머신러닝을 이용한 소프트웨어 취약점 자동 분류 장치 및 방법

기술 개요
Overview

① 적용분야
소프트웨어 취약점 분석 및 공격 탐지 시스템

② 기술요약
취약점 분석 대상인 소프트웨어 바이너리 파일에 따라 상이한 학습 데이터 셋과 상이한 학습 알고리즘을 통해 생성된 복수의 취약점 분류 모델 중 하나를 선택하여 취약점 분석을 수행함으로써, 단일 학습 알고리즘과 단일 학습 데이터 셋 사용으로 인한 과적합 문제를 해결함과 동시에 취약점 분류의 정확성을 향상시킬 수 있음

③ 특허 권리 범위
독립항은 복수의 취약점 데이터 셋 중 하나와 복수의 학습 알고리즘 중 하나를 이용한 학습을 통해 생성된 복수의 취약점 분류 모델 중 소프트웨어 바이너리 파일에 대한 취약점 분류 모델을 결정하여 취약점 분석을 수행하되, 소프트웨어 바이너리 파일과 복수의 취약점 데이터 셋 사이의 유사도와 복수의 취약점 분류 모델 각각의 분류 정확도에 기초하여 취약점 분류 모델을 결정하는 특징을 포함하고 있음
종속항들은 취약점 분류 모델 결정의 구체적 방식, 취약점 분류 모델 생성을 위해 이용되는 학습 알고리즘의 종류, 취약점 분류 모델 생성을 위해 이용되는 학습 알고리즘과 취약점 데이터 셋의 선택 방식 등에 대한 상세 기술요소들을 포함하고 있음

기술의 목적
퍼징(fuzzing), 기호 실행(symbolic execution) 등을 이용하는 종래 소프트웨어 취약점 분석 기술은 취약점 탐지에 많은 시간이 소요되고 데이터에 대한 전문가 수준의 이해를 요구함. 또한, 머신 러닝 또는 딥 러닝 알고리즘을 이용하는 종래 인공지능 기반 소프트웨어 취약점 분석 기술은 학습 알고리즘의 종류, 학습 데이터의 양 및 종류에 따라 학습된 취약점 분류 모델의 성능 차이가 발생하고 여러 학습 데이터 셋을 이용하여 학습하더라도 과적합(overfitting) 문제가 발생할 수 있음

해결 방안
각각 상이한 취약점 데이터 셋과 상이한 학습 알고리즘을 이용하여 학습된 복수의 취약점 분류 모델을 생성한 후, 복수의 취약점 분류 모델 중 취약점 분석 대상인 소프트웨어 바이너리 파일에 적합한 취약점 분류 모델을 선택하여 취약점 분석을 수행함

기술의 특장점
퍼징, 기호 실행 등을 이용한 종래 소프트웨어 취약점 분석 기술에 비해 분석 정확성과 속도를 향상시킬 수 있고, 종래 인공지능 기반 소프트웨어 취약점 분석 기술이 가지고 있는 과적합 문제와 학습 알고리즘의 종류, 학습 데이터의 양 및 종류에 따른 분석 정확성 문제를 해결할 수 있음

기술적용 시 기업의 이점

다수의 학습 알고리즘과 학습 데이터셋의 조합을 이용하여 사용자의 요구와 취약점 분석 대상인 소프트웨어에 적합한 취약점 분석 모델을 결정하여 취약점 분석을 수행할 수 있으며, 이를 통해 취약점 분석의 정확성과 성능을 향상시킬 수 있음

SWOT분석
Analysis

S 강점

다수의 학습 알고리즘과 다수의 학습 데이터 셋의 조합을 통해 사용자 요구와 취약점 분석 대상 소프트웨어의 특성에 적합한 취약점 분석이 가능하며, 단일 학습 알고리즘과 학습 데이터 셋을 이용하는 종래 인공지능 기반 취약점 분석 기술이 가지고 있는 과적합 문제와 정확도 하락 문제를 해소 가능함

W 약점

다양한 특성의 소프트웨어 및 다양한 공격 유형에 대응하기 위해 보다 다양한 학습 알고리즘과 학습 데이터 셋을 확보할 필요가 있으며, 취약점 분석 성능이 취약점 분석 모델 학습에 이용된 학습 알고리즘과 학습 데이터 셋에 의존적임

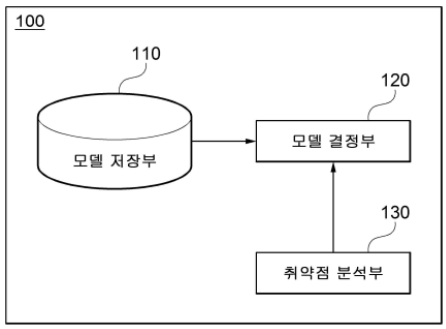
O 기회요인

최근 오픈소스를 활용한 소프트웨어 개발이 보편화되면서 소프트웨어 공급망 관리의 중요성이 높아지고 있고 소프트웨어에 내재된 보안 취약점에 대한 공격을 포함한 사이버 보안 위험이 증가하고 있어 소프트웨어 취약점 분석을 통한 보안성 강화에 대한 필요성과 수요가 높아지고 있음

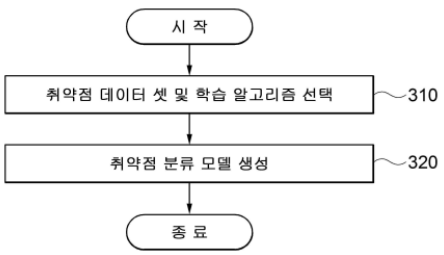
T 위험요인

해외 기업들에 의해 시장이 주도되고 있으며, 다양한 국내외 업체들에 의해 다양한 분석 기술이 상용화되고 있어 시장 경쟁이 심화되고 있음

대표도면
Drawing



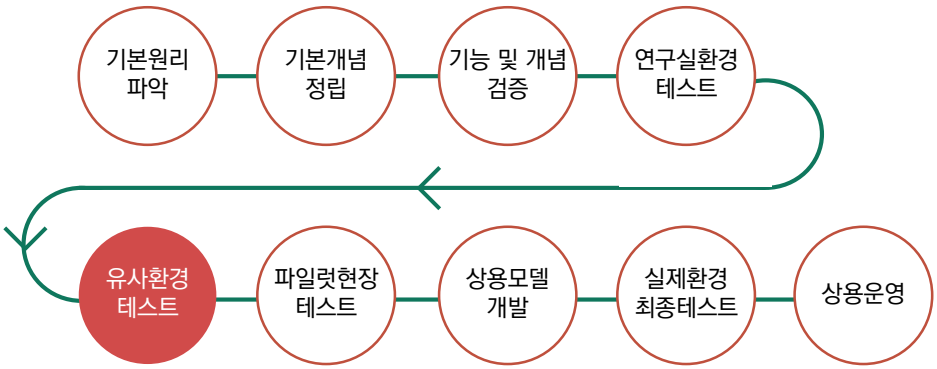
〈 소프트웨어 취약점 분류 장치의 블록도 〉



〈 소프트웨어 취약점 분류 방법의 순서도〉

기술의 완성도
Technology Readiness level

● : 현재 단계입니다.



특허현황
Patent status

발명의 명칭	출원번호	등록번호	출원국가
소프트웨어 취약점 분류 장치 및 방법	10-2019-0110679 (2019.09.06.)	10-2074909 (2020.02.03.)	한국

기술키워드
Keyword

한글키워드	영문키워드
소프트웨어, 보안, 취약점, 학습, 머신러닝, 딥러닝	Software, Security, Vulnerability, Machine Learning, Deep Learning, Training

발명자
Inventor Info.

교수명	윤주범
소속	세종대학교 정보보호학과
연구분야	SW 취약점 분석, AEG(Automatic Exploit Generation), 네트워크 보안기술 등
E-mail	jbyun@sejong.ac.kr
웹사이트	http://home.sejong.ac.kr/~jbyun/