

제안요청서

□ AI 기반 실시간 사이버 위협 탐지 및 대응 자동화 기술 개발 및 실증

실증 과제	AI 기반 실시간 사이버 위협 탐지 및 대응 자동화 기술 개발 및 실증
실증 개요	○ (개요) ①내·외부망이 명확히 분리되지 않은 일반 기업의 업무 환경 및 ②개인 PC·스마트폰 환경과 같이 제3자와 소통하고 외부 서버와 상시 연결되는 네트워크 환경을 대상으로 AI 기반 실시간 사이버 위협 탐지 및 대응 자동화 기술 실증 추진 - 특히, 제3자와 상시 통신하여 정보 주체의 사전 동의를 확보하기 어려운 환경에서 발생하는 개인정보 수집·분석 문제에 대해 보안 목적에 한정한 합리적 적용 기준 마련을 위한 실증을 추진 ○ (목표) AI 기반 실시간 사이버 위협 탐지 기술의 안전성·신뢰성을 종합적으로 검증하여, 향후 AI 보안 기술의 제도적 기준 마련을 위한 데이터를 제시하고 합리적 운영 기준을 수립하는 것을 목표로 진행 - 온라인으로 진행되는 대화·커뮤니케이션, 자료·링크전송 과정을 실시간으로 사이버 보안 AI가 모니터링 및 학습하여 이상 징후 발생 시 사전 차단 및 사후 추적 ○ (필요성) 예기치 못한 시스템 오류나 해킹으로 인한 개인정보 유출은 막대한 피해를 야기하나 기존 룰베이스 기반 보안 시스템은 관리자 권한을 악용할 때나 해커를 분별하기 어려워 사전 보안 조치에 한계 존재 - 최근 발생한 온라인 쇼핑몰의 개인정보 대량 유출 사례와 같이 대량의 개인정보가 비정상적으로 조회되거나 노출되는 상황을 AI가 실시간으로 학습하고 탐지하여, 사고 발생 즉시 자동으로 차단하는 기술 도입의 필요성이 대두 - 현행 「개인정보보호법」 및 「정보통신망법」에는 네트워크 트래픽·로그 데이터를 수집·저장·분석하는 과정에서 '동의' 및 '목적 제한' 조항이 존재하여 AI 학습·탐지용 데이터 활용 제약 - 가명정보 처리 시 특정 행위자(예: 내부자 계정, 침해 계정)를 식별하기 어렵기 때문에 계정 탈취나 내부자 위협 탐지 등 보안 목적 달성에 지장 - 제3자와 소통하는 네트워크 환경에서 수집되는 여러 정보주체의 개인정보 수집·이용의 포괄적 동의 확보가 불가능 ※ 기존 백신·엔드 포인트 등 보안 솔루션은 데이터를 수집 및 학습하지 않고 파일 형식 및 확장자 등을 기준으로 필터링만 하기 때문에 제3자 개인정보 수집·처리 문제가 발생하지 않음

실증 대상

- **(실증대상)** ①보안 정책 및 전담 인력이 충분히 확보되지 않은 일반 기업(B2B)와 ②개인 이용자 PC·스마트폰 단말(B2C)을 대상으로 외부 네트워크와 상시 연결되는 환경을 대상으로 AI기반 보안 솔루션 실증 추진
- (중소·중견기업 전용 보안) 보안 전문가가 부족해 해킹에 취약한 기업을 대상으로 회사 서버에 오가는 모든 데이터를 보안 AI 실시간 모니터링 하는 실증 진행(내부 근로자는 개인정보 수집·이용 동의 후 실증)
- (개인용 보안 서비스 환경) 일반 개인 이용자가 사용하는 PC 및 모바일 (Android, iOS)에 설치된 보안 애플리케이션을 통해 실시간으로 유입되는 위협 콘텐츠 및 앱 활동 로그를 AI가 분석(사용자 개인정보 수집·이용 동의 후 실증)
- 실제 산업 현장에서 가장 일반적으로 나타나는 구조로, 현행 법령상 제3자 개인정보 수집·분석에 대한 포괄 동의 확보가 불가능

< 데이터 수집 및 처리 유형 예시 >

분야	주요 데이터	구체적 예시
일반 산업	- 내부 시스템 ↔ 외부 파트너사·공급망과의 통신 트래픽 - 사용자 웹클라우드 서비스 이용에 따른 외부 도메인/외부 SaaS 접속 트래픽 로그 - 외부에서 다운로드·업로드가 발생하는 파일 전송·교환 트래픽 로그	협력사가 계약서·발주서 등을 이메일이나 메신저로 전송하는 경우, 해당 파일은 전달 과정에서 기업 내부 네트워크를 경유하게 되어 협력사 직원 등 제3자의 동의 여부와 관계없이 실시간 모니터링 중인 사이버 위협 탐지 AI가 개인정보를 수집·학습하게 됨
금융 분야	- 금융거래 트래픽 로그 - 인증 트래픽 로그(외부 단말 기반 인증 요청 흐름) - 외부 결제사/외부 시스템 API 통신 트래픽 (카드사·PG사·금융기관 연동)	고객이 은행 상담직원에게 카카오톡·기업 메신저·이메일 등을 통해 통장 사본이나 신분증 이미지 등을 제출하는 경우, 해당 파일은 고객의 별도 동의 여부와 관계없이 실시간 모니터링 중인 사이버 위협 탐지 AI가 개인정보를 수집·학습하게 됨
개인 PC 및 모바일	- 개인 서비스 (B2C) 모바일 활동 로그 - 개인 PC 단말 활동 로그	개인 이용자가 모바일 메신저, SNS를 통해 수신한 피싱 의심 링크(URL) 및 첨부파일 실행 시 발생하는 프로세스 트래픽, 브라우저 및 앱 내 광고 팝업, 스팸 문자 등을 통해 유입되는 악성 스크립트 및 개인정보 탈취 목적의 위장 웹페이지 접속 로그 등 개인정보를 실시간 모니터링 중인 사이버 위협 탐지 AI가 수집·학습하게 됨

실증 내용

< 데이터 수집·분석 기반 기술 및 기대효과 >

분야	주요 데이터	적용기술	기대 효과
일반 산업	내부 시스템 ↔ 외부 파트너 간 통신 트래픽 로그 (메일, 파일 전송 등)	AI 기반 네트워크 행위 분석 및 파일 전송 이상 탐지 기술	외부 협력사와의 데이터 교환 과정에서 발생할 수 있는 정보 유출 및 비인가 파일 전송 행위 조기 탐지
	SaaS 및 클라우드 서비스 접속 트래픽 로그	AI 기반 애플리케이션 식별 및 접근 행위 분석 기술	업무용 SaaS 사용 현황을 기반으로 비인가 서비스 접근 및 데이터 유출 가능성 식별
	외부 다운로드 /업로드 파일 전송 트래픽 로그	AI 기반 파일 전송 패턴 분석 및 행위기반 이상 탐지 기술	대용량 파일 전송 및 비정상 다운로드 행위에 대한 실시간 탐지 및 대응 가능
금융 분야	금융 거래 및 인증 트래픽 로그	AI 기반 인증 패턴 분석 및 이상 행위 탐지 기술	비정상 로그인 및 금융 정보 탈취 시도 등 이상 인증 행위 조기 탐지
	외부 결제/API 통신 트래픽 (카드사, PG 등)	API 호출 행위 분석 및 이상 트랜잭션 탐지 기술	외부 연계 구간에서 발생하는 위·변조 요청 및 비정상 API 호출 탐지
개인 PC 및 모바일	개인 모바일 및 PC 기반 서비스 이용 트래픽 로그	AI 기반 사용자 OS/브라우저 식별 기술 URL/접속 패턴 분석 기술	피싱 사이트 접속, 악성 URL 클릭 등 사용자 기반 보안 위협 사전 차단
	메신저, 이메일, SNS 기반 통신 트래픽 로그	AI 기반 메신저, 메일 SNS 트래픽 식별 기술 콘텐츠 전송 패턴 분석 기술	악성 URL 및 첨부파일을 통한 개인 정보 탈취 및 계정 탈취 시도 탐지
	내부 자산의 외부 통신 트래픽 (VPN, 스캔 등)	AI 기반 프로토콜 분석 및 비정상 통신 행위 탐지 기술	비인가 VPN, 스캔 행위 등 내부 자산 기반 위협 행위 가시성 확보 및 대응 강화

○ 네트워크 데이터 수집 및 학습 데이터 활용

- (데이터 다양성 확보) 네트워크 전 구간에서 발생하는 트래픽 로그(SSL/TLS 포함)를 수집하며, 외부 제3자와 접속하는 트래픽 (외부 단말 접속, 외부 사용자 요청, 외부 API 호출 등)을 포함한 메타데이터셋을 구축·활용
- (학습데이터 검증) 내부·외부 혼합 트래픽 환경에서 수집된 데이터셋을 AI 탐지 모델 학습용으로 구성하고, 제3자 기반 트래픽이 혼재된 상황에서의 메타데이터 품질·표준화 수준을 검증
- (탐지 성능 향상) 실제 트래픽 기반 데이터셋 적용 시 AI 모델의 탐지율·재현율이 제3자에서 발생하는 이상 행위에 대해서도 향상되는지 평가

○ AI 기반 네트워크 위협 탐지

- (탐지 정확도 평가) 실시간 트래픽 환경에서 제3자 기반 이상 행위를 탐지하는 정확도를 검증하며, 또한 트래픽 내 전송/유입되는 파일에 대해 AI 모델 기반 악성 파일 탐지 제공 및 Anti-Virus 엔진 기반 악성 파일 탐지 및 탐지명 제공이 가능한지 함께 확인
- (암호화 트래픽(SSL/TLS 등) 가시성 확보) 외부 웹사이트와 통신하는 제3자 암호화 트래픽(SSL/TLS)에서도 애플리케이션-위협 식별이 가능한지 확인
- (실시간성 검증) 위협 이벤트 발생 시 탐지→경고→대응까지의 지연 시간을 측정하며, 특히 위협 트래픽에서 악성 파일 탐지 및 탐지명 제공 결과가 실시간 경고/대응 흐름에 지연 없이 반영되는지 측정

○ 사이버 위협 인텔리전스(CTI) 연계 및 자동 대응 기술

- (CTI 데이터 연계) 자사 TI를 기반으로 신뢰도 검증이 가능한 외부 위협 인텔리전스(VirusTotal 등 TI 제공업체) 피드와 탐지 이벤트를 실시간 연동하여, 외부 공격·외부 IP/도메인·제3자 위협과의 상관 분석을 수행
- (자동화 검증) 탐지된 내부/외부 위협 이벤트를 기반으로 자동 대응 정책이 실행되는지 검증하고, 특히 반복적 외부 위협(스캔·브루트포스·C2 통신 등)에 대한 자동화 효율성을 측정
- (대응 속도 평가) CTI 교차 정보를 활용할 때 외부-내부 혼합 공격의 탐지→대응 주기 단축률을 검증

○ LLM 기반 설명·분석 지원 및 의사결정 신뢰성 검증

- (분석 자동화) LLM 기반 요약·설명 기능을 활용하여 탐지 이벤트의 원인·영향 범위를 자동 해석하고, 제3자 기반 외부 이벤트(외부 접속·외부 트래픽 흐름 등)에 대한 설명 가능성을 검증
- (신뢰성 검증) 보안 담당자가 LLM이 제공한 분석 결과를 활용할 때 외부 제3자 기반 이벤트 관련 의사결정의 정확도·신뢰성이 향상되는지 평가
- (효율성 향상) 평균 분석시간(MTTR) 단축 효과 및 반복 분석 자동화 비율을 측정해 LLM 기반 설명 기능의 운영적 효율성을 검증

○ AI 오남용 방지 및 사생활 보호 장치

- 본 실증은 근로자 및 개인 감시 목적이 아닌, 사이버 위협 탐지 목적에 한정하여 데이터 처리

	- 개인 식별 가능 정보는 최소 수집가능한 범위까지 가명 처리 원칙 적용 - AI 분석 결과는 인사·평가·징계 목적에 활용하지 않도록 내부 통제 규정 마련 - 데이터 접근 권한을 최소화하고, 로그 감사 체계 운영 - 데이터 처리 범위 및 목적을 내부 구성원에게 사전 고지하여 투명성 확보
기대효과	○ (활용 기준 확보) AI 기반 실시간 위협 탐지 및 자동 대응 기술의 수집-분석-대응 전 주기 실증을 수행하고, 특히 제3자·외부 서비스와 접하는 네트워크 트래픽 환경에서의 안전성·신뢰성을 검증함으로써 향후 AI 보안 기술 인증·평가체계에 적용 가능한 기술적 활용 기준을 마련함 ○ (정책 목표 달성) 보안 목적의 개인정보 활용이 제3자 기반 트래픽 및 외부 연계 데이터 환경에서도 실효적으로 가능함을 실증하여, AI 보안 기술 인증체계 및 개인정보보호법 적용에 대한 합리적 제도 개선 방향을 제시 - 이를 통해 국가 차원의 지능형 보안 기술 활성화 및 정책 목표 달성에 기여 ○ (이용자 편익) AI 자동화 기반 탐지·대응 기술의 적용으로 보안 담당자의 업무 효율성과 대응 속도를 향상시키고, 외부 공격·제3자 기반 위협을 포함한 전 범위 위협 대응 능력을 강화하여 기업·기관의 사이버 위협 피해를 최소화하는 데 기여
특례지원 내용	○ AI 기반 실시간 위협 탐지 및 자동화 대응 기술의 실증을 위해, 보안 목적에 한정한 개인정보 수집·분석 시 사전 포괄 동의 확보가 구조적으로 곤란한 환경에서, 보안 목적에 한정한 최소 범위 내 처리 기준을 마련할 수 있도록 특례 적용, 가명정보 처리 목적에 '보안 목적(위협 탐지 및 대응)'을 추가할 수 있도록 특례 적용 ○ 정보통신망법상 정보보호 관리체계(ISMS·ISMS-P) 인증 의무대상자는 특례 사업 수행 과정에서도 개인정보보호 관련 인증체계 준수 필요 ○ 과학기술정보통신부 「AI 윤리 가이드라인」의 설명가능성·투명성 원칙을 존중하되, 보안 AI의 특성을 고려하여 결과 중심 설명 기준을 적용하고, 핵심 탐지 알고리즘의 보안성을 유지할 수 있는 범위 내에서 합리적 조정 방안을 검토

<관련 규제>

- ① 실시간 위협 탐지를 위해 네트워크 트래픽 및 사용자 로그를 수집해야 하나, 현행 법령상 모든 정보주체의 사전 동의가 필요하여 실시간 수집·분석 적용 제한(개인정보보호법 제15조)
 - 보안 목적의 실시간 위협 탐지를 위한 네트워크 트래픽 분석 및 사용자 행위 분석에는 현행 법령상 적용 한계가 존재
 - 특히, 제3자와 소통하는 네트워크 환경에서 수집되는 여러 정보주체의 개인정보 수집·이용의 포괄적 동의 확보가 불가능
- ② 가명정보 활용 목적이 통계작성·과학적 연구·공익적 기록보존으로 한정되어 있어, 보안 분석 및 위협 탐지 등 보안 목적의 데이터 활용이 불가능한 상황임 (개인정보보호법 제28조의2)
- ③ 현행 ISMS-ISMS-P 제도에는 AI 기술 활용을 위한 개인정보 수집 및 이를 활용한 보안시스템에 대한 내용 부재 (정보통신망법 제47조)
- ④ AI 시스템의 설명가능성과 투명성 확보를 요구하나, 보안 AI의 경우 탐지 로직이 노출될 시 공격자가 이를 악용할 우려가 있어 보안 시스템의 특수성이 반영되지 못하는 한계가 있음 (AI 윤리 가이드라인, 과학기술정보통신부 2020)