DEOBFUSCATOR

Research Proposal

2025.03.15

컴퓨터융합학부 202002514 안상준
인공지능학과 202202487 박혜연
컴퓨터융합학부 202202602 손예진

Contents

01 연구 주제 이름

바이너리 프로그램에서 제어 구조를 식별하는 도구 개발

가상화 난독화를 역난독화 할 수 있는 모델의 개발

LLM 학습을 위한 역난독화 데이터셋을 생성, LLM이 수행한 역난독화를 신뢰성 있게 사용할 수 있다는 것을 입증하고자 함

가상화 난독화

가상화 기반의 난독화된 악성코드의 증가로 인한 분석의 어려움 증가

패턴 탐색

악성코드를 분석하기 위한 패턴 탐색

LLM 학습

데이터셋 부족 및 신뢰성 문제로 난독화/역난독화 등의 보안 분야에는 사용되지 못한 LLM 학습

목적

LLM을 이용하여 가상화 난독화된 코드를 역난독화하는 것

가상화된 코드 분석을 통한 원본 코드의 구조 확인할 수 있는데, 이를 이용하여 위/변조된 프로그램이나 난독화된 악성코드를 탐지하여 피해를 막고자 한다.

그러한 과정에 LLM을 사용하여 더 빠른 분석을 가능하게 할 수 있을 것으로 예측한다.

04 탐구 내용 및 기대 결과

보안 분야에서 LLM이 사용될 수 있음을 입증
이를 활용함으로써 코드 분석에 필요한 시간의 감소

사용하기 어려울 것으로 예측되는 데이터셋에 대하여,
직접 데이터셋을 구축하여 LLM의 학습에 활용
추후 다른 연구에도 사용될 수 있을 것

05 프로젝트 관련 학습 계획

학습할 내용	기간	역할 분담
난독화, 컴파일러, 바이너리코드 등 프로젝트에 필요한 기본 지식	03.01 ~ 03.31	박상준, 손예진, 박혜연
IDA 사용법	03.01 ~ 03.31	박상준, 손예진, 박혜연
바이너리 코드에서의 switch, while 구조	03.01 ~ 03.31	박상준, 손예진, 박혜연
난독화 도구 실습 및 난독화된 코드 분석	04.01 ~ 04.30	박상준, 손예진
최신 LLM활용 방법 학습 및 적용	04.01 ~ 04.30	박혜연

06 연구 일정 계획

구분	특징1	특징2
Switch문 탐지를 위한 특징적인 코드 패턴	03.01 ~ 03.31	박상준, 손예진, 박혜연
LLM 학습을 위한 데이터 생성 및 전처리 방안	04.01 ~ 04.30	박상준, 손예진, 박혜연
LLM을 활용하여 바이너리 코드의 switch-loop 구조 식별 방법 제안	05.01 ~ 05.31	박상준, 손예진, 박혜연
제안된 방법에 대한 실험 및 개선	06.01 ~ 06.30	박상준, 손예진, 박혜연

DEOBFUSCATOR

감사합니다

컴퓨터융합학부 202002514 안상준

인공지능학과 202202487 박혜연

컴퓨터융합학부 202202602 손예진