사물인터넷융복합논문지 Vol. 12, No. 3, pp. 85-95, 2026 https://doi.org/10.20465/KIOTS.2026.12.3.085

SecuBERT: AI 기반 네트워크 이상행위 자동탐지 및 시각화 프레임워크

이형우*
한신대학교 AISW대학 교수

SecuBERT: AI-Based Network Anomaly Auto-Detection and Visualization Framework

Hyung-Woo Lee*
Professor, School of Computing and Artificial Intelligence, Hanshin University

요 약 최근 네트워크 환경의 복잡성 증가와 보안 위협의 다양화로 인해 다양한 보안 장비 및 서비스에서 생성되는 로그의 양이 급증하고 있으며, 정상 행위와 비정상 행위를 신속 정확하게 식별하는 것은 침해사고 대응 및 보안 관제의 핵심 과제이다. 그러나 기존의 탐지 방식은 로그 표현의 다양성과 행위 문맥의 차이를 충분히 반영하지 못한다는 문제점이 발생한다. 이에 본 연구에서는 프로파일 기반 의미 유사도 분석과 규칙 보정을 결합한 BERT 기반 네트워크 로그 이상행위 탐지 프레임워크 SecuBERT를 제안한다. 제안 프레임워크는 정상 및 비정상 행위를 설명하는 프로파일 문장을 사전에 정의하고, 입력 로그를 BERT 임베딩 공간에 매핑한 뒤 각 프로파일과의 코사인 유사도를 계산하여 행위 유형을 판정한다. 또한 키워드 힌트 기반 점수와 포트 및 프로토콜 특성에 따른 보정 규칙을 결합하여 탐지 결과의 실용성과 탐지 정확도를 향상시켰다. 제안한 SecuBERT 시스템은 대규모 라벨링 데이터셋이 제한적인 환경에서도 효과적으로 네트워크 이상탐지 기능을 제공하며, 향후 실제 운영 환경에서 수집되는 네트워크 이벤트 정보 기반 자동화된 탐지 시스템으로 확장 가능하다.

주제어 : BERT, 네트워크 로그, 이상행위 탐지, AI 기반 자동 탐지, 시각화, 유사도

Abstract With the increasing complexity of modern network environments and the diversification of security threats, the volume of logs generated by various security devices and services has grown rapidly, making the rapid and accurate identification of normal and abnormal behaviors a critical task in incident response and security monitoring. However, existing detection approaches have limitations in adequately capturing the diversity of log representations and the contextual differences of behaviors. To address this issue, this paper proposes SecuBERT, an improved BERT-based network log analysis system. The proposed framework defines profile sentences describing normal and abnormal behaviors, maps input logs into the BERT embedding space, and classifies events by measuring semantic similarity to the predefined profiles. In addition, keyword hint-based scoring and correction rules based on port and protocol characteristics are incorporated to improve the practicality and detection accuracy of the results. The proposed SecuBERT system demonstrated that effective network log-based anomaly detection is possible even in environments where large-scale labeled datasets are limited, and it also showed the potential to be extended into an AI-based automated detection system for real-world operational network log environments.

Key Words : BERT, Network Log, Anomaly Detection, AI based Detection, Visualization, Similarity

이 논문은 한신대학교 학술연구비 지원에 의해 연구되었음.
*교신저자 : 이형우(hwlee@hs.ac.kr)
접수일 2026년 05월 04일 수정일 2026년 06월 08일 심사완료일 2026년 06월 17일

1. 서론

정보시스템은 웹 서비스, 클라우드 인프라, 원격접속 환경, 내부 업무 시스템, API 기반 마이크로서비스 등으로 빠르게 확장되고 있으며, 이에 따라 네트워크 트래픽과 보안 이벤트 역시 기하급수적으로 증가하고 있다. 이러한 환경에서는 방화벽, IDS/IPS, VPN 게이트웨이, DNS 서버, 프록시 서버, 웹 서버, 데이터베이스 서버 및 다양한 보안 솔루션이 대량의 로그를 지속적으로 생성한다. 로그는 시스템 동작 상태와 사용자 활동, 보안 이벤트를 기록하는 중요한 근거 자료이며, 보안 관제, 침해사고 분석, 정책 준수 점검, 사후 포렌식 등의 과정에서 핵심적인 역할을 한다[3,5,8].

그러나 실제 보안 운영 환경에서 네트워크 이상탐지 과정은 다음과 같은 이유로 어렵다. 첫째, 로그의 양이 방대하다. 대규모 기관 또는 기업 환경에서는 하루에도 수십만 건에서 수백만 건 이상의 로그가 생성될 수 있다. 둘째, 생성되는 로그의 구조가 서로 다양하다. 웹 서버, VPN, DNS, 내부 서비스 로그는 서로 다른 포맷과 표현을 사용하며, 같은 유형의 이벤트라도 장비 제조사나 설정 방식에 따라 기록 형식이 달라질 수 있다. 셋째, 로그는 표면적으로는 단순한 텍스트처럼 보이지만, 실제로는 정상 행위와 비정상 행위를 구분하는 문맥 정보가 포함되어 있다. 예를 들어 인증 실패 로그는 단일 건으로 보면 단순 오류일 수 있으나, 반복적으로 발생하거나 외부 원격지에서 짧은 시간 내 다수 발생하는 경우 공격 징후로 해석되어야 한다. 따라서 네트워크를 대상으로 공격 여부를 판단하는 과정은 상당히 어렵다.

기존의 보안 로그 분석은 주로 규칙 기반 또는 시그니처 기반 방식에 의존해 왔다. 이러한 방식은 전문가가 정의한 패턴과 비교하여 이미 알려진 공격을 빠르게 식별하는 데 효과적이다. 그러나 로그 표현이 조금만 달라져도 탐지가 누락될 수 있고, 새롭게 등장하는 공격 유형이나 변형된 공격 패턴을 유연하게 처리하기 어렵다. 또한 규칙의 수가 많아질수록 관리 비용이 증가하며, 서로 다른 장비와 시스템 환경에 규칙을 일관되게 적용하는 것도 쉽지 않다[4,16].

이러한 한계를 극복하기 위한 대안으로 머신러닝과 딥러닝 기반 로그 분석 기법이 연구되고 있다. 특히 최근의 자연어처리(NLP: Netural Language Processing) 기술은 로그를 단순 문자열 집합이 아니라 의미를 가진 텍스트 데이터로 해석하는 접근을 가능하게 하였다. 그중 BERT(Bidirectional Encoder Representations from Transformers)[1]는 양방향 문맥 정보를 학습하는 대표적인 사전학습 언어모델로서, 문장 분류와 의미 유사도 분석에서 우수한 성능을 보여주었다. BERT를 로그 분석에 적용하면, 동일하지 않은 표현이라도 의미적으로 유사한 이벤트를 같은 범주로 인식할 수 있으며, 이는 기존 문자열 매칭 기반 방식보다 높은 유연성을 제공한다[6,7,9].

그럼에도 불구하고 보안 로그 분석에 BERT를 직접 적용하는 데에는 현실적인 제약이 존재한다. 무엇보다 실제 운영 로그에 대한 대규모 라벨링 데이터셋을 확보하기 어렵고, 모든 공격 유형에 대한 지도학습 모델을 구축하는 것은 비용이 크다. 또한 완전한 블랙박스 모델은 탐지 결과의 근거를 설명하기 어렵기 때문에 보안 분석가가 결과를 신뢰하고 활용하는 데 한계가 있을 수 있다. 따라서 의미 기반 분석의 장점을 유지하면서도, 도메인 지식을 함께 반영할 수 있는 해석 가능한 구조가 필요하다[11,12,15].

이에 본 연구에서는 BERT 기반 네트워크 로그 이상행위 탐지 프레임워크(SecuBERT)를 제안하며, (1) 프로파일 문장 기반 의미 유사도 탐지, (2) 키워드 및 포트/프로토콜 보정 규칙 결합, (3) 이상탐지 결과에 대한 시각화 기능을 제공하였다. SecuBERT는 로그 이벤트를 BERT 임베딩 공간에 사상한 뒤, 사전에 정의된 정상 및 비정상 프로파일과의 유사도를 비교하여 판정을 수행한다. 여기에 `failed login`, `payload`, `dns tunnel`, `resource exhaustion` 등과 같은 로그 표현과 SSH, DNS, SMB/WinRM, HTTPS와 같은 포트 및 프로토콜 특성을 반영하는 규칙 기반 보정값을 결합하여 네트워크에서의 이상 유무를 판단하였다. 또한 결과를 CSV와 시각화 그래프로 출력함으로써, 단일 이벤트 수준의 판정과 전체 로그 집합의 경향을 동시에 확인할 수 있도록 하였다. 정상 및 다양한 공격 시나리오를 반영한 총 100개 파일, 304개 이벤트 규모의 샘플 로그셋을 구성하였고, 분석 결과를 CSV 파일과 그래프 형태로 시각화하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 통해 기존 로그 이상탐지 기법과의 차별성을 정리하고, 3장에서는 SecuBERT의 입력 데이터, 프로파일 설계, 유사도 계산 및 규칙 보정 방식을 설명한다. 4장에서는 데이터 구성 및 성능 평가 결과를 제시하고, 5장에서는 제안 기법의 확장성과 타당성 분석 후, 6장에서는 결론과 향후 연구과제를 제시한다.

2. 관련연구

2.1 로그 전처리 및 파싱 기반 이상탐지

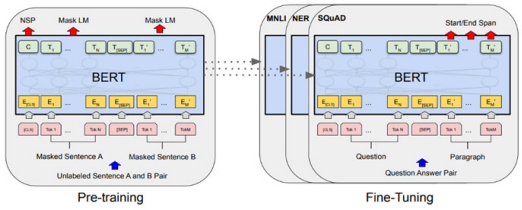
로그 기반 이상탐지는 시스템 운영 안정성, 장애 진단, 침해사고 탐지 측면에서 중요한 연구 주제로 다루어져 왔다. 초기 연구는 시그니처 기반 탐지와 통계 기반 탐지에 집중되었으며, 이후 전통적 머신러닝과 딥러닝 기법으로 확장되었다. 최근에는 Transformer 및 BERT 계열 언어모델을 활용하여 로그의 문맥과 의미를 직접 반영하려는 연구가 활발히 진행되고 있다. 로그 분석의 초기 단계에서는 비정형 로그 메시지를 정형화 가능한 템플릿 형태로 변환하는 과정이 중요하게 다루어졌다. He 등[4]이 제안한 Drain은 고정 깊이 파스 트리를 사용하여 대규모 로그를 온라인 방식으로 빠르게 파싱하였다. Drain은 이후 다양한 로그 이상탐지 연구에서 전처리 도구로 활용되었다. 그러나 파싱 기반 접근은 템플릿 추출 과정에서 로그 원문의 의미가 손실될 수 있고, 새로운 유형의 로그가 나타날 때 파서 또는 템플릿 사전을 갱신해야 하는 부담이 존재한다. 특히 보안 로그는 장비 및 운영 환경에 따라 문장 표현의 차이가 심하므로, 템플릿 기반 접근만으로는 일반화에 한계가 있다[5,8].

2.2 머신러닝 및 시계열 기반 이상탐지

전통적인 로그 이상탐지 연구는 통계적 특성과 순차 패턴을 중심으로 이루어졌다. DeepLog[3]는 LSTM을 이용하여 정상 로그 시퀀스의 다음 이벤트를 예측하고, 예측 오류를 기반으로 이상을 감지하는 구조를 제안하였다. 이후 LogAnomaly[5]는 로그 시퀀스의 순차 정보뿐 아니라 이벤트 발생의 정량적 특성까지 함께 고려하여 DeepLog의 한계를 보완하였다. ConAnomaly[8], LightLog[10], WDLog[14]와 같은 연구 역시 로그 의미와 시계열성을 함께 반영하기 위한 다양한 구조를 제시하였다. 그러나 이들 연구는 원문 의미 반영의 제약 또는 파싱 의존성의 문제를 완전히 해소하지는 못하였다.

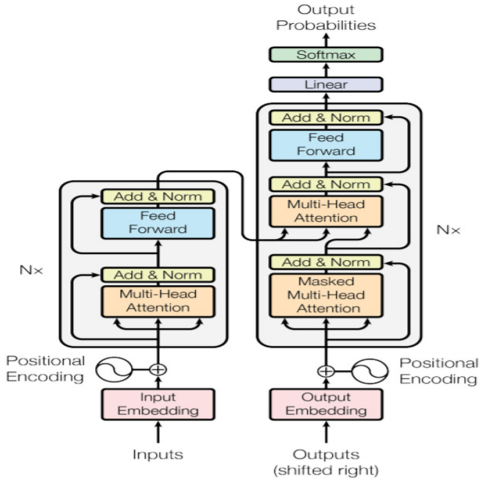
2.3 Transformer 및 BERT 기반 로그 이상탐지

Transformer 구조는 장거리 의존성과 문맥 정보를 효과적으로 반영할 수 있는 자기주의 메커니즘을 도입함으로써 자연어처리 분야에 큰 변화를 가져왔다[2]. BERT[1]는 이러한 Transformer 기반 구조를 바탕으로 양방향 문맥 표현을 학습하며, 문장 분류와 의미 유사도 분석에서 높은 성능을 보였다.



[Fig. 1] BERT Pre-training & Fine-Tuning Architecture

로그 분석 분야에서도 BERT 기반 접근이 빠르게 확산되었다. LogBERT[6]는 자기지도학습 기반 구조를 이용하여 정상 로그 시퀀스를 학습하고, 마스킹 예측 및 시퀀스 수준 판정을 통해 이상을 감지하였다. BERT-Log[7]는 사전학습 언어모델을 활용한 로그 이상탐지의 가능성을 제시하였고, Sentence-BERT 기반 연구[9]는 로그 문장을 보다 정교한 의미 공간에 매핑하는 방법을 제안하였다. 이외에도 LTAnomaly[11], LAnoBERT[12], Logformer[13], LogCSS[15] 등은 Transformer 또는 BERT를 활용하여 로그 의미와 구조적 특성을 함께 고려하려는 다양한 시도를 보여준다.



[Fig. 2] BERT Input/Output Structure

2.4 네트워크 침입탐지와 언어모델 기반 접근

네트워크 보안 분야에서는 전통적으로 패킷, 플로우, 세션 단위의 수치 특징을 기반으로 한 침입탐지 연구가 활발히 수행되어 왔다. 최근에는 BERT 기반 구조를 네트워크 IDS에 적용하는 연구도 발표되고 있다. Yang and Peng[16]은 BERT 기반 네트워크 침입탐지 구조를 제안하여 기존 특징 엔지니어링 중심 접근의 한계를 완

화하고자 하였다. 그러나 이러한 연구의 다수는 구조화된 특징 벡터나 패킷 수준 데이터를 다루며, 실제 운영 환경에서 수집되는 텍스트 로그 자체를 직접 분석하는 연구와는 차이가 있다.

2.5 본 연구의 차별성

기존 연구는 시계열 예측, 문맥 기반 표현 학습, 마스크드 언어모델링 등 다양한 접근을 통해 성능 향상을 이루었으나, (1) 해석 가능성 부족, (2) 대규모 라벨 데이터 의존성, (3) 실제 운영 로그에 대한 적용성 한계를 문제점으로 내포하고 있다. 이에 본 연구에서는 이러한 한계를 보완하기 위해, 네트워크 로그 파일을 직접 입력으로 사용하고, BERT 임베딩과 프로파일 유사도 비교를 기반으로 하며, 힌트 점수와 포트/프로토콜 기반 보정을 결합하고, 결과를 시각화까지 포함하여 제공한다는 점에서 기본 연구[6,7,16]와 차별성을 갖는다. 따라서 기존 연구는 시퀀스 모델링 또는 masked language modeling 중심이지만, 본 연구는 프로파일 문장과 로그 간 의미 유사도 비교를 핵심 판정 메커니즘으로 사용하고, 여기에 키워드 및 포트/프로토콜 규칙을 결합한 하이브리드 접근을 제안한다.

3. 제안하는 SecuBERT 기반 네트워크 로그 이상탐지 및 시각화 시스템

3.1 SecuBERT 프레임워크 구조

본 연구에서 제안하는 `SecuBERT`는 네트워크 로그 파일을 입력으로 받아, 이벤트 수준에서 정상과 비정상을 판별하고, 그 결과를 표와 시각화 그래프 형태로 제공하는 것을 목표로 한다. 로그 파일 내 i번째 이벤트 e_i는 텍스트 메시지인 x_i와 메타데이터 m_i로 구성된 이벤트 $e_i = (x_i, m_i)$를 원소로 갖는 집합 $D = \{e_1, e_2, \cdots, e_N\}$으로 정의할 수 있다. 이때 메타데이터 m_i에는 출발지 주소, 목적지 주소, 포트, 프로토콜, 서비스명, 타임스탬프 등 로그 분석에 필요한 구조적 정보가 포함될 수 있다. 따라서 각 이벤트에 대해 예측(판별)하고자 하는 레이블은 $y_i \in \{0,1\}$로 정의할 수 있으며, $y_i = 0$은 정상(Normal), $y_i = 1$은 비정상(Abnormal)을 의미한다.

따라서 제안하는 모델은 (1) 로그 파일에 대한 입력 및 전처리 과정 이후에, 각각의 이벤트 e_i에 대해서 (2) BERT 기반 프로파일 유사도 측정, (3) 키워드 힌트 기반 점수 측정 및 (4) 포트/프로토콜 기반 보정 과정을 수행한다. 그리고 측정된 결과를 토대로 최종 판정 및 위험도를 산정하고 (5) 최종적으로 결과 저장 및 시각화 과정을 수행한다.

3.2 SecuBERT 시스템 입력 데이터

SecuBERT는 보안 운영 환경에서 흔히 수집되는 로그 파일을 직접 입력으로 사용한다. 이는 정형화된 데이터셋에 한정되지 않고, 실제 운영 환경의 다양한 로그 형태를 최대한 수용하기 위한 설계이다. 시스템은 로그 파일을 읽어 각 이벤트를 추출한 뒤, BERT 기반 의미 표현을 생성하고, 정상 행위 및 비정상 행위 프로파일과의 유사도를 비교하여 판정을 수행한다. 이후 각 이벤트별 결과를 종합하여 파일 단위 및 전체 집합 단위의 분포를 그래프로 시각화한다.

입력 데이터는 보안장비 및 네트워크 서비스 로그에서 추출한 총 100개 파일, 304개 이벤트를 사용하였다. 입력된 로그 이벤트는 토크나이저를 통해 최대 128 토큰으로 정규화되며, BERT 모델에 입력된다. 본 구현에서는 기본적으로 Tramsformer를 토대로 BERT 기반 이상탐지 방식을 사용하되, 보안 도메인에 대한 추가 사전학습 모델이 존재할 경우 이를 선택적으로 사용할 수 있도록 하였다.

3.3 정상 및 비정상 프로파일 분석 및 유사도 측정

정상 행위를 설명하는 i번째 프로파일 문장인 $p_i^{(N)}$으로 구성된 정상 프로파일 집합을 $P_{norm} = \{p_1^{(N)}, p_2^{(N)}, \cdots, p_M^{(N)}\}$로, 그리고 비정상 행위를 설명하는 j번째 프로파일 문장 $p_j^{(A)}$으로 구성된 비정상 프로파일 집합을 $P_{anom} = \{p_1^{(A)}, p_2^{(A)}, \cdots, p_K^{(A)}\}$로 각각 정의할 수 있다. 이때 입력 로그 이벤트의 텍스트 x_i와 각각의 프로파일 문장은 BERT 인코더 $f_\Theta(x_i)$를 통해 임베딩 벡터로 변환된다.

이때 BERT의 출력은 토큰 단위 은닉 벡터들의 집합으로 주어지므로, BERT의 마지막 은닉 상태에 대해 attention mask를 고려한 mean pooling($\overline{M}$) 방식을 수행하여 문장 단위 임베딩 벡터를 생성한다.

$$\Psi_e = \overline{M}(BERT(e_i)) \quad (1)$$

여기서 `$e_i$`는 입력 로그 이벤트, `$\Psi_e$`는 정규화된 의미 벡터이다. SecuBERT는 로그 이벤트를 미리 정의된 정상 및

비정상 프로파일과 비교한다. 정상 프로파일은 정상 웹 접근, 정상 DNS 조회, 정상 사용자 인증, 정상 내부 서비스 트래픽, 정상 파일 전송을 설명하는 문장 집합으로 구성된다. 비정상 프로파일은 포트 스캔, 브루트포스 로그인, 명령제어 통신, 데이터 유출, 악성 파일 다운로드, 내부 확산, 서비스 거부 공격, DNS 이상 행위를 설명하는 문장 집합으로 구성된다. 이는 라벨링 데이터셋이 부족한 상황에서도 BERT의 의미 유사도 기능을 활용하기 위함이다.

정상 프로파일과 비정상 프로파일 각각에 대해 설명문을 BERT 임베딩 공간에 사상한 뒤, 입력 로그 이벤트와의 유사도를 계산한다.

$$S_{norm}(e_i) = \max(\cos(\Psi_e, \Psi_p)),\ p \in P_{norm} \quad (2)$$

$$S_{anom}(e_i) = \max(\cos(\Psi_e, \Psi_q)),\ q \in P_{anom} \quad (3)$$

따라서 두 유사도의 차이를 이용하면 의미 기반 이상 마진을 정의할 수 있다.

$$M(e_i) = S_{anom}(e_i) - S_{norm}(e_i) \quad (4)$$

이때 $M(e_i)$가 양수이며 크면 클수록 해당 이벤트 e_i는 정상보다 비정상 프로파일과 더 높은 의미적 유사성을 갖는다고 해석할 수 있으며, 반대로 $M(e_i)$가 0보다 작거나 작으면 작을수록 정상 프로파일에 더 가깝다고 판단할 수 있다.

3.4 키워드 힌트 및 포트/프로토콜 기반 보정

BERT 유사도는 로그 문장의 전반적 의미를 효과적으로 반영하지만, 'failed login', 'payload', 'resource exhaustion'과 같이 보안 운영상 즉각적인 판단이 필요한 키워드는 별도 가중치를 통해 보강할 필요가 있다.

먼저 비정상 키워드 집합을 K_{anom}, 정상 키워드 집합을 K_{norm}이라고 할 때 키워드 힌트 점수는 다음과 같이 정의할 수 있다.

$$H(e_i) = \sum_{r \in K_{anorm}} \omega_r^{(anom)} I(r \subset x_i) - \sum_{s \in K_{norm}} \omega_s^{(norm)} I(s \subset x_i) \quad (5)$$

이때, $\omega_r^{(anom)}$과 $\omega_s^{(norm)}$은 각각 비정상/정상 키워드의 중요도에 해당하는 가중치 값을 의미한다. 즉, 비정상 키워드가 많이 포함될수록 점수는 증가하고, 정상 행위를 암시하는 키워드가 포함될수록 점수는 감소한다.

SSH 기반 로그인 실패, DNS 기반 긴 질의 또는 알고리즘적 도메인, SMB/WinRM 기반 원격 명령 실행, 승인된 HTTPS 접근 등은 네트워크 차원에서 중요한 단서이다. 본 연구는 이러한 맥락을 반영하기 위해 포트/프로토콜 기반 보정값을 도입하였다. 포트·프로토콜 기반 보정은 SSH, DNS, SMB/WinRM, HTTPS 등 서비스 맥락을 고려하여 동일 키워드라도 서비스 종류에 따라 위험도를 차등 반영하도록 설계하였다. 포트·프로토콜 기반 보정 점수는 메타데이터 m_i가 특정 보안 규칙을 만족하는지 여부에 따라 계산한다. 포트/프로토콜 집합을 R이라 할 때, 보정 점수는 다음과 같다.

$$P(e_i) = \sum_{q \in R} \omega_q I(\phi_q(m_i) \equiv 1) \quad (6)$$

$\phi_q(m_i)$는 이벤트의 메타데이터가 q번째 규칙을 만족할 경우 1을 반환하는 함수이며, ω_q는 해당 규칙의 가중치이다. 예를 들어 짧은 시간 동안 다수 목적지 포트에 대한 반복 접근, 비정상적인 DNS 쿼리 길이, 관리 포트에 대한 외부 접근, 내부망 간 측면 이동 징후 등은 양의 보정값이 부여된다.

3.5 최종 판정, 위험도 산정 및 시각화

이제 의미 유사도 점수, 키워드 힌트 점수, 포트·프로토콜 보정 점수를 통합하여 이벤트에 대한 최종 이상 점수를 산출할 수 있다.

$$F(e_i) = \alpha M(e_i) + \beta H(e_i) + \gamma P(e_i) \quad (7)$$

여기서 α, β, γ는 각각 의미 유사도, 키워드 힌트, 포트·프로토콜 정보의 상대적 기여도를 조절하는 결합 가중치이다. 따라서 식 (7)은 단순한 문장 분류 모델이 아니라, 의미 기반 표현과 도메인 규칙을 함께 결합하는 하이브리드 기반 이상행위 탐지 구조임을 의미한다.

각각의 이벤트에 대한 이상 여부는 식 (7)로부터 계산된 이상 점수 $F(e_i)$와 임계값 τ를 비교하여 다음과 같이 결정한다.

$$\hat{y} = \begin{cases} 1, \text{if } F(e_i) > \tau \\ 0, otherwise \end{cases} \quad (8)$$

$\hat{y}$는 이벤트 e_i에 대한 정상/비정상 예측 레이블이며, τ는 이상행위 판정을 위한 임계값으로 0.015 값을 사용하였다. 실제 운영 환경에서 이벤트는 단순히 정상/비정상으로 구분하는 것보다는 위험 수준을 연속적인 값으로 제시하는 것이 유용할 수 있다. 따라서 다음과 같이 정규화된 위험 점수 $R(e_i)$를 측정하고 이를 토대로 사전에 정의된 두 개의 임계값 δ_1, δ_2를 기준으로 `low`, `medium`, `high` 수준의 위험도를 산정하도록 하였다.

$$Risk(e_i) = \begin{cases} Low, & R(e_i) < \delta_1 \\ Medium, & \delta_1 \le R(e_i) < \delta_2 \\ High, & R(e_i) \ge \delta_2 \end{cases} \quad (9)$$

즉, 제안 모델은 단순한 형태의 이상탐지 모델이 아니라, 네트워크 로그 이벤트를 위험 수준 기반으로 우선순위화할 수 있는 분석 도구로도 활용될 수 있다.

최종 결과는 이벤트 단위 CSV 파일로 저장되며, 위험도 분포도, 예측 카테고리 분포도, 정상·비정상 이벤트 비율 그래프, 성능 평가 지표 그래프 등 다양한 시각화 결과로 출력하도록 설계하였다. 이는 단순히 분류 성능 수치를 제시하는 수준을 넘어, 보안 분석가가 탐지 결과를 직관적으로 해석하고 후속 분석에 활용할 수 있도록 지원한다는 점에서 의미가 있다.

따라서 제안 기법은 대규모 라벨 데이터가 부족한 환경에서도 적용 가능한 프로파일 기반 의미 유사도 이상탐지 모델로 볼 수 있으며, 동시에 규칙 기반 탐지의 실용성과 딥러닝 기반 의미 표현의 일반화 가능성을 함께 결합한 하이브리드 접근이라 할 수 있다. 다음 장에서는 이러한 제안 기법의 성능을 검증하기 위한 실험 환경과 평가 결과를 제시한다.

4. 실험 결과 및 성능 분석

4.1 실험 환경 및 데이터 구성

본 장에서는 제안한 SecuBERT의 성능을 검증하기 위해 데이터 구성, 비교 실험, 구성요소 제거 실험, 오분류 분석, 시각화 결과를 단계적으로 제시한다. 특히 본 연구에서는 정상 로그와 비정상 로그를 함께 포함하는 네트워크 이벤트 데이터를 대상으로, 제안 모델이 실제 보안 관제 환경에서 얼마나 안정적으로 위협을 탐지할 수 있는지를 중점적으로 평가하였다. 또한 분석 대상 파일 수가 증가함에 따라 탐지 성능이 어떻게 변화하는지를 추가적으로 분석함으로써, 제안 기법의 확장성과 실용성을 함께 검토하였다.

실험은 Python 기반 환경에서 수행하였으며, 사전학습 언어모델을 활용하여 로그 문장의 의미적 표현을 추출하고 이를 정상 행위 프로파일 및 이상 행위 프로파일과 비교하는 방식으로 분류를 수행하였다. 입력 데이터는 텍스트 파일, 로그 파일, CSV 파일 형태로 구성되었으며, 각 파일에는 시간 정보, 출발지 및 목적지 주소, 프로토콜, 포트 정보, 그리고 이벤트 메시지가 포함되어 있다. 모델은 각 로그 이벤트를 독립적인 분석 단위로 처리하며, 이벤트 단위의 판정 결과를 통해 파일 단위의 전체 위험 성향도 함께 확인할 수 있도록 설계하였다.

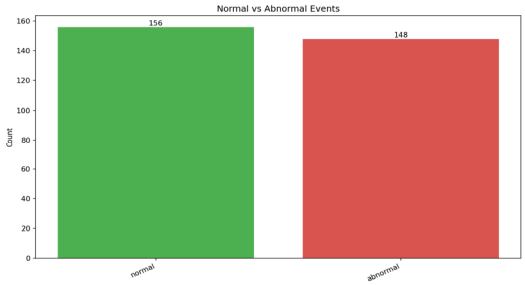
제안 모델은 정상 행위와 비정상 행위에 대한 의미 기반 프로파일을 사전에 정의하고, 입력 로그가 각 프로파일과 어느 정도 유사한지를 계산하여 최종 판정 결과를 도출한다. 여기에 인증 실패, 포트 스캔, 악성 페이로드 다운로드, 데이터 유출, DNS 이상행위와 같은 보안 관점의 핵심 패턴을 반영함으로써 단순 유사도 비교의 한계를 보완하였다. 이러한 구조는 단순 키워드 탐지보다 문맥을 더 잘 반영하며, 동시에 보안 분석 관점의 해석 가능성을 제공한다.

4.2 실험 결과

실험에서는 표 1과 같이 총 100개의 로그 파일을 대상으로 분석을 수행하였으며, 상세 분석 과정에서 총 304개의 이벤트가 처리되었고, 분석 가능한 이벤트는 총 285개였으며, 그림 3과 같이 정상과 비정상 이벤트 분포가 구성되며 만일 정상/비정상이 혼합되어 있어 명확한 판정이 어려운 이벤트 19개는 최종 평가에서 제외하였다.

[Table 1] List of Collected Network Information

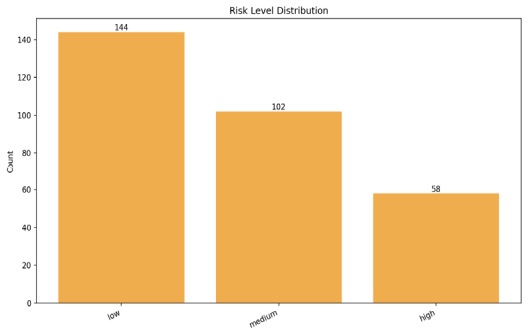
Category	Number of File	Number of Event
Text File	39	117
Log File	41	123
CSV File	20	64
Total	100	304



[Fig. 3] Normal vs Abnormal Events

그림 4의 위험도 분포 분석 결과, 전체 304개 이벤트 중 low위험도는 144건(47.37%), medium은 102건(33.55%), high는 58건(19.08%)으로 나타났다. low위험도 이벤트는 모두 정상 로그에 해당하였으며, high 위험도 이벤트는 모두 비정상 로그로 분류되어 제안 모델이 정상과 고위험 이상행위를 비교적 명확하게 구분하고 있음을 확인하였다. 또한 medium 위험도 구간에는 정상과 비정상 이벤트가 함께 포함되어 경계적 사례를 반영하는 특성을 보였다. 이러한 결과는 제안 모델이 단순

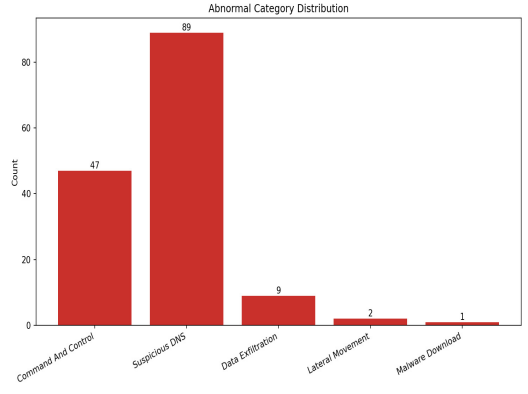
한 이진 분류를 넘어, 이벤트의 위협 수준에 따라 우선순위를 부여할 수 있음을 보여주며, 실제 보안 관제 환경에서 효율적인 대응 판단에 활용될 수 있음을 알 수 있었다.



[Fig. 4] Risk Level Distribution

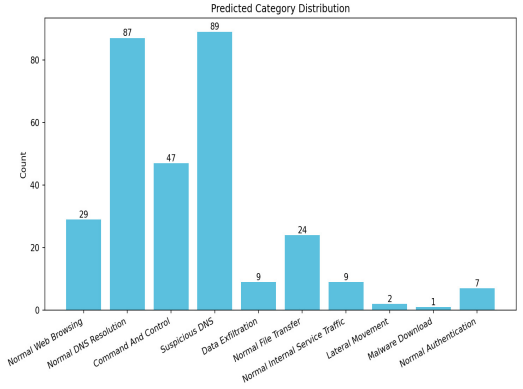
4.3 이상탐지 결과

일반적인 이상탐지 시스템이 단순히 정상과 비정상만을 구분하는 경우, 운영자는 모든 이상 이벤트를 동일한 수준으로 처리해야 하는 부담이 있다. 그러나 그림 4에 제시된 low/medium/high의 3단계 위험도 체계 뿐만 아니라 그림 5와 같이 비정상 이벤트에 대한 분류 결과를 제시하여, 보안 담당자는 이벤트를 우선순위에 따라 선별적으로 대응할 수 있도록 하였다.



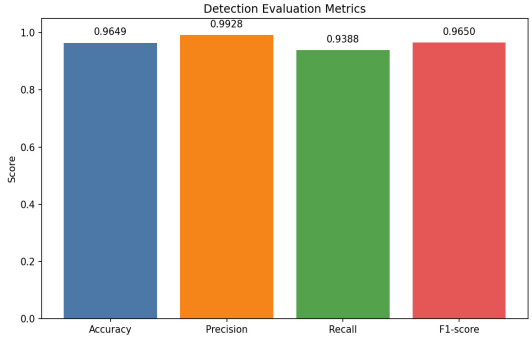
[Fig. 5] Abnormal Category Distribution

그림 6의 예측 카테고리 분포를 분석한 결과, 전체 304개 이벤트 중 Suspicious DNS가 89건(29.28%)으로 가장 높은 비중을 차지하였고, Normal DNS Resolution이 87건(28.62%)으로 뒤를 이었다. 또한 Command And Control은 47건(15.46%)으로 나타나, DNS 기반 이상행위와 명령제어 관련 이벤트가 주요 탐지 대상으로 분류되었음을 확인할 수 있었다. 정상 범주에서는 Normal Web Browsing 29건(9.54%), Normal File Transfer 24건(7.89%)이 비교적 높은 비중을 보였다. 이러한 결과는 제안 모델이 정상 및 비정상 로그를 의미적으로 구분하면서도, 특히 DNS 이상행위와 명령제어형 위협에 민감하게 반응하고 있음을 보여준다. 반면 Lateral Movement와 Malware Download와 같은 카테고리는 낮은 빈도를 보여, 해당 유형의 데이터 수가 상대적으로 적거나 다른 이상 범주에 비해 보수적으로 분류되었을 가능성을 시사한다. 전체적으로 예측 카테고리 분포는 제안 모델이 주요 보안 위협 유형을 효과적으로 식별하고 있음을 나타낸다.



[Fig. 6] Predicted Category Distribution

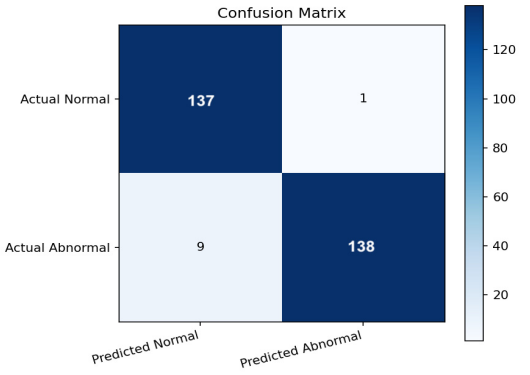
성능 평가는 정확도, 정밀도, 재현율, F1-score과 같은 네 가지 대표 지표를 사용하였고 각각의 지표에 대한 측정 결과는 다음 그림 7과 같다.



[Fig. 7] Detection Evaluation Metrics

본 연구에서는 이러한 네 가지 지표를 통해 제안 모델이 단순히 많이 탐지하는 수준을 넘어서, 오탐과 미탐 사이에서 어떤 성향을 보이는지 함께 분석하였다. 전체 실험 결과, 제안 모델은 정확도 0.9649, 정밀도 0.9928, 재현율 0.9388, F1-score 0.9650의 성능을 나타냈다. 이는 제안 모델이 전체적으로 높은 수준의 탐지 성능을 확보하고 있음을 보여준다. 특히 정밀도가 0.9928로 매우 높게 나타난 점은 모델이 비정상으로 판정한 이벤트 대부분이 실제 위협 이벤트였음을 의미한다.

그림 8과 같이 혼동행렬 관점에서 보면, 정상 이벤트를 정상으로 올바르게 분류한 경우는 137건, 비정상 이벤트를 비정상으로 올바르게 탐지한 경우는 138건이었다. 반면 정상 이벤트를 비정상으로 잘못 판정한 경우는 1건에 불과하였고, 실제 비정상 이벤트를 정상으로 판단한 경우는 9건이었다. 이는 특정 비정상 로그가 정상 서비스 패턴과 의미적으로 유사하거나 명시적 공격 단서가 부족했기 때문으로 판단된다.



[Fig. 8] Confusion Matrix

정확도와 F1-score가 모두 0.96 이상으로 나타난 것은 제안 모델이 전체적으로 균형 잡힌 성능을 보여준다는 점을 의미한다. 특히 보안 분야에서는 단순 정확도보다 재현율과 정밀도의 균형이 중요하므로, 본 연구의 결과는 실제 침해 탐지 시스템으로의 적용 가능성을 뒷받침하는 근거가 된다. 정밀도가 높은 모델은 관제 인력에게 과도한 알람 부담을 주지 않는다는 장점이 있고, 재현율이 일정 수준 이상 확보되면 실제 위협 상황에 대한 탐지 누락도 줄일 수 있기 때문이다.

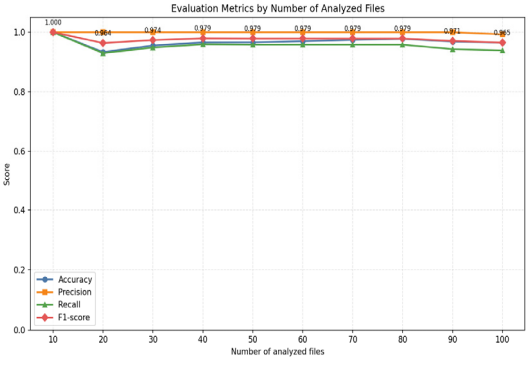
5. 제안 기법의 확장성 및 타당성 분석

5.1 분석 대상 변화에 따른 성능 측정

본 연구에서는 제안 모델의 안정성과 확장성을 평가하기 위해, 그림 9와 같이 분석 대상 파일 수를 10개 단위로 점진적으로 증가시키면서 성능 지표가 어떻게 변화하는지도 함께 관찰하였다. 이는 실제 운영 환경에서 데이터가 지속적으로 축적될 때, 모델 성능이 일관되게 유지되는지를 확인하기 위한 분석이다.

파일 수 증가 실험은 학습량 증가 효과가 아니라 분석 대상 로그 수 증가에 따른 성능 안정성을 확인하기 위한 것으로, 각 구간은 동일한 정상/비정상 비율을 유지하도록 구성하였다. 파일 수가 20개로 증가하면서 정확도는 0.9333, F1-score는 0.9636으로 감소하였고, 이후 30개와 40개 구간에서는 다시 0.95 이상 수준으로 회복되었다.

40개 파일 이후부터는 지표가 전반적으로 안정화되는 경향(정확도 0.9658, F1-score 0.9794)을 보였고, 50개 파일에서도 유사한 수준이 유지되었다. 60개, 70개, 80개 파일 구간에서는 정확도가 각각 0.9697, 0.9744, 0.9778로 나타나면서 오히려 소폭 향상되는 모습도 관찰되었다. 이는 데이터 수가 증가함에 따라 다양한 정상 및 비정상 패턴이 누적되어, 모델의 판정 결과가 특정 사례에 과도하게 영향을 받지 않고 점차 안정화된 것으로 볼 수 있으며, 최종 100개 파일 구간에서는 정확도 0.9649, 정밀도 0.9928, 재현율 0.9388, F1-score 0.9650으로 수렴하였다.



[Fig. 9] Evaluation Matrics by Number of Analyzed Files

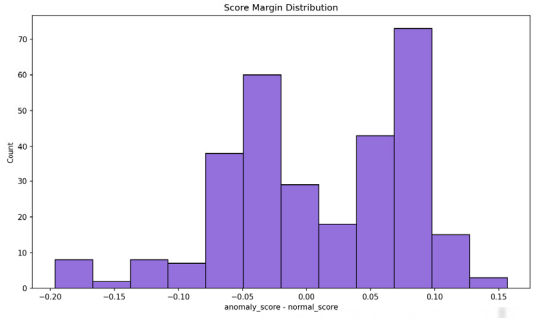
이러한 변화 양상은 제안 모델이 전체적으로 높은 안정성을 유지하고 있음을 의미한다. 초기에는 작은 표본 특성에 따라 지표가 다소 크게 변할 수 있지만, 데이터가 충분히 누적된 이후에는 성능이 일정 범위 내에서 안정화되는 모습을 보였다. 특히 F1-score가 대부분의 구간에서 0.96 이상을 유지했다는 점은, 모델이 다양한 데이터 규모에서도 비교적 일관된 탐지 품질을 제공함을 의미한다.

5.2 이상행위 자동탐지 결과 분석

SecuBERT는 대규모 라벨 데이터가 부족한 환경에서도 BERT 기반 의미 유사도 분석을 활용하여 네트워크 로그의 정상/비정상 여부를 판별할 수 있음을 보여준다. 특히 본 실험 결과는 제안한 보안 로그 탐지 모델이 의미 기반 문장 표현과 보안 규칙성 보강을 결합함으로써 높은 탐지 성능을 달성할 수 있음을 보여준다. 특히 정밀도가 매우 높게 유지된 것은 실제 운영 환경에서 오탐 경보를 줄이는 데 유리한 장점이 있다.

반면 재현율은 정밀도보다 다소 낮게 나타났으며, 일부 비정상 이벤트를 정상으로 분류하는 경향이 확인되었다. 이는 공격 유형이 정상 행위와 유사한 표현을 포함하거나, 이벤트 메시지가 짧고 문맥 정보가 부족한 경우에 발생할 가능성이 있다.

그림 10과 같이 이상탐지 점수 차이(anomaly_score - normal_score)는 최소 −0.1964에서 최대 0.1567까지 분포하였으며, 평균은 0.0105, 중앙값은 0.0103으로 나타났다. 전체 304개 이벤트 중 음수 구간은 144건, 양수 구간은 160건으로 확인되어, 정상 이벤트와 이상 이벤트가 비교적 균형 있게 분포하고 있음을 알 수 있다. 또한 임계값 0.015를 초과한 이벤트는 148건, 고위험 기준인 0.08이상인 이벤트는 58건으로 나타났다. 이는



[Fig. 10] Score Magin Distribution

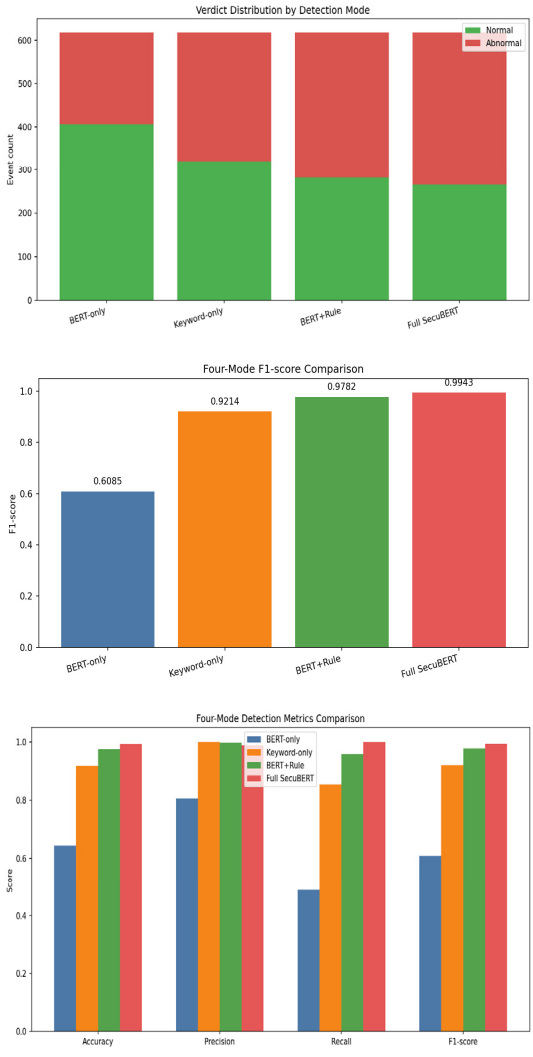
제안 모델이 단순히 정상과 이상을 구분하는 데 그치지 않고, 이상 점수와 정상 점수의 차이를 통해 위협 강도를 단계적으로 반영하고 있으며 해당 이벤트를 더 명확한 이상행위로 판단하고 있음을 알 수 있다.

5.3 제안 기법의 타당성 검증

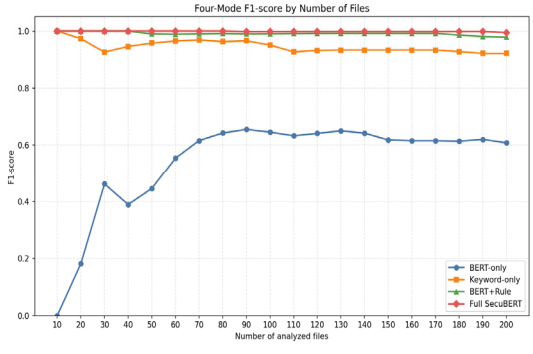
본 연구에서 제안한 SecuBERT의 핵심은 BERT 기반 의미 유사도, 키워드 힌트, 포트·프로토콜 기반 보정 점수를 결합한 최종 이상탐지 함수 $F(e_i)$이다. 따라서 $F(e_i) = \alpha M(e_i) + \beta H(e_i) + \gamma P(e_i)$를 구성하는 각 요소에 대해 분석해 보면, $M(e_i)$는 BERT 기반 프로파일 유사도 마진, $H(e_i)$는 키워드 기반 힌트 점수, $P(e_i)$는 포트/프로토콜 기반 보정 함수를 의미하므로, 이를 각각 (1) BERT-only 방식, (2) Keyword_only 방식, (3) BERT+Rule 방식 그리고 (4) Full SecuBERT 방식의 네 가지 형태로 단계적으로 분리할 수 있다. 각각의 네 가지 모델/단계별로 기존 100개의 로그 파일을 총 200개의 로그 파일로 확대하여 본 연구에서 제시한 기법의 타당성 및 우수성을 입증하고자 하였다.

그림 11과 같이 각각 (1) BERT-only 방식, (2) Keyword_only 방식, (3) BERT+Rule 방식 그리고 (4) 제안한 SecuBERT 방식별로 검출된 (A) 정상/비정상 이벤트 개수 분포 비교, (B) 각각의 방식별로 각각 Accuracy, Precision, Recall 및 F1-score 성능 지표에 대한 비교 결과를 나타낸다. 분석 결과 제안한 SecuBERT 방식이 (1)~(3) 방식보다 가장 우수한 성능을 보이는 것을 알 수 있다.

그림 12와 같이 네 가지 모델에 대해 분석 대상 파일을 10개 단위로 증가하여 최대 200개 파일을 대상으로 분석하였을 경우 F1-score 값의 변화를 측정해 보았다. 실험 결과 아래 그림과 같이 제안한 SecuBERT 방식이 대상 파일을 변화해 나가더라도 (1) BERT-only 방식 보다는 월등한 성능을 나타내며, (2) Keyword_only 및 (3) BERT+Rule 기반 방식보다 가장 우수한 성능을 보이는 것을 알 수 있다. 이를 통해서 본 연구에서 제안한 SecuBERT 기법이 확장성과 함께 이론적/실제적 측면에서 타당성을 제공한다는 것을 확인할 수 있었다.



[Fig. 11] Classification & Performace Comparison



[Fig. 12] Four Model F1-score Comparison

6. 결론

본 연구에서는 프로파일 기반 의미 유사도 분석과 규칙 보정을 결합한 BERT 기반 네트워크 로그 이상행위 탐지 및 시각화 프레임워크 SecuBERT를 제안하였다. SecuBERT는 로그 이벤트를 BERT 임베딩 공간에 사상한 후, 정상 및 비정상 프로파일과의 의미 유사도를 계산하고, 여기에 키워드 힌트와 포트/프로토콜 기반 보정 규칙을 결합하여 최종 판정을 수행한다. 또한 분석 결과를 CSV와 그래프로 시각화함으로써 보안 분석가의 해석 가능성과 실무 활용성을 높였다.

따라서 본 연구에서 제시한 내용은 보안 로그 분석 측면에서 BERT 계열 모델이 단순 분류기 이상의 역할을 할 수 있음을 보여준다. 즉, SecuBERT는 의미 기반 임베딩 도구, 준분류기, 설명 가능한 분석 프레임워크의 역할을 동시에 수행한다. 이는 실제 운영 환경에서 모델 신뢰성과 해석 가능성이 중요하다는 점을 고려할 때 의미 있는 접근이다.

향후 연구로는 보안 도메인 특화 추가 사전학습을 적용한 BERT 계열 모델을 활용함으로써 로그 의미 표현의 품질을 더욱 향상시킬 수 있다. 그리고 SIEM 또는 스트리밍 로그 처리 환경과 연동하여 대단위 실시간 이상탐지 체계에 효율적으로 확장 가능할 것으로 기대된다.

REFERENCES

[1] J.Devlin, M.W.Chang, K.Lee, and K.Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2019, pp. 4171-4186.

[2] A.Vaswani, N.Shazeer, N.Parmar, J.Uszkoreit, L.Jones, A.N.Gomez, L.Kaiser, and I.Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017, pp. 5998-6008.

[3] M.Du, F.Li, G.Zheng, and V.Srikumar, “DeepLog: Anomaly detection and diagnosis from system logs through deep learning,” in Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS), Dallas, TX, USA, 2017, pp. 1285-1298.

[4] P.He, J.Zhu, Z.Zheng, and M.R.Lyu, “Drain: An online log parsing approach with fixed depth tree,” in Proc. IEEE Int. Conf. Web Services (ICWS), Honolulu, HI, USA, 2017, pp. 33-40.

[5] W.Meng, Y.Liu, Y.Zhu, S.Zhang, D.Pei, Y.Liu, Y.Chen, R.Zhang, S.Tao, P.Sun, and R.Zhou, "LogAnomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs," in Proc. 28th Int. Joint Conf. Artificial Intelligence (IJCAI), Macao, China, 2019, pp. 4739-4745.

[6] H.Guo, S.Yuan, and X.Wu, "LogBERT: Log anomaly detection via BERT," arXiv preprint arXiv:2103.04475, 2021, pp. 1-11.

[7] S.Chen and H.Liao, "BERT-Log: Anomaly detection for system logs based on pre-trained language model," Applied Artificial Intelligence, Vol. 36, No. 1, pp. 1-19, 2022.

[8] D.Lv, N.Luktarhan, and Y.Chen, "ConAnomaly: Content-based anomaly detection for system logs," Sensors, Vol. 21, No. 18, pp. 1-22, 2021.

[9] C.Hu, X.Sun, H.Dai, H.Zhang, and H.Liu, "Research on log anomaly detection based on Sentence-BERT," Electronics, Vol. 12, No. 17, pp. 1-19, 2023.

[10] V.H.Le, H.Zhang, et al., "LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge," Computer Networks, Vol. 203, pp. 1-14, 2022.

[11] D.Han, M.Sun, M.Li, and Q.Chen, "LTAnomaly: A transformer variant for syslog anomaly detection based on multi-scale representation and long sequence capture," Applied Sciences, Vol. 13, No. 13, pp. 1-19, 2023.

[12] Y.Lee, J.Kim, and P.Kang, "LAnoBERT: System log anomaly detection based on BERT masked language model," Applied Soft Computing, Vol. 145, pp. 1-16, 2023.

[13] F.Hang, W.Guo, H.Chen, L.Xie, C.Zhou, and Y.Liu, "Logformer: Cascaded transformer for system log anomaly detection,"CMES - Computer Modeling in Engineering & Sciences, Vol. 136, No. 1, pp. 517-529, 2023.

[14] W.Niu, X.Liao, S.Huang, Y.Li, X.Zhang, and B.Li, "A robust wide & deep learning framework for log-based anomaly detection,"Applied Soft Computing, Vol. 153, pp. 1-17, 2024.

[15] Z.Li, X.Tu, H.Gao, S.Huang, and Z.Ma, "LogCSS: Log anomaly detection based on BERT-CNN with context-semantics-statistics features," Journal of Information Security and Applications, 2024, pp. 1-17.

[16] Y.Yang and X.Peng, "BERT-based network for intrusion detection system," EURASIP Journal on Information Security, Vol. 2025, No. 11, pp. 1-18, 2025.

이 형 우(Hyung-Woo Lee) [종신회원]

- 1994년 2월 : 고려대학교 컴퓨터학과 (학사)
- 1996년 2월 : 고려대학교 컴퓨터학과 (석사)
- 1999년 2월 : 고려대학교 컴퓨터학과 (박사)
- 1999년 3월 ~ 2003년 2월 : 백석대학교 정보통신학부 교수
- 2003년 3월 ~ 현재 : 한신대학교 AISW대학 교수

〈관심분야〉
사물인터넷, 정보보호, 모바일 보안 및 디지털 포렌식, 지능형 사이버공격 대응