스마트 빌딩 냉난방공조 시스템에서 LLM 기반 침입 탐지용 자동 프롬프트 생성 방법 설계

박수현, 안석현, 김도익, 조성제, 김홍근

KCC 2025
2025.07.02

컴퓨터보안 및 OS 연구실

단국대학교 DANKOOK UNIVERSITY

Computer Security & OS LAB

INDEX

01 서론

서론

❖ 스마트 빌딩: 냉난방공조(HVAC), 조명, 차양 등의 다양한 건물 기능을 관리하고 제어하는 연결된 장치 및 소프트웨어의 네트워크

❖ 스마트 빌딩의 구성 요소 중 하나인 BAS(Building Automation System)는 건물 기능 제어에 초점

- 원래의 BAS는 폐쇄형 구조 ⇒ 보안 고려 X
- 원격 통신과 클라우드 기반 제어 기술 확산으로 급격히 개방되며 보안에 취약

⇒ BAS 대상 침입 탐지 시스템(Intrusion Detection System, IDS)에 대한 요구 증가

출처 : https://www.cisco.com/c/en/us/solutions/smart-building/what-is-a-smart-building.html

서론

❖ 침입 탐지 모델 (Intrusion Detection System)

- **NIDS (Network IDS)**
  - 규칙 기반: 사전의 정의된 규칙에 따라 네트워크 트래픽을 분석하여 침입 여부 판단 (ex. Snort IDS)
    - 장점: 높은 탐지 정확도, 탐지 근거 설명 가능
    - 단점: 유지보수가 복잡함, 제로 데이 공격 탐지 불가
  - 데이터 기반: 과거 트래픽 데이터를 학습하여 정상 및 비정상 패턴을 식별하는 ML or DL 기반 탐지
    - 장점: 트래픽의 동적 특성을 반영한 대응 가능
    - 단점: 탐지 결과에 대한 설명이 어려움, 오탐 및 미탐의 위험

❖ **LLM 기반 침입 탐지 모델 적용 가능성**을 평가하기 위한 연구

- 기대 효과: 네트워크 환경 변화에 유연한 대처, 쉬운 유지보수, 탐지 결과에 대한 설명 가능성

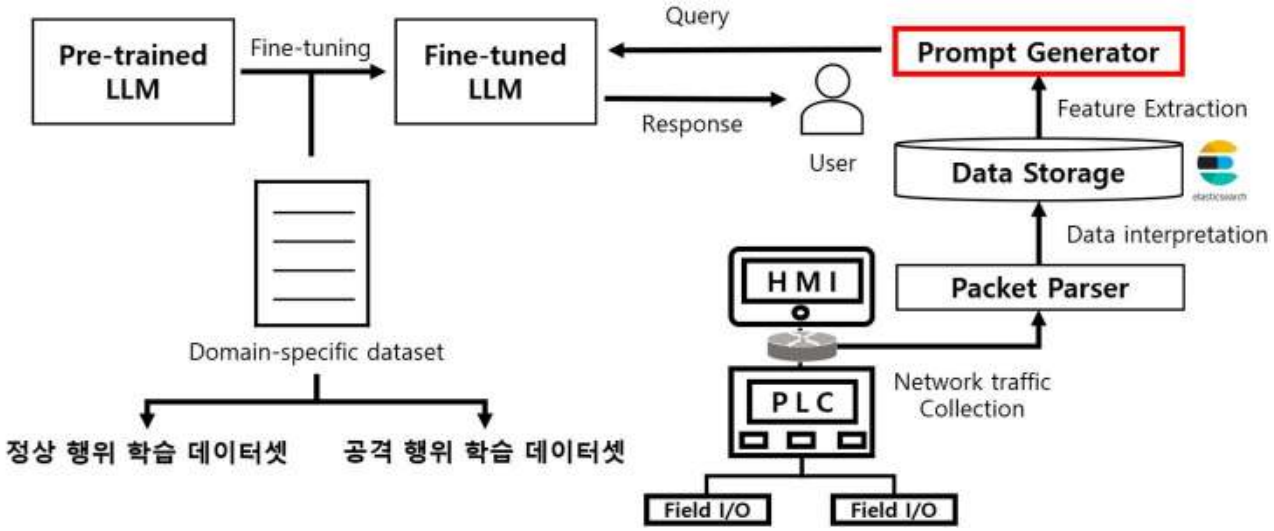
LLM 기반 네트워크 침입 탐지 시스템

LLM 기반 네트워크 침입 탐지 시스템

❖ 패킷 파서, 데이터 스토리지, 프롬프트 생성기, 파인튜닝 된 LLM으로 구성

❖ 흐름

1. 네트워크 트래픽 데이터를 수집 후 전처리하여 스토리지에 저장
2. 이를 바탕으로 프롬프트를 생성하여 정상 및 공격 행위에 대해 학습한 파인튜닝 된 LLM에게 전달
3. LLM은 프롬프트를 입력 받아 추론을 통해 탐지 결과를 응답

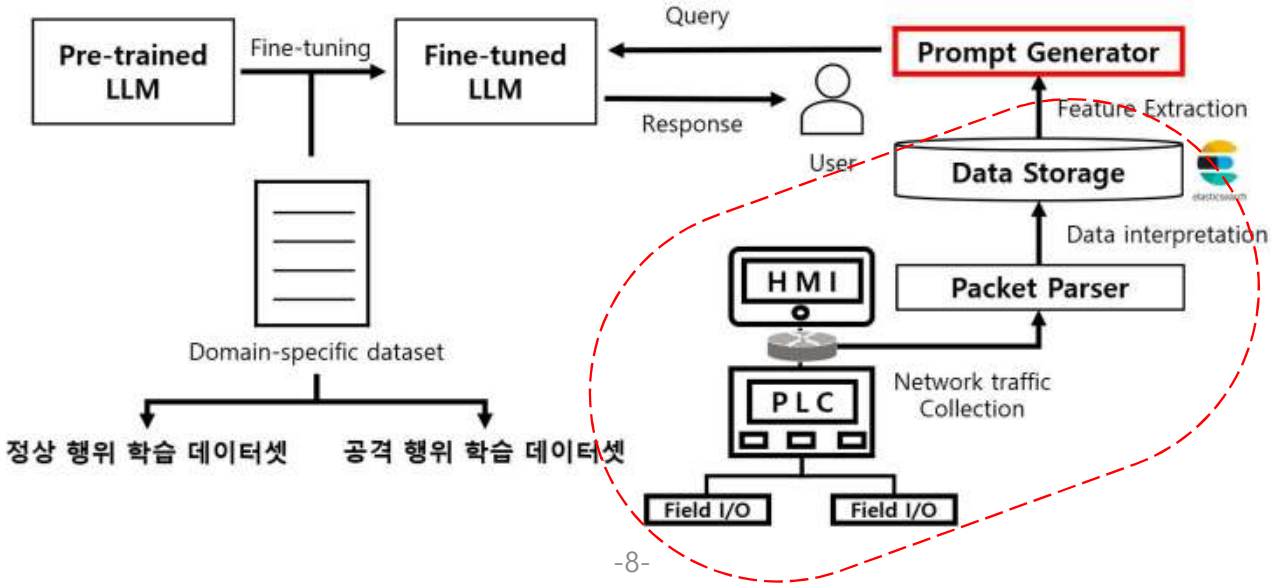


LLM 기반 네트워크 침입 탐지 시스템

❖ 패킷 파서, 데이터 스토리지, 프롬프트 생성기, 파인튜닝 된 LLM으로 구성

❖ 흐름

1. 네트워크 트래픽 데이터를 수집 후 전처리하여 스토리지에 저장
2. 이를 바탕으로 프롬프트를 생성하여 정상 및 공격 행위에 대해 학습한 파인튜닝 된 LLM에게 전달
3. LLM은 프롬프트를 입력 받아 추론을 통해 탐지 결과를 응답

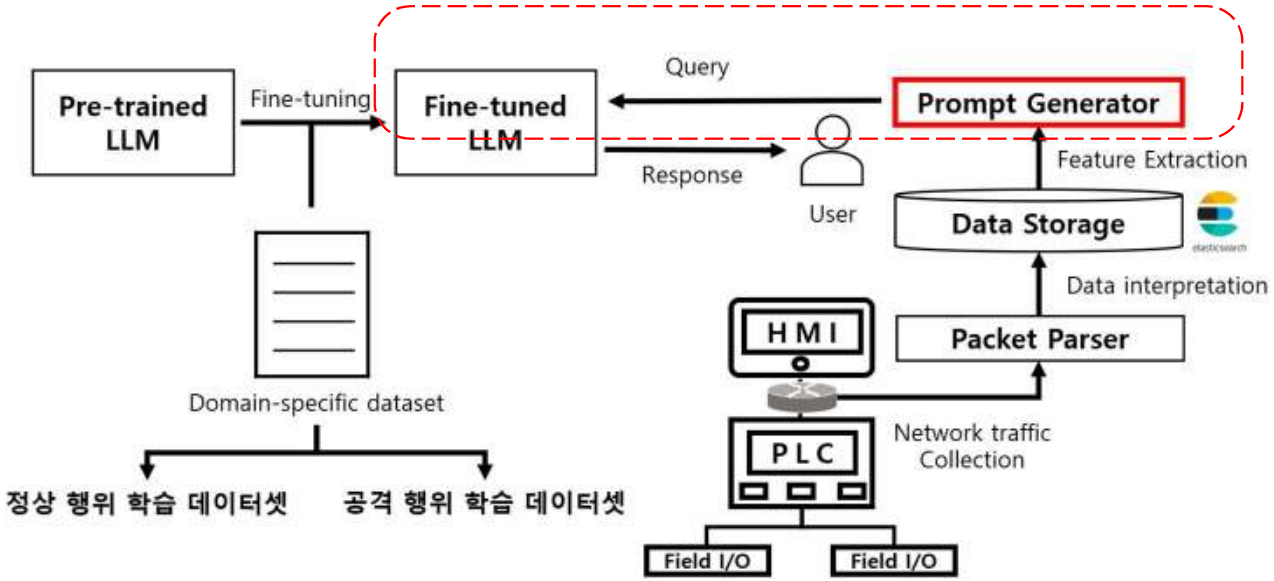


LLM 기반 네트워크 침입 탐지 시스템

❖ 패킷 파서, 데이터 스토리지, 프롬프트 생성기, 파인튜닝 된 LLM으로 구성

❖ 흐름

1. 네트워크 트래픽 데이터를 수집 후 전처리하여 스토리지에 저장
2. 이를 바탕으로 프롬프트를 생성하여 정상 및 공격 행위에 대해 학습한 파인튜닝 된 LLM에게 전달
3. LLM은 프롬프트를 입력 받아 추론을 통해 탐지 결과를 응답

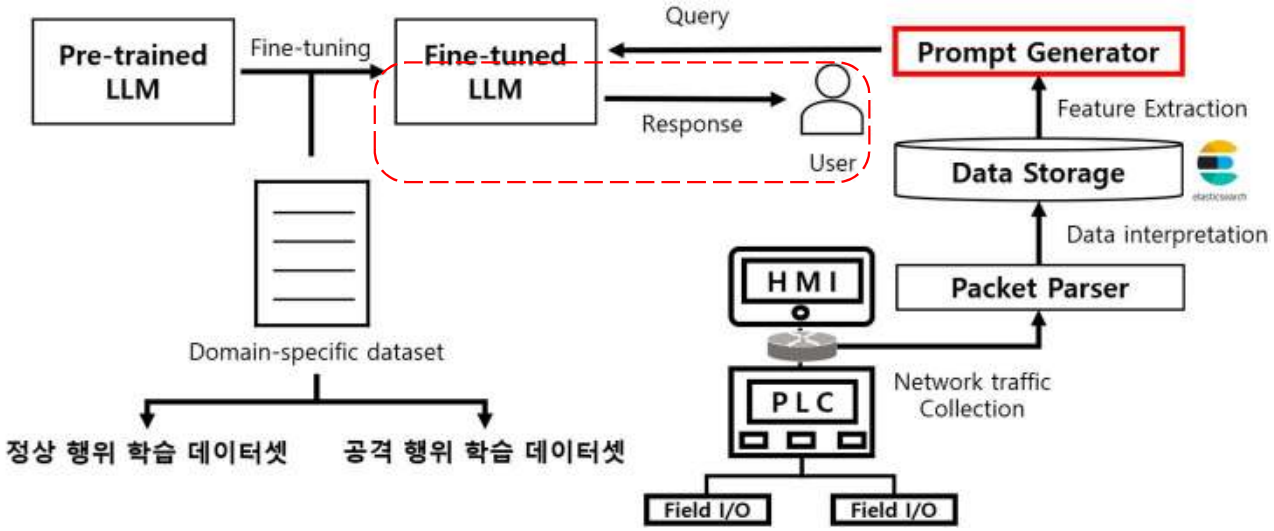


LLM 기반 네트워크 침입 탐지 시스템

❖ 패킷 파서, 데이터 스토리지, 프롬프트 생성기, 파인튜닝 된 LLM으로 구성

❖ 흐름

1. 네트워크 트래픽 데이터를 수집 후 전처리하여 스토리지에 저장
2. 이를 바탕으로 프롬프트를 생성하여 정상 및 공격 행위에 대해 학습한 파인튜닝 된 LLM에게 전달
3. LLM은 프롬프트를 입력 받아 추론을 통해 탐지 결과를 응답



파인튜닝 모델

- 스마트빌딩 HVAC 시스템의 정상 및 공격 행위에 대한 학습 데이터로 LLaMA 3.1을 파인튜닝
- 정상 행위 데이터: ICSSIM[7]을 변형하여 개발한 HVAC 가상 시뮬레이터(IBEEMS SIM)의 제어 네트워크 동작을 기반으로 수집
  - [7] Dehlaghi-Ghadim, Alireza, et al., "ICSSIM—a framework for building industrial control systems security testbeds," Computers in Industry 148, 103906, 2023.
- 공격 행위 데이터: MITRE ATT&CK for ICS의 공격 전술(tactic), 기법(technique), 설명을 바탕으로 생성

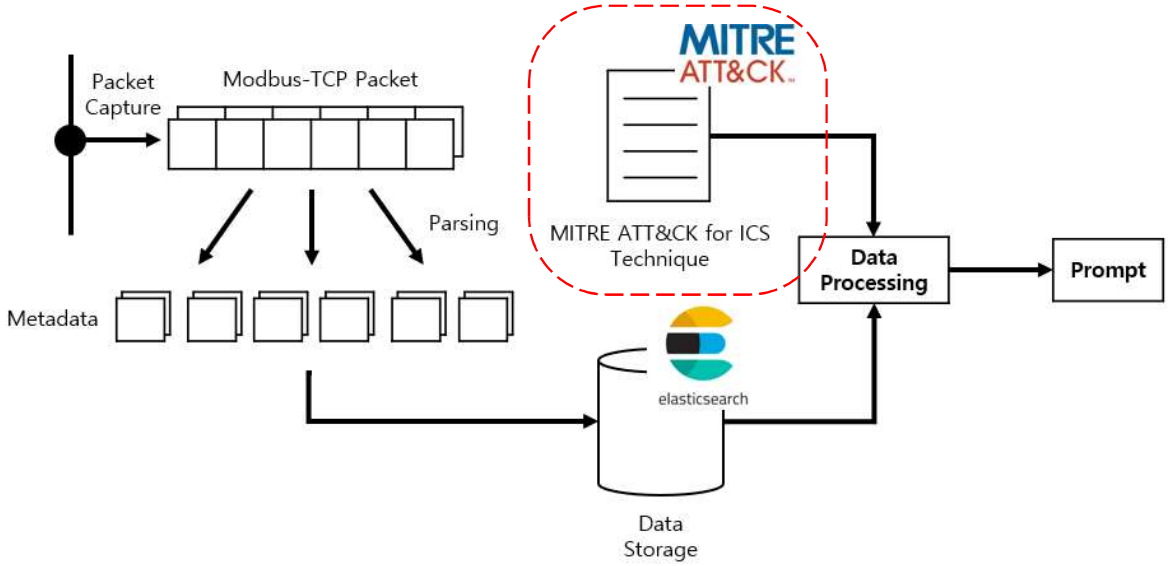
03 프롬프트 생성

프롬프트 생성 과정

❖ 프롬프트 생성 과정

- 준비사항: MITRE ATT&CK for ICS에 포함한 공격 탐지에 사용할 질문(p_Question) 생성

1. 패킷 수집 후 저장: 수집한 Modbus-TCP 패킷을 분할하여 해석하고 데이터 스토리지에 저장
2. 프롬프트 개발: MITRE ATT&CK for ICS 공격 기법을 분석하여 지정된 패킷의 필드를 데이터 스토리지에서 추출 후 가공하여 자연어 질의 형태의 프롬프트 생성

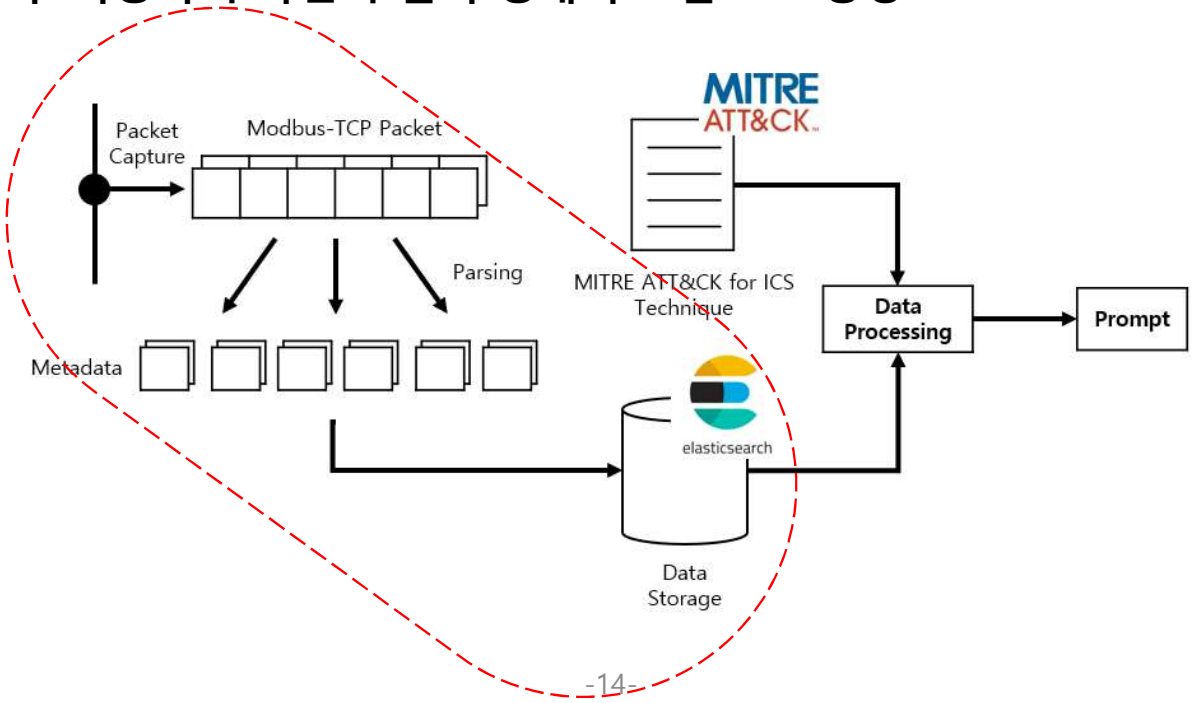


프롬프트 생성 과정

❖ 프롬프트 생성 과정

- 준비사항: MITRE ATT&CK for ICS에 포함한 공격 탐지에 사용할 질문(p_Question) 생성

1. 패킷 수집 후 저장: 수집한 Modbus-TCP 패킷을 분할하여 해석하고 데이터 스토리지에 저장
2. 프롬프트 개발: MITRE ATT&CK for ICS 공격 기법을 분석하여 지정된 패킷의 필드를 데이터 스토리지에서 추출 후 가공하여 자연어 질의 형태의 프롬프트 생성

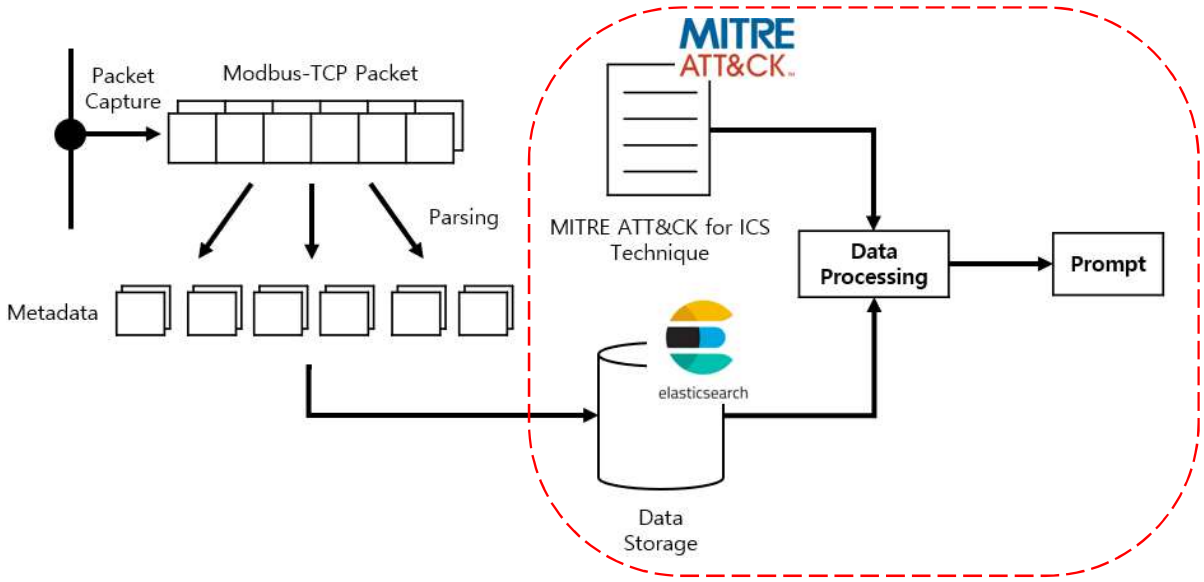


프롬프트 생성 과정

❖ 프롬프트 생성 과정

- 준비사항: MITRE ATT&CK for ICS에 포함한 공격 탐지에 사용할 질문(p_Question) 생성

1. 패킷 수집 후 저장: 수집한 Modbus-TCP 패킷을 분할하여 해석하고 데이터 스토리지에 저장
2. 프롬프트 개발: MITRE ATT&CK for ICS 공격 기법을 분석하여 지정된 패킷의 필드를 데이터 스토리지에서 추출 후 가공하여 자연어 질의 형태의 프롬프트 생성



프롬프트 생성 과정 - 준비사항

❖ 네트워크 탐지 가능한 공격 식별 및 질의문 개발

- **공격 식별**
  - 네트워크 환경에서 탐지 가능한 MITRE ATT&CK for ICS 공격 기법을 표와 같이 식별
- **질의문 개발**
  - HVAC 시스템에서 공격 대상 자산인 HMI와 PLC가 침해되었을 때, 네트워크 수준에서 탐지 가능한 공격 기법과 탐지에 사용할 질의문 개발
  - 질의문은 프롬프트 개발 단계에서 결합됨

탐지 가능한 공격 기법 및 탐지 질문(p_question)

Technique name (Technique ID)	Question
Adversary-in-the-Middle (T0830)	Is there a value that falls outside the normal range, a response from a device other than expected, or more than one response to a single request?
Denial of Service (T0814)	Did a large number of packets occur in a short period of time?
Exploitation of Remote Services (T0866)	Check whether there is any packet using function code 43/13, and if found, determine whether it is normal for the source device of that packet to perform that function code operation.
Modify Parameter (T0836)	Identify any communications that involve write-related function codes (6 or 10), and determine whether the source device is unauthorized to use these function codes or if any of the written values fall outside the acceptable range.
Spoof Reporting Message (T0856)	Does a single device respond twice to a read request for the same variable within a very short time interval with different values? If so, would such redundant and inconsistent responses be considered unexpected behavior?
Unauthorized Command Message (T0855)	Identify any communications that involve write-related function codes (6 or 10), and verify whether the source device is authorized to use those function codes under normal operation.

프롬프트 생성 과정 – 패킷 수집 후 저장

❖ 프롬프트 생성을 위한 전처리

- 수집한 Modbus-TCP 패킷을 필드별로 분할하여 저장
  - 타임스탬프, IP, 데이터 필드 등을 자연어 형태로 변환
  - 타임스탬프(epoch time) → UTC time
  - IP → 장치명
  - 데이터 필드 → 완전히 해석된 데이터 값
- Elasticsearch 기반의 데이터 스토리지 구성


```
"_index": "modbus250407",
"_id": "qVmnDZYBvUgt9JHKdx82",
"_score": 1,
"_source": {
  "timestamp": "2025-04-07T00:30:00.058099",
  "source_device": "plc1",
  "destination_device": "hmi1",
  "source_port": 502,
  "destination_port": 49868,
  "transaction_id": 3233,
  "protocol_id": 0,
  "length": 7,
  "unit_id": 1,
  "function_code": 3,
  "data": {
    "type": "response",
    "tag_name": "cooling_water_supply_value",
    "value": 28
  }
}
```

프롬프트 생성 과정 – 프롬프트 개발

❖ 고려사항

- 단일 트래픽 기반 탐지 방식의 경우 네트워크 패킷의 데이터 필드의 비정상 유무 정도만 판단 가능
  - Function code를 이용한 공격이나 DoS 같은 복수 트래픽을 통해 탐지해야 하는 공격을 탐지할 수 없음

⇒ 해결방안: 일정 주기마다 모여진 **복수의 트래픽을 기반**으로 프롬프트를 생성하여 비정상 여부 판단

❖ 행위 문맥 생성

- 일정 시간 프레임 동안 HMI와 PLC의 네트워크 행위를 표현하는 행위 문맥(behavior context) 생성
- HVAC 가상 시뮬레이터에서 PLC-HMI간에 0.013초 동안 발생한 패킷 200개에 대한 행위 문맥 생성

(1) At **00:31:00.026979**, HMI1 send a function code 3 request to PLC1 to read cooling_water_supply_value.

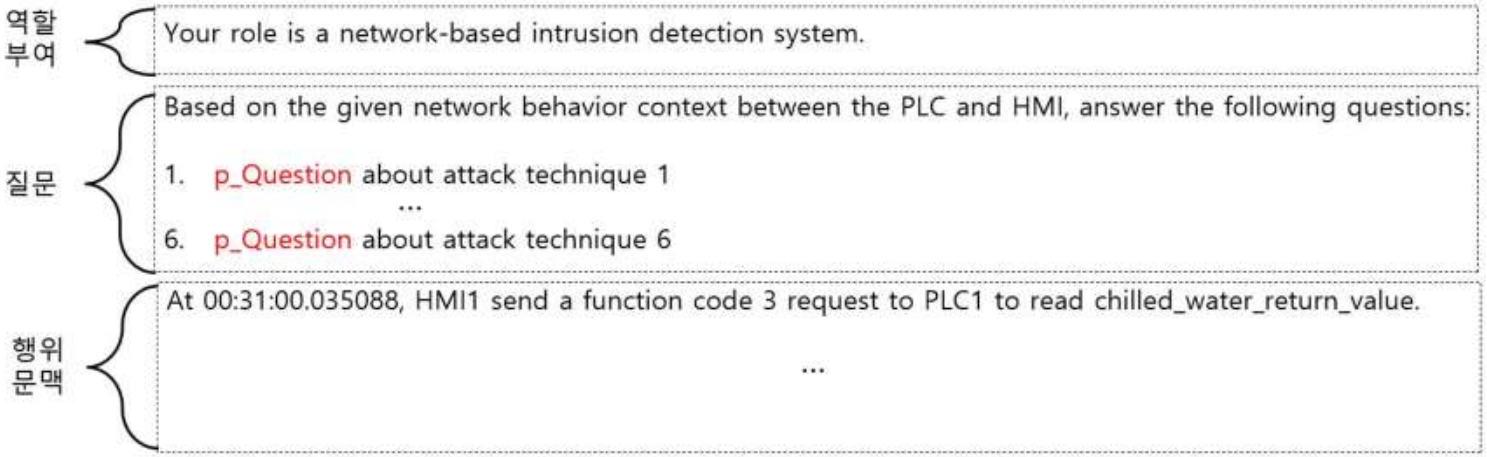
(중략)

(200) At **00:31:00.038917**, PLC2 respond to HMI1 with function code 3, value 3.0 for heating_valve_mode.

프롬프트 생성 과정 – 프롬프트 개발

❖ 프롬프트 템플릿 설계

- 역할 부여: 네트워크 기반 침입 탐지 시스템이라는 역할 부여
- 질문: PLC와 HMI의 행위 문맥을 바탕으로 6가지 공격 기법 발생 여부를 확인하는 질문
- 행위 문맥: 공격 기법의 발생 여부 판단



프롬프트 템플릿

p_Question

Technique name (Technique ID)	Question
Adversary-in-the-Middle (T0830)	Is there a value that falls outside the normal range, a response from a device other than expected, or more than one response to a single request?
Denial of Service (T0814)	Did a large number of packets occur in a short period of time?
Exploitation of Remote Services (T0866)	Check whether there is any packet using function code 43/13, and if found, determine whether it is normal for the source device of that packet to perform that function code operation.
Modify Parameter (T0836)	Identify any communications that involve write-related function codes (6 or 10), and determine whether the source device is unauthorized to use these function codes or if any of the written values fall outside the acceptable range.
Spoof Reporting Message (T0856)	Does a single device respond twice to a read request for the same variable within a very short time interval with different values? If so, would such redundant and inconsistent responses be considered unexpected behavior?
Unauthorized Command Message (T0855)	Identify any communications that involve write-related function codes (6 or 10), and verify whether the source device is authorized to use those function codes under normal operation.

04 실험 및 결과

최종 실험 프롬프트 시나리오

❖ 공격 시나리오가 적용된 200개의 패킷에서 공격 탐지를 얼마나 해내는지 확인

- **[Exploitation of Remote Services (T0866)]**

✓ HMI1이 PLC2에게 function code 43/13 요청을 보냄. 응답으로 S7-1200 모델을 사용함을 확인. 공격자는 검색을 통해 PLC2가 지멘스사의 S7-1200 모델의 PLC 모델임을 확인.

- **[Adversary-in-the-Middle (T0830)] & [Modify Parameter (T0836)] & [Unauthorized Command Message (T0855)]**

✓ HMI1이 heating_valve_status에 write 요청을 통해 1(on)으로 바꿈.

- **[Spoof Reporting Message (T0856)]**

✓ 하나의 heating_valve_status 값의 read 요청에 아주 짧은 시간(약 0.01초) 내에 2개의 응답이 발생하게 되었다.

- **[Denial of Service (T0814)]**

✓ 0.02초 사이에 200개의 패킷이 발생함.

실험 및 결과

❖ 질문

▪ 프롬프트 템플릿에 따라 프롬프트 생성

역할 부여	Your role is a network-based intrusion detection system. Based on the given network behavior context between the PLC and HMI1, answer the following questions:
질문	1. Is there a value that falls outside the normal range, a response from a device other than expected, or more than one response to a single request? (중략) 6. Identify any communications that involve write-related function codes (6 or 10), and verify whether the source device is authorized to use those function codes under normal operation.
행위 문맥	At 00:31:00.026979, HMI1 send a function code 3 request to PLC1 to read cooling_water_supply_value. (중략) At 00:31:00.038917, PLC2 responded to HMI1 with function code 3, value 3.0 for `heating_valve_mode`.

실험 및 결과

❖ 답변

Adversary-in-the-Middle (T0830)	1. **The response from PLC2 to HMI1 at 00:31:00.035924 with function code 6 instead of the expected function code 3 for `return_air_temperature_value` is abnormal.** This inconsistency in function codes could suggest tampering or an error in communication.
Denial of Service (T0814)	2. **A phenomenon was observed where HMI1 continuously sends requests to PLC1 and PLC2 for a short duration. In the current environment, more than 100 requests per minute may indicate a Denial of Service (DoS) attack.** The system may block the source or continue normal operation depending on the circumstances.
Exploitation of Remote Services (T0866)	3. At 00:31:00.028034, PLC2 responded to HMI1 with a function code 43/13 response. **According to the context, the presence of function code 43/13 can be associated with an Exploitation of Remote Services (T0866) attack.** This suggests that there might be an attempt to exploit remote services on the PLC.
Modify Parameter(T0836)	4. At 00:31:00.029434, HMI1 sends a function code 6 request to PLC2 to write heating_valve_status with a value of 1. **However, since HMI1 is expected to perform read operations (function code 3), the use of function code 6 here is suspicious and could indicate modification attempts.**
Spoof Reporting Message (T0856)	5. No, the values are within the normal range, and all responses originate from the expected devices. **However, there is a case where two responses with different values occurred within a very short time interval for a single variable read request, which may indicate a communication issue or a potential security breach.**
Unauthorized Command Message (T0855)	6. Identifying unauthorized write commands is crucial for maintaining system integrity. **In this case, HMI1 sent a write command to PLC2 to change the status of heating_valve to ON, which is not part of its normal operational behavior.**

실험 및 결과

❖ 답변 일부 분석

- **[Exploitation of Remote Services (T0866)]**

✓ 시나리오: HMI1이 PLC2에게 function code 43/13 요청을 보냄. 응답으로 S7-1200 모델을 사용함을 확인. 공격자는 검색을 통해 PLC2가 지멘스사의 S7-1200 모델의 PLC 모델임을 확인.

✓ **답변: At 00:31:00.028034, PLC2 responded to HMI1 with a function code 43/13 response. According to the context, the presence of function code 43/13 can be associated with an Exploitation of Remote Services (T0866) attack.**

- **[Spoof Reporting Message (T0856)]**

✓ 시나리오: 하나의 heating_valve_status 값의 read 요청에 아주 짧은 시간(약 0.01초) 내에 2개의 응답이 발생하게 되었다.

✓ **답변**: **However, there is a case where two responses with different values occurred within a very short time interval for a single variable read request, which may indicate a communication issue or a potential security breach.**

- **[Denial of Service (T0814)]**

✓ 시나리오: 0.02초 사이에 200개의 패킷이 발생함.

✓ **답변**: **A phenomenon was observed where HMI1 continuously sends requests to PLC1 and PLC2 for a short duration. In the current environment, more than 100 requests per minute may indicate a Denial of Service (DoS) attack.**

05 결론 및 향후연구

결론 및 향후 연구

Conclusion

- 스마트빌딩 HVAC 네트워크 환경에서 MITRE ATT&CK 공격 기법 탐지를 위한 질의 프롬프트를 네트워크로부터 자동 생성하는 방법 제안
- 생성된 프롬프트를 파인튜닝된 LLaMA 3.1에 입력하고 답변을 분석한 결과, 네트워크 트래픽 기반 탐지가 가능한 6개의 공격 기법을 정확히 탐지
- 탐지된 공격 기법에 대한 설명을 제공함으로써 해석 가능한 탐지 기술 개발의 가능성을 보임

Future Work

- PLC, HMI 외에도 다양한 자산을 탐지 대상에 포함
- Modbus-TCP 이외의 산업용 이더넷 통신 프로토콜을 분석 대상으로 추가

Acknowledgement

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제임 (No. RS-2021-KP002461).

또한 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 학석사연계ICT핵심인재양성 사업의 연구결과로 수행되었음 (IITP-2023-00259867).

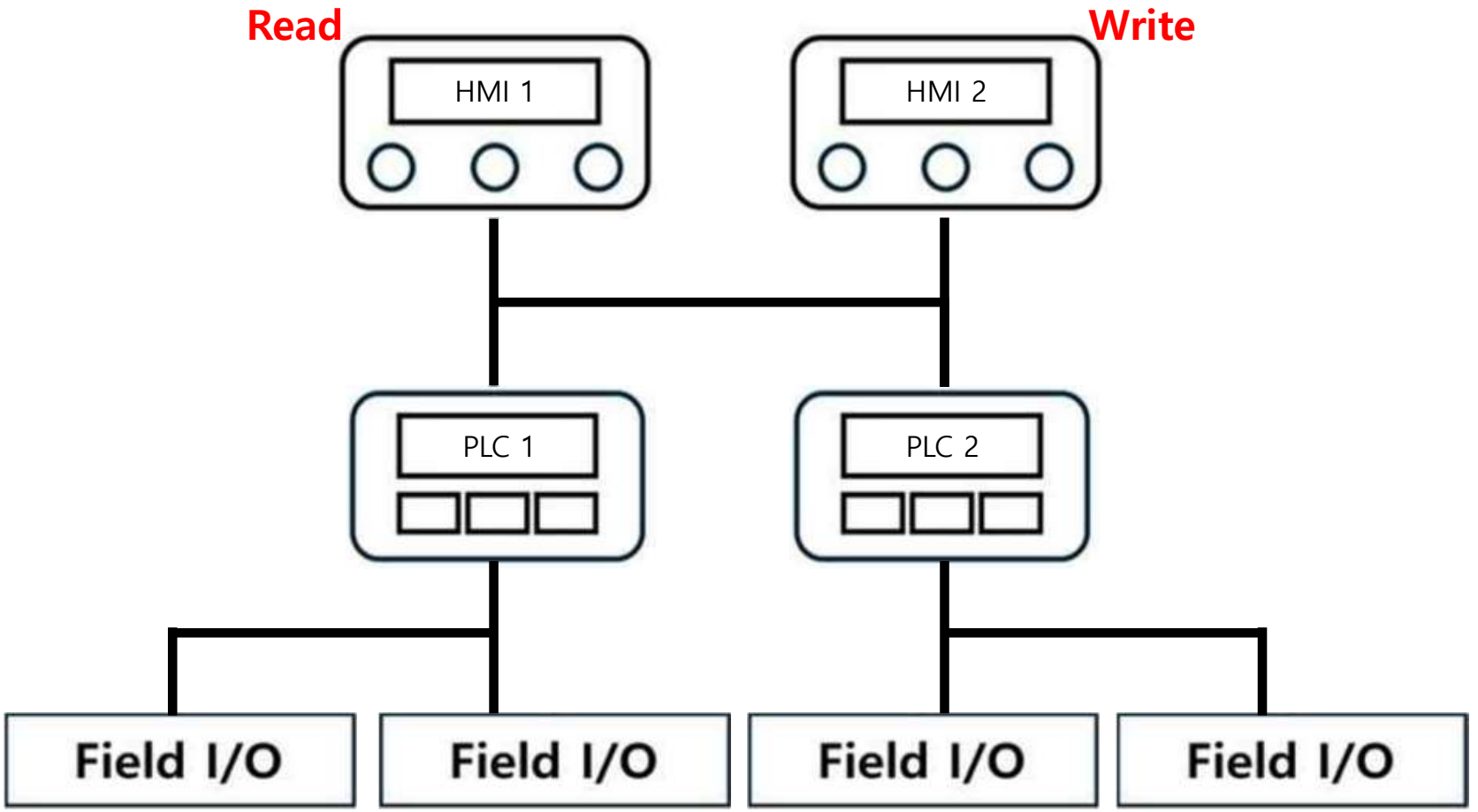
또한 본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (No. 2024-0-00035).

Thank you

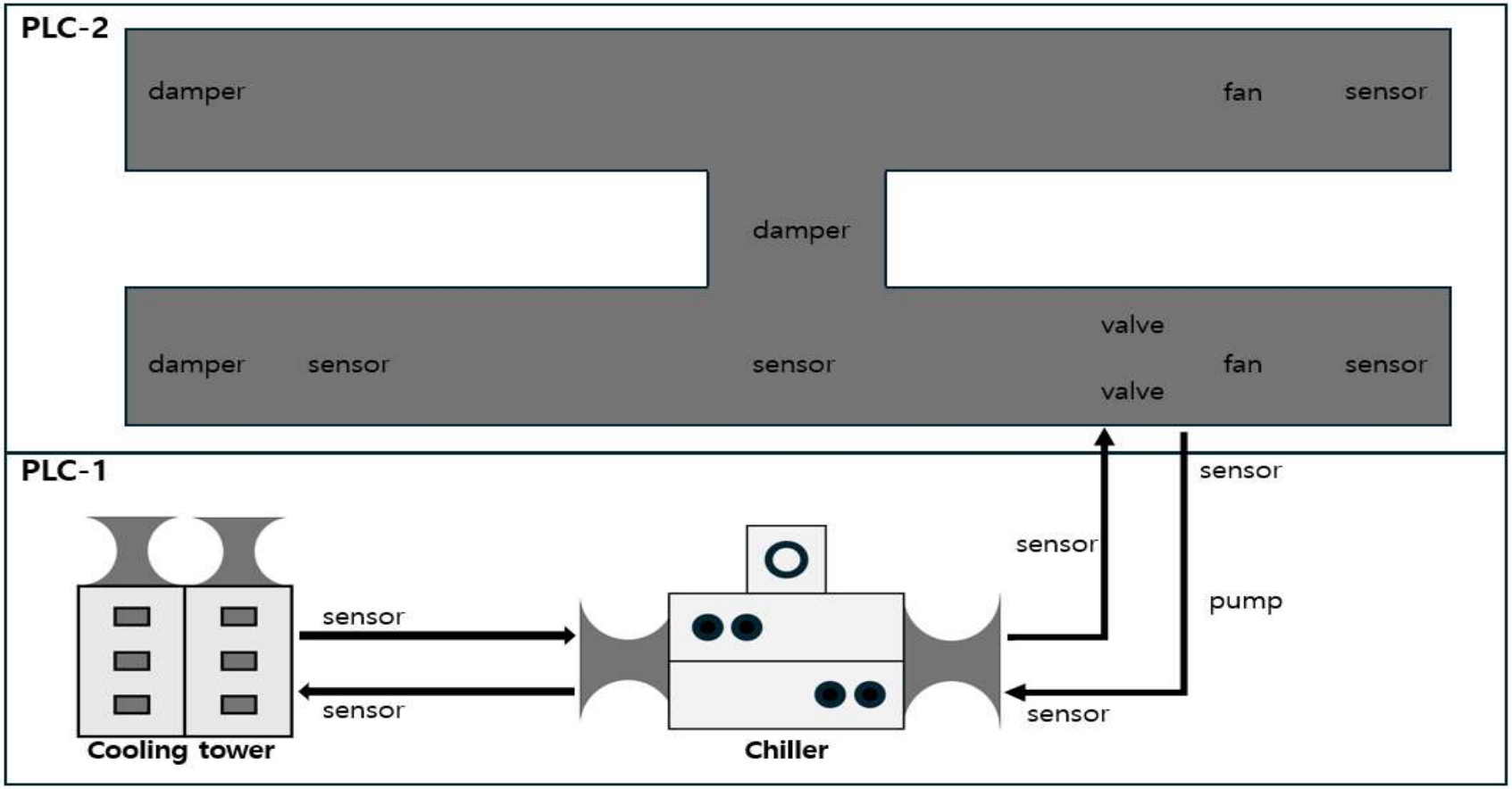
Q&A

Appendix.

시뮬레이션 환경 장치 구성



시뮬레이션 환경 구조


PLC-2
damper
fan
sensor
damper
damper
sensor
sensor
valve
valve
fan
sensor
PLC-1
sensor
sensor
pump
sensor
sensor
sensor
Cooling tower
Chiller

시뮬레이션 환경

❖ 시뮬레이션 환경 (ICSSIM 변형)

- HMI1의 동작 예시

[HMI1 - 04:56:47] (Latency 0.468ms)

cooling_water_supply	Max: 40.0 Min: 20.0	28.0002
cooling_water_return	Max: 40.0 Min: 20.0	29.0002
chilled_water_supply	Max: 30.0 Min: 5.0	23.0002
chilled_water_return	Max: 30.0 Min: 5.0	23.0002
chilled_water_exit	Max: 30.0 Min: 5.0	24.0002
cooling_water_circulation_pump	Auto	X
chilled_water_circulation_pump	Auto	X
inside_co2	Max: 1100.0 Min: 0.0	500.0
inside_temperature	Max: 60.0 Min: -20	27.4998
return_air_temperature	Max: 60.0 Min: -20	27.4998
outside_air_temperature	Max: 60.0 Min: -20	0.0
mixing_air_temperature	Max: 60.0 Min: -20	27.4998
exhaust_air_temperature	Max: 60.0 Min: -20	27.4998
return_fan	Auto	X
supply_fan	Auto	X
exhaust_air_damper	Auto	X
outside_air_damper	Auto	X
mixing_air_damper	Auto	>>>
heating_valve	Auto	X
cooling_valve	Auto	X