

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

Шабанова Бориса Михайловича

на диссертационную работу Логуновой Влады Игоревны
**«Разработка методов гибридного фаззинга для приложений
процессорных архитектур Байкал-М и RISC-V 64»,**

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

На данный момент технологическая инфраструктура в виде вычислительных устройств и необходимого программного обеспечения (ПО) имеет фундаментальное значение в науке, промышленности, бизнес-процессах и повседневной жизни. Развиваются различные аппаратные платформы, языки программирования и инструменты для разработки программ. Сложность разрабатываемых программ неуклонно растет, как и объем обрабатываемых ими данных. Соответственно, широкое распространение ПО напоминает об актуальности вопросов безопасности его применения.

При анализе и тестировании программ необходимы средства моделирования входных данных, не ограниченных фиксированным числом предопределенных тестовых вариантов. Чтобы увеличить тестовое покрытие возможных путей исполнения программы применяются автоматизированные инструменты. Наиболее распространенным способом тестирования программ с автоматической генерацией входных данных является фаззинг, с помощью которого на вход программе подаются случайно созданные входные данные соответствующих типов. Однако перебор всех допустимых наборов данных невозможен из-за огромного количества комбинаций. Поскольку фаззинг не учитывает структуру и внутреннюю логику программы, то с помощью него сложно протестировать маловероятные пути исполнения, реализующиеся лишь для малой части возможных входных данных.

С целью увеличения доли покрываемого кода за счет таких применяется динамическая символьная интерпретация программы с конструированием математической модели исполнения программы. В этом случае входные данные заменяются символьными переменными, а инструкции программы интерпретируются как преобразования над этими переменными. Такой вид анализа дает больше точности и позволяет находить редкие пути выполнения, поскольку позволяет использовать значительно больше информации о программе на этапе инструментации. Поэтому его применение вместе с фаззинг-тестированием приводит в росту покрытия кода и называется гибридным фаззингом. Целесообразность применения символьного анализа для гибридного фаззинга требует от интерпретатора высокой производительности, наличия современных оптимизаций и методов моделирования как, например, разрешение косвенной адресации переходов.

Одним из важных аспектов для промышленного применения методов динамической символьной интерпретации является возможность анализа исполняемых программ, не требующего наличия исходного кода. Большинство

существующих символьных инструментов либо позволяет работать с бинарным кодом программ только для архитектуры x86, либо используют исходный код. Разработка новых подходов к моделированию преобразований машинных инструкций процессорных архитектур Байкал-М и RISC-V является **актуальной**, так как такие подходы позволяют включить в область применимости гибридного фаззинга программное обеспечение данных аппаратных платформ, особенно учитывая то, что эти платформы будут использованы для отечественных доверенных программно-аппаратных решений.

Теоретическая и практическая значимость полученных результатов, их научная новизна

В работе предложены новые методы моделирования зависимостей выполнения программы от символьных данных для архитектур RISC-V и AArch64:

1. Метод символьной интерпретации набора целочисленных инструкций архитектуры, RISC-V, включающего сокращенные инструкции и псевдоинструкции. Разработанный метод реализует символьную семантику инструкций, посредством конструирования SMT-формулы, описывающей преобразование символьного контекста на уровне выполнения отдельной инструкции RISC-V.

2. Методы динамической символьной интерпретации бинарного кода архитектур Байкал-М (AArch64) и RISC-V 64. Методы предоставляют возможность точного определения границ для множественных условных переходов, что позволяет выявлять соответствующие целевые адреса исполняемого кода в процессе динамической символьной интерпретации и ускорять исследование новых путей выполнения при применении разработанных методов в контексте гибридного фаззинг-тестирования.

Теоретическая значимость работы заключается в разработанных методах динамической символьной интерпретации бинарного кода процессорных архитектур Байкал-М (AArch64) и RISC-V 64, а также методе символьной интерпретации набора целочисленных инструкций RISC-V на основе конструирования абстрактных синтаксических деревьев их операционной семантики. Для вышеуказанных архитектур была проведена экспериментальная оценка эффективности применения методов динамической символьной интерпретации в контексте фаззинг-тестирования.

Практическая значимость работы заключается в том, что разработанные в данной диссертации методы динамической символьной интерпретации бинарного кода архитектур Байкал-М (AArch64) и RISC-V 64 реализованы в программных инструментах ИСП РАН и эксплуатируются в различных проектах и организациях. Метод символьной интерпретации набора целочисленных инструкций архитектуры RISC-V, включающего сокращенные инструкции и псевдоинструкции, был интегрирован в открытую символьную библиотеку Triton, благодаря чему он доступен широкому кругу разработчиков.

Структура диссертации

Диссертация состоит из Введения, пяти глав, Заключения и списка литературы из 91 наименования. Общий объем диссертации 117 страниц.

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 (впервые было реализовано конкретно-символьное исполнение, или конколлик), SAGE (символьный анализ на уровне бинарного кода для архитектуры x86), QSym (гибридный подход, совмещающий параллельную работу фаззинга с символьной интерпретацией) и другие. Приводятся основные распространенные подходы к обеспечению переносимости при символьном анализе программ, предназначенных для различных платформ. Приводится описание разработанного в ИСП РАН инструмента Sydg позволяющего осуществлять конкретно-символьную интерпретацию.

В Главе 2 предлагается метод динамической символьной интерпретации бинарного кода программ архитектуры Байкал-М (AArch64). Приводится описание архитектурно-зависимых компонентов, необходимых для проведения конкретно-символьного анализа данной архитектуры. Приводится описание косвенных переходов, их типичное представление в бинарном коде программ, скомпилированных для архитектуры Байкал-М (AArch64). При выявлении потенциальных косвенных переходов описывается способ поиска границ соответствующей таблицы переходов, а также способ коррекции символьных адресных выражений в случае встроенных сдвигов арифметических инструкций и добавление полученных ограничений в предикаты путей программы.

В Главе 3 приводится описание метода символьной интерпретации набора целочисленных инструкций архитектуры RISC-V, включающего сокращенные инструкции и псевдоинструкции. Приводится описание набора архитектурных расширений, обрабатываемых предлагаемым методом. Приводится описание модуля разрешения архитектурных зависимостей открытой символьной библиотеки Triton, в рамках которого были реализованы конструкторы операционной семантики инструкций архитектуры RISC-V в виде абстрактных древовидных выражений, а также раскрываются особенности встраивания в данный модуль разработанного метода. Описываются аспекты моделирования сокращенных инструкций и псевдоинструкций. Приводится пример символьной эмуляции на основе разработанного метода для решения задачи захвата флага.

Глава 4 посвящена разработанному новому методу динамической символьной интерпретации бинарного кода программ архитектуры RISC-V 64, данный метод опирается на метод, приведенный в Главе 3, с помощью которого выполняется конструирование символьной семантики отдельной инструкции. Приводится описание архитектурно-зависимых компонентов, требуемых для моделирования символьных зависимостей при конкретно-символьном анализе программ архитектуры RISC-V 64. Приводится описание механизмов работы системы динамической инструментации DynamoRIO и их адаптации для работы с вызовами функций в бинарном коде архитектуры RISC-V 64. Описываются

ограничения при анализе косвенных переходов в бинарном коде данной архитектуры и алгоритм распознавания таких переходов в условиях частичной информации о наборе инструкций базового блока, поступившего на исполнение, а также метод определения размеров соответствующей таблицы.

В главе 5 приводится экспериментальная оценка реализации разработанных методов динамической символьной интерпретации в рамках инструмента Sydr. Оценка для каждой архитектуры включает сравнение при проведении динамической символьной интерпретации как самостоятельного вида анализа и при ее использовании вместе с инструментом фаззинг-тестирования в контексте гибридного фаззинга. Результаты экспериментов показывают, что использование разработанных методов позволяет находить новые пути исполнения программы, повышая, тем самым, покрытие.

В Заключении приведены основные результаты работы, а именно: разработаны методы динамической символьной интерпретации бинарного кода программ архитектуры Байкал-М и архитектуры RISC-V 64, реализованы алгоритмы определения возможных направлений косвенных переходов для бинарного кода вышеуказанных архитектур, разработан метод символьной интерпретации набора целочисленных инструкций архитектуры RISC-V, включающего сокращенные инструкции и псевдоинструкции, разработанные методы реализованы в инструментах Triton и Sydr, проведена оценка их эффективности.

Апробация результатов работы

Основные результаты диссертационного исследования прошли апробацию на профильных конференциях: X и XII научно-практические конференции OS DAY в Москве в 2023 и 2025 гг. и Международная конференция «Иванниковские Чтения» в Иркутске в 2025 г.

По теме диссертации автором опубликовано, пять работ, индексируемых в изданиях из перечня ВАК (две публикации) и Web of Science и Scopus (три публикации), что удовлетворяет требованиям п. 11 Положения о присуждении ученых степеней. Получено одно свидетельство о государственной регистрации программы для ЭВМ. Одна научная статья подготовлена и опубликована автором единолично, остальные – в соавторстве.

Основные результаты диссертационной работы и положений, выносимых на защиту, достаточно полно изложены в работах, опубликованных соискателем.

Замечания

1. В Главе 1 утверждается, что производительность символьной эмуляции сокращает возможности её практического применения для реальных приложений и в гибридном фаззинге, однако в самой работе не приводится оценка производительности данного подхода или ссылки на соответствующие оценки в работах других авторов.

2. В текстовом описании работы алгоритма сканирования базового блока для распознавания табличного перехода для архитектуры RISC-V 64 не хватает

указаний на названия соответствующих методов его формального представления (Алгоритм 1, стр. 80,81).

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

Заключение

Автореферат правильно отражает содержание диссертации и позволяет достаточно точно оценить основные полученные результаты, степень их обоснованности и достоверности.

Диссертационная работа В. И. Логуновой по теме «Разработка методов гибридного фаззинга для приложений процессорных архитектур Байкал-М и RISC-V 64» 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, Логунова Влада Игоревна, заслуживает присуждения ученой степени кандидата технических наук по вышеуказанной специальности.

Официальный оппонент,
доктор технических наук,
член-корреспондент РАН,
руководитель отделения
Федерального государственного
бюджетного учреждения
«Национальный исследовательский
центр «Курчатовский институт»

Шабанов Борис Михайлович

«25» ноября 2025 г.