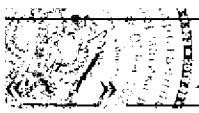
«УТВЕРЖДАЮ»
Директор Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», академик РАН


И.А. Соколов

«11» 11 2022 г.

ОТЗЫВ

ведущей организации – Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской Академии Наук» (ФИЦ ИУ РАН) на диссертационную работу Вишнякова Алексея Вадимовича на тему «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 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Актуальность темы

Непрерывно ведущаяся в мире разработка программного обеспечения, проводимая как отдельными программистами, так и целыми коллективами специалистов ИТ-компаний, неизбежно приводит к проявлению новых ошибок и уязвимостей в создаваемом ПО. Компании внедряют безопасный цикл разработки ПО с целью обнаружения ошибок в программных продуктах до их ввода в эксплуатацию. Методы динамической символьной интерпретации бинарного кода позволяют обнаруживать программные дефекты и порождать входные данные для их воспроизведения. Однако существующие решения обнаружения ошибок при помощи символьной интерпретации либо анализируют исходный код, либо применяют статический анализ, либо же ограничиваются одним типом ошибок. Согласно проведенного в работе анализа, большинство из них мало доступно или непригодно для промышленного применения.

В диссертационной работе Вишнякова А.В. предлагается новый автоматизированный комплексный метод обнаружения ошибок (выхода за границу массива, целочисленного переполнения, деления на нуль и др.) в контексте гибридного фаззинга. Таким образом, тема диссертационной работы Вишнякова А.В. является актуальной.

Характеристика содержания диссертационной работы

Диссертация имеет общий объём 131 страницу и состоит из введения, пяти глав, заключения и списка литературы из 111 источников.

Во введении описывается актуальность проблемы, формируются цель и задачи диссертационной работы, указана научная новизна результатов исследования, раскрыта их теоретическая и практическая значимость и приводятся основные положения, выносимые на защиту.

В первой главе приводится обзор работ по теме диссертации, а также рассматриваются существующие инструменты символьной интерпретации. Приводится обзор методов преодоления избыточной и недостаточной помеченности, поиска ошибок при помощи символьной интерпретации, а также делаются выводы о применимости описанных методов.

Вторая глава посвящена описанию предложенного автором алгоритма слайсинга предиката пути, который позволяет устранять избыточные ограничения из предиката пути. В главе исследуются формальные свойства и характеристики алгоритма, доказываются теоремы о конечности и корректности алгоритма, осуществляется оценка асимптотической сложности алгоритма. Далее приводятся примеры преодоления избыточной и недостаточной помеченности с помощью приведённого алгоритма. В итоге, осуществляется экспериментальная оценка эффективности алгоритма, которая показала, что алгоритм слайсинга предиката пути повышает точность генерируемых входных данных во время динамической символьной интерпретации.

В третьей главе рассматривается разработанный метод моделирования семантики функций, который позволяет пропускать символьную интерпретацию библиотечных функций, описывая их семантику символьными формулами. В главе описываются модели для более 30 функций стандартной библиотеки: динамического выделения памяти, копирования, печати в стандартный поток вывода, строкового сравнения и поиска, преобразования строки в число. Экспериментальная оценка разработанного метода показала, что происходит ускорение динамической символьной интерпретации и открытия новых путей выполнения программы.

Четвёртая глава содержит описание разработанных методов построения предикатов безопасности и автоматизированного поиска ошибок при помощи символьных предикатов безопасности в гибридном

фаззинге. В главе представлены методы построения предикатов безопасности для обнаружения ошибок целочисленного переполнения, выхода за границы массива, деления на нуль и других. Приводится экспериментальная оценка точности методов построения предикатов безопасности на наборе тестов Juliet. Разработанные методы показали общую точность 95.59 % для 11 типов ошибок CWE (15772 теста). В результате апробации предложенного автоматизированного метода поиска ошибок в контексте гибридного фаззинга было найдено 17 новых ошибок в 10 различных проектах с открытым исходным кодом (OpenJPEG, Poppler, miniz, unbound и др.). Результаты продемонстрировали состоятельность метода и его способность обнаруживать новые дефекты в сложных программных системах.

В пятой главе приводится описание деталей реализации предложенных методов в программных инструментах.

В заключении сформулированы основные результаты диссертации.

Новизна и достоверность полученных результатов

Основные научные результаты, полученные в диссертационной работе, являются новыми и заключаются в следующем:

1. разработан алгоритм слайсинга предиката пути, устраняющий избыточные ограничения из предиката пути;
2. разработан метод построения предикатов безопасности для обнаружения ошибок выхода за границу массиву, целочисленного переполнения и деления на нуль;
3. разработан метод автоматизированного поиска ошибок при помощи символьных предикатов безопасности в контексте гибридного фаззинга.

Достоверность результатов подтверждается формальным исследованием свойств представленного алгоритма слайсинга предиката пути (доказаны теоремы о его конечности и корректности, проведена оценка асимптотической сложности), экспериментальной оценкой точности и скорости алгоритма. Точность метода построения предикатов безопасности измерена на наборе тестов Juliet. Качественная оценка автоматизированного метода поиска ошибок при помощи символьных предикатов безопасности проведена в результате обнаружения новых дефектов в различных проектах с открытым исходным кодом.

Теоретическая значимость и практическая ценность

Теоретическая значимость результатов диссертации заключается в разработанных методах построения предикатов безопасности для обнаружения различных видов ошибок, которые применимы для анализа сложных программных систем в контексте гибридного фаззинга. Более того, разработан алгоритм слайсинга предиката пути, который позволяет

устранять избыточные ограничения во время динамической символьной интерпретации. Формально исследованы свойства алгоритма и доказаны соответствующие теоремы. Практическая значимость полученных результатов заключается в том, что предложенные методы позволяют обнаруживать новые программные ошибки и порождать входные данные для их подтверждения. Автоматически выделяются истинно положительные срабатывания. Представленные методы встраиваются в системы непрерывной интеграции в рамках безопасного цикла разработки ПО и тем самым способствуют повышению безопасности разрабатываемого ПО. Методы позволили обнаружить новые программные дефекты и используются в Центре доверенного искусственного интеллекта ИСП РАН.

Замечания по работе

Несмотря на общую положительную оценку диссертационной работы, следует отметить несколько недостатков:

1. в третьей главе не раскрыт вопрос обработки ошибочных ситуаций при моделировании семантики функций;
2. в четвёртой главе не указывается время построения предикатов безопасности и обнаружения ошибок.

Заключение по работе

Отмеченные недостатки не снижают положительной оценки диссертационной работы. Диссертация Вишнякова А.В. является законченной научно-квалификационной работой, в которой решена актуальная задача разработки нового автоматизированного комплексного метода обнаружения ошибок в контексте гибридного фаззинга. Диссертация соответствует паспорту научной специальности 2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (пп. 1, 2). Результаты диссертации в полной мере отражены в публикациях автора, в том числе две – в журналах из перечня ВАК, две – в журналах Web of Science и Scopus, и апробированы на крупных научных конференциях. Автореферат содержит в краткой форме все основные результаты, полученные в диссертации, и соответствует содержанию диссертации.

Диссертационная работа Вишнякова А.В. полностью соответствует всем требованиям ВАК РФ, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук, а Вишняков Алексей Вадимович заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Отзыв на диссертационную работу рассмотрен и утвержден на заседании секции Ученого совета ФИЦ ИУ РАН 21 ноября 2022, протокол №4. Присутствовало на заседании 16 человек. Результаты голосования: принято единогласно.

Руководитель Отделения ФИЦ ИУ РАН, д.ф.-м.н.

Синицин В.И.

«21» ноября 2022 г.

«Подпись Синицина В.И. заверяю»

Ученый секретарь ФИЦ ИУ РАН, д.т.н.

В.Н. Захаров