

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Батузова Кирилла Андреевича

«Исследование и разработка методов оптимизации программ для систем динамической двоичной трансляции»,

представленную к защите на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность темы исследования

Технология динамической двоичной трансляции – это единственный способ эффективно (в отличие от интерпретации) исполнять двоичный код программного обеспечения одной процессорной архитектуры на другой процессорной архитектуре, отличающейся от первоначальной. При проведении оптимизации кода программ во время динамической двоичной трансляции необходимо учитывать не только эффективность полученного кода, но и время, затраченное на компиляцию и оптимизации. Диссертационная работа Батузова К.А. посвящена *быстрым алгоритмам* машинно-независимых оптимизаций, распределения регистров и трансляции векторных инструкций и их реализации в системе динамической двоичной трансляции QEMU, что определяет ее *актуальность*.

Структура работы

Диссертация состоит из введения, пяти глав и заключения. Полный объем диссертации составляет 105 страниц. Список литературы содержит 60 наименований.

Во **введении** определяется цель работы, обосновывается её актуальность, формулируются основные научные результаты.

В **первой главе** приводится обзор предметной области, дается обзор систем динамической двоичной трансляции, алгоритмов оптимизации программ и распределения регистров. Поскольку в работе сделан упор на создании предельно быстрых алгоритмах оптимизации, многие вычислительно сложные алгоритмы исключены из предложенной реализации. Проведенный анализ показывает, что стандартный метод решения задачи анализа потока данных, используемый при проведении машинно-независимых оптимизаций, не учитывает особенности графа потока управления, получаемого при динамической двоичной трансляции. Распространенные алгоритмы распределения регистров требуют построения дополнительного промежуточного представления с частично распределенными

регистрами, что является нежелательным во время динамической двоичной трансляции. Дополнительный положительный эффект на производительность целевого кода оказывает возможность эффективно преобразовывать гостевые векторные операции в целевые векторные операции. Для проведения экспериментов был выбран полносистемный эмулятор QEMU, так как он реализует наиболее общий случай динамической двоичной трансляции и обладает открытым исходным кодом.

Во **второй главе** описывается *метод проведения машинно-независимых оптимизаций* во время динамической двоичной трансляции. Основным способом проведения машинно-независимых оптимизаций является сведение их к задаче анализа потока данных. В случае произвольного графа потока управления задача анализа потока данных может быть решена итеративным алгоритмом. Для ациклических графов потока управления, возникающих во время динамической двоичной трансляции, можно выбрать такой порядок просмотра базовых блоков, что окончательное решение будет найдено за одну итерацию стандартного алгоритма. В работе данное утверждение формулируется в виде теоремы и приводится его доказательство. Используя полученный таким способом однопроходный алгоритм решения задачи анализа потока данных, в работе строятся алгоритмы продвижения констант и копий и сворачивания константных выражений. На большинстве реальных задач из пакета SPECint2000 демонстрируется незначительный прирост производительности с использованием предложенных в работе оптимизаций

В **третьей главе** описывается *однопроходный алгоритм локального распределения регистров*. Описанный алгоритм основан на эвристическом алгоритме «первый дальний» (англ. Furthest-First), выбирающем для освобождения в случае нехватки регистр с наиболее дальним использованием. Изложенный в работе *новый* алгоритм отличается тем, что он также отслеживает те регистры, содержимое которых будет неизбежно сохранено в память до его последующего использования, и освобождает такие регистры как можно раньше. Результаты использования предложенного алгоритма сопоставлены с используемым в QEMU алгоритмом распределения регистров и свидетельствуют о перспективности его объединения с глобальным распределением регистров.

В **четвёртой главе** описывается *однопроходный алгоритм глобального распределения регистров*. Данный алгоритм сначала выбирает регистры, хранящие значения переменных на границах базовых блоков, а затем в этих ограничениях производит локальное распределение регистров внутри базовых блоков. В работе формулируются и доказываются теоремы, которые подтверждают корректность и эффективность описанного алгоритма. Результаты использования данного *нового* алгоритма в сочетании с новым алгоритмом локального распределения регистров не дают заметного прироста производительности целевого кода, что объясняется

слишком малыми регионами оптимизации в QEMU с малым числом базовых блоков, в результате чего алгоритм глобального распределения регистров применяется только к 10% регионов оптимизации.

В пятой главе описывается *метод трансляции векторных инструкций* одной процессорной архитектуры в векторные инструкции другой процессорной архитектуры. Кроме совпадения семантики исходной и целевой инструкций, необходимо также обеспечить корректное обращение к частям векторного регистра. Во-первых, векторные регистры могут перекрываться, и изменение значения одного из них ведет к изменению значения другого. Во-вторых, в случае, когда выразить векторную инструкцию с помощью других векторных инструкций не удалось, она эмулируется последовательностью скалярных инструкций, которые обращаются к элементам вектора по одному. В работе формулируются ограничения, при которых данная задача может быть эффективно решена, и строится консервативный алгоритм, который дает точное решение в случае выполнения сформулированных ограничений и безопасную аппроксимацию в случае их нарушения. В этой главе продемонстрированы наиболее заметные положительные результаты оптимизации целевого кода.

В **заключении** приводятся основные результаты диссертационной работы, которые выносятся на защиту:

- Разработан быстрый метод решения задачи анализа потока данных для ациклических графов потока управления.
- Разработан новый однопроходный алгоритм локального распределения регистров.
- Разработан новый однопроходный алгоритм глобального распределения регистров.
- Разработан эффективный метод выражения векторных инструкций одной процессорной архитектуры через векторные инструкции другой процессорной архитектуры.

Практическая значимость диссертации заключается в том, что все предложенные методы и алгоритмы были реализованы в эмуляторе с открытым исходным кодом QEMU и протестированы на примерах реальных программ. Проведенные эксперименты демонстрируют повышение производительности и снижение количества сохранений значений регистров в память. Для предложенных методов сформулированы и доказаны *теоремы о корректности и скорости сходимости*, что подтверждает *достоверность и обоснованность результатов работы*.

В работе присутствуют следующие недостатки:

1. В результатах экспериментов в работе приводится только общее время работы эмулятора. Разбиение его на время работы гостевого кода и время работы транслятора дало бы более подробную информацию о поведении системы.
2. В третьей главе приводятся только результаты профилирования, показывающие, что количество сохранений значений регистров в память уменьшилось, но не приводятся измерения того, как это повлияло на скорость работы эмулятора.
3. Приведенная в первой главе формальная модель распределения регистров явным образом нигде далее в работе не используется.

В целом указанные замечания не влияют на положительную оценку работы.

Автореферат полно и правильно отражает содержание диссертационной работы.

Диссертационная работа соответствует всем требованиям ВАК РФ, предъявляемых к диссертациям на соискание учёной степени кандидата физико-математических наук, а её автор, Батузов Кирилл Андреевич, заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Кандидат технических наук, старший научный сотрудник, начальник отделения «Системы программирования» публичного акционерного общества «ИНЭУМ им. И.С. Брука»

В.Ю. Волконский

Подпись кандидата технических наук Волконского В.Ю. заверяю, зам. генерального директора «ИНЭУМ им. И.С. Брука»

В.М. Фельдман

«25» 04 2018 г.