대규모 언어 모델과 검색 증강 생성 기법을 사용한 스마트 빌딩용 침입 탐지 시스템

김도익, 안석현, 박수현, 조성제, 김홍근
단국대학교 모바일시스템공학과, 단국대학교 인공지능융합학과, 단국대학교 소프트웨어학과, 동국대학교 국제정보보호대학원 정보보호학과
{doik0920, seokhyun, parksh, sjcho}@dankook.ac.kr, hgkim4044@gmail.com

1. 서론

스마트 빌딩은 편리성과 에너지 효율성을 제공하지만 사이버 공격에 취약함.

o 기존 IDS
- 시그니처/오용탐지 → 새로운 유형 공격에 취약
- 이상 탐지 → 높은 오탐률

o 해결 방안: 스마트 빌딩의 HVAC의 복잡한 도메인 특성을 반영하고, 명확한 설명이 가능한 탐지 시스템

o 연구 목표: 도메인 특화 지식을 반영한 정밀 탐지를 위해, 외부 문서 기반 추론이 가능한 RAG 기반 스마트 빌딩 침입 탐지 기법 개발

2. 데이터셋

o 정상 데이터: ICSSIM 을 변형한 가상 시뮬레이터 IBEEMS SIM 에서 수집한 제어 네트워크 행위

o 공격 데이터: MITRE ATT&CK for ICS 네트워크 기반 공격 행위

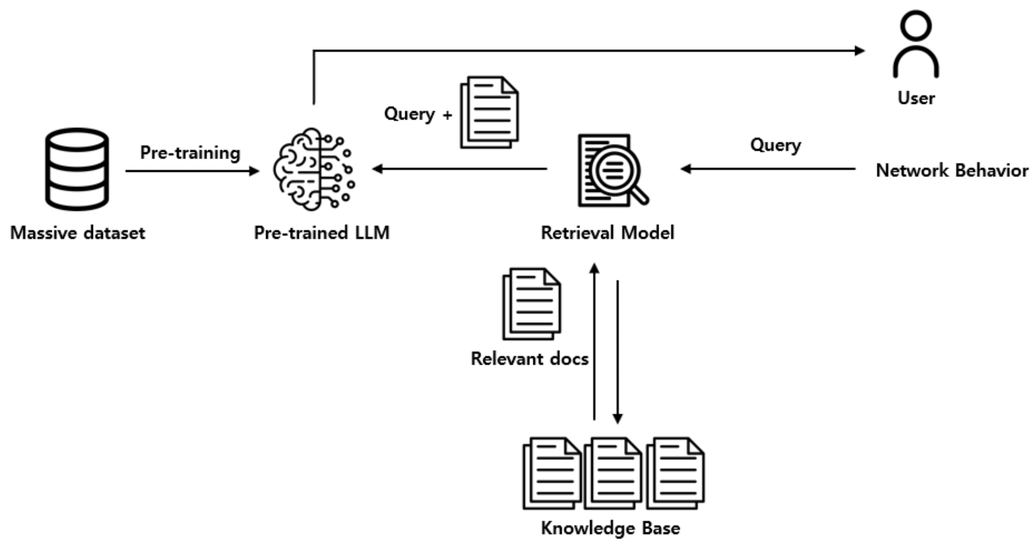
o 구축된 데이터는 RAG 시스템의 **지식 베이스에 저장**

3. 시스템 동작 순서

Step 1. 쿼리 입력
자연어 형태의 네트워크 행위의 쿼리 입력

Step 2. 문서 검색
쿼리를 임베딩하여 벡터 검색 엔진이 지식베이스의 관련 문서를 검색

Step 3. LLM 응답 생성
검색된 문서 + 쿼리를 입력받아, LLM이 침입 여부 판단 및 근거 설명



4. 실험 방법

벤치마크 데이터셋
- 모든 실험은 동일한 평가 쿼리 100개를 사용하여 수행됨

1. 임베딩 모델

- 의미적 표현력과 연산 효율성을 고려하여 3개 임베딩 모델 비교 평가
- 모델별 평가 쿼리에 대해 상위 5개 문서와의 코사인 유사도 평균을 측정하여 비교

임베딩 모델	Top5 평균 유사도
e5-base-v2	0.85
gte-base-v2	0.84
all-mpnet	0.50

- 실험 결과 e5-base-v2가 평균 0.8504로 가장 높은 유사도를 보였으며, HVAC 데이터셋과 높은 의미적 적합성을 유지

2. 벡터 검색 엔진
- HNSW + 코사인 유사도 기반으로 쿼리당 상위 5개 문서를 빠르게 검색
- 평균 검색 지연 시간(ms)을 성능 지표로 측정
- RAG 기반 실시간 침입 탐지 시스템에서 빠른 문서 검색과 문서의 정확도는 핵심요소

유형	검색 엔진	평균 지연 시간(ms)
벡터 검색 라이브러리	FAISS	0.1166
벡터 데이터 베이스	Chroma	4.2811
	Elasticsearch	9.3435
	Redis	0.8245
	Weaviate	6.1109
	Milvus	3.9275
	Qdrant	2.6764

- FAISS가 평균 0.1166ms로 가장 빠른 응답 속도를 기록하여 실시간 응용에 적합
- 검색된 문서의 신뢰성과 일관성이 우수

3. 대규모 언어 모델(LLM)

- 온프레미스에서 사용 가능한 LLM 4종 평가
- LLM 모델별 RAG 적용/미적용 Precision과 Recall 차이를 비교하고, 응답을 통해 설명력을 평가

Model	Precision (Non-RAG)	Recall (Non-RAG)	Precision (RAG)	Recall (RAG)
LLaMA 8B	0.48	0.64	0.98	0.97
Mistral 7B	0.21	0.18	0.94	0.96
Gemma 7B	0.67	0.80	1.00	0.98
Falcon 7B	0.52	0.54	0.84	0.86

정상/비정상 판단 예시

질문: HMI1이 PLC2로 Function Code 14 (Read File Record)를 전송. 정상인가?
RAG 모델 응답 요약:
- HMI1은 일반적으로 PLC2에 Function Code 3 (read)명령만 전송
- Function Code 14는 일반적인 읽기 작업과 관련이 없는 것으로 판단
- Function Code 3 이외의 명령은 비정상 행위로 간주됨
- 정상 동작과 불일치 -> **비정상 행위로 판단**

o 실험 결과:
- RAG 모델은 검색된 문서를 바탕으로 문맥 기반 추론 응답 수행
- 단순 생성이 아닌 상황에 맞는 논리적 판단 제공
- RAG 적용 시 성능 향상 (Mistral 7B 기준)
  - Precision: +73%p
  - Recall: +78%p
- RAG 미적용 모델 대비 설명 가능성과 정밀도 모두 향상

5. 결론

- 스마트 빌딩 HVAC 제어 환경에 **특화된 RAG 기반 침입 탐지 시스템**
- RAG 미적용 모델은 판단 유보 또는 모호한 응답 빈도 높음
- RAG 구조는 도메인 문서를 기반으로 판단
- **탐지 정확도뿐 아니라, 시스템의 신뢰성과 해석 가능성 향상에 효과적**

● 한계점
- 본 연구는 시뮬레이터 기반 네트워크 행위 데이터셋에 한정됨
- 실제 환경에서의 공격 시나리오 반영 부족

6. Acknowledgement

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제임 (No. RS-2021-KP002461). 또한 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 학석사연계ICT 핵심인재양성사업의 연구결과로 수행되었음 (IITP-2023-00259867). 또한 본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (No. 2024-0-00035).