JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 7, pp. 117-126, July 2026
https://doi.org/10.9708/jksci.2026.31.07.117

A Two-Stage Framework for Cybersecurity Scenario Matching Using Embedding-Based Top-K Filtering and LLM Contextual Verification

Jihun Jang*, Jungpyo Hong**

*Staff, Dept. of Information and Technology, Korea Institute of Materials Science, Changwon, Korea
**Associate Professor, Dept. of Information and Communication Engineering, Changwon National University, Changwon, Korea

[Abstract]

Traditional Security Information and Event Management and Security Orchestration, Automation, and Response systems are widely used to detect cyber threats. However, rule-based detection methods have difficulty identifying complex attacks that require contextual understanding. This study proposes a two-stage threat scenario matching system that combines vector embedding and Large Language Model analysis. First, the system retrieves the Top-K most similar threat scenarios using embedding-based similarity search. Then, the LLM compares the candidate scenarios with the input security alerts and selects the best match. Experimental results using simulated breach data demonstrated that the detection accuracy, which was 91.54% in the embedding stage, was significantly enhanced through the LLM's contextual re-verification, successfully identifying all threat scenarios with zero misclassifications.

▸ **Key words**: Security Automation, Alert Sequence Analysis, Security Threat Detection, Vector Embedding, LLM

[요 약]

지능화되는 사이버 위협에 대응하기 위해 정보보호 이벤트 통합관리(SIEM)와 보안 오케스트레이션·대응자동화(SOAR) 솔루션이 널리 활용되고 있다. 그러나 기존 규칙 기반 탐지 방식은 장기간에 걸친 공격의 전체 맥락을 반영하지 못해 우회 공격 식별에 한계가 있다. 또한 딥러닝 기반 탐지 기법은 높은 성능이 기대되지만 학습 데이터 확보와 모델 학습 비용으로 인해 실무 적용에 제약이 존재한다. 본 연구에서는 이를 극복하기 위해 임베딩 거리 기반 Top-K 필터링과 대규모 언어 모델(LLM)의 맥락 검증을 결합한 2단계 보안위협 시나리오 매칭 시스템을 제안한다. 제안 시스템은 기관의 보안위협 시나리오를 벡터 임베딩해 지식베이스를 구축하고, 입력된 보안 경보 시퀀스와의 유사도 검색으로 상위 K개 후보를 선별한다. 이후 LLM이 맥락을 검증해 최종 시나리오를 선택한다. 실험결과 1차 벡터 검색에서는 k=3일 때 98.46%의 Top-k 정확도를 달성했으며, LLM 기반 맥락 검증을 통해 벡터 공간의 유사성으로 발생한 오분류를 모두 교정할 수 있음을 확인하였다.

▸ **주제어**: 보안 자동화, 경보 분석, 위협 탐지, 벡터 임베딩, 대규모 언어모델

• First Author: Jihun Jang, Corresponding Author: Jungpyo Hong
*Jihun Jang (jangjh0305@kims.re.kr), Dept. of Information and Technology, Korea Institute of Materials Science
**Jungpyo Hong (hansin@changwon.ac.kr), Dept. of Information and Communication Engineering, Changwon National University
• Received: 2026. 05. 20, Revised: 2026. 06. 05, Accepted: 2026. 07. 14.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

지능화·고도화되는 사이버 위협과 기관의 전산 인프라 팽창은 대량의 보안 경보를 발생시키고 있으며, 이는 인력 중심의 수동 보안 관제의 한계를 야기하고 있다. 이에 대응하여 정보보호 이벤트 통합관리(이하 SIEM) 및 보안 오케스트레이션·대응 자동화(이하 SOAR) 등 자동화 기반의 대응 체계가 적극적으로 도입되는 추세이다. 그러나 이러한 솔루션들은 조건문이나 임계치 기반으로 동작하여 정교한 우회 공격이나 장기간에 걸쳐 맥락적으로 진행되는 지능형 지속 위협(APT)을 식별하는 데 기술적 한계를 보인다. 이를 보완하기 위해 머신러닝·딥러닝 기반의 위협 탐지 모델 연구가 활발히 진행 중이나, 실제 운영 환경에서는 양질의 학습 데이터 확보가 어려워 기관별 특화된 환경에 부합하는 모델 구축에 많은 제약이 따른다. 이에 본 연구에서는 별도의 추가 모델 학습 없이도 이기종 보안 시스템의 경보 맥락을 분석하여 유연하고 정확하게 위협을 식별하는 새로운 자동화 탐지 파이프라인을 제안하고자 한다.

본 연구는 데이터 부족과 모델 학습 비용 문제로 AI 기반 보안 체계 도입을 망설이는 기관에 즉각적인 대안을 제시하며, 온프레미스(On-Premise) 인프라를 활용함으로써 내부 보안 로그의 외부 유출 없이 고도화된 위협 탐지 및 대응 자동화 체계를 신속하게 구축할 수 있음을 증명했다는 점에서 학술적·실무적 의의를 지닌다.

본 논문의 구성은 다음과 같다. 제2장에서는 보안 자동화 시스템, 위협헌팅과 시나리오 모델링 등 연구의 이론적 배경을 설명한다. 제3장에서는 전통적인 보안위협 분석과 최근 활발히 연구되는 AI 기반 보안위협 분석기법과 관련된 선행연구를 분석하고 기존 연구의 한계와 본 연구의 필요성을 제시한다. 제4장에서는 벡터 임베딩 기반 후보 시나리오 선정과 LLM 기반 맥락 검증으로 구성된 제안 기법의 전체 아키텍처와 동작과정을 설명한다. 제5장에서는 제안 기법의 성능을 검증하기 위한 실험 환경을 구성하고 단계별 검증 결과와 구성요소 제거 실험을 통한 제안기법의 정확도와 계산 효율성을 평가한다. 마지막으로 제6장에서는 연구 결과를 정리하고 향후 연구방향을 제시한다.

II. Preliminaries

컴퓨팅 역량의 비약적인 발전과 디지털 전환 수요의 급증으로 인해 사이버 공격 기법 역시 고도화·지능화되는 추세이다. 이로 인해 전통적인 인적 자원 중심의 수동 보안 관제 체계는 다량의 오탐(False Positive) 발생에 따른 경보 피로와 침해사고 대응 지연이라는 물리적 한계에 직면하게 되었다[1]. 이러한 한계를 극복하고자 최근 보안 담당자들은 수집된 위협 데이터를 단일 플랫폼에서 실시간으로 통합 분석하고, 탐지된 위협에 즉각적으로 대응할 수 있는 보안 자동화 시스템 구축 연구에 집중하고 있다.

1. Security Automation System

현재 보안 관제 현장에서 활발히 운용 중인 대표적인 보안 자동화 시스템으로는 SIEM과 SOAR 플랫폼이 존재한다. SIEM과 SOAR는 독립적으로 작동하기보다, SIEM의 탐지 기능과 SOAR의 대응 기능이 유기적으로 연계되어 보안 운영의 효율성을 극대화하는 방향으로 발전해 왔다[2].

이러한 보안 자동화 시스템의 기술적 진화 과정과 자동화 수준에 따른 단계별 특징을 요약하면 Table 1.과 같다.

Table 1. Maturity Levels of Security Automation Systems

Level	Platform	Approach
1	Legacy System	Human-centric manual review of individual logs
2	SIEM	Rule-based correlation analysis of heterogeneous logs
3	SOAR	Event-driven security alert sequence analysis
4	AI-SOC	AI Driven context-aware verification

2. Threat Hunting

방화벽, 침입차단시스템 등 전통적인 정보보호 시스템은 알려진 공격 패턴인 시그니처 기반의 수동적 방어 메커니즘을 기반으로 하고 있다. 이러한 방식은 제로데이 취약점을 악용한 공격에 대해 탐지 공백을 야기할 수 있다. 위협 헌팅은 이를 극복하기 위해, 내부 인프라에 이미 위협 행위자가 침투해 있다는 가정을 전제로 수행되는 선제적·능동적 보안 대응 방법론이다. 위협 헌팅의 프로세스는 Fig 1.와 같이 가설 수립 및 검증 단계로 구성된다.

위협 헌팅은 지속적인 가설 검증과 탐지 시나리오의 고도화를 통해 기관 내 잔존 위협을 최소화하는 것을 목적으로 한다[3]. 보안 담당자의 분석 역량과 정보보호 시스템 간의 유기적인 연결을 형성함으로써, 수동적 방어 중심의 보안 패러다임을 기계적·자동화된 능동적 방어 영역으로 확장하는 중추적인 역할을 수행한다.

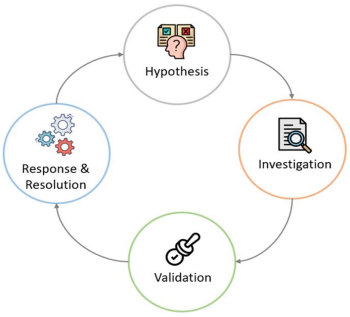

Fig. 1. Process of Threat Hunting

3. Threat Scenario Modeling

위협 시나리오 모델링은 잠재적 위협과 공격 경로를 체계적으로 식별하고 분석하기 수단으로써 Fig 2.와 같이 대상 추상화, 위협 식별, 공격 시나리오 분석, 위험도 평가, 대응책 도출 과정을 통해 수행된다. 특히 공격 시나리오 분석 단계에서는 개별 보안 이벤트를 독립적으로 분석하는 것이 아니라 공격 행위 간의 연관성과 인과관계를 고려하여 공격 경로를 모델링한다. 이를 통해 복합적인 공격 흐름을 이해하고 잠재적 보안위협을 체계적으로 분석할 수 있는 것으로 알려져 있다[4].

모델링된 위협 시나리오는 다양한 정보보호 시스템으로부터 생성되는 보안경보를 실제 공격 전술과 연결하기 위한 분석 요소으로 활용되며 위협 헌팅 과정에서 도출된 공격 징후와 연계하여 분석 효율성 향상과 보안 자동화 체계의 기반을 제공한다.

Fig. 2. Process of Threat Modeling

III. Related Work

본 장에서는 보안경보 분석, 위협 시나리오 식별, 보안 자동화와 관련된 주요 선행 연구를 전통적 분석 기법과 최근 활발하게 연구되고 있는 AI 기반 분석 기법으로 구분하여 고찰한다. 이를 통해 기존 연구들의 성과와 한계를 분석하고, 기관의 보안 운영 환경에 적용 가능한 보안위협 시나리오 매칭 시스템의 필요성을 제시한다.

1. Traditional Analysis Methods

1.1 Security Alert Sequence Analysis

보안 경보 시퀀스 분석은 지능형 지속 위협(APT)과 같이 단계적으로 진행되는 공격 행위를 탐지하기 위한 대표적인 연구 분야이며, 개별 보안 이벤트를 독립적으로 분석하는 대신, 시간 순서에 따라 발생한 경보들의 연관성과 공격 흐름을 시퀀스 형태로 모델링하여 공격자의 행위를 식별한다. 이를 위해 순환신경망, 장단기 기억 네트워크(LSTM), Transformer 기반 모델 등을 활용하여 경보 간의 관계를 학습하고 이상 행위를 탐지하는 연구가 활발히 수행되어 왔다. 대표적으로 대규모 시계열 로그 데이터로부터 내부자 이상 행위를 탐지하기 위해 LSTM 기반 잡음 제거 오토인코더를 활용하여 시퀀스 정보를 잠재 벡터 형태로 추출하고, 이를 Local Outlier Facto 및 Isolation Forest와 같은 이상 탐지 기법과 결합하는 방법이 제안되었다[5]. 해당 연구는 시퀀스 정보를 효과적으로 압축함으로써 오탐률을 감소시키고 복합적인 이상 행위를 탐지할 수 있음을 보였다.

그러나 이러한 비지도 학습 기반 접근법은 정상 패턴과 상이한 시퀀스를 탐지하는 데에는 효과적이지만, 탐지된 이벤트가 실제로 어떠한 공격 시나리오에 해당하는지 설명하기 어렵다는 한계를 가진다. 또한 시퀀스 길이와 잠재 벡터 차원에 따라 탐지 성능이 크게 변동하며, 실제 운영 환경에서 요구되는 위협 맥락 분석 기능은 충분히 제공하지 못한다.

1.2 Security Scenario Matching

보안 시나리오 매칭은 복잡한 공격 과정을 사전에 정의된 시나리오와 비교하여 위협을 식별하는 접근 방식이다. 기존 연구에서는 보안 전문가의 지식과 과거 침해사고 사례를 기반으로 공격 시나리오를 정형화하고 이를 탐지 체계에 반영하는 방법들이 제안되었다[6]. 또한 MITRE ATT&CK 프레임워크를 활용하여 국가 배후 공격 조직의 침투 과정과 산업기술 유출 사례를 공격 단계별로 매핑하고 대응 전략을 수립하는 연구도 수행되었다[7].

그러나 이러한 접근법은 주로 ATT&CK 전술·기술(TTP) 수준의 분류와 시각화에 초점을 두고 있어, 실제 운영 환경에서 발생하는 보안경보 시퀀스를 조직이 정의한 보안 위협 시나리오와 맥락적으로 연결하는 데에는 한계가 있었다. 또한 대부분 전문가의 규칙 정의와 수동 분석에 의존하여 대규모 경보 환경에서 자동화된 시나리오 매칭을 수행하기 어렵다는 제약이 존재한다.

2. AI-driven Analysis Methods

2.1 LLM-based Security Operations Center Automation

최근에는 대규모 언어모델의 강력한 문맥 이해 능력과 추론 능력을 활용하여 보안관제 업무를 자동화하려는 연구가 활발히 진행되고 있다. 특히 LLM은 서로 다른 출처에서 발생한 보안 이벤트를 종합적으로 해석하고, 개별 경보를 상위 수준의 공격 시나리오와 연결할 수 있어 복잡한 보안 위협 분석에 적합한 기술로 주목받고 있다. 최근에는 AI 에이전트, 강화학습 기반 대응 전략, 지식 그래프를 결합한 자율형 SOC 아키텍처가 제안되었으며, 이를 통해 알려지지 않은 위협에 대한 적응형 탐지 및 대응 체계를 구축하고자 하였다[8].

이러한 연구는 보안관제의 전 과정을 자동화할 수 있는 미래 지향적 비전을 제시하였다는 점에서 의미가 있다. 그러나 복잡한 지식 그래프 구축과 다중 에이전트 운영에 필요한 높은 계산 비용으로 인해 일반적인 기업 환경이나 온프레미스 환경에서 즉시 적용하기에는 현실적인 제약이 존재한다.

2.2 RAG-based Security Analysis

대규모 언어모델의 환각 문제를 완화하고 최신 위협 정보를 활용하기 위해 검색 증강 생성을 보안 분석에 적용하는 연구가 활발히 진행되고 있다. 이와 관련하여 정적 분석 기법과 Fine-tuning, 그리고 벡터 데이터베이스 기반 RAG를 결합하여 소프트웨어 취약점(CWE/CVE)을 탐지하는 하이브리드 분석 체계를 제안하고 Faithfulness 지표를 통해 생성된 결과의 사실 일관성을 평가함으로써 보안 분석 결과의 신뢰성을 정량적으로 검증한 사례가 있다[9].

이러한 접근법은 외부 지식을 활용하여 LLM의 추론 신뢰성을 향상시켰다는 점에서 의의가 있으나, 주로 취약점 분석이나 문서 기반 질의응답에 초점을 두고 있다. 따라서 실시간으로 생성되는 다수의 보안경보와 공격 행위 간의 맥락적 관계를 종합적으로 분석하고, 이를 조직이 정의한 보안위협 시나리오와 연결하는 문제는 충분히 다루지 못하였다. 또한 검색된 지식의 유사도에 크게 의존하기 때문에, 유사한 공격 패턴이 다수 존재하는 환경에서는 정확한 시나리오 식별이 어려울 수 있다는 한계가 있다.

2.3 Adaptive Playbooks for SOAR

SOAR는 다양한 보안 솔루션을 통합하고 표준화된 대응 절차인 플레이북을 자동 수행함으로써 운영 효율을 향상시키는 기술이다. 최근 연구에서는 고정된 대응 절차의 한계를 극복하기 위해 침해사고의 특성에 따라 대응 과정을 동적으로 재구성하는 적응형 플레이북 생성 기법이 제안되었다[10]. 해당 연구는 기존 정적 플레이북 대비 평균 대응 시간을 약 73% 단축할 수 있음을 보고하였다.

그러나 적응형 플레이북 연구는 새로운 위협에 대응하기 위한 플레이북 분기 생성과 대응 절차 최적화에 초점을 맞추고 있다. 따라서 현재 발생한 보안 이벤트가 조직이 보유한 보안위협 시나리오와 얼마나 부합하는지를 분석하거나, 경보의 맥락을 기반으로 적절한 위협 시나리오를 식별하는 기능은 상대적으로 부족하다. 결과적으로 대응 자동화 수준은 향상될 수 있으나, 플레이북 적용의 근거가 되는 위협 시나리오 판단 과정은 여전히 제한적으로 다루어지고 있다는 한계가 존재한다.

3. Summary and Remaining Challenges

기존 연구들은 보안 경보 시퀀스 분석, 시나리오 기반 위협 모델링, RAG 기반 보안 분석, SOAR 자동화, 그리고 LLM 기반 자율형 SOC 구축 등 다양한 관점에서 보안 운영의 자동화를 발전시켜 왔다. 보안경보 시퀀스 분석 연구는 대규모 보안 로그로부터 이상 행위를 효과적으로 탐지할 수 있음을 입증하였으며, 시나리오 기반 연구는 공격 절차를 체계적으로 모델링하고 분류할 수 있는 기준을 제

Table 2. Summary of Related work

Domain	Skill	Achievements	Limitations
Alert Analysis	LSTM, LOF/IF	Detected anomalous behaviors from security alert sequences	Unable to associate detected anomalies with specific threat scenarios
Scenario Matching	MITRE ATT&CK	Formalized threat scenarios and modeled multi-stage attack procedures	Lacks automated scenario matching for large-scale alert streams
RAG based Analysis	RAG Pipeline	Improved reliability of security analysis through external knowledge retrieval	Limited capability to correlate security alerts with specific scenarios
Adaptive Playbooks	CTI, Threat DB	Dynamically adapted response workflows and improved SOAR operational efficiency	Focuses on response optimization rather than identifying appropriate threat scenario
LLM based	AI agents	Proposed autonomous SOC architectures capable of adaptive threat analysis	High operational complexity hinder practical deployment

공하였다. 또한 RAG 기반 분석은 LLM의 신뢰성을 향상시켰고, SOAR 및 자율형 SOC 연구는 보안 대응의 자동화 가능성을 제시하였다.

그러나 기존 접근법들은 기관 내에서 발생하는 보안경보를 기관의 보안위협 시나리오와 연결하여 분석하는 문제를 충분히 해결하지 못하였다. 시퀀스 분석 기반 연구는 이상 행위를 탐지할 수 있으나 이를 구체적인 공격 시나리오와 연계하여 설명하기 어렵고, 시나리오 기반 연구는 여전히 전문가의 규칙 정의와 수동 분석에 크게 의존한다. 또한 RAG 기반 분석과 LLM 기반 자동화 연구는 풍부한 맥락 분석 능력을 제공하지만, 기관 환경에 부합할 수 있도록 최적화하기에는 비용과 운영 복잡성이 높다는 한계가 존재한다. 이러한 주요 선행연구의 특징과 한계를 비교한 결과는 Table 2.에 요약하였다.

따라서 본 연구에서는 보안경보와 사내 보안위협 시나리오 간의 의미적 유사성을 기반으로 후보 시나리오를 선정하고, LLM을 활용하여 경보 맥락과 시나리오의 적합성을 검증하는 2단계 보안위협 시나리오 매칭 기법을 제안한다. 이를 통해 기존 시퀀스 분석의 설명력 부족 문제와 규칙 기반 시나리오 매칭의 한계를 보완하고, 복잡한 자율형 SOC 아키텍처 없이도 실시간 보안관제 환경에서 활용 가능한 경량형 위협 시나리오 식별 체계를 제안하고자 한다.

IV. The Proposed Method

전통적인 딥러닝 기반 보안 모델은 정교하게 레이블링된 대규모 학습 데이터셋이 필수적으로 요구되나, 정보보호 분야의 특성 상 기관이 보유한 실제 로그 데이터나 침해사고 데이터는 외부 공유가 제한되어 양질의 대규모 데이터셋을 확보하여 범용 보안 모델을 구축하는 데에는 현실적인 한계가 존재한다[11].

본 연구에서는 이러한 문제를 해결하기 위해 기존 AI 기반 탐지 모델들이 직면하고 있는 데이터 획득의 폐쇄성과 모델 학습에 소요되는 높은 비용과 관련된 문제를 해결한 보안위협 탐지방안 도출에 초점을 맞추었다.

1. System Design and Architecture

1.1 On-Premises Based Analyzer

기관이 수집하는 보안 경보와 보안위협 시나리오는 IP 주소, 보유 정보보호 시스템 목록, 네트워크 구성과 같이 대외 유출 시 기관 정보보호 관리체계에 큰 위협이 될 수 있는 민감 정보들로 구성되어 있다. 이러한 정보들을 외부 SaaS 기반 LLM API를 통해 분석하게 될 경우 데이터 전송 과정에서의 유출 위험과 외부 모델의 학습 데이터로 활용될 가능성이 존재한다.

제안하는 방법은 기관의 보안 데이터의 기밀성을 원천적으로 보호하고 모든 데이터 처리가 기관 내부에서 수행될 수 있도록 온프레미스 환경에서 동작하도록 설계하였다. 이를 통해 데이터의 외부 유출을 근본적으로 차단하고, 기관이 요구하는 수준의 분석체계를 구현할 수 있다.

1.2 Two-Stage Analysis Architecture

기관에서 AI를 활용하여 업무 개선을 추진하는 과정에서 가장 걸림돌이 되는 사항은 '기술 및 IT 인프라의 부족'과 '비용 부담'으로 알려져 있다[12]. 제안하는 방법에서는 제한적인 연산 인프라를 가진 기관에서도 자동화된 보안위협 탐지체계를 구축할 수 있도록 단계별 분석 아키텍처를 설계하였다.

수집된 보안경보 시퀀스와 기관이 보유한 보안위협 시나리오 전체 텍스트에 대해 초기 단계에서부터 LLM을 통한 검증을 수행한다면 전체 시나리오에 대한 맥락적 검증이 가능하다는 장점이 있다. 그러나 이 방식은 기관이 위협헌팅을 수행하며 보안위협 시나리오가 확장될수록 프롬프트의 길이가 지나치게 길어지게 만들 가능성이 있다. 최근 연구에서는 프롬프트의 길이가 증가할수록 모델이 중간에 위치한 정보를 제대로 식별하지 못하고, 결과적으로 환각 현상의 심화와 답변 품질 저하로 귀결되는 원인이 되는 것으로 확인되었다[13]. 또한 프롬프트의 길이가 길어질수록 이를 처리하고 이해하기 위한 LLM 모델의 크기도 커질 수밖에 없으며 이는 더욱 많은 가용 자원을 요구한다.

제안하는 방법은 벡터 임베딩 기반의 1차 검색과 LLM 기반의 전체 맥락 검증의 2차 검색으로 단계를 구분하여 LLM이 처리하는 프롬프트의 길이를 최소화하였다. 이러한 아키텍처를 통해 제한된 컴퓨팅 환경에서도 답변 품질을 향상시키고 환각 현상을 최소화할 수 있도록 설계하였다.

2. System Architecture

제안하는 방법의 전체 아키텍처는 SIEM으로부터 수집된 보안 경보를 기관이 보유한 보안위협 시나리오와 맥락적으로 비교하기 위해 Fig 3.와 같이 2단계의 탐색 구조를 채택하였다. 제안된 아키텍처에서는 GloVe-100을 이용하여 보안경보와 보안위협 시나리오를 동일한 의미 공간의 벡터로 변환한 뒤 FAISS 인덱스에 저장하여 이후 거리 기반 후보 시나리오 선정에 활용된다. 선정된 후보 시나리오들을 LLM에 입력하여 시퀀스와 맥락적 적합성을 검증한

다. 시스템 구현에 사용된 주요 모델과 플랫폼은 Table 3.과 같다.

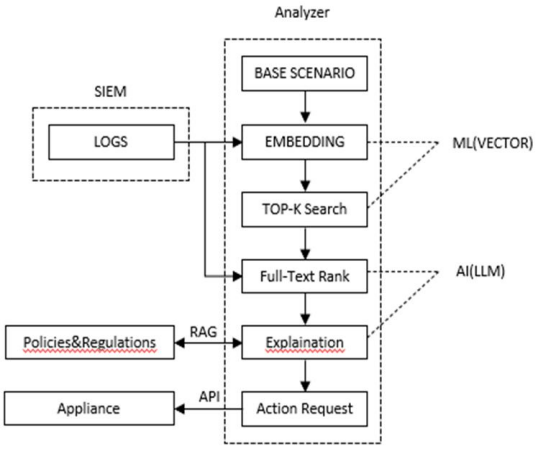

Fig. 3. Proposed Method System Architecture

Table 3. Technical specifications of models and platforms

Embedding	GloVe-100
Library	FAISS
LLM	Qwen2.5-14B-Instruct
Platform	vLLM

2.1 Knowledge Base and Data Preprocessing

기관에서는 정보보호시스템들로부터 수집되는 다종·다량의 정보보호 이벤트들을 통합 관리하기 위해 SIEM을 활용하여 기관 환경에 부합하는 '보안 경보'로 정규화하여 재정의하여 운용하고 있으며, 이렇게 정의된 보안 경보를 기반으로 SOAR의 플레이북 또는 자체 대응체계를 통한 대응 행위를 수행하고 있다.

이러한 환경을 모사하기 위해서 제안하는 방법에서는 SIEM에 적용할 수 있는 Table 4.와 같은 형식의 100가지 보안경보 목록을 정의하고, 해당 목록을 활용하여 기관에서 발생 가능한 13종의 보안위협 시나리오를 Table 5.와 같은 형식으로 설계하였다.

Table 4. Cybersecurity Alert Format

Format	TITLE(metadata):IP:Agent
Example	Connection failed from untrusted IP(192.0.1.27):10.10.20.5:Firewall

Table 5. Example of a Cybersecurity Threat Scenario

No.	SC_001
Title	Unauthorized Access and Persistence Establishment
Alerts	· Failed connection attempt from an untrusted IP · Failed connection attempt from an untrusted IP · Successful connection established from an untrusted IP · Administrative privilege escalation granted · CRONTAB scheduled task created

2.2 Vector Embedding-Based Top-K Retrieval

정의된 보안 시나리오 목록을 벡터 임베딩 모델을 활용하여 벡터 변환을 수행한 뒤 FAISS에 저장해두고, 입력된 보안 경보 시퀀스를 동일한 임베딩 모델을 통해 변환을 수행하여 각 시나리오의 변환벡터와 L2 Norm 계산을 수행한다. 이 중 가장 근접한 K개의 후보 시나리오를 다음 단계에 활용될 수 있도록 선정한다.

또한, 단순 오경보와 보안 시나리오에 부합하지 않는 경보 시퀀스인 경우 후속 검증 단계에서 참고할 수 있도록 수식 1과 같이 '확실성(Certainty)' 지표를 추가하여 전달한다. 확실성 지표는 거리를 평균으로 스케일링한 뒤 음의 지수함수를 적용하고 Softmax 연산을 수행함으로써 후보 시나리오들 간 상대적으로 얼마나 근접한지 비교할 수 있도록 활용된다.

$$C_j = \frac{e^{-\frac{d_j}{\bar{d}}}}{\sum_{i=1}^{k} e^{-\frac{d_i}{\bar{d}}}} \times 100 \,(\bar{d} = mean(d)) \qquad (1)$$

2.3 LLM-Based Context Comparison and Validation

벡터 임베딩 기반 후보 시나리오 선정 단계를 거쳐 입력된 K개의 원본 시나리오와 수집된 원본 보안경보 시퀀스를 LLM 모델을 통해 검증함으로써 정의된 보안위협 시나리오와 가장 부합하는 최종 시나리오를 선정한다.

LLM 모델의 답변 일관성을 확보하기 위해 Table 6.와 같은 요건을 반영하여 구조화된 프롬프트를 적용하여 시나리오 최종 선정, 판단 근거, 공격 분석 결과, 후속조치 방안을 제안하도록 설계하였다.

Table 6. Components of the LLM Prompt

Context	1) List of candidate scenarios 2) Certainty indicators 3) Security alert sequence
Const.	1) Select exactly one option strictly from the candidate scenarios 2) Prohibit the generation of any unlisted scenarios 3) Select exclusively from the predefined list of information security system mitigation actions
Logic	1) Correlate and compare the operational flow and context between the event sequence and cyber threat scenarios 2) Define the handling procedure when the certainty level is approximately 30% 3) Propose an optimal security system countermeasure tailored to the identified threat
Output	1) Specify the formatting rules for the selected scenario, logical justification, attack analysis, and risk assessment 2) Describe the essential verification requirements and actionable remediation steps 3) Incorporate a toggle/option for recommending security policy enforcement rules

V. Experiments

1. Environment and Datasets

본 연구에서 제안하는 탐지방안의 자원 효율성을 검증하기 위해 보편적으로 기관에서 활용하고 있는 워크스테이션급 사양의 서버를 Table 7.과 같이 구성하였다.

이를 통해 고사양 GPU 클러스터와 같은 대규모 인프라를 구성하기 어려운 대다수 기관에서도 효과적인 탐지체계를 구성하는데 기여할 수 있을 것으로 판단된다.

실험을 위해 100종의 보안 경보를 정의하고, 이를 조합하여 Table 8.과 같이 13종의 보안 위협 시나리오 및 표준 시퀀스(Ground Truth)를 구성하였다. 테스트 데이터로는 실제 보안 관제 현장에서 발생하는 오경보 상황을 모사하기 위해, 표준 시퀀스에 상관관계가 없는 보안 경보를 무작위로 혼입하여 시나리오별로 10개씩 제작하여 결과적으로 총 130개의 평가 데이터셋을 구축하였다.

Table 7. Experimental Specifications

H/W	**CPU**	Intel(R) Xeon(R) Gold 6442Y
	GPU	NVIDIA L40S
	RAM	512GB
	Storage	2TiB
S/W	**OS**	Ubuntu 22.04
	CUDA	12.4

Table 8. Threat Scenarios Used in the Experiments

SC_01	Unauthorized Access and Persistence Establishment
SC_02	Malicious File Creation and Outbound Communication
SC_03	Account Creation, Privilege Escalation, and Service Registration
SC_04	Web Shell Upload, Execution, and Persistence Maintenance
SC_05	System Information Discovery and Data Exfiltration
SC_06	Indicator Defense Evasion and Persistence Establishment
SC_07	Backdoor Persistence Configuration and Outbound Connection Attempts
SC_08	Privilege Escalation and Critical System File Modification
SC_09	Network Reconnaissance and DoS Attack Attempts
SC_10	Phishing-based Malware Infiltration
SC_11	Internal Reconnaissance and Lateral Movement for Malware Propagation
SC_12	Website Vulnerability Exploitation and Mass Data Exfiltration
SC_13	Anomalous Connection and Internal Network Discovery Detection

2. Top-K Retrieval Stage

1차 후보 시나리오 선별 단계의 유효성을 검증하기 위해 인공지능 연구 전반에서 모델의 분류 성능을 측정하기 위해 보편적으로 채택되는 Top-k 정확도를 평가지표로 활용하였다.

구축된 테스트 데이터를 대상으로 Top-k 정확도를 분석한 결과 Fig 4.와 같은 결과를 얻었다. k를 3으로 설정하였을 때 98.46%의 높은 분류 정확도를 기록하였으며, 이를 통해 1차 단계에서 3개의 후보 시나리오를 선정할 경우 2차 검증 단계로 전달되는 컨텍스트의 길이를 최소화하는 동시에 탐지 방안의 전체적인 신뢰도를 보장할 수 있는 최적의 임계치임을 확인하였다.

3. LLM-Based Validation Stage

1차 단계에서 확실성 수치상 상위 순위로 선정되지 못한 후보 시나리오에 대해 LLM을 통해 전체 맥락적 의미를

해석하여 정답을 교정하도록 검증하였으며, 실험 결과 재검증을 통해 전량 정답 시나리오를 선정함을 확인하였다.

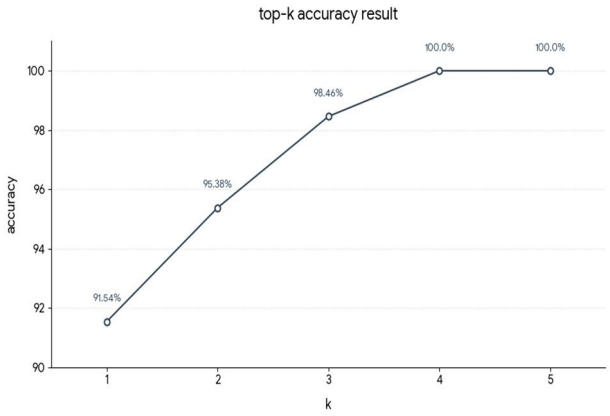

Fig. 4. Top-K accuracy result

Fig 5.에서와 같이 실제 테스트 사례 분석 결과 입력된 보안 경보 시퀀스는 접속시도와 권한상승 및 지속성 확보와 관련된 보안경보 위주로 구성되어 있다. 1차 검색 단계에서는 SC_03이 확실성 37.41%로 가장 높은 유사도를 보이지만, SC_01과 37.23%로 차이가 0.18% 내외로 근소하여 오탐을 유발하였다. 이는 두 시나리오 모두 '권한 상승'이라는 공통 요소를 포함하고 있어 벡터 공간 상 거리가 가깝게 형성되었기 때문으로 판단된다.

그러나 LLM은 시나리오의 세부 패턴과 로그 시퀀스를 대조하여 이를 교정하였는데, SC_03의 핵심 지표인 '사용자 계정 생성'이 부재하며, SC_01의 '초기침투-지속성 확보' 흐름과 부합함을 도출하였다.

결과적으로 단순히 확실성 순위가 높은 시나리오를 배제하고 보안경보 시퀀스가 보유한 핵심 과정을 가장 잘 설명하는 보안위협 시나리오를 최종 정답으로 재분류함을 확인함으로써 제안하는 탐지방안이 실제 보안 관제 현장에서도 오탐률을 낮출 수 있는 실무적 방법임을 증명하였다.

4. Ablation Study

제안하는 보안위협 시나리오 매칭 기법은 임베딩 기반 후보 시나리오 검색 단계와 LLM 기반 맥락 검증 단계로 구성된 2단계 구조를 채택한다. 이러한 구조는 임베딩 검색의 높은 검색 효율성과 LLM의 우수한 문맥 이해 능력을 상호 보완적으로 활용하기 위하여 선택되었다. 본 절에서는 제안하는 방법을 구성하는 각 요소가 탐지 성능 향상에

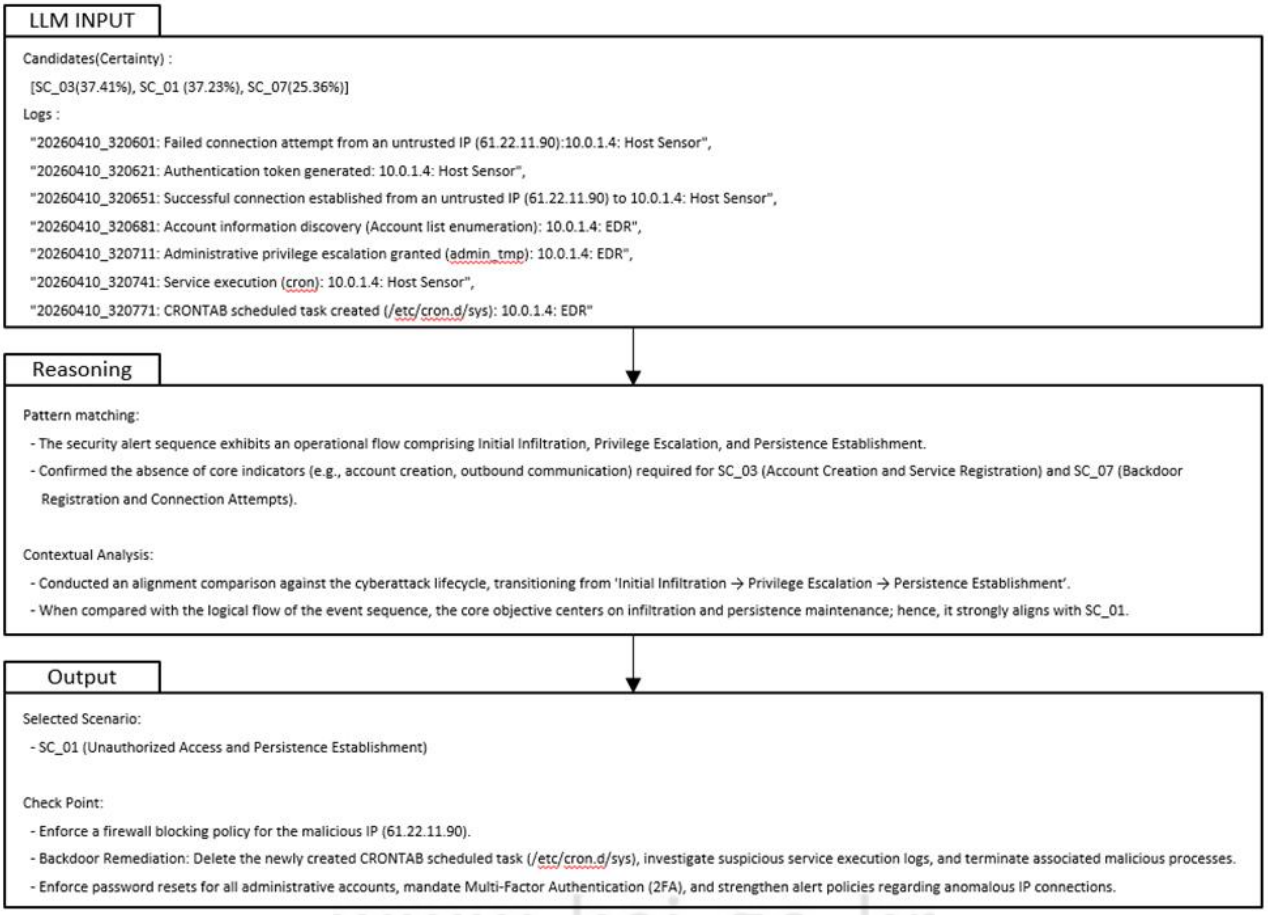

Fig. 5. Results of the LLM Analysis

미치는 영향을 분석하기 위하여 구성 요소를 제거하여 실험하였다. 실험 대상은 Table 9.와 같이 세 가지 방법으로 구성하였다.

Table 9. Configurations Used in the Ablation Study

Method	Embedding	LLM
Method 1	O	X
Method 2	X	O
Method 3 (Proposed)	O	O

실험은 동일한 130개의 테스트 데이터셋과 동일한 서버 환경에서 수행하였다. 또한 실제 보안관제 환경에서의 적용 가능성을 평가하기 위해 정확도, 평균 입력 토큰 수, 평균 처리 시간을 함께 측정하였다. 이를 통해 탐지 정확도와 실제 운영 환경에서 중요한 계산 비용과 추론 효율성을 고려하여 종합적으로 비교하였다.

실험 결과는 Table 10.와 같다. 방법 1.은 임베딩 기반 유사도 검색만을 수행하여 가장 낮은 계산 비용으로 후보 시나리오를 즉시 반환할 수 있었다. 그러나 k=1로 설정한 경우 정확도는 91.54%에 그쳤으며 k=4까지 후보군을 확장한 이후에야 모든 정답 시나리오가 후보 내에 포함되었다. 이는 실제 운영환경에서는 최종 시나리오를 선택하기 위하여 추가적인 전문가의 판단이 필요하여 완전한 자동화에는 한계가 있음을 나타낸다. 이러한 오분류는 '권한 상승', '지속성 확보', '내부 정찰'과 같이 서로 유사한 공격 절차를 공유하는 시나리오 간 의미적 차이를 임베딩 유사도만으로 충분히 구분하지 못하기 때문으로 판단된다.

반면 방법 2.는 모든 보안위협 시나리오를 동시에 비교·분석하기 때문에 높은 시나리오 식별 정확도를 보였다. 그러나 모든 시나리오 정보를 하나의 프롬프트에 포함해야 하므로 입력 토큰 수가 크게 증가하였으며, 이에 따라 처리 시간이 증가하는 것으로 나타났다. 이러한 경향은 기관에서 관리하고 있는 보안위협 시나리오의 수량에 비례하여 프롬프트 길이가 선형적으로 증가하기 때문에 계산 비용과 확장성 측면에서 한계를 가진다.

제안하는 방법은 임베딩 검색을 통해 후보 시나리오를 Top-K로 제한함으로써 LLM이 분석해야 하는 탐색 공간을 효과적으로 축소하였다. 이어 후보 시나리오만을 대상으로 LLM이 보안경보 시퀀스의 발생 순서와 맥락적 관계를 종합적으로 분석하여 임베딩 기반 검색에서 발생할 수 있는 의미적 모호성과 오분류를 교정하였다. 결과적으로 제안하는 방법은 방법 2. 대비 적은 입력 토큰과 추론 비용으로 동일한 시나리오 식별 정확도를 유지하면서도 평균 입력 토큰 수는 약 65% 감소하였으며, 평균 처리 시간은 약 71% 감소시켰다. 또한 분석할 시나리오가 설정한 k 값에 따른 후보군으로 제한되므로 기관이 보유한 보안위협 시나리오가 증가하더라도 일정한 프롬프트 길이를 제공함으로써 유연한 확장 가능성을 가지고 있다.

이와 같은 결과는 제안하는 2단계 보안위협 시나리오 매칭 구조가 단순히 두 기법을 결합한 것이 아니라, 임베딩 검색을 통한 탐색 공간 축소와 LLM 기반 의미 검증이 상호 보완적으로 작동함으로써 정확도와 계산 효율성을 동시에 확보할 수 있음을 보여준다. 따라서 제안 기법은 제한된 컴퓨팅 자원을 사용하는 기관 환경에서도 실시간 보안관제에 적용 가능한 경량형 AI 기반 보안위협 시나리오 매칭 기법으로 활용 가능함을 확인하였다.

VI. Conclusions

본 연구에서는 기관 환경에 부합하는 양질의 데이터셋 확보의 어려움과 모델 학습 비용 문제로 인해 도입에 현실적인 제한이 존재하는 기존 AI 기반 보안 탐지체계의 한계를 극복하기 위해 벡터 임베딩 기반 검색과 LLM 기반 검증을 결합한 2단계 보안위협 탐지방안을 제안하였다.

실험결과 1차 벡터 검색 단계에서 Top-k 정확도 분석결과 k값을 3으로 설정하였을 때 98.46%로 높은 정확도를 확보하였으며, 이후 LLM 기반 재검증을 통해 벡터 거리 기반으로는 구분이 어려운 시나리오 간의 맥락적 차이를 보완하여 최종적으로 정확한 시나리오로 분류해낼 수 있음을 확인하였다. 또한 구성 요소 제거 실험을 통해 제안하는 2단계 구조가 임베딩 기반 검색만을 활용한 방법보다 높은 식별

Table 10. Performance Comparison of the Compared Methods

Method	Accuracy	Avg. Prompt Tokens	Avg. Processing Time (s)
Method 1	91.54%(k=1) 100%(k=4)	N/A	X
Method 2	100%	1,620	11.6
Method 3	100%	570	3.4

정확도를 달성하였으며, 전체 시나리오를 LLM에 직접 입력하는 방법과 비교하여 평균 입력 토큰 수와 평균 처리 시간을 크게 감소시키면서도 동일한 정확도를 유지함을 확인하였다. 이를 통해 제안하는 방법이 제한된 연산 자원을 보유한 기관 환경에서도 유연하게 적용이 가능한 AI 기반 보안 자동화 체계로 활용될 수 있음을 입증하였다.

향후 연구에는 최종 분류된 보안위협 시나리오에 대하여 검색 증강 생성 기술을 적용하여 기관이 자체적으로 보유하고 있는 사이버공격 대응지침, 기관 내규, 자산 정보 등과 연계하여 맞춤형 대응 조언을 제공하는 방향으로 연구를 확장하고자 한다. 또한 LLM이 제안하는 정보보호 시스템 정책을 수행할 수 있도록 API 연계를 진행하여 지능형 대응 자동화 체계로 발전시키고자 한다.

REFERENCES

[1] T. Ban, T. Takahashi, S. Ndichu, and D. Inoue, "Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response," Applied Sciences, Vol. 13, No. 11, Art. No. 6610, May 2023. DOI: 10.3390/app13116610.

[2] R. Kremer, P. N. Wudali, S. Momiyama, T. Araki, J. Furukawa, Y. Elovici, and A. Shabtai, "IC-SECURE: Intelligent System for Assisting Security Experts in Generating Playbooks for Automated Incident Response," arXiv preprint, arXiv:2311.03825, Nov. 2023. DOI: 10.48550/arXiv.2311.03825.

[3] H. Ryu and I.-R. Jeong, "A Study on the Establishment of Threat Hunting Concept and Comparative Analysis of Defense Techniques," Journal of The Korea Institute of Information Security and Cryptology, Vol. 31, No. 4, pp. 793-799, Aug. 2021. DOI: 10.13089/JKIISC.2021.31.4.793.

[4] J. Gim, J. Kwak, K. Cho, and S. Kim, "Analysis of the Problems of Domestic Studies on Threat Modeling," Journal of The Korea Institute of Information Security & Cryptology, Vol. 35, No. 2, pp. 397-413, Apr. 2025.

[5] K. Sim and K. Kim, "Insider Anomaly Behavior Detection Method Using an Unsupervised Learning-Based Autoencoder," Journal of Digital Contents Society, Vol. 24, No. 8, pp. 1929-1936, Aug. 2023. DOI: 10.9728/dcs.2023.24.8.1929.

[6] M. Hong and Y. Lee, "A Study on the Scenario-Based Evaluation System for Predicting Corporate Security Accidents," Journal of the Korea Academia-Industrial Cooperation Society, Vol. 24, No. 3, pp. 409-415, Mar. 2023. DOI: 10.5762/KAIS.2023.24.3.409.

[7] G. H. Ahn, J. H. Oh, S. R. Yeo, and W. H. Park, "MITRE ATT&CK-based Framework for Preventing Industrial Technology Leakage," Review of KIISC, Vol. 31, No. 3, pp. 29-38, Jun. 2021.

[8] R. Sugumar, "Next-Generation Security Operations Center (SOC) Resilience: Autonomous Detection and Adaptive Incident Response Using Cognitive AI Agents," International Journal of Technology, Management and Humanities, Vol. 10, No. 2, pp. 62-76, Jun. 2024.

[9] J. Lee, "Design and Implementation of a Vulnerability Detection System Using RAG-based LLM Response Generation," Master's thesis, Chungnam National University, Feb. 2026.

[10] E. Özgözler, A. Varol, and İ. Tanrıverdi, "Event-Driven Agentic SOC (ED-ASOC): Supervisor-Based LLM Framework for Dynamic Incident Response and SOAR Orchestration," Proceedings of the 14th International Symposium on Digital Forensics and Security (ISDFS), Boston, MA, USA, pp. 1-5, 2026. DOI: 10.1109/ISDFS69419.2026.11458922.

[11] M. Lopez-Ledezma and G. Velarde, "Cyber Security Data Science: Machine Learning Methods and Their Performance on Imbalanced Datasets," Digital Management and Artificial Intelligence: Proceedings of the Fourth International Scientific-Practical Conference (ISPC 2024), pp. 529-542, 2025. DOI: 10.1007/978-3-031-88052-0_45.

[12] Korea Chamber of Commerce and Industry (KCCI) and Korea Institute for Industrial Economics & Trade (KIET), "Survey on the Status of AI Technology Utilization by Domestic Enterprises," Press Release, 2024. Available: https://www.korcham.net/nCham/Service/Economy/appl/KcciReportDetail.asp?SEQ_NO_C010=20120938915&CHAM_CD=B001

[13] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," Transactions of the Association for Computational Linguistics (TACL), Vol. 12, pp. 157-173, 2024. DOI: 10.1162/tacl_a_00638.

Authors

Jihun Jang received B.S. degree in Computer Science from the Republic of Korea Naval Academy, Korea, in 2018. He is currently pursuing the M.S. degree in Information and Communication Engineering at Changwon National University, Korea. He is currently an Cybersecurity Specialist with the Korea Institute of Materials Science. His research interests include AI security, artificial intelligence, cybersecurity policy design and strategy&planning.

Jungpyo Hong received Ph.D. degree in the School of Electrical Engineering from the Korea Advanced Institute of Science and Technology, Korea, in 2016. He is currently an Associate Professor of the Department of Information and Communication Engineering at ChangwonNational University. His research interests include noise reduction, array signal processing, and artificial intelligence.