ОТЗЫВ

официального оппонента на диссертационную работу

Бородина Алексея Евгеньевича

«Межпроцедурный контекстно-чувствительный статический анализ для поиска ошибок в исходном коде программ на языках Си и Си++»,

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

Диссертационная работа Бородина А. Е., выполненная в Институте системного программирования РАН, посвящена разработке методов статического анализа для поиска ошибок для использования в жизненном цикле разработки программного обеспечения. Поиск ошибок с помощью статического анализа имеет ряд преимуществ: одновременный анализ многих путей выполнения, диагностика места ошибки и поиск ошибок на редко исполняемых путях, которые плохо покрываются при тестировании и динамическом анализе. Вместе с тем статические анализаторы выдают ложные срабатывания, т. е. предупреждения, которые описывают ситуации в исходном коде, не являющиеся ошибочными. Разработка статического анализатора является поиском компромисса между количеством пропущенных ошибок, уровнем ложных срабатываний и скоростью анализа. В работе рассматривается задача построения анализатора с высокой скоростью, достаточной для анализа проектов в миллионы строк кода за 6 часов, и с низким уровнем ложных срабатываний (менее 30%). К анализатору не предъявляется требования поиска всех возможных ошибок. Разработанные алгоритмы предполагается реализовать в инструменте статического анализа Svace, разрабатываемого в ИСП РАН, с целью улучшения существующего анализатора для языков Си и Си++.

Диссертация состоит из введения, шести глав и заключения. Объем диссертации составляет 137 страниц, включая 12 таблиц и 26 рисунков.

Во введении обосновывается актуальность работы, формулируются цели работы, приводятся выносимые на защиту положения.

Первая глава состоит из шести разделов. В первом разделе описываются особенности использования статических анализаторов для поиска ошибок в жизненном цикле разработки программного обеспечения, приводятся требования к анализатору для такого использования. Во втором разделе описывается различие между ошибками времени выполнения программы и

ошибками в исходном коде. Третий раздел содержит описывает возможность отказа от корректности анализа при поиске ошибок. В четвертом разделе рассматривается синтаксический анализ на примере анализатора, реализованного в инструменте Svace. Пятый раздел содержит обзор существующих семантических анализаторов. Все рассмотренные анализаторы используют анализ на основе резюме. Рассмотрены как коммерческие анализаторы, так и анализаторы, разработанные для научных целей. В шестом разделе описывается инструмент Svace, рассматривается задача улучшения существующего анализатора семантического уровня для поиска ошибок в исходном коде на языках Си и Си++. Целью улучшения является возможность анализа проектов размером в миллионы строк кода при высоком уровне истинных срабатываний.

Во второй главе описывается язык svace0, являющийся внутренним представлением анализатора. Приводится структурная операционная семантика языка и пример простой функции на языке Си и языке svace0.

В третьей главе описывается внутрипроцедурный анализ. Рассматривается задача анализа только для некоторых состояний на входе в функцию: такие состояния, где отсутствуют алиасы среди переменных и ячеек памяти. Анализ является модификацией символьного исполнения, при котором в точках слияния путей объединяются состояния для разных путей. Для повышения точности анализа предложен механизм нумерации значений, доступных через указатели.

Четвертая глава посвящена детекторам для поиска ошибок. Анализ разделен на ядро и расширения. Ядро анализа выполняет общие действия, такие как анализ указателей, нумерация значений. В расширениях выполняется анализ интересующих свойств программы и реализация детекторов. В главе приводится описание критерия выдачи предупреждений и описываются несколько детекторов для поиска нулевых указателей разной сложности. Более сложные детекторы позволяют найти большее количество ошибок.

Пятая глава посвящена межпроцедурному анализу. В главе описывается общая схема анализа на основе резюме, алгоритмы создания и трансляции резюме в контексты вызова функции. В конце главы описывается расширение анализа для поиска несоответствий между конструкторами и деструкторами классов в языке Си++. Описываются преимущества такого подхода по сравнению с общим анализом.

В шестой главе описывается реализация алгоритмов в инструменте Svace, и приводится оценка скорости и качества анализа. Анализ производится для биткода LLVM, который генерируется компилятором Clang. При этом возникает проблема соответствия биткода LLVM исходному коду. Для решения этой проблемы был выполнен ряд модификаций компилятора

Clang. Оценка скорости анализа производится для 37 проектах с открытым исходным кодом, среди которых 5 проектов имеют размер больше 5 миллионов строк кода. Анализ каждого проекта не превышает 5 часов, таким образом, возможно его использование во время ночной сборки. Оценка выданных предупреждений производилась для ОС Tizen и Android 5.0.2. Оценивались предупреждений для множества детекторов, среди которых детекторы поиска утечек памяти и ресурсов, разыменований нулевых указателей, доступа к массивам с неправильной проверкой индекса, некорректного освобождения памяти, несоответствий конструкторов и деструкторов. Для большинства детекторов более 70% предупреждений являются истинными.

Заключение диссертации содержит основные результаты работы: разработан алгоритм внутрипроцедурного анализа функции; на его основе разработан межпроцедурный алгоритм анализа; алгоритмы реализованы в подсистеме инструмента Svace для поиска ошибок в программах на языках Си и Си++.

Особо следует отметить математические результаты диссертации, представленные теоремами в разделах 3.5 и 4.8 диссертации. Теоремы 1 и 2 третьей главы диссертации показывают корректность проводимого внутрипроцедурного анализа в рамках сделанных автором работы предположений. Теорема 3 диссертации (раздел 4.8) предлагает способ упрощения формул логики высказываний, используемых для обеспечения чувствительности к путям в ходе анализа. Такое упрощение является существенным условием достижения масштабируемости межпроцедурного анализа при анализе больших программ.

В работе следует отметить следующие недостатки:

1. Недостаточно уделяется внимания описанию особенностей анализа специфических для языка Си++ конструкций (отличных от языка Си), например, исключениям, шаблонным функциям;
2. В четвертой главе приводится описание реализации детекторов для поиска разыменований нулевых указателей разной степени сложности: используя прямой анализ, используя прямой и обратный анализ, с чувствительностью к путям. Не хватает сравнительного анализа приведенных методов, выявляющего их достоинства и недостатки, области применимости и т.п.;
3. В разделе 6.5 приведены замеры времени и скорости анализа, согласно которым анализ для проектов на языке Си выполняется быстрее, чем для проектов на языке Си++. Желательно пояснить также возможные причины такого поведения анализа.

Отмеченные недостатки не влияют на общую положительную оценку работы.

Диссертационная работа является завершенным научным исследованием, в которой автору удалось решить поставленные перед диссертационным исследованием задачи и разработать инструмент статического анализа для поиска ошибок в программах на языках Си и Си++. Результаты диссертации представлены в статьях автора в журналах и докладывались на российских и международных конференциях. Автореферат правильно и полно отражает содержание диссертации и оформлен надлежащим образом.

Диссертационная работа отвечает требованиям ВАК РФ, предъявляемым к кандидатским диссертациям по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, а ее автор, Бородин Алексей Евгеньевич, заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.11.

Официальный оппонент, д.ф.-м.н., профессор Терехов Андрей Николаевич

«19» мая 2016 года

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»,

199034, Санкт-Петербург, Университетская наб. д.7-9.

+7 (812) 328-20-00, spbu@spbu.ru, http://spbu.ru

Подпись д.ф.-м.н. А.Н. Терехова удостоверяю.