연구개발계획요구서(RFP)

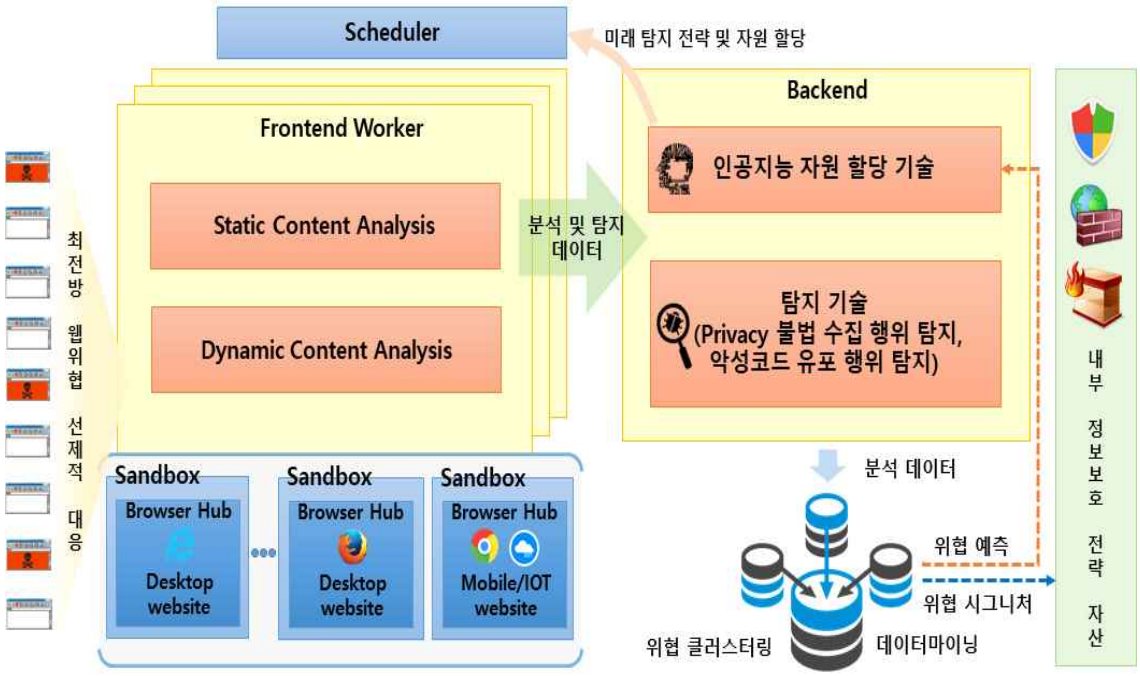
과제명: 인공지능을 이용한 웹 위협 탐지 및 예측 자동화 시스템

1. 개요

가. 기술의 개념 및 정의

o 기존 웹 위협 탐지시스템의 한계점을 개선하고, 증가하는 웹 위협에 대응하기 위하여 인공지능 기반 불법 악성코드 유포 및 개인정보 강제수집 행위를 탐지, 분석 및 예측하는 기술 연구

o 폭발적으로 증가하는 웹사이트로 부터 지속적으로 고도화되는 웹 위협에 대응하기 위해 타켓 사이트를 고속으로 탐지하는 동적크롤러 기술과 머신러닝 기반의 웹 빅데이터 분석 및 웹 위협 탐지 및 예측 기술을 개발



[인공지능 기반 자동화된 웹 위협 탐지 및 예측 시스템 개념도]

나. 기술의 중요성/필요성 및 시급성

o 기술의 중요성/필요성

- 웹을 통해 악성코드를 유포하는 방식은 간단하면서 강력한 공격 기법으로 불특정 다수 또는 대상을 표적화하여 공격할 수 있으나, 기존의 시스템 으로는 효과적으로 이러한 웹 위협을 탐지 및 차단하는 데는 한계가 있기 때문에 효과적 대응을 위한 새로운 개념의 기술개발이 중요하며 시급히 요청되고 있음.
- 현재의 탐지방법으로 국내외 수많은 타켓 웹사이트를 전수 조사하는 것은 현실적으로 불가능하기 때문에, 효과적으로 탐색 및 제한된 자원을 효율적으로 수행할 수 있는 기술개발이 필요함.

o 기술개발의 시급성

- 최근 웹을 통한 사이버 공격은 가장 위협적이며 지속적으로 증가 및 고도화되고 있는 추세로, 효과적으로 대처 가능한 관련기술이 시급히 필요함.
- 최근 구글, 페이스북, 네이버 등 개인정보를 활용한 온라인 광고 플랫폼의 성공으로 인해, 타 온라인 서비스들도 광범위하고 다양한 방법으로 Privacy 정보를 수집 및 활용하게 되었고, 더 많은 광고 수익을 위해 공격적이고 불법적으로 Privacy를 수집하는 문제가 사회적 문제가 되고 있기 때문에 시급한 대처가 필요함.

다. 연구개발 최종목표

o 응용연구

항목		목표 성능	비고
•동적 크롤러	• 대상 브라우저	• 3개 이상	• IE, Chrome, Edge 등
	• 동시 실행 브라우저 수	• 20개 이상	• i7급 PC 1개 활용시
	• 10만개 도메인 분석 주기	• 1일 1.5회	• i7급 PC 1개 활용시 • 1depth 기준
• 악성코드		• 정탐율 75% 이상	• 멀웨어 등 최근 3년간 유포된

유포 행위 탐지		악성코드 유포행위 데이터 기준
• 기계학습 기반 악성 코드 유포 행위 예측	• 정탐율 70% 이상	• 멀웨어 등 최근 3년간 악성코드 유 유포행위 분석 데이터 기준
• 자바스크립트 난독화 여부 자동 판단	• 정탐율 90% 이상	
• 탐색 우선순위 자율 결정	• 150초 이상 단축	• 100개 웹사이트중 20개 사이트에 악성코드 유포기준
• 개인정보 강제 수집 행위 탐지	• 정탐율 75% 이상	• 최근 3년간 개인정보 강제 수집행위 기준
• 바이너리 난독화 여부 자동 판단	• 정탐율 90% 이상	• 1,000종의 unique한 말웨어 대상 • MS.Virusign 등 잘 알려진 데이터 기준
• 바이너리 난독화 자동화 해제	• 난독화 해제율 75% 이상	

o 시험개발

항목		목표 성능	비고
•동적 크롤러	• 대상 브라우저	• 5개 이상	• IE, Chrome, Firefox, Opera 등
	• 동시 실행 브라우저 수	• 30개 이상	• i7급 PC 1개 활용시
	• 10만개 도메인 분석 주기	• 1일 2회	• i7급 PC 1개 활용시 • 1depth 기준
• 악성코드 유포 행위 탐지		• 정탐율 90% 이상	• 멀웨어 등 최근 3년간 유포된 악성코드 유포행위 데이터 기준
• 기계학습 기반 악성 코드 유포 행위 예측		• 정탐율 87% 이상	• 멀웨어 등 최근 3년간 악성코드 유 유포행위 분석 데이터 기준
• 탐색 우선순위 자율 결정		• 300초 이상 단축	• 100개 웹사이트중 20개 사이트에 악성코드 유포기준
• 개인정보 강제 수집 행위 탐지		• 정탐율 90% 이상	• 최근 3년간 개인정보 강제 수집행위 기준
• 바이너리 난독화 자동화 해제		• 난독화 해제율 85% 이상	• 1,000종의 unique한 말웨어 대상 • MS.Virusign 등 잘 알려진 데이터 기준

2. 국내외 기술현황 및 전망

가. 국내 기술동향 및 전망

o 개인정보 침해 탐지 및 방지 연구(웹 컨텐츠 수집, 개인정보 수집 및 활용 등)

o 리얼 브라우저 기반 웹 크롤러를 이용한 개선된 악성 웹사이트 탐지 기법 (2016 정보보호학회)

o 빅데이터 분석 기반의 정보 검색을 위한 웹 크롤러 (2017 한국디지털콘텐츠학회)

o 웹 크롤링 기반 개인정보 유출 방지 시스템 설계 (2016 한국인터넷정보학회)

나. 국외 기술동향 및 전망

o A cloud-based web crawler architecture (2015 ICIN)

o Collecting internet malware based on client-side honeypot (2008 ICYCS)

o Efficient Method for Analyzing Malicious Websites by Using Multi- Environment Analysis System (2017 AsiaJCIS)

3. 연구개발 계획

가. 단계별 연구개발 목표

o 민·군수용

구분	연구개발 목표	연구개발 내용	주요결과물
응용연구	1.개요 다.항의 연구개발 최종 목표 참조 (응용연구)	• 동적크롤러 설계 • 악성코드 유포행위 탐지방안 연구 • 악성코드 유포 예측 알고리즘 연구 • 자바스크립트 난독화 여부 자동판단 연구 • 개인정보 강제 수집행위 탐지방안 연구 • 바이너리 난독화 자동화 해제방안 연구	5. 연구개발 결과 제시물 및 평가 항목 가.항의 연구개발 결과 최종 제시물 참조
시험개발	1.개요 다.항의 연구개발 최종 목표 참조 (시험개발)	• 동적크롤러 성능 보완 • 악성코드 유포행위 탐지방안 성능보완 • 악성코드 유포 예측 알고리즘 성능보완 • 개인정보 강제 수집행위 탐지성능 보완 • 바이너리 난독화 자동화 해제성능 보완	5. 연구개발 결과 제시물 및 평가 항목 가.항의 연구개발 결과 최종 제시물 참조

① 연구개발 목표를 달성하기 위해 수행하는 연구개발 내용 및 결과물은 추가 제안 가능

② 최종목표의 달성 여부는 공인시험기관의 시험성적서를 평가에 반영하여 판단

③ 위 표는 시험개발단계에서 요구되는 연구내용의 예시이며, 과제 신청시 본 문서의 [1-다.연구개발 최종 목표] 항목을 참고하여 최종 목표 달성을 위한 연차별 목표를 연구개발계획서에 제시

나. 사업기간 및 연구개발비

o 사업기간 : 시험개발 3.5년(응용 2년, 시험 1.5년)

o 총 연구개발비(정부출연금) : 30.0억원 이내 (응용 18억원, 시험 12억원)

4. 적용 및 파급효과

가. 적용분야

o 민수분야

- 정부 공공기관 및 민간기관 주요 웹사이트 보호 및 모니터링에 적용
- 급증하는 불법적 개인정보 수집 방지시스템에 적용

o 군수분야

- 국방부, 합참 및 전술제대 주요 웹사이트 보호 및 모니터링에 적용
- 급증하는 불법적 군 개인정보 수집 방지시스템에 적용

나. 파급효과

o 기술적 측면

- 동적 웹 분석 및 위협 탐지기술 분야의 혁신적인 원천기술 확보
- 웹 기반 공격적 Privacy 정보 수집행위 탐지 및 대응기술과 체계적인 모니터링 기술 획득
- 페이스북 및 구글 등 글로벌 온라인 플랫폼에서 수행하는 불법 개인정보 수집 및 Privacy 침해에 대응 가능한 기술 확보
- 인공지능개념을 적용한 탐색전략 예측 및 자원할당 등을 자동으로 수행하는 최적화된 차세대 웹 크롤러의 원천 기술 확보

o 경제·산업적 측면

- 인공지능 개념을 적용한 웹 위협 탐지 및 위협 예측기술은 최근 민 및 군수 분야에서 발생하는 불법적 국가 및 산업 기밀유출을 차단하여 경제적, 산업적으로 막대한 국가이익 창출이 기대됨.
- 세계적으로도 체계적인 개인정보 불법 수집 및 모니터링 기반기술이 아직 미미한 상태에서, 머신러닝 기반의 웹 빅데이터 분석기술과 자동화된 웹 위협 탐지 및 예측 기술을 개발함으로써 국내는 물론 해외 진출을 통한 경제적 및 산업 발전 효과가 클 것으로 예상됨.

o 군사적 측면

- 군 특성상 가장 중요하게 다루는 보안관점에서, 인공지능 개념을 적용한 웹 위협 탐지 및 위협 예측기술은 보안강화 관점에서 매우 중요하게 활용될 것으로 판단됨.
- 본 과제에서 획득된 기술은 웹사이트에 직접 방문하여 행위정보를 수집 및 분석하기 때문에 사이버 첩보전에서 최적의 방어 기술로 활용할 수 있으며, 수집 및 분석 데이터를 이용 및 재가공을 통한 최선의 공격 기술로 활용이 가능할 것으로 판단됨.

5. 연구개발 결과 제시물 및 평가항목

가. 연구개발 결과 최종 제시물

o 인공지능을 이용한 웹 위협 탐지 및 예측 자동화 시스템 시제 1식

※ 시제 세부내용은 CDR시 확정

o 기술자료 1식

- 동적크롤러 설계보고서
- 악성코드 유포행위 탐지방안 설계보고서
- 악성코드 유포 예측 알고리즘 설계보고서
- 기계학습 기반 탐색 우선순위 결정 알고리즘 설계보고서

- 개인정보 강제 수집행위 탐지방안 설계보고서
- 바이너리 난독화 자동 해제방안 설계보고서
- 기타 H/W, S/W 설계보고서

※ 구체적 기술자료 산출물은 제안서에 추가 기술

o 공인시험기관 수행 최종목표 시험성적서 1부

나. 연구개발 결과 평가항목

o 응용연구

항목		평가 내용	비고
•동적 크롤러	• 대상 브라우저	• 3개 이상	• 시험
	• 동시 실행 브라우저 수	• 20개 이상	• 시험
	• 10만개 도메인 분석 주기	• 1일 1.5회	• 시험
• 악성코드 유포 행위 탐지		• 정탐율 75% 이상	• 시험
• 기계학습 기반 악성 코드 유포 행위 예측		• 정탐율 70% 이상	• 시험
• 자바스크립트 난독화 여부 자동 판단		• 정탐율 90% 이상	• 시험
• 탐색 우선순위 자율 결정		• 150초 이상 단축	• 시험
• 개인정보 강제 수집 행위 탐지		• 정탐율 75% 이상	• 시험
• 바이너리 난독화 여부 자동 판단		• 정탐율 90% 이상	• 시험
• 바이너리 난독화 자동화 해제		• 난독화 해제율 75% 이상	• 시험

o 시험개발

항목		평가 내용	비고
• 동적 크롤러	• 대상 브라우저	• 5개 이상	• 시험
	• 동시 실행 브라우저 수	• 30개 이상	• 시험
	• 10만개 도메인 분석 주기	• 1일 2회	• 시험
• 악성코드 유포 행위 탐지		• 정탐율 90% 이상	• 시험
• 기계학습 기반 악성 코드 유포 행위 예측		• 정탐율 87% 이상	• 시험
• 탐색 우선순위 자율 결정		• 300초 이상 단축	• 시험
• 개인정보 강제 수집		• 정탐율 90% 이상	• 시험

행위 탐지		
• 바이너리 난독화 자동화 해제	• 난독화 해제율 85% 이상	• 시험

※ 상세한 평가방법 및 절차는 TRR시 확정

6. 참여 요건

가. 추진 체계 요건

o 주관연구기관 및 참여기관 : 제7조제2항 및 동법 영 제14조제2항 각 호에 해당하는 기관 또는 단체

o 응용연구 및 시험개발의 경우에는 주관연구기관 또는 참여기관에 1개 이상의 기업 참여 필수(제27조제4항)

o 기업분담율 : 민·군기술협력사업 공동시행규정 제27조

나. 연구책임자의 자격 및 과제 신청요건

o 연구책임자의 자격 : 관련분야의 연구 경험이 풍부한 중견 연구자를 책임자로 선임하여 연구의 최종목표를 달성할 수 있도록 계획, 업무프로세스 정립, 원활한 추진 및 조정과 과제관리를 수행할 수 있어야 한다.

o 과제 신청요건 : 주관연구기관은 제안한 연구개발 목표를 충분히 달성할 수 있는 연구팀을 구성하여야 하며, 필요시 컨소시엄을 구성할 수 있다.

다. 기타

o 최종평가는 공인시험기관의 성적서를 반영하여 평가

※ 민·군기술협력사업 공동시행규정 제31조(별표9)

o 년차평가는 매년 11월 수행을 가정하여 계획수립

7. 참고문헌

※ 제안서 작성시 인용된 참고자료 기술

8. 과제 문의사항 연락처

소속	성 명	연락처
민군협력진흥원	김도선	042-607-6016