『2025 제3회 KISIA 정보보호 개발자 해커톤』

개발기획서

팀명	DS(디에스)
프로젝트명	DeepScan
프로젝트 소개	AI SSL DLP:LLM 프롬프트의 의미를 실시간 분석하여 민감정보 유출을 사전 차단하고, 시계열 · 다자간 조합을 분석해 사후 감시까지 가능한 AI SSL DLP 시스템
팀 소개 및 팀원별 역할	

1 추진 배경 및 필요성

1.1 기존 SSL DLP의 구조적 한계

기존 SSL DLP는 SSL 트래픽을 MITM 방식으로 복호화한 후, 사전에 정의된 정규표현식, 키워드 리스트 등을 통하여 탐지하는 구조이다. 그러나 다음과 같은 한계를 지닌다:

(1)**의미 기반 탐지 불가능**:

키워드 매칭 또는 단순한 패턴 기반 분석에 그쳐, “o1ㅇ-“, "하나three4" 등 **의도적으로 변형된 데이터**는 탐지하지 못함.

(2)**오탐 및 미탐** 발생:

문맥을 고려하지 않는 기계적 탐지 방식은 정상 행위를 오탐하거나, 중요한 기밀 유출을 놓치는 미탐을 초래함.

(3)**시계열 및 다중 사용자 분산 전송 탐지 불가**:

일부 내부자들은 개인정보를 **시점별, 사용자별로 나누어 전송하는 방식**으로 탐지를 회피하는데, 기존 시스템은 이러한 행위 기반 분석 기능이 부재함.

1.2 AI 기반 SSL DLP의 필요성

기존 SSL DLP 한계를 극복하지 못할 경우, **국가 기밀의 은닉·분산 유출**을 효과적으로 차단하지 못해 국가 안보가 위협받고, 사회적 혼란과 경제적 피해가 가중된다. 단순 기술 문제를 넘어 **국가 체계적 리스크와 직결**되며, AI 기반 고도화된 SSL DLP 도입은 국가 안정과 지속 가능한 정보 보호를 위한 필수 전략이다.

1.2.1 국가 기밀 정보 유출에 따른 안보 위협 심화

(1) 국가 중요 인프라 및 전략 정보 유출 위험 증대

- SSL 트래픽 복호화 후에도 **단순 키워드 탐지**에 의존하는 기존 DLP는 **변형·우회 정보 유출을 85% 이상을 식별하지 못하는** 실정
- 2022년 미 국가안보국(NSA) 보고서에 따르면, 내부자 및 고도화된 공격으로 인한 기밀 정보 유출 사건이 전체 보안 사고의 35%를 차지하며, 이 중 **60% 이상이 암호화 우회 기법을 활용**함

- 미국 국토안보부에 따르면, 2023년 한 해 동안 에너지 인프라에 대한 사이버 공격 중 42%가 내부자 및 은닉된 트래픽을 통해 이뤄짐

(2)국가 주요 인프라의 데이터 유출은 국가 안보에 직결되는 치명적 위협

- 한국 정부는 2024년 '국가정보보호백서'를 통해 내부자 위협 및 **암호화 트래픽 우회 공격 증가**를 심각한 안보 위협으로 경고

1.2.2 사회 혼란 및 경제적 피해 가중

(1)국가 기밀 유출에 따른 불확실성 및 사회 불안 가중

- 국가 정책, 군사 정보, 주요 인사 개인정보 유출시, 여론 조작, 불신 조장, 사회 혼란 발생
- 2017년 '워터링홀' 공격 사례에서 기밀 유출 후 3개월 내 민심 급변동 및 주요 산업 분야 불안정 발생 보고됨
- 경제적 손실 및 복구 비용 급증
- 2 023년 Ponemon Institute2에 따르면, 개인정보 유출 사고 시 평균 경제적 피해액은 3,920만 달러에 달하며, **국가 기밀 유출 사고는 이보다 최대 4배 이상 높은 피해** 추정
- 국가 기밀 유출 시 **시스템 복구, 보안 강화, 법적 대응 비용 증가**로 사회·경제 전반에 막대한 부담 초래

2 주요 기능 및 개발환경

2.1 주요 기능

기존 한계를 극복하고 고도화된 내부 위협 탐지를 수행하기 위해 AI 기반 SSL DLP 솔루션의 필요성이 제기된다. AI SSL DLP는 다음과 같은 혁신적 기능을 기반으로 설계된다:

(1)**의미 기반 탐지 (Semantic-aware NLP)**

:자연어 처리(NLP) 기반 문맥 분석을 통해 "유재석 → 유군", "O1ㅇ → 010" 등 우회·은닉 표현도 실질적으로 식별 가능

(2)**시계열 탐지 및 메모리 기반 분석**

: 트래픽의 시간적 흐름을 추적하고 세션 단위로 저장하여, "010", "2345", "6789"가 1일 간격으로 분산 전송된 개인정보임을 통합적으로 식별 가능.

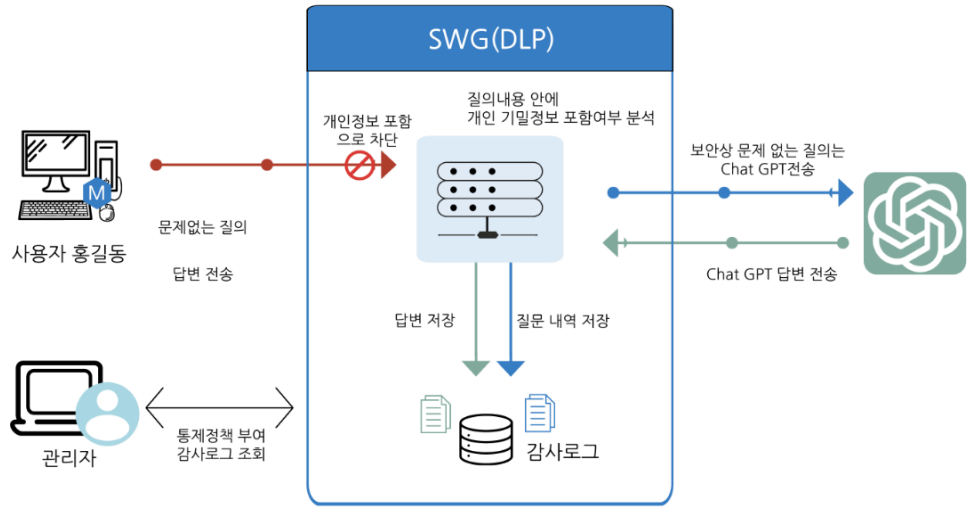
(3)**다중 사용자 조합 행위 분석**

:여러 사용자가 역할 분담하여 "010", "1234", "5678"을 나누어 전송하는 경우, 접속 로그 및 프롬프트 상관 분석을 통해 집합적으로 탐지 가능.

(4)**생성형AI 서비스 프롬프트 탐지 및 차단**

:외부 AI 서비스에 전달되는 프롬프트 의미 분석을 선행하여 유출 가능성 있는 요청을 사전 차단하고, **1주일 내 사용자별 프롬프트 조합이 개인정보 유출에 해당**하는 경우 관리자에게 경고를 발송함.

내용 기반 (주민 번호 민감정보 포함시 차단)및 경보



2.2 개발환경 및 기술 스택

2.2.1 시스템 아키텍처 및 인프라

(1)마이크로서비스 기반 클라우드 네이티브 아키텍처

- Docker 컨테이너, Kubernetes 클러스터 운영으로 유연한 확장성과 고가용성 확보
- 서비스별 독립 배포, 무중단 업데이트 및 자동 복구 메커니즘 탑재

(2)데이터 파이프라인

- Apache Kafka를 통한 실시간 고속 데이터 스트리밍 및 처리
- 데이터 인게스천(ingestion), 정제, 변환, AI 모델 서빙을 위한 파이프라인 분리 및 최적화

(3)데이터 저장소

- **시계열 DB**: TimescaleDB (PostgreSQL 기반) 활용, 대규모 트래픽 이벤트 시간대별 저장 및 분석
- **그래프 DB**: Neo4j를 이용한 사용자-행위-데이터 관계성 저장 및 쿼리 최적화
- **로그 분석**: Elasticsearch + Kibana 조합으로 실시간 로그 집계 및 시각화 제공

2.2.2 AI 모델링 및 서빙 플랫폼

(1)자연어 처리

- Hugging Face Transformers 라이브러리 기반의 대형 사전학습 모델(PLM) 활용
- 도메인 특화 파인튜닝 수행 (한국어 개인정보 도메인 말뭉치 구축 및 활용)
- 커스텀 토크나이저 및 사전 처리 모듈로 비표준 문자 처리 강화

(2)시계열 이상 탐지

- LSTM, Transformer 기반 시퀀스 이상 탐지 모델 구현
- Autoencoder 및 GAN 기반 비지도 학습으로 정상-비정상 행위 분리

(3)그래프 기반 이상 탐지
- PyTorch Geometric, DGL(Deep Graph Library) 활용하여 GNN 모델 구축
- 다중 사용자 협업 및 연계 유출 시나리오 자동 분류 및 경고
(4)모델 서빙
- TensorFlow Serving, TorchServe를 통한 RESTful API 실시간 모델 추론 서비스 제공
- Kubernetes 내 모델 자동 스케일링 및 A/B 테스트 지원

2.2.3. SSL 트래픽 처리 및 복호화
(1)MITM Proxy 설계 및 구현
- 오픈소스 기반 mitmproxy 또는 Envoy 필터 확장으로 SSL/TLS 세션 복호화
- 고성능 패킷 캡처 및 페이로드 추출, 인라인 전처리 모듈 탑재
(2)멀티프로토콜 지원
- HTTPS, SMTPS, FTPS 등 주요 TLS 기반 프로토콜 완벽 지원
- 비대칭 키 교환 프로토콜 분석 및 세션 식별 최적화
(3)패킷 재조립 및 세션화
- TCP 세그먼트 재조립 및 페이로드 정합성 검사
- 세션 단위 데이터 누적 및 실시간 분석 적용

2.2.4. 보안성 및 운영체계
(1)데이터 보호
- 전송 및 저장 데이터 AES-256 암호화 적용
- 키 관리 시스템(KMS)과 HSM(Hardware Security Module) 연동
(2)접근 통제
- RBAC 기반 권한 분리 및 MFA 적용으로 관리자 및 사용자 권한 엄격 관리
- 컴플라이언스 준수
- 개인정보보호법, GDPR 등 법률 규정에 부합하는 데이터 처리 정책 반영
- 감사 로그 보관 및 변경 불가성 확보

2.2.5. 모니터링 및 경고 체계
(1)실시간 대시보드
- Grafana/Kibana 기반 사용자 맞춤형 시각화
- 이상행위, 탐지 결과, 사용자 세션 모니터링
(2)통합 경고 시스템
- 이메일, SMS, SIEM 연동 자동 경고 발송
- 탐지 레벨별 자동 분류 및 우선순위 지정
(3)리포팅 및 감사

- 일별/주별/월별 상세 탐지 보고서 자동 생성
- 관리자용 감사 추적 로그 및 포렌식 지원 자료 제공

2.3 개발 언어 및 주요 API

2.3.1 주요 프로그래밍 언어

- **Python**: AI 모델 개발, 데이터 처리, API 서버 구축
- **Go / Rust**: 네트워크 트래픽 캡처 및 처리 모듈, 고성능 실시간 스트리밍
- **TypeScript (React)**: 대시보드 및 웹 UI 개발

2.3.2 주요 라이브러리 및 프레임워크

- **NLP**: Hugging Face Transformers, KoNLPy, SpaCy
- **시계열 분석**: TensorFlow, PyTorch, Prophet
- **그래프 신경망**: PyTorch Geometric, DGL
- **네트워크**: Scapy, mitmproxy API
- **데이터베이스**: TimescaleDB, Neo4j, Elasticsearch REST API

2.4. 개발 및 운영 인프라

- **클라우드 인프라**: AWS, GCP, Azure 중 선택, IaC(Infrastructure as Code) 기반 자동 프로비저닝
- **CI/CD 파이프라인**: Jenkins, GitLab CI, ArgoCD를 통한 코드 품질 검증 및 무중단 배포
- **컨테이너 오케스트레이션**: Kubernetes로 클러스터 관리, 자동 스케일링 및 헬스 체크
- **로그 및 메트릭 수집**: Prometheus, Fluentd, ELK 스택 활용 실시간 성능 및 보안 모니터링

2.5. 기술적 차별성 및 기대효과

- 기존 키워드·패턴 기반 SSL DLP 대비 탐지 정확도 및 민감도 대폭 향상
- 분산 및 은닉 개인정보 유출 행위를 통합·연계 분석하여 실시간 차단 가능
- 다중 사용자 협업 공격, 생성형 AI 프롬프트 유출 등 최신 위협에도 선제 대응
- 고성능, 확장성 높은 아키텍처로 대규모 엔터프라이즈·정부기관 환경에 최적화

3 활용방안(공공성) 및 기대효과

3.1 공공기관 · 국가기관 내 실질적 적용 가능성

기존 키워드 기반 시스템은 의도적 변형, 시점 분산, 다중 사용자 유출 협업 시나리오에 무력하지만, 해당 솔루션은 문맥·행위 기반 복합 분석을 통해 실시간으로 탐지 및 차단 가능하다. AI 기반 SSL DLP는 다음과 같은 기관 및 부처에 실질적인 다음과 같은 사회적·행정적 효용을 창출한다

- **행정안전부·국가정보원**: 내부 문서 전송, 메신저 기반 AI 사용 이력 등에 대한 시맨틱 분석 및 시계열 기반 개인정보 유출 탐지
- **법무부·검찰청·경찰청**: 수사 정보 또는 민감한 개인정보의 외부 전송을 실시간으로 모니터링 및 차단
- **국방부·과학기술정보통신부**: 국가 전략 기술, 무기체계 관련 데이터의 암호화 트래픽 기반 외부 유출 탐지

3.2 기존 솔루션의 한계 및 보완사항

항목	기존 시스템	AI SSL DLP
의미 기반 탐지	정규표현식 기반 단순 키워드 매칭	Transformer 기반 문맥 의미 인식 탐지
분산 유출 인식	단일 세션 단위 분석 한정	시계열 메모리 + 사용자 간 연계 분석
은닉 기법 대응력	특수문자, 비표준 표현에 취약	커스텀 토크나이저 및 GAN 기반 비정형 탐지
확장성/유연성	중앙 집중식 모놀리식 구조	MSA + Kubernetes 기반 클라우드 네이키브 설계

3.3 사회 · 국가적 기대효과

(1)국가 기밀 유출 방지

고위험 행위의 조기 탐지 및 차단으로 인하여 사이버 침해 후 복구 비용 감소, 사회적 신뢰도 및 정보 주권 강화

(2)사고 대응 비용 절감

Ponemon Institute(2023)에 따르면, 사전 탐지 기반 보안 시스템은 평균 사고당 대응 비용을 31% 이상 절감하며, 대응 시간은 평균 51% 단축됨

(3)보안 거버넌스 고도화

내부자 유출, 암호화 트래픽 우회 등 기존 규제 대상 외 행위에 대한 실질적 통제 수단 확보, 국가 보안 규정 및 감사 체계 강화

(4)공공 신뢰 회복 및 민간 확산 기반 확보

공공기관 우선 도입 후, 금융기관·의료기관·대형 통신사업자 등으로 확산 가능. 규제 기반 보안 수준 향상과 민간 도입 인센티브로 이어질 수 있음.