DEOBFUSCATOR

Project Brainstorming Result

2025.03.26

14조
안상준, 손예진, 박혜연
지도교수 : 조은선

Contents

01 아이디어 발산

회의모습

01 아이디어 발산

난독화	tigress
역난독화	ollvm
llm	vmprotect
바이너리	github
loop	c
switch	gcc
가상화	clang
if else	optimize
pin tool	동적 분석
jump	정적 분석
table	IDA
fine	OllyDbg
tuning	ai
prompting	feature
chat gpt	selection
claude	vectorize

02 아이디어 수렴

핵심 개념	정의 및 하위 개념 목록
AI	1. LLM : claude, chat gpt claude version : 3.7, 3.5 chat gpt : 4.5, 4, o1, o3-mini, o3-mini-high LLaMA : 3.1, 3.2, 3.3 Gemini, PaLM, LaMDA 2. Training : fine tuning, prompting, 새로운 모델 구축 fine tuning : binary에서 feature를 추출하여 학습, max length를 조절하여 최적의 length탐색 prompting : ChatGpt, Claude를 이용하여, zero shot이나 in context learning을 사용 feature : Pin Tool을 이용하여 trace를 뽑아 feature로 사용 새로운 모델 구축 : transformer등과 같은 모델 구조를 사용해서 학습 3. Preprocessing : feature selection, vectorize feature selection : instruction sequence 에서 ai 학습에 불필요한 정보 제거로 모델의 효율성을 높임 vectorize : 기존 자연어 처리 방식을 사용하지 않고 바이너리 코드에 맞는 vectorize방식 사용
도구	1. 분석 : IDA, OllyDbg, PinTool, objdump, ghidra, PEView, QEMU 2. 난독화 : Ollvm, Tigress, VMProtect, Themida, CodeVirtualizer 3. 컴파일러 : gcc, clang, optimize, javac, icc, -o2, -oz, -ofast
난독화	1. 기법 : flattening, opaque predicate, control flow, renaming flattening : 함수의 기본 블록을 하나의 큰 switch 문 안에 배치하여 원래의 제어 흐름을 숨김 control flow : 제어 흐름 난독화는 프로그램의 실행 흐름을 복잡하게 만들어 코드 분석을 어렵게 renaming : 변수, 메서드, 클래스의 이름을 의미 없는 문자열로 변경하여 코드의 가독성을 낮춤 opaque predicate : 항상 참 또는 거짓으로 평가되는 조건문을 삽입하여 불필요한 분기를 만듦
제어구조	1. loop, switch, if-else

02 아이디어 수렴

바이너리에서
구조를 식별하는
LLM 모델 개발

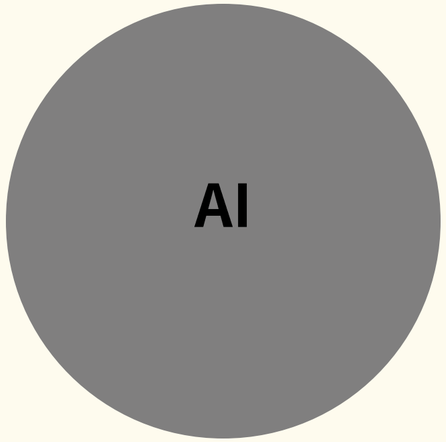
AI

도구

난독화

제어구조

02 아이디어 수렴



LLM	Claude : 3.7, 3.5 Chat GPT : 4.5, 4, o1, o3-mini, o3-mini-high LLaMA : 3.1, 3.2, 3.3 Gemini, PaLM, LaMDA
Training	fine tuning : binary에서 feature를 추출하여 학습, max length를 조절하여 최적의 length를 탐색 prompting : Chat GPT, Claude를 이용하여, zero shot이나 in context learning을 사용 feature : Pin Tool을 이용하여 trace를 뽑아 feature로 사용 새로운 모델 구축 : transformer 등과 같은 모델 구조를 사용해서 학습
Preprocessing	feature selection : instruction sequence 에서 AI 학습에 불필요한 정보 제거로 모델의 효율성을 높임 vectorize : 기존 자연어 처리 방식을 사용하지 않고 바이너리 코드에 맞는 vectorize 방식 사용

02 아이디어 수렴

도구

분석	IDA, OllyDbg, PinTool, objdump, Ghidra, PEView, QEMU
난독화	Ollvm, Tigress, VMProtect, Themida, CodeVirtualizer
컴파일러	gcc, clang, javac, icc, optimize, -O2, -Oz, -Ofast

02 아이디어 수렴

난독화

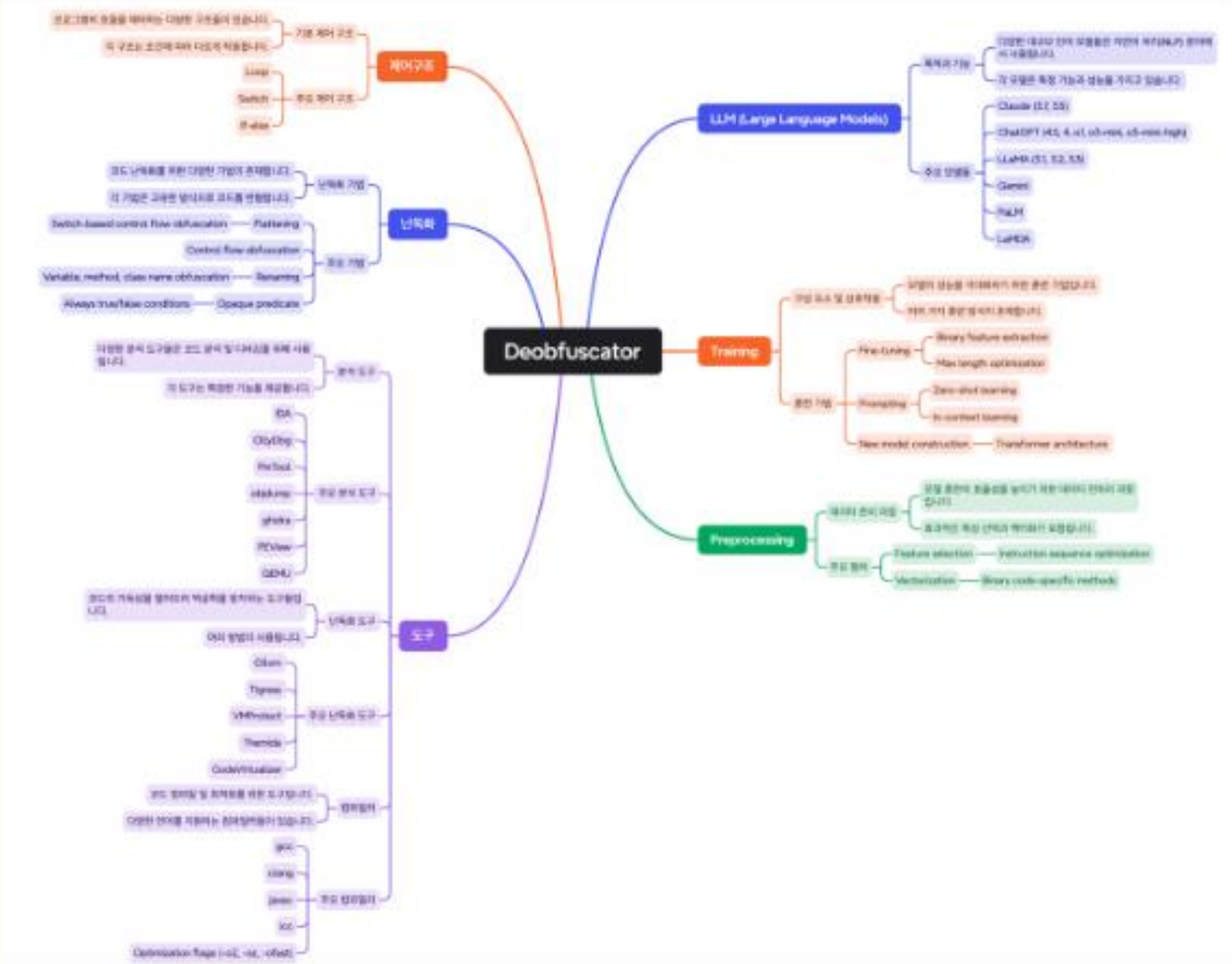
flattening	함수의 기본 블록을 하나의 큰 switch 문 안에 배치하여 원래의 제어 흐름을 숨김
control flow	프로그램의 실행 흐름을 복잡하게 만들어 코드 분석을 어렵게 함
renaming	변수, 메서드, 클래스의 이름을 의미 없는 문자열로 변경하여 코드의 가독성을 낮춤
opaque predicate	항상 참 또는 거짓으로 평가되는 조건문을 삽입하여 불필요한 분기를 만듦

02 아이디어 수렴

제어구조

if-else	cmp, test je, jne, jg, jl
switch	jmp eax
loop	loop: dec ecx jnz loop
call	call function ret

03 시각화


Deobfuscator
LLM (Large Language Models)
Training
Preprocessing
난독화
도구
제어구조
Switch-based control flow obfuscation
Flattening
Control flow obfuscation
Variable, method, class name obfuscation
Renaming
Always true/false conditions
Opaque predicate
Loop
Switch
If-else
IDA
ObjDbg
PinTool
radare2
gdb4s
PEView
GDHU
Obfuscator-LLVM
Tigress
VMProtect
Themida
CodeVirtualizer
gcc
clang
javac
icc
Claude (3.7, 3.5)
ChatGPT (4.5, 4.o1, o3-mini, o3-mini-high)
LLaMA (3.1, 3.2, 3.3)
Gemini
PaLM
LaMDA
Fine-tuning
Prompting
Zero-shot learning
In-context learning
New model construction
Transformer architecture
Feature selection
Instruction sequence optimization
Vectorization
Binary code-specific methods

DEOBFUSCATOR

감사합니다

14조
안상준, 손예진, 박혜연
지도교수 : 조은선