소프트웨어 등록증

등록번호	**ASSET_0010228**
등록일	2023년 11월 24일
등록구분	국가연구개발사업 성과물, 소프트웨어(기술상세정보)
등록기관	모아소프트 서울특별시 송파구 오금로 422, 502호
등록명	LSTM 모델 기반 AI 악성 도메인 판별 프로그램

NTIS 소프트웨어 ID	202311020063	**성과발생연도**	2023년
과제고유번호	9991008586	**과제기준년도**	2022년
과제명	AI 기반 IP/서버 자동 보안 관제 시스템 개발		

위의 소프트웨어는 국가연구개발사업을 통해 창출된 성과로서,
소프트웨어 분야 연구성과 관리·유통 전담기관에 등록되었음을 증명합니다.

2024년 07월 09일

nipa
National IT Industry Promotion Agency

정보통신산업진흥원
충청북도 진천군 덕산면 정통로 10

LSTM 모델 기반 AI 악성 도메인 판별 프로그램

기술명

LSTM 모델 기반 AI 악성 도메인 판별 프로그램

수행기관

회사명	모아소프트	사업자등록번호	2158187398
대표자	장주수	홈페이지	http://www.moasoftware.co.kr
회사대표전화	0269452180	주소	서울특별시 송파구 오금로 422, 502호

기술개요

개발 배경
• 1980년대 초 악성 코드가 최초 발생하였으며 현재까지 지속적으로 심각한 피해를 끼치고 있음
• 악성 코드는 컴퓨터 시스템과 네트워크를 심각하게 훼손 시킬 뿐만 아니라, 사용자의 개인 정보와 기업의 기밀 데이터를 노출하는 등 심각한 사이버 범죄를 일으키는 주범
• 악성 코드 및 도메인 탐지에 주로 사용되는 시그니처 및 행위 기반 같은 전통적 탐지 방법은 DGA(Domain Generation Algorithm)로 인한 악성 도메인을 차단 하는 데 한계를 지님
• DGA 악성 도메인을 효과적으로 탐지하기 위한 새로운 방법 모색 및 도입

개발 필요성
• 악성 코드는 데이터 유출 등의 심각한 사이버 범죄를 일으키는 주범
• 최근 악성 코드 작성자는 악성 코드의 주 거점 지대인 C&C 서버를 은폐 시키고, 추적을 피하고자 보다 발전된 방법인 DGA(Domain Generation Algorithm, 도메인을 일정 주기에 따라 무작위로 대량 생산하는 방법)을 악성 코드 내에 심어 악성 도메인 판별 및 추적을 어렵게 하고 있음
• 이에 따라, 최신 발전 AI 모델인 LSTM 모델을 활용하여 악성 도메인을 자동 판별하는 방법이 가장 효과적인 대응 방안

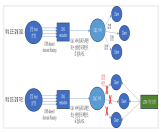
기술 개념
• DGA 알고리즘을 구현하여 악성 도메인을 무작위로 생산
• 악성 도메인과 정상 도메인 데이터를 기반으로 구축된 LSTM 기반의 AI 모델로 학습 및 테스트를 진행
• AI Deep Learning 기술을 이용하여 정산 도메인과 악성 도메인의 구분하고 어떤 도메인이 악성 도메인 인지를 판별
• 일정 수준 이상의 판별 정확도(AUC; Area Under Curve)로 해당 도메인의 판별 신뢰도를 평가

기술 특징
• DGA 기술 구현
• 다양한 AI모델을 사용하여 구축된 고유의 AI 모델로 정확도 99%이상의 악성 도메인 판별
• 교차 검증 및 5-fold방식을 사용하여 보다 검증된 학습 및 테스트 방식 사용
• 차별화된 데이터 저장 및 사용 방식으로 학습 및 테스트 속도 개선
• 학습 및 테스트 결과에 대한 다양한 시각적 자료 제공

대표 기능
• 악성 코드 14종(banjori, corebot, cryptolocker, dircrypt, kraken, lockyv2, pykspa, qakbot, ramdo, ramnit, simda, matsnu, suppobox, gozi)에 대한 악성 도메인 무작위 생성
• 정상 도메인 및 악성 도메인 데이터 전처리
• LSTM기반 AI모델로 데이터를 학습 및 테스트하고, 악성 도메인 판별 및 예측
• 판별 결과에 대한 정확도 계산

적용 분야
• 사이버 보안 분야
• 국방 및 방산 보안 분야
• 시계열 등의 특성을 지닌 데이터를 통하여 AI 예측이 필요한 분야
• 기타 인공지능 적용 및 보안 관련 분야

[참고 그림 : 악성코드의 역사]
[참고 그림 : 해당 기술의 필요성]

등록한 특허

발명명칭	출원번호	출원일	등록번호	등록일
자료가 없습니다.				

등록한 표준

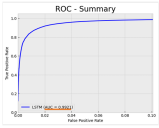
표준명칭	등록번호	등록일
자료가 없습니다.		

경쟁력

개발 기술에 관련된 국내/외 기술 현황
• DGA의 DNS based domain fluxing 방식을 사용하여 매일 수많은(수 백 개에서 수 만 개까지) 도메인 주소를 무작위로 생성하는 경우, 보안 장비나 사법 기관에서 C&C서버 접속을 막는 것이 현실적으로 어려움. 악성 도메인으로 인해 시스템 고장 또는 사고 발생 시 많은 희생과 고비용을 초래하고, 복구와 신뢰 회복에 많은 자원과 노력이 소모되기 때문에 높은 신뢰성을 요구함
• 최근의 국제 입찰에서는 시스템의 신뢰성을 중요시하여 제작 요청서 및 제안서에 신뢰성 조건을 충족시키도록 요구하고 있음. 현재 이러한 중요성을 인식하고 다양한 분야에서 신뢰성 관리 체계에 대한 도입과 연구가 국내에서 활발히 진행되고 점차 국제 수준에 걸맞게 변하는 추세지만 아직 대부분의 연구에서 악성 도메인 차단에 대한 미흡한 신뢰도를 보여줌
• 현재 사람이 수동적으로 악성 도메인을 일일이 차단하기 어려울 뿐만 아니라, 기존 시그니처나 행위 기반 방식의 악성 도메인 자동 차단 방식은 새롭게 변형되거나 은폐 시도하는 도메인의 경우 차단률이 매우 낮음
• 많은 국내외에서 LSTM 모델로 인한 악성 도메인 차단 방식이 연구되고 있으나, 신뢰도 측면에서 많이 미흡한 결과를 보여주고 있음

타 기술 대비 해당 기술 특징 및 장점
• 악성 도메인 판별을 위한 자체 AI 모델 구축
• 악성 도메인을 판별을 위해 사용된 본 모델은 AUC 기준 정확도 99.21%이상 달성(그림 '본 기술 AUC 정확도' 참조) 기술하여 국내/외에서 연구된 기존 LSTM 모델보다 훨씬 높은 성능 및 신뢰도를 지님
• 본 기술은 공용어인 영어를 기반으로 작성되어 전 세계에서 사용 가능
• Pickle, JSON, CSV 파일 등의 활용으로 빠르게 데이터를 저장하고 활용하여 전체적으로 데이터 처리 속도가 높음

[참고 그림 : 본 기술 AUC 정확도]



기술이전 실적 및 방법

기술이전 실적유무	No		
희망이전 유형		담당자명	정주원
담당자 전화번호	010-9954-7133	담당자 이메일	jwjung@moasoftware.co.kr

개발 목표

• 상업용 AI 기반 IP/서버 자동 보안 관제 시스템 구축 기반 마련
• 최적의 AI 모델로 정확도(AUC) 99% 이상 정확도로 악성 도메인 판별 기술을 생산하고, 악성 도메인을 완벽히 차단하여 사이버 범죄 근절 방안 마련

타겟고객 및 시장

• 현재 타겟 고객
 - 한국항공우주산업, 한화에어로스페이스, 한화시스템, 현대로템, 현대위아, 한화디펜스, S&T 모티브, STX 엔진, STX 조선해양, 대한항공, 한화, 한진중공업, 대우조선해양, LIG넥스원, LS엠트론, 빅텍, 데일리시큐, 스텔라사이버, WHOIS 등 방산 협력 업체 및 사이버 보안 관련 업체

• 가망 타겟 고객
 - 금융, 항공, 전기전자, 자동차, 플랜트 및 장치 산업 종사 업체, 기타 사이버 보안이 요구되는 방산 관련 중소 업체 혹은 연구 기관

• 잠재 타겟 고객
 - 방산, 금융, 자동차, 항공 우주, 의료 산업, 각종 엔지니어 및 관리자 등

• 미래 타겟 고객
 - Pratt & Whitney, Boeing, Lockheed Martin, Northrop Grumman, Raytheon Technologies, Bell Textron, Rockwell International, Microsoft, Google, Intel, Ndivia 등 국외 방산 및 사이버 보안 관련 업체

• 국내외 경쟁 기업의 기술 개발 현황 및 제공 시장 규모
 - 국내외 경쟁 기업의 기술 개발 현황: attack graph기반 공격 경로 시뮬레이션, 유사 도메인 검색을 통해 잠재적 피싱 도메인 탐지, 크롤링을 통한 도메인 탐지, 블랙리스트 기법 등을 주로 사용하여 악성 도메인 사전 탐지
 - 하지만 이는 이미 알려진 악성 도메인에 대해서만 유효하며, 알려지지 않은 악성 도메인을 사전 탐지하는 데 한계가 있음
 - 알려지지 않은 도메인을 유추 할 수 있다 하더라도 그 정확도 및 신뢰성이 본 기술에 비해 낮은 수준을 보임
 - 기술 제공 시장 규모: 국내외 방산 및 사이버 보안 관련 시장 규모는 그림 '국내외 시장 규모' 참조

[참고 그림 : 국내외 시장 규모]

세부구성기술

• 동작 원리
 - 무작위로 생성된 도메인 중에서 악성 도메인을 판별하고 정확도를 출력

• 내부 구조 (그림 'LSTM기반 모델 가상환경 구조' 참조)
 - DGA generator와 DGA predictor로 구성

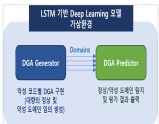
• 대표 기능 및 서비스 (그림 '내부 동작 프로세스' 참조)
 - 악성 도메인 판별
 - 판별 정확도 출력

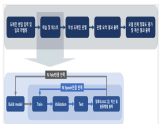
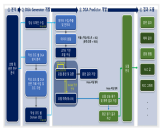
• 세부 기술
 - 세부 기능 및 역할 (그림 'LSTM기반 모델 가상환경 구조' 참조)
  (1) DGA Generator: 악성 코드 별 DGA를 구현하여 대량의 정상 및 악성 도메인을 임의 생성
  (2) DGA Predictor: 학습 및 테스트를 통해 정상 및 악성 도메인 탐지하고 평가 결과를 출력

 - 세부 동작 원리 (그림 '내부 동작 프로세스' 및 '모델 세부 구조' 참조)
  (1) 정상 도메인 데이터는 Alexa Top 1M에서 수집, 악성 도메인 데이터는 본 모델 가상 환경 내에 구현하여 동적 생성
  (2) 악성 도메인과 정상 도메인을 수집하여 학습 및 테스트를 진행
  (3) 무작위로 생성된 도메인들 중에서 악성 도메인을 판별하고 정확도 출력

 - 기타 설명은 그림 '내부 동작 프로세스' 참조
 - 세부 내부 구조: 그림 '모델 세부 구조' 참조

[참고 그림 : LSTM기반 모델 가상환경 구조]
[참고 그림 : 내부 동작 프로세스]
[참고 그림 : 모델 세부 구조]



개발기간 및 소요공수

개발기간(시작일)	2022-08-01	개발기간(종료일)	2024-07-31
총사업비	150,000,000		

년도별 사업비 내역

• 1차년도 : 71,000,000
• 2차년도 : 71,000,000
• 3차년도 : 8,000,000

개발환경(시스템환경, 도구, 개발언어, 제약사항)

• 기술 구현에 필요한 하드웨어(개발 장비 및 개발자 장비 동일)
 - 노트북 및 데스크탑, 혹은 서버(CPU Intel Xeon Gold X 2, 메모리 256GB, 디스크 2TB, 그래픽 Quadro RTX A5000 X4 스펙 이상)

• 기술 구현에 필요한 소프트웨어(개발 장비 및 개발자 장비 동일)
 - Python Editor(개발 당시 Visual Studio Code 사용)

• 기술 구현 시 생산성 및 품질 제고를 위해 활용한 지원 도구
 - 서버(CPU Intel Xeon Gold X 2, 메모리 256GB, 디스크 2TB, 그래픽 Quadro RTX A5000 X4 스펙): AI 모델 학습 및 테스트 용
 - Jupyter notebook: AI 학습 및 테스트에 필요한 주요 라이브러리 탑재

• 사용 언어
 - Python

• 사전에 고려한 기술적 제한 사항
 - 실제 발생된 악성 도메인 수집을 위한 개발 인력 및 시간적 제약
 - 실시간 웹 개발 구현의 개발 인력 및 시간적 제약

• 가정 및 종속 사항
 - 구현된 대표 악성 도메인 종류들이 악성 도메인의 대부분 특성을 다 포괄하고 있다고 가정

• 제외된 기능 및 서비스
 - 사용자가 판별할 임의의 도메인을 수동 입력하지 않고, 판별할 도메인이 시스템에 의해 자동으로 다량 입력됨

• 개발 기술의 제약 사항
 - 압축 파일 안에 있던 파일들은 모두 같은 폴더 안에 있어야 함
 - 최초 실행 전 'results.pkl' 이나, 'traindata.pkl' 파일이 생성되어 있지 않아야 함. (실행 전 'results.pkl' 이나, 'traindata.pkl' 파일이 남아 있게 되면, 이전에 학습하거나 테스트한 결과 값을 불러와 출력. 즉, 새로운 데이터에 대한 학습이 다시 진행되지 않으며, 새로운 데이터 에 대한 학습 과정도 보여지지 않음. 그러나 만약 이전 학습에 대한 결과 값을 다시 확인하고 싶으면 해당 파일을 삭제하지 않은 채 다시 프로그램을 실행시키면 됨)
 - 실행 전 numpy, keras, sklearn, tensorflow, matplotlib 라이브러리가 설치되어 있어야 함

• 개발 기술의 한계
 - 가상 환경에서 구현되었기 때문에 실시간 악성 도메인을 판별하기 어려움. 추후 웹으로 구현이 필요함
 - 악성 코드 14종에 대한 악성 도메인 판별이 가능. 그 이외의 악성 도메인을 판별하기 위해서는 더 많은 데이터 수집이 필요함

개발관리 및 적용방법론

개발방법론 유형	

개발관리 및 적용방법론 상세설명

• 적용 개발 방법론
 - 개발 방법론명: Agile 방법론
 - 방법론 선정 사유 및 배경: Agile 방법론은 프로세스에 얽매이지 않고 코딩 작업을 통해 끊임없이 새로운 프로토타입을 제시하는 방법. Agile 방법론 중 XP(eXtreme Programming) 방식을 일부 사용하여 프로토타입을 빠른시간안에 주기적으로 제작하고, 제작된 프로토타입을 반복 테스트하고 수정함으로써 소프트웨어를 철저히 관리하고 고객이 가장 만족할 수 있는 방향으로 소프트웨어를 개발. 또한 프로그램 개발 초기에 요구사항을 명확히 하여 개발 기간 동안 요구 사항 변경을 최소로 하고, 여러 사람이 아닌 적은 인원으로 개발을 진행하여 공동 작업량을 줄임. 이에 따라, Agile 방법론의 단점이 상쇄되어 Agile 개발 방법론을 적용하는 것이 가장 최적의 선택으로 작용

• 형상 관리 활용 도구 및 목적
 - 형상 관리 활용 도구: Microsoft Teams 및 GitLab
 - 형상 관리 활용 목적: 문서 및 프로그램 버전 관리, 이력 관리, 일정 관리 등

• 기술 개발 과정 상에서 발생하는 산출물 및 공유 가능 산출물
 - 최종 산출물 결과서 1종

적용한 표준

• 방위사업청 '무기체계 SW개발 및 관리 매뉴얼'
 - 한국 국방 무기체계 SW 개발을 위해 따라야하는 SW 개발 프로세스 표준. 개발 프로세스별 제출해야 하는 산출물과 역할의 업무가 제시되어 있음

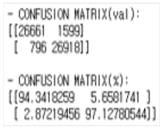
테스트 기준

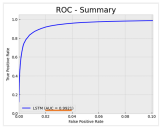
• 초기 정의된 테스트 목표 대비 실제 테스트 결과 (그림 '본 기술 AUC 정확도' 참조)
 - 초기 정의 테스트 목표치: AUC정확도 90% 이상
 - 실제 테스트 목표 달성치: AUC정확도 99% 이상

세부 품질 점검 (그림 '실제 전체 품질 평가 결과(Confusion Matrix)' 및 '본 기술 AUC 정확도' 참조)
• (1) Epoch 단위별
 - loss 값과 AUC정확도 평가: Epoch가 증가될수록 loss가 감소하고 AUC는 증가되는 지 평가
 - Confusion Matrix인한 평가: Epoch가 증가될수록 FN(False Negative)과 FP(False Positive)가 감소하고, TN(True Negative)과 TP(True Positive)가 증가하는 지 평가

• (2) 전체 품질 평가
 - ROC그래프 및 AUC정확도 평가: ROC그래프가 우상향하고, AUC 정확도가 0.9~1 사이에 위치하여 90%이상의 성능이 보이는 지 평가
 - Confusion Matrix인한 평가: TP와 TN의 비율이 90%가 넘는 지 평가

[참고 그림 : 실제 전체 품질 평가 결과(Confusion Matrix)]
[참고 그림 : 본 기술 AUC 정확도]





테스트 방법

테스트 기법 및 방법 (그림 '본 모델 테스트 절차' 참조)
• (1) 데이터 셋 분할
 - 데이터 셋을 학습용 데이터 셋과 테스트용 데이터 셋으로 나누어 모델의 성능 평가
 - 테스트용 데이터 셋은 과대 적합을 방지하고 모델의 일반화 성능을 평가하기 위해 사용
 - 8:2 비율로 학습용 데이터 셋과 테스트 데이터 셋으로 나누며, 나누기 전에 랜덤하게 섞어 데이터 편향 방지

• (2) 학습용 데이터 셋 재 분할
 - 학습용 데이터 셋을 다시 95:5 비율로 훈련용 데이터 셋과 검증용 데이터 셋으로 나누어 모델의 학습 과정을 최적화
 - 검증용 데이터 셋은 모델의 하이퍼파라미터 튜닝 등에 활용

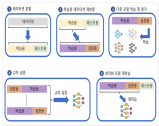
• (3) 모델 학습 및 평가
 - 모델 학습 과정에서 검증용 데이터 셋을 활용하여 모델의 성능 평가 및 최적 상태의 모델 선정

• (4) 교차 검증
 - 모델의 성능을 더 객관적으로 평가하기 위해 학습용 데이터 셋을 5등분하고 5번 반복하는 교차 검증 수행
 - 이를 통해 모델의 성능 평가 결과를 계산하여 모델의 안정성 확인

• (5) 최적화 모델 종합 평가
 - 최적의 모델을 보다 신뢰성있게 평가하기 위해, 최적의 모델에서의 테스트 평가 결과를 다시 정밀 평가

• 테스트 지원 도구
 - Jupyter notebook

[참고 그림 : 본 모델 테스트 절차]



인증 현황

인증기관	인증구분	인증번호	인증일	인증서첨부
자료가 없습니다.				

제품구성

주요기능

제품 특장점

사용 환경

제품관리정책

부속자료(사용자 매뉴얼, 백서, 데모, 브로셔 등)