ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора физико-математических наук Ильина Вячеслава Анатольевича на диссертацию Вишнякова Алексея Вадимовича «Поиск ошибок в бинарном коде методами динамической символьной интерпретации», представленную к защите на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 (05.13.11) «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Диссертационное исследование А.В. Вишнякова направлено на решение проблемы автоматического поиска ошибок в программном обеспечении в ее современном статусе. При этом, под ошибками понимаются *эксплойты* – ошибки, которые могут быть использованы как уязвимости для достижения вредоносного эффекта в процессе функционирования данного ПО. Эта проблема имеет долгую историю, на протяжении которой «непрерывно» появляются новые типы уязвимостей и, соответственно, разрабатываются новые методы и подходы их поиска и устранения. Это связано как с быстрым ростом количества программных продуктов, доступных для массового практического использования, так и с резким усложнением их структуры, применяемых математических методов и алгоритмических решений, а также постоянно появляющихся новых программно-аппаратных архитектур. Новый этап развития этой проблемы, потенциально имеющий взрывной характер, связан с массовым внедрением в повседневную жизнь цифровых технологий (*Интернет Вещей* ++). Если говорить о современном состоянии этой проблемы в контексте проблематики диссертационного исследования А.В. Вишнякова, то можно отметить следующие технологические достижения:

- стандарт SDL по безопасному циклу разработки ПО в ИТ индустрии;
- *фаззинг* - динамический метод анализа ПО за счет генерации новых входных данных;
- фаззинг с обратной связью по покрытию;
- *гибридный фаззинг* – фаззинг с обратной связью по покрытию, динамической символьной интерпретацией и с учетом семантики данного ПО.

Одним из ключевых направлений развития этих технологий является преодоление избыточной и недостаточной помеченности (данных, зависящих от пользовательского ввода), когда (булевый) предикат пути по трассе машинных инструкций содержит чрезмерное количество ограничений, или часть необходимых ограничений отсутствуют. Основным здесь является анализ методов поиска ошибок с помощью символьной интерпретации. Соответственно, среди существующих разработок, релевантных проблематике диссертации А.В. Вишнякова, можно отметить следующие программные инструменты: QSYM - реализация гибридного фаззинга; и KLEE – поиск ошибок во время символьной интерпретации и снижение избыточной помеченности на основе разбиения на независимые подмножества ограничений в предикате пути; SAVIOR — фреймворк для гибридного фаззинга; IntScope - обнаружения ошибок целочисленного переполнения методами

статического анализа бинарного кода. В диссертации отмечены следующие их аспекты, которые в настоящее время принимают характер серьезных недоработок:

- KLEE - работает с исходным кодом и требует сборки анализируемого кода специальным образом, осуществляет полностью символьную интерпретацию без реального запуска программы (проблемы с масштабированием скорости и потребления памяти);
- SAVIOR - работает с исходным кодом и использует файлы из корпуса фаззера в качестве входных данных для динамической символьной интерпретации, встреченные инструкции эмулируются через KLEE (снижение скорости и точности генерируемых входных данных) в отличие от подхода, когда конкретные значения берутся из реального выполнения программы;
- IntScope – наиболее релевантен разработкам в обсуждаемой диссертационной работе, в которой применяется статический анализ бинарного кода (значит, не генерирует входные данные для воспроизведения ошибок) и символьную интерпретацию для поиска ошибок целочисленного переполнения. Однако код этого инструмента закрыт.

Таким образом, при высокой степени разработанности проблемы автоматического поиска эксплойтов, в последние годы выявился ряд недоработок, на устранение которых была направлена диссертационная работа А.В. Вишнякова. Несомненно, тема диссертационной работы Вишнякова А.В. актуальна и отвечает современным запросам ИТ индустрии разработки надежного программного обеспечения.

Целью диссертационного исследования является разработка метода автоматического поиска ошибок в бинарном коде в парадигме гибридного фаззинга, с сочетанием динамической символьной интерпретации и с обратной связью по покрытию.

По полученным в диссертационной работе результатам отметим ключевые аспекты, определяющие их научную новизну. Предложен алгоритм слайсинга предиката пути для устранения избыточных ограничений, полученные в результате динамической символьной интерпретации бинарного кода. Для этого алгоритма доказаны теоремы о его конечности, корректности, и проведена оценка его вычислительной сложности. Алгоритм слайсинга предиката пути позволяет устранить избыточные ограничения во время символьной интерпретации и повысить точность порождаемых входных данных. Разработан метод построения предикатов безопасности для ошибок деления на ноль, выхода за границу массива и целочисленного переполнения во время динамической символьной интерпретации пути выполнения программы. Разработан метод автоматизированного поиска ошибок при помощи символьных предикатов безопасности для применения после гибридного фаззинга. Разработан метод поиска ошибок в контексте гибридного фаззинга, позволяющий автоматически выявлять истинно положительные ошибки.

Результаты, полученные в ходе диссертационной работы, представлены в шести публикациях, две из которых вышли в научных журналах, рекомендованных ВАК, две в журналах, индексируемых WoS/Scopus, а также две публикации в тезисах

научной конференции. Получено два Свидетельства о государственной регистрации программ для ЭВМ.

Положения, вынесенные диссертантом на защиту, основаны на полученных результатах, и представляют большой интерес для дальнейших прикладных разработок в области автоматического поиска программных ошибок. Их достоверность определяется корректностью постановки задач, использованием современных математических и ИТ методов, формальным исследованием свойств представленного алгоритма слайсинга предиката пути, экспериментальными оценками точности методов (в т.ч. на наборе тестов Juliet), а также обнаружением новых ошибок в проектах с открытым исходным кодом широко используемых в литературе.

Диссертация включает в себя Введение, пять Глав, Заключение и список Литературы из 111 наименований. Общий объем диссертации 131 страниц, 1 рисунок и 9 таблиц.

Во Введении формулируется цель диссертационной работы, обосновывается ее актуальность, теоретическая и практическая значимость работы, сформулированы научные результаты диссертационной работы.

В Первой Главе приводится обзор работ по теме диссертации.

Вторая Глава посвящена предложенному алгоритму слайсинга предиката пути, который позволяет устранять избыточные ограничения из предиката пути. доказаны теоремы о конечности и корректности алгоритма, произведена оценка его асимптотической сложности. Показано, что алгоритм позволяет повысить скорость и точность генерации новых входных данных.

В Третьей Главе представлен разработанный метод моделирования семантики функций. Предложены семантические модели для более 30 функций стандартной библиотеки. Показано ускорение динамической символьной интерпретации и открытия новых путей выполнения программы.

В Четвертой Главе описаны разработанные методы построения предикатов безопасности и автоматизированного поиска ошибок с их помощью в парадигме гибридного фаззинга. Представлены решения проблемы характерные для бинарного кода, в частности, предложен алгоритм определения знаковости арифметической операции по бинарному коду. Представлены результаты оценки метода построения предикатов безопасности на наборе тестов Juliet - для ряда типов ошибок получено повышение точности в разы по сравнению с известными в литературе инструментами.

В Пятой Главе представлены детали программной реализации предложенных методов.

В Заключении приведены основные результаты работы.

Диссертация написана ясным языком, по главам сделаны обоснованные выводы. Автореферат правильно отражает содержание диссертации и полученные результаты.

По диссертации можно сделать замечание - не рассмотрен вопрос об ограничениях по типам ошибок, к которым можно применять разработанные методы. Это

замечание, однако, не снижает общую положительную оценку диссертации А.В. Вишнякова.

Приходим к заключению, что диссертация Вишнякова Алексея Вадимовича по теме «Поиск ошибок в бинарном коде методами динамической символьной интерпретации» является самостоятельным научным трудом, выполненном на высоком научном уровне, и является законченной научной работой. Основное содержание диссертации отражено в опубликованных статьях, доложено на научных конференциях. Полученные научные результаты имеют практическое значение для разработки новых методов автоматического поиска ошибок в бинарных кодах разрабатываемого ПО. Диссертация А.В. Вишнякова соответствует требованиям п. 9 Положения ВАК РФ о присуждении ученых степеней, ее содержание соответствует паспорту специальности 2.3.5 (05.13.11) «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей». Таким образом, соискатель Вишняков Алексей Вадимович заслуживает присвоения ученой степени кандидата физико-математических наук по специальности 2.3.5 (05.13.11) «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент:
Главный научный сотрудник
КК НБИКС-природоподобных технологий
ФГБУ Национальный исследовательский
центр «Курчатовский институт» (г. Москва),
доктор физ.-мат. наук

В.А. Ильин
28 ноября 2022 г.