УТВЕРЖДАЮ
Заместитель директора по научной работе
ИПМ им. М.В. Келдыша РАН,
член-корреспондент РАН, д.ф.-м.н., профессор
М.В. Якобовский
«04» октября 2023 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»
на диссертацию Куца Даниила Олеговича
на тему «Метод моделирования косвенной адресации в рамках динамической символьной интерпретации»,
представленную к защите на соискание ученой степени кандидата технических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Практическая значимость и актуальность работы

Диссертационная работа Куца Д.О. посвящена развитию методов анализа программ для увеличения покрытия программ тестами для проведения более глубокого тестирования, обнаружения большего числа ошибок. Автоматизация тестирования является критической в современной индустрии программирования. От ее успехов существенно зависит как стоимость разработки программных систем, так и отсутствие убытков при их эксплуатации благодаря отсутствию критических ошибок в программах. Реализованные соискателем методы расширяют и дополняют программный инструмент Sydr, разработанный и используемый в ИСП РАН для промышленного тестирования программ.

Особенностью инструмента Sydr и данной работы является анализ программ не на языках высокого уровня, а бинарных кодов, полученных после компиляции. Это повышает практическую значимость, поскольку позволяет автоматизировать тестирование программ, для которых недоступны исходные коды, а также сборок программ, созданных на разных языках. С другой стороны, любой анализ бинарных кодов намного сложнее, чем на языках высокого уровня и более продвинутых методов. Одна из

существенных трудностей, на преодоление которой направлена данная работа, — продолжение анализа после операций с косвенной адресацией, как чтений данных, так и переходов.

В диссертационной работе проведено исследование и предложен новый метод моделирования косвенной адресации при анализе бинарных программ методом динамической символьной интерпретации. Символьная интерпретация — это один из широко используемых методов динамического анализа программ, которая позволяет исследовать и порождать описания классов путей исполнения тестируемой программы. Для этого строится математическая модель исполнения программы, в которой входные данные заменены на символьные переменные.

Существенным недостатком классических алгоритмов символьной интерпретации при моделировании машинных команд является учет лишь явно выраженных, прямых зависимостей по данным. Однако в бинарном коде часто встречаются косвенные зависимости по данным, возникающие из-за доступа по адресам, хранимым и выбираемым из памяти машины. При символьной интерпретации такие ситуации представлены доступом по адресу, зависящему от символьных переменных, диапазоны значений которых в существующих методах не известны анализатору. Это является значительным препятствием для эффективной и плодотворной генерации тестов, исполнение которых проходит через такие точки программы.

Варианты решения данной проблемы были предложены в различных существующих инструментах динамической символьной интерпретации. В диссертационной работе рассмотрены как популярные в настоящее время инструменты, объединяющие рандомизированное порождение тестов с символьной интерпретацией, т.н. гибридные фаззеры (QSYM, SymQEMU, SymCC, Fuzzolic), так и некоторые инструменты анализа бинарного кода (angr, KLEE, Mayhem). В работе разбираются различные аспекты работы этих инструментов и выделен ряд критичных недоработок, на устранение которых и направлено данное диссертационное исследование.

В диссертации представлены новые методы построения более точных предикатов на символьные переменные в путях, проходящих через косвенные адресации, и тем самым повышена точность проводимого анализа в целом. В частности, анализ косвенных переходов, организованных через таблицу переходов в памяти программы, полученных в результате компиляции условных переходов с множеством ветвей (switch), позволяет

продолжать анализ путей исполнения программы после перехода.

Разработанные методы и алгоритмы реализованы как расширение, новые подсистемы в инструменте динамической символьной адресации Sydr, и используются в промышленном тестировании, выполняемом в ИСП РАН. В диссертации представлены результаты экспериментального исследования, показывающие увеличение покрытия программ тестами и тем самым повышение качества автоматического тестирования.

Таким образом, тема диссертационной работы Куца Д.О. является актуальной, а результаты обладают высокой практической значимостью.

Общая характеристика работы

Диссертация состоит из введения, четырех глав, заключения и списка литературы из 75 наименований. Общий объем диссертации 113 страниц, 5 рисунков и 7 таблиц.

В первой главе приводится обзор работ по теме диссертации и описание метода динамической символьной интерпретации. Приводится описание существующих методов обработки косвенной адресации, рассматриваются их достоинства и недостатки.

Во второй главе приводится описание предлагаемого метода поиска и моделирования косвенных переходов. Рассматривается проблема символьной интерпретации косвенных переходов, приводятся примеры с такими переходами. Описывается предлагаемый алгоритм поиска косвенных переходов в бинарном коде. Приводится описание разработанного эвристического подхода к определению границ таблицы переходов в памяти программы. Дается экспериментальная оценка эффективности разработанного метода на совокупности тестовых задач, которая показала, что разработанный метод позволяет увеличить число выявленных путей исполнения в тестируемой программе.

Третья глава посвящена описанию предлагаемого метода моделирования чтений по символьно вычисляемому адресу. Описывается подход определения области памяти, к которой может осуществляться доступ по символьному адресу. Приводится описание разработанного алгоритма построения символьных ограничений для моделирования чтения по символьно вычисляемому адресу. В работе рассматривается несколько способов построения символьных ограничений, производится экспериментальная оценка для определения наиболее эффективного способа. Экспериментально продемонстрировано, что применение

разработанного метода позволяет значительно увеличить число выявленных символьных условных переходов в анализируемой программе, что приводит к увеличению достижимого покрытия программ тестами.

Четвертая глава содержит описание деталей реализации предложенных методов в программных инструментах, разработанных в ИСП и экспериментальному исследованию производительности SMT-решателей, используемых в разработанных алгоритмах.

В заключении сформулированы основные результаты диссертационной работы.

Основные научные результаты и их значимость для науки и практики

Разработанные в диссертационной работе Куца Д.О. методы и алгоритмы поиска и моделирования косвенных переходов, моделирования чтений памяти по символьно вычисляемому адресу и программный инструмент, реализующий предложенные методы, обладают как теоретической, так и практической значимостью.

Значимость для науки этих результатов заключается в том, что предложенные методы обладают новизной и дополняют существующие методы анализа машинных кодов так, что позволяют выполнять более глубокий анализ путей выполнения программ после косвенных переходов и чтений памяти по символьному адресу.

Практическая значимость разработанных Куцем Д.О. методов, реализованных как расширение программного инструмента Sydr, продемонстрирована повышением точности анализа методом динамической символьной интерпретации и увеличением тестового покрытия исследуемых программ. Реализованные методы применяются в проектах ИСП РАН и в Центре доверенного искусственного интеллекта ИСП РАН, в составе инструмента Sydr внедрены и используются для безопасной разработки ПО в ООО «Код Безопасности».

Замечания

Экспериментальное исследование проведено на наборах существующих, хорошо отлаженных приложений и демонстрирует увеличение покрытия кода тестами. Предполагается, что это позволит порождать новые тесты, выявляющие ошибки, не обнаруженные ранее другими методами. Реализованные методы уже используются для промышленного тестирования расширенным инструментом Sydr с обнаружением ошибок. Однако не приводятся данные, сколько ошибок обнаружено новыми

методами в сравнении с числом ошибок, обнаруживаемых с помощью инструмента Sydr без данного расширения. Это дало бы более впечатляющую экспериментальную характеристику разработанных методов.

Заключение

Указанный недостаток не влияют на общую положительную оценку диссертационной работы Куца Даниила Олеговича. Диссертационная работа «Метод моделирования косвенной адресации в рамках динамической символьной интерпретации» является законченным научным исследованием по актуальной тематике, выполненной на высоком уровне. Название диссертации соответствует основному содержанию диссертации. Автореферат достаточно полно отражает содержание работы.

Работа удовлетворяет всем требованиям ВАК РФ, предъявляемым к работам на соискание степени кандидата технических наук, а ее автор, Куц Даниил Олегович, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Отзыв на диссертацию подготовлен на основании заключения структурного подразделения «Отдел программного обеспечения высокопроизводительных вычислительных систем и сетей» по результатам проведенного обсуждения диссертации и заслушивания доклада Д.О. Куца в ИПМ им. М.В. Келдыша РАН на заседании семинара «Программирование» им. М.Р. Шура-Бура 27 июня 2023 г.

Старший научный сотрудник
к.ф.-м.н. Ю.А. Климов

Сведения об организации:
Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»
Адрес: 125047, город Москва, Миусская пл., д. 4
Телефон: +7 499 978-13-14
E-mail: office@keldysh.ru
Веб-сайт: https://www.keldysh.ru