2020 IT 21
Global Conference
Digital New Deal
Technology Essentials
디지털 뉴딜 기술 핵심

Session 5-5

제어시스템 보안 데이터셋 HAI 1.0

김형천 실장 (국가보안기술연구소)

[요약문]

주요 기반시설을 구성하고 있는 제어시스템을 대상으로 하는 보안위협은 지속적으로 증가하고 있으며, 이에 대응하기 위한 가장 좋은 방법으로 인공지능(AI) 및 머신러닝(ML) 기반의 위협탐지 방법이 우선적으로 검토되고 있다. 이러한 AI 및 ML 기반의 연구를 위해서는 관련 데이터셋이 반드시 필요하지만, 제어시스템 운영환경의 특성을 반영하고 다양한 유형의 공격상황이 포함된 데이터셋은 매우 부족한 상황이다.

이에 여러 제조사의 보일러, 터빈, 양수시스템, DCS, PLC 등으로 구성된 테스트베드 기반으로 제어시스템 운영환경의 다양한 운전상황을 시뮬레이터를 연계하여 수집한 정상운전 데이터와 다양한 유형의 공격상황을 재현한 공격 데이터를 포함하는 HAI 데이터셋을 개발하고 공개하였다.

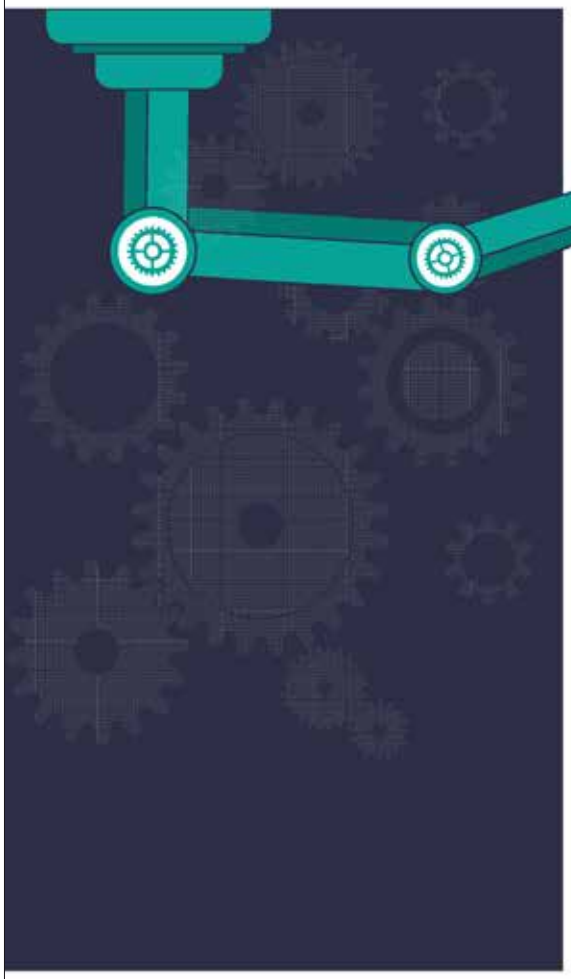
본 발표에서는 HAI 1.0 데이터셋에 대한 자세한 설명과 활용에 대해 설명한다.

[발표자 약력]

1999년 고려대학교 전산학과 학사
2001년 고려대학교 전산과학 전공 석사
2011년 고려대학교 정보보호 전공 박사
2001년 ~ 현재 국가보안기술연구소 책임연구원
2011년 ~ 현재 국가보안기술연구소 실장

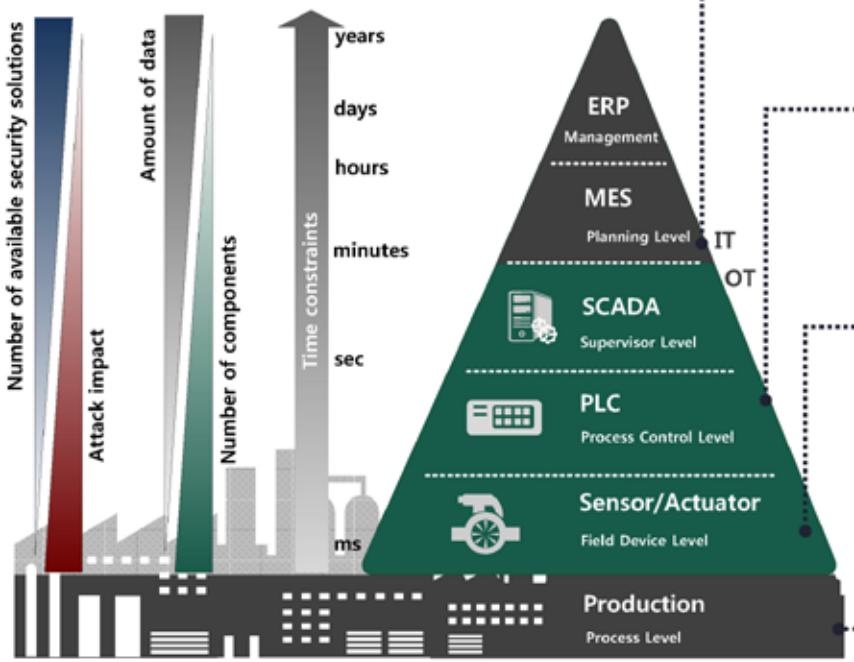
관심분야 : 소프트웨어 보안, 클라우드 컴퓨팅 보안, 운영체제 보안, 인공지능 기반 보안위협 탐지


제어시스템 보안 데이터셋
HAI 1.0



CONTENTS

ICS 보안현황



연결접점 증가 | 다양한 보안위협에 노출

- 생산성 향상 기술: Digital Transformation, Smart factory, Big data, Cloud, AI
- 연결성 증가: 데이터 유출·사이버 공격에 노출

제조사 특화 | 범용 보안장비 사용한계

- 제조사별 다양한 프로토콜 사용
- 제어로직 구현 도구·방식의 차이
- 가용성 최우선에 따른 보안 솔루션 적용 한계

IIoT 도입 증가 | 잠재적 보안위협 증가

- 다종 IoT 장치 → 취약점 분석 한계
- 다수 IoT 장치 → 취약점 패치 문제
- 저사양 IoT 장치 → 보안 솔루션 적용 한계

대규모화·복잡화 | 피해 파급력 증가

- 사이버-물리영역 간의 상호작용의 복잡도 증가
- 상호작용 연결고리를 통한 물리적 피해 가능

ICS 보안현황

"1년간 6~10회 공격을 받은 기관 수가 2017년에 비해 2019년 200% 증가"

(출처 : SANS 2019 State of OT/ICS Cybersecurity Survey)

우크라이나 대정전 사태(15.12)

BlackEnergy3

- APT 공격 수행
- IT망 경유 OT망 침투

다양한 환경, 이기종 기기 대상
다수 사이트 장기간 공격

제어시스템 웜(16.3)

PLC-Blaster

- 제어기기 대상 공격 및 전파

제어기기 및 현장장치 직접 공격
Hidden Interaction Chain

제어시스템 랜섬웨어 (17.2)

LogicLocker

- Cross-vendor
- 수평적/수직적 네트워크 전파 가능

IT영역 위협의 OT영역 전이
제어시스템 전 영역 대상 확산

ICS 보안현황

상호작용 연결고리를 통한 피해 발생 가능

Hidden Inter-app Interaction Chain

162 가지

어플리케이션간
직·간접적 상호작용
연결고리 존재

High Risky Interaction Chain

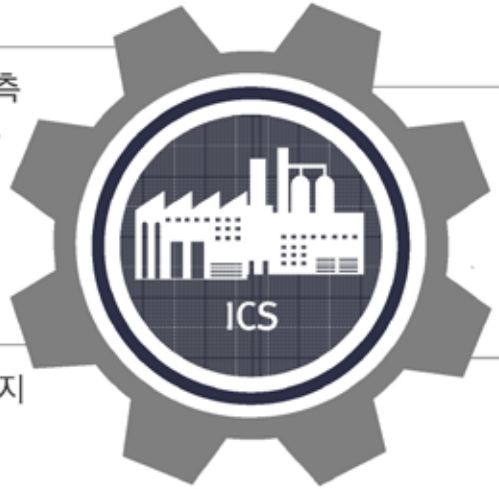
32 가지

현실적인 피해 발생 가능한
위협적인 상호작용
연결고리 존재

* W. Ding and H. Hu,"On the Safety of IoT Device Physical Interaction Control", CCS 2018

인공지능 기반 ICS 보안연구

현장 데이터 활용 및 인공지능 솔루션 도입 증가



스마트 통합관제

- 정확한 비즈니스 성과 예측
- 안전 및 운영 리스크 감소
- 현장 파악 투명성 확보

운영 최적화

- 현 운영 상태 실시간 시각화
- 운전정보 기반 성능 최적화 운전
- 자동화 장비로 생산성 증대

이상 감지

- 과거 데이터 기반 이상 조기 감지
- 주요 상태(진동,온도, 압력 등) 감시/진단 분석으로 사고 예방

예지 정비

- 잔여 수명 예측으로 유지보수 효율화
- 에너지 효율 최적화
- 정비인력 관리 효율화

인공지능 사용 목적

딥러닝 기술 적용하기 전과 후는
"청동기 시대와 철기 시대와 같다."

– NAVER LABs 김정희 부장 –

자동화

- 데이터만으로 컴퓨터가 스스로 특징을 만들어 냄
- 수많은 반복학습을 통해 스스로 판단

정확성

- 확률을 높이고 오차를 줄임
- 음성/영상 인식, 자연어 처리 등에서 최고 성능 제공

인공지능 사용 목적

Google

이미지 인식

28% vs. 5%

Error Rate : 2012 vs 2015

Google

음성 인식

23% vs. 8%

Error Rate : 2013 vs 2015

facebook

물체 인식

30% ⇧

출처: Google, Facebook 해당 기사 자료 정리

인공지능 연구 국외현황

빅데이터 보유 기업이 인공지능 기술 선도

Alphabet

- 2001년부터 AI 기술 확보 33조원 투자
- 자율주행,사진,음성,스마트홈에 AI 적용

facebook

- 딥스페이스 : 얼굴 인식 알고리즘 개발
- 가상비서 서비스'M '공개

Microsoft

bing Azure OneDrive

- 인공지능 개인비서 코타나 서비스
- 딥러닝 기반 채팅봇 테이 개발

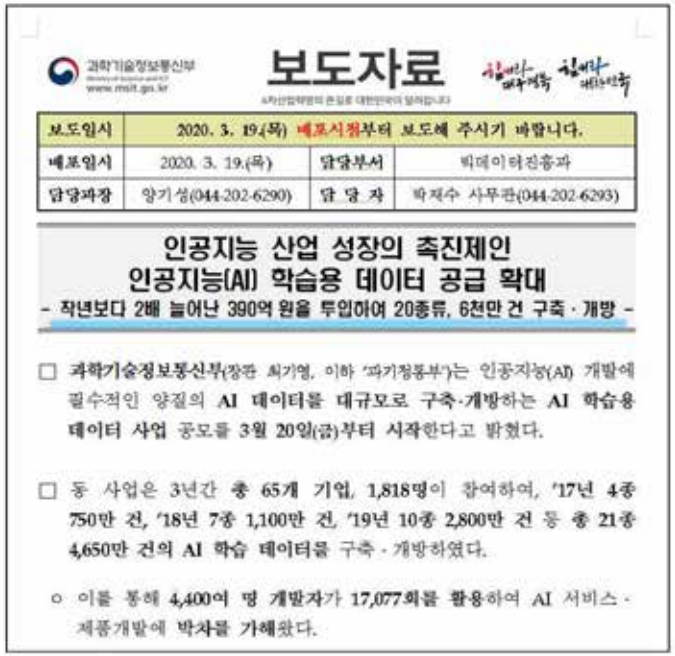
IBM

- 자체보유 빅데이터가 부족한 IBM
- 빅데이터 보유 기업들과 M&A 제휴

- 기상예측,이미지인식 및 감성분석
- EMR 데이터분석을 통한 치료법 권고

인공지능 연구 국내현황

국가주도 인공지능 학습용 데이터 사업 지원 확대



과학기술정보통신부 www.msit.go.kr

보도자료

보도일시	2020. 3. 19.(목) 배포시점부터 보도해 주시기 바랍니다.		
배포일시	2020. 3. 19.(목)	담당부서	빅데이터진흥과
담당과장	양기성(044-202-6290)	담 당 자	박재수 사무관(044-202-6293)

인공지능 산업 성장의 촉진제인
인공지능(AI) 학습용 데이터 공급 확대
- 작년보다 2배 늘어난 390억 원을 투입하여 20종류, 6천만 건 구축·개방 -

□ **과학기술정보통신부**(장관 최기영, 이하 '과기정통부')는 인공지능(AI) 개발에 필수적인 양질의 **AI 데이터**를 대규모로 구축·개방하는 **AI 학습용 데이터 사업** 공모를 **3월 20일(금)부터 시작**한다고 밝혔다.

□ 동 사업은 3년간 **총 65개 기업, 1,818명**이 참여하여, **'17년 4종 750만 건, '18년 7종 1,100만 건, '19년 10종 2,800만 건** 등 **총 21종 4,650만 건의 AI 학습 데이터**를 구축·개방하였다.

○ 이를 통해 **4,400여 명 개발자**가 **17,077회**를 **활용**하여 AI 서비스·제품개발에 **박차를 가해**왔다.

인공지능의 핵심 요소

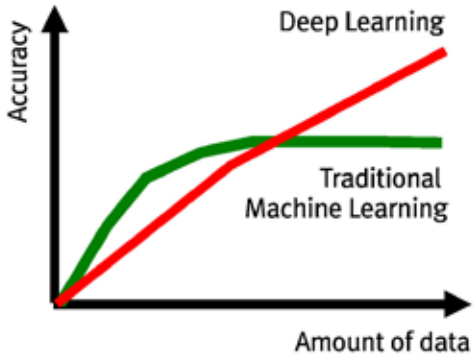
양질의 빅데이터

- 학습을 위한 재료로 학습 결과의 품질을 결정
- 딥러닝: 빅데이터에 특화된 알고리즘

+

평가 방법

- 분별 가능한 것 : 사물 구분, 이미지 조작, 번역 등
- 정해진 규칙 : 체스, 바둑, 게임 등



Reddit

3.7 billion comments from online discussions on many topics

↓

726 million examples

OpenSubtitles

over 400 million lines of subtitles from movies and TV

↓

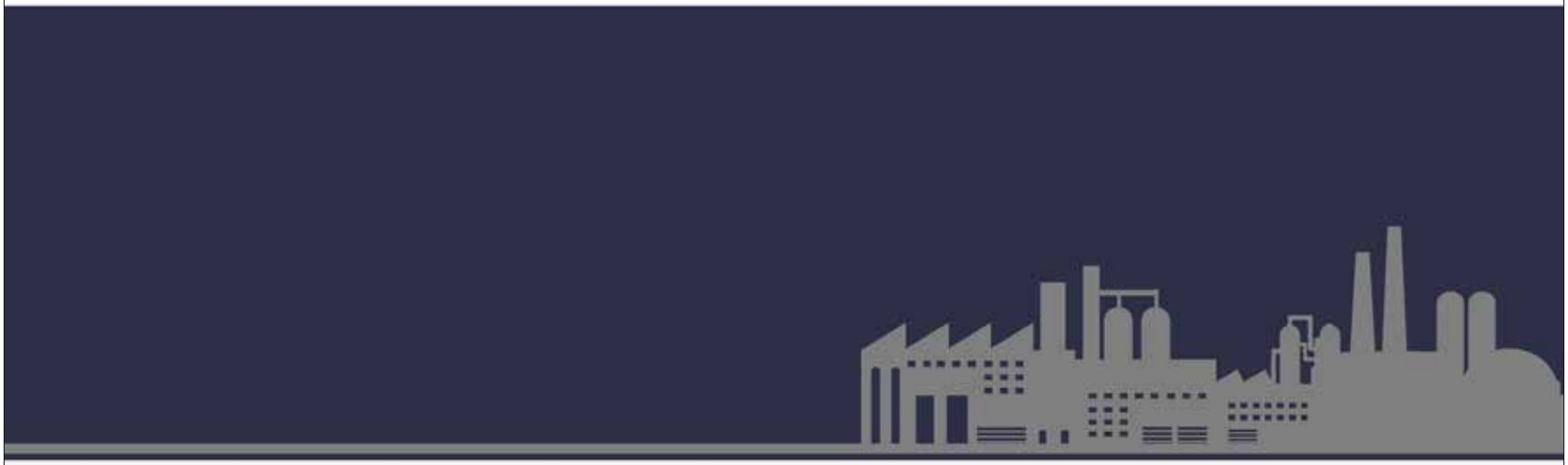
316 million examples

AmazonQA

over 3.6 million product question answers

↓

3.6 million examples



ICS 보안 데이터셋 HAI v1.0

HIL-based Augmented ICS

ICS 보안연구의 어려움: 데이터 부재

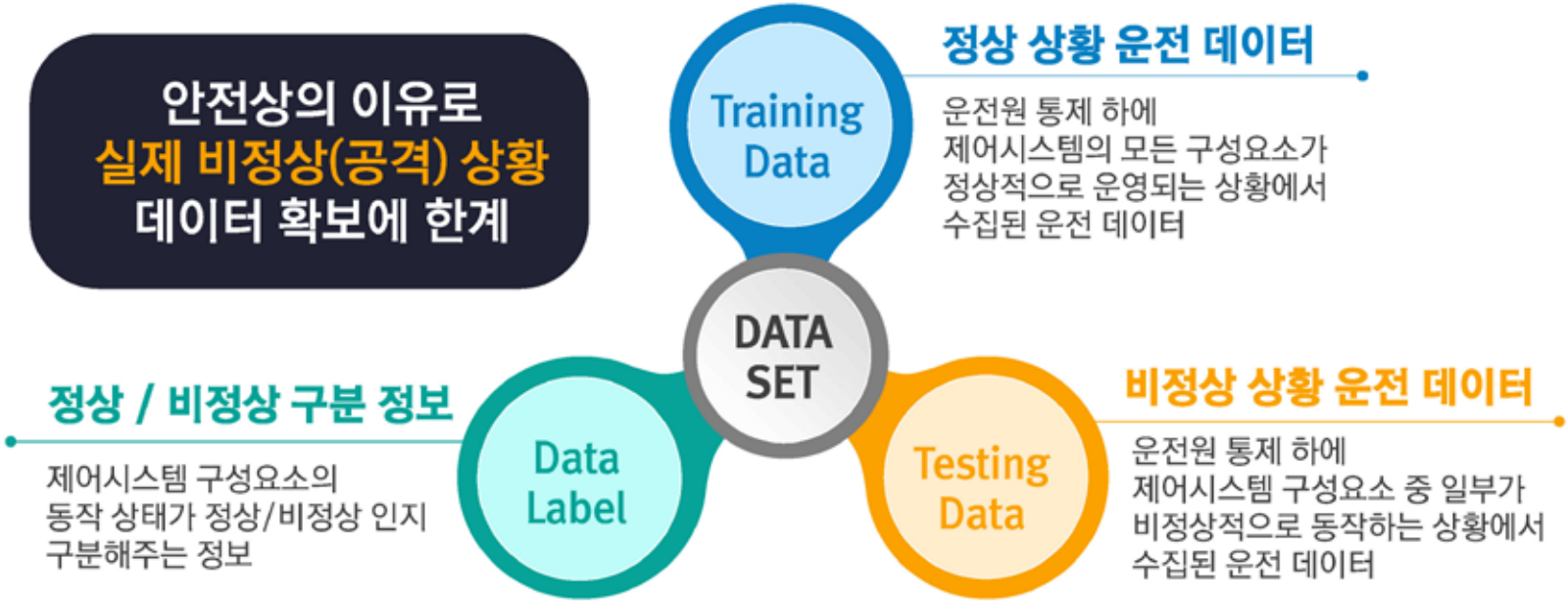
실제 제어시스템 대상

- 안전 또는 보안의 이유로 비공개 등의 접근 제약
- 실제 위협 상황에 대한 실증 불가능
- 비협조적인 제조사

테스트베드 대상

- 구축/관리를 위한 많은 비용 발생
- 전문인력을 지속적으로 배치하기 어려움
- 연구 단계에서 상시 활용 제한

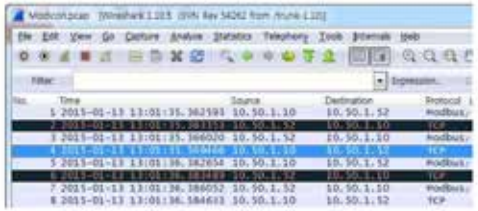
ICS 보안 데이터셋

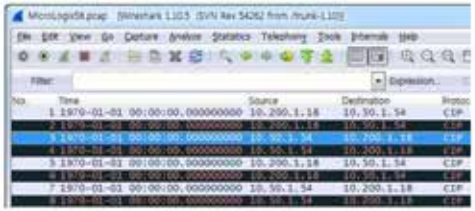

안전상의 이유로
실제 비정상(공격) 상황
데이터 확보에 한계
Training Data
정상 상황 운전 데이터
운전원 통제 하에
제어시스템의 모든 구성요소가
정상적으로 운영되는 상황에서
수집된 운전 데이터
DATA SET
Data Label
정상 / 비정상 구분 정보
제어시스템 구성요소의
동작 상태가 정상/비정상 인지
구분해주는 정보
Testing Data
비정상 상황 운전 데이터
운전원 통제 하에
제어시스템 구성요소 중 일부가
비정상적으로 동작하는 상황에서
수집된 운전 데이터

ICS 보안 데이터셋 개발 고려사항

시간 동기화

- 계층적 데이터 전송 과정에서 시간 지연 발생
- 디바이스 간 시간 동기화가 안됨





데이터 라벨링

- 정상과 비정상의 정확한 구분 필요
- 학습 및 탐지 성능을 결정하는 핵심 정보

시스템 응답 안정화

- 초기 기동, 정비 등 불안정 응답 구간 존재
- 공격 후 시스템 응답 안정화 필요

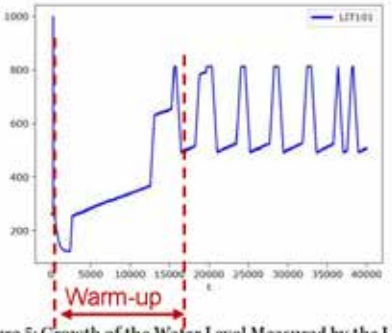


Figure 5: Growth of the Water Level Measured by the LIT101 Sensor During the Initial Warm Up Period

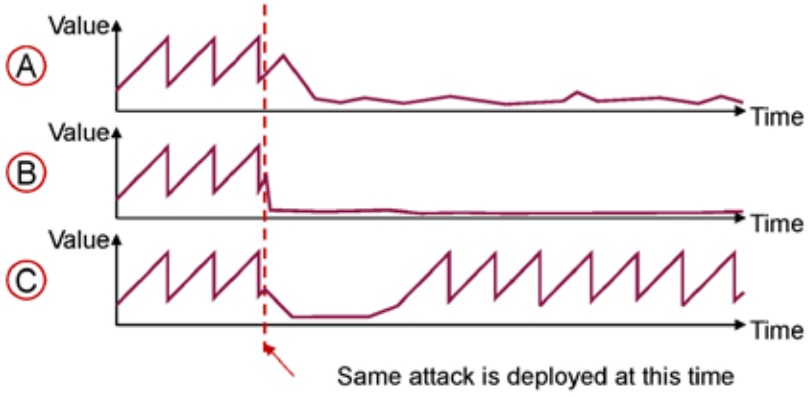
Kravchik, M., Shabtai, A.: Detecting Cyberattacks in Industrial Control Systems Using Convolutional Neural Networks ArXiv e-prints (Jun 2018)

ICS 보안 데이터셋 개발 고려사항

동일 공격 반복 수행 필요

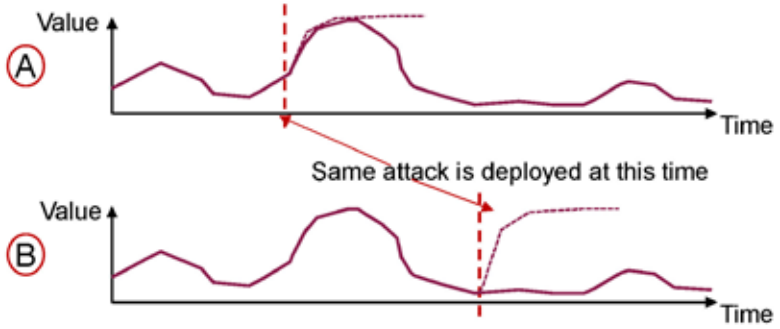
공격 시점 외부 환경요인에 따라 시스템 응답 패턴 상이

(온·습도, 압력, 발전 부하량 등)

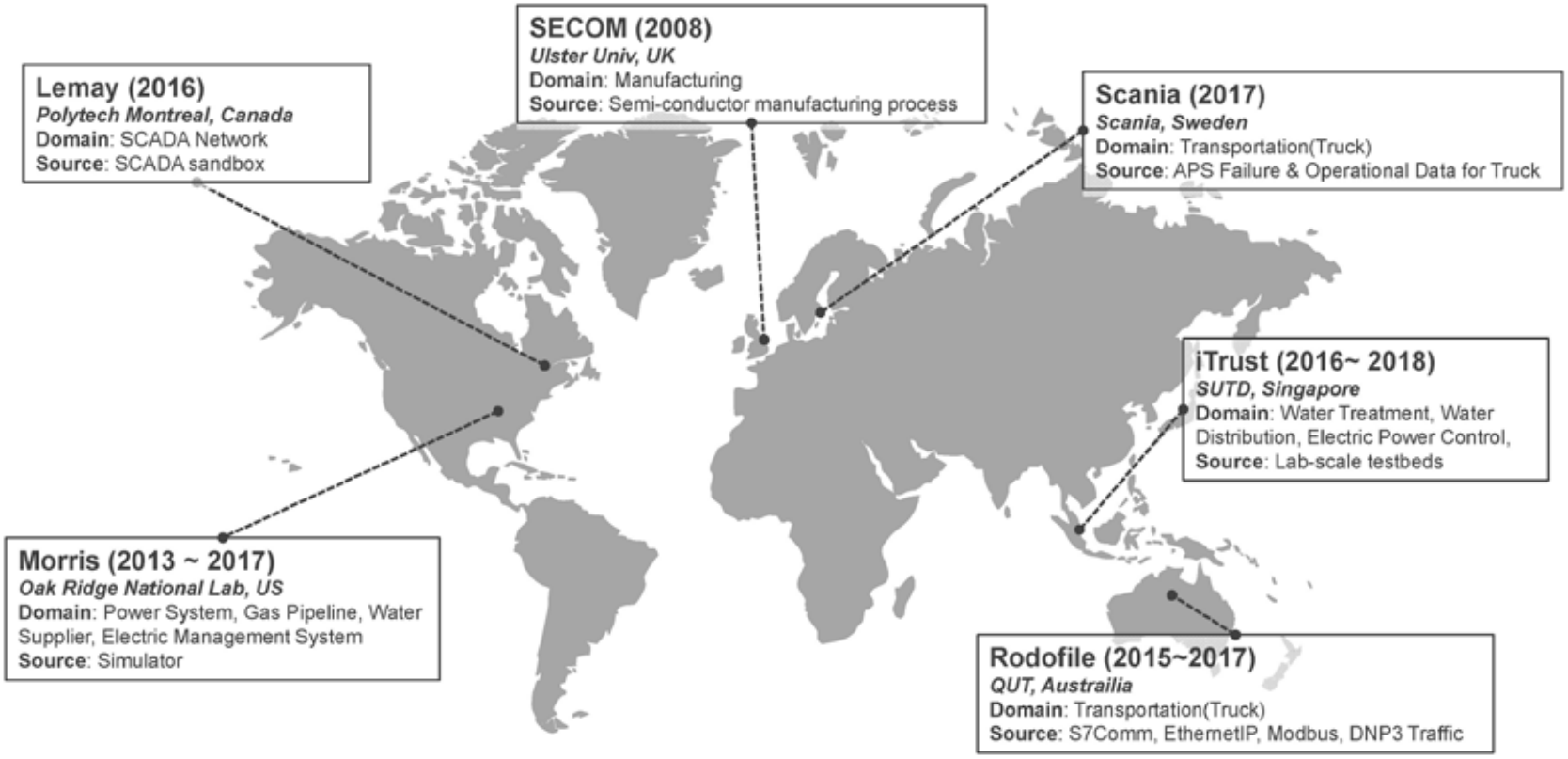


공격 시점 내부 상태에 따라 시스템 응답 패턴 상이

(압력, 속도, 유압 등)



ICS 보안 데이터셋 국외 현황



ICS 보안 데이터셋 국외 현황

#	Dataset ID	URL
1	Morris	https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets
2	iTrust	https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/
3	Scania	https://archive.ics.uci.edu/ml/datasets/APS+Failure+at+Scania+Trucks
4	SECOM	https://archive.ics.uci.edu/ml/datasets/SECOM
5	Rodofile	https://github.com/qut-infosec/2017QUT_S7comm
6	Lemay	https://github.com/antoine-lemay/Modbus_dataset
7	ICS-pcap	https://github.com/automayt/ICS-pcap
8	4SICS	https://www.netresec.com/?page=PCAP4SICS
9	S4x15CTF	https://www.netresec.com/?page=DigitalBond_S4
10	DEFCON23	https://media.defcon.org/DEF%20CON%2023/DEF%20CON%2023%20villages/DEF%20CON%2023%20ics%20village/DEF%20CON%2023%20ICS%20Village%20packet%20captures.rar

ICS 보안 데이터셋: iTrust (SUTD: Singapore Univ. of Tech. and Design)

SWaT
Secure Water Treatment

- **정수시설 테스트베드**
  - 6개의 PLC, 6단계 정수처리
- **약 11일치 데이터 제공**
  - 36종 공격 수행, 51개 포인트/초
- **현재 가장 널리 활용되는 데이터셋**
  - 유명학회(CCS, S&P, NDSS, CCS 등)
  - 공격 의도/영향력 분석에 한계

WADI
Water Distribution

- **배수시설 테스트베드**
  - 2개의 PLC, 6개 소비용 물탱크 배수
  - SWaT의 정수 사용
- **약 16일치 데이터 제공**
  15종 공격 수행, 103개 포인트/초

EPIC
Electric Power and Intelligent Control

- **전력망 테스트베드**
  - 발전-송전-마이크로그리드-스마트홈
  - 최대 30kW 발전 (10kWx3기)
  - SWaT, WADI 전원 공급
- **8가지 구동 시나리오**
  - 정상 상황만 각 30분 수집

ICS 보안 데이터셋 활용

연구 개발에 활용

- 학습 기반 기술 연구의 핵심
- 자체 테스트를 통한 문제점 개선

기술/제품 성능비교

- 기술/제품별 장단점 분석
- 동일환경에서 객관적 성능 비교

보안제품 요구사항 제시

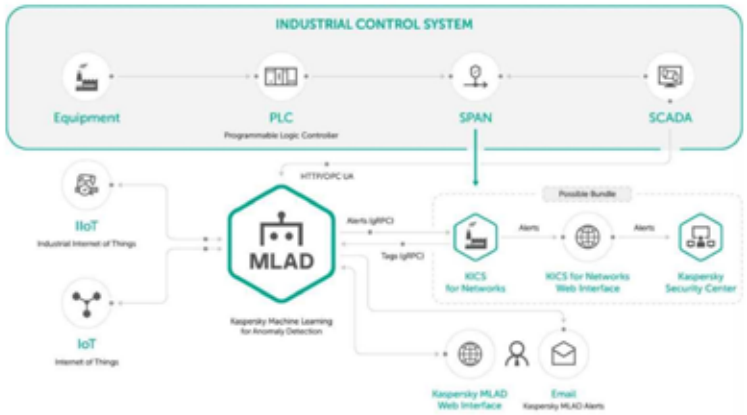
- 보호 대상 시스템의 정보 공개 문제 해소
- 유사 데이터셋으로 최소한의 성능요구 가능

Kaspersky의 SWaT 데이터 활용

"Anomaly Detection for Water Treatment System based on Neural Network with Automatic Architecture Optimization", ICML 2018

2018년도 연구논문 발표

MLAD : Machine Learning for Anomaly Detection



2019년도 Hannover Messe* 전시

* R&D, 산업 자동화, IT, 산업공급, 생산기술, 서비스 및 에너지, 환경 기술산업을 위한 전시회

HAI 개발 목표

공개된 보안 데이터셋의 한계 극복을 통한 신뢰성 있는 데이터셋

수작업에 의존적

- 장기간 연속적인 공격이 어려움
- 동일 상황의 반복 재현이 어려움
- 공격 정밀도의 한계 존재
- 수작업에 따른 오류 가능성 증가

분석 정보 부족

- 공격 시점과 대상 정보만 제공
- 공격 정도의 수치적 비교 불가
- 공격 후 정상화 여부 판단 불가

단순한 공격

- 제어 임계 값 단순 증감 공격
- 센서 값의 단순 증감 공격
- 액츄에이터 제어명령 On/Off

HAI 개발 요약

HAI 1.0
정상 데이터: 약 10일
공격 데이터: 38종 (5.5일)
(단일 14종, 복합 38종)

2. 제어시스템 보안 데이터셋 개발 도구 개발

제어기기 공격 자동화

- 제어기기 포인트 값 조작
- 제어루프 기반 공격 은닉
- 명령 실행 스케줄링

SCADA 운전 자동화

- Setpoint 명령 전달
- 제어응답 안정화 확인
- 명령 실행 스케줄링

1. 제어시스템 테스트베드 구축

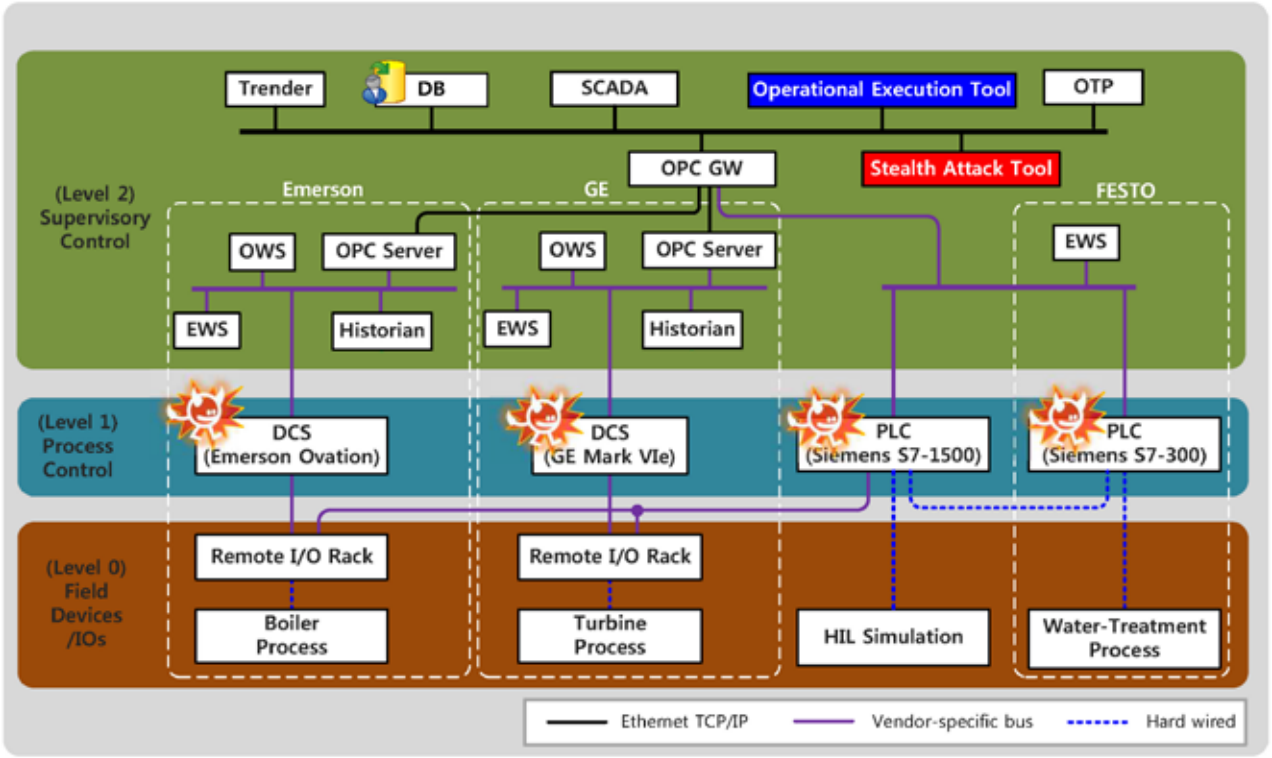
3. 데이터 수집 시스템 구축

- 운전 데이터 수집·취합
- 공격 라벨 자동 생성
- OPC-UA 기반
(이기종 환경 통합)

1. 제어시스템 테스트베드

HIL 시뮬레이터 기반 화력/양수발전 모사 제어시스템 테스트베드 구축

2. 제어시스템 보안 데이터셋 개발 도구



2.1 SCADA 운전 자동화

시스템 안전 을 보장하는 24시간 무인 운전 환경 제공

운영 현황

- 전체 발전량, 부하량 모니터링
- 발전 시설별 발전량, 주파수 모니터링

운전 감시

- 시스템 시작/정지 명령
- Set-point 명령
- 센서/액추에이터 모니터링

자동 운전

- Set-point별 스케줄링 설정
- 시스템 안정화 여부 감시

2.2 제어기기 공격 자동화

제어기기 특화 은닉공격 재현 가능한 공격 자동화 환경 제공

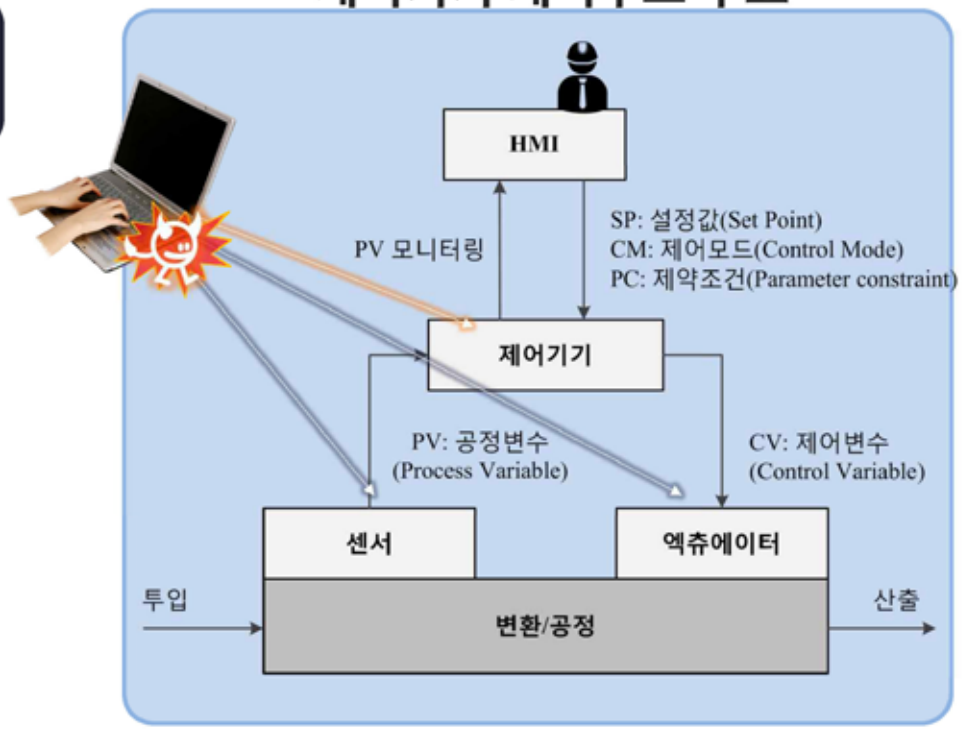
제어기기 제어로직 위변조

- 운전원의 제어명령 조작
- 제어기기 출력 조작
- HMI 데이터 조작을 통한 공격 은닉

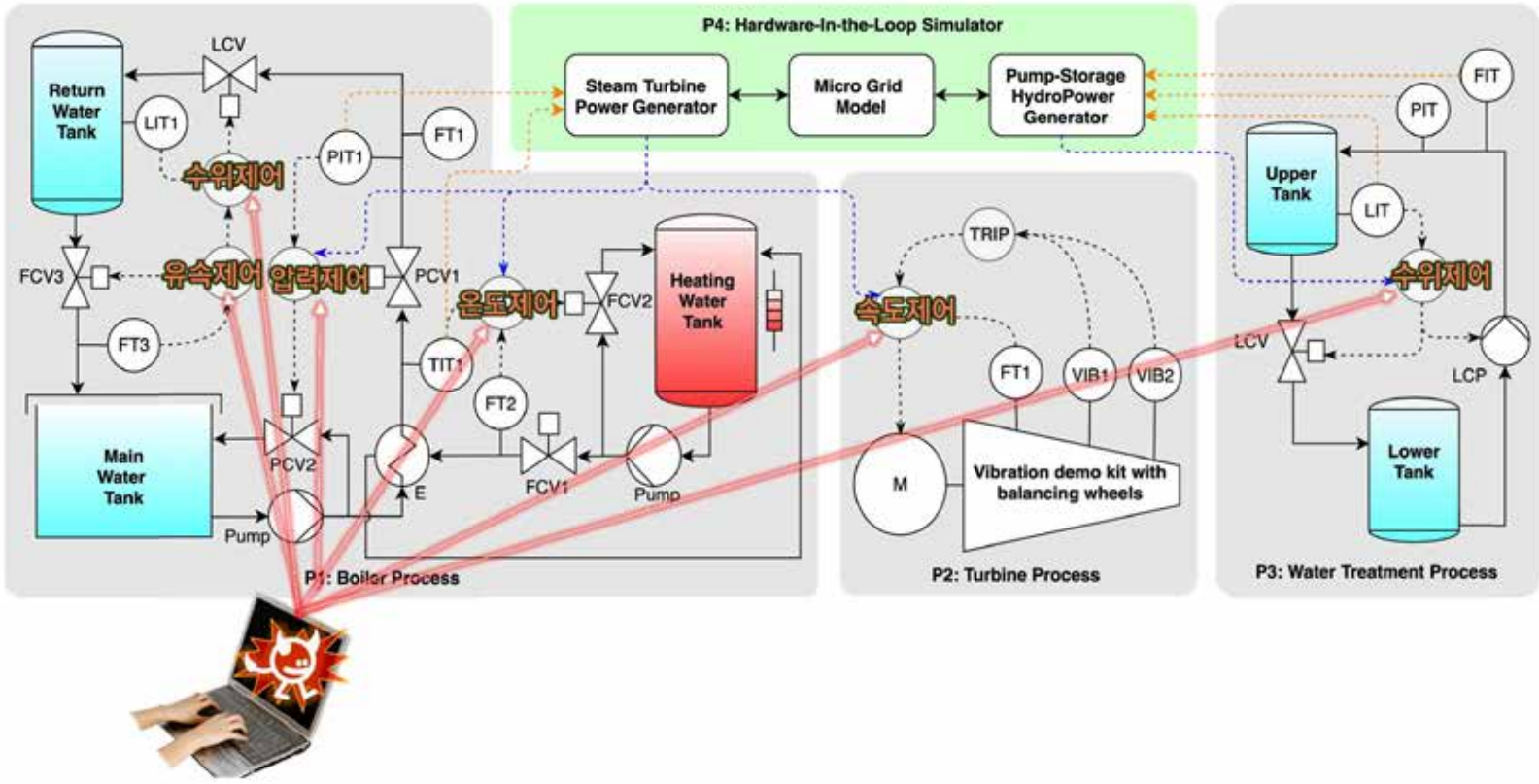
현장장치 입출력 위변조

- 센서 값(PV) 조작
- 액츄에이터 제어명령(CV) 조작

제어기기 제어루프 구조



2.3 제어기기 공격 자동화



2.4 제어기기 공격 자동화

Algorithm 1: Pseudo-code for a PCL attack

Input: target control loop, forced variables and values, PV response prevention, time limits
Output: attack logs
Data: a latest PV snapshot in steady-state

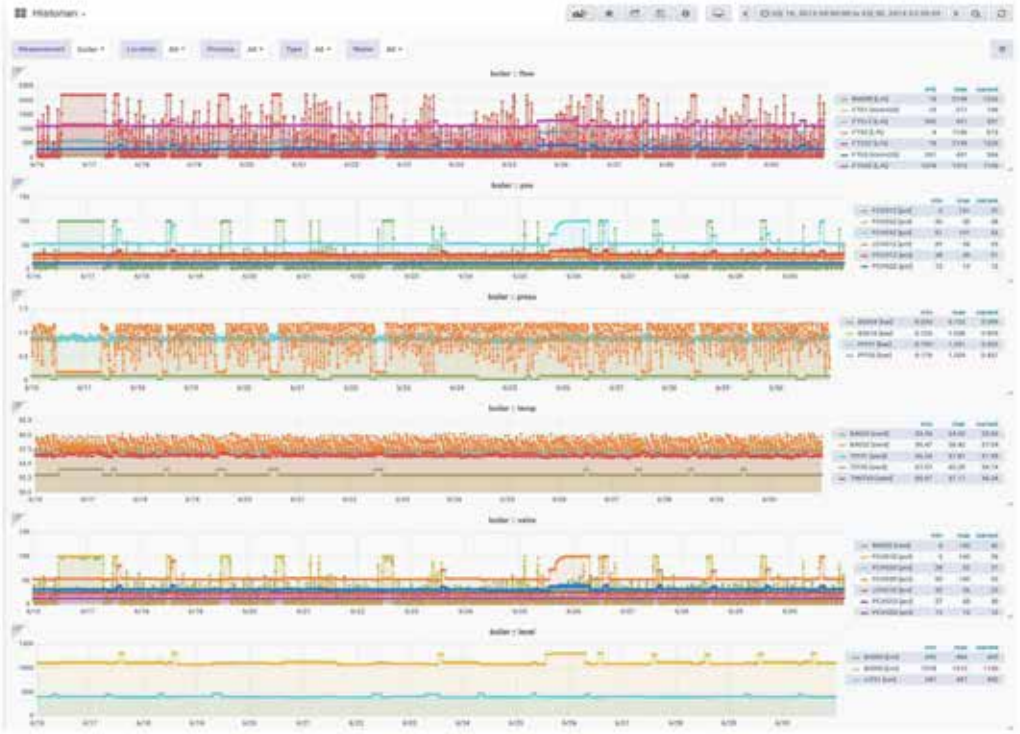
```
Store the current SP, PV and CO recovery values
Generate a attack sequence that is linearly increase to
    the forced value and decrease to the recovery value
    over time limits for each forced variable
Wait until the target control loop is steady state
while any attack sequence remains do
    if the PV response prevention is activated then
        Replay the PV using the snapshot data
    if the SP attack sequence remains then
        Force the SP value from the attack sequence
    if the CO attack sequence remains then
        Force the CO value from the attack sequence
    Record the logs of the attack sequence
```

Table 1: PCL Attack primitives for HAI testbed

No	Attack ID	Target	SP	CO	PV-RP
1	AP-P1PC-SP	P1.PC	V		
2	AP-P1PC-SPRP				V
3	AP-P1PC-CO		V		
4	AP-P1PC-CORP			V	V
5	AP-P1FC-SP	P1.FC	V		
6	AP-P1FC-SPRP				V
7	AP-P1LC-SP	P1.LC	C		
8	AP-P1LC-SPRP				V
9	AP-P1LC-CO		C	V	
10	AP-P1LC-CORP				V
11	AP-P2SC-SP	P2.SC	V		
12	AP-P2SC-SPRP				V
13	AP-P3LC-SP1CO1	P3.LC	V_{LH}	V_{LCV}	
14	AP-P3LC-SP2CO2		V_{LL}	V_{LCP}	

No			ation
1	AP		secs
2	AP		sec
3	AP		secs
4	AP		secs
5	AP		secs
6	AP		secs
7	AP		secs
8	AP-P1PC-CO	19-10-30 11:37	100 secs
9	AP-P1PC-CORP	19-10-30 12:30	100 secs
10	AP-P2SC-SPRP	19-10-30 14:30	100 secs
11	AP-P3LC-SP2CO2	19-10-30 15:35	100 secs
12	AP-P3LC-SP1CO1	19-10-30 16:33	100 secs
13	AP-P2SC-SP	19-10-31 08:42	100 secs
14	AP-(P1PC-SPRP, P2SC-SPRP)	19-10-31 10:30	100 secs
15	AP-(P1PC-CO, P2SC-SP)	19-10-31 11:33	100 secs
16	AP-P2SC-SPRP [Repeat No.11]	19-10-31 13:25	100 secs
17	AP-(P1LC-CORP, P2SC-SPRP)	19-10-31 14:30	100 secs
18	AP-(P1FC-SP, P2SC-SP)	19-10-31 15:41	100 secs
19	AP-(P1PC-SP, P3LC-SP1CO1)	19-10-31 16:30	100 secs
20	AP-(P1LC-SPRP, P3LC-SP1CO1)	19-11-01 09:29	100 secs
21	AP-(P1LC-CO, P3LC-SP1CO1)	19-11-01 10:41	100 secs
22	AP-P3LC-SP1CO1 [Repeat No.12]	19-11-01 11:23	100 secs
23	AP-(P1FC-SPRP, P1LC-SP)	19-11-01 12:31	100 secs
24	AP-(P1PC-CO, P1FC-SPRP)	19-11-01 13:41	100 secs
25	AP-P1PC-SP [Repeat No.4]	19-11-01 14:23	100 secs
26	AP-(P1FC-SP, P1PC-SPRP)	19-11-01 15:31	100 secs
27	AP-P1FC-SPRP	19-11-01 16:18	100 secs
28	AP-(P1PC-SPRP, P3LC-SP2CO2)	19-11-01 17:20	100 secs
29	AP-(P1FC-SP, P1PC-SP)	19-11-04 15:31	100 secs
30	AP-(P1PC-SP, P3LC-SP2CO2)	19-11-04 17:20	100 secs
31	AP-(P1LC-CO, P3LC-SP2CO2)	19-11-05 09:30	380 secs
32	AP-(P1FC-SP, P3LC-SP2CO2)	19-11-05 10:20	290 secs
33	AP-P2SC-SP [Repeat No.11]	19-11-05 11:23	340 secs
34	AP-(P2SC-SP, P3LC-SP2CO2)	19-11-05 12:30	340 secs
35	AP-(P1LC-SP, P2SC-SP)	19-11-05 14:45	2880 secs
36	AP-(P1PC-SP, P1LC-SP)	19-11-05 16:20	330 secs
37	AP-P1LC-CO [Repeat No.7]	19-11-05 17:23	310 secs
38	AP-(P1PC-CO, P1LC-CO)	19-11-06 08:58	310 secs

3. 데이터 수집

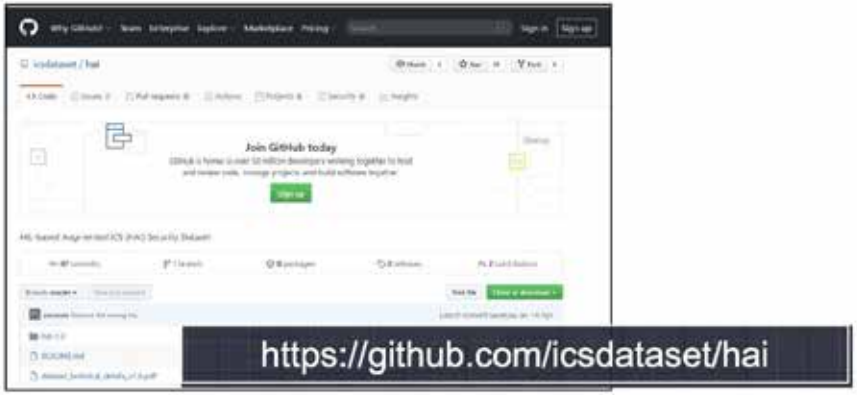


데이터 수집량

549.7MB
(59 포인트/초)

보일러: 31 포인트
터 빈: 12 포인트
수처리: 16 포인트

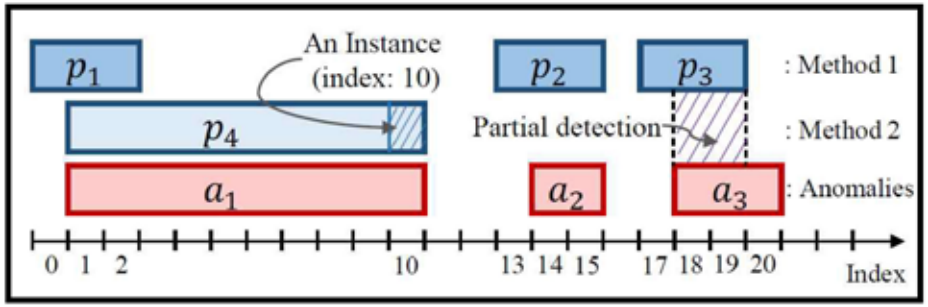
4. 데이터셋 공개 (2020. 2. 7.)




성능평가 TaPR
Time-series aware Precision/Recall

시계열 데이터 이상탐지 성능평가

부분 공격 탐지에 적합한 성능평가 방법 부재

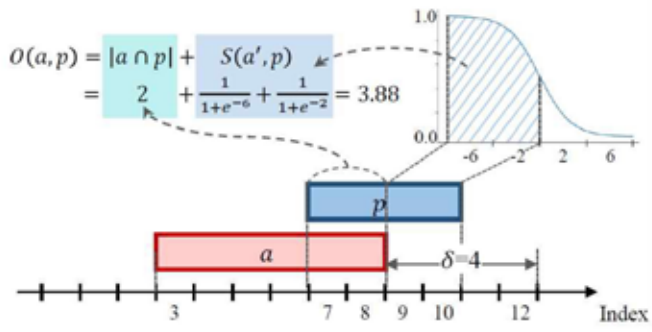


실제 탐지성능: Method1 〉 Method2

기존 평가방법: Method1 〈 Method2

TaPR 성능평가

기존 평가방법 Partial Detection 문제점 개선



- 평가방안을 우수학회에 발표

  (게재) CIKM 2019

  ACM International Conference on Information and Knowledge Management

  (제출) ICDM 2020

  European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database

- 성능평가 도구를 github에 공개

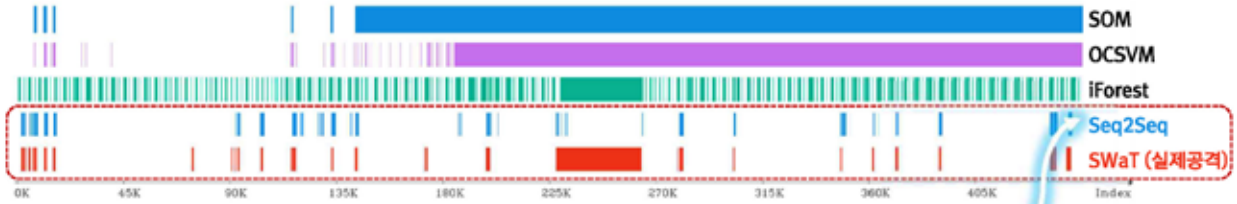
  https://github.com/saurf4ng/TaPR

* 황원석, 윤정한, 김종욱, 김형천, "Time-Series Aware Precision and Recall for Anomaly Detection," CIKM, 2019

* 황원석, 윤정한, 김종욱, 김형천, "Time-Series Aware Precision and Recall for Anomaly Detection – Enhanced Metrics Addressing the Antinomy, Obscurity, and Inflexibility," PKDD, 2020 (Submitted)

TaPR 성능평가

iTrust SWaT 데이터셋 대상 성능평가 비교

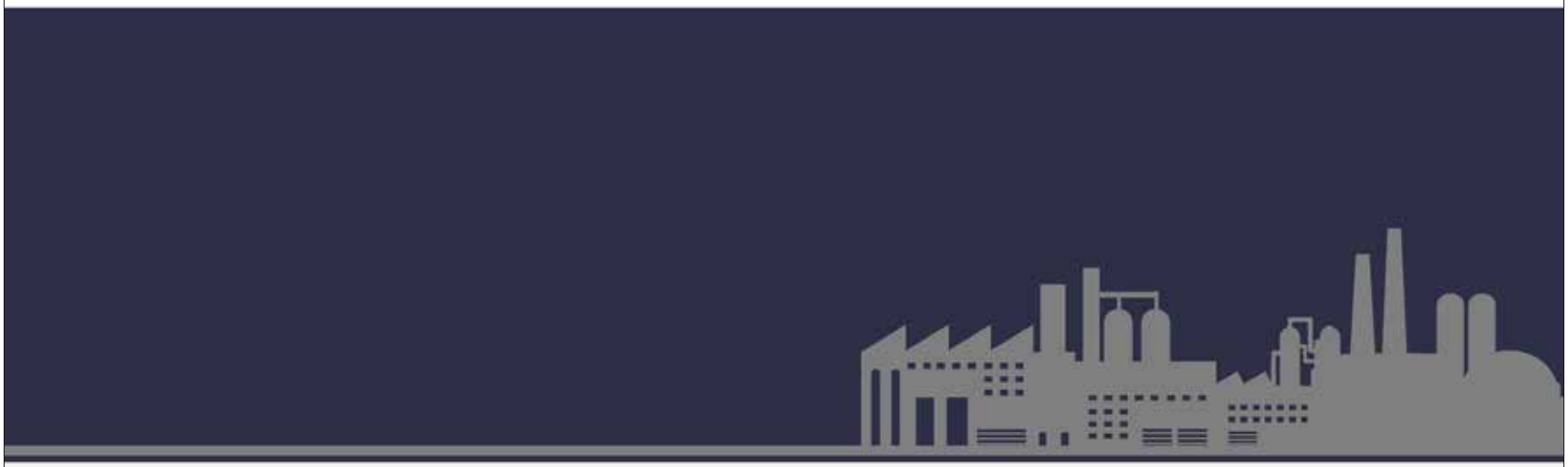


평가척도 정확성 개선

가장 좋은 탐지 결과
(다양한 공격 탐지 & 적은 오탐)

비교척도	1등	2등	3등	4등
제안한 평가척도 (TaPR) *	Seq2Seq (0.44)	SOM (0.29)	iForest (0.05)	OCSVM (0.05)
Precision &Recall	iForest (0.43)	Seq2Seq (0.35)	OCSVM (0.28)	SOM (0.28)
TSAD**	SOM (0.73)	Seq2Seq (0.46)	OCSVM (0.23)	iForest (0.07)

*황원석, 윤정한, 김종욱, 김형천, "Time-Series Aware Precision and Recall for Anomaly Detection", CIKM 2019
**Nesime Tatbul et al., "Precision and Recall for Time Series," *Advances in Neural Information Processing Systems*, 2018



마무리

보안 데이터셋: 제어시스템 보안연구의 기반

제어시스템 보안 연구를 위한 환경 구축

- 활용 가능한 연구대상 절대 부족
- 제품 개발/성능평가용 데이터셋 개발

보안위협 탐지기술 개발

- ICS 보안 데이터셋 및 평가방안
- 머신러닝 기반 이상행위 탐지 기술

데이터셋 및 평가방안 공유

- (연구자) 보안 연구 활성화
- (개발자) 제품 개발 성능 비교
- (사용자) 보안 성능 요구사항

향후 계획

☑ 보안위협 탐지 온라인 경연대회 (8월 ~ 9월)

※ 학습 데이터는 정상 상황 데이터만 제공
※ 1등 1,000만원, 2등 500만원, 3등 300만원 수준

☑ 운영환경 변경/공격 추가: HAI 2.0 공개

※ 보안위협 탐지 경연대회 참가자 대상 선공개(예정)


감사합니다