

ОПРЕДЕЛЕНИЕ ФУНКЦИОНАЛЬНОСТИ СЕМАНТИЧЕСКИХ ЯДЕР В БИНАРНОМ КОДЕ

Ильина Анна Александровна

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: ilina@ispras.ru

Научный руководитель — Курмангалеев Шамиль Фаимович

В современном мире актуальным направлением является аудит программного обеспечения с целью проведения сертификации. Основные ограничения включают отсутствие доступа к исходному коду и недостаточную документацию функциональности. Это увеличивает время анализа и создает потребность в автоматизированных методах выделения функциональности, ориентированных на бинарный код. Одним из вариантов решения представленной проблемы является выделение семантических ядер (участков кода, реализующих определенную семантику).

Описание метода. В данной работе предлагается метод, позволяющий разделить исследуемый бинарный файл на семантические ядра и определить их функциональность. Предлагаемый алгоритм состоит из нескольких этапов: 1) выделяются функции, которые будут рассматриваться как источники семантических ядер, 2) бинарный код делится на семантические ядра с помощью частичных трасс выполнения, называемых **трасслетами** (которые создаются с помощью инструментов Angr [1] и Crusher [2]), и выделения подграфов из графа вызовов программы, в соответствии с выбранными на предыдущем этапе источниками, 3) для полученных семантических ядер определяется входящая в них функциональность с помощью больших языковых моделей благодаря определению входящих реализаций различных алгоритмов.

Результаты. Тестирование метода производилось на репозитории TheAlgorithms [3], который состоит из реализаций распространённых алгоритмов. В таблице 1 продемонстрированы полученные значения True Positive.

Кроме того, проводилось тестирование на СУБД Microsoft Access 97, состоящей из 10,539 функций. В результате удалось обнаружить встроенную разработчиками программного обеспечения игру Magic Eight Ball, являющуюся недокументированной возможностью.

Категория алгоритмов	True Positive (%)
client_server	77,78
conversions	56,25
games	50
hash	80
math	45
misc	45,45
numerical_methods	50
searching	77,78
sorting	81,48

Таблица 1: Результаты тестирования.

Литература

1. Shoshitaishvili Y. SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis // IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 2016, P.138-157.
2. ИСП Crusher: комплекс динамического анализа программ.
<https://www.ispras.ru/technologies/crusher>
3. TheAlgorithms:
<https://github.com/TheAlgorithms>