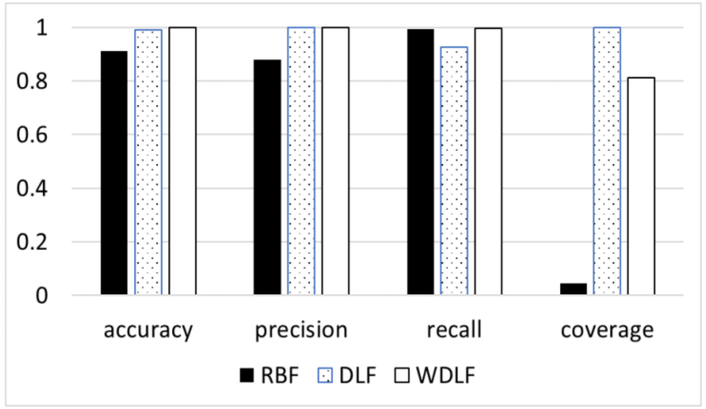




(a) Model 1 [9] (b) Model 2 [13] (c) Model 3 [14] (d) Model 4 [15]

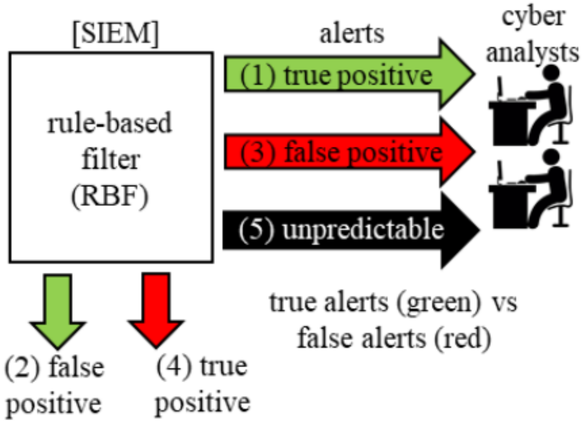


# AI 보안관제 1.0

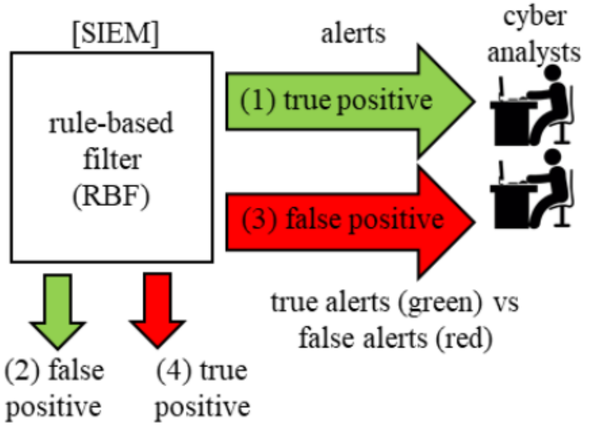
- 지도학습: 정오탐분류 연구사례
  - 웨이버(Waiver) 적용
  - 신규 이벤트 제외



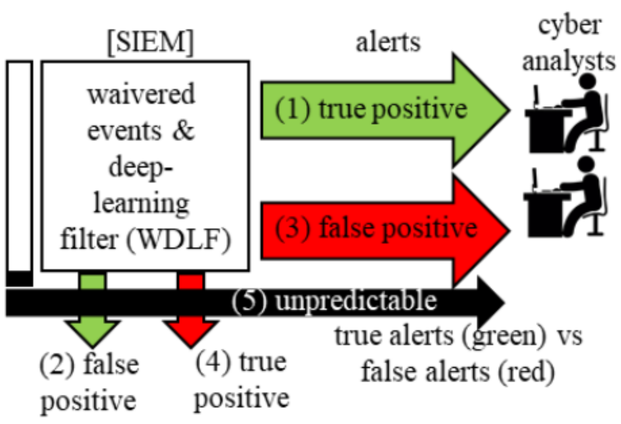





(a) current handcrafted rule-based filter



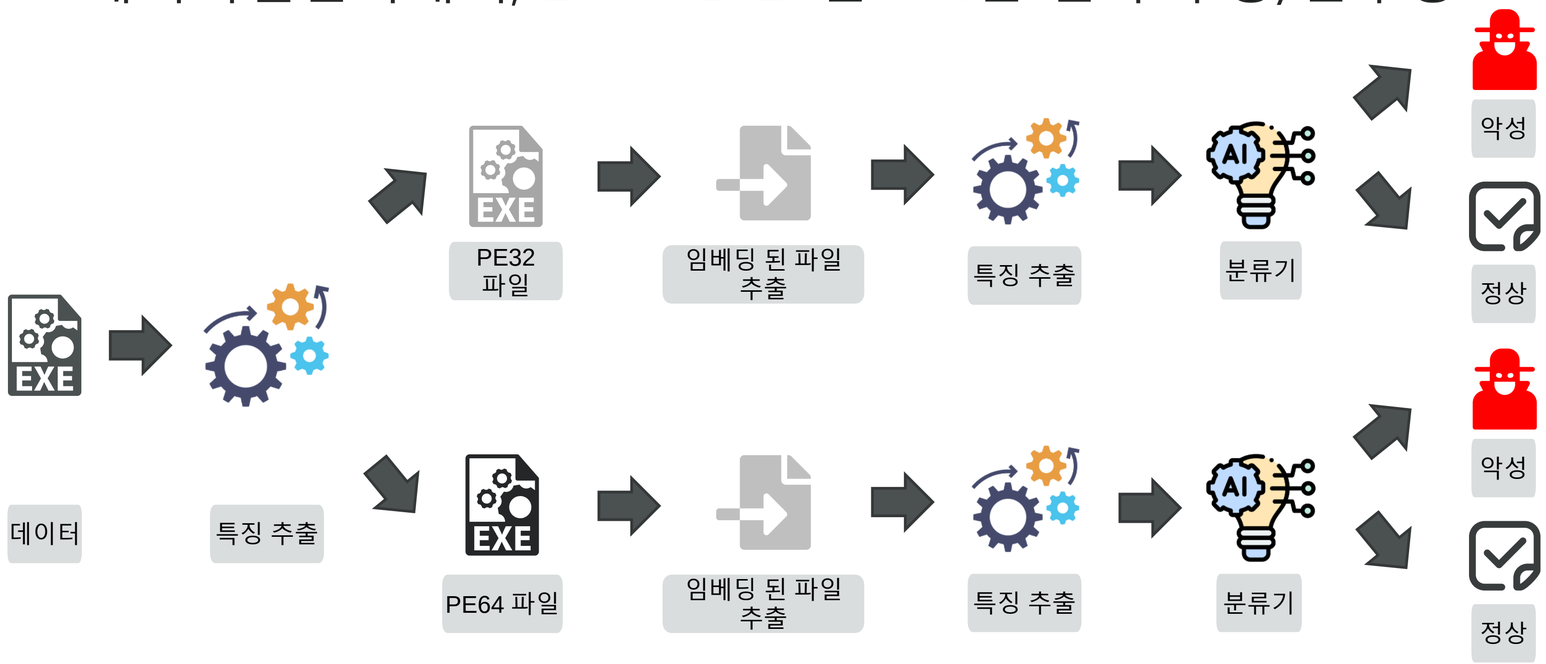

(b) deep-learning filter on raw packets inside events

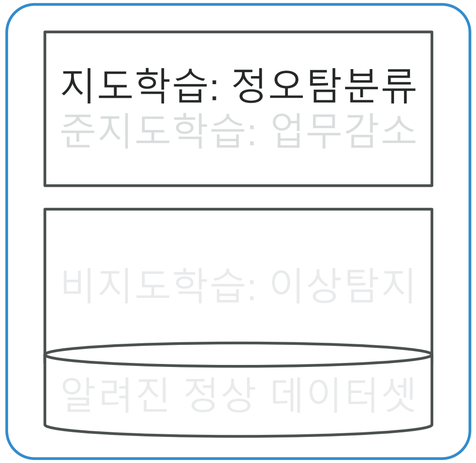


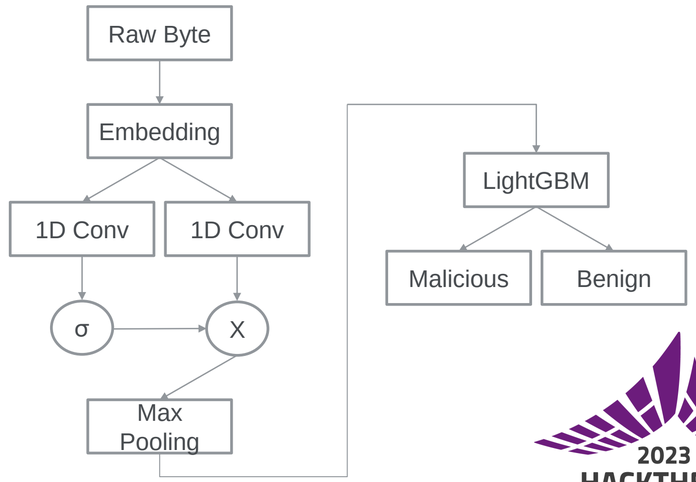

(c) waivered events and deep-learning filter

# AI 보안관제 1.0

- 지도학습: 정오탐분류 연구사례
  - 악성코드탐지
    - ✓PE파일, 문서형(PDF, OOXML, HWP), 안드로이드 앱
  - 데이터챌린지대회, 2017~2020년도 4년 연속 우승/준우승

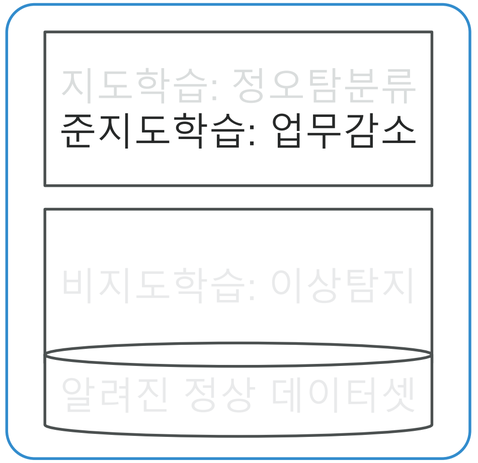


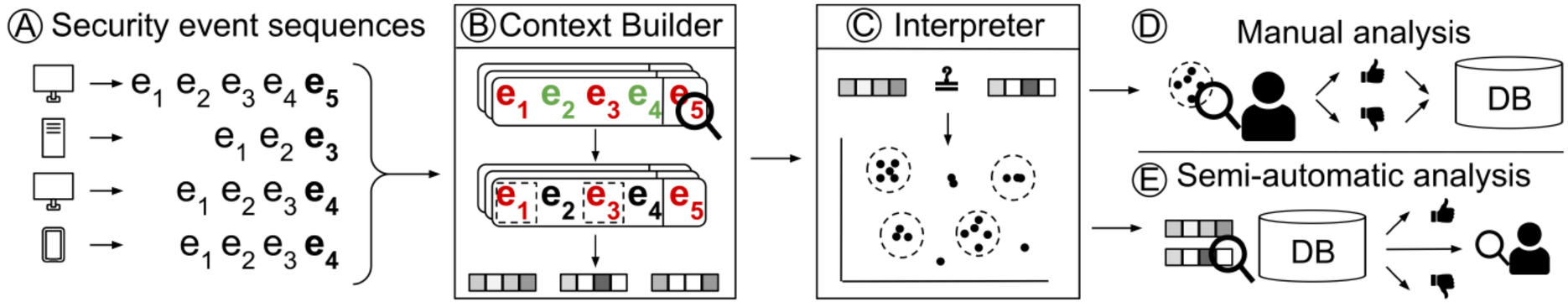



# AI 보안관제 1.0

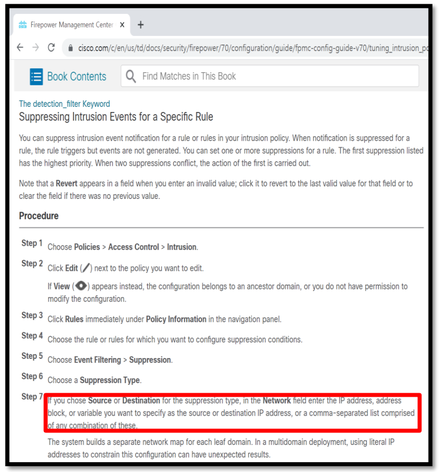
- 준지도학습: 업무감소 연구사례
  - 국내는 IPS 이벤트 콘텐트 + 지도학습
  - 해외는 IPS 이벤트 시퀀스 + 비지도학습
    - ✓T. Ede, et al., "DEEPCASE: Semi-Supervised Contextual Analysis of Security Events," IEEE Security & Privacy'22
    - ✓비지도학습 딥러닝 모델의 어텐션 벡터 (이벤트 이름 시퀀스 예측)
  - 기술 한계
    - ✓비지도학습 패러미터 최적화 까다로움
    - ✓데이터 많아지면 클러스터링 느려짐



# AI 보안관제 1.0

- 준지도학습: 업무감소 연구사례
  - RAID: Reducing Alert Fatigue in Network Intrusion Detection, 논문 심사 중
  - Alert Fatigue 해결을 위해 탐지 제외(filtering) 정책 추가
    - ✓(이벤트명, IP주소) 기준 탐지 제외
    - ✓F. Kokulu, et al., ‘‘Matched and mismatched SOCs,’’ ACM CCS’19
      - 해외: 과감한 탐지 제외 정책 권장, 국내: 미탐 우려
  - 연구 목표
    - ✓Fine-Grained Filtering
    - ✓준지도학습으로 (반)자동화
  - 기존 기술
    - ✓Coarse-Grained Filtering
    - ✓탐지 제외 정책 수작업 추가



Firepower Management Center

cisco.com/c/en/us/td/docs/security/firepower/70/configuration/guide/fpmc-config-guide-v70/tuning_intrusion_p

Book Contents | Find Matches in This Book

The detection_filter Keyword

Suppressing Intrusion Events for a Specific Rule

You can suppress intrusion event notification for a rule or rules in your intrusion policy. When notification is suppressed for a rule, the rule triggers but events are not generated. You can set one or more suppressions for a rule. The first suppression listed has the highest priority. When two suppressions conflict, the action of the first is carried out.

Note that a **Revert** appears in a field when you enter an invalid value; click it to revert to the last valid value for that field or to clear the field if there was no previous value.

**Procedure**

Step 1 Choose **Policies** > **Access Control** > **Intrusion**.

Step 2 Click **Edit** (✎) next to the policy you want to edit.

If **View** (👁) appears instead, the configuration belongs to an ancestor domain, or you do not have permission to modify the configuration.

Step 3 Click **Rules** immediately under **Policy Information** in the navigation panel.

Step 4 Choose the rule or rules for which you want to configure suppression conditions.

Step 5 Choose **Event Filtering** > **Suppression**.

Step 6 Choose a **Suppression Type**.

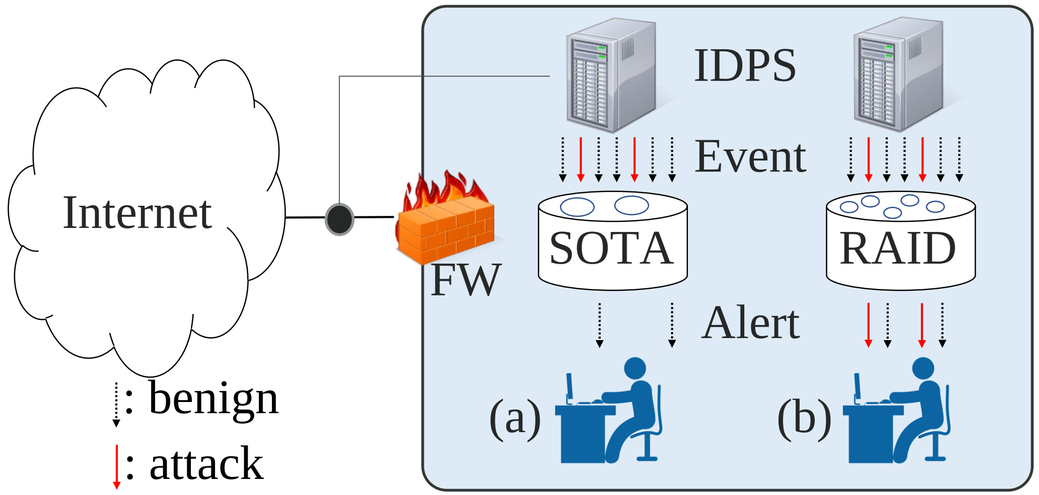
Step 7 If you chose **Source** or **Destination** for the suppression type, in the **Network** field enter the IP address, address block, or variable you want to specify as the source or destination IP address, or a comma-separated list comprised of any combination of these.

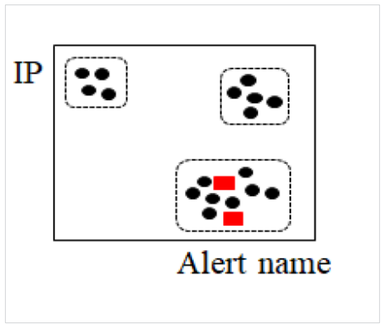
The system builds a separate network map for each leaf domain. In a multidomain deployment, using literal IP addresses to constrain this configuration can have unexpected results.

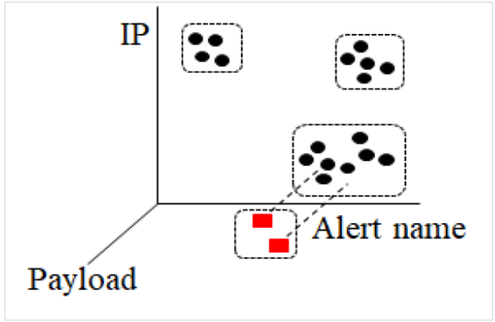
**Step 7** If you chose **Source** or **Destination** for the suppression type, in the **Network** field enter the IP address, address block, or variable you want to specify as the source or destination IP address, or a comma-separated list comprised of any combination of these.

# AI 보안관제 1.0

- 준지도학습: 업무감소 연구사례
  - RAID: Reducing Alert Fatigue in Network Intrusion Detection, 논문 심사 중

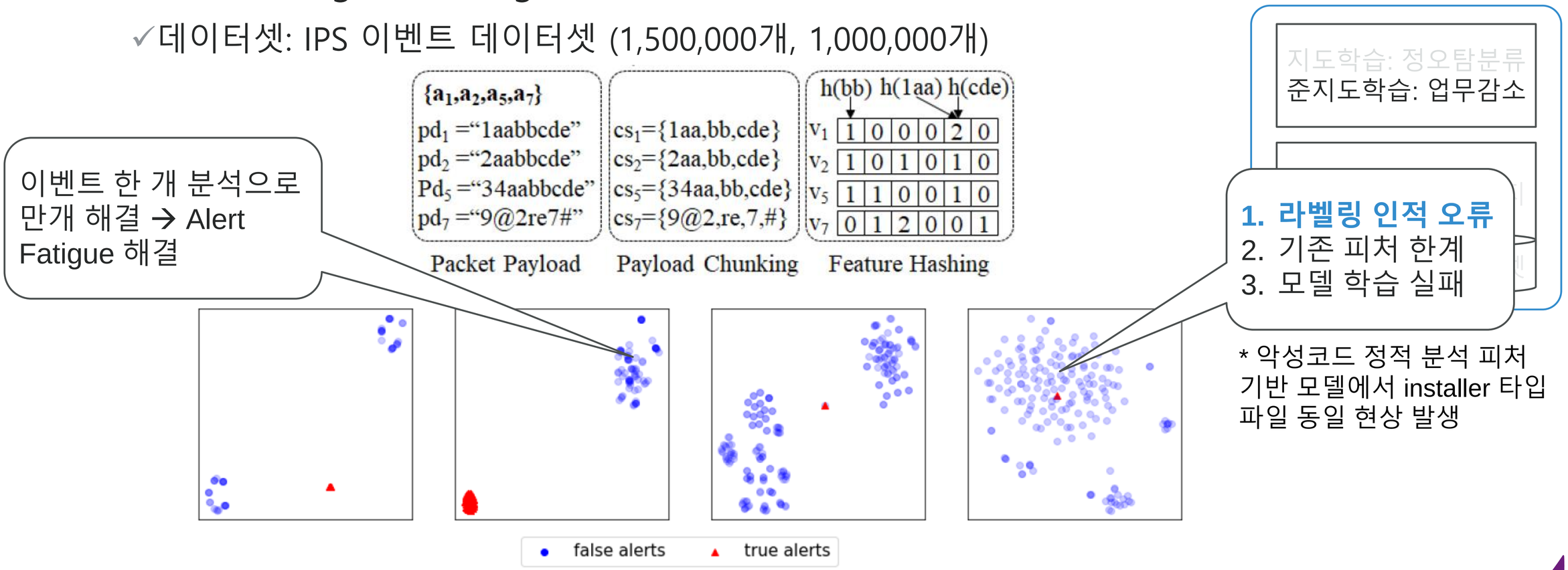



# AI 보안관제 1.0

- 준지도학습: 업무감소 연구사례
  - RAID: Reducing Alert Fatigue in Network Intrusion Detection, 논문 심사 중
    - ✓데이터셋: IPS 이벤트 데이터셋 (1,500,000개, 1,000,000개)




이벤트 한 개 분석으로 만개 해결 → Alert Fatigue 해결

지도학습: 정오탐분류
준지도학습: 업무감소

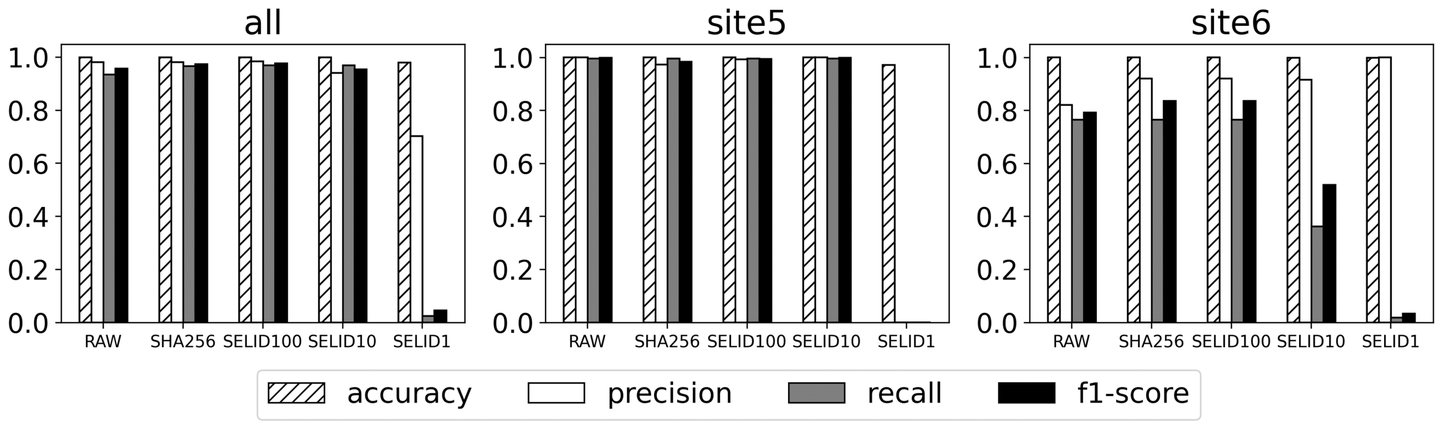
1. **라벨링 인적 오류**
2. 기존 피처 한계
3. 모델 학습 실패

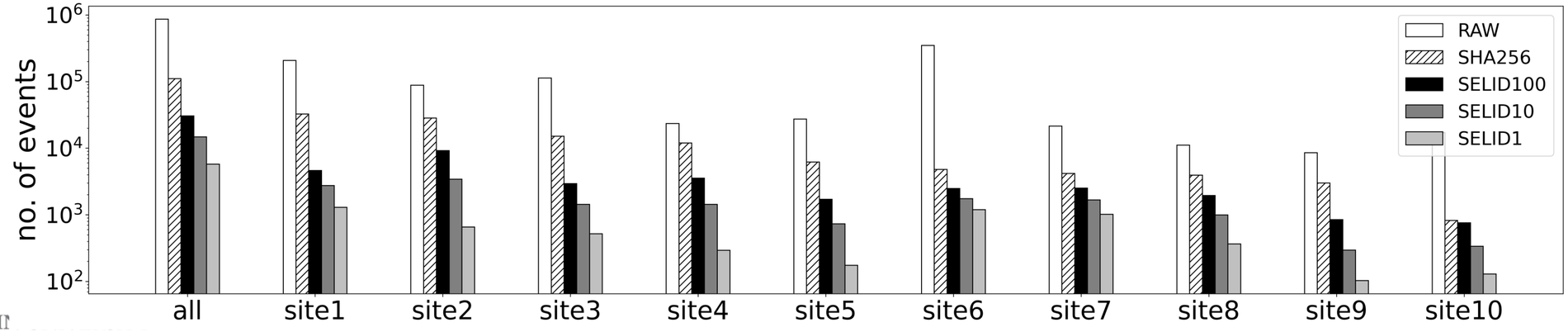
* 악성코드 정적 분석 피처 기반 모델에서 installer 타입 파일 동일 현상 발생

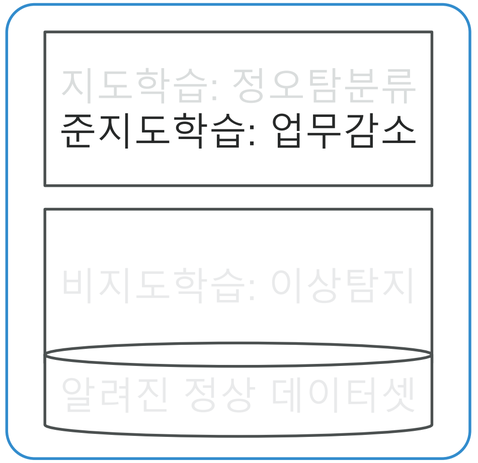
t-SNE visualization of RAID clustering results. Each plot shows a group of alerts of the same $IIP||an$.

# AI 보안관제 1.0

- 준지도학습: 업무감소 연구사례
  - 데이터 중복성 제거 기술 활용
    - ✓Selective Event Labeling for Intrusion Detection datasets, CISC-W'22

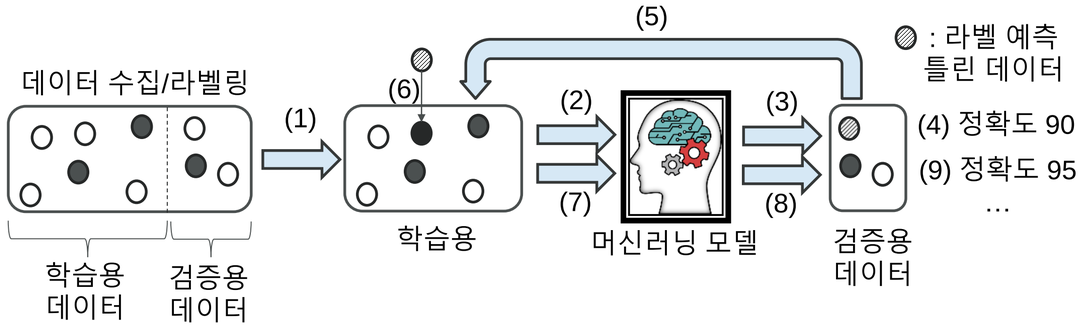


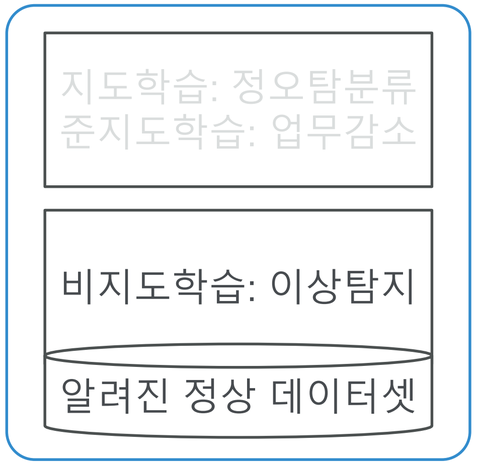



# AI 보안관제 1.0

- 비지도학습: 이상탐지 (라벨 오류)
  - 쉬운 데이터셋 지도학습 → 성능 한계 발생?




2021 AI+Security 우수논문·아이디어 공모전 최우수상(과학기술정보통신부 장관상),
“MaaD: 머신러닝을 이용한 보안데이터 라벨 디버거“




✓Dataset

- 이벤트명 분류 (약 5,000개, 대분류 3개, 소분류 28개)
  - F1-score 0.9 → 0.98

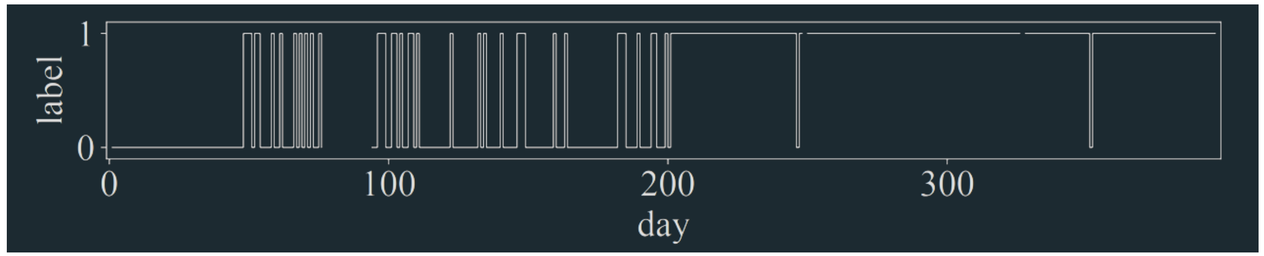
| | | |
|---|---|---|
| GNU C Library | BoF | 3 |
| GNU C Library | BoF.A | 3 |
| GNU C Library | BoF.B | 1 |

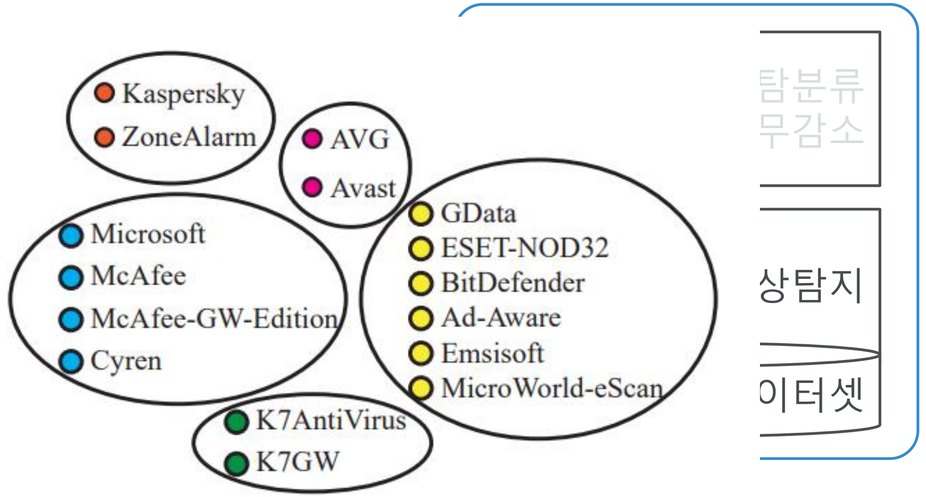
- IPS 침입탐지 이벤트 (약 1,500,000개)

```
--GET /web_resources/js/jquery_ui.js?20191127&vEhF=6587 AND 1=1 UNION ALL SELECT 1 NULL '<script>alert("XSS")</script>' table_name FROM
-GET /web_resources/js/jquery-1.11.2.min.js?20191127&PMqe=2856 AND 1=1 UNION ALL SELECT 1 NULL '<script>alert("XSS")</script>' table_nai
```

# AI 보안관제 1.0

- 비지도학습: 데이터감사/라벨오류
  - S. Zhu, et al., “Measuring and Modeling the Label Dynamics of Online Anti-Malware Engines,” USENIX Security'20



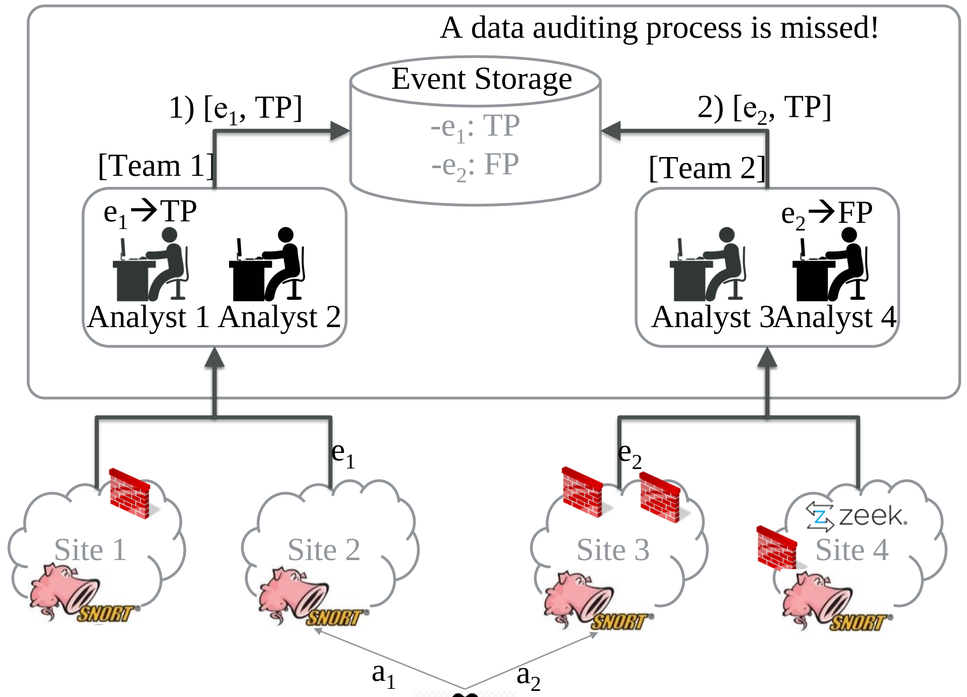

  - MLOps vs DevOps
    - ✓Andrew Ng (이구아나 사진 라벨링)
      - https://www.youtube.com/watch?v=06-AZXmwHjo

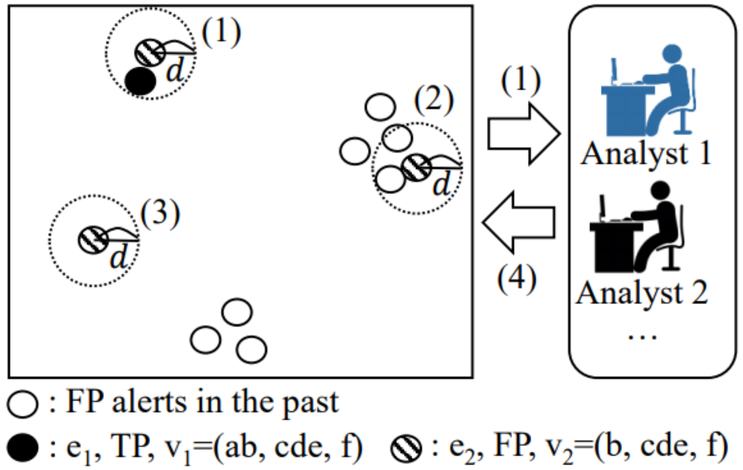


1. 라벨 불일치 발생 가능
2. **감사 기능 필요!**

# AI 보안관제 1.0

- 비지도학습: 데이터감사/라벨오류
  - 데이터 감사 제도와 기술 필요성
  - "Data Auditing for Intelligent Network Security Monitoring," IEEE Communications Magazine, 2023




지도학습: 정오탐분류
준지도학습: 업무감소

비지도학습: 이상탐지

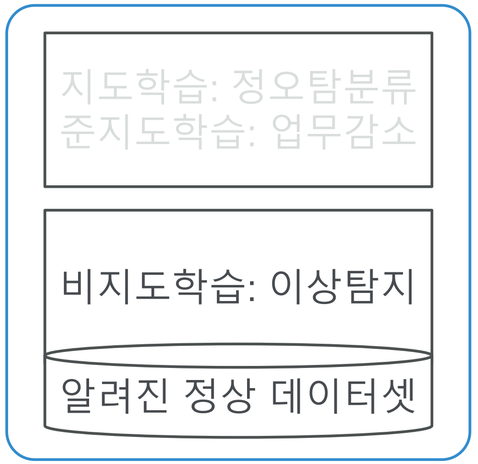
알려진 정상 데이터셋

# AI 보안관제 1.0

- 비지도학습: 네트워크 플로우 이상탐지
  - 현장 문제
    - ✓IDS/IPS 기반 보안관제
    - ✓암호화된 트래픽 증가
      - 네트워크 플로우 정보만 활용 가능
    - ✓엔드포인트 로그/이벤트 활용 불가능
      - 기술적 문제가 아닌 정책적 문제

| 구분 | 보안이벤트 | 침해대응 |
|---|---|---|
| 2012년 | 1,165,780,019 | 2,093 |
| 2013년 | 4,457,431,724 | 2,611 |
| 2014년 | 7,669,366,325 | 2,329 |
| 2015년 | 9,299,910,213 | 2,423 |
| 2016년 | 5,142,182,012 | 1,671 |
| 2017년 | 6,782,338,158 | 863 |
| 2018년 | 5,826,722,829 | 438 |
| 2019년 | 2,013,667,894 | 503 |
| 2020년 | 3,799,502,860 | 596 |
| 계 | 46,156,902,034 | 13,527 |

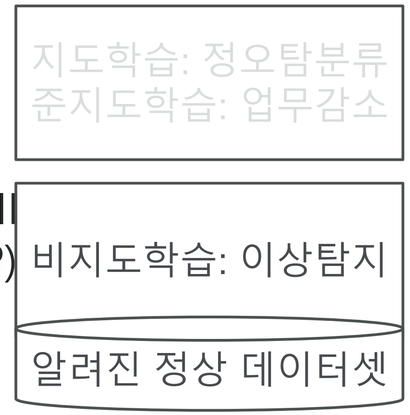
이윤수 외 5인, "**코로나19에 따른 사이버위협 및 대응기술 동향**, 한국정보보호학회지 2021




  - 연구 목표
    - ✓네트워크 플로우 정보 기반의 이상 탐지 기술 개발
    - ✓다양한 네트워크 환경에서 최소한의 추가 설정으로 동작 가능
    - ✓DDoS, Port scanning, Brute-forcing tool (password guessing), data leakage 등 다양한 공격 시도 탐지

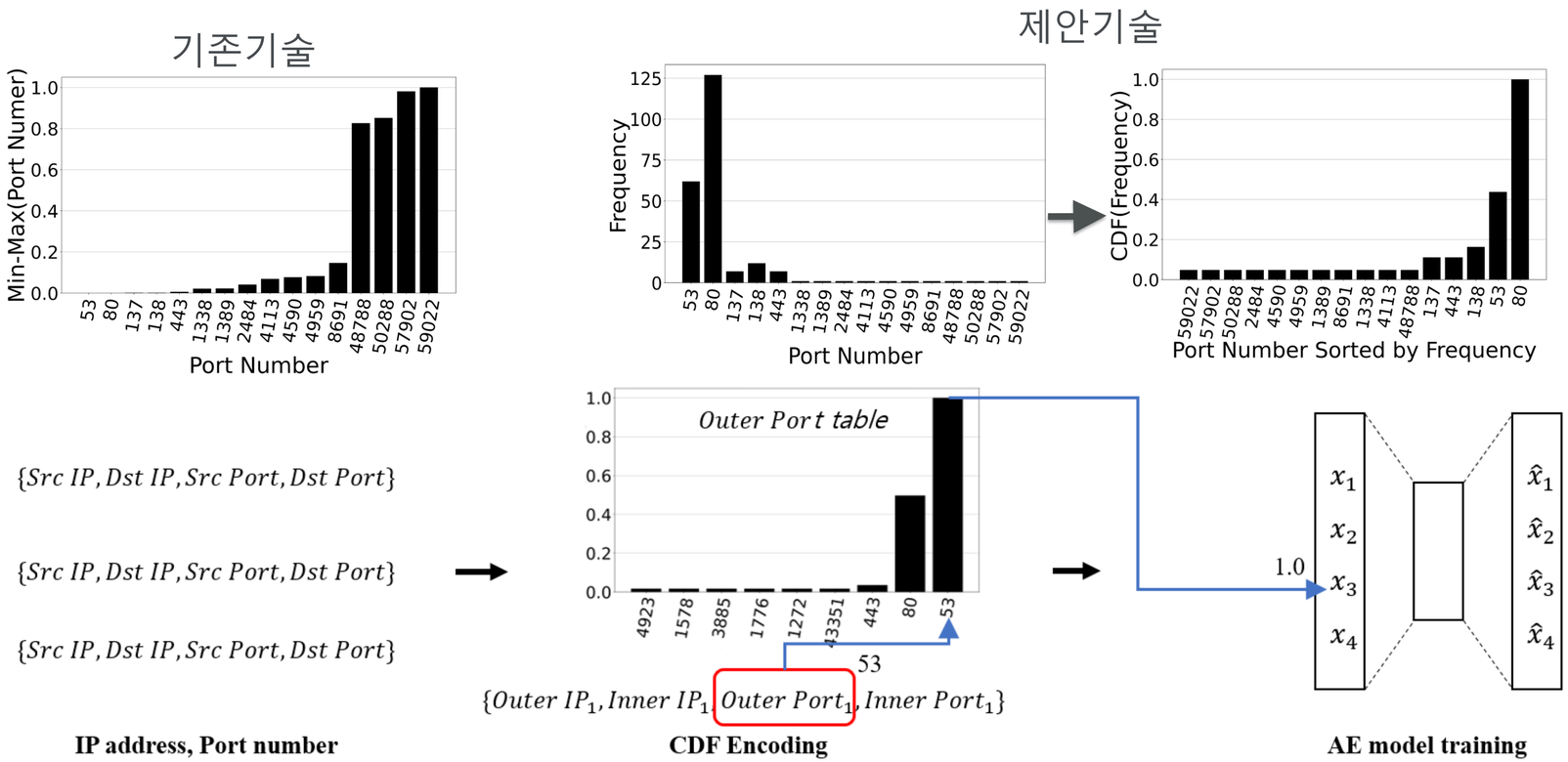
# AI 보안관제 1.0

- 비지도학습: 네트워크 플로우 이상탐지
  - 기존 기술
    - ✓L. Ertoz, et al.,"The MINDS – Minnesota Intrusion Detection System," 2003
      - 피처 (numerical) + 머신러닝 모델 (SVM)
      - 타임 윈도우 단위 피처 생성
        - **sIP, dIP, sPort, dPort**, protocol flags, # of bytes, # of packets, card(sIP,dIP), card(dIP,sIP), count(sIP||dPort), count(dIP||sPort), card(sIP,dIP), card(dIP,sIP) count(sIP||dPort), count(dIP||sPort)
    - ✓S. Jan, et al., "Throwing Darts in the Dark? Detecting Bots with Limited Data using Neural Data Augmentation," IEEE Symposium on Security and Privacy 2020
      - Frequency encoding 소개. URL, 쿠키 등에 적용
      - 피처 (numerical, categorical) + 딥러닝 모델 (Auto-encoder) + data augmentation



# AI 보안관제 1.0

- 비지도학습: 네트워크 플로우 이상탐지
  - Frequency Encoding, CDF (Cumulative Distribution Function) 정규화
    - ✓원격주소 프로파일링과 딥러닝을 이용한 네트워크 이상탐지 연구, 제6회 금융보안원 논문공모전 대상, 2022.




지도학습: 정오탐분류
준지도학습: 업무감소

비지도학습: 이상탐지

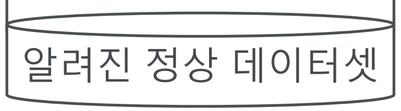
알려진 정상 데이터셋

# AI 보안관제 1.0

- 비지도학습: 네트워크 플로우 이상탐지
  - 딥러닝 이상 탐지를 위한 IP 주소와 포트 번호 인코딩, CISC-W'22
    - ✓4차원 피처 (source/destination IP + Port) 사용
    - ✓경쟁 기술은 40~80차원 피처 사용

지도학습: 정오탐분류
준지도학습: 업무감소

비지도학습: 이상탐지




| Dataset | CIC-IDS-2017 | | | CIC-IDS-2018 | | | Ton-IOT | | |
|---|---|---|---|---|---|---|---|---|---|
| Metric | Pre. | Rec. | F1 | Pre. | Rec. | F1 | Pre. | Rec. | F1 |
| VAE | 0.50 | 0.91 | 0.65 | **0.94** | 0.76 | 0.85 | 0.89 | 0.95 | 0.92 |
| MemAE | 0.58 | 0.95 | 0.72 | 0.92 | 0.64 | 0.75 | 0.81 | **0.98** | 0.89 |
| CDF | **0.87** | **0.98** | **0.92** | 0.86 | **0.94** | **0.90** | **0.93** | **0.98** | **0.96** |

# AI 보안관제 2.0

- **AI for Security, 실전 업무 투입 가능?**
  - **아직은 좀…**
- 보안 실무 적용은 100% 정확도, 또는 명확한 **판단 근거** 요구
  - XAI (eXplainable AI)
- 과거 유사했던 사례/데이터 검색
  - ChatGPT 이슈

- Data-Driven Security
- AI for Security

MIT Technology Review
Published by DMK

인공지능

Why you shouldn't trust AI search engines

## AI 검색엔진을 신뢰하면 안 되는 이유

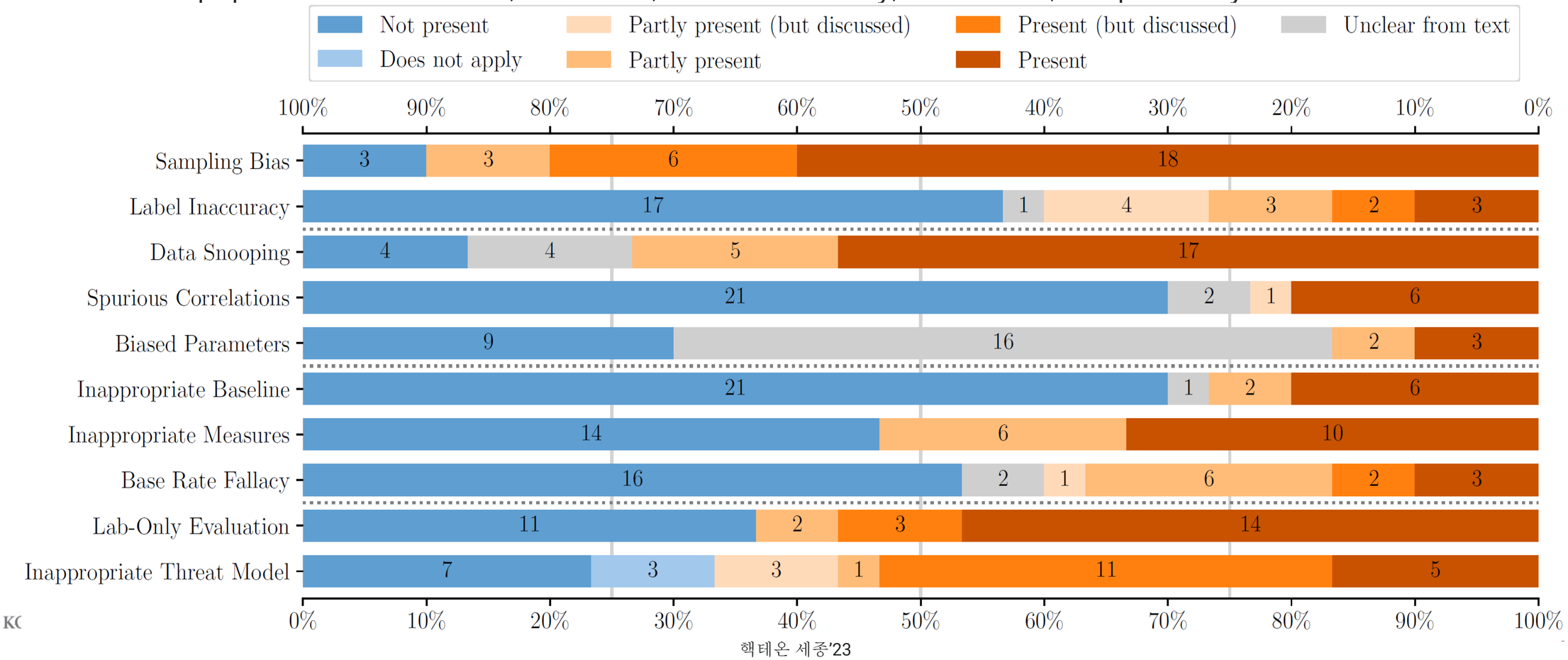
AI 언어모델은 종종 거짓을 사실처럼 제시하는, 헛소리를 내뱉는 기계로 악명이 자자하다. 사실 여부가 중요한 검색 기능과 결합할 경우 위험한 상황이 연출될 수 있다.

Melissa Heikkilä 2023년 3월 28일

빅테크가 최근에 범한 실수들이 AI 기반 검색엔진의 실패를 의미하지 않는다. 구글과 마이크로소프트가 그들의 AI가 생성한 결과를 더 정확하게 만들 수 있는 한 가지 방법은 인용문을 제공하는 것이다. 과거 구글 AI 윤리팀의 공동 책임자였으며 현재 AI 스타트업 허깅페이스(Hugging Face)의 연구원이자 윤리학자인 마가렛 미첼(Margaret Mitchell)은 출처를 연결하면 사용자들이 검색엔진이 어디서 정보를 가져왔는지 알 수 있게 될 것이라고 설명한다.

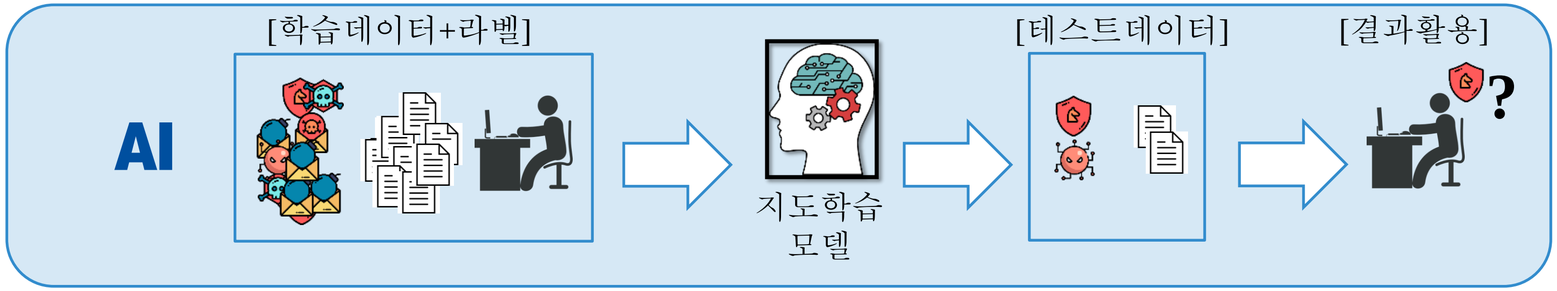
# AI 보안관제 2.0

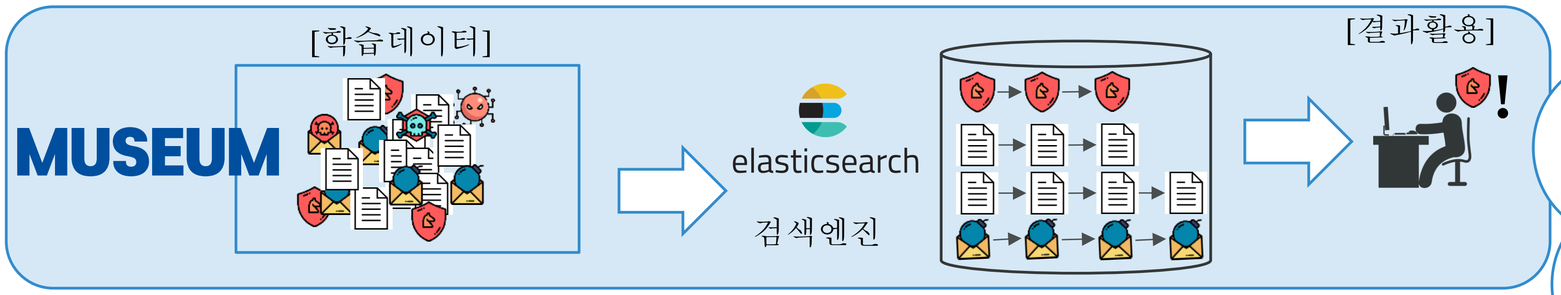
- 2020's: 인공지능 보안 분야 적용 시 유의사항
  - D. Arp, et al., "Dos and Don'ts of Machine Learning in Computer Security," USENIX Security'22
    - ✓30 papers from ACM CCS, IEEE S&P, USENIX Security, and NDSS, for past 10 years

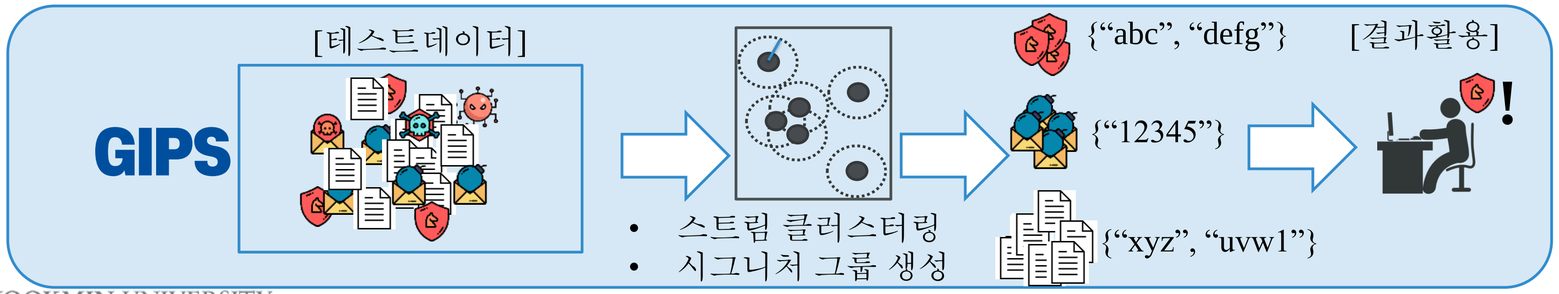


# AI 보안관제 2.0

- 악성/공격/이상 탐지 근거 제시 필요



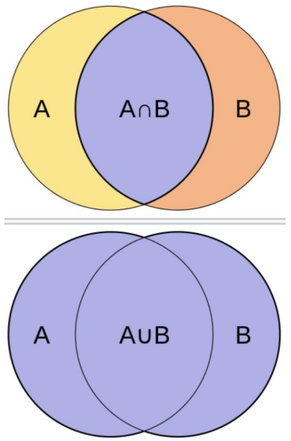



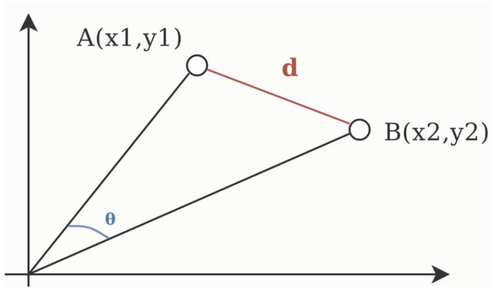

- MUSEUM: 악성코드 실시간 검색 기술 (과거 데이터)
- GIPS: 보안 이벤트 시그니처 그룹 생성 기술 (현재 데이터)

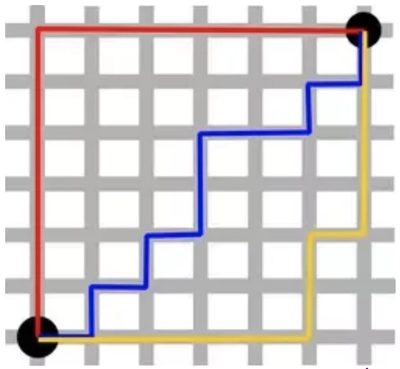
# MUSEUM: 악성코드 실시간 검색 기술

- 신기술 연구 필요성
  - 비슷한 파일(메시지, 롱스트링, …) 신속 정확하게 찾을 수 있는 확장성 있는 기술 필요
  - **파일**: 큰 파일, 크기 2배 이상 차이 파일도 처리 가능 (SSDEEP 한계 극복)
  - **신속**: pairwise comparison, O(n) → Inverted index, O(1)
  - **정확**: Jaccard similarity, Cosine Similarity, Euclidean distance, Minkowski distance, …
  - **확장**: Elasticsearch platform
- But, how to?

elasticsearch



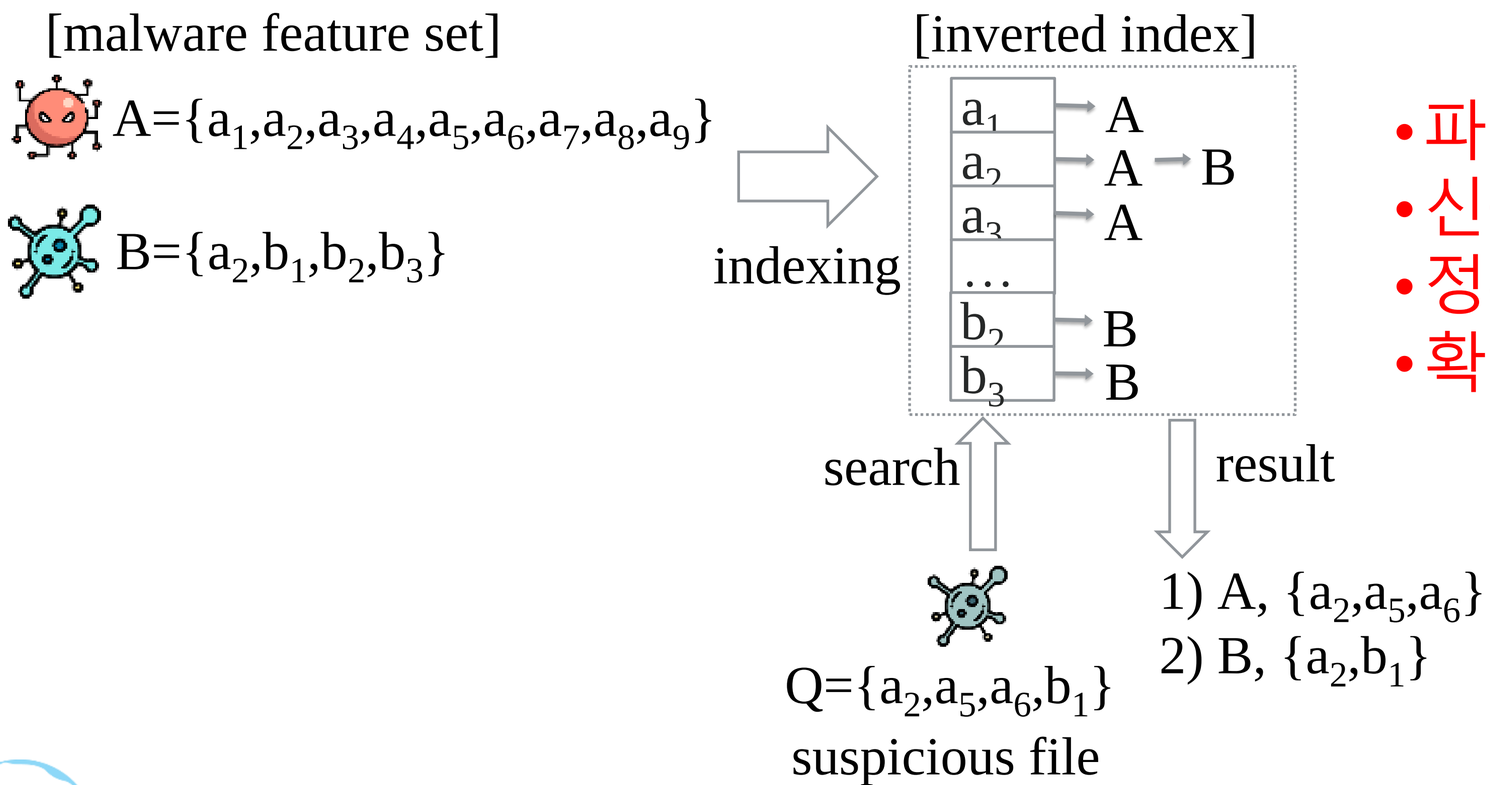



https://medium.com/@gshriya195/top-5-distance-similarity-measures-implementation-in-machine-learning-1f68b9ecb0a3

# MUSEUM: 악성코드 실시간 검색 기술

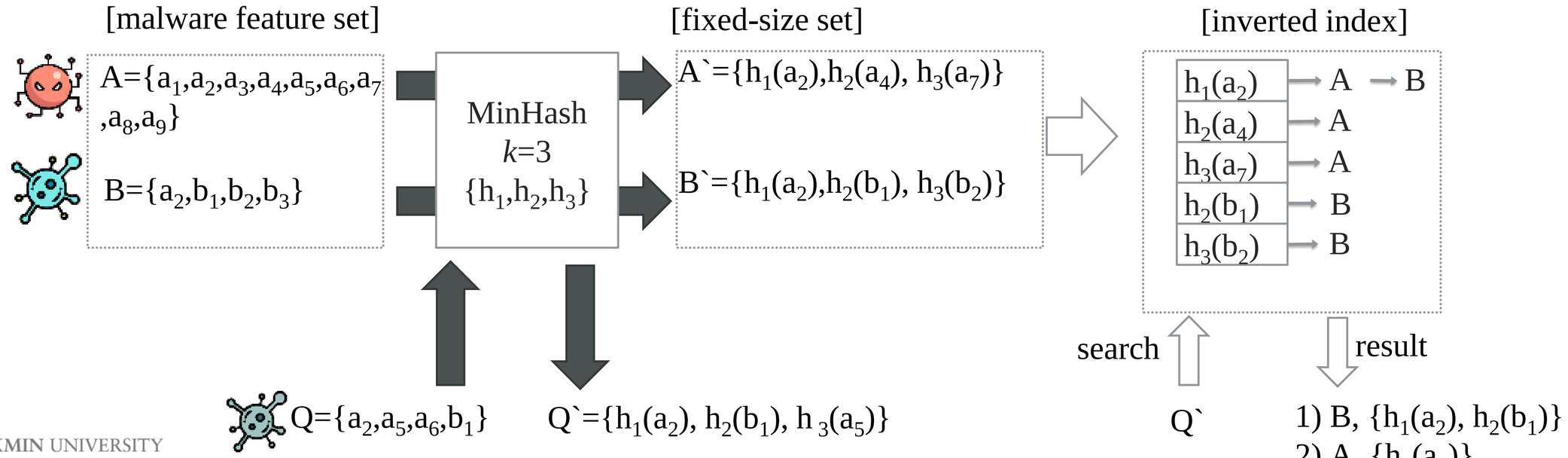
- Naïve approach
  - Elasticsearch for security big-data (ex: malware files)




- 파일 (X): 큰 파일, 크기 2배 이상 차이
- 신속 (O): Inverted index, O(1)
- 정확 (X): Jaccard similarity
- 확장 (O): Elasticsearch platform

# MUSEUM: 악성코드 실시간 검색 기술

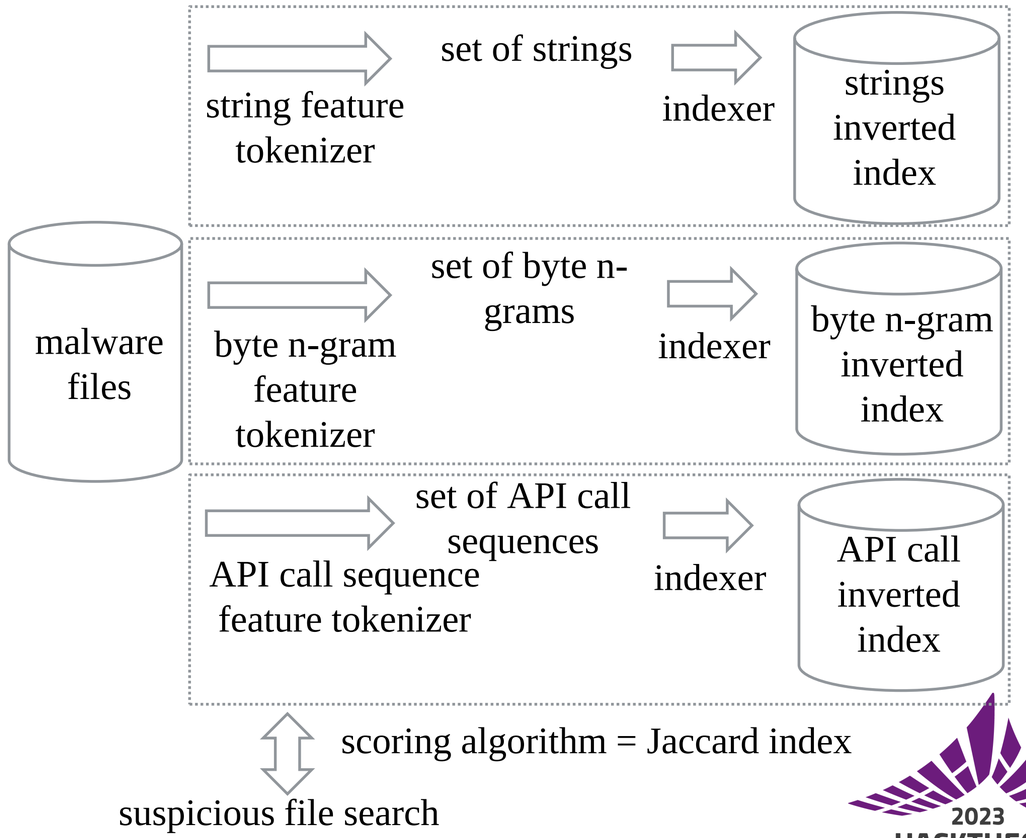
- Multifaceted-Search Engine Using MinHash sampling
  - D. Kim, J. Hur and M. Yoon, "Scalable and Multifaceted Search and Its Application for Binary Malware Files," in *IEEE Access*, vol. 9, pp. 112770-112779, 2021.
  - "**이분 그래프 기반의 악성코드 빅데이터 자동분석 플랫폼 연구**", 2018 디지털 금융혁신과 금융보안 공모전 대상 (금융감독원장상)
  - 특허 출원 (국내/미국)
  - Broder, Andrei Z. "On the resemblance and containment of documents." *Proceedings. Compression and Complexity of SEQUENCES 1997 (Cat. No. 97TB100171)*. IEEE, 1997.



# MUSEUM: 악성코드 실시간 검색 기술

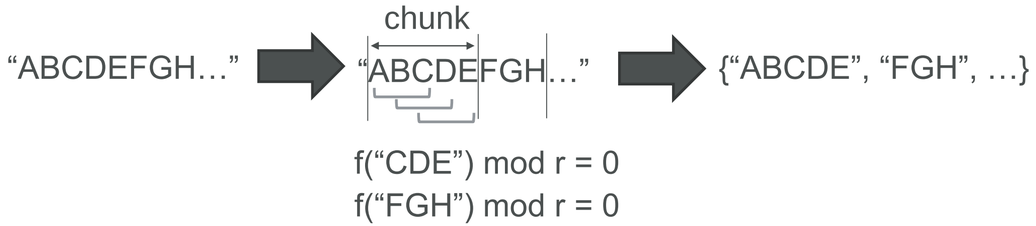
- Multifaceted-Search Engine Using MinHash sampling
  - 장점
    - **파일 (O): 큰 파일, 크기 2배 이상 차이**
    - **신속 (O): Inverted index, O(1)**
    - **정확 (O): Jaccard similarity**
    - **확장 (O): Elasticsearch platform**
    - **Multifaceted**
  - Advanced version
    - ✓MinHash: number of hash functions = 1, smallest *k*
    - ✓Min-Max hash: min(h(A)), max(h(A))
  - Rival scheme
    - ✓ssdeep for indexing

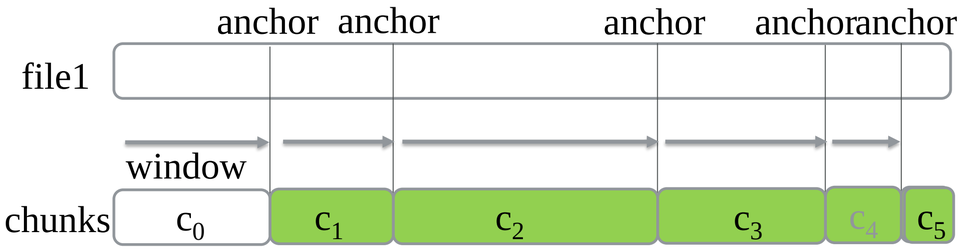
- Static analysis (PE header, disassembled code, strings, …)
- Dynamic analysis (Sandboxing, API sequence, …)

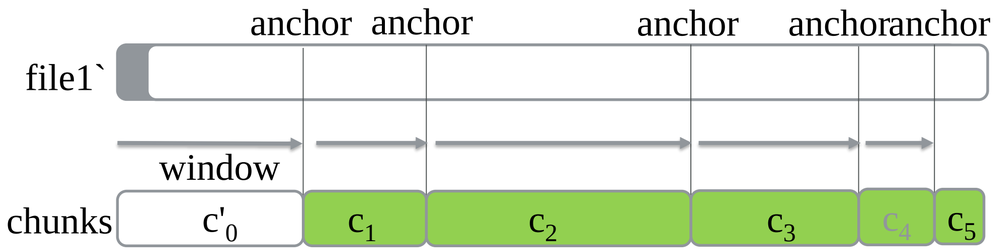


# MUSEUM: 악성코드 실시간 검색 기술

- Multifaceted-Search Engine Using MinHash sampling
  - **Sequence type only → set type**
    - ✓텍스트/ASCII/알려진 프로토콜 패킷 →NLP (Natural Language Processing) 토큰(token) 생성
    - ✓바이너리 → CDC(Content-Defined Chunking)




  - ✓DLP (Data Leakage Prevention) 직접 활용 가능
    - J. Hur, H. G. Shon, Y. J. Kim and M. Yoon, "Packet Chunking for File Detection," in IEEE/ACM Transactions on Networking, 2022

# MUSEUM: 악성코드 실시간 검색 기술

- 실험 결과

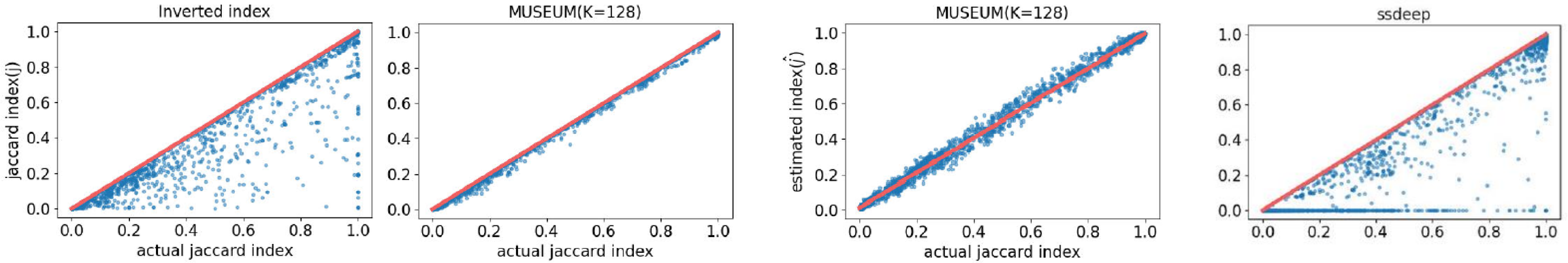



**FIGURE 4.** Accuracy comparison of inverted indexing and MUSEUM where dataset1 and the ASCII string feature are used. Each point represents a query file, whose $x$-coordinate is the true jaccard index between the query file and the most similar malware file; the $y$-coordinate of the left and middle plots is the jaccard index between the query file and the first ranked search result while the right plot uses the estimator of equation (4). If the search is perfect, all points would be on $y = x$.

| Type | | dataset1 | dataset2 |
|---|---|---|---|
| Features | | Byte stream (n-gram), Byte stream (AE), ASCII string | Byte stream (AE) |
| Indexing | Period | 2018.10.01 | 2018.10.01 ~ 2019.02.01 |
| | no.of files | 68,858 | 12,000,000 |
| Query | Period | 2018.10.02 | 2019.02.02 |
| | no.of files | 3,000 | 10,000 |

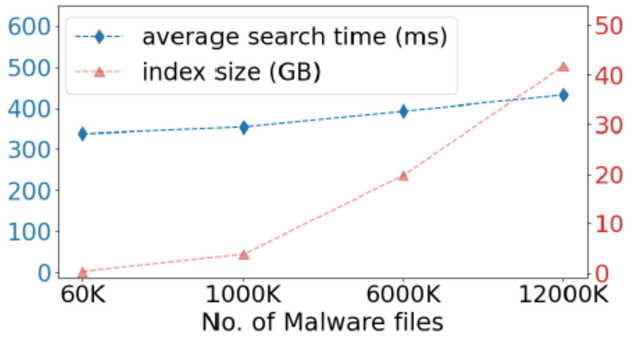



**FIGURE 6.** Comparison of throughput and indexing space with different collection sizes of malware files. Dataset1, 2 and byte stream (AE chunking) feature are used.

# GIPS: 보안 이벤트 시그니처 그룹 생성 기술

- 실시간 공통 시그니처 생성을 통한 근거 제시
  - 제로데이 공격 방어
  - 공격 탐지 근거 제시 (설명 가능 일부 충족)

| 구분 | | 탐지 정확성 | 자동화 | 제로데이 공격 방어 | 비고 |
|---|---|---|---|---|---|
| 오용탐지 (misuse detection) | 수작업 탐지규칙 생성 | 상 | 하 | 하 | Snort, 상용장비, 최다 활용 기술 |
| | 탐지 시그니처 추출/생성 | 중 | 중 | **상** | Early Bird, Triple-heavy-hitter, **단기 집중 반복 공격만** 탐지 가능 (예: Worm, DDoS) |
| 이상탐지 (anomaly detection) | 지도학습 | 중 | 중 | 중 | Classification, CNN/RNN, Random Forest, 연구 활발 |
| | 비지도학습 | 하 | 중 | **상** | Clustering, 오토인코더, Isolation Forest, **정확성 한계** |

$d_0$="Sdh$9DkhttpFm"
$d_1$="httpjhh57^dhgfnmFG"
$d_2$="httpabroot12admin74"
$d_3$="httpHf8root121d34admin@"
$d_4$="http&dU4rootGyadminA2"
$d_5$="pdh%dh9*dhg$fhrdO3"
⋮

(a) 데이터 스트림

$d_0$="Sdh$9DkhttpFm"
$d_1$="httpjhh57^dhgfnmFG"
$d_2$="httpabroot12admin74"
$d_3$="httpHf8root121d34admin@"
$d_4$="http&dU4rootGyadminA2"
$d_5$="pdh%dh9*dhg$fhrdO3"
⋮

(b) 수작업 탐지규칙 생성

$d_0$="Sdh$9DkhttpFm"
$d_1$="httpjhh57^dhgfnmFG"
$d_2$="httpabroot12admin74"
$d_3$="httpHf8root121d34admin@"
$d_4$="http&dU4rootGyadminA2"
$d_5$="pdh%dh9*dhg$fhrdO3"
⋮

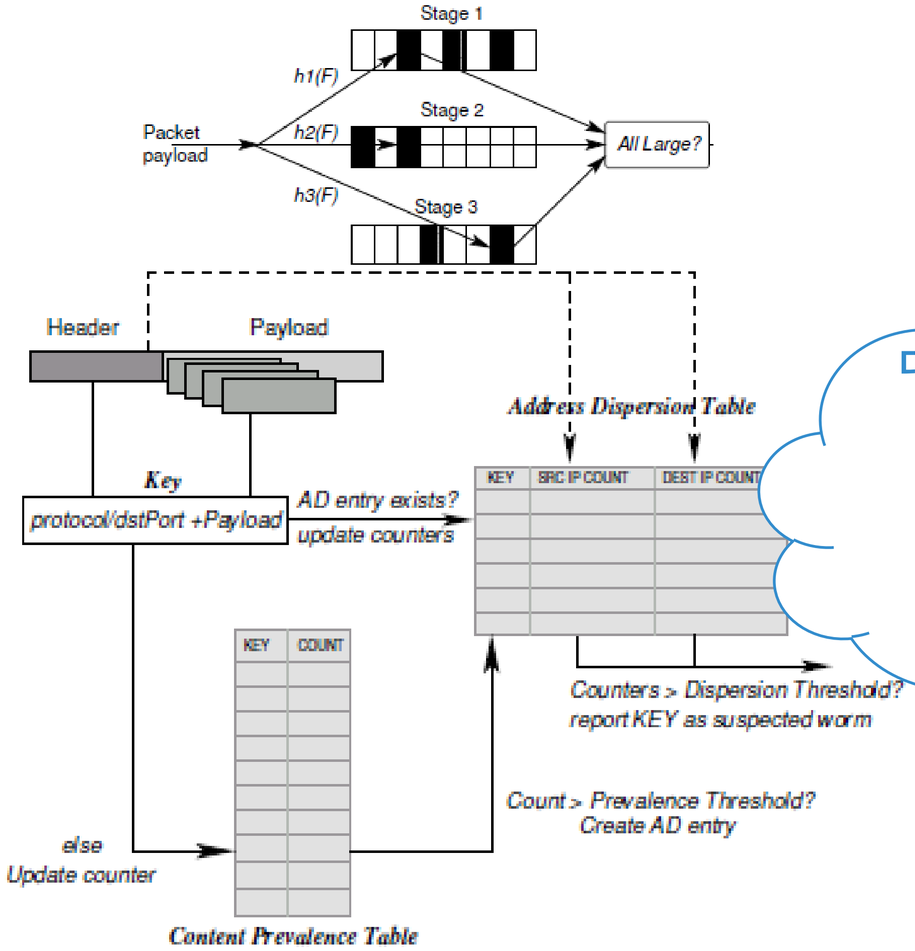
(c) 탐지 시그니처 추출/생성

$d_0$="Sdh$9DkhttpFm"
$d_1$="httpjhh57^dhgfnmFG"
$d_2$="httpabroot12admin74"
$d_3$="httpHf8root121d34admin@"
$d_4$="http&dU4rootGyadminA2"
$d_5$="pdh%dh9*dhg$fhrdO3"
⋮

(d) 이상적 기술 (실시간 처리, 다수 시그니처 조합 추출/생성)

# GIPS: 보안 이벤트 시그니처 그룹 생성 기술

- "시그니처 추출/생성" 기존 기술 및 한계

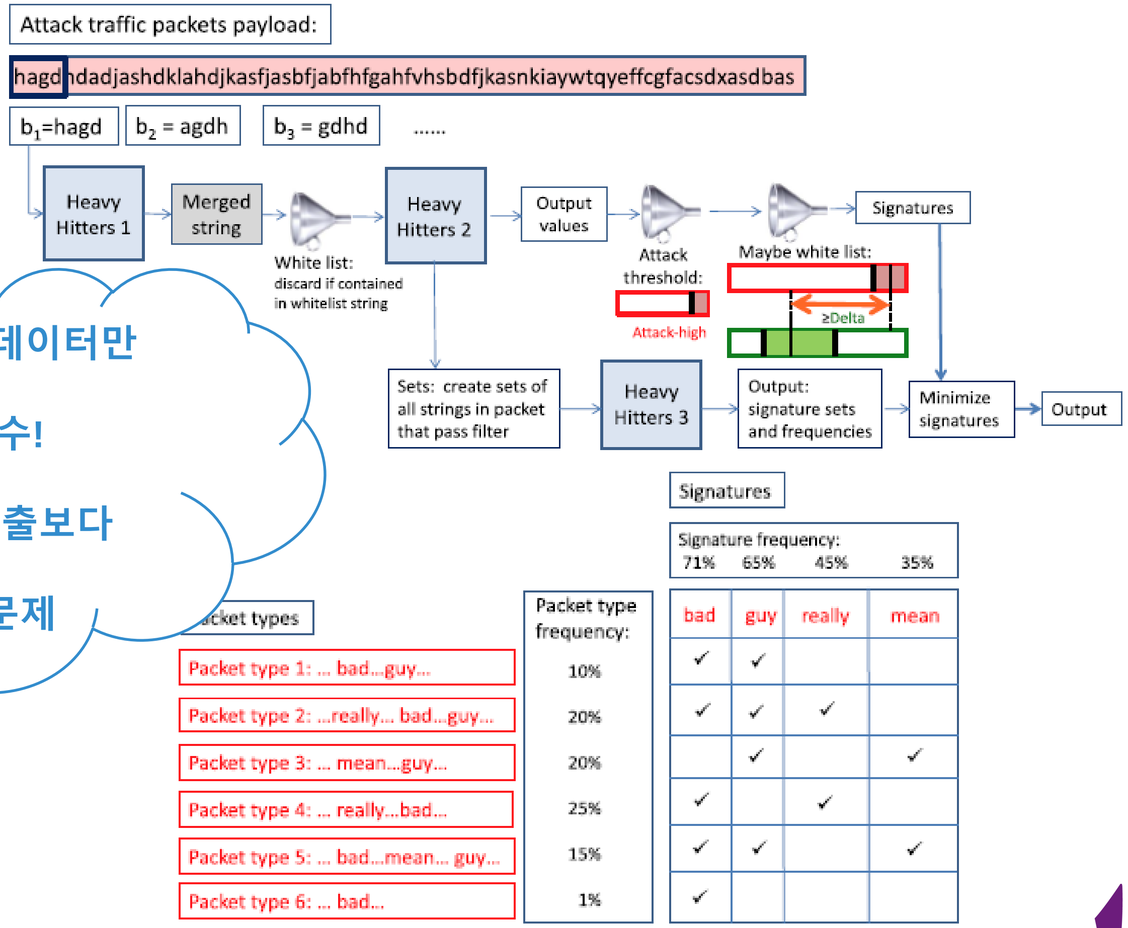



대량의 유사 데이터만

수집 필수!

시그니처 추출보다

어려운 문제

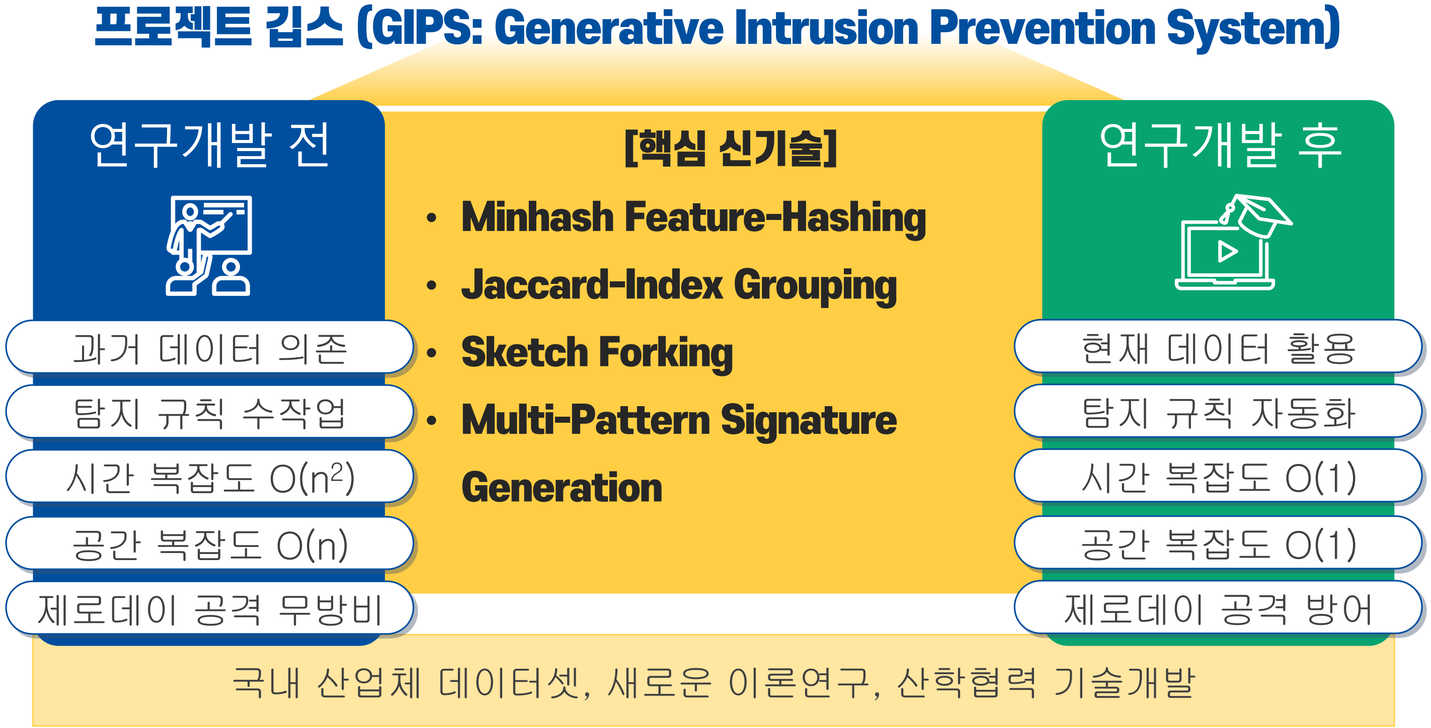



Singh, S., Estan, C., Varghese, G., & Savage, S. (2004, December). Automated Worm Fingerprinting. In *OSDI* (Vol. 4, pp. 4-4).

Y. Afek, A. Bremler-Barr and S. L. Feibish, "Zero-Day Signature Extraction for High-Volume Attacks," in IEEE/ACM Transactions on Networking, vol. 27, no. 2, April 2019

# GIPS: 보안 이벤트 시그니처 그룹 생성 기술

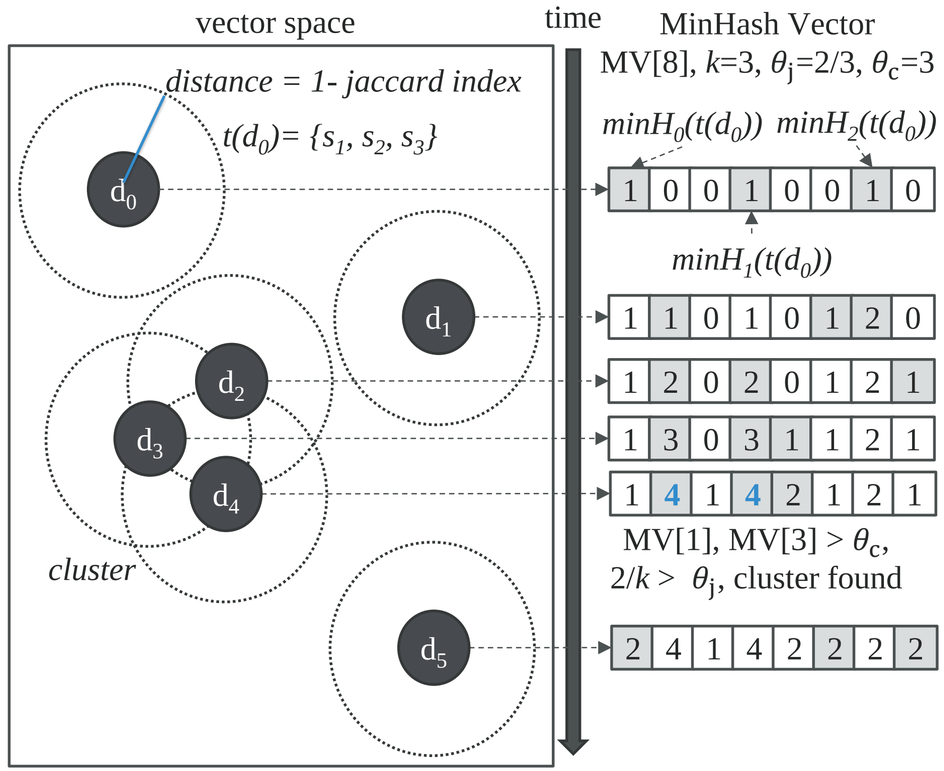
- GIPS: Generative Intrusion Prevention System
  - HyungBin Seo and MyungKeun Yoon, “Generative Intrusion Detection and Prevention on Data Stream,” under review
  - 제로데이 방어를 위한 생성형 보안기술 연구, 한국연구재단 중견연구과제, 2023.3~2026.2, PI 윤명근

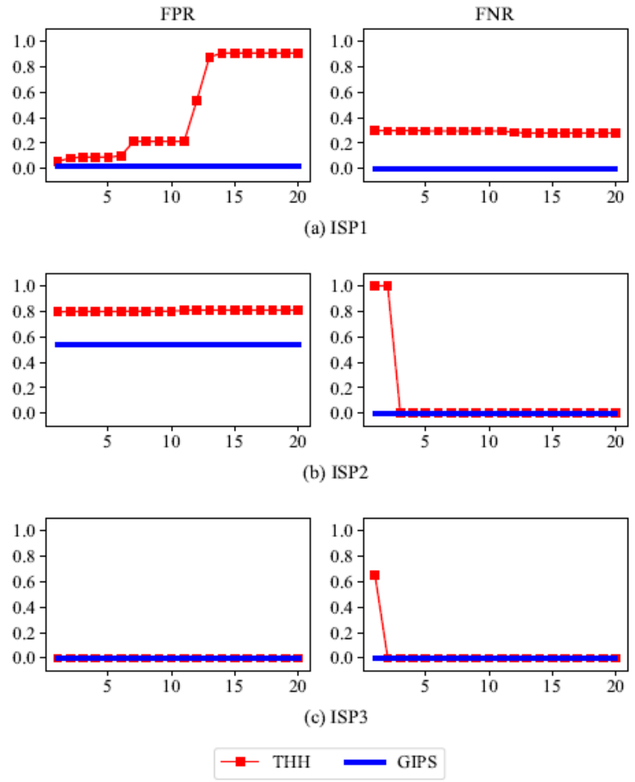


# GIPS: 보안 이벤트 시그니처 그룹 생성 기술

- GIPS: Generative Intrusion Prevention System
  - HyungBin Seo and MyungKeun Yoon, “Generative Intrusion Detection and Prevention on Data Stream,” under review, 특허출원

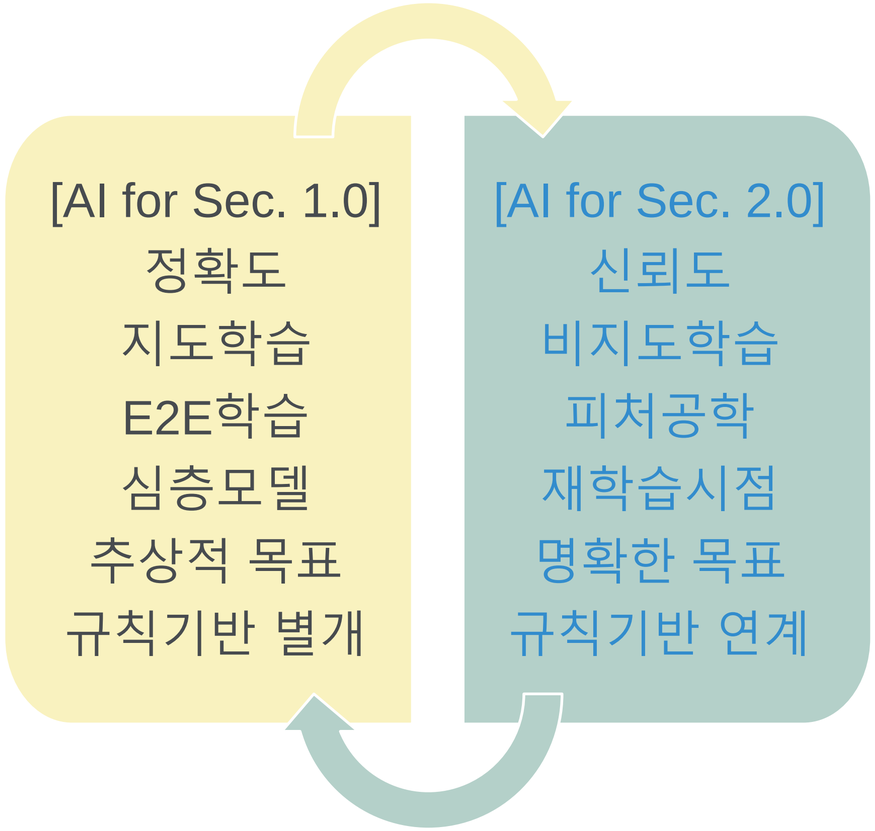
$d_0$=“Sdh$9DkhttpFm”
$d_1$=“httpjhh57^dhgfnmFG”
$d_2$=“httpabroot12admin74”
$d_3$=“httpHf8root121d34admin@”
$d_4$=“http&dU4rootGyadminA2”
$d_5$=“pdh%dh9*dhg$fhrdO3”
⋮



# 결론

- **AI=데이터기반 알고리즘=도구≠목적**
  - AI/데이터분석 이용한 업무 (반)자동화
  - 규칙기반 문제해결 알고리즘
- AI 보완 기술 적극적 활용
  - 탐지/판단 근거 생성형 보안기술 필요
  - 제로데이 공격 대응 가능
- 보안도메인 지식 반영 필수
- *"...the strength of machine-learning tools is finding activity that is similar to something previously seen..."*
  - R. Sommer and **V. Paxson**, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," *2010 IEEE Symposium on Security and Privacy*, 2010




# Q & A

# 감사합니다!

mkyoon@kookmin.ac.kr

https://infosec.kookmin.ac.kr