ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Дудиной Ирины Александровны

«Поиск ошибок переполнения буфера в исходном коде программ с помощью символьного выполнения»,

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

Актуальность темы

Повышение сложности и размеров программ и ограничения времени, отводимого на их разработку, неизбежно приводят к ошибкам в программах. К числу ошибок, обнаружение которых на стадии разработки программы представляет особенный практический интерес, относятся ошибки переполнения буфера. Ошибки переполнения буфера не только приводят к непредсказуемому поведению программы, но и являются источником серьезных уязвимостей, эксплуатируемых злоумышленниками. Архитектуры современных вычислительных машин не позволяют выявлять подобные ошибки аппаратными средствами, поэтому приходится использовать программные средства.

Диссертационная работа Дудиной И. А. посвящена *актуальной* задаче – разработке статических методов обнаружения ошибок переполнения буфера в программах на языке С, удовлетворяющих следующим требованиям:

- масштабируемости, т.е. возможности анализа программ большого размера – до нескольких млн. строк исходного кода,
- качеству анализа, т.е. хорошему покрытию ошибочных ситуаций и небольшому количеству ложных срабатываний.

Структура работы

Диссертационная работа Дудиной И. А. состоит из введения, семи глав, заключения и одного приложения. Полный объем диссертации составляет 145 страниц. Список литературы содержит 76 наименований.

Во введении обосновывается актуальность темы исследования, цель и задачи, определяется научная новизна и основные положения работы, выносимые на защиту, а именно:

- определение ошибочной ситуации, используемое для формального построения

алгоритма обнаружения переполнения буфера;

- метод внутрипроцедурного чувствительного к путям анализа для поиска переполнения буфера известного в момент компиляции размера при помощи символьного выполнения и основанный на нём метод контекстно-чувствительного межпроцедурного анализа, использующий резюме;
- два подхода к поиску переполнения буфера произвольного размера;
- методы поиска переполнения буфера при работе со строками и обработке данных, полученных из недоверенных источников.

В первой главе приводится широкий обзор существующих подходов к обнаружению переполнения буфера в исходном коде, проводится глубокий анализ требований к современному детектору переполнения буфера на основе изучения популярных тестовых наборов и известных уязвимостей из реальных проектов. На основании проведенного анализа требований делается важный вывод о необходимости использования чувствительного к путям и к контексту анализа для повышения его качества.

Во второй главе вводится и обосновывается выбор формального определения ошибки переполнения буфера, которое используется для построения чувствительного к путям анализа.

Третья глава начинается описанием модельного языка, для программ на котором формально излагается алгоритм поиска ошибок. Далее рассматривается алгоритм *внутрипроцедурного* анализа для поиска переполнения буфера известного в момент компиляции размера. Приводится общее описание абстрактного состояния анализа и алгоритма его продвижения по графу потока управления. Излагается подход к построению достаточных условий переполнения буфера на основании определения ошибки, сформулированного во второй главе, сводящийся к проверке на выполнимость свободной от кванторов формулы теории битовых векторов, которая может быть эффективно выполнена с помощью SMT-решателя. Для инструкций модельного языка приводятся передаточные функции, преобразующие абстрактное состояние. Доказывается теорема о корректности построенных достаточных условий ошибки для функций, удовлетворяющих ряду условий.

Четвёртая глава посвящена построению *межпроцедурного* анализа с помощью метода резюме на основании алгоритма внутрипроцедурного анализа, предложенного в третьей главе. Рассматриваются подходы для построения достаточных условий ошибки для случаев, когда вычисление индекса располагается в нескольких различных функциях, и когда происходит обращение к буферу, определённому в другой функции.

Доказана теорема о корректности построенных достаточных условий ошибки для межпроцедурного случая.

В пятой главе описываются расширения рассмотренного алгоритма анализа, позволяющие осуществлять поиск переполнения буфера, возникающего при обработке данных, полученных из недоверенного источника, и переполнений при работе со строками. Это важная часть работы, позволяющая покрыть значительный класс ошибок переполнения буфера.

В шестой главе приводятся два подхода к поиску переполнения буфера произвольного размера. Первый подход заключается в применении метода анализа возможных значений индекса для анализа значений размера выделенного буфера. Второй подход основывается на проверке определения из второй главы как формулы с кванторами в теории битовых векторов напрямую с помощью SMT-решателя.

Седьмая глава посвящена описанию реализации детекторов, результатов тестирования на реальных проектах и тестовых наборах и сравнению с инструментом статического анализа Infer.

В заключении формулируются основные полученные результаты исследования и направления для дальнейшей работы.

Оценка научной новизны и достоверности

Новые научные результаты исследования, предложенные автором, заключаются в следующем:

- предложено формальное определение ситуации переполнения буфера, учитывающее возможный контракт функции в отсутствие явной информации о контракте от программиста,
- на основе предложенного определения разработан масштабируемый алгоритм межпроцедурного контекстно-чувствительного анализа для поиска переполнения буфера известного в момент компиляции размера, различающий пути выполнения,
- разработаны методы обнаружения переполнения буфера произвольного размера: на основе алгоритма для буферов константного размера и путём проверки по определению напрямую,
- предложены расширения базового алгоритма для поиска переполнений буфера при работе со строками и данными из недоверенного источника.

Достоверность результатов обосновывается корректностью представленных в работе математических доказательств, успешной реализацией предложенных

алгоритмов в промышленном инструменте статического анализа, согласованностью результатов с другими подходами, приведёнными в библиографическом обзоре. Результаты диссертационного исследования докладывались на семи российских и международных конференциях, опубликованы в 8 печатных работах (в том числе 6 работ в журналах из перечня ВАК).

Значимость для науки и практики результатов диссертационного исследования

Теоретическая значимость результатов работы заключается в предложенных алгоритмах межпроцедурного чувствительного к контексту и путям анализа программ на языке С для обнаружения переполнения буфера, масштабируемых на программные системы из миллионов строк исходного кода.

Практическая значимость результатов исследования состоит в том, что реализованные в рамках статического анализатора Svace детекторы переполнения буфера используются для проверки реальных программных систем и позволяют обнаруживать ошибки данного типа на стадии разработки.

Недостатки и замечания по диссертационной работе

- нет явных оценок алгоритмической сложности представленных в работе алгоритмов, хотя приведенные данные о времени работы анализатора свидетельствуют о допустимой сложности;
- небольшое несоответствие: при описании основных характеристик работы говорится, что инструмент Svace не содержит модуль поиска переполнений буфера, чувствительного к путям, и поэтому пропускает много ошибок, а в главе 7 рассматривается использование разработанного в рамках диссертационной работы модуля как уже включенного в инструмент Svace;
- в разделе 6.1 диссертации недостаточно подробно изложен алгоритм межпроцедурного анализа для поиска переполнения буферов произвольного размера;
- в главе 7 дано сравнение результатов предложенного в работе анализатора с наиболее интересным статическим анализатором Infer, но не объясняется, за счет чего предложенный в работе метод обладает столь заметным преимуществом в части обнаружения ошибок переполнения буфера;
- в разделах 7.1.2 и 7.3 приводятся причины возникновения ложных срабатываний и методы их подавления, однако отсутствует анализ причин пропусков реальных ошибок;

- используемый при анализе результатов термин «варианты потока» (flow variant) не достаточно конкретизирован, поэтому не совсем понятно, как он соотносится с приведенным в приложении термином «варианты ошибки» (functional variant, flow type).

Отмеченные недостатки не снижают высокий научно-технический уровень работы и не влияют на общую высокую положительную оценку представленной диссертации.

Соответствие содержания автореферата и диссертации

Автореферат полно и правильно отражает содержание диссертационной работы.

Заключение

Диссертационное исследование Дудиной И. А. является законченной научной работой, обладает актуальностью, научной новизной и практической значимостью, проведено на высоком научно-техническом уровне.

Диссертация соответствует всем требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а её автор, Дудина Ирина Александровна, заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент
кандидат технических наук, начальник отделения «Системы программирования» Публичного акционерного общества «Институт электронных управляющих машин им. И. С. Брука», Российская федерация, 119334, Москва, ул. Вавилова, д. 24

тел. +7 (499) 135-89-49

адрес электронной почты: ineum@ineum.ru

В.Ю. Волконский

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

В.М. Фельдман

26 апреля 2019 г.