가상 머신 기반으로 난독화된 실행파일의 구조 및 원본의미 추출 동적 방법

이성호, 한태숙

김영철

2015. 12. 30.

요약

- 난독화 기술이 악성 코드 보호에 악용
- 가상 머신 기반으로 난독화된 악성 코드는 분석의 어려움
- 난독화된 프로그램을 분석하는 동적 분석 기반의 프레임워크 제안

서론

- 코드 난독화
  : 코드를 이해하기 어렵게 만들어 분석을 난해하게 만듦
  : 악성 코드에 이용되어 악성 코드 검출을 방해

- 프로그램 의미를 기반으로 악성 코드 탐지
  프로그램 행동을 바탕으로 스파이웨어 탐지

서론

- 가상 머신 기반 난독화
  : 프로그램을 임의의 바이트 코드로 변경
  : 가상 머신을 통해 코드를 실행하여 분석이 어려움

- 가상 머신의 구조를 추출
  원본 프로그램의 의미를 유추하는 분석 기법 제안

서론

1. 프로그램의 동적 트레이스 추출
2. 중간언어 변환, 제어 흐름 그래프 구축,
   분석을 통해 가상 머신의 구성요소 추출
3. 원본 프로그램의 동적 트레이스 복원, 제어 흐름 그래프 복원

관련 연구

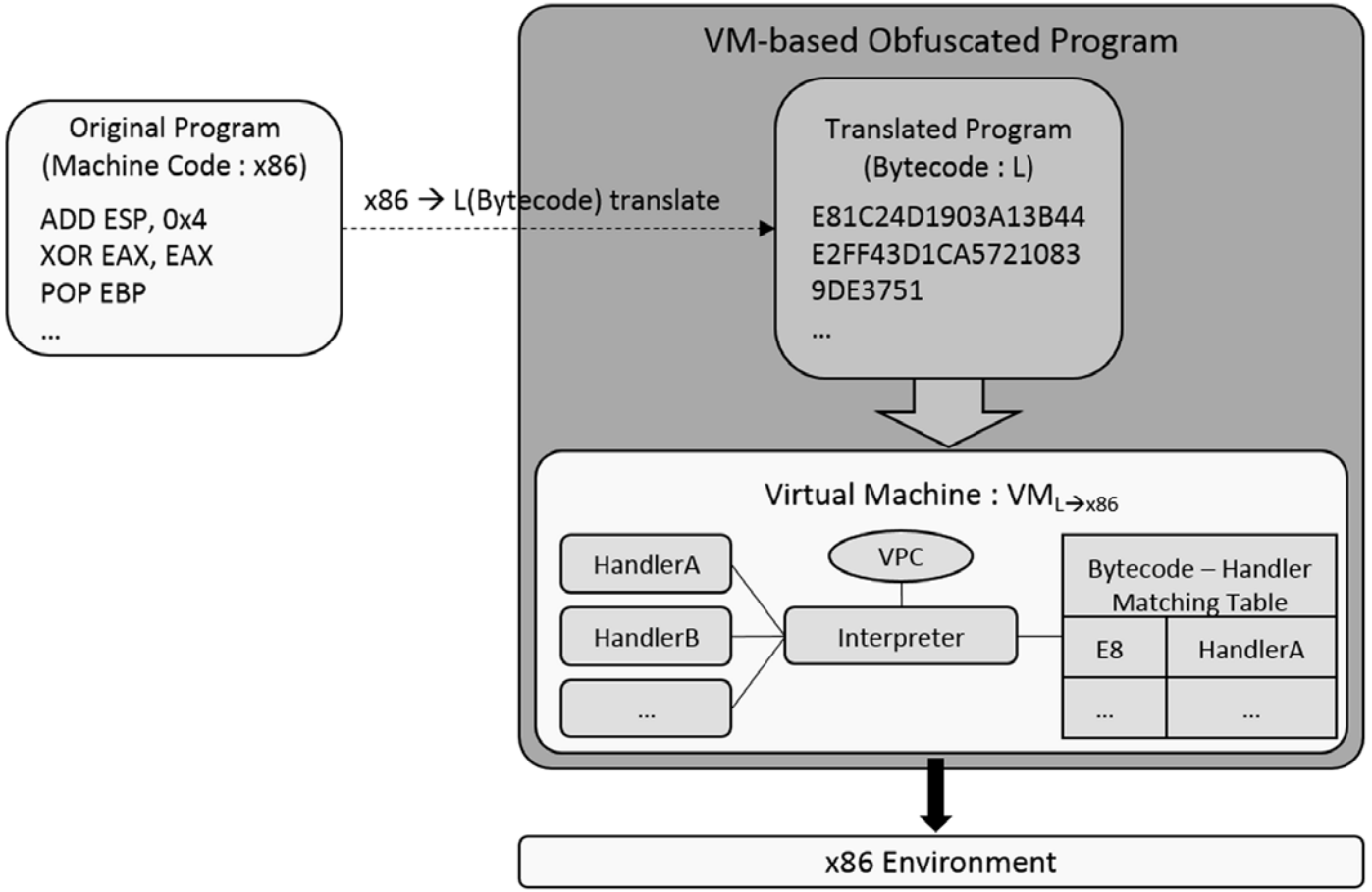
- Rolf Rolles
  VMProtect가 생성한 가상 머신의 구조를 직접 분석

- Monirul Sharif
  동적 트레이스 추출하여 자동으로 가상 머신 구조 분석

- Kevin Coogan
  동적 트레이스에서 시스템 콜 관련 명령어와 인자를 추출하여 원본과 의미적으로 같은 동적 트레이스 복원 기법 소개

- Jason Raber
  동적 트레이스를 반복적인 클러스터링을 하여 바이트 코드 핸들러로 프로그램 제어를 넘겨주는 인터프리터를 제거하는 방법 제안

실행파일의 가상 머신 구조 추출 기법



실행파일의 가상 머신 구조 추출 기법

1. 동적 트레이스 추출

- [실행된 명령, 주소 + 브랜치 대상 주소 + 메모리 입출력 주소]
- Pin을 이용하여 동적 트레이스 추출 도구 ‘Tracer’ 개발
- 헥사 코드를 동적 트레이스 출력 방식으로 채택

실행파일의 가상 머신 구조 추출 기법

2. 중간 언어 구조

- Binary Analysis Platform에서 사용하는 BIL(BAP IL)을 채택
- BIL
  복잡한 의미의 x86 명령을 단순한 명령 리스트로 나타냄
  메모리 입출력에 관한 별도의 구문을 가짐

BIL	Modified BIL
Load (exp, exp, exp, τ_{reg})	Load (exp, exp, integer (integer), exp, τreg)
Store (exp, exp, exp, exp, τ_{reg})	Store (exp, exp, integer (integer), exp, exp, τreg)

<BIL & 수정된 BIL>

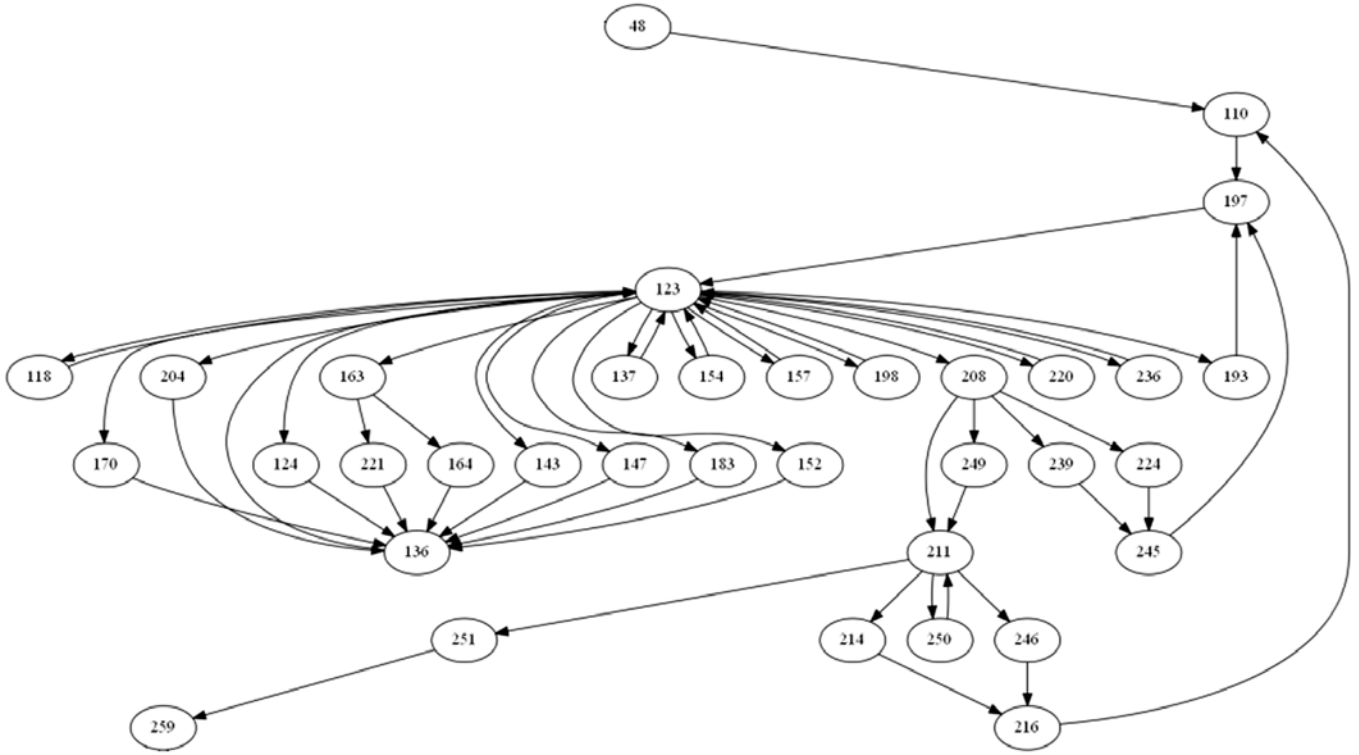
실행파일의 가상 머신 구조 추출 기법

3. 가상 머신 구조 추출

- 제어 흐름 그래프 구축을 위한 규칙
  R1. 베이직 블록 정의
  R2. 베이직 블록 분화
  R3. 제어 흐름 보정

실행파일의 가상 머신 구조 추출 기법

3. 가상 머신 구조 추출



<제어 흐름 그래프>

실행파일의 가상 머신 구조 추출 기법

3. 가상 머신 구조 추출

- 인터프리터 추출 방법
  : 제어 흐름 그래프에서 가장 많음 후임 노드를 가진 노드

실행파일의 가상 머신 구조 추출 기법

3. 가상 머신 구조 추출

- 가상 프로그램 카운터 추출 방법
  : 인터프리터에 쓸모없는 명령들이 다수 존재
  : 6개의 규칙을 정의하여 역방향 슬라이싱 수행

실행파일의 가상 머신 구조 추출 기법

3. 가상 머신 구조 추출

- 가상 프로그램 카운터 추출 방법