논문 2024-4-30 http://dx.doi.org/10.29056/jsav.2024.12.30

스마트 빌딩용 대규모 언어 모델 기반 침입 탐지 모델

안석현*, 박수현**, 김도익***, 조성제**† , 김홍근****

A Large Language Model-based Intrusion Detection Model for Smart Buildings

Seokhyun Ann*, Suhyeon Park**, Dolk Kim***, Seong-je Cho**† , HongGeun Kim****

요 약

스마트 빌딩의 빌딩 자동화 시스템은 편리성과 효율성을 제공하지만 보안 취약성으로 인해 사이버 공격에 노출되기 쉽다. 이를 해결하기 위해 본 논문에서는 스마트 빌딩 환경에 특화된 대규모 언어 모델 기반 침입 탐지 모델을 제안한다. 냉난방공조(HAVC) 시스템을 대상으로 정상 및 공격 행위를 식별하여 학습 데이터셋으로 개발하고, LLaMA 3 8B 모델을 파인 튜닝하였다. 정상 데이터는 물리 프로세스와 HMI, PLC, 센서 등의 장치 정보를 기반으로 개발되었으며, 공격 데이터는 MITRE ATT&CK for ICS 매트릭스로부터 네트워크에서 탐지 가능한 공격 전술과 기법을 반영하였다. 파인 튜닝된 모델은 네트워크 공격 행위에 기반하여 생성한 프롬프트 셋에 대해 90%의 정확도를 보였으며, 도메인 지식이 없는 기본 모델 대비 우수한 성능을 보였다. 일부 부정확한 탐지는 학습 데이터 부족에 기인하며, 추가 데이터셋 개발과 재학습을 통해 정확도 개선이 가능할 것으로 예상된다. 본 연구는 스마트 빌딩 환경에서 대규모 언어 모델 기반 침입 탐지 시스템의 효과적인 적용 가능성을 입증한다.

Abstract

Smart building automation systems provide convenience and efficiency but remain vulnerable to cyberattacks. This study proposes an intrusion detection model specifically designed for smart buildings using a large language model. Targeting HVAC systems, normal and malicious behaviors were identified to create training datasets, and the LLaMA 3 8B model was fine-tuned. Benign datasets were developed from physical processes and devices such as HMIs, PLCs, and sensors, while malicious datasets incorporated attack tactics and techniques from the MITRE ATT&CK for ICS matrix. The fine-tuned model demonstrated 90% accuracy on evaluation prompts, significantly outperforming the base model. Inaccuracies were attributed to limited training data, suggesting that additional dataset development and retraining could further enhance performance. These results demonstrate the effectiveness of large language models in intrusion detection for smart building environments.

한글키워드 : 스마트 빌딩, 침입 탐지 시스템, 빌딩 자동화 시스템, 대규모 언어 모델, 파인 튜닝

keywords : smart building, intrusion detection system, building automation system, large language model, fine tuning

* 단국대학교 일반대학원 인공지능융합학과
** 단국대학교 소프트웨어학과
*** 단국대학교 모바일시스템공학과
**** 동국대학교 국제정보보호대학원 정보보호학과
† 교신저자: 조성제(email: sjcho@dankook.ac.kr)
접수일자: 2024.12.08. 심사완료: 2024.12.13.
게재확정: 2024.12.20.

1. 서 론

스마트 빌딩은 HVAC(Heating, Ventilating, and Air Conditioning), 조명, 화재 감지, 보안, UPS(Uninterruptible Power Supply) 등 다양한

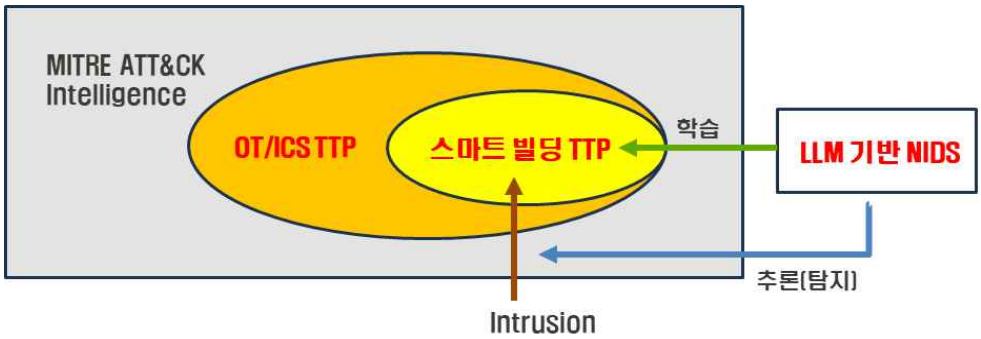

그림 1. 스마트 빌딩 환경의 대규모 언어 모델 기반 네트워크 침입탐지시스템
fig 1. Network Intrusion Detection System based on LLM in Smart Building Environment

하위 시스템을 통합적으로 운영하며, 에너지 절약, 운영 비용 절감, 안전 강화 등의 이점을 제공한다[1]. 이러한 빌딩 자동화 시스템(Building Automation System, BAS)은 스마트 빌딩 운영의 핵심이지만, OT(Operation Technology) 시스템의 특성상 사이버 공격에 취약하다[2, 3, 4]. OT 시스템은 안정성과 자동화에 중점을 두어 개발되었으나, 외부 네트워크와의 연결 시 보안 측면에서 상당한 위험을 내포한다[5].

대표적인 사례로 2019년 대만 TSMC의 랜섬웨어 공격과 2021년 콜로니얼 파이프라인 사건은 OT 시스템의 사이버 보안이 얼마나 중요한지 보여준다. TSMC의 경우 약 48시간 동안 시스템이 중단되며 3,000억 원의 경제적 손실을 입었고, 콜로니얼 파이프라인은 약 5,500마일의 파이프라인 운영이 마비되었다[6]. 이러한 사례는 스마트 빌딩의 사이버 보안을 강화하기 위한 침입 탐지 시스템(Intrusion Detection System, IDS)의 필요성을 대두시켰다. IDS는 네트워크 트래픽 및 시스템 활동을 실시간으로 모니터링하여 비정상 행위나 악성 공격을 탐지한다.

침입 탐지 시스템은 크게 오용 탐지(Misuse Detection)와 이상 탐지(Anomaly Detection)로 나뉜다[7]. 오용 탐지는 알려진 공격 패턴을 데이터베이스와 비교하여 탐지하는 방식으로, 기존 공격 탐지에는 효과적이나 새로운 유형의 공격에는 취약하다. 반면, 이상 탐지는 정상적인 동작 패턴과 다른 행위를 비정상으로 간주하여 탐지하는 방식으로, 알려지지 않은 공격 탐지에는 효과적이지만 높은 오탐율이 단점이다[8, 9].

이와 같은 문제를 해결하기 위해 본 논문에서는 대규모 언어 모델(Large Language Model, LLM)을 활용하여 스마트 빌딩 환경에 특화된 침입 탐지 시스템의 적용 가능성을 탐색한다. LLM은 방대한 텍스트 데이터를 학습하여 언어의 문맥적 관계를 이해하고 자연스러운 텍스트를 생성할 수 있는 기술로[10, 11], 대표적으로 OpenAI의 GPT-4와 META의 LLaMA 3 등이 있다. 특히, LLaMA 3는 80억 개 및 700억 개 매개변수를 가지며, 사전 학습과 파인튜닝(Fine-tuning)을 통해 다양한 도메인에 적용 가능하다. 본 논문에서는 그림 1과 같이 스마트 빌딩의 OT 네트워크 환경을 대상으로 LLM 기반의 네트워크 침입 탐지 시스템을 개발한다. 스마트 빌딩 OT 네트워크는 IoT(Internet of Things) 기반의 내장형 시스템으로 구성되어, 침입탐지를 위한 에이전트를 설치할 수 없기 때문에 네트워크 트래픽을 기반으로 한 침입탐지 시스템을 고려하였다. 사전 학습된 대규모 언어 모델에 스마트 빌딩 환경에 특화된 정보를 학습 데이터로 개

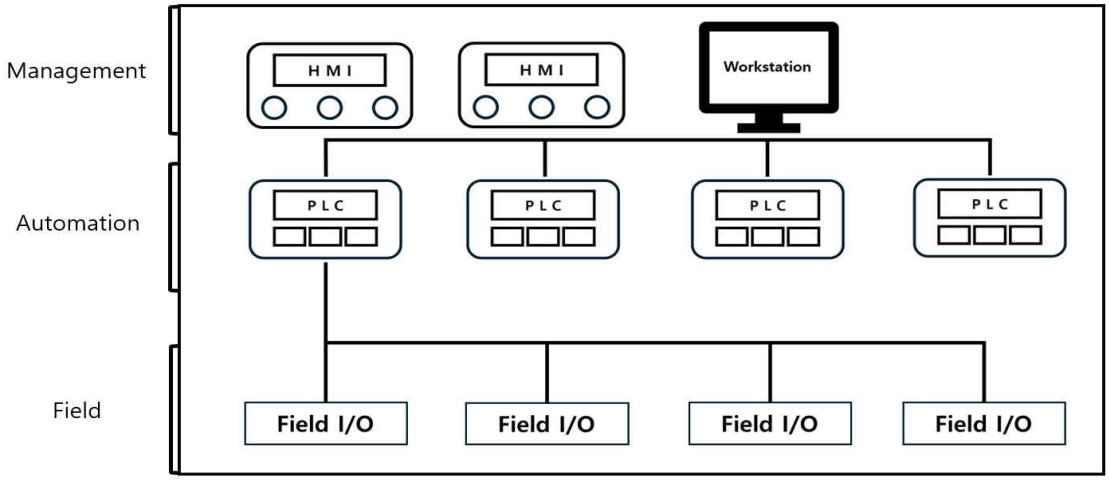

그림 2. 빌딩 자동화 시스템의 구조
fig 2. Architecture of A Building Automation System

발하여, 침입을 판정할 수 있도록 파인튜닝을 한다.

본 논문의 구성은 다음과 같다. 2장에서는 빌딩 자동화 시스템과 HVAC 시스템의 구조를 설명하고, 3장에서는 침입 탐지 모델의 설계와 동작 시나리오를 제시한다. 4장에서는 실험 설계와 결과를 분석하며, 5장에서는 결론과 향후 연구 방향을 논의한다.

2. 스마트 빌딩 HVAC 시스템

2.1 빌딩 자동화 시스템 구조

빌딩 자동화 시스템은 빌딩 내 다양한 시스템을 통합적으로 제어하고 관리하는 OT의 한 유형으로, 거주자의 편안함을 유지하고 에너지 소비를 줄이며, 운영 및 유지 관리 비용을 감소시키는 동시에 보안을 강화하고 장비 상태를 모니터링하는 등의 기능을 제공한다[12, 13]. 이러한 시스템은 그림 2와 같이 필드 계층(Field Layer), 자동화 계층(Automation Layer), 관리 계층(Management Layer)으로 구성된다[14, 15].

필드 계층은 물리적 환경에 가장 가까운 계층으로, 빌딩 내의 센서와 액추에이터가 위치한다. 센서는 온도, 습도, CO_2 농도와 같은 환경 데이터를 모니터링하며, 액추에이터는 냉온수 순환펌프, 환기팬, 댐퍼 등의 장치를 제어하여 환경 조건을 조정한다. 이 계층은 다양한 프로토콜을 사용하거나 간단한 전기 신호를 통해 자동화 계층의 컨트롤러와 통신하며, 물리적 데이터를 실시간으로 제공한다.

자동화 계층은 필드 계층의 데이터를 수집하고 제어를 실행하는 핵심 역할을 담당한다. 이 계층에는 PLC(Programmable Logic Controller)나 DDC(Direct Digital Controller)가 포함되며, 센서로부터 데이터를 전달받아 사전에 정의된 로직에 따라 액추에이터를 제어한다. 또한, 자동화 계층은 네트워크를 통해 다른 컨트롤러 또는 관리 계층과 데이터를 교환하며, 빌딩 내 물리적 프로세스를 효율적으로 운영한다.

관리 계층은 빌딩 자동화 시스템에서 상위 계층으로, 전체 시스템의 데이터를 통합하고 분석하

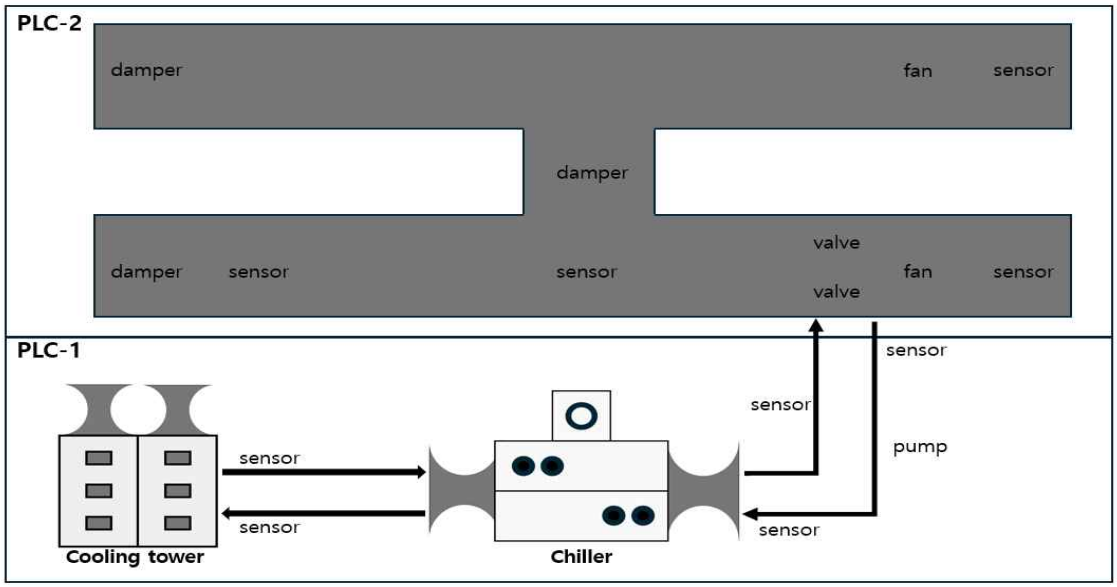

그림 3. HVAC 시스템 구성
fig 3. HVAC System Construction

며 사용자가 시스템을 제어할 수 있는 인터페이스를 제공한다. 관리 계층에서는 HMI (Human-Machine Interface)를 통해 센서 및 액추에이터 데이터를 시각적으로 확인하고 제어할 수 있으며, 시스템의 로그를 기록하여 유지보수 및 결함 감지를 지원한다. 또한, 전체 시스템 상태를 분석하여 최적의 에너지 소비를 유지하도록 돕는다.

2.2 스마트 빌딩 HVAC 시스템 구조

빌딩 자동화 시스템 내 HVAC(Heating, Ventilation, and Air Conditioning) 시스템은 그림 3과 같다. 필드 계층에서는 온도, 습도, CO_2 농도를 측정하는 센서와 냉온수 순환펌프, 환기팬, 혼합 댐퍼 등의 액추에이터가 상호 작용하며, 자동화 계층에서는 PLC가 데이터를 처리하여 냉난방 밸브 제어, 냉각수 흐름 조절, 공조 작업을 수행한다. 관리 계층에서는 HMI를 통해 시스템의 상태를 모니터링하고, 필요에 따라 제어 명령을 내려 환경 조건을 조정한다.

HVAC 시스템은 빌딩 자동화 시스템의 계층 구조를 통해 유기적으로 작동하며, 거주자의 편안함을 유지하고 에너지 소비를 효율적으로 관리하는 데 기여한다. 이러한 구조는 시스템의 안정적인 운영과 효과적인 관리에 필수적인 역할을 한다[16, 17].

3. 실험 설계

3.1 HVAC 물리 프로세스

침입 탐지 시스템이 작동할 HVAC 시스템은 빌딩 자동화 시스템에 특화된 환경으로 구성되었다. 이 환경은 냉난방 및 공조 기능을 포함하며, 각각의 물리적 프로세스는 PLC-1과 PLC-2에 의해 제어된다. PLC는 다양한 센서와 액추에이터를 통해 데이터를 수집하고 제어하며, HMI는 시스템 상태를 모니터링하거나 제어 명령을 설정하는 역할을 수행한다.

구체적으로, 해당 시스템은 냉온수기 1대, 냉각탑 1대, PLC 2대, HMI 2대, 여러 센서 및 액추에이터로 구성되며, Modbus 프로토콜을 사용하여 각 구성 요소 간의 통신이 이루어진다. 냉난방 제어는 PLC-1이 담당하며, 냉온수 공급온도, 환수온도, 출구온도와 냉각수 공급온도 및 환수온도 등의 센서 데이터를 기반으로 냉온수 순환펌프와 냉각수 순환펌프를 제어한다.

공조 시스템은 PLC-2가 담당하며, 환기온도, 외기온도, 혼합온도, 실내온도, 실내 CO_2 농도 등의 센서 데이터를 수집하여 이를 바탕으로 환기팬, 급기팬, 배기댐퍼, 외기댐퍼, 혼합댐퍼, 난방밸브 및 냉방밸브를 제어한다.

HMI는 침입 탐지 시스템의 주요 인터페이스로 작동하며, HMI-1을 통해 사용자는 센서 데이터 및 액추에이터 상태를 시각적으로 확인할 수 있다. 또한, HMI-2는 제어 명령을 통해 액추에이터 상태를 직접 변경할 수 있다. 각 센서 및 액추에이터의 값에 대한 세부 설명은 표 1과 표 2와 같다.

3.2 침입 탐지 모델

본 연구에서는 빌딩 자동화 시스템에서 발생하는 정상 행위와 공격(악성) 행위를 탐지하기 위해 정상/공격 데이터 수집, 학습 데이터 생성, 대규모 언어 모델 학습 단계를 순서로 침입 탐지 모델을 구현하였다.

스마트 빌딩용 침입 탐지 모델을 개발하기 위해 개발 환경과 조건에 따라 적합한 기반 모델 선정이 필요하다. 본 연구에서는 한정된 컴퓨팅 자원과 소규모 데이터 등의 제약을 고려하여 LLaMA 3의 80억 개의 매개변수를 가진 모델을 사용하였다.

정상/공격 데이터는 MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) for ICS(Industrial Control System)[18], 빌딩 자동화 시스템 환경과 장치 정보와 오픈 소스를 활용하여 수집하였다. 여기에는 빌딩 자동화 시스템의 프로토콜, 공격 기법, 네트워크 정보(IP 주소, MAC 주소, 포트 번호와 장치 스펙 등)가 포함된다. 수집된 데이터는 전처리 과정을 통해 부정확한 정보와 사용할 수 없는 정보들을 제거하여 분석 가능한 형태로 변환되었다.

학습 데이터셋 생성 단계에서는 수집된 텍스트 데이터를 전문가의 검토를 통해 질문과 답변 형식으로 개발하였다. HVAC 시스템의 자산(PLC, HMI, Sensor, Actuator), 물리 프로세스의 네트워크 행위 등의 특성을 반영하여 정상 행위 데이터셋을 다음과 같은 사례를 포함하여 생성하였다.

- Question: How is the HVAC system in BAS HVAC system configured?
- Answer: The HVAC system in Smart Building BAS includes PLC-1, PLC-2, HMI-1, HMI-2, multiple sensors, and actuators. PLCs periodically read data from sensors, and HMIs provide an interface to monitor data values. Communication occurs using the Modbus/TCP protocol, and devices are connected in a star network topology.

- Question: What data is exchanged between the cooling water circulation pump and PLC-1 in Smart Building BAS?
- Answer: The cooling water circulation pump sends ON/OFF signals to PLC-1, with ON corresponding to 1 and OFF to 0. Deviations from these expected values could signal malfunctions or transmission errors.

표 1. HVAC 시스템 공조기 관련 센서 및 액추에이터 목록
Table 1. List of sensors and actuators related to an air handling unit in the HVAC system

Name	Type	Range	PLC
Return_air_temperature_value	Input	-20 ~ 60℃	2
Outside_air_temperature_value	Input	-20 ~ 60℃	
Mixing_air_temperature_value	Input	-20 ~ 60℃	
Exhaust_air_temperature_value	Input	-20 ~ 60℃	
Exhaust_air_damper_state	Output	0 ~ 100%	
Outside_air_damper_state	Output	0 ~ 100%	
Mixing_air_damper_state	Output	0 ~ 100%	
Exhaust_air_damper_mode	Control	0 ~ 100% / Auto	
Outside_air_damper_mode	Control	0 ~ 100% / Auto	
Mixing_air_damper_mode	Control	0 ~ 100%	
Heating_valve_state	Output	0 ~ 100%	
Heating_valve_mode	Control	0 ~ 100% / Auto	
Cooling_valve_state	Output	0 ~ 100%	
Cooling_valve_mode	Control	0 ~ 100% / Auto	
Return_fan_state	Output	On/Off	
Return_fan_mode	Control	On/Off	
Supply_fan_state	Output	On/Off	
Supply_fan_mode	Control	On/Off/Auto	
Inside_temperature	Input	-20 ~ 60℃	
Inside_CO_2	Input	0 ~ 1100CO2	

표 2. HVAC 시스템 냉온수 및 냉각수 관련 센서 및 액추에이터 목록
Table 2. List of sensors and actuators related to cooling water and chilled water in the HVAC system

Name	Type	Range	PLC
Chilled_water_supply_value	Input	05 ~ 30℃	1
Chilled_water_return_value	Input	05 ~ 30℃	
Chilled_water_exit_value	Input	05 ~ 30℃	
Cooling_water_supply_value	Input	20 ~ 40℃	
Cooling_water_return_value	Input	20 ~ 40℃	
Chilled_water_circulation_pump_status	Output	On/Off	
Chilled_water_circulation_pump_mode	Contorl	On/Off/Auto	
Cooling_water_circulation_pump_status	Output	On/Off	
Cooling_water_circulation_pump_mode	Control	On/Off/Auto	

표 3은 생성된 정상 행위 데이터셋으로 스마트 빌딩 정의 및 구조, HVAC 시스템 정의 및 구조, 정의된 HVAC 물리 프로세스로 구분하였다.

표 3. 정상 행위 데이터셋
Table 3. Normal behavior Dataset

분류 기준	개수
스마트 빌딩 정의 및 구조	13
HVAC 시스템 정의 및 구조	22
HVAC 물리 프로세스	87
합계	122

악성 행위 데이터셋은 MITRE ATT&CK for ICS 공격 기법을 기반으로 네트워크에서 탐지 가능한 공격 행위 등의 특성을 반영하여 다음과 같은 사례를 포함하여 생성하였다.

- Question: What does the technique Modify Program (Technique ID: T0889) involve in the context of adversarial actions against smart building HVAC control systems?
- Answer: The Modify Program technique (Technique ID: T0889) involves adversaries modifying or adding a program on a controller to affect its interaction with physical processes, peripheral devices, and other network hosts. This can be done through methods like Program Download, online edit, and program append, allowing for changes in instructions and logic within Programmable Logic Controller (PLC) and similar programming elements.

- Question: What are some of the targeted assets in a smart building HVAC control system that may be affected by the exploitation of remote services (Technique ID: T0866)?
- Answer: Some of the targeted assets in a smart building HVAC control system include Control Server, Data Historian, Human-Machine Interface (HMI), and Programmable Logic Controller (PLC).

생성된 악성 행위 QA는 표 4는 공격 전술(Tactic)별 공격 기법(Technique)에 대해 생성된 악성 행위 데이터셋이다.

표 4. 공격자의 기법 기반 악성 행위 데이터셋
Table 4. Adversarial techniques-based malicious behavior Dataset

Tactic	Technique ID	Technique name	개수
Initial Access	T0866	Exploitation of remote services	6
	T0822	External remote services	11
	T0886	Remote services	7
Execution	T0858	Change Operating Mode	6
	T0807	Command-Line Interface	11
	T0871	Execution through API	3
	T0823	Graphical User Interface	4
	T0821	Modify Controller Tasking	5
	T0853	Scripting	11
	T0863	User Execution	6
Persistence	T0891	Hardcoded Credentials	5
	T0889	Modify Program	6
	T0839	Module Firmware	6
	T0857	System Firmware	6
Evasion	T0872	Indicator Removal on Host	5
	T0856	Spoof Reporting Message	5
Privilege Escalation	-	-	-
Discovery	T0842	Network Sniffing	11
	T0846	Remote System Discovery	5
	T0888	Remote System Information Discovery	6
	T0887	Wireless Sniffing	8
Lateral Movement	T0812	Default Credentials	6
	T0867	Lateral Tool Transfer	4
	T0843	Program Download	8

Collection	T0830	Adversary-in-the-Middle	14
	T0802	Automated Collection	5
	T0868	Detect Operating Mode	5
	T0801	Monitor Process State	5
	T0861	Point & Tag Identification	4
	T0845	Program Upload	6
Command and Control	T0885	Commonly Used Port	4
	T0884	Connection Proxy	3
	T0869	Standard Application Layer Protocol	5
Inhibit Response Function	T0800	Active Firmware Update Mode	3
	T0878	Alarm Suppression	6
	T0803	Block Command Message	3
	T0804	Block Reporting Message	4
	T0805	Block Serial COM	5
	T0892	Change Credential	5
	T0814	Denial of Service	15
	T0816	Device Restart/Shutdown	6
	T0835	Manipulate I/O image	5
	T0838	Modify Alarm Settings	6
Impair Process Control	T0806	Brute Force I/O	3
	T0836	Modify Parameter	4
	T0855	Unauthorized Command Message	3
Impact	-	-	-
합계			270

4. 모델 평가

MITRE ATT&CK for ICS 모델에서 제시된 공격 전술별 공격 기법을 기반으로 특정 공격 기법이 실행되었을 때 발생할 수 있는 네트워크 행위를 기반으로 생성된 프롬프트를 사용하여 파인튜닝된 대규모 언어 모델의 침입 탐지 성능을 평가하였다. 평가에 사용된 질문 예는 다음과 같으며, 30개를 개발하여 평가하였다.

(1) Quetsion: HMI2 sent a broadcast packet to PLC2 using a protocol other than Modbus TCP. Is this behavior normal?

(2) Question: A PLC used destination port 502 but communicated with HMI using a protocol other than Modbus. Is this behavior normal?

(3) Question: HMI2 made repeated connection attempts to multiple external IPs in a short time. Is this behavior normal?

라마3 기본 모델은 30개의 질문에 대해 응답 공백(<|end_of_text|>)을 반환했으며, 이는 모델이 도메인 지식이 부족하여 적절한 응답을 생성하지 못했음을 나타낸다. 반면, 파인튜닝된 라마3 모델은 빌딩 자동화 시스템에서 정상 및 악성 행위를 반영한 도메인 특화 데이터셋으로 학습되었기 때문에 정상 행위와 악성 행위를 효과적으로 구분하였다. 특히, 모델의 성능 평가는 학습 데이터와는 별개로, MITRE ATT&CK for ICS 모델에서 제시된 공격 기법을 기반으로 설계된 프롬프트를 사용하여 이루어졌다. 이러한 프롬프트는 침입 행위에 따라 발생할 수 있는 네트워크 이상 행위를 묘사하며, 모델이 학습 데이터의 단순 암기가 아닌 일반화된 판단 능력을 보유하고 있는지를 검증하기 위해 설계되었다.

평가 데이터는 실제 시나리오를 반영하여 총 30개의 질문으로 구성되었으며, 파인튜닝된 모델의 응답은 정상 행위와 악성 행위를 효과적으로 분류하는 데 중점을 두었다. 예를 들어, 다음과 같은 질문과 응답 예시이다.

(1) Question: HMI1 communicated with PLC2 using a protocol other than Modbus TCP (SSH, SMB, RDP). Is this behavior normal?

- Base Model : <|end_of_text|>
- Fine-tuning Model: This behavior is not normal. If HMI1 communicates with PLC2 using protocols other than Modbus TCP, it could indicate malicious actions by an attacker

(2) Question: HMI1 transmitted 124 bytes of data to an external IP. Is this behavior normal?

- Base Model : <|end_of_text|>
- Fine-tuning Model: This is not normal behavior. An average HMI typically transmits less than 100 bytes of data per request. Excessive data transmission could indicate malicious actions by an attacker.

표 5와 같이 평가 결과 파인튜닝된 모델은 30개의 질문 중 27개의 질문에 대해서 정상 행위는 정상 행위로 악성 행위는 악성 행위로 올바르게 구분했다. 특히, 모델은 정상 행위와 악성 행위를 구분할 뿐만 아니라, 악성 행위로 판단될 경우 이에 대한 구체적인 근거를 제시하며 응답하였다. 예를 들어, 정상적인 HMI의 데이터 전송 패턴(평균 100바이트 이하)을 기준으로 데이터 전송량의 이상 여부를 판단하거나, Modbus 이외의 프로토콜 사용이 악성 행위와 관련될 가능성을 제시하였다. 이러한 응답은 단순한 예/아니오 형태를 넘어, 도메인 지식을 기반으로 한 설명을 제공함으로써 침입 탐지 시스템의 해석 가능성을 높인다.

표 5. 평가 결과
Table 5. Evaluation results

실제 \ 예측	정상 행위	악성 행위
정상 행위	15(TN)	0(FP)
악성 행위	3(FN)	12(TP)
Accuracy	90%	

공격을 탐지하지 못한 경우, 공격 탐지에 대한 프롬프트 관련된 도메인 지식이 충분히 제공되지 않았음을 확인하였으며, 이를 보완하는 경우에는 탐지율의 향상이 예상된다.

5. 결론 및 향후 연구

본 논문에서는 대표적인 OT 환경인 스마트 빌딩 환경을 대상으로 대규모 언어 모델 기반의 침입 탐지 모델을 제안하였다. 파인 튜닝된 라마 3 모델은 기본 라마3 모델과 비교하여 정상 행위와 악성 행위에 대해 구분할 수 있어 정확도와 오탐율 측면에서 효과적인 성능을 보였다.

그러나, 파인 튜닝에 사용된 학습 데이터 셋이 제한적이기 때문에 모델이 학습할 수 있는 정상과 악성 행위의 다양성에 한계가 있을 수 있다. 또한, 30개의 질문에 대한 평가로 구성되어 있어 모델의 성능을 폭넓게 검증하지 못했다. 따라서, 향후 연구에서는 파인 튜닝에 사용되는 데이터의 규모를 확대하여 다양한 악성 행위에 대해 학습하고 평가 개수를 늘려 보다 체계적인 성능 검증을 수행할 예정이다. 이를 통해, 스마트 빌딩과 같은 OT 시스템에서 사이버 보안에 특화된 침입 탐지 시스템으로 대규모 언어 모델의 적용 가능성을 높이고자 한다.

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구과제임(No. RS-2021-KP002461).
또한, 2021년도 정부(과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(no. 2021R1A2C2012574).
또한, 과학기술정보통신부 및 정보통신기획평가원의 학석사연계ICT핵심인재양성사업의 연구결과로 수행되었음(IITP-2023-00259867).

참 고 문 헌

[1] S. Ann, S. -J. Cho and H. Kim, “A Preliminary Study on an Intrusion Detection Method using Large Language Models in Industrial Control Systems”, 2024 Fifteenth International Conference on Ubiquitous and Future Networks (ICUFN), Budapest, Hungary, 2024, pp. 600-602, doi: 10.1109/ICUFN61752.2024.10625633.

[2] H. Kanamaru, “Requirements for IT/OT Cooperation in Safe and Secure IACS”, 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Chiang Mai, Thailand, 2020, pp. 39-44, doi: 10.23919/SICE48898.2020.9240295.

[3] M. Bristow. (2021). A SANS 2021 Survey: OT/ICS Cybersecurity, SANS Institution. https://www.sans.org/white-papers/SANS-2021-Survey-OTICS-Cybersecurity/.

[4] Fortinet Inc. (2022). State of Operational Technology and Cybersecurity Report, https://www.fortinet.com/content/dam/fortinet/assets/analyst-reports.

[5] Dean Parsons. (2022). The State of ICS/OT Cybersecurity in 2022 and Beyond. SANS Institution. https://www.sans.org/whitepapers/state-ics-ot-cybersecurity-2022-beyond/

[6] J. Beerman, D. Berent, Z. Falter and S. Bhunia, “A Review of Colonial Pipeline Ransomware Attack”, 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), Bangalore, India, 2023, pp. 8-15, doi: 10.1109/CCGridW59191.2023.00017.

[7] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat and S. Venkatraman, “Deep Learning Approach for Intelligent Intrusion Detection System”, in IEEE Access, vol. 7, pp. 41525-41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[8] Garcia-Teodoro, Pedro, et al. “Anomaly-based network intrusion detection: Techniques, systems and challenges”, computers & security 28.1-2 (2009): 18-28, doi:10.1016/j.cose.2008.08.003.

[9] R. Vinayakumar, K. P. Soman and P. Poornachandran, “Applying convolutional neural network for network intrusion detection”, 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 1222-1228, doi: 10.1109/ICACCI.2017.8126009.

[10] Shafee, Samaneh, Alysson Bessani, and Pedro M. Ferreira. “Evaluation of llm chatbots for osint-based cyber threat awareness”, arXiv preprint arXiv:2401.15127 (2024).

[11] Ferrag, Mohamed Amine, et al. “Generative AI and Large Language Models for Cyber Security: All Insights You Need”, arXiv preprint arXiv:2405.12750 (2024).

[12] W. Kastner, G. Neugschwandtner, S. Soucek and H. M. Newman, “Communication systems for building automation and control”, in Proceedings of the IEEE, vol. 93, no. 6, pp. 1178-1203, June 2005, doi: 10.1109/JPROC.2005.849726.

[13] Domingues, Pedro, et al. “Building automation systems: Concepts and technology review”, Computer Standards & Interfaces 45 (2016): 1-12, doi: 10.1016/j.csi.2015.11.005
[14] Ciholas, Pierre, et al. “The security of smart buildings: a systematic literature review”, arXiv preprint arXiv:1901.05837 (2019).
[15] Graveto, Vitor, Tiago Cruz, and Paulo Simões. “Security of Building Automation and Control Systems: Survey and future research directions”, Computers & Security 112 (2022): 102527, doi: 10.1016/j.cose.2021.102527.
[16] G. Goddard, J. Klose and S. Backhaus, “Model Development and Identification for Fast Demand Response in Commercial HVAC Systems”, in IEEE Transactions on Smart Grid, vol. 5, no. 4, pp. 2084-2092, July 2014, doi: 10.1109/TSG.2014.2312430.
[17] Fong, Kwong Fai, Victor Ian Hanby, and Tin-Tai Chow. “HVAC system optimization for energy management by evolutionary programming”, Energy and buildings 38.3 (2006): 220-231, doi: 10.1016/j.enbuild.2005.05.008.
[18] “MITRE ATT&CK(Adversarial Tactics, Techniques, and Common Knowledge)”, https://attack.mitre.org/

저 자 소 개

안석현(Seokhyun Ann)

2024.8 단국대학교 소프트웨어학과 학사
2024.9-현재 단국대학교 인공지능융합학과 석사과정
<주관심분야> 운영기술 보안, AI를 활용한 악성코드 탐지 등

박수현(Suhyeon Park)

2020.3-현재 단국대학교 소프트웨어학과
<주관심분야> 운영기술 보안, AI 보안

김도익(Doik Kim)

2021.3-현재 단국대학교 모바일시스템공학과
<주관심분야> 운영기술 보안, AI 보안

조성제(Seong-je Cho)

1989.2 서울대학교 컴퓨터공학과 공학사
1991.2 서울대학교 컴퓨터공학과 공학석사
1996.8 서울대학교 컴퓨터공학과 공학박사
1997.3-현재 단국대학교 소프트웨어학과/
컴퓨터학과 교수
<주관심분야> 디지털포렌식, 시스템 보안 및 악성코드 분석, 인공지능 보안, 시스템 소프트웨어 등

김홍근(Honggeun Kim)

1994.5 한국전산원
1996.5 한국인터넷진흥원
<주관심분야> 컴퓨터보안, 정보보호