ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, члена-корреспондента РАН
Шабанова Бориса Михайловича
на диссертационную работу Куца Даниила Олеговича
по теме **«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»

Актуальность темы

В настоящее время электронно-вычислительные устройства и необходимое для них программное обеспечение широко используются в науке, промышленности, бизнесе, частной жизни. Развиваются различные аппаратные платформы, языки программирования, средства создания программ. С каждым годом увеличивается сложность создаваемых программ и объемы обрабатываемых данных. Конечно, распространение программного обеспечения приводит к актуализации вопроса о его безопасности.

Для тестирования программ и анализа их уязвимости необходимы инструменты моделирования входных данных, с помощью которых можно проверить все возможные ветви исполнения программы. Ввиду возрастающей сложности программ, разветвленности логики их исполнения и разнообразия обрабатываемых данных создание тестовых наборов, обеспечивающих адекватное покрытие исполняемого кода, становится сложной задачей. Создание тестовых наборов без привлечения средств автоматизации не представляется возможным.

Одним из распространенных способов тестирования программ путем автоматической генерации входных данных является фаззинг, с помощью которого на вход программе подаются случайно созданные элементы данных соответствующих типов. При этом полная генерация всех допустимых наборов данных невозможна из-за огромного количества комбинаций. Так как фаззинг не учитывает структуру и внутреннюю логику

программы, то с помощью него сложно добиться удовлетворительного покрытия.

Для увеличения доли покрываемого кода применяют динамическую символьную интерпретацию программы, использующую математическую модель исполнения программы. В этом случае входные данные заменяются символьными переменными, и инструкции программы интерпретируются с использованием этих символьных переменных. Это дает возможность вычислять условия переходов между базовыми блоками программы в символьном виде и инвертировать эти условия, увеличивая таким образом покрытие программного кода (путем вычисления входных данных, необходимых для изменения протекания потока управления программы по другому пути).

Основным препятствием для тестирования программы с помощью символьной интерпретации является наличие неявных зависимостей между операциями, а также косвенных переходов, то есть таких переходов, условие для которых вычисляется не явно, а считывается из некоторой области памяти (например, из таблицы переходов).

Поиск и моделирование косвенных переходов и косвенной адресации является актуальной задачей, так как это способно повысить покрытие программного кода тестируемой программы. Разработка новых подходов к моделированию косвенных переходов и косвенной адресации крайне актуально, так как данная задача является ресурсоемкой, и прямые методы ее решения способны существенно замедлить процесс символьной интерпретации.

Теоретическая и практическая значимость полученных результатов, их научная новизна

В работе предложены новые методы моделирования следующих видов неявных зависимостей, направленные на увеличение покрытия исполняемого кода:

1. Метод поиска и моделирования косвенных переходов. В качестве косвенных переходов рассматривается прежде всего исполняемый код, полученный из оператора ветвления switch. В таком случае, при достаточно большом количестве альтернатив ветвления оптимизирующий компилятор может создать переход по адресу, загруженному из ячейки памяти с

нужным смещением от базы таблицы переходов. Предложен метод детектирования возможного косвенного перехода и поиска границ таблицы переходов для таблиц двух видов: таблицы с абсолютными адресами и таблицы со смещениями адреса. Предложенный метод поиска границ таблицы переходов позволяет обнаружить альтернативные адреса перехода и инвертировать переход.

2. Метод моделирования чтения памяти по символьно вычисляемому адресу. Метод состоит из определения границ доступа к памяти (с помощью выбора фиксированного участка памяти, бинарного поиска границ с использованием SMT-решателя и синтаксического анализа символьного выражения адреса) и конструирования ограничения пути на основе определенных границ доступа к памяти. В качестве способов формирования ограничения пути рассматривались вложенные if-then-else деревья, двоичные деревья поиска и линеаризованные двоичные деревья поиска.

Теоретическая значимость работы состоит в разработке новых методов поиска и моделирования косвенных переходов и моделирования чтений по символьно вычисляемому адресу. Методы позволяют учитывать косвенную адресацию в символьной модели выполнения программы и обеспечивают более высокую точность динамической символьной интерпретации бинарных программ.

Практическая значимость полученных результатов подтверждается наличием четырех зарегистрированных программ для ЭВМ, реализующих предложенные методы и алгоритмы. Разработанные в данной диссертации методы реализованы в программных инструментах ИСП РАН и эксплуатируются в различных проектах и организациях.

Структура диссертации

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

Во Введении формулируются цели и задачи диссертационной работы, обосновывается актуальность темы исследования, теоретическая и практическая значимость работы, выделяется научная новизна и основные положения, выносимые на защиту.

В Главе 1 приводится обзор предметной области по теме диссертации. Рассматривается метод динамической символьной интерпретации, приводится его описание. Рассматриваются различные инструменты, реализующие динамическую символьную интерпретацию программ, в частности DART (в котором впервые было реализовано конкретно-символьное исполнение, или конколик), Mayhem (в котором был реализован гибридный режим динамического анализа, совмещающий в себе последовательное моделирование исполнения программы на разных наборах данных и одновременное моделирование нескольких путей исполнения), QSYM (реализующий гибридный подход, совмещающий фаззинг с символьной интерпретацией) и другие. Приводятся основные распространенные подходы к моделированию косвенной адресации в программах. Приводится описание инструмента Sydr разработки ИСП РАН для осуществления конкретно-символьной интерпретации.

Во Главе 2 предлагается новый метод поиска и моделирования косвенных переходов. Приводится описание косвенных переходов, их типичное представление в бинарном коде, а также конструкции языка, порождающие косвенные переходы. Приводится описание таблицы переходов с двумя вариантами представления адресов (абсолютные адреса и смещения от некоторого базового адреса). Описывается алгоритм поиска потенциальных косвенных переходов. По результату поиска потенциального косвенного перехода описывается метод поиска границ соответствующей таблицы переходов, а также конструирование символьных ограничений для ветвей переходов и добавление ограничений в предикат пути программы. В конце главы приводятся результаты экспериментальной проверки реализованного алгоритма. Результаты экспериментов показывают, что использование нового алгоритма позволяет находить новые пути исполнения программы, повышая, таким образом, ее покрытие.

В Главе 3 приводится описание нового метода моделирования чтений памяти по символьно вычисляемому адресу. Данный метод состоит из двух частей. Первой частью является определение области памяти, из которой производится чтение по символьно вычисляемому адресу. Для данной операции предлагается три подхода: выбор фиксированного участка памяти, бинарный поиск границ с использованием SMP-решателя и синтаксический

анализ символьного выражения адреса. Второй частью метода является символьная интерпретация операции чтения, которая заключается в построении соответствующего предиката пути. Для построения ограничений предлагается три подхода: вложенные if-then-else деревья, двоичные деревья поиска и линеаризованные двоичные деревья поиска. Для предложенного метода моделирования чтений памяти по символьно вычисляемому адресу были проведены эксперименты, продемонстрировавшие, что использование нового алгоритма приводит к открытию новых путей исполнения программы, которые были недоступны ранее.

Глава 4 посвящена программной реализации методов обнаружения и моделирования косвенных обращений, описанных в главах 2 и 3. Приводится описание инструмента динамической интерпретации Sydr, в рамках которого были реализованы результаты работы, а также особенности встраивания в него новых разработанных алгоритмов.

В заключении приведены основные результаты работы, а именно: разработан метод поиска и моделирования косвенных переходов, разработан метод моделирования чтения памяти по символьно вычисляемому адресу, разработанные методы реализованы в коде программного продукта Sydr.

В целом диссертация Д.О. Куца является законченным исследованием, представляет решение актуальных задач, объединенных общим подходом, обеспечивающим возможность моделирования косвенной адресации в рамках динамической символьной интерпретации.

Апробация результатов работы

Результаты, полученные в ходе диссертационной работы, представлены в четырех публикациях, удовлетворяющих требованиям п.11 Положения о присуждении ученых степеней и индексируемых в Web of Science и Scopus. Получено четыре Свидетельства о государственной регистрации программ для ЭВМ. Одна научная статья подготовлена и опубликована автором единолично, остальные – в соавторстве с научным руководителем и другими авторами. Основные результаты диссертационной работы и положений, выносимых на защиту, достаточно полно изложены в работах, опубликованных соискателем.

Основные результаты диссертационного исследования прошли апробацию на всероссийских и международных открытых конференциях: Открытая конференции ИСП РАН в 2017, 2019 и 2020 гг.; Международная конференции Ivannikov Memorial Workshop 2021.

Замечания

К содержанию работы могут быть сделаны следующие замечания:

1. При оценке разработанных методов автор производит сравнение только между двумя версиями программного инструмента Sydr, в составе которого были реализованы данные методы. В работе нет экспериментального сопоставления с существующими инструментами, приведенными в обзоре работ в Главе 1.

2. Общее замечание по описанию алгоритмов в работе. В диссертации упомянуто несколько алгоритмов, которые напрямую влияют на увеличение покрытия программы путем обработки косвенных обращений: обнаружение косвенных переходов, распознавание границ таблицы переходов, определение области памяти для считывания по символьно вычисляемому адресу. Все эти алгоритмы описаны в виде текста. Тогда как в формальном виде описан только алгоритм линеаризации (Алгоритм 1).

3. В списке литературы в позиции № 65 опечатка в адресе репозитория (в ссылке указано имя репозитория sydr_benchmark, тогда как должно быть sydr-benchmark). Из-за этого переход по ссылке в источнике невозможен.

Указанные замечания не являются критическими и не снижают научную и практическую ценность проведенных исследований.

Заключение

Содержание автореферата отражает суть диссертационной работы и позволяет достаточно ясно оценить основные полученные результаты и степень их обоснованности и достоверности.

Диссертационная работа Куца Даниила Олеговича по теме «Метод моделирования косвенной адресации в рамках динамической символьной интерпретации» является законченным научным исследованием, основное содержание диссертации отражено в опубликованных статьях и обсуждено на научных конференциях.

Диссертационная работа отвечает требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, ее содержание соответствует паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», а ее автор, Куц Даниил Олегович, заслуживает присуждения ученой степени кандидата технических наук по вышеуказанной специальности.

Официальный оппонент,
член-корр. РАН, д.т.н., доцент,
директор Межведомственного суперкомпьютерного
центра РАН (МСЦ РАН),
заместитель директора по научной работе
Федерального государственного учреждения
«Федеральный научный центр
Научно-исследовательский институт
системных исследований
Российской академии наук» Б.М.Шабанов

Подпись Шабанова Б.М. удостоверяю
Начальник отдела кадров МСЦ РАН В.В. Шишкина
«3» сентября 2023 г.